

Identification of CT46/HORMAD1, an immunogenic cancer/testis antigen encoding a putative meiosis-related protein

Yao-Tseng Chen¹, Charis A. Venditti², Gregory Theiler³, Brian J. Stevenson³, Christian Iseli³, Ali O. Gure², C. Victor Jongeneel³, Lloyd J. Old², and Andrew J. G. Simpson²

¹Weill Medical College of Cornell University, New York, NY, USA

²Ludwig Institute for Cancer Research, New York Branch, New York, NY, USA

³Office of Information Technology, Ludwig Institute for Cancer Research, and Swiss Institute of Bioinformatics, Lausanne, Switzerland

Keywords: human, ESTs, cancer/testis, tumor antigens, HORMAD1, mRNA, tissue distribution

Transcripts with ESTs derived exclusively or predominantly from testis, and not from other normal tissues, are likely to be products of genes with testis-restricted expression, and are thus potential cancer/testis (CT) antigen genes. A list of 371 genes with such characteristics was compiled by analyzing publicly available EST databases. RT-PCR analysis of normal and tumor tissues was performed to validate an initial selection of 20 of these genes. Several new CT and CT-like genes were identified. One of these, *CT46/HORMAD1*, is expressed strongly in testis and weakly in placenta; the highest level of expression in other tissues is <1% of testicular expression. The *CT46/HORMAD1* gene was expressed in 31% (34/109) of the carcinomas examined, with 11% (12/109) showing expression levels >10% of the testicular level of expression. *CT46/HORMAD1* is a single-copy gene on chromosome 1q21.3, encoding a putative protein of 394 aa. Conserved protein domain analysis identified a HORMA domain involved in chromatin binding. The CT46/HORMAD1 protein was found to be homologous to the prototype HORMA domain-containing protein, Hop1, a yeast meiosis-specific protein, as well as to asy1, a meiotic synaptic mutant protein in *Arabidopsis thaliana*.

Introduction

CT antigens represent potential cancer vaccine targets for a wide range of tumor types (1). The *MAGE*, *BAGE*, and *GAGE* gene families were the first CT antigens to be identified. This was achieved on the basis of the autologous CD8+ T cell responses they elicited in cancer patients (2). Subsequently, it was recognized that CT antigens also elicit antibody responses in cancer patients, and a further series of CT antigen genes was identified by SEREX (serological analysis of recombinant cDNA expression libraries of human tumors) (3, 4). The SEREX-defined CT antigens include the *SSX* gene family, *SCP1*, *NY-ESO-1*, *CT7*, *HOM-TES-85*, *CAGE*, *CAGE1*, and *NY-SAR-35* (5). More recently, CT antigens have also been sought by identifying genes with restricted CT mRNA expression pattern, regardless of their immunogenicity. This strategy has resulted in the identification of *LAGE-1*, *CT9*, *CT10*, and *SAGE* by representational difference analysis, and several other CT antigen genes, including *CT15*, *CT16*, *FATE*, and *TPTE*, by EST database mining (1). In the present study, we continued our search for new CT antigens by analyzing EST databases for genes with predominant expression in testis, and then evaluating their expression in tumors by RT-PCR. Of the 20 CT candidate genes analyzed, we identified *CT46/HORMAD1* as a novel CT antigen gene that encodes a putative meiosis-related protein.

Results

Selection of CT candidate genes by EST-based database analysis

The Ludwig Institute for Cancer Research (LICR) Transcriptome database (6) was analyzed, and transcripts with somatic tissue: testicular EST ratios of <5% ($P = 0.05$) were selected, resulting in a list of 371 genes. Twelve of the 371 genes had already been described in the literature as having a CT expression pattern, including seven listed in the recently compiled CT database (*SPANXA1/CT11.1*, *MAGEA2/CT1.2*, *GAGED2/CT12.1*, *BORIS/CT27*, *HAGE/CT13*, *AF15q14/CT29*, and *TDRD1/CT41.1*; 5). The remaining 359 genes were evaluated with on-line bioinformatics tools to confirm the testis-specificity of the mRNA transcript and to seek evidence of expression in cancer cell lines or tissues. Two hundred-thirty genes were found to have either ESTs present in more than two somatic tissues, no ESTs in any cancer cDNA libraries (except germ cell tumors), or inadequate data in the database. All such genes were eliminated. A sample of 20 genes was then selected from among the remaining 129 genes based on their higher testis:normal EST ratios and the presence of ESTs from more than one type of cancer. The mRNA distribution of these genes in normal tissues was analyzed by RT-PCR (Table 1).

Identification of four CT or CT-like genes by RT-PCR

Ten of the twenty genes selected showed ubiquitous expression in all twelve normal tissues examined, and three showed differential expression, with at least moderate expression in two or more somatic tissues (Table 2). Seven genes remained as potential CT genes, including four true testis-specific genes (*BOLL*, *PRM2*, *LOC440934*, *LOC151273*) and three genes with limited and/or weak expression in somatic tissues (*CPXCR1*, *C10orf94*, and *HORMAD1*).

The expression of these 7 genes was then evaluated in 29 cell lines consisting of 15 melanomas, 4 small cell lung cancers (NCI-H82, -H128, -H187, -H740), 3 non-small cell lung cancers (SK-LC-5, -14, -17), 3 colon cancers (SW403, HCT15, LS174T), 1 renal cancer (SK-RCC-1), 1 hepatocellular carcinoma (SK-HEP-1), 1 bladder cancer (T24), and 1 sarcoma (SW982). Melanoma expresses known CT antigens more frequently than most other tumor types (5). The other cell lines have previously been typed

© 2005 by Yao-Tseng Chen

and have been shown to express one or more known CT genes (data not shown).

The expression profile of the 7 potential CT genes in this selected “CT-rich” cell line panel is summarized in Table 3. Three genes—*BOLL*, *PRM2*, and *LOC151273*—showed no expression in any of the twenty-nine cell lines, indicating that these genes, although having cancer-derived ESTs in GenBank, are rarely expressed in cancer. The other four genes, *CPXCR1*, *C10orf94*, *LOC440934*, and *HORMAD1*, showed moderate to strong expression in one or more cell lines, which identified them as new CT or CT-like genes. The entire process of analyzing the expression pattern of the 20 genes by RT-PCR is summarized in Figure 1.

Of these four genes, *CPXCR1* and *C10orf94* showed moderate to strong mRNA expression in normal brain tissue by RT-PCR. *LOC440934* was only expressed in five of seven cell lines derived from lung cancer (including four from small cell lung cancer), and was not found to be expressed in any of the other twenty-two cell lines from other cell lineages. *CPXCR1*, *C10orf94*, and *LOC440934* are thus likely to be differentiation antigens,

exhibiting concurrent strong expression in testis but not in other somatic tissues, rather than true CT genes. This phenomenon has previously been observed, for example in the case of NY-BR-1, which is a breast differentiation antigen that is also expressed in testis (7). The products of *CPXCR1* and *C10orf94* are not likely to be useful as targets for cancer vaccines, as the concomitant brain expression raises the concern of antineuronal autoimmunity. On the other hand, the *LOC440934* gene product might be of value as a vaccine target for lung cancer.

In comparison to these three genes, *HORMAD1* was expressed in three melanoma cell lines and two nonmelanoma cell lines, and thus appears to be a new CT gene. This gene was designated CT46, following our proposed CT nomenclature system (5).

Quantitative RT-PCR (qRT-PCR) analysis of CT46/HORMAD1 expression

To confirm the qualitative RT-PCR data on cell lines and to further evaluate the expression of CT46/HORMAD1 in tumor tissues, qRT-PCR was performed. In addition to showing strong

Table 1

CT candidate genes selected for RT-PCR validation.

LICR No.	Gene Name	Ensembl No.	UniGene No.	Ref Seq. No.	Chromosome	Gene Description
HTR004485	<i>BOLL</i>	ENSG00000152430	Hs.169797	NM_033030	2q33	Boule-like (Drosophila)
HTR010660	<i>PRM2</i>	ENSG00000122304	Hs.2324	NM_002762	16p13	Protamine 2
HTR016539	<i>LOC440934</i>	N.A.	Hs.238964	BC033986	2q36	Hypothetical gene supported by BC008048
HTR051763	<i>LOC151273</i>	N.A.	Hs.244783	BC039382	2q32	New Ets-related factor (LOC151273)
HTR013870	<i>CPXCR1</i>	ENSG00000147183	Hs.458292	NM_033048	Xq21	CPX chromosome region, candidate 1
HTR055775	<i>C10orf94</i>	ENSG00000171772	Hs.549231	NM_130784	10q26	chromosome 10 open reading frame 94
HTR007567	<i>HORMAD1/CT46</i>	ENSG00000143452	Hs.298312	NM_032132	1q21	Hypothetical Protein DKFZp434A1315
HTR016783	<i>FLJ33768</i>	ENSG00000176363	Hs.177927	NM_173610	15q23	Hypothetical protein FLJ33768
HTR015705	<i>PCSK4</i>	ENSG00000115257	Hs.46884	NM_017573	19p13	Proprotein convertase subtilisin/Kexin type 4
HTR011589	<i>FSCN3</i>	ENSG00000106328	Hs.128402	NM_020369	7q31	Fascin homolog 3, actin-bundling protein, testicular
HTR009020	<i>HCFC2</i>	ENSG00000111727	Hs.506558	NM_013320	12q23	Host cell factor C2
HTR005822	<i>MGC26979</i>	ENSG00000164953	Hs.116240	NM_153704	8q22	MGC26979 hypothetical protein
HTR007542	<i>SCML2</i>	ENSG00000102098	Hs.495774	NM_006089	Xp22	Sex comb on midleg-like 2 (Drosophila)
HTR005702	<i>DEPDC1B</i>	ENSG00000035499	Hs.482233	NM_018369	5q12	DEP domain containing 1B
HTR009187	<i>YBX2</i>	ENSG00000006047	Hs.380691	NM_015982	17p11-13	Germ cell specific Y-box binding protein
HTR009044	<i>NYD-SP14</i>	ENSG00000137473	Hs.378893	NM_031956	4q31	NYD-SP14 protein
HTR006938	<i>NEK2</i>	ENSG00000117650	Hs.153704	NM_002497	1q32	NIMA (never in mitosis gene a)-related kinase 2
HTR001543	<i>TP53TG3</i>	ENSG00000180598	Hs.513537	NM_015369	16p13	TP53TG3 protein
HTR002199	<i>MBNL3</i>	ENSG00000076770	Hs.105134	NM_133486	Xq26.2	Muscleblind-like 3 (Drosophila)
HTR007263	<i>FLJ14904</i>	ENSG00000143194	Hs.180191	NM_032858	1q24	Hypothetical Protein FLJ14904

expression of *CT46/HORMAD1* in testicular tissue, qualitative RT-PCR (Table 2) showed weak expression of *CT46/HORMAD1* in brain, breast, colon, spleen, and placenta. This data was

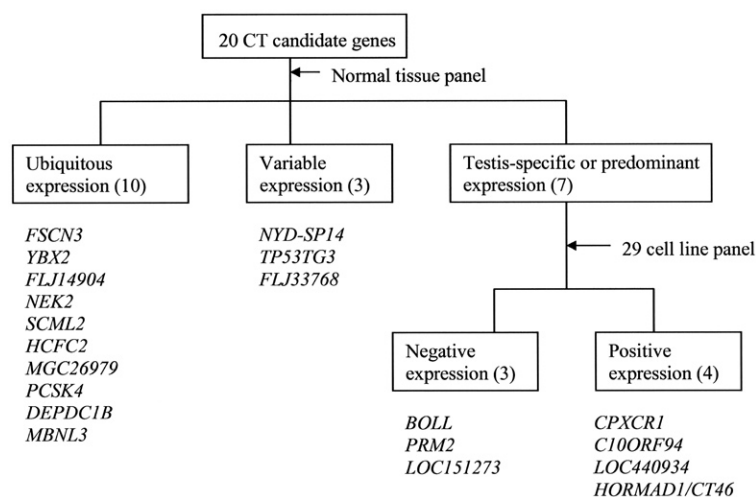
confirmed by qRT-PCR. Among 11 nontesticular normal tissues, the highest expression was seen in placental tissue, with 0.76% of the testicular expression, followed by spleen (0.55%),

Table 2

CT candidate gene expression in normal tissues.

Gene	Tissue											
	Brain	Breast	Colon	Kidney	Liver	Lung	Pancreas	Placenta	Prostate	Sk. Muscle	Spleen	Testis
<i>BOLL</i>	-	-	-	-	-	-	-	-	-	-	-	+++
<i>PRM2</i>	-	-	-	-	-	-	-	-	-	-	-	+++
<i>LOC440934</i>	-	-	-	-	-	-	-	-	-	-	-	+++
<i>LOC151273</i>	-	-	-	-	-	-	-	-	-	-	-	+
<i>CPXCR1</i>	++	-	-	-	-	+	-	-	-	+	-	+++
<i>C10orf94</i>	+++	-	-	-	-	-	-	-	-	-	-	+
<i>HORMAD1/CT46</i>	+	+	+	-	-	-	-	+	-	-	+	+++
<i>FLJ33768</i>	+	+	-	++	-	-	-	+	-	-	-	+++
<i>PCSK4</i>	++	+	+	++	++	+	+	-	+	-	-	+++
<i>FSCN3</i>	++	++	++	++	+	+	+	++	++	NT	+	+++
<i>HCFC2</i>	++	++	++	++	++	++	++	++	++	++	++	++
<i>MGC26979</i>	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
<i>SCML2</i>	++	++	++	++	++	++	++	++	++	++	++	++
<i>DEPDC1B</i>	+++	+++	+++	+++	+++	+++	+++	+++	+++	NT	+++	+++
<i>YBX2</i>	+++	+++	+++	+++	+++	-	+++	-	+++	NT	+	+++
<i>NYD-SP14</i>	+++	++	+	+++	-	+++	-	+	+	NT	+	+++
<i>NEK2</i>	+++	+++	+++	++	+	+++	+++	+++	+++	+	+++	+++
<i>TP53TG3</i>	++	+	+	+	-	+	+	+	+	-	+	+
<i>MBNL3</i>	++	++	++	++	++	++	++	++	++	++	++	+
<i>FLJ14904</i>	+++	++	+++	-	+	+	++	+++	+++	+	+	+++

Figure 1



Flowchart summarizing the analysis of 20 CT candidate genes. Gene expression in normal tissues and cancer cell lines was evaluated by qualitative RT-PCR. *Ubiquitous expression* corresponds to the presence of PCR products of similar intensity in most or all of the normal tissues examined, as judged by ethidium bromide staining on agarose gels. *Variable expression* means significant expression in at least several normal tissues, and *testis-specific or -predominant expression* means significant PCR products were observed only in testis, or in testis and not more than two additional normal tissues (refer to the text and Table 2).

Table 3

CT candidate gene expression in cell lines.

Cell Line	Gene						
	<i>BOLL</i>	<i>PRM2</i>	<i>LOC151273</i>	<i>CPXCRI</i>	<i>C10orf94</i>	<i>LOC440934</i>	<i>HORMAD1/CT46</i>
SK-MEL-3	-	-	-	-	-	-	-
SK-MEL-10	-	-	-	-	-	-	-
SK-MEL-12	-	-	-	-	-	-	+
SK-MEL-14	-	-	-	-	-	-	-
SK-MEL-21	-	-	-	-	-	-	-
SK-MEL-24	-	-	-	-	-	-	++
SK-MEL-28	-	-	-	-	-	-	-
SK-MEL-36	-	-	-	-	-	-	-
SK-MEL-37	-	-	-	-	-	-	-
SK-MEL-49	-	-	-	-	-	-	-
SK-MEL-55	-	-	-	-	-	-	-
SK-MEL-80	-	-	-	-	-	-	++
SK-MEL-95	-	-	-	-	-	-	-
SK-MEL-108	-	-	-	-	-	-	-
SK-MEL-128	-	-	-	-	-	-	-
NCI-H82	-	-	-	-	+++	++	-
NCI-H128	-	-	-	-	-	++	++
NCI-H187	-	-	-	-	+	++	-
NCI-H740	-	-	-	+	-	++	-
SK-LC-5	-	-	-	-	-	+	-
SK-LC-14	-	-	-	+	-	-	++
SK-LC-17	-	-	-	-	-	-	-
SW403	-	-	-	++	-	-	-
HCT-15	-	-	-	+	-	-	-
LS174T	-	-	-	+++	-	-	-
SK-RCC-1	-	-	-	-	-	-	-
SK-HEP-1	-	-	-	+	-	-	-
T24	-	-	-	-	-	-	-
SW982	-	-	-	-	-	-	-
testis	+++	+++	+	+++	+	+++	+++

and colon (0.23%). Other normal tissues expressed *CT46/HORMAD1* mRNA at <0.1% of testicular expression, including breast (0.046%) and brain (0.044%).

Quantitative RT-PCR analyses of cell lines confirmed the qualitative PCR data. Thus, of the 15 melanoma cell lines tested, the 3 positive lines - SK-MEL-12, -24, and -80 - expressed *CT46/HORMAD1* at 2.85%, 6.39%, and 8.33% of the testicular expression level, respectively. All other melanoma cell lines found to be negative by qualitative RT-PCR had *CT46/HORMAD1* mRNA levels that were <0.02% the testicular expression level. There is thus 100% concordance between the qualitative and quantitative RT-PCR results. Since these two assays utilized primers derived from different regions of the genes, this data validated the expression data of *CT46/HORMAD1* in normal tissue and in cell lines.

The expression of *CT46/HORMAD1* in additional tumor cell lines and tumor specimens was then examined by qRT-PCR and is summarized in Figure 2. We observed weak, moderate, and strong *CT46/HORMAD1* expression by qualitative RT-PCR to be approximately equivalent to >0.1%, >1%, and >10% of testicular expression as measured by qRT-PCR. Based on these cut-off values, moderate to strong *CT46/HORMAD1* expression (>1% testicular level) was seen in 14/30 (47%) non-small cell lung cancer specimens, 4/11 (36%) breast cancer specimens, 7/20 (35%) esophageal cancer specimens, 5/18 (28%) endometrial cancer specimens, 3/15 (20%) bladder cancer specimens, and 1/15 (7%) colon cancer specimens. Similar levels of expression were also seen in 4/12 (25%) small cell lung cancer cell lines and 2/17 (12%) colon cancer cell lines, but not in neuroblastoma cell lines (0/5). In total, 34/109 (31%) tumor specimens showed

>1% testicular level of expression, with 12/109 (11%) exhibiting strong (>10%) expression of *CT46/HORMAD1*.

CT46/HORMAD1 protein is immunogenic in cancer patients

BLAST analysis of the *CT46/HORMAD1* sequence against the patent database showed that a partial *CT46/HORMAD1* cDNA sequence had previously been identified by Obata et al. (GenBank Accession No. AX053429) by SEREX analysis of breast cancer with autologous patient serum. This indicates that *CT46/HORMAD1* is immunogenic and capable of eliciting spontaneous antibody responses in cancer patients.

The *CT46/HORMAD1* gene and its products

CT46/HORMAD1 is a single-copy gene, located on chromosome 1q21.3, that spans 22.8 kb and encodes an mRNA of 1880 bp (excluding the polyA tail). An intronless pseudogene was also identified on chromosome 6q12-14.1 (GenBank Accession No. AL132673), with 93% sequence identity to the *CT46/HORMAD1* cDNA sequence.

RT-PCR and DNA sequencing of testicular *CT46/HORMAD1* cDNA revealed two transcript variants. The predominant, full-length *CT46/HORMAD1* transcript consists of 13 exons, whereas the alternative transcript variant lacks exon 4 (64 bp). The major transcript encodes a putative protein of 394 aa, with the translational initiation site located in exon 2. If the same initiation site is used for transcript variant 2, the encoded protein would only be 60 aa in length, due to a frameshift in the open reading frame resulting from the missing 64 bp. Alternatively, this minor, shorter transcript may be translated from a new initiation site in exon 3, with a putative 323 aa protein, of which the carboxyl 313 residues are identical to those of the main product. A search for conserved protein domains identified a HORMA domain comprising the entire length of the full-length 394-aa sequence (KOG4652, HORMA domain; and pfam02301, HORMA domain) (Figure 3). Indeed, while this study was ongoing, the Human Genome Organization (HUGO) named the gene *HORMAD1*, recognizing it as a HORMA domain-containing protein. HORMA (for *H*op1p, *R*ev7p, and *M*AD2) domain proteins are involved in modulating chromatin structure and dynamics. Specifically, it has been suggested that the HORMA domain recognizes chromatin states that result from DNA double-strand breaks or nonattachment to the mitotic spindle and acts as an adaptor to recruit other proteins (8). Hop1, the prototype HORMA domain protein, is a yeast meiosis-specific protein, with which *CT46/HORMAD1* shares 25.8% homology over its 215 aa sequence. Although it is not certain whether *CT46/HORMAD1* is the human Hop1 ortholog, the presence of the HORMA domain, the similarity to Hop1 and asy1 (*Arabidopsis thaliana*, meiotic asynaptic mutant protein; 27.65% similarity over 260 residues), together with the germ cell-restricted expression of *CT46/HORMAD1*, all point to *CT46/HORMAD1* being a meiosis-related protein.

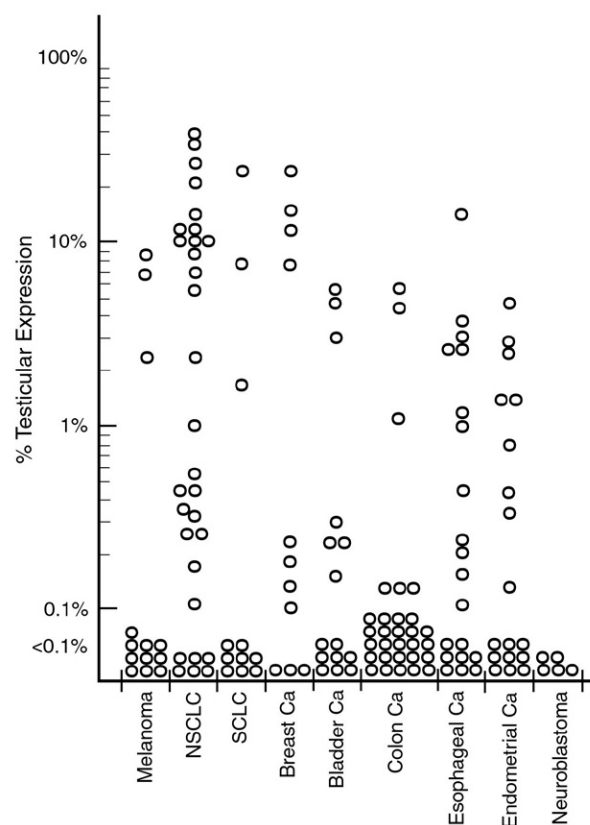
***CT46/HORMAD1* is highly conserved across species**

Homology searches using predicted *CT46/HORMAD1* protein sequences identified orthologs in other primates (*Macaca fascicularis*, GenPept Accession No. BAB63133), as well as rodents (*Mus musculus*, RefSeq Accession No. NP_080765; *Rattus norvegicus*, RefSeq Accession No. XP_228333). All are hypothetical proteins predicted from cDNA sequences. Each of the cDNAs was derived from testis, indicating conserved testis-specific transcription.

The available monkey cDNA sequence (GenBank Accession No. AB070034) is a partial sequence encoding the carboxyl 298 residues, with 98.3% (293/298) sequence identity to human *CT46/HORMAD1*. The mouse and rat counterparts are full-length sequences, with predicted proteins of 374 aa and 391 aa, respectively. The mouse protein shows 78% sequence identity to *CT46/HORMAD1* (89% similarity allowing conservative amino acid changes), and the rat protein has 72% identity to *CT46/HORMAD1*, with 83% sequence similarity, including conservative changes.

In addition to identifying these ortholog genes, the protein homology search identified additional meiotic synapsis proteins, including the meiotic synapsis protein from rice [GenPept Accession No. BAD00095, from *Oryza sativa* (japonica cultivar-group)] and the Asy1 meiotic protein from Chinese kale (GenPept Accession No. AAN37925), further supporting the hypothesis that *CT46/HORMAD1* is an evolutionarily conserved meiotic protein.

Figure 2



Expression of *CT46/HORMAD1* in tumor cell lines and specimens. The expression level was determined by real-time RT-PCR and was expressed as a percentage of the testicular expression level. Each open circle represents one sample. The melanoma, small cell lung cancer (SCLC), and neuroblastoma samples were cell lines, whereas RNA from primary tumor specimens was used for the non-small cell lung cancer (NSCLC), breast cancer, bladder cancer, esophageal cancer, endometrial cancer, and colon cancer samples.

MGC26710 is a human protein homologous to CT46/HORMAD1

Among human proteins, *MGC26710* is most similar to CT46/HORMAD1. The *MGC26710* gene is located on chromosome 22q12 and encodes a putative protein of 307 aa (RefSeq Accession No. NM_152510). Its similarity to CT46/HORMAD1 lies in the N-terminal HORMA domain, with 54% sequence identity in approximately the first 240 residues, which has 72% similarity, including conservative changes (Figure 3).

The mRNA expression of *MGC26710* in normal tissues was evaluated by qualitative RT-PCR. The results indicated tissue-restricted expression, with strong expression in testis, liver, and brain tissues, weak expression in kidney tissue, and no or minimal expression in 8 other normal tissues. Examination of the cancer cell lines showed moderate to strong expression in three of twenty-one cell lines tested (NCI-H82, SK-LC-14, and T24), which did not coincide with CT46/HORMAD1 expression. *MGC26710* is thus a differentially expressed gene, but differs from CT46/HORMAD1 in its normal and tumor-tissue expression profile.

Discussion

Through analysis of genes with predominant expression in testis, we have identified CT46/HORMAD1 as a novel CT antigen. Twenty-seven ESTs from normal tissues corresponding to CT46/HORMAD1 were found in GenBank, twenty-three from testis and four from brain tissue. By comparison, 9 ESTs derived from tumor tissue were found, including 4 from germ cell tumors, 4 from breast cancer, and 1 from lung cancer. The EST distribution thus suggested that CT46/HORMAD1 is a germ cell-specific gene that can be activated in non-germ cell malignancies, which is characteristic of CT antigen genes. Our experimental data confirm this impression, revealing CT46/HORMAD1 expression in lung, breast, esophageal, endometrial, bladder, and colon cancers. Although qRT-PCR detected amplification products in a few somatic tissues, we could not formally exclude the possibility that this was the result of amplifying contaminating genomic DNA, as the intronless pseudogene is highly homologous, even in the region where the trans-intronic primers and probe were derived. Even if mRNA

Figure 3

Conserved domain database 22426:
KOG4652, HORMA domain [Chromatin structure and dynamics]

Conserved domain length = 324 residues, 100% aligned

CT46	15	VFPNKISTEHQSLVLVKRLLA VSVSCITYLRGIFPECA YTRYLD--DLCVKILREDKNC PG--STQLVKWMLGC
KOG4652	1	TLPNGLNEKQSLFEMTRLL YVAISTILRERGI FPEEYFKDRYVDGNLLVMTLLRRQDAPEGRLVSWLEKGV---
CT46	85	YDALQKKYLRMVVLAVYTNPEDPQTISECYQFKFKYTNNGLMDFISK N-----QSNES SMLSTD-TKKASILL
KOG4652	73	HDAIRQKLLKLSL-VITESEDPEDI-EVYIFSFVYDEEGSVSARIN YINGINGQSSKAFELSQLSMDTTRRQFAKL
CT46	154	IRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPDYQPPGFKDGDCEGVIFEGEPMYLN VGEVSTPFHIFKVKVTT
KOG4652	146	IRKLHICTQLLEPLPQ-GLILSMRLYYTERVPPDYQPEGFKDSTRAFYTL PVNPEQINIGAVSTPHHKGFVKVL-
CT46	229	ERERMENIDSTILSPKQIKTPFQKILRDKDVEDEQEHYTSDDLDIETKMEEQEKNPASSELEEPSLVCEEDEIMR
KOG4652	219	-----SDATDSMEKAER-----T
CT46	304	SKESPDL SISHSQVEQLVNKTS ELDMSKTRSGKVFQNKMANGNQPVKSSKENRKRSQHESGR---IVLHHFDS
KOG4652	232	DKISDDP-FDLILVQQLNKSEADKSFQSEKTT SITPNVLGNPLVPVDQSEEDLLKSDSPGTGRCSCECGLDV
CT46	376	SSQESVPKRRKFSEPK EHI
KOG4652	306	SKQASVPKTRKSCRKTEHG

Homology between CT46 and MGC26710 hypothetical protein

Identities = 136/249 (54%), with conservative changes = 180/249 (72%)

CT46	1	MATAQLQR-----TPMSALVFPNKISTEHQSLVLVKRLLA VSVSCITYLRGIFPECA YTRYLDLDCVKILREDK
MGC26710	1	MATAQLSHCITIHKASKETVFP SQITNEHESLKMVKKLFATSI SCITYLRGLFPES SYGERHLLDSLKILREDK
CT46	71	NCPGSTQLVKWMLGCYDALQKKYLRMVVLAVYTNPEDPQTISECYQFKFKYTNNGLMDF--ISK NQSNES SMLS
MGC26710	76	KCPGSLHIIRWIQGC FDALEKRYLRMAVLTLYTDPMGSEKVTEMYQFKFKYTKEGATMDFDSSHSSSTSFS EGTNN
CT46	144	TDTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPDYQPPGFKDGD-CEGVIFEGEPMYLN VGEVST
MGC26710	151	EDIKKASVLLIRKLYILMQDLEPLPNNVLTMLKHYNAVTPHDYQPLGFKEGVNSHFLFDKEPINVQVGFVST
CT46	218	PFHIFKVKVTTTERERMENIDSTIL 241
MGC26710	226	GFHSMKVKVMT EATKVIDLENNLF 249

Homology between CT46/HORMAD1, KOG4652, and MGC26710. Amino acid sequence alignment between CT46/HORMAD1 and the prototype HORMA domain-containing protein KOG4652 (top), and between CT46/HORMAD1 and the homologous MGC26710 hypothetical protein (bottom). Identical sequences are indicated by the amino acid residue, conservative amino acid changes by plus signs, and gaps by dashes.

were expressed in somatic tissues, our data demonstrated that the level of expression in all somatic tissues is <1% that of testicular expression. Similar low-level expression has also been observed for other CT antigens (1), which does not preclude their use as targets for cancer vaccines.

It has been observed that CT antigens can be separated into two groups, based on whether or not they are located on chromosome X. Chromosome X has been shown to contain an unusually high number of testis-specific genes (9, 10), some of which are CT antigen genes. CT antigen genes belonging to this group include *MAGE*, *GAGE*, *NY-ESO-1*, *SSX*, *XAGE*, *SPANX*, and the recently identified *CT45* (11). These genes are almost always members of multigene families, with highly similar members derived from recent gene duplication events. In contrast, most CT antigen genes not located on chromosome X are single-copy genes. *CT46/HORMAD1* is a new member of the latter group.

Although the function of *CT46/HORMAD1* remains to be experimentally validated, the predicted protein contains a HORMA-domain, and is thus likely to be involved in regulating chromatin structure and dynamics. More specifically, *CT46/HORMAD1* is highly similar to meiotic proteins, consistent with its tissue-specific expression in germ cells. This likely association with meiosis is of particular interest, as other meiosis-related proteins have also been found to be CT antigens, including Spo11 and SCP-1 (synaptonemal complex protein 1) (12). We have speculated that expression of such meiosis-specific proteins in somatic cells may lead to genome instability and thus contribute to tumor progression (13).

Although this study resulted in the identification of *CT46/HORMAD1*, EST-based analyses in general are not particularly effective at identifying tissue-specific genes, including CT genes. Most genes that appeared to have a testis-specific or testis-predominant expression pattern based on EST data exhibited broad-spectrum expression in multiple somatic tissues upon RT-PCR analysis with gene-specific primers. One major reason for this is the underrepresentation of certain types of normal tissues in the EST database. At the time of our "Virtual Northern" analysis, 1,989,425 ESTs were included in the cDNA pool from "normal" tissues. This included 224,322 ESTs from brain and 92,259 from testis, whereas fewer than 1% of the ESTs were derived from pancreas (7614 ESTs), ovary (8152), spleen (16,164), or colon (17,509). As a result, there are frequently no ESTs for genes with low abundance transcripts in these tissues in the database. In comparison, recently developed gene-profiling techniques, such as massively parallel signature sequencing (MPSS), appear to promise a more comprehensive coverage of rare transcripts, at least at the present time. In parallel to the current study, we have also taken the massively parallel signature sequencing approach to identify new CT antigen genes, and this has led to the identification of more than a dozen novel CT genes, including *CT45*, a gene family located on chromosome Xq26 (11).

Abbreviations

CT, cancer/testis; qRT-PCR, quantitative RT-PCR

Acknowledgements

This work was supported by funding from the Cancer Research Institute (to Y.-T. Chen, C. V. Jongeneel, and A. O. Gure) through the Cancer Antigen Discovery Collaborative.

References

1. Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, Chen YT. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy [review]. *Immunol Rev* 2002; **188**: 22-32. (PMID: 12445278)
2. Boon T, van der Bruggen P. Human tumor antigens recognized by T lymphocytes [review]. *J Exp Med* 1996; **183**: 725-9. (PMID: 8642276)
3. Sahin U, Tureci O, Schmitt H, Cochlovius B, Johannes T, Schmits R, Stenner F, Luo G, Schober I, Pfreundschuh M. Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc Natl Acad Sci U S A* 1995; **92**: 11810-3. (PMID: 8524854)
4. Old LJ, Chen YT. New paths in human cancer serology. *J Exp Med* 1998; **187**: 1163-7. (PMID: 9547328)
5. Scanlan MJ, Simpson AJ, Old LJ. The cancer/testis genes: review, standardization, and commentary. *Cancer Immun* 2004; **4**: 1. (PMID: 14738373)
6. Stevenson BJ, Iseli C, Beutler B, Jongeneel CV. Use of transcriptome data to unravel the fine structure of genes involved in sepsis. *J Infect Dis* 2003; **187 Suppl 2**: S308-14. (PMID: 12792844)
7. Jager D, Stockert E, Gure AO, Scanlan MJ, Karbach J, Jager E, Knuth A, Old LJ, Chen YT. Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. *Cancer Res* 2001; **61**: 2055-61. (PMID: 11280766)
8. Aravind L, Koonin EV. The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair [review]. *Trends Biochem Sci* 1998; **23**: 284-6. (PMID: 9757827)
9. Wang PJ, McCarrey JR, Yang F, Page DC. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 2001; **27**: 422-6. (PMID: 11279525)
10. Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res* 2004; **14**: 1861-9. (PMID: 15466286)
11. Chen Y-T, Scanlan MJ, Venditti CA, Chua R, Theiler G, Stevenson BJ, Iseli C, Gure AO, Vasicek T, Strauberg RL, Jongeneel CV, Old LJ, Simpson AJG. Identification of cancer/testis-antigen genes by massively parallel signature sequencing. *Proc Natl Acad Sci U S A* 2005; **102**: 7940-5. (PMID: 15905330)
12. Tureci O, Sahin U, Zwick C, Koslowski M, Seitz G, Pfreundschuh M. Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens. *Proc Natl Acad Sci U S A* 1998; **95**: 5211-6. (PMID: 9560255)
13. Old LJ. Cancer/testis (CT) antigens - a new link between gametogenesis and cancer. *Cancer Immun* 2001; **1**: 1. (PMID: 12747762)
14. Kelso J, Visagie J, Theiler G, Christoffels A, Bardien S, Smedley D, Otgaar D, Greyling G, Jongeneel CV, McCarthy MI, Hide T, Hide W. eVOC: a controlled vocabulary for unifying gene expression data. *Genome Res* 2003; **13**: 1222-30. (PMID: 12799354)

15. SAGE Anatomic Viewer. URL: <http://cgap.nci.nih.gov/SAGE/AnatomicViewer>
16. NCBI BLAST. URL: <http://www.ncbi.nlm.nih.gov/BLAST/>
17. Ensembl Genome Browser. URL: <http://www.ensembl.org>

Materials and methods

Tumor tissue specimens and cell lines

Tumor tissue specimens were obtained from the Departments of Pathology at the Weill Medical College of Cornell University and at Memorial Sloan-Kettering Cancer Center, following protocols approved by their institutional review boards. Cell lines were obtained from the cell-line bank maintained at the New York Branch of the Ludwig Institute for Cancer Research.

EST-based identification of genes with a predominantly CT expression pattern

The LICR Transcriptome database (6) was used to search for genes showing a predominantly CT expression pattern, which we refer to as CT-like genes. This relational database documents clusters of transcript sequences (including ESTs) aligned to the genome, and the fine structure of the genes from which they are derived (6). A set of controlled vocabularies (eVOCs) (14) is used to describe the origin of EST libraries contributing to the database, allowing reliable searches for genes with specific tissue-expression patterns. The version of the Transcriptome database used during this study was based on Build 30 of the NCBI assembly of the human genome.

Three pools of ESTs were derived from the database. Pool A contained ESTs derived from cDNA libraries of normal adult tissues excluding testis, ovary, placenta, pooled normal tissues, and normal tissues of unknown origin. Pool B included ESTs from libraries of any cancer type except testicular cancer. Finally, pool C contained libraries from normal testis. Normalized and subtracted libraries, as well as small libraries (less than 600 ESTs), were excluded, in an attempt to avoid nonrepresentative EST data.

Genes showing an expression level in normal tissues (pool A) below 5% of that observed in normal testis (pool C), but which are also found in cancers (pool B), were retrieved. Fisher's exact test was applied to test the significance of the representational difference observed between pools A and C for the putative CT genes, and genes with $P < 0.05$ were retained. This list contained 371 candidates, including several genes already listed in the CT database (5).

In silico analysis

To select the most promising candidates among the 371 CT genes identified, the expression profiles of each gene in normal and tumor tissues were evaluated using a combination of the SAGE Anatomic Viewer and its Virtual Northern tool (15) and database searches using BLASTN (16). The objective of the analysis was to identify Unigene clusters containing ESTs derived from testis as well as from non-germ cell tumors, but with limited expression in somatic tissues. Once a Unigene cluster was determined to be a likely CT candidate, the intron-exon structure of the corresponding gene was defined using tools at the NCBI Web site. This information was then used to design trans-intronic primers for RT-PCR.

For specific genes of interest, for example, *CT46/HORMAD1*, various tools on the NCBI Web site were used for protein similarity searches, the identification of conserved domains, and the prediction of possible transcript variants and proteins. Gene identifiers were retrieved from the Ensembl database (17) in order to maintain a consistent naming convention; short names were assigned to each new gene identified in the project, using Human Gene Nomenclature Committee (HGNC)-approved symbols whenever possible.

Qualitative RT-PCR

For RT-PCR analysis of normal tissue expression, a panel of normalized cDNA (MTC panels I and II, BD Biosciences, Palo Alto, CA) derived from 16 normal tissues was used. Tissues included in these panels were brain, colon, heart, kidney, leukocytes, liver, lung, ovary, pancreas, placenta, prostate, skeletal muscle, small intestine, spleen, thymus, and testis.

In order to evaluate gene expression in tumor cell lines, total RNA was prepared by standard guanidinium thiocyanate-CsCl gradient method, and 2 μ g was used in a 20 μ l reverse transcription reaction. Two microliters of the synthesized cDNA was then used per 25 μ l PCR reaction. PCRs were set up using a commercial master mix (Platinum Taq Supermix, Invitrogen, Carlsbad, CA) and 35 cycles of amplification, each consisting of 15 s at 94°C, 1 min at 60°C, and 1 min at 72°C. The PCR products were visualized by 1% agarose gel electrophoresis and ethidium bromide staining.

Quantitative RT-PCR

Quantitative RT-PCR was performed using an ABI PRISM 7000 Sequence Detection System (Applied Biosystems, Foster City, CA). Normal testis total RNA was obtained commercially (Ambion, Austin, TX). Tumor tissue total RNA was prepared using Trizol reagents (Invitrogen). Two micrograms total RNA was used per 20 μ l reverse transcription reaction, and 2 μ l cDNA was then used for each 25 μ l PCR. The reactions were set up in duplicate sets, and the level of expression was determined as the abundance relative to that in the normal testis sample. For this purpose, a standard curve was established for each PCR plate, consisting of testicular cDNA in four-fold serial dilutions. Forty-five two-step cycles of amplification were performed, with each cycle consisting of 15 s at 95°C and 1 min at 60°C. The RNA quality of the cell lines and tissues was evaluated by separate control amplification of *GUS* and *GAPDH* transcripts. All specimens included in the final analysis have Ct values differing by less than four cycles, indicating similar cDNA quality and quantity.

Contact

Address correspondence to:

Yao-Tseng Chen, MD, PhD
Department of Pathology
C458A
Weill Medical College of Cornell University
1300 York Avenue
New York, NY 10021
USA
Tel.: + 1 212 746-6472
Fax: + 1 212 746-4483