

**UTILIZAÇÃO DOS DADOS DE SEQÜÊNCIAS  
EXPRESSAS NA IDENTIFICAÇÃO E  
CARACTERIZAÇÃO DE NOVOS ANTÍGENOS  
TUMORAIS *CANCER/TESTIS***

**FABIANA BETTONI**

**Tese apresentada à Fundação Antônio  
Prudente para a obtenção do Título de  
Doutor em Ciências**

**Área de concentração: Oncologia**

**Orientadora: Dra. Anamaria Aranha  
Camargo**

**São Paulo  
2007**

FUNDAÇÃO ANTÔNIO PRUDENTE  
  
Anamaria Rodrigues Alves Kuninari  
Diretora de Escola  
RG 6.080.138

FUNDAÇÃO ANTÔNIO  
PRUDENTE

## FICHA CATALOGRÁFICA

Preparada pela Biblioteca do Centro de Tratamento e Pesquisa  
Hospital do Câncer A. C. Camargo

Bettoni, Fabiana

**Utilização dos dados de seqüências expressas na identificação e caracterização de novos antígenos tumorais *Cancer/Testis*** / Fabiana Bettoni – São Paulo, 2007.

125p.

Tese (Doutorado)-Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências-Área de Concentração: Oncologia.

Orientadora: Anamaria Aranha Camargo

Descritores: 1. SEQUÊNCIAS ETIQUETAS EXPRESSAS. 2. ANTÍGENOS DE TUMORES. 3. IMUNOTERAPIA. 4. GENES/identificação. 5. BIOLOGIA COMPUTACIONAL.

*“O real é o mar. Nele há escritores que nadam  
e há escritores que mergulham, mas a água é a mesma.”*

*José Saramago*

## DEDICATÓRIA

*Aos meus pais, Élio e Anelita, e ao meu irmão, André, pelo carinho, apoio e amor nos momentos mais difíceis, sempre me ajudando a superar os obstáculos com força e coragem. Por tudo o que renunciaram nestes anos para que este sonho pudesse se tornar realidade, posso dizer, com toda certeza, que **ESSA VITÓRIA É NOSSA!** Muito obrigada por vocês existirem! **AMO VOCÊS!***

*À minha orientadora Dra. Anamaria Aranha Camargo por ter sido muito mais do que um exemplo de excelente profissional. Obrigada por todos os ensinamentos e conselhos nos momentos difíceis, pela paciência e também por ter sido mãe e amiga em tempo integral. Em especial, agradeço por ter acreditado comigo neste sonho e por hoje podermos vê-lo realizado.*

## AGRADECIMENTOS

Ao Raphael Parmigiani pela disponibilidade e paciência em me ensinar boa parte de tudo que aprendi nesses quatro anos. Muito obrigada pela sincera amizade, pelos conselhos dentro e fora do laboratório e, acima de tudo, por me incentivar sempre!

À Mariana Brait por ter sido muito mais que uma amiga. Muito obrigada por sempre me apoiar e também ouvir e compreender todas as minhas reclamações e medos nos momentos mais difíceis. Em especial, agradeço a ajuda e as sugestões que contribuíram para a finalização deste trabalho e também a toda a sua família por sempre me acolher com muito carinho.

À Lilian Pires por ter sido uma grande amiga e por ter acompanhado todo o desenvolvimento deste trabalho sempre disposta a ajudar e a tentar solucionar os problemas principalmente quando decidimos entrar no mundo da proteômica. Em especial, agradeço ao Vinícius que continua sendo meu único sobrinho e que com suas brincadeiras tornaram meus dias muito mais alegres e divertidos.

À Márcia Dellamano pela amizade sincera, pelo ombro amigo, pelas risadas e momentos de diversão e também pela troca de conselhos dentro e fora do laboratório. Muito obrigada por sempre torcer por mim!

À Rose Santiago pela amizade, pelos momentos de diversão e também por ser minha cunhada preferida.

À Ana Paula Felipe pela amizade, pelo ombro amigo, pelas risadas e pelos momentos de descontração e diversão que me ajudaram muito durante a finalização deste trabalho.

À Anna Christina Salim pela disponibilidade em me ajudar sempre que aquelas dúvidas “sem sentido” apareciam e também pela amizade e pelo apoio nos momentos de dificuldade.

Ao Daniel Vidal pela amizade e também pela ajuda no laboratório e pelas discussões científicas.

À Elisângela Monteiro pela amizade, pelos momentos de descontração e pela ajuda no dia-a-dia.

À Felícia Cavalher pela disponibilidade e ajuda com os experimentos de *Southern-blot* e também pela amizade.

Ao Fernando Camargo Filho pelo desenvolvimento do *pipeline* de seleção dos candidatos, pelas discussões sobre bioinformática e biologia molecular e pela amizade.

À Maria Cristina Costa pela disponibilidade em realizar as reações de seqüenciamento.

Ao Murilo Geraldo pela colaboração no início do desenvolvimento deste trabalho.

A todos os amigos do Laboratório de Biologia Molecular, Anamaria, Ana Paula, Andréa, Anna Christina, Camila, Daniel, Daniela, Elisa, Érico, Felícia, Lilian, Maria Cristina, Maria Izabel, Mônica, Newton, Ricardo, Tamara e Valéria pela amizade, apoio e ajuda oferecidos no desenvolvimento deste trabalho.

À Waleska Martins pela ajuda com a extração dos RNAs e também pela amizade dentro e fora do laboratório.

À Dra. Dirce Carraro pela amizade e pelo apoio no início do desenvolvimento deste trabalho.

À Marilene Lopes pela ajuda e atenção nos momentos de dificuldade durante os experimentos de clonagem, expressão e purificação das proteínas recombinantes.

Ao Dr. Flávio Henrique-Silva da Universidade Federal de São Carlos pela atenção e disponibilidade em me ajudar com os experimentos de expressão e purificação das proteínas recombinantes e também por ceder as proteínas de cana-de-açúcar (CaneCPI-2 e CaneCPI-3).

À Patrícia Sávio pela ajuda e disposição durante os experimentos de clonagem e expressão das proteínas recombinantes.

Ao Dr. Sandro de Souza e todo o seu Laboratório de Biologia Computacional pela disponibilidade e ajuda com o Banco de Dados do Transcriptoma Humano. Em especial ao Pedro Galante pela importante ajuda com a análise dos antígenos CTs já conhecidos e também na construção do Banco de Dados Map4\_bh e à Natanja Kirschbaum pela ajuda na análise do Banco de Dados Map4\_bh.

À Dra. Helena Brentani pela contribuição na análise do Banco de Dados do Transcriptoma Humano e também na construção do Banco de Dados Map4\_bh.

Aos demais colegas do Instituto Ludwig pela convivência e amizade.

À Dra. Bettina Malnic, Dra. Eugenia Constanzi Strauss e ao Dr. Alex Fiorini, pelo acompanhamento de meu trabalho e pelas contribuições através de críticas e sugestões.

À Dra. Ana Lúcia Nascimento do Centro de Biotecnologia do Instituto Butantan, em especial à Fernanda, pela disponibilidade e ajuda na utilização da *French® Pressure Cell Press* (Sim-Aminco®) em seu laboratório.

Às amigas Milena, Paula, Juliana e Denise pelo apoio em todos esse anos de amizade e, principalmente, por entenderem minha ausência e os tempos sem notícias.

A todos os velhos e novos amigos que estiveram ao meu lado nestes quatro anos. Obrigada pela força, pelo ombro amigo, pelas risadas, pelos momentos de diversão e também pela ajuda na finalização deste trabalho. Sem a amizade de vocês tudo teria sido mais difícil.

A todos os funcionários do Instituto Ludwig pela amizade, convivência diária e também pelo apoio técnico e administrativo. Em especial à Izabel Carneiro e Sirléia Miranda pelas conversas e risadas entre um experimento e outro e à Ana Cláudia, pela atenção e disponibilidade em me ajudar com os problemas do dia-a-dia.

À Luciana Pitombeira e Ana Maria Kuninari, pelo trabalho desenvolvido junto à secretaria de pós-graduação e por todo apoio e disponibilidade recebidos para o desenvolvimento deste trabalho nesta Fundação.

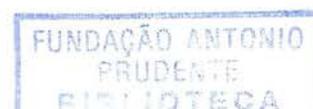
À Suely Francisco e todos os funcionários da biblioteca da Fundação Antonio Prudente pela disponibilidade e atenção na organização deste material bibliográfico.

Ao Prof. Dr. Luiz Fernando Lima Reis pela coordenação deste curso de pós-graduação.

Aos Profs. Dr. Ricardo Brentani e à Dra. Luisa Villa pela direção do Instituto Ludwig e por permitirem a realização deste trabalho em um centro de excelência na pesquisa sobre o câncer.

À FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) pela bolsa de estudo concedida (Processo 03/05193-0).

À Escola Especializada “Schwester Heine”, em especial à professora Tânia Siqueira e a todas as crianças deste hospital, por me mostrarem o verdadeiro valor de



um sorriso e por me ensinarem diariamente como vale a pena acreditar em nossos sonhos.

A todas as pessoas que por eventual falha de memória não mencionei, minhas sinceras desculpas e o meu muito obrigado.

## RESUMO

Bettoni F. **Utilização dos dados de seqüências expressas na identificação e caracterização de novos antígenos tumorais *Cancer/Testis***. São Paulo; 2007. [Tese de Doutorado-Fundação Antonio Prudente]

Os antígenos *Cancer/Testis* (CTs) são um subgrupo de antígenos tumorais com expressão restrita a testículo normal e diferentes tipos de tumores. Estes antígenos são capazes de induzir resposta imune humoral e celular em pacientes com câncer e, devido ao seu restrito padrão de expressão, são considerados candidatos ideais para o desenvolvimento de vacinas e imunização passiva. Neste trabalho, uma estratégia computacional baseada em dados de expressão provenientes de seqüências expressas (ESTs) foi utilizada para a identificação de novos antígenos CT. Seqüências expressas (mRNA e ESTs) foram alinhadas a seqüência genômica humana permitindo o agrupamento de seqüências expressas derivadas de um mesmo gene. Considerando a origem tecidual das seqüências expressas correspondentes a cada um dos genes foi possível definir um padrão de expressão *in silico* e selecionar genes candidatos a novos antígenos CT. No total foram identificados 1255 genes candidatos representados por ESTs que apresentassem splicing e fossem derivadas de bibliotecas de testículo e/ou tecido tumoral dos quais 93 foram selecionados para validação experimental do seu padrão de expressão. A validação experimental do padrão de expressão foi realizada através de RT-PCR em 21 tecidos normais, 17 linhagens celulares tumorais e 160 amostras de nove diferentes tipos de tumores. Com base nesta análise, foram identificados cinco candidatos a antígenos CT freqüentemente expressos em diferentes tipos de tumores. O candidato CT101 apresentou-se expresso em 41% das amostras tumorais, e a maior freqüência de expressão foi observada em glioblastomas (70%). O candidato CT704 apresentou-se expresso em 65% das amostras tumorais com uma alta freqüência de expressão entre os tumores de pulmão (93%). O candidato CT802 apresentou-se expresso em 20% das amostras tumorais com maior expressão nos tumores de útero (50%). O candidato CT809 mostrou-se expresso em 24% das amostras tumorais e apresentou expressão predominante nos tumores de pulmão (50%). O candidato CT1001

apresentou-se expresso em 18% das amostras tumorais sendo mais frequentemente expresso nos tumores gástricos (33,3%). Os candidatos CT101, CT802 e CT809 para os quais já havia uma seqüência completa de mRNA disponível em bancos de dados públicos, foram selecionados para avaliar a presença de resposta imune humoral em pacientes com câncer. Para tanto, as respectivas proteínas recombinantes foram expressas em sistema bacteriano e utilizadas em ensaios de *Immunoblotting*. Anticorpos anti-CT101 [transcrito correspondente ao gene PASD1 (*PAS domain containing protein 1*)] foram encontrados em 41% das amostras de plasma de pacientes com câncer de útero e em 43,3% dos indivíduos sadios. A freqüência de anticorpos anti-CT802 [transcrito correspondente ao gene ASZ1 ou GASZ (*Germ cell-specific ankyrin, SAM and basic leucine zipper domain containing protein 1*)] observada nas amostras de plasma de pacientes com câncer de útero foi de 22,7% enquanto dentre os indivíduos sadios a freqüência foi de 6,7%. Já a freqüência de anticorpos anti-CT809, correspondente ao gene FAM46D (*Family with sequence similarity 46, member D*), foi de 7,7% nos pacientes com câncer de pulmão e 6,7% nos indivíduos sadios. Além disso, utilizamos a técnica de RACE para caracterizar os candidatos CT704 e CT1001 que não apresentavam seqüência completa de mRNA disponível em banco de dados públicos. Em conjunto esses resultados demonstram que a estratégia *in silico* utilizada neste trabalho foi eficiente na identificação de 5 antígenos CTs sendo que para 3 destes foi possível detectar resposta imune humoral em pacientes com câncer.

## SUMMARY

Bettoni F. [Utilization of the expressed sequences data to the identification and characterization of new Cancer/Testis antigens]. São Paulo; 2007. [Tese de Doutorado-Fundação Antonio Prudente]

Cancer/testis (CT) antigens are a subgroup of tumor antigens with a restricted expression in normal testis and in different types of tumors. These antigens are capable of eliciting humoral and cellular immune response in cancer patients and due to their restricted expression pattern they are considered promising candidates for the development of vaccines and passive immunotherapy. In this work, a computational approach based on expression data from expressed sequence tags (ESTs) was used to identify novel CT antigens. Expressed sequences (mRNA and ESTs) were aligned against the human genome sequence, allowing the clustering of the sequences derived from a same gene. Considering the tissue origin of the expressed sequences corresponding to each gene it was possible to define an *in silico* expression pattern and to select novel CT antigen candidates. A total of 1255 candidate genes represented by spliced ESTs derived from testis and/or tumoral cDNA libraries were identified and 93 of them were selected for experimental validation of their expression pattern. The experimental validation of the expression pattern was carried out by RT-PCR in 21 normal tissues, 17 tumor cell lines and 160 samples derived from 9 different types of tumors. Based on this analysis we were able to identify five CT antigens candidates frequently expressed in different types of tumors. Candidate CT101 was expressed in 41% of tumor samples and the highest frequency of expression was observed in glioblastomas (70%). Candidate CT704 was expressed in 65% of tumor samples with a high expression frequency among lung tumors (93%). Candidate CT802 was expressed in 20% of tumor samples with higher expression in uterus tumors (50%). Candidate CT809 was expressed in 24% of tumor samples and was predominantly expressed in lung tumors (50%). Candidate CT1001 was expressed in 18% of the tumor samples being more frequently expressed in gastric

tumors (33.3%). Candidates CT101, CT802 and CT809 for which a full-length sequence were already available in public databases were selected to evaluate the presence of humoral immune response in cancer patients. The respective recombinant proteins were expressed in a bacterial system and were used in immunoblotting assays. Anti-CT101 antibodies [transcript corresponding to PASD1 gene (PAS domain containing protein 1)] were observed in 41% of plasma samples from patients with uterus tumor and in 43.3% of plasma from healthy individuals. The frequency of anti-CT802 antibodies [transcript corresponding to ASZ1 or GASZ gene (Germ cell-specific ankyrin, SAM and basic leucine zipper domain containing protein 1)] observed in plasma samples from patients with uterus tumor was 22.7% while among healthy individuals the frequency was 6.7%. Finally, the frequency of anti-CT809 antibodies, that corresponds to FAM46D gene (Family with sequence similarity 46, member D) was 7.7% among patients with lung tumors and 6.7% among healthy individuals. Moreover, we performed RACE experiments to characterize candidates CT704 and CT1001 that did not present a full-length mRNA sequence available in public databases. Taken together, these results showed that the *in silico* strategy used in this work was efficient in the identification of 5 CT antigens and for 3 of them it was possible to detect humoral immune response in cancer patients.

## LISTA DE FIGURAS

<b>Figura 1</b>	Esquema representando a construção do Banco de Dados do Transcriptoma Humano (Map4).	27
<b>Figura 2</b>	Esquema representativo de um <i>cluster</i> presente no Banco de Dados Map4.	28
<b>Figura 3</b>	Visualização das quatro ontologias do eVOC utilizadas na anotação das bibliotecas de cDNA disponibilizadas pelo GenBank.	30
<b>Figura 4</b>	Interface gráfica disponibilizada pelo programa BLAT.	33
<b>Figura 5</b>	Seqüência genômica disponibilizada pelo programa BLAT onde se localiza o candidato analisado.	33
<b>Figura 6</b>	Mapa do vetor pET-28a.	44
<b>Figura 7</b>	Mapa do vetor pET-32a.	45
<b>Figura 8</b>	Seqüência referente à fase aberta de leitura do transcrito CT101.	51
<b>Figura 9</b>	Seqüência referente à fase aberta de leitura do transcrito CT802.	51
<b>Figura 10</b>	Seqüência referente à fase aberta de leitura do transcrito CT809.	52
<b>Figura 11</b>	Representação esquemática do alinhamento entre a seqüência genômica e as seqüências disponíveis em bancos de dados públicos para os transcritos CT704 (A) e CT1001 (B).	61
<b>Figura 12</b>	Expressão dos candidatos CT101, CT704, CT802, CT809, CT1001 e do gene GAPDH nas amostras de tecidos normais após 10 minutos de exposição.	77
<b>Figura 13</b>	Expressão dos candidatos CT101, CT704, CT802, CT809, CT1001 e do gene GAPDH nas linhagens celulares tumorais após 10 minutos de exposição.	78
<b>Figura 14</b>	Representação gráfica do perfil de expressão por RT-PCR dos cinco candidatos a antígenos CT nas amostras de tumores.	80
<b>Figura 15</b>	Figura esquemática da estrutura gênica das formas de <i>splicing</i> do transcrito PASD1.	82
<b>Figura 16</b>	Extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT101.	84

<b>Figura 17</b> <i>Immunoblotting</i> com extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT101.	85
<b>Figura 18</b> Identificação de anticorpos anti-CT101 em plasma de pacientes com câncer de útero por <i>Immunoblotting</i> .	86
<b>Figura 19</b> Extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT802 fusionada com Tioredoxina (Trx).	88
<b>Figura 20</b> <i>Immunoblotting</i> com extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT802 fusionada com Tioredoxina.	89
<b>Figura 21</b> Digestão enzimática da proteína CT802 com enterokinase.	90
<b>Figura 22</b> Identificação de anticorpos anti-CT802 em plasma de pacientes com câncer de útero por <i>Immunoblotting</i> .	91
<b>Figura 23</b> Extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT809.	92
<b>Figura 24</b> <i>Immunoblotting</i> com extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT809.	92
<b>Figura 25</b> Identificação de anticorpos específicos em plasma de pacientes com câncer de pulmão por <i>Immunoblotting</i> .	94
<b>Figura 26</b> Representação esquemática do alinhamento entre a seqüência genômica e as seqüências geradas por RACE para extensão do transcrito CT704.	95
<b>Figura 27</b> Representação esquemática do alinhamento entre a seqüência genômica e a seqüência gerada por RACE para extensão do transcrito CT1001.	96

## LISTA DE TABELAS

<b>Tabela 1</b>	Comparação entre o número de seqüências e <i>clusters</i> presentes nos bancos de dados Map4 e Map4_bh.	31
<b>Tabela 2</b>	Amostras tumorais utilizadas para avaliação da expressão dos candidatos selecionados.	37
<b>Tabela 3</b>	Enzimas utilizadas para a clonagem sítio dirigida dos candidatos CT101, CT802 e CT809 e os respectivos iniciadores para amplificação da fase aberta de leitura com o tamanho do fragmento amplificado pelos mesmos.	46
<b>Tabela 4</b>	Iniciadores utilizados nas reações de RACE para gerar a seqüência completa dos transcritos CT704 e CT1001.	60
<b>Tabela 5</b>	Lista dos 20 antígenos CT já descritos na literatura utilizados para a definição dos critérios para seleção dos <i>clusters</i> candidatos.	64
<b>Tabela 6</b>	Listas geradas pelo <i>pipeline</i> e respectivos critérios utilizados para seleção dos candidatos.	69
<b>Tabela 7</b>	<i>Clusters</i> selecionados para o processo de validação experimental.	70
<b>Tabela 8</b>	Número de <i>clusters</i> que apresentaram expressão em testículo.	71
<b>Tabela 9</b>	Candidatos que apresentaram expressão em testículo e que seguiram para a segunda fase do processo de validação experimental.	72-3
<b>Tabela 10</b>	Número de <i>clusters</i> que apresentaram expressão exclusiva em testículo e em testículo e tecidos normais do sistema nervoso.	75
<b>Tabela 11</b>	Dados de expressão nas linhagens celulares tumorais analisadas dos 6 candidatos que apresentaram um perfil de expressão característico de um antígeno CT ou CTB.	76
<b>Tabela 12</b>	Número de <i>clusters</i> que apresentaram expressão em pelo menos uma linhagem celular analisada.	76
<b>Tabela 13</b>	Perfil de expressão dos cinco candidatos a antígenos CT nas amostras tumorais testadas.	79

## LISTA DE ABREVIATURAS

<b><math>\Omega</math></b>	Ohm
<b><math>\mu\text{g}</math></b>	Micrograma
<b><math>\mu\text{l}</math></b>	Microlitro
<b><math>\mu\text{m}</math></b>	Micrômetro
<b><math>\mu\text{M}</math></b>	Micromolar
<b>ADAM</b>	<i>A Desintegrin and Metalloproteinase</i>
<b>ATCC</b>	<i>American Type Culture Collection</i>
<b>BLAST</b>	<i>Basic Local Alignment Search Tool</i>
<b>BLAT</b>	<i>BLAST-Like Alignment Tool</i>
<b>BSA</b>	Albumina Sérica Bovina (do inglês <i>Bovine Serum Albumin</i> )
<b>cDNA</b>	DNA complementar
<b>CEA</b>	Antígeno carcinoembrionário (do inglês <i>Carcinoembryonary Antigen</i> )
<b>CG</b>	<i>CircleGrow</i>
<b>CGAP</b>	<i>Cancer Genome Anatomy Project</i>
<b>ChT</b>	<i>Choriotrofolasto-testis</i>
<b>CT</b>	<i>Cancer Testis</i>
<b>CTB</b>	<i>Cancer-Testis-Brain</i>
<b>dbEST</b>	Banco de Dados de ESTs (do inglês <i>EST database</i> )
<b>DEPC</b>	Di-etil pirocarbonato
<b>DNA</b>	Ácido desoxirribonucléico
<b>dNTP</b>	Deoxinucleotídeo
<b>DO</b>	Densidade Óptica
<b>EBV</b>	vírus Epstein-Barr
<b>EDTA</b>	Ácido etilenodiaminotetracético disódico
<b>ELISA</b>	<i>Enzyme-Linked Immunosorbent Assay</i>
<b>EST</b>	Etiqueta de seqüência expressa (do inglês <i>Expressed Sequence Tags</i> )
<b>FAM46D</b>	<i>Family with sequence similarity 46, member D</i>
<b>g</b>	Grama
<b>GAPDH</b>	Gliceraldeído 3-fosfato desidrogenase

<b>GASZ</b>	<i>Germ cell-specific ankyrin, SAM and basic leucine zipper domain ] containing protein 1</i>
<b>HCGP</b>	Projeto Genoma Humano do Câncer
<b>hMLH1</b>	<i>Human mut-L homologue-1</i>
<b>HPV</b>	Papiloma vírus humano
<b>HLA</b>	Antígeno leucocitário humano (do inglês <i>Human Leucocyte Antigen</i> )
<b>IFN-<math>\gamma</math></b>	Interferon- $\gamma$
<b>Ig</b>	Imunoglobulina
<b>IL</b>	Interleucina
<b>INCA</b>	Instituto Nacional do Câncer
<b>IPTG</b>	$\beta$ -D-tiogalactopiranosídeo de Isopropila
<b>JNK</b>	<i>c-Jun NH2-terminal protein kinase</i>
<b>kb</b>	Kilobases
<b>kDa</b>	Kilodaltons
<b>LAK</b>	<i>Lymphokine Activated Killer cells</i>
<b>M</b>	Molar
<b>mA</b>	Miliampere
<b>Map4_bh</b>	<i>Map4_best hits</i>
<b>MHC</b>	Complexo principal de histocompatibilidade (do inglês <i>Major histocompatibility complex</i> )
<b>ml</b>	Mililitro
<b>MPSS</b>	<i>Massively Parallel Signature Sequencing</i>
<b>mM</b>	Milimolar
<b>mRNA</b>	RNA mensageiro
<b>NCBI</b>	Centro Nacional de Informação em Biotecnologia (do inglês <i>National Center for Biotechnology Information</i> )
<b>ng</b>	Nanograma
<b>NK</b>	<i>Natural Killer</i>
<b>nm</b>	Nanômetro
<b>nM</b>	Nanomolar
<b>ORESTES</b>	<i>Open Reading Frame ESTs</i>
<b>ORF</b>	Fase aberta de leitura (do inglês <i>Open Reading Frame</i> )

<b>PASD1</b>	<i>PAS domain containing protein 1</i>
<b>pb</b>	Pares de base
<b>PBS</b>	<i>Phosphate Buffered Saline</i>
<b>PBST</b>	<i>Phosphate Buffered Saline Tween</i>
<b>PCR</b>	Reação em cadeia da polimerase (do inglês <i>Polimerase Chain Reaction</i> )
<b>RACE</b>	<i>Rapid Amplification of cDNA Ends</i>
<b>RDA</b>	Análise representativa diferencial (do inglês <i>Representational Difference Analysis</i> )
<b>RNA</b>	Ácido ribonucléico
<b>RNase</b>	Ribonuclease
<b>rpm</b>	rotações por minuto
<b>RT</b>	Transcrição reversa (do inglês <i>Reverse Transcriptase</i> )
<b>SAGE</b>	<i>Serial Analysis of Gene Expression</i>
<b>SAM</b>	<i>Sterile alpha motif</i>
<b>SDS</b>	<i>Sodium lauryl sulfate</i>
<b>SEREX</b>	<i>Serological Analysis of Recombinant tumor cDNA Expression Libraries</i>
<b>SSC</b>	<i>Saline Sodium Citrate</i>
<b>TAE</b>	Tris Acetato EDTA (do inglês <i>Tris Acetate EDTA</i> )
<b>TGF-<math>\beta</math></b>	<i>Transforming growth factor <math>\beta</math></i>
<b>TIGR</b>	<i>The Institute for Genomic Research</i>
<b>UCSC</b>	<i>University of California Santa Cruz</i>
<b>UV</b>	Ultravioleta

## ÍNDICE

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>1.1</b>	Imunologia de Tumores	4
<b>1.2</b>	Antígenos Tumorais	7
<b>1.3</b>	Antígenos CT	9
<b>1.3.1</b>	Técnicas imunológicas usadas para identificação de antígenos CT	13
<b>1.3.2</b>	Técnicas não imunológicas usadas para identificação de antígenos CT	15
<b>1.4</b>	Banco de Dados do Transcriptoma Humano	19
<b>2</b>	<b>OBJETIVOS</b>	<b>23</b>
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>25</b>
<b>3.1</b>	Identificação de novos antígenos CT	26
<b>3.1.1</b>	Construção do Banco de Dados do Transcriptoma Humano (Map4)	26
<b>3.1.2</b>	Construção do Banco de Dados Map4_best hits (Map4_bh)	30
<b>3.1.3</b>	Construção do pipeline de seleção de novos antígenos CT	31
<b>3.1.4</b>	Inspeção manual dos candidatos selecionados	32
<b>3.2</b>	Avaliação do padrão de expressão dos candidatos selecionados	34
<b>3.2.1</b>	Tecidos normais	34
<b>3.2.2</b>	Linhagens celulares tumorais	35
<b>3.2.3</b>	Amostras de tumores de pacientes do Banco de Tumores do Hospital AC Camargo	36
<b>3.2.4</b>	Avaliação da contaminação com DNA genômico nos RNAs extraídos	38
<b>3.2.5</b>	Tratamento do RNA total com DNaseI	39
<b>3.2.6</b>	Síntese de cDNA	39
<b>3.2.7</b>	RT-PCR	40

3.2.8	Confirmação da identidade dos fragmentos de RT-PCR correspondentes aos candidatos validados na análise de expressão por sequenciamento	41
3.2.9	<i>Southern-blot</i>	42
3.3	Expressão das Proteínas Recombinantes	43
3.3.1	Clonagem da fase aberta de leitura dos candidatos com seqüência completa de mRNA em vetor de expressão	43
3.3.2	Indução da expressão das proteínas recombinantes	52
3.3.3	Deteção da expressão das proteínas recombinantes através de <i>Immunoblotting</i>	53
3.3.4	Teste de solubilidade das proteínas recombinantes	54
3.3.5	Purificação das proteínas recombinantes	55
3.3.6	Exclusão da proteína Tio redoxina fusionada com a proteína recombinante CT802	56
3.4	Deteção de anticorpos contra as proteínas recombinantes em plasma de pacientes	57
3.4.1	Amostras utilizadas	57
3.4.2	<i>Immunoblotting</i>	58
3.5	Obtenção da seqüência completa dos candidatos a antígenos CTs representados exclusivamente por ESTs	59
4	<b>RESULTADOS</b>	62
4.1	Seleção de genes candidatos com padrão de expressão característico de antígenos CT no Banco de Dados do Transcriptoma Humano (Map4)	63
4.1.1	Análise da distribuição de ESTs nos <i>clusters</i> correspondentes a antígenos CT conhecidos do Banco de Dados do Transcriptoma Humano (Map4)	63
4.1.2	Seleção automática de candidatos a novos antígenos CT no Map4_bh	68
4.1.3	Inspeção manual e seleção de <i>clusters</i> candidatos para validação experimental	69

4.2	Validação experimental dos candidatos selecionados	70
4.2.1	Análise da expressão dos candidatos selecionados em testículo	70
4.2.2	Análise de expressão em tecidos normais dos candidatos com expressão positiva em testículo	74
4.2.3	Análise de expressão em linhagens celulares de origem tumoral dos candidatos com expressão exclusiva em testículo e expressão em testículo e tecidos normais do sistema nervoso	75
4.2.4	Análise de expressão dos candidatos em amostras de tumores de pacientes provenientes do Banco de Tumores do Hospital AC Camargo	79
4.3	Análise da resposta humoral contra o candidato CT101 em pacientes com câncer	81
4.3.1	Expressão e purificação da proteína recombinante	81
4.3.2	Deteção de anticorpos anti-CT101 em plasma de pacientes com câncer	85
4.4	Análise da resposta humoral contra o candidato CT802 em pacientes com câncer	87
4.4.1	Expressão e purificação da proteína recombinante	87
4.4.2	Deteção de anticorpos anti-CT802 em plasma de pacientes com câncer	89
4.5	Análise da resposta humoral contra o candidato CT809 em pacientes com câncer	91
4.5.1	Expressão e purificação da proteína recombinante	91
4.5.2	Deteção de anticorpos anti-CT809 em plasma de pacientes com câncer	93
4.6	Obtenção da seqüência completa dos candidatos CT704 e CT1001	94
4.6.1	CT704	94
4.6.2	CT1001	96
5	<b>DISCUSSÃO</b>	98
6	<b>CONCLUSÕES</b>	112

**ANEXOS**

- Anexo 1** Número de acesso da seqüência referente a cada candidato selecionado e a seqüência dos iniciadores referentes a cada um com o respectivo tamanho do produto amplificado.
- Anexo 2** “Antibodies against the cancer-testis antigen CTSP-1 are frequently found in prostate cancer patients and are an independent prognostic factor for biochemical-recurrence after surgery.” Artigo científico submetido à revista Prostate.
- Anexo 3** Parmigiani RB, Bettoni F, Vibranovski MD, et al. Characterization of a *cancer/testis* (CT) antigen gene family capable of eliciting humoral response in cancer patients. **Proc Natl Acad Sci U.S.A.** 2006; 103:18066-71.

# ***INTRODUÇÃO***



## 1 INTRODUÇÃO

O câncer é considerado um dos maiores problemas de saúde mundial e uma das mais importantes causas de mortalidade e morbidade na população. Dados da Sociedade Americana de Câncer estimam que, em 2007, 1.444.920 novos casos da doença ocorram nos Estados Unidos sendo esperado 559.650 óbitos por câncer para o mesmo ano neste país (JEMAL et al. 2007). Já no Brasil, o Ministério da Saúde estimou que, em 2003, ocorreram aproximadamente 134.691 óbitos por câncer em todo o país. Segundo estimativas do Instituto Nacional do Câncer (INCA), em 2006 ocorreriam mais de 472.050 novos casos da doença (Ministério da Saúde 2005).

Além da ressecção cirúrgica do tumor, quando esta é possível, tratamentos adjuvantes, como quimioterapia e radioterapia, são amplamente utilizados no tratamento da doença. No entanto, devido à sua inespecificidade, tais tratamentos afetam também as células normais (FERRANTE et al. 1999) produzindo muitos efeitos colaterais. Desta forma, estratégias alternativas como a imunoterapia representam grandes promessas para um tratamento mais eficaz e específico contra o câncer.

A imunoterapia contra o câncer engloba diferentes modalidades terapêuticas que se dividem em imunização ativa ou passiva. Na imunização ativa, podemos citar: a injeção de peptídeos antigênicos, proteínas recombinantes ou vetores virais ou bacterianos contendo cDNAs correspondentes a antígenos tumorais; vacinas compostas por células dendríticas purificadas dos pacientes e incubadas com antígenos tumorais ou transfectadas com os genes que codificam estes antígenos;

tratamento com citocinas e a administração de substâncias inflamatórias. Tais estratégias levam ao aumento da resposta imune contra o tumor, no entanto, a grande dificuldade destas técnicas está no fato de que a indução da resposta imune deve ser forte o suficiente para erradicar todas as células tumorais o que nem sempre é alcançado (ABBAS et al. 2000).

Já a imunização passiva envolve a administração de células T que reconhecem especificamente antígenos presentes em células tumorais ou de anticorpos específicos para antígenos tumorais. Tal estratégia apresenta uma resposta rápida quando comparada com a imunização ativa, entretanto não pode ser considerada tão duradoura quanto esta. Dentre as metodologias utilizadas podemos citar a utilização de anticorpos monoclonais que podem ou não ser empregados sob a forma de imunocombinados. Além disso, outra técnica empregada é a administração de células ativadas *in vitro* onde leucócitos periféricos do paciente são coletados, estimulados *in vitro* por exposição a IL-2 e novamente reinfundidos ao paciente. Este tratamento dá origem a uma população de células, denominadas células LAK (*Lymphokine Activated Killer cells*), que exibe maior capacidade tumoricida. Posteriormente, é realizada a administração contínua de IL-2 o que leva à expansão *in vivo* destes linfócitos, que persiste enquanto IL-2 é administrada. Uma variação desta metodologia consiste em isolar linfócitos infiltrantes do tumor ao invés dos leucócitos. Vale citar que estudos realizados em pacientes com tumores metastáticos apresentaram resultados animadores (ABBAS et al. 2000).

## 1.1 IMUNOLOGIA DE TUMORES

Evidências de que o sistema imune humano poderia atuar de forma a restringir o crescimento de tumores foram obtidas por volta de 1890 quando o cirurgião William B. Coley observou que pacientes com câncer que contraíam infecções bacterianas apresentavam, em alguns casos, uma regressão significativa da massa tumoral devido à estimulação inespecífica do sistema imune. Em 1909, Paul Ehrlich propôs que o sistema imune reprimia o crescimento de carcinomas descrevendo a hipótese da imunovigilância, formalmente introduzida por Burnet e Thomas 50 anos mais tarde (DUNN et al. 2004). Esses achados clínicos preliminares foram, posteriormente, comprovados por estudos histológicos que demonstraram a presença de infiltrados mononucleares compostos por linfócitos T, células *natural killer* e macrófagos circundando muitos tumores, além da presença de proliferação linfocitária nos linfonodos que drenam locais de crescimento tumoral (ABBAS et al. 2000).

Atualmente, mais de cem anos após as primeiras observações, e frente aos inúmeros avanços obtidos na área da imunologia do câncer, não existem mais dúvidas de que o sistema imune humano seja capaz de reconhecer proteínas de origem tumoral e de que a imunovigilância possa limitar, até certo ponto, a cinética de crescimento de alguns tumores. Originalmente, a teoria da imunovigilância contra o câncer preconizava que esta fosse essencialmente uma função protetora exercida pela resposta imune adaptativa nos momentos iniciais da transformação celular. No entanto, estudos recentes têm demonstrado que o sistema imune protege o paciente contra o desenvolvimento de um tumor, mas devido à seleção de tumores com baixa

imunogenicidade, também é capaz de promover o crescimento tumoral (SHANKARAN et al. 2001). Dada a alta taxa de mitose das células tumorais e sua instabilidade genética, podem ocorrer mutações ou deleções de genes que codificam determinados antígenos tumorais. Se estes antígenos não forem necessários para o crescimento do tumor ou para a manutenção do fenótipo transformado, esta perda pode representar uma vantagem para a seleção das células antígeno-negativas. Assim, este duplo efeito do sistema imune no desenvolvimento de tumores levou alguns pesquisadores a proporem um modelo de edição imunológica (*immunoediting*) que compreende três fases: eliminação, equilíbrio e escape (DUNN et al. 2002, 2004).

A fase de eliminação representa o conceito original da imunovigilância do câncer e se a mesma for bem sucedida esta fase representa todo o processo de edição imunológica sem que ocorra a progressão para as fases subsequentes. A resposta inicia-se quando células do sistema imune reconhecem o tumor e as moléculas de interferon- $\gamma$  (IFN- $\gamma$ ) produzidas desencadeiam uma cascata de reações do sistema imune resultando na morte de células tumorais. As células tumorais mortas são processadas pelas células dendríticas que migram para o linfonodo e apresentam os antígenos tumorais para as células T CD4<sup>+</sup> e CD8<sup>+</sup>. Em seguida, células T tumor-específicas migram para o tumor reconhecendo e destruindo as células que expressam tais antígenos tumorais. Desta forma, a fase de eliminação deve ser um processo contínuo que precisa ser repetido a cada momento que surgem células tumorais antigenicamente distintas.

Já na fase de equilíbrio, a mais longa das três fases, o sistema imune e as variantes das células tumorais que sobreviveram à fase de eliminação entram em um

equilíbrio dinâmico. Embora a maioria das células tumorais seja destruída, novas variantes surgem carregando diferentes mutações que conferem às mesmas a capacidade de evadir o sistema imune. Os componentes do sistema imune exercem uma potente e contínua pressão seletiva sobre as células tumorais geneticamente instáveis e mutantes, suficiente para contê-las, porém não o bastante para eliminá-las completamente. Assim, o processo de equilíbrio resulta em uma nova população de clones tumorais com reduzida imunogenicidade, originada a partir de uma população heterogênea inicial que foi “esculpida” pelo sistema imune.

Desta forma, na fase de escape, as variantes das células tumorais selecionadas na fase de equilíbrio podem crescer em um ambiente imunologicamente intacto. Este escape dos mecanismos de defesa do organismo normalmente ocorre quando alterações genéticas e epigenéticas nas células tumorais conferem resistência à detecção e/ou eliminação pelo sistema imune, permitindo a expansão dos tumores e tornando-os clinicamente detectáveis.

Os tumores adquirem a capacidade de escapar dos mecanismos de defesa imunológica do paciente por diversos mecanismos. Provavelmente, um dos mais importantes é a diminuição ou perda da expressão de moléculas de HLA (*Human Leucocyte Antigen*) classe I nas células tumorais, o que dificulta o seu reconhecimento pelos linfócitos T citotóxicos (ABBAS et al. 2000). Além disso, também pode haver diminuição da expressão de  $\beta_2$ -microglobulina, ou ainda, de componentes da maquinaria de processamento de antígenos (CHEEVER et al. 1995). Dentre outros mecanismos também podemos citar a secreção pelas células tumorais de produtos que suprimem a resposta imune anti-tumoral (como o TGF- $\beta$  e IL-10) e o não reconhecimento das células tumorais pelos linfócitos T citotóxicos devido à

ausência de expressão de moléculas co-estimulatórias (como B7) ou de HLA classe II, necessárias para a ativação de linfócitos *T-helper*.

Frente a todos esses obstáculos, torna-se importante o desenvolvimento de terapias que se baseiam na estimulação do sistema imune para que seja capaz de reconhecer as células tumorais e, desta forma, combater o tumor de modo eficiente. O desenvolvimento de pesquisas no campo da imunologia de tumores engloba desde o estudo dos mecanismos imunológicos que destroem as células tumorais até o desenvolvimento de abordagens imunológicas para detectar, diagnosticar e tratar os inúmeros tipos de câncer. Nesse sentido, uma das principais atividades de pesquisa nesta área consiste na identificação de antígenos tumorais capazes de induzir uma resposta imune em pacientes com câncer e que possam, dessa maneira, ser utilizados como alvo na imunoterapia.

## 1.2 ANTÍGENOS TUMORAIS

Na década de 90, VAN DER BRUGGEN et al. (1991) descreveram pela primeira vez a clonagem de um antígeno tumoral, denominado MAGE-1 (*Melanoma Antigen 1*), o qual era capaz de desencadear uma resposta imune específica dependente de linfócitos T citotóxicos em pacientes com melanoma. Mais de uma década após esta descoberta, o fato das células tumorais expressarem antígenos que estimulam a resposta imune tem sido claramente demonstrado tanto em modelos experimentais como em pacientes com câncer. Desta maneira, uma grande variedade de antígenos tumorais reconhecidos por linfócitos B e T tem sido identificada (SAHIN et al. 1997; OLD e CHEN 1998).

Os antígenos tumorais são definidos como proteínas com padrão de expressão predominantemente em tumores e que necessariamente são reconhecidas pelo sistema imune. Tais antígenos são, atualmente, agrupados nas seguintes categorias:

1. Antígenos codificados por genes mutados. Ex: p53, CDK4 e MUM-1 (COULIE et al. 1995; WOLFEL et al. 1995; SCANLAN et al. 1998);
2. Proteínas de vírus oncogênicos. Ex: oncoproteínas de HPV e EBV (LENNETTE et al. 1995; TINDLE 1996);
3. Antígenos superexpressos em tecidos tumorais. Ex: HER2/neu, anidrase carbônica e eIF-4gamma (FISK et al. 1995; SAHIN et al. 1995; BRASS et al. 1997);
4. Antígenos *cancer/testis* os quais estão silenciados em todos os tecidos normais, exceto testículo, mas que são expressos em diferentes tipos de tumores. Ex: MAGE-A1, NY-ESO-1 e SSX-1 (VAN DER BRUGGEN et al. 1991; CHEN et al. 1997; GURE et al. 1997);
5. Antígenos de diferenciação, expressos em tipos celulares específicos e em seus respectivos tumores. Ex: Melan A, gp100 e Tirosinase de melanócitos (BRICHARD et al. 1993; KAWAKAMI et al. 1994).

Dentre estas classes, valem ressaltar os antígenos com restrito padrão de expressão, os antígenos *cancer/testis* (antígenos CT) e os antígenos de diferenciação, por serem considerados candidatos promissores para o desenvolvimento de imunoterapia contra o câncer.

### 1.3 ANTÍGENOS CT

Os antígenos CT (CTs) constituem até o momento um grupo composto por 93 membros organizados em 47 famílias gênicas e apresentando como principal característica um perfil de expressão em tecidos normais restrito a tecidos gametogênicos e em tecidos tumorais abrangendo um grande número de tipos histológicos (CHEN et al. 1998). Dentre os tecidos normais adultos, vale ressaltar que estes antígenos estão expressos exclusivamente nas células germinativas de testículo. Além disso, dentre outras características em comum, podemos citar a localização de mais de 50% dos mesmos no cromossomo X e a presença de extensas famílias gênicas.

Apesar desta classe de antígenos apresentar expressão predominante em tecidos gametogênicos, é importante ressaltar que alguns de seus membros encontram-se expressos em outros tecidos normais, porém, com um nível bastante inferior ao encontrado nas células gametogênicas e tumores (CREW et al. 1995; CHEN et al. 1997; LETHE et al. 1998). Assim, a partir desta observação surgiram subcategorias como os antígenos CTB (*Cancer-Testis-Brain*) e os antígenos ChT (*Choriotrofolasto-testis*) (DALMAU et al. 1999; SCANLAN et al. 2002a).

Ainda com relação à expressão dos CTs, outras características quanto ao padrão de expressão vêm sendo descritas na literatura. A expressão da proteína no tumor é heterogênea podendo estar restrita a uma pequena quantidade de células em determinadas regiões ou de maneira dispersa, mas também presente em todo o tumor (JUNGBLUTH et al. 2000, 2001). A ativação da expressão gênica no tumor pode estar associada à hipometilação da região promotora e/ou acetilação de histonas

(SCANLAN et al. 2002a; GRUNWALD et al. 2005). E, além disso, ainda que controversa, parece haver uma correlação entre a expressão desses antígenos e a progressão tumoral. Enquanto GOYDOS et al. (2001) detectaram a expressão do antígeno NY-ESO-1 em 10% das amostras de melanoma primário (n=20) e em 47% das lesões metastáticas (n=32), PENG et al. (2005) encontraram uma maior frequência de expressão deste antígeno em carcinomas hepatocelular estadio I e II do que nos estadios III e VI. Já BARROW et al. (2006) não demonstrou diferença de expressão do NY-ESO-1 entre amostras de melanoma em diferentes estadios. Este mesmo grupo, no entanto, demonstrou que a expressão de MAGE-A1 e MAGE-A4 foi maior dentre as 90 amostras de melanoma avançado, 51% e 44% respectivamente, do que entre as 251 amostras de melanoma primário, 20% e 9% respectivamente. Tal correlação entre a expressão de membros da família MAGE e melanomas avançados também foi encontrada por BRASSEUR et al. (1995). Neste caso, a expressão dos genes MAGE-1, 2, 3 e 4 foi detectada, respectivamente, em 16%, 41%, 36% e 11% dos casos de melanoma primário (n=100) e em 48%, 70%, 76% e 22% de melanomas metastáticos (n=145).

Com exceção de alguns CTs que apresentam um papel importante no desenvolvimento das células germinativas, a função da maioria destes antígenos no tecido normal ainda é desconhecida (SIMPSON et al. 2005). Ainda não se sabe se a expressão destes antígenos nos tumores contribui para o fenótipo maligno ou apenas reflete a reativação por desmetilação ou alterações na cromatina de genes normalmente silenciados, o que poderia fazer com que fossem eliminados pela pressão do sistema imune.

Dentre os antígenos CT com função conhecida em tecido normal citamos vários exemplos. O SCP-1 faz parte do complexo sinaptoneal envolvido no pareamento dos cromossomos homólogos durante a meiose (MEUWISSEN et al. 1992), enquanto o OY-TES-1 é o precursor da proteína sp32 de ligação a acrosina, envolvida no empacotamento desta na cabeça do espermatozóide (BABA et al. 1994). Já o CT15/Fertilina $\beta$  é um membro da família protéica ADAM (*a desintegrin and metalloproteinase protein*), sendo esta formada por proteases e/ou moléculas de adesão de superfície celular. Acredita-se que o CT15 esteja envolvido na interação entre o óvulo e o espermatozóide através da ligação com  $\alpha_6\beta_1$  integrina (VIDAEUS et al. 1997). Além disso, o SPAG9 demonstrou estar envolvido na via de sinalização de MAPK interagindo com moléculas JNK (*c-Jun NH2-terminal protein kinase*) e a proteína Sp17 media interações célula-célula, como também apresenta um papel na motilidade dos espermatozoides (SURI 2006). Já o BORIS está envolvido no cancelamento do *imprinting* durante o final da mitose na espermatogênese.

Por outro lado, alguns estudos indicam que a expressão de determinados antígenos poderia ter um papel fundamental na tumorigênese. Um estudo realizado por MONTE et al. (2006) demonstrou que proteínas MAGE-A podem recrutar histonas deacetilases para o p53 causando hipoacetilação de histonas próximas aos sítios de ligação da proteína p53 ao DNA e hipoacetilação da própria proteína p53. Por inibirem a ativação de p53, há uma redução na resposta ao dano de DNA dependente de p53 em células que apresentam um aumento de expressão de proteínas MAGE-A. Além disso, ainda com relação à família MAGE, estudos funcionais realizados com membros ubiquamente expressos (Necdina e MAGE-D1) sugerem que certos membros da família estejam envolvidos no controle do ciclo celular e

apoptose, funções importantes no processo de formação e desenvolvimento do câncer.

Devido à imunogenicidade dos antígenos CT em pacientes com câncer e ao seu padrão de expressão restrito às células germinativas presentes nos testículos, estes antígenos podem ser considerados candidatos ideais para o desenvolvimento de imunoterapia contra o câncer. Isso porque o testículo é considerado um órgão imunologicamente protegido já que suas células não expressam moléculas de HLA classe I e, desta forma, não são capazes de apresentar antígenos para os linfócitos T (LUCAS et al. 1999). Além disso, o bloqueio físico imposto pela barreira hemato-testicular impede a passagem de anticorpos para tal órgão. Sendo assim, abordagens imunoterápicas que tenham como alvo estes antígenos não reagem, teoricamente, com nenhum tecido normal atacando especificamente as células tumorais.

No entanto, os antígenos CT também possuem algumas características relacionadas ao seu padrão de expressão nas células tumorais que podem dificultar a aplicação da imunoterapia contra o câncer (SCANLAN et al. 2002a). Além da expressão heterogênea da proteína no tumor, podemos também destacar: (a) a frequência de expressão de um antígeno CT específico varia entre os diferentes tipos de tumor, sendo o melanoma geralmente de maior frequência, seguido por tumores de bexiga, pulmão, mama e próstata, (b) a frequência de expressão dos diferentes CTs em um determinado tipo de tumor é altamente variável, e, além disso, (c) há uma tendência da expressão destes antígenos estar agrupada, ou seja, certos tumores expressam múltiplos CTs enquanto outros não apresentam expressão de antígenos CT.

Uma das possibilidades para contornar o problema da heterogeneidade de expressão destes antígenos em tumores é o desenvolvimento de vacinas polivalentes contendo diferentes epítomos derivados do maior número possível de antígenos CT. Desta forma, a identificação de novos antígenos CT, assim como a caracterização dos seus respectivos padrões de expressão nos diferentes tipos de tumores, é um passo fundamental para o sucesso da imunoterapia contra o câncer.

### **1.3.1 Técnicas imunológicas usadas para identificação de antígenos CT**

A identificação de antígenos tumorais iniciou-se na década de 60 com a análise de soro heteroimune, obtido a partir de coelhos e outros animais imunizados com tumores humanos (DAY 1965). O desafio, geralmente não alcançado, era remover os anticorpos que reagiam contra antígenos de tecidos normais os quais diferiam entre as espécies. Mesmo apresentando tais limitações, dois antígenos importantes foram identificados com esta metodologia: a  $\alpha$ -fetoproteína e o antígeno carcinoembrionário (CEA) (ABELEV et al. 1963; GOLD e FREEDMAN 1965).

Grandes avanços na caracterização de antígenos tumorais foram obtidos com o desenvolvimento da tecnologia de hibridomas e a produção de anticorpos monoclonais, no entanto, a identificação de um real antígeno tumoral humano não era tão eficiente e grande parte dos candidatos identificados acabava sendo um antígeno de diferenciação presente também no tecido normal. Além disso, após a identificação dos antígenos tumorais, ainda era necessário saber se o sistema imune do paciente os reconhecia como tal.

Para solucionar esta questão foi desenvolvida uma técnica imunológica denominada *Autologous Typing*. O objetivo desta técnica era restringir a análise aos

reagentes autólogos (células tumorais, soro e linfócitos do mesmo paciente) para eliminar a contribuição dos aloantígenos observada nas reações com os anticorpos monoclonais. Para tanto, o cultivo de células tumorais era necessário, o que limitou a análise a tipos de tumores que poderiam ser adaptados a cultura *in vitro* com mais facilidade, tais como: leucemia, linfomas, melanoma, tumor renal e cerebral (CAREY et al. 1976; PFREUNDSCHUH et al. 1978; UEDA et al. 1979). Embora muito promissora, a técnica revelou que apenas uma pequena parte dos pacientes desenvolve anticorpos autólogos específicos para antígenos tumorais de superfície e que, mesmo quando presentes, os anticorpos são geralmente insuficientes para a caracterização molecular dos antígenos identificados.

Somente em 1995, com o desenvolvimento de uma técnica inovadora, a identificação de antígenos tumorais tornou-se mais eficiente. A técnica, denominada SEREX (*Serological Analysis of Recombinant tumor cDNA Expression Libraries*), consiste em isolar antígenos tumorais que elicitam resposta imune com altos títulos de imunoglobulina G (IgG) em pacientes com câncer (SAHIN et al. 1995). Para tal, é feita a biópsia do tumor, com posterior extração de RNA e síntese de cDNA (DNA complementar). Em seguida, este cDNA é clonado em vetor de expressão e a biblioteca resultante é rastreada com o soro do paciente em busca de clones reativos. Os clones positivos são, então, seqüenciados e verifica-se se há similaridade com algum gene já descrito. Paralelamente, é feita também uma avaliação da expressão destes candidatos em tecidos normais e tumorais e, ainda, a medida da imunogenicidade dos mesmos, através da determinação da frequência de anticorpos nos soros de indivíduos sadios e de pacientes com o mesmo tipo de tumor (OLD e CHEN 1998). Vale ressaltar que esta técnica é capaz de verificar a resposta imune

humoral contra antígenos tumorais intracelulares até então pouco explorada na sorologia do câncer. Além disso, SAHIN et al. (1997) identificaram os antígenos MAGE-A1 e Tirosinase originalmente clonados pela técnica de *Autologous Typing* indicando que SEREX também pode detectar antígenos tumorais que elicitam resposta imune mediada por linfócitos T citotóxicos.

### **1.3.2 Técnicas não imunológicas usadas para identificação de antígenos CT**

Devido ao restrito padrão de expressão gênica observado nos antígenos CT, metodologias “não-imunológicas”, baseadas na análise do perfil de expressão de diferentes transcritos, também têm sido utilizadas na identificação desses antígenos. Técnicas como a análise representativa diferencial (*Representational Difference Analysis – RDA*), o *differential display* e a hibridização em *arrays* de cDNA e oligonucleotídeos vêm sendo utilizadas com sucesso na identificação de transcritos que são expressos em testículo normal e diferentes tipos de tumores. Entretanto, técnicas complementares, tais como RT-PCR (*Reverse Transcriptase – Polimerase Chain Reaction*) e *Northern-blot* são necessárias para a confirmação do padrão de expressão dos genes encontrados. Com a aplicação destas técnicas, vários antígenos CT foram identificados, tais como LAGE-1, CT10, CTp11, SPANX-C1, MMA-1A, DBL/MCF2 (LETHE et al. 1998; ZENDMAN et al. 1999, 2002; DE WIT et al. 2002; SEGAL et al. 2005).

Ainda tomando por base o perfil de expressão gênica dos CTs e frente ao grande número de seqüências expressas disponibilizadas em bancos de dados públicos sob a forma de ESTs (*Expressed Sequence Tags – seqüências parciais de uma das extremidades de um clone de cDNA*) está cada vez mais acessível e prática

a utilização de ferramentas de bioinformática para a avaliação *in silico* do padrão de expressão de qualquer transcrito. A produção de ESTs teve suas origens em 1980, mas somente no início da década de noventa elas começaram a ser produzidas em larga escala (ADAMS et al. 1995; KORENBERG et al. 1995; HILLIER et al. 1996). Até 27 de abril de 2007, 8.109.025 ESTs humanas obtidas a partir de diversos órgãos e tecidos foram depositadas no *GenBank*, derivadas, principalmente, do *Merck Gene Index Project* (ECKMAN et al. 1998; WILLIAMSON 1999), do *Cancer Genome Anatomy Project* (CGAP) (STRAUSBERG et al. 2000) e do Projeto Genoma Humano do Câncer (HCGP) (CAMARGO et al. 2001). A disponibilidade desse grande número de seqüências parciais possibilita a reconstituição da seqüência dos transcritos através de montagens com o auxílio de programas de computador e a determinação *in silico* do padrão de expressão através da informação do tecido de origem da biblioteca de cDNA (REYMOND et al. 2002).

Os programas de computador como, por exemplo, CAP3 (HUANG e MADAN 1999) e TIGR *Assembler* (QUACKENBUSH et al. 2000) identificam trechos de similaridade e sobreposição entre a seqüência das ESTs, produzindo uma seqüência consenso (*contig*) que representa o transcrito. Alguns grupos já utilizam essa estratégia para a construção de bancos de dados representando os diferentes transcritos humanos; destacando-se o Unigene (SCHULER 1997) e o TIGR Gene Index (QUACKENBUSH et al. 2000). No entanto, a grande limitação dessa metodologia diz respeito aos artefatos gerados durante o processo de montagem das seqüências. Os artefatos ocorrem principalmente devido a: i) baixa qualidade das seqüências, que dificulta a determinação precisa de sobreposições, ii) presença de formas alternativas de *splicing*, que pode dificultar o agrupamento de seqüências

derivadas de um mesmo gene (GALANTE et al. 2004), iii) existência de famílias gênicas com grande similaridade em nível de nucleotídeos e cujas seqüências podem ser artificialmente agrupadas em um mesmo consenso (QUACKENBUSH et al. 2000). Outro problema detectado nos bancos de dados de seqüências expressas é a presença de uma significativa porção de seqüências derivadas de contaminação das bibliotecas de cDNA com DNA genômico.

No caso dos CTs, tendo conhecimento da origem tecidual das ESTs correspondentes a um dado transcrito é possível selecionar candidatos que possuam seqüências expressas derivadas apenas de bibliotecas de cDNA de testículo normal e de qualquer tecido tumoral. Essa metodologia permite uma seleção rápida, simples e barata de um número significativo de candidatos sem a necessidade da utilização de técnicas imunológicas convencionais. Após essa análise preliminar *in silico*, os candidatos selecionados podem ser validados experimentalmente através da avaliação do padrão de expressão gênica e do perfil de imunogenicidade utilizando soros de pacientes.

Em um trabalho pioneiro utilizando dados de ESTs, SCANLAN et al. (2002b) foram capazes de identificar e validar experimentalmente três novos antígenos CT através da seleção de transcritos no banco de dados do Unigene. Nesse trabalho foram selecionados transcritos cujas seqüências expressas correspondentes eram derivadas somente de bibliotecas de cDNA de testículo normal e de tecido tumoral. Utilizando essa metodologia, foram selecionados 1325 diferentes transcritos que apresentavam seqüências expressas de testículo e tecido tumoral, sendo que 61 representavam transcritos conhecidos e 1264 transcritos parciais ainda não caracterizados. Grande parte dos transcritos parciais estava representada por um

número reduzido de ESTs (na sua maioria apenas uma EST) o que sugere que muitos desses candidatos não correspondam realmente a transcritos humanos e que tenham sido derivados de contaminação com DNA genômico nas bibliotecas de cDNA. A seleção desses transcritos artefatuais aumenta o trabalho de validação experimental e diminui a eficiência de validação. Neste contexto, do total de 1264 transcritos parciais, 73 foram selecionados para análise do perfil de expressão e apenas três novos CTs (CT15, CT16 e CT17) foram identificados (SCANLAN et al. 2002b).

Além dos problemas relacionados com eventuais contaminações de bibliotecas de cDNA com DNA genômico, esse tipo de análise também apresenta limitações relacionadas aos artefatos gerados durante o processo de montagem das seqüências expressas já discutidos anteriormente. Dentre os vários problemas existentes, os artefatos de montagem associados à presença de famílias gênicas são de especial relevância. Hoje sabemos que grande parte dos antígenos CT conhecidos constituem extensas famílias gênicas (ARADHYA et al. 2001) e que nem todos os membros dessas famílias apresentam o padrão de expressão característico dos CTs pois muitos deles apresentam ESTs derivadas de tecidos normais. Devido à alta similaridade ao nível de nucleotídeos existente entre diferentes membros de uma família de CTs, é possível que ESTs correspondentes a um determinado membro sejam erroneamente agrupadas junto com ESTs de outro membro. Esse fato dificulta a determinação do real padrão de expressão com bases em dados de ESTs, podendo levar a exclusão de candidatos que apresentem alta similaridade com outros transcritos expressos também em tecidos normais. Neste contexto, SCANLAN et al. (2002b) verificaram também que a análise executada falhou na identificação de nove antígenos CT conhecidos, todos eles membros de extensas famílias gênicas. É

importante ressaltar neste momento que pelo menos 5 desses antígenos conhecidos apresentam ESTs derivadas de bibliotecas de tecidos normais.

A partir deste trabalho, vários outros estudos têm se baseado em banco de dados públicos para a identificação de novos antígenos CTs (DONG et al. 2003, 2004). Alguns trabalhos estão partindo da análise de expressão de genes previamente conhecidos por apresentarem, a partir de bancos de dados de ESTs, uma expressão específica em testículo e, desta forma, são capazes de caracterizá-los como novos antígenos CTs (DONG et al. 2003). Já o estudo desenvolvido por CHEN et al. (2005), por exemplo, baseou-se em uma metodologia diferente. Foram selecionados, a partir do banco de dados de MPSS (*Massively Parallel Signature Sequencing*), genes que apresentavam expressão pelo menos duas vezes maior em testículo quando comparada com outros tecidos normais. De um total de 1056 genes selecionados, 15 genes correspondiam a antígenos CTs já conhecidos. Assim, 1041 genes foram analisados com base em bancos de dados públicos de ESTs sendo que foram considerados candidatos a antígenos CTs aqueles genes cujas ESTs não correspondiam a mais do que dois tecidos normais. Baseando-se neste critério foi possível identificar 202 genes, sendo que 36 correspondiam a genes sem intron e foram excluídos da validação experimental. Por fim, a partir da análise do perfil de expressão destes candidatos foi possível identificar cinco novos antígenos CTs (CHEN et al. 2005).

A exemplo destes trabalhos, este projeto tem como objetivo identificar novos antígenos CT através de uma análise preliminar *in silico*, seguida de validação experimental dos candidatos selecionados através da análise do perfil de expressão e imunogenicidade. Para a análise *in silico*, nós utilizaremos o Banco de Dados do

Transcriptoma Humano construído e disponibilizado pelo Laboratório de Biologia Computacional do Instituto Ludwig.

#### 1.4 BANCO DE DADOS DO TRANSCRIPTOMA HUMANO

O Banco de Dados do Transcriptoma Humano contém informações sobre as regiões transcritas do genoma humano. Tais regiões foram definidas *in silico* através do alinhamento entre seqüências de cDNA (ESTs e clones de cDNA completamente seqüenciados) e seqüências genômicas. O alinhamento entre as seqüências transcritas e a seqüência genômica foi feito com o programa BlastN sendo, para tanto, utilizada a montagem do genoma humano fornecida pelo Centro Nacional de Informação em Biotecnologia (NCBI) como fonte da informação genômica e seqüências transcritas humanas de ORESTES provenientes do HCGP e de mRNA e ESTs presentes no banco de dados GenBank disponibilizado pelo Centro Nacional de Informação em Biotecnologia (NCBI) (<http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html>).

Desta forma, com base neste banco de dados é possível delimitar todas as regiões transcritas para um determinado clone genômico, assim como extrair várias informações, tais como: a identidade das ESTs alinhadas, a origem tecidual e o método utilizado na geração de cada uma das ESTs além das coordenadas de alinhamento das mesmas. As coordenadas referentes aos exons e introns definidas através dos alinhamentos foram indexadas tanto em relação à seqüência genômica, quanto ao transcrito. Todas as informações sobre o mapeamento foram armazenadas em um banco de dados relacional MySQL.

O agrupamento das seqüências transcritas em *clusters* foi feito com base nas coordenadas dos exons definidas na seqüência genômica, de forma que dois transcritos foram considerados parte de um mesmo *cluster* (e, portanto, derivadas de um mesmo transcrito) quando compartilhavam as coordenadas de pelo menos um exon. Utilizando esta estratégia foram definidos 318.275 *clusters*, dos quais 21.306 incluíam seqüências de cDNA completas. Essa estratégia de agrupamento evita parcialmente os artefatos gerados através do alinhamento direto de seqüências transcritas, geralmente associados à baixa qualidade das seqüências, à presença de extensas famílias gênicas no genoma e às formas alternativas de *splicing*.

O alinhamento das seqüências expressas contra a seqüência genômica fornece ainda uma maneira indireta de eliminar o problema relacionado à presença de contaminação das seqüências expressas com DNA genômico. Isso porque o alinhamento entre seqüências expressas e seqüências genômicas geralmente apresentam interrupções devido à presença de introns na seqüência genômica os quais são removidos das seqüências expressas após a ocorrência do *splicing* (GALANTE et al. 2004). Assim sendo, a presença de interrupções nos alinhamentos podem ser utilizadas para indicar se uma determinada seqüência expressa apresenta ou não *splicing*. A presença de *splicing* nas seqüências expressas catalogadas em nosso banco de dados, por sua vez, pode ser utilizada como um critério de seleção de candidatos para validação, já que a presença de *splicing* constitui uma evidência concreta de transcrição eliminando a hipótese de contaminação com DNA genômico.

Para a procura por novos CTs foi construído em nosso laboratório um *pipeline* que realizou automaticamente buscas no Banco de Dados do Transcriptoma. Para a construção do *pipeline* foi utilizado como base o comportamento quanto ao

perfil de expressão de CTs já conhecidos e bem caracterizados e que se encontram em nosso banco de dados. Foram feitas análises quanto à distribuição de ESTs de diferentes tecidos de cada um destes CTs com o objetivo de gerar um *pipeline* mais próximo da realidade quanto à seleção de candidatos a novos antígenos CT.

Desta forma, acreditamos que a utilização do Banco do Transcriptoma Humano pode, em parte, ter eliminado os problemas relacionados com a contaminação de DNA nas seqüências de ESTs, assim como minimizado os artefatos gerados durante o processo de montagem e agrupamento das seqüências expressas. Assim sendo, o Banco do Transcriptoma Humano permitiu uma seleção de candidatos mais rápida e eficiente para a fase de validação experimental.

## ***OBJETIVOS***



## 2 OBJETIVOS

1. Validar experimentalmente uma estratégia computacional desenvolvida para a identificação de genes com padrão de expressão característico de antígenos CTs e que leva em consideração para a seleção de candidatos o tecido de origem das seqüências expressas (ESTs) correspondentes a um determinado transcrito.
2. Validar experimentalmente o perfil de expressão em tecidos normais, linhagens tumorais e amostras de diferentes tipos de tumores de genes candidatos selecionados computacionalmente.
3. Produzir a proteína recombinante correspondente a cada um dos antígenos CT validados segundo o padrão de expressão e que apresentem uma seqüência completa depositada em bancos de dados públicos.
4. Avaliar a presença de anticorpos específicos contra as proteínas recombinantes dos CTs validados em plasma de pacientes com câncer através de *Immunoblotting*.
5. Gerar a seqüência completa de novos genes validados como um novo antígeno CT segundo o padrão de expressão e que estejam representados exclusivamente por ESTs em nosso banco de dados.

## ***MATERIAIS E MÉTODOS***



### 3 MATERIAIS E MÉTODOS

#### 3.1 IDENTIFICAÇÃO DE NOVOS ANTÍGENOS CT

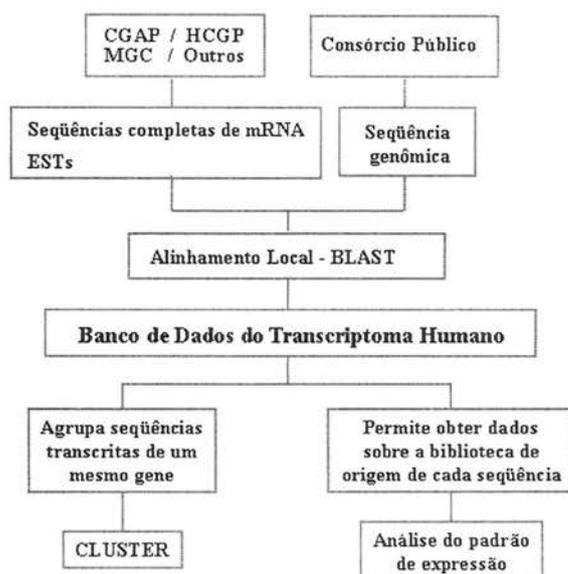
##### 3.1.1 Construção do Banco de Dados do Transcriptoma Humano (Map4)

O banco de dados utilizado para a identificação de candidatos a antígenos CT foi construído pelo Laboratório de Biologia Computacional do Instituto Ludwig e contém informações relacionadas ao alinhamento entre seqüências expressas (mRNA e ESTs) e a seqüência genômica humana (Figura 1).

Para a construção do banco Map4, os dados correspondentes às seqüências expressas foram obtidos a partir de diversas fontes: seqüências de ORESTES provenientes do HCGP e seqüências de mRNA e ESTs depositadas no GenBank disponibilizado pelo Centro Nacional de Informação em Biotecnologia (NCBI) (<http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html>). As 3.475.517 seqüências expressas, sendo 52.903 correspondentes a seqüências de mRNA e 3.422.614 correspondentes a ESTs, foram filtradas quanto à presença de contaminação com seqüências de vetores, bactérias e fungos e os elementos repetitivos foram mascarados utilizando o *software* PFP (Paracel, Pasadena, CA).

O mapeamento das seqüências expressas em relação à seqüência genômica foi feito com base na similaridade entre as mesmas utilizando o programa BLASTN (NCBI versão 2.0.12), também disponibilizado pelo NCBI. Os dados gerados pelo BLASTN foram filtrados para eliminar todos os alinhamentos que não continham uma região alinhada com cobertura de pelo menos 45% do tamanho total da EST ou

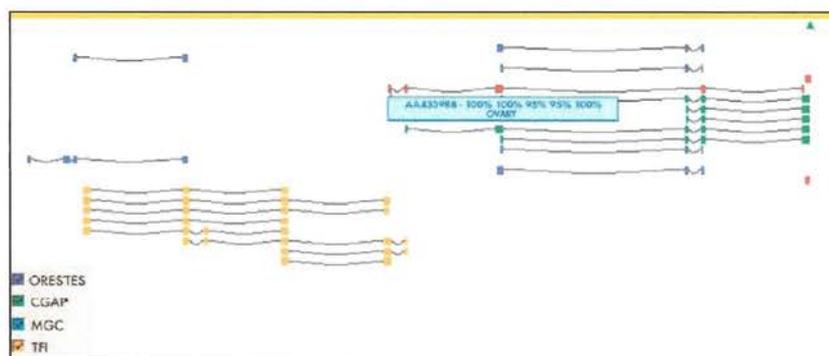
55% do tamanho total de uma seqüência completa de mRNA, com no mínimo 93% de identidade (GALANTE et al. 2004). Os dados dos alinhamentos entre as seqüências expressas (ESTs e mRNAs) e a seqüência genômica foram armazenados em um banco de dados relacional MySQL.



**Figura 1** - Esquema representando a construção do Banco de Dados do Transcriptoma Humano (Map4).

Os dados armazenados neste banco de dados relacional foram utilizados para agrupar seqüências expressas derivadas de um mesmo gene em *clusters* de seqüências. Para o agrupamento foram levadas em consideração as coordenadas dos alinhamentos das seqüências expressas em relação à seqüência genômica de forma que duas seqüências expressas foram consideradas parte de um mesmo *cluster* quando as coordenadas dos alinhamentos de pelo menos um exon fossem comuns às duas seqüências (Figura 2). No total foram gerados 318.275 *clusters* sendo que

21.306 correspondiam a *clusters* compostos por pelo menos uma seqüência completa de mRNA.



**Legenda:** Os exons simbolizam as linhas grossas e os introns as linhas finas. A seqüência genômica está representada em amarelo e as seqüências expressas são representadas em diferentes cores de acordo com o projeto de origem (MGC, ORESTES, CGAP ou TFI). Para cada seqüência transcrita é possível obter os dados correspondentes à biblioteca de origem assim como o alinhamento de cada exon com relação à seqüência genômica.

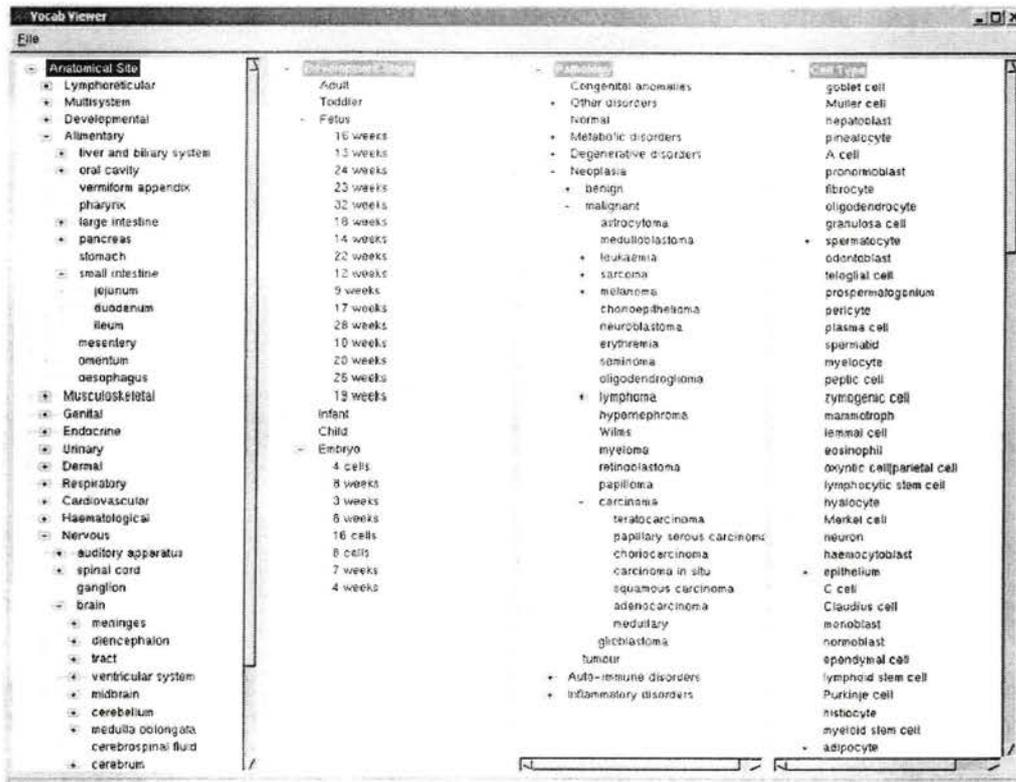
**Figura 2** - Esquema representativo de um *cluster* presente no Banco de Dados Map4.

Assim, através de consultas no banco Map4 é possível fazer um levantamento de todas as seqüências expressas que alinham em um determinado clone genômico. Informações adicionais sobre a orientação da seqüência (5' ou 3'), origem tecidual e coordenadas dos alinhamentos também podem ser obtidas. De especial importância para o nosso trabalho são as informações referentes à origem tecidual das seqüências expressas, pois a partir dessas informações é possível inferir o padrão de expressão de um determinado transcrito. As informações sobre a origem tecidual das seqüências expressas foram obtidas diretamente do dbEST nos campos *tissue* e *organ* presentes na descrição de cada biblioteca de cDNA. Como não existe um campo específico para a classificação da biblioteca em normal ou tumoral foram classificadas como tumorais todas as bibliotecas que nos dados de anotação de tecido ou órgão apresentaram as palavras *tumor*, *cancer*, *leukemia* ou 'oma' no final da

FUNDAÇÃO ANTÔNIO BRUDENTE  
*Almeida*  
 Ana Maria Rodrigues Alves Kuninhan  
 Diretora de Escola  
 RG 8.080.138

descrição, como por exemplo adenoma, carcinoma e glioblastoma. Todas as outras bibliotecas foram classificadas como derivadas de tecidos normais.

Em 2003, KELSO et al. publicaram um trabalho propondo um novo sistema para a anotação das bibliotecas de cDNA disponibilizadas no domínio público, o eVOC. O sistema consistia em quatro ontologias (sistema anatômico, tipo celular, patologia e estágio de desenvolvimento) interligadas e capazes de classificar uma biblioteca com precisão (Figura 3). Desta forma, programas computacionais foram desenvolvidos para identificar os termos específicos de cada biblioteca em cada uma das quatro ontologias e, em seguida, foi realizada uma inspeção manual das anotações para complementar os dados encontrados. Quando uma informação de determinada biblioteca não era descrita na ficha descritiva disponibilizada por bancos de dados públicos, os pesquisadores responsáveis pela submissão das seqüências eram contatados para se obter mais informações. Assim sendo, os dados de anotação gerados pelo eVOC foram adicionados ao Map4 complementando as informações sobre a origem tecidual das seqüências expressas já presentes no banco.



**Legenda:** Para cada ontologia (sistema anatômico, estágio de desenvolvimento, patologia e tipo celular) foram definidos termos formando árvores hierárquicas

**Fonte:** KELSO et al. (2003)

**Figura 3** – Visualização das quatro ontologias do eVOC utilizadas na anotação das bibliotecas de cDNA disponibilizadas pelo GenBank.

### 3.1.2 Construção do Banco de Dados Map4\_best hits (Map4\_bh)

O Banco de Dados Map4\_best hits (Map4\_bh) foi criado a partir do banco Map4 para solucionar o problema referente aos múltiplos alinhamentos de seqüências expressas no genoma humano. Desta forma, o Map4\_bh contém apenas o melhor alinhamento de cada seqüência expressa no genoma humano sendo que este foi escolhido com base na extensão do alinhamento em relação ao tamanho total da

seqüência multiplicado pela identidade média dos alinhamentos no caso de seqüências com mais de um exon.

Como o Map4\_bh difere do Map4 apenas nos critérios de alinhamento das seqüências expressas no genoma humano, o número de seqüências expressas presentes no banco é o mesmo. No entanto, o número total de *clusters* foi reduzido para 276.586 sendo que 17.788 correspondem a *clusters* compostos por pelo menos uma seqüência completa de mRNA (Tabela 1).

**Tabela 1** – Comparação entre o número de seqüências e *clusters* presentes nos bancos de dados Map4 e Map4\_bh.

	<i>Map4</i>	<i>Map4_bh</i>
Número total de seqüências expressas	3.475.517	3.475.517
Número total de <i>clusters</i>	318.275	276.586
<i>Clusters</i> compostos por pelo menos uma seqüência completa de mRNA	21.306	17.788

### 3.1.3 Construção do pipeline de seleção de novos antígenos CT

Para a procura por genes candidatos com padrão de expressão característico de antígenos CT, o técnico em bioinformática Fernando Camargo Filho construiu uma série de programas, utilizando a linguagem Perl, que constituíram o *pipeline* de seleção dos candidatos. Este *pipeline* realizou, automaticamente, buscas no banco de dados Map4\_bh utilizando como critério de seleção o tecido de origem de cada uma das ESTs de um determinado *cluster*. Mais especificamente, foram selecionados no Map4\_bh *clusters* constituídos por seqüências expressas derivadas de bibliotecas de testículo e/ou tecido tumoral. Uma vez que a presença de *splicing*, caracterizada por uma interrupção no alinhamento entre a seqüência expressa e a genômica, permite excluir a possibilidade de contaminação com DNA genômico nas bibliotecas de

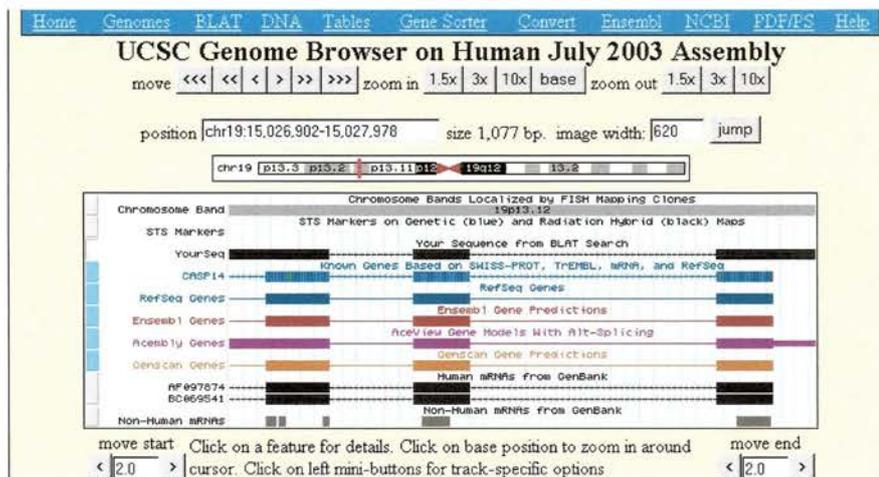
cDNA, a mesma também foi utilizada como critério de seleção dos genes candidatos a novos antígenos CT.

#### 3.1.4 Inspeção manual dos candidatos selecionados

A inspeção manual dos *clusters* selecionados automaticamente pelo *pipeline* foi realizada com o intuito de confirmar a informação obtida a partir do banco de dados sobre a presença de *splicing* nas ESTs uma vez que artefatos de alinhamento, causados por erros de seqüenciamento e problemas na montagem da seqüência genômica, podem levar a ocorrência de interrupções no alinhamento as quais podem ser interpretadas equivocadamente como um “falso *splicing*”. A ocorrência de *splicing* foi confirmada manualmente através da identificação dos sítios aceptores e doadores de *splicing* conservados (GT/AG) na seqüência genômica que flanqueava um exon. Além disso, dentre os *clusters* selecionados no Map4\_bh alguns já correspondiam a antígenos CT conhecidos. Desta forma, durante o processo de inspeção manual, *clusters* candidatos que correspondiam a antígenos CT já descritos foram excluídos do processo de validação experimental.

Para a análise manual de cada *cluster* candidato, foi escolhida como seqüência referência a seqüência expressa com o maior tamanho. As seqüências referências foram novamente alinhadas contra a seqüência do genoma humano disponibilizada pela Universidade de Santa Cruz – Califórnia (UCSC – montagem de julho de 2003) (<http://www.genome.ucsc.edu>), utilizando o programa BLAT. O resultado do alinhamento pode ser visualizado através de uma interface gráfica disponibilizada pelo programa que permite a verificação da orientação da seqüência, a presença de *splicing* e, conseqüentemente, da localização de introns e exons

(Figura 4). Além disso, também é possível visualizar a seqüência genômica onde se localiza o candidato analisado para confirmar a presença dos sítios conservados de *splicing* (Figura 5).



**Legenda:** A seqüência genômica é mostrada em cinza, no alto da figura. A EST alinhada é mostrada em preto sendo os exons representados por barras e os introns por linhas. Possíveis seqüências de genes conhecidos, predição gênica, mRNAs humanos e não humanos alinhados com a seqüência da EST são mostrados abaixo da seqüência da EST analisada.

**Figura 4** - Interface gráfica disponibilizada pelo programa BLAT.

```

caaaatcttc ccaggcaaat atggtagagg gaaagagcac ataggagggt 149406811
gtgctcataa agtaggataa tgggaattgt attttttaga ggtacaacag 149406861
gtataactgt acatggctga ggtggggcag agcataagct cactggagggt 149406911
aaggagatag aggaactgaa cagtattaac acagataaaa aagtttactc 149406961
tctccctctt cccctccttt tccctccac ctcctctctc cctctcctgc 149407011
ag CTGAGA TtGTGGCAA AAAATTATTA AGCCTCTGCG CTGATGAAGA 149407061
GAAAGATGAA GTCTACCAA AGATTATCT CAAATTTCTC TTAGTAAACT 149407111
CA gtatgta tatttttaaa agtagtaaat tcttaggttt taaggacaaa 149407161
gtaacaatcg ctcaatggat taatttttaa aaagatatta gagcacatac 149407211
ggctgttgat catatattaa ttactttta tataatatta caaactctat 149407261
tcttagtgtt aaaaaaatct ctcctccatt aaatcttatt cctcaccaca 149407311
gaccctgagt attgtagcat ttgtacaatt gcatttagtc tatataaag 149407361
atggaccatt gtatatagcg tcttagctgg tgccctactc tactgatata 149407411
acatagctag ctatgtaagt gctaagtctc gtggaaagaa aataatatt 149407461
ttgtaaaatc taatcgaaga aagacattgg tcatatgaaa taatccagta 149407511
gaataggtta aagtacaagt gtgactctgt ttgtattc ttctcttttt 149407561
actac AAC CATATTGAAT TTTGCTGTC TTTAAAAGA GGAAATGTCG 149407611
AACATGGTGA TAGTTCTGCT TACGAAAACG TGAATTTAT TGTGAATGTA 149407661
AGAGATATT GTAAT gtaa gcagttctgt ttattttatc tatatgacat 149407711
ttatgagtca tgtagcttca cactactaac ataaatgctc ttctgcttgg 149407761
tacttgcaaa ctgctcgata tccctatggc cacagtttct gactactcac 149407811
attacttaga cttttaaaaa ttaacagtga aattacaag gtgggggaca 149407861
ctgagttggt aaggatggct ctttcttttt ataaaacaag gaaatgattg 149407911
caaaaatata acaaaaagaag ttaatgcat atccaagctt aaaataaagg 149407961
ggactctgtg ggtctatgtg taccagaagt taagtttaa agcagtatga 149408011
cagtggtaag aagcagctga ttttgggtg gctcgggtgg attttaagtc 149408061

```

**Legenda:** As letras em preto representam as seqüências dos introns e em azul dos exons. Os números à direita indicam a posição no genoma e os círculos vermelhos indicam a presença dos sítios conservados de *splicing* (GT/AG).

**Figura 5** – Seqüência genômica disponibilizada pelo programa BLAT onde se localiza o candidato analisado.

## 3.2 AVALIAÇÃO DO PADRÃO DE EXPRESSÃO DOS CANDIDATOS SELECIONADOS

A análise do padrão de expressão tecidual dos candidatos foi realizada através de reações de RT-PCR. Iniciadores específicos para cada candidato foram desenhados e testados em cDNAs de diferentes tecidos. As moléculas de cDNA foram sintetizadas a partir de RNA total obtido de 21 tecidos normais disponibilizados comercialmente pela CLONTECH<sup>®</sup>, 17 linhagens celulares tumorais derivadas de 11 tecidos diferentes e 160 amostras de tumores do Banco de Tumores do Hospital AC Camargo derivadas de 9 diferentes tecidos.

### 3.2.1 Tecidos normais

Devido à dificuldade de obtenção de tecidos normais, foi utilizado um painel de RNAs de tecidos humanos normais comercializado pela CLONTECH<sup>®</sup>. Estes RNAs nos foram enviados em Etanol e 0,08M Acetato de Sódio e para a utilização os mesmos foram precipitados e ressuspensos em água DEPC.

Os RNAs são provenientes de 21 tecidos diferentes: cérebro, mama, pulmão, próstata, intestino delgado, cerebelo, colon, cérebro fetal, fígado fetal, coração, rim, placenta, glândula salivar, músculo esquelético, medula espinhal, baço, estômago, timo, traquéia, útero e testículo. Antes de serem utilizados para a síntese de cDNA, a qualidade destes RNAs foi avaliada em gel de agarose.

A integridade dos RNAs foi visualizada aplicando-se 1µg de RNA total em gel de 1% agarose. Antes de ser aplicado, o RNA foi desnaturado a 65°C por cinco minutos, sendo mantido em condição desnaturante em tampão de amostra contendo

Uréia (2X TAE, 30% Glicerol, 7M Uréia e traços de Azul de Bromofenol). Todos os procedimentos foram feitos com materiais próprios para uso exclusivo de RNA e livres de RNase. A coloração foi feita com brometo de etídio e o gel visualizado em luz ultravioleta (UV). Foram considerados íntegros, os RNAs que apresentaram as bandas correspondentes aos RNAs ribossômicos 28S e 18S bem evidentes, tendo a primeira o dobro da intensidade da segunda.

### 3.2.2 Linhagens celulares tumorais

As linhagens celulares tumorais disponibilizadas pela ATCC<sup>®</sup> (*American Type Culture Collection*) foram cultivadas em meio de cultura apropriado segundo especificações da própria ATCC<sup>®</sup> (<http://www.atcc.org>). Foram cultivadas 17 linhagens celulares derivadas de 11 tecidos tumorais diferentes: HL-60 (leucemia aguda mielocítica), SCABER (carcinoma de células escamosas da bexiga), A172 (glioblastoma), T98G (glioblastoma multiforme), MCF-7 (adenocarcinoma de mama), MDA-MB-436 (adenocarcinoma de mama), SW480 (adenocarcinoma coloretal), FaDu (carcinoma de células escamosas da faringe), H1155 (adenocarcinoma de pulmão), H358 (adenocarcinoma de pulmão), A2058 (melanoma), SKMel-28 (melanoma maligno), SaOS-2 (osteossarcoma), DU145 (carcinoma de próstata), PC3 (adenocarcinoma de próstata), CaSki (carcinoma epidermóide da cérvix uterina) e Hela (adenocarcinoma de útero).

Para cada extração de RNA foi necessária uma massa mínima de  $3 \times 10^6$  células correspondendo, em média, a quatro garrafas médias com células confluentes. A extração foi realizada através da técnica de Sedimentação em Cloreto de Césio (CHIRGWIN et al. 1979). Inicialmente, o meio de cultura foi aspirado e 9ml de

solução de lise (4M Isotiocianato de guanidina / 25mM Citrato de Sódio – pH 7 / 0,1M  $\beta$ -mercaptoetanol) foram adicionados à garrafa de cultura (75cm<sup>2</sup>). Em seguida, o lisado celular foi homogeneizado e transferido para um tubo de ultracentrífuga contendo 4 ml de solução de Cloreto de Césio (5,7M CsCl / 25mM NaAc) e centrifugado a 150000xg por 17 horas a 20°C (rotor SW40Ti, Beckman®). Após a centrifugação, o RNA formou um precipitado no fundo do tubo e o sobrenadante contendo proteínas e DNA foi descartado. A parede interna do tubo foi limpa e o RNA foi solubilizado em 50 a 200 $\mu$ l de água biodestilada estéril tratada com DEPC. Antes de serem utilizados para a síntese de cDNA, a qualidade dos RNAs foi avaliada em gel preparativo de agarose como descrito no item 3.2.1.

A quantificação do RNA obtido foi feita através da leitura de uma alíquota da amostra em espectrofotômetro com comprimento de onda equivalente a 260nm, considerando-se que 1DO<sub>260nm</sub> equivale a 40 $\mu$ g/ml de RNA. A relação entre as leituras realizadas a 260 e 280nm foi utilizada como parâmetro na estimativa do grau de contaminação do RNA por proteínas. Para que seja considerado puro, o valor desta relação deve ser superior a 2,0.

### **3.2.3 Amostras de tumores de pacientes do Banco de Tumores do Hospital AC Camargo**

As amostras de tumores foram solicitadas junto ao Banco de Tumores do Hospital AC Camargo tendo sido utilizadas apenas amostras cedidas após o consentimento informado dos pacientes e aprovação do Comitê de Ética do Hospital (Projeto de Pesquisa 548/03). Todas as precauções pertinentes para manter o sigilo e a confidencialidade dos dados dos pacientes foram adotadas. As amostras de RNA de

glioblastoma foram gentilmente cedidas pelo Dr. Gregory Riggins da Universidade de Duke nos Estados Unidos. Estes glioblastomas são *xenografts* originados a partir da injeção de células tumorais humanas em camundongos.

As 160 amostras utilizadas nesse estudo foram derivadas de nove tipos diferentes de tumor: colon, estômago, glioblastoma, mama, melanoma, próstata, pulmão, útero e tireóide (Tabela 2).

**Tabela 2** – Amostras tumorais utilizadas para avaliação da expressão dos candidatos selecionados.

<i>Tecido</i>	<i>Número de amostras</i>
Colon	18
Estômago	9
Glioblastoma	13
Mama	24
Melanoma	18
Próstata	21
Pulmão	14
Útero	19
Tireóide	24

A extração de RNA das amostras de tumores foi feita utilizando-se o reagente Trizol (Invitrogen®). Para tanto, os tumores, ainda congelados, foram imersos em 1ml do reagente Trizol sobre uma placa de petri estéril e cortados em pequenos fragmentos com o auxílio de um bisturi estéril. Posteriormente, estes fragmentos foram transferidos para tubos cônicos de poliestireno de 5ml (Falcon®) aos quais foi acrescido 1ml do reagente Trizol. Com auxílio de um Polytron (Kinematica® AG), as amostras foram completamente homogeneizadas e o protocolo foi seguido conforme

as especificações do fabricante. A quantificação do RNA obtido foi realizada conforme descrito no item 3.2.2 e, antes de serem utilizados para a síntese de cDNA, a qualidade dos RNAs foi avaliada em gel de agarose como descrito no item 3.2.1.

#### **3.2.4 Avaliação da contaminação com DNA genômico nos RNAs extraídos**

Para se ter a garantia de que os RNAs extraídos não estavam contaminados com DNA genômico, 200ng de RNA total foram amplificados por PCR utilizando iniciadores específicos para seqüências intrônicas do gene hMLH1 (*human mut-L homologue 1*) (*Forward* - 5'TGG TGT CTC TAG TTC TGG3' e *Reverse* - 5'CAT TGT TGT AGT AGC TCT GC3') de maneira que a ocorrência de amplificação indicasse contaminação com DNA genômico (SILVA et al. 2003). A reação de amplificação foi realizada em presença de tampão 1x Taq DNA polimerase, 1,6mM MgCl<sub>2</sub>, 0,2mM dNTPs, 0,5µM de cada iniciador e 1 unidade de Taq DNA polimerase (GIBCO/BRL<sup>®</sup>) em um volume final de 25µl. A reação teve uma desnaturação inicial de 5 minutos a 94°C, seguida de 35 ciclos de desnaturação a 94°C por 45 segundos, anelamento dos iniciadores a 55°C por 45 segundos e extensão a 72°C por 1 minuto e um tempo de extensão final de 6 minutos a 72°C. Caso o RNA estivesse contaminado com DNA, seria amplificado um fragmento de aproximadamente 250pb visualizado em gel de 8% poliacrilamida corado com prata. Nas amostras em que foi detectada a contaminação realizou-se o tratamento com DNase (Promega<sup>®</sup>).

### 3.2.5 Tratamento do RNA total com DNaseI

O RNA total, quando necessário, foi tratado com DNase livre de RNase (Promega<sup>®</sup>), utilizando 10 unidades da enzima para tratar 50µg de RNA por 60 minutos a 37°C. A reação foi interrompida pela adição de uma solução de neutralização (0,05M EDTA, 1,5M NaOAC e 1% SDS) e o RNA foi extraído com fenol/clorofórmio e precipitado a -20°C por 18 horas através da adição de 250mM NaCl e etanol absoluto com um volume igual a duas vezes o volume inicial. Após centrifugação, o RNA foi solubilizado em água DEPC. A quantificação foi feita através da leitura de uma alíquota da amostra no espectrofotômetro a 260 e 280nm conforme descrito anteriormente. As amostras tratadas foram novamente submetidas a amplificação com iniciadores específicos para seqüências intrônicas do gene hMLH1 para confirmar a eliminação da contaminação.

### 3.2.6 Síntese de cDNA

A síntese da primeira fita de cDNA foi realizada a partir de 1µg de RNA total, com iniciadores oligo(dT)<sub>12-18</sub>, utilizando o kit *Superscript<sup>TM</sup> Preamplification System for First-Strand Synthesis*, seguindo as instruções do fornecedor. Para avaliar a eficiência de síntese, o cDNA foi amplificado com iniciadores desenhados em diferentes exons do gene GAPDH (Forward - 5'CTG CAC CAC CAA CTG CTT A3' e Reverse - 5'CAT GAC GGC AGG TCA GGT C3'). As amplificações foram realizadas em um volume final de 10µl na presença tampão 1X Taq DNA polimerase, 1,0mM MgCl<sub>2</sub>, 0,1mM dNTPs, 0,3µM de cada iniciador e 1 unidade de Taq DNA polimerase, utilizando como molde 1µl do cDNA sintetizado. A reação foi realizada com uma desnaturação inicial de 4 minutos a

94°C, seguida de 22 ciclos de desnaturação a 94°C por 45 segundos, anelamento dos iniciadores a 55°C por 45 segundos e extensão de 1 minuto a 72°C. Os produtos da amplificação foram analisados em gel de 8% poliacrilamida corado com prata.

### 3.2.7 RT-PCR

O cDNA sintetizado foi utilizado como molde em reações de PCR com iniciadores específicos para cada candidato selecionado. Os iniciadores foram desenhados em exons diferentes, tornando possível a distinção entre a amplificação do DNA genômico e a amplificação do cDNA, com o auxílio do programa Primer3 (<http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>) (Anexo 1). Além disso, a temperatura de anelamento dos iniciadores e a possível existência de estruturas secundárias e formação de dímeros foram analisadas no programa Oligotech<sup>®</sup>.

As RT-PCRs foram realizadas em presença de tampão 1x Taq DNA polimerase, 1,0mM MgCl<sub>2</sub>, 0,1mM dNTPs, 0,3μM de cada iniciador e 1 unidade de Taq DNA polimerase, utilizando como molde 1μl de cDNA em um volume final de 25μl. As reações tiveram 40 ciclos sendo que a temperatura de anelamento foi determinada pelo programa computacional Oligotech<sup>®</sup> a partir das seqüências dos iniciadores. Em seguida, os produtos amplificados foram visualizados em gel de 8% poliacrilamida corado com prata.

O processo de validação experimental através da análise de expressão por RT-PCR foi dividido em 4 fases. Em um primeiro momento, os candidatos foram analisados quanto à expressão em testículo, sendo que apenas os que apresentaram um resultado positivo seguiram em nossa análise. A segunda fase consistiu na análise de expressão no painel de 20 tecidos normais obtidos comercialmente através da

CLONTECH® e os candidatos que não apresentaram expressão em nenhum dos tecidos normais analisados seguiram para a próxima fase. Na terceira fase analisamos o perfil de expressão em 17 linhagens tumorais e, por fim, na última fase analisamos a expressão dos mesmos candidatos em 160 amostras de tumores de pacientes do Banco de Tumores do Hospital AC Camargo.

Durante a primeira fase da análise de expressão dos candidatos selecionados, ou seja, a análise da expressão em testículo, as reações de RT-PCR foram padronizadas visando otimizar a especificidade das mesmas. Os problemas de amplificação se resumiram a: (1) inespecificidade da reação, (2) ausência de amplificação e (3) amplificação de um produto com tamanho diferente do esperado. Desta forma, o processo de padronização consistiu em alterações na concentração de Cloreto de Magnésio, temperatura de anelamento dos iniciadores e adição de 1M Betaína à reação de amplificação. Além disso, os fragmentos com tamanho diferente do esperado também foram seqüenciados como descrito no próximo item.

### **3.2.8 Confirmação da identidade dos fragmentos de RT-PCR correspondentes aos candidatos validados na análise de expressão por seqüenciamento**

Os fragmentos de RT-PCR dos candidatos que apresentaram o padrão de expressão característico de antígeno CT foram seqüenciados para confirmação da especificidade da reação de RT-PCR. As reações de seqüenciamento foram realizadas utilizando o kit *Dynamic™ ET Terminator Cycle Sequencing* (Amersham Biosciences®) disponível para o seqüenciamento automático no seqüenciador ABI Prism 3100 DNA Sequencer (Applied Biosystems®) segundo instruções do fornecedor. As seqüências obtidas para cada candidato foram alinhadas através do

programa *BLAST2 sequences* contra a seqüência da EST utilizada para o desenho dos iniciadores.

### 3.2.9 *Southern-blot*

Para se obter uma maior sensibilidade e especificidade em nossa análise, os produtos de RT-PCR das amostras normais e de linhagens celulares dos candidatos validados foram re-analisados por *Southern-blot*. Inicialmente, os produtos de RT-PCR foram fracionados em gel de 1% agarose corado com brometo de etídio e depois transferidos para membranas de nylon (Hybond™-Amersham Biosciences®). Antes da transferência, o gel foi incubado em solução desnaturante (1,5M NaCl e 1M NaOH) por 45 minutos e, em seguida, duas vezes de 30 minutos cada em solução de neutralização (0,5M Tris pH 7,5 e 1,5M NaCl), sob agitação constante. Após a transferência por capilaridade com tampão 10X SSC, as membranas foram lavadas em tampão apropriado (2X SSC) e, em seguida, foi feito o *cross-link* do DNA à membrana com luz UV (120mJ/cm<sup>2</sup>).

Como sonda foi utilizado um produto de RT-PCR referente à amplificação em testículo de cada um dos candidatos, sendo que todos os produtos foram previamente seqüenciados. A sonda foi marcada através da incorporação de nucleotídeos [ $\alpha$ -<sup>32</sup>P]dCTP (Amersham Biosciences®). A marcação foi feita utilizando-se o *Random Primers DNA Labeling System* (Invitrogen®), seguindo as instruções do fabricante. Depois de marcada, a sonda foi purificada em coluna de Sepharose 50 (Invitrogen®) para a eliminação dos nucleotídeos não incorporados. Posteriormente, a sonda foi hibridizada com as membranas a 65°C durante 16 horas em solução de hibridização (0,25M Na<sub>2</sub>HPO<sub>4</sub>, 7% SDS, 1% BSA e 1mM EDTA).

Após a incubação, as membranas foram lavadas em solução de lavagem (0,25M  $\text{Na}_2\text{HPO}_4$ , 1% SDS e 1mM EDTA) durante 30 minutos a 65°C por duas vezes. Por fim, as mesmas foram expostas a filme fotográfico (Hyperfilm™ MP – Amersham Biosciences®) a -70°C por diferentes períodos de tempo.

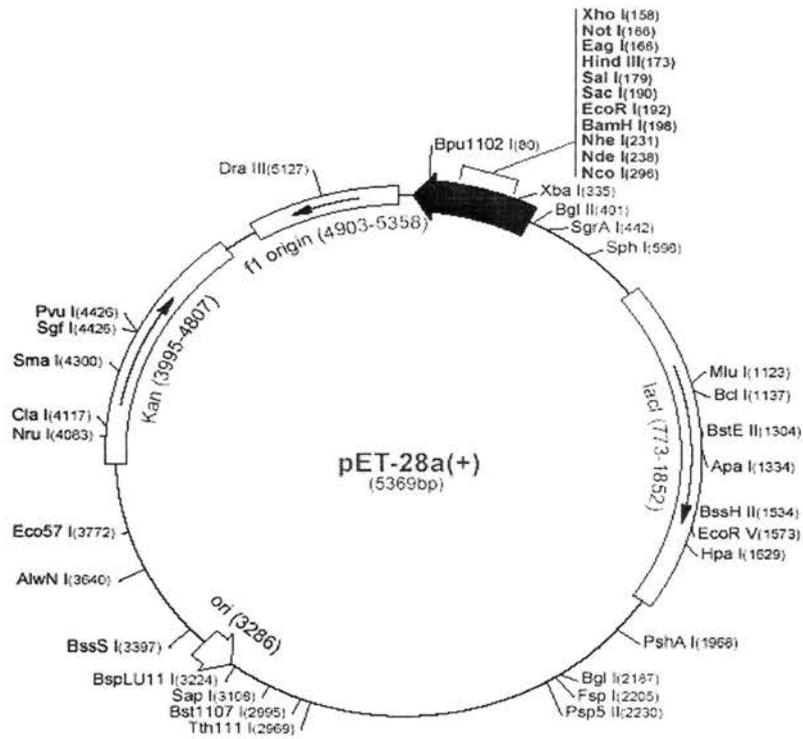
### **3.3 EXPRESSÃO DAS PROTEÍNAS RECOMBINANTES**

Os candidatos validados como possíveis novos antígenos CTs e que apresentavam uma seqüência completa de mRNA depositada em bancos de dados públicos foram selecionados para avaliação da resposta humoral em pacientes com câncer. Para tanto, as proteínas recombinantes correspondentes aos candidatos CT101, CT802 e CT809 foram expressas em sistema bacteriano, purificadas e, então, utilizadas na avaliação da presença de anticorpos em plasma de pacientes com câncer através de ensaios de *Immunoblotting*.

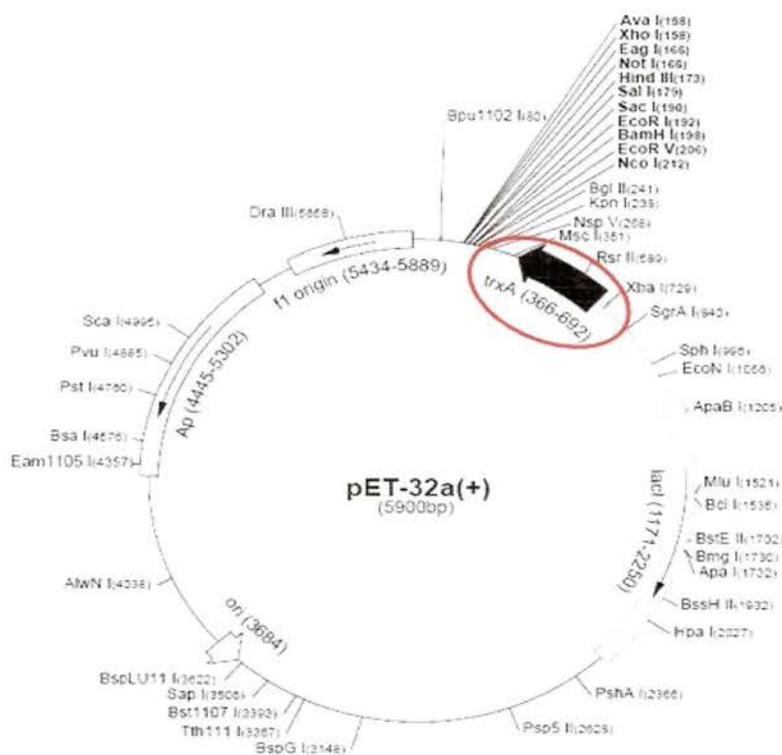
#### **3.3.1 Clonagem da fase aberta de leitura dos candidatos com seqüência completa de mRNA em vetor de expressão**

Para a clonagem das fases abertas de leitura dos candidatos CT101 e CT809 foi utilizado o vetor de expressão pET-28a (Novagen®), cujo mapa está esquematizado na Figura 6. Para o candidato CT802 não foi possível a clonagem neste vetor, uma vez que uma das enzimas utilizadas (Sall) apresentou uma baixa eficiência na digestão do vetor e do fragmento contendo a fase aberta leitura deste candidato. Sendo assim, para a clonagem do mesmo foi necessário utilizar outro

vetor de expressão disponível no laboratório, pET-32a (Novagen®), cujo mapa está esquematizado na Figura 7.



**Figura 6** – Mapa do vetor pET-28a.



**Legenda:** O círculo em vermelho destaca a presença da seqüência referente à proteína Tioredoxina *upstream* aos sítios de clonagem.

**Figura 7** – Mapa do vetor pET-32a.

Vale ressaltar que estes vetores possuem um sistema de fusão da proteína recombinante a uma seqüência formada por 6 resíduos de histidinas consecutivos (*His-tag*) que permite sua rápida e eficiente purificação em coluna de agarose carregada com níquel. Além disso, também possuem um controle eficiente da expressão do gene de interesse, diminuindo a expressão basal da proteína de interesse antes da indução por IPTG.

As enzimas selecionadas para a clonagem sítio dirigida da fase aberta de leitura dos candidatos não possuem sítios de restrição na sequência do respectivo fragmento a ser clonado. Os iniciadores utilizados na amplificação dos insertos foram construídos de maneira a permitir a clonagem dos fragmentos em fase de leitura e de maneira orientada. Além disso, no caso do CT101 foi construído um iniciador reverso (PROT101R) posterior ao fim da fase aberta de leitura para facilitar a amplificação da mesma em reações de *nested PCR* uma vez que esta possui 2.322pb. O produto da amplificação com este iniciador foi, então, utilizado como molde para a PCR com os iniciadores que flanqueiam a fase aberta de leitura deste candidato. As enzimas utilizadas para a clonagem de cada um dos candidatos assim como a sequência dos iniciadores utilizados para a amplificação das fases abertas de leitura estão descritos na Tabela 3.

**Tabela 3** – Enzimas utilizadas para a clonagem sítio dirigida dos candidatos CT101, CT802 e CT809 e os respectivos iniciadores para amplificação da fase aberta de leitura com o tamanho do fragmento amplificado pelos mesmos.

<i>Candidato</i>	<i>Enzimas</i>	<i>Iniciadores</i>	<i>Fragmento amplificado (pb)</i>
CT101	EcoRI / HindIII	PTN101F 5' CGGAATTCATGAAGATGAGAGGGGAAAAG 3' PTN101R 5' CCCAAGCTTTTAGCACGGCTTATTGAGTC 3' PROT101R 5' CATGCAGACCTCATTGGCCTG 3'	1ª reação – 2.387 2ª reação – 2.322
CT802	EcoRV / EcoRI	PTN802F32 5' GATATCATGGCGGCGAGCGCGCTGCGA 3' PTN802R32 5' CGGAATCTTATTTCCTCTGGAAAGTTAG 3'	1.428
CT809	EcoRI / HindIII	PTN809F 5' CGGAATTCATGTCTGAAATCAGATTCACC 3' PTN809R 5' CCCAAGCTTTAACTCATACCATTTGATCC 3'	1.170

**Legenda:** Em cinza estão destacados os sítios referentes a cada enzima e em rosa as bases adicionadas como âncora para as mesmas. Em azul estão destacados os códons de iniciação (ATG) nos iniciadores *forward* e de terminação (TTA) nos iniciadores reverso.

A síntese de cDNA de testículo foi realizada conforme descrito no item 3.2.6. As reações para amplificação da fase aberta de leitura do CT101 foram realizadas em presença de tampão 1x Taq *Platinum Hi-Fidelity*, 2,0mM MgSO<sub>4</sub>, 0,1mM dNTPs, 0,4μM de cada iniciador e 1 unidade de Taq *Platinum Hi-Fidelity* (Invitrogen®), utilizando como molde 1μl de cDNA de testículo para a reação com o iniciador externo e 1μl desta reação diluída 1:10 para a PCR com os iniciadores que flanqueiam a fase aberta de leitura. A reação iniciou-se com desnaturação a 94°C por 2 minutos seguida de 40 ciclos de desnaturação a 94°C por 1 minuto, anelamento dos iniciadores a 58°C por 1 minuto e extensão de 2 minutos a 68°C e uma extensão final de 6 minutos a 68°C. Em seguida, o produto amplificado de 2.322pb foi visualizado em gel de 1% agarose corado com brometo de etídio.

A reação para amplificação da fase aberta de leitura do CT802 foi realizada em presença de tampão 1x Taq *Platinum Hi-Fidelity*, 2,0mM MgSO<sub>4</sub>, 0,1mM dNTPs, 0,5μM de cada iniciador, 1M Betaína e 1 unidade de Taq *Platinum Hi-Fidelity* (Invitrogen®), utilizando como molde 1μl de cDNA de testículo. Já a reação para amplificação da fase aberta de leitura do CT809 foi realizada em presença de tampão 1x Taq *Platinum Hi-Fidelity*, 2,0mM MgSO<sub>4</sub>, 0,1mM dNTPs, 0,5μM de cada iniciador e 1 unidade de Taq *Platinum Hi-Fidelity* (Invitrogen®) também utilizando-se como molde 1μl de cDNA de testículo. As condições de reação foram as mesmas utilizadas para a amplificação da fase aberta de leitura do CT101 sendo que o produto amplificado do CT802 correspondia a 1.428pb e o produto do CT809 a 1.170pb.

Com o intuito de obter uma massa suficiente para a digestão com as enzimas de restrição e posterior purificação destes produtos, as reações foram feitas em

quadruplicatas e o volume total da amplificação de cada um dos candidatos foi concentrado através de precipitação com 3M Acetato de Sódio pH 5,2 e duas vezes o volume de etanol absoluto. Em seguida, os produtos foram digeridos com as enzimas selecionadas para a clonagem.

Os fragmentos foram digeridos primeiramente com a enzima HindIII (New England – Biolabs<sup>®</sup>), no caso dos candidatos CT101 e CT809, e EcoRV (New England – Biolabs<sup>®</sup>), no caso do candidato CT802. Para as reações foram utilizadas 5 unidades da enzima, em tampão apropriado, durante 4 horas a 37°C. Os produtos das digestões foram purificados com fenol-clorofórmio e precipitados com etanol. Em seguida, foram submetidos a uma nova reação de digestão com a enzima EcoRI (New England – Biolabs<sup>®</sup>) sob as mesmas condições. Ao final, os produtos das digestões foram purificados em gel de 1% agarose utilizando-se o *QIAquick Gel Extraction Kit Protocol* (QIAGEN<sup>®</sup>) segundo as recomendações do fabricante. Paralelamente, os vetores pET-28a e pET-32a também foram digeridos sob as mesmas condições com as mesmas enzimas HindIII/EcoRI e EcoRV/EcoRI, respectivamente. Posteriormente, os vetores digeridos foram purificados a partir do gel de 1% agarose como descrito anteriormente.

Aproximadamente 50ng do vetor e 100ng de cada um dos insertos, todos duplamente digeridos com as respectivas enzimas de restrição, foram adicionados a uma reação de ligação contendo 5 unidades de T4 DNA ligase (New England – Biolabs<sup>®</sup>) em tampão apropriado e volume final de 10µl. As reações foram incubadas a 14°C por 16 horas e, em seguida, os produtos das ligações foram dialisados em filtro de nitrocelulose (Millipore<sup>®</sup>) para a remoção do excesso de sal.

Posteriormente, 2µl da diálise foram utilizados para transformação em bactéria *E. coli* cepa DH10B (Stratagene®) eletrocompetente. Após o choque elétrico (resistência de 200Ω, capacitância de 25mF e voltagem de 1.8Kv) em eletroporador *Gene Pulser* (BioRad®), as bactérias foram incubadas durante 1 hora em 1ml de meio CG (CircleGrow – Bio101®) a 37°C sob agitação constante de 200rpm. Em seguida, foram semeadas em placa de CG-ágar (1,5% de ágar) contendo o antibiótico apropriado para cada vetor utilizado, ou seja, 25µg/ml de Kanamicina para o pET-28a e 100µg/ml de Ampicilina para o pET-32a. As placas foram incubadas por 16 horas a 37°C em estufa seca.

A presença do inserto de interesse foi verificada através de PCR de algumas colônias isoladas de cada placa. As reações foram realizadas em presença de tampão 1x Taq Polimerase com MgCl<sub>2</sub> (Phoneutria®), 0,25mM dNTPs, 0,5µM de cada iniciador referente ao promotor T7 presente nos vetores (T7 PROM 5'TAA TAC GAC TCA CTA TAG GG3' e T7 TERM 5'GCT AGT TAT TGC TCA GCG G3') e 1 unidade de Taq Polimerase (Phoneutria®). As reações iniciaram com desnaturação a 94°C por 4 minutos seguidos de 35 ciclos de desnaturação a 94°C por 1 minuto, anelamento dos iniciadores a 55°C por 1 minuto e extensão de 2 minutos a 72°C e uma extensão final de 7 minutos a 72°C. Em seguida, a presença do inserto de interesse nas colônias analisadas foi visualizada em gel de 1% agarose corado com brometo de etídio.

As colônias que apresentaram o inserto de interesse foram novamente coletadas da placa e crescidas em 3ml de CG acrescido do antibiótico apropriado durante 16 horas a 37°C para posterior purificação de DNA plasmidial em pequena escala. Os plasmídeos foram purificados utilizando-se o kit *Wizard™ Plus SV*

*Minipreps* (Promega<sup>®</sup>) seguindo as recomendações do fabricante e, em seguida, foram seqüenciados, como descrito no item 3.2.8, com os iniciadores T7 PROM e T7 TERM para confirmação da inserção do fragmento em fase no vetor. No caso do candidato CT802, devido à presença da seqüência correspondente a proteína Tioredoxina no vetor pET-32a (Figura 7), foi necessária a utilização de um iniciador interno (pET32F 5'CGC TGC TGC TAA ATT CGA ACG3') ao T7 PROM e mais próximo à seqüência de interesse para o seqüenciamento da mesma. Devido ao tamanho dos fragmentos, foi necessário ainda utilizarmos iniciadores internos para seqüenciar toda a extensão da fase aberta de leitura dos transcritos e desta forma, poder confirmar que a seqüência foi estendida corretamente sem alteração, inserção ou deleção de bases que poderiam alterar a fase aberta de leitura dos mesmos. As figuras 8, 9 e 10 indicam a posição dos iniciadores utilizados no seqüenciamento ao longo da fase aberta de leitura correspondente aos transcritos CT101, CT802 e CT809, respectivamente.

**ATG**AAGATGAGAGGGGAAAAGAGAAGAGACAAAGTCAATCCAAAAAGCTCTCAAAGGAAATTAACCTGGATTCCA  
 TCATTTCTACCTATGATTACTTCAACCAAGTGACGCTACAGTTATTAGATGGCTTTATGATTACACTGAGCACAGAT  
 GGAGTGATCAATTTGTGTGGCTGAAAACATCTCTCTCTCTTGGACATTTAC**AGCTGAGATTGTGGGCAAAAAAT**  
 TAITAAGCCTTCTGCCTGATGAAGAGAAAGATGAAGTCTACCAAAAGATTATTCTCAAATTTCTTTACTAACTCA  
 GAAACACATATTAATTTTGTCTGTCATTTAAAAAGAGAAAATGTGCAACATGGTGATAGTTCTGCTTACGAAAACGT  
 GAAATTTATTGTGAATGTAAGAGATATTTGTAATGAGTTTCTGTGGTCTTTAGTGGCTGTTTTCAGCCACCTCTG  
 TGCTGACTTTGTGTCATGTTCTCAGGAGGATCGGCTTTATCTTGTGGGAAATGTTTGCATTCTCAGGACTCAGCT  
 CCTGCAGCAACTTTACACTTCAAAGGCAGTCAGTGATGAAGCTGTACTTACACAAGATTCAGATGAGGA**ACCTTTT**  
**GTGGGAGAGCTCAGTAGCTCTCAAGGTCAAAGAGGACACACTAGCATGAAAGCCGTGTACGTTGAACCCGCTGCT**  
 GCTGTCTGTGCTGTCTATCTCAGACGACCAAAATGATATTGAGAGGTTGAGCAGTATGGTCCACAAGAAAACGT  
 TCACATGTTTGTAGATTCTGATTCAACTTATTGCTCCAGTACAGTTTTCTGGATACTATGCCTGAATCTCCAGCCTT  
 ATCCTTGCAAGACTTTTCAGGTTGAGCCTGAGGTGAATCCATTGTACAGGGCAGACCCAGTGGAC**CTGGAGTTCTCG**  
**GTGGATCAGGTGACTCAGTGGACCAGGAGGGCCCAATGGAACAGCAGGACCCAGAGAACCCAGTTGCCCCGTTG**  
 GACCAGGCAGGCGCTGATCCAGTGGATCCAGGACTCAGTGGACCTGGGGCTGCTGGCCGAAGTGCAGC  
 CATTACAGCCATCATCACCAGTTGCATATGACATCATTAGCCAGGAACTGGAAGTGAAGAAAGTTGAAGGAGCA  
 GCTAGAAGAGAGGACTTGTGTTGCTGCATGATGCCATCCAAAACAGCAGAATGCATTGGAATT**GATGATGGATCAC**  
**CTTCAGAAGCAGCCAAACACATTACGCCACGTTGTCAATCTCTGATCTCCAATCTTCGGAGGACAGTGC**  
 AACAGAAAACAACAGCTGGGCAAGTGAAGCGGCTCTCCACATCCCAAGGACGTCAGTGTTTCTGTGTTTATCT  
 TTATCCAATCTCTCAAAAACACTGGGGAGCTTCAGGAGCCTTGTGTTGCCCTTCAACCAGCAGCAACTGGTGCAGCA  
 AGAACACACCTGAAGGAGCAGCAGCGGAGCTGCGGGAGCAGCTGCAACAGCTGAGAGAGC**AAAAGGAAGGTGC**  
**AGAAGCAGAAGAAGATGCAGGAGAAGAAGAAGCTGCAGGAGCAGAAAATGCAGGAGAAGAAGAAGTGCAGGAG**  
 CAGAGCGGGCAAAAAGAAGAAGAAGCTACAGGAGCGGAAGAAGTGCAGGGCAGATGCTACAGAAAAGAGCCAGA  
 GGAGGAGCAGCAGAAAGCAGCAGTTGCAAGAGCAGCCACTGAAGCATAATGTCATCGTGGGGAATGAGAGGGTGA  
 GATATGCTGCAAAAACCCAGCTGACGTATCTGTGCCCTTGCAATCACCTGTTAGATTTTTAC**AGGCCCAACCCA**  
**TTGTTCTGTCCAGAGAGCAGCTGAACAACAGCCCTTGGCTTCTATCAAGATGAAAACCTGTGGGCAACAGGAAGA**  
 TGAGAGTCAAAGTTTTATCCTGAGGCGTATCAAGGGCCCCCGTGAACCAGCTGCCATTGATAGATACCTCAAAC  
 CTGAGGCAATTTCTTCTCCAGCATTCTCAGTTTCCATAACTTCAGACTCAACCATAAGCACCCCTGGAGACCCAC  
 AGGATACATCCGGCTTTGGCAAGAGTTGTCTGATTCACCTCGGCTCTGTGTCAGAGTGAACACTTGGCTTGCAGT  
 AGCAGGGCACCTGCAGGGCAACCCACTACCATCAGGTGCAAGTTTCTGAGGTAGGAGTGCAGGGACCTCTGA  
 TCCACAGGCTTTCAAGGCCTGTGCATACCAGCCAGACAGATGAGATCTGGGAGCAGACCAGATTGATGCCTG  
 CAGAGCAACGTGACTCAAATAAGCCGT**CTAA**

**Legenda:** As bases destacadas em vermelho correspondem ao códon de iniciação (ATG) e terminação (TAA) da sequência. As sequências de bases destacadas em azul correspondem aos iniciadores construídos para o sequenciamento.

**Figura 8** – Sequência referente à fase aberta de leitura do transcrito CT101.

**ATG**CGCGGAGCGCGCTGCGAGGCCTGCCAGTGGCTGGCGGAGGCGAGAGTAGCGAGAGCGAGGATGATGGCTGG  
 GAGATTGGGTATCTCGACCGGACGCTCAGAAAATTGAAAAGGCTATTACCCATTGAAGAAAAGAAAAGAAAATTA  
 AGAAAAGCAATGACCAATCGGAGATGTTTCAATGGTCCAGGAGCTCTAGATTCTGGCATTAGTGTAGATTTCAAAC  
 CAGTATGGATGGACTCCCTTATGTATGCTGTAGTGTGCCAATGCAGAGCTGGTTCGGGCTCTTTGGACAGAGG  
 TGCT**AAATGCAAGCTTTGAG**AAGGATAAGCAAAGTATTTTGATAACTGCATGTTCTGCTCATGGCTCAGAGGAACAG  
 ATCTTGAAGTGTGTAAGACTACTACTTTCAAGAAATGCTGATCCAAATGTTGCTTGTAGGAGACTTATGACCCCAAT  
 CATGTATGCTGCTCGAGATGGTACACCCAGGTTGTTGCTCTCCTTGTGCTCATGGAGCAGAAGTTAATACCCAGG  
 ATGAGAATGGTTACACTGCTTTAACGTGGGCAGCACGTCAGGGTCATAAAAATATAGTTTTGAAGTTGCTTGAAC**TT**  
**GGAGCTAATAAAAATG**CTACAAACCAAAGATGGAAGATGCAAGTGAAGTTCAAAAGAAACAAACATCATGA  
 GATCTTCAACTTACTTTCTTTTACTTTAAATCCATTGGAAAGGAAAACCTCAACAGCTAACTAAAAGAAGACACTATTTG  
 TAAAATATTGACAACAGATTCTGATAGAGAAAAGATCACATTTTATGTTTCATATACAGCATTTGGAGATCTGGAAG  
 TATTTTTACATGGTCTTGGACTTGAACATATGACAGATTTACTAAAAGGAAAGGATATAACGTTAAGACATCTTT**G**  
**ACCATGAGGGAAGAT**GAATTTACAAAGAATGGAATTACCAGTAAAGACCAGCAGAAAATCTGGCTGCTCTTAAA  
 GAACTACAGGTAGAAGAGATACAATTTGGAGAGCTATCTGAAGAGACAAAGTTGGAAATCAGTGGTGATGAGTTCC  
 TCAACTTTCTTCAAATTAATAAACAGTGTGGCCATTTAATAACAGCTGTACAGAATGTTATTACTGAGTTACCTG  
 TAAATTTCAAAGATAACACTGGAATGGGCTTCTCCCGAAATTTACTTCAAGTTTGTGAAGAATTGGTTAATAAT  
 GTTGAAGATTGAGTGAAGGCTGCAAACTAAAAGACCTAATTCAAAAGTTGCAAAATGAACGGGAAAATGATC  
 CAACTCATATACAATTAAGGGAAGAAGTATCTACATGGAATAGTGAATTTGAAGAGGACAGCTATTACCATATG  
 CGGATTCGGTTTTCTTTTCAATTTGCAAGCTA**ACTTTCCAGAGGAAAATAA**

**Legenda:** As bases destacadas em vermelho correspondem ao códon de iniciação (ATG) e terminação (TAA) da sequência. As sequências de bases destacadas em azul correspondem aos iniciadores construídos para o sequenciamento.

**Figura 9** – Sequência referente à fase aberta de leitura do transcrito CT802.

**ATG**TCTGAAATCAGATTCACCAATCTCGCTTGGGATCAAGTTATAACACTGGATCAAGTGTTAGATGAAGTAA  
 TTCCAATTCATGGAAAGGGGAATTTCCCCACAATGGAGGTAAAACCAAAGACATCATTCATGTTGTGAAAGA  
 TCAACTCATAGGGCAAGGAATTATTGTTAAAGATGCCAGATTGAATGGTTCCGTAGCAAGTTACATACTTGCA  
 AGCCACAATGGAATCAGCTATAAGGATCTGGACGTTATTTTGGTGTGAGCTTCCAGGTAACGAAGAATTC  
**AGGTTGTTAAAAGATG**CAGTTCTAGACTGTCTACTTGACTTTTTACAAAAGATGTAAGAAGGAAAAGCTCTC  
 CCCAGATATCATGAAAGACGCTTACGTACAGAAATTGGTCAAGGTTTGCAATGGGCATGATTGTTGGAGTCTT  
 ATCTCCCTTTCAAATAACACTGGGAAGAATTTAGAACTAAAATTTGTGAGTTCACTCAGACGGCAGTTTGAATT  
 TAGTGTAGATTCCCTTCAAATGTTTTGGATCCCATGTTAGACTTCTACAGTGACAAAAATGCCAAGCTAACCA  
 AAGAATCCTATCCTGTTG**GGTAG**CTGAAAGCATGTATGGAGACTTCCAGGAAGCAATGACACATTTGCAAC  
 ACAAGCTCATATGTACCAGGAAACCTGAAGAGATTAGAGGTGGTGGCCTTCTGAAGTACTGCAGCTTGCTGCT  
 TCATGGCTTCAAGCCAGCCTGTATGTCAGAAATCAAAAACCTAGAACGTTATATGTGCTCTAGATTCTTTATTG  
 ATTTTCTCATATAGAAGAACAGCAAAAAGAAAATTGAATCATACCTCCACAACCATTTCATAGGTGAAGGAAT  
 GACCAAGTATGACTACCTTATGACCTTGATGGAGTTGTGAATGAAAGCACTGTTGCTCATGAGTTATGAAA  
 GAAGACAGATTCTCCACCTGATCACCATGATGGCTTTGAAAAGTACTTGGAGAACTAAATATTCTACCCAATAC  
 ACAAAGGTAACCTTGCTTTATCAGCCTGCTCCGTACTTTGCAGCTGAGGCAAGGTACCCTATTATGTAATAC  
 CTGAGCCACCCCGTTAGCTTCCAGCCATACCACCCACTGCACCTTCGTGGATCAAATGGTATGAGT**TAA**

**Legenda:** As bases destacadas em vermelho correspondem ao códon de iniciação (ATG) e terminação (TAA) da sequência. As seqüências de bases destacadas em azul correspondem aos iniciadores construídos para o seqüenciamento.

**Figura 10** – Seqüência referente à fase aberta de leitura do transcrito CT809.

### 3.3.2 Indução da expressão das proteínas recombinantes

Os plasmídeos contendo as seqüências confirmadas por seqüenciamento foram utilizados para transformação de bactérias *E. coli* cepa BL-21 Rosetta (Stratagene®) quimiocompetentes. Inicialmente, 50ng de cada plasmídeo foram adicionados às bactérias e incubados em gelo durante 20 minutos. Após o choque térmico a 42°C por 1 minuto, as bactérias foram incubadas por 1 hora em 1ml de meio CG a 37°C sob agitação constante de 200rpm. Em seguida, foram semeadas em placa de CG-ágar (1,5% de ágar) contendo 25µg/ml de Cloranfenicol e o antibiótico apropriado para cada vetor utilizado. As placas foram incubadas por 16 horas a 37°C em estufa seca.

Quatro colônias transformantes de cada candidato foram inoculadas e incubadas durante 16 horas a 37°C em 3ml de meio CG contendo os antibióticos apropriados. A partir deste pré-inóculo, as bactérias foram crescidas em 10ml de meio CG contendo antibiótico até atingirem a D.O.<sub>600</sub>=0,6. Neste momento, a expressão das proteínas recombinantes foi induzida através da adição de 0,4mM

IPTG ( $\beta$ -D-tiogalactopiranosídeo de Isopropila) por 4 horas a 37°C sob agitação constante de 150rpm. O cultivo de bactérias foi centrifugado a 4000xg por 10 minutos a 4°C. O sobrenadante foi descartado e o *pellet* de bactérias solubilizado em 300 $\mu$ l de tampão de amostra (240mM TrisHCl pH 6,8, 0,8% SDS, 200mM  $\beta$ -mercaptoetanol, 40% Glicerol e traços de Azul de Bromofenol). Os extratos brutos foram fracionados em gel de 12% poliacrilamida com SDS corado em solução de *Coomassie Blue* [50% metanol, 10% ácido acético e 0,25% Coomassie R250 (Amersham Biosciences®)].

### **3.3.3 Detecção da expressão das proteínas recombinantes através de *Immunoblotting***

As proteínas do extrato bruto de bactéria foram separadas através de eletroforese em gel de 12% poliacrilamida com SDS, para o candidato CT809, ou 8%, poliacrilamida no caso dos candidatos CT101 e CT802. Em seguida, foram transferidas para membranas de nitrocelulose (Hybond™ P - Amersham Biosciences®) durante 90 minutos sob corrente constante de 0,8mA/cm<sup>2</sup> no sistema *Nova blot* (Amersham Biosciences®) em tampão de transferência (39mM Glicina, 48mM TrisHCl pH 7,4, 0,037% de SDS e 20% de metanol). Para avaliação da eficiência de transferência, as membranas de nitrocelulose foram coradas com 0,1% *Ponceau* (Sigma®) para a visualização das proteínas.

As membranas foram, então, bloqueadas durante 1 hora a temperatura ambiente com PBST (137mM NaCl, 2,5mM KCl, 10mM Na<sub>2</sub>PO<sub>4</sub>, 2mM KH<sub>2</sub>PO<sub>4</sub> e 0,05% de Tween 20) contendo 5% de leite desnatado liofilizado (Molico®, Nestlé). Posteriormente, as membranas foram incubadas por 1 hora a temperatura ambiente

com anticorpo primário anti-*His-tag* (Amersham Biosciences®) na diluição 1:10.000 em PBST contendo 5% de leite. Após três lavagens de 10 minutos cada com PBST, as membranas foram incubadas por 1 hora a temperatura ambiente com anticorpo secundário anti-IgG de camundongo (Amersham Biosciences®) conjugado com peroxidase na diluição de 1:10.000 em PBST contendo 5% de leite. Após três lavagens de 10 minutos cada com PBST, as reações foram reveladas com a solução *ECL Western blotting analysis system* (Amersham Biosciences®) e as membranas expostas a filmes radiográficos (Hyperfilm™ MP - Amersham Biosciences®).

### 3.3.4 Teste de solubilidade das proteínas recombinantes

Para o teste de solubilidade das proteínas recombinantes foram utilizados 5ml de cada cultura de células após indução da expressão das proteínas recombinantes, conforme descrito anteriormente. As bactérias foram sedimentadas por centrifugação a 10000xg por 2 minutos a temperatura ambiente. O *pellet* de células foi então solubilizado em 10ml de tampão de lise (100mM NaCl, 10mM Tris pH 8,0, 50mM NaH<sub>2</sub>PO<sub>4</sub>).

Posteriormente, as células foram lisadas com o auxílio de uma *French*® *Pressure Cell Press* (Sim-Aminco®) a 1000psi por quatro vezes. Em seguida, o lisado celular foi centrifugado a 16000xg por 5 minutos a 4°C. O sobrenadante, contendo as proteínas solúveis, e o *pellet*, contendo as proteínas insolúveis, foram separados e reservados. Para a solubilização das proteínas insolúveis, o *pellet* foi ressuspendido em 10ml de tampão de lise contendo 6M uréia. Foi adicionado tampão de amostra (240mM Tris pH 6,8, 0,8% SDS, 200mM β-mercaptoetanol, 40% Glicerol e traços de Azul de Bromofenol) a uma alíquota da porção solúvel e outra da

porção insolúvel e os produtos finais foram fracionados em gel de 12% poliacrilamida com SDS, para o candidato CT809, ou 8% poliacrilamida, no caso dos candidatos CT101 e CT802, e corados em solução de *Coomassie blue*.

### 3.3.5 Purificação das proteínas recombinantes

Para a produção de uma grande quantidade de proteína recombinante, o clone de bactéria que apresentou uma maior expressão dentre os quatro analisados para cada proteína, foi crescido em 500ml de meio CG e a expressão da proteína foi induzida nas mesmas condições descritas anteriormente. Em seguida, a cultura de células foi centrifugada a 16000xg por 10 minutos a 4°C. O *pellet* de células foi solubilizado em 20ml de tampão de lise e, posteriormente, as células foram lisadas com o auxílio de uma *French® Pressure Cell Press* (Sim-Aminco®) a uma pressão de 1000psi por quatro vezes. Após a lise, o lisado foi centrifugado a 16000xg por 10 minutos a 4°C. O sobrenadante contendo as proteínas solúveis já prontas para serem purificadas foi reservado enquanto o *pellet* contendo as proteínas insolúveis presentes em corpos de inclusão foi ressuspensionado em 10ml de tampão de lise contendo 6M uréia.

Em seguida, as soluções passaram por uma resina de agarose carregada com Níquel (*Ni-NTA resin* – Invitrogen®) para purificação das proteínas recombinantes. Assim, 3ml de resina foi empacotado em uma seringa de 10ml contendo lã de vidro. As colunas foram equilibradas com tampão de lise, no caso das proteínas solúveis, e tampão de lise-uréia, no caso das proteínas insolúveis. Em seguida, os sobrenadantes contendo as proteínas recombinantes foram aplicados em sua respectiva coluna. Posteriormente as colunas foram lavadas com 10ml do mesmo tampão. Após esta

lavagem, seguiram-se eluições seriadas utilizando diferentes concentrações de Imidazol (Sigma®). As eluições foram feitas com 5ml de tampão de lise, no caso das proteínas solúveis, e 5ml de tampão de lise-uréia, no caso das proteínas insolúveis, contendo 10, 25, 50, 100 e 500mM Imidazol. As frações foram coletadas separadamente e, em seguida, analisadas em gel de poliacrilamida corado com *Coomassie blue*.

Após verificar quais frações continham a maior parte de cada uma das proteínas recombinantes, as mesmas foram reunidas e dialisadas para a remoção total do Imidazol e parcial da uréia, no caso das proteínas insolúveis. As diálises foram feitas a 4°C em membrana de celulose (*Dialysis Tubing-12KDa-Sigma®*) contra o mesmo tampão, mas reduzindo-se gradativamente a concentração de uréia (4M, 2M e 1M) a cada 6 horas de diálise, no caso das proteínas insolúveis.

Ao final da diálise foi feita a dosagem das proteínas com o reagente de Bradford (Biorad®). A dosagem foi feita a 595nm em leitor de ELISA (*Microplate Reader Benchmark – Biorad®*).

### **3.3.6 Exclusão da proteína Tioredoxina fusionada com a proteína recombinante CT802**

Como já descrito anteriormente, a fase aberta de leitura correspondente ao candidato CT802 foi clonada no vetor pET-32a permitindo a expressão da proteína CT802 em fusão com a proteína Tioredoxina. Anticorpos anti-tioredoxina são comuns em pessoas com alterações na tireóide e para a detecção de anticorpos contra a proteína CT802 recombinante nos plasmas de pacientes com câncer foi necessário remover a Tioredoxina da proteína recombinante.

Na construção utilizada, entre a Tioredoxina e a proteína CT802 recombinante havia um sítio de enterokinase e, desta forma, optamos pela digestão com esta enzima para a remoção da Tioredoxina. Para tal utilizamos o kit *Enterokinase Cleavage Capture Kit* (Novagen<sup>®</sup>) tendo sido utilizada 1 unidade da enzima enterokinase para clivar cada 100µg da proteína recombinante. A digestão foi realizada em tampão apropriado por 16 horas a temperatura ambiente e foi avaliada em gel de 12% poliacrilamida com SDS corado com *Coomassie blue*. Após a clivagem, o produto digerido foi purificado em uma coluna de agarose (*EKapture™ Agarose*) a qual possui afinidade pela enzima enterokinase. Desta forma, o filtrado continha as proteínas CT802 e Tioredoxina separadas e o mesmo foi utilizado para a detecção de anticorpos anti-CT802 em plasmas de pacientes após a dosagem da proteína com o auxílio do reagente de Bradford (Biorad<sup>®</sup>). A dosagem foi feita a 595nm em leitor de ELISA (*Microplate Reader Benchmark – Biorad<sup>®</sup>*).

### **3.4 DETECÇÃO DE ANTICORPOS CONTRA AS PROTEÍNAS RECOMBINANTES EM PLASMA DE PACIENTES**

#### **3.4.1 Amostras utilizadas**

Amostras de plasma de pacientes com câncer de útero (22 amostras) e pulmão (13 amostras) foram solicitadas junto ao banco de tumores do Hospital A.C. Camargo. Além disso, foram utilizados como controle negativo plasmas de 30 doadores do banco de sangue do Hospital A.C. Camargo. Todas as amostras foram utilizadas somente após o consentimento informado dos pacientes e a aprovação do

Comitê de Ética do Hospital. Todas as precauções necessárias para manter o sigilo e a confidencialidade dos dados dos pacientes foram adotadas.

### 3.4.2 *Immunoblotting*

As proteínas recombinantes purificadas foram fracionadas através de eletroforese em gel de 12% poliacrilamida com SDS, para a proteína CT809, ou 8% poliacrilamida, no caso das proteínas CT101 e CT802. Em seguida, foram transferidas e imobilizadas em membranas de nitrocelulose como descrito no item 3.3.3. Em cada canaleta dos géis foi aplicado 500ng de proteína recombinante. Para avaliação da eficiência de transferência, as membranas de nitrocelulose foram coradas com *Ponceau* 0,1% (Sigma®) permitindo a visualização das proteínas.

As membranas foram bloqueadas durante 1 hora a temperatura ambiente com PBST contendo leite 5%. Posteriormente, as membranas foram incubadas, por 1 hora a temperatura ambiente, com o plasma na diluição 1:100 em leite 1%. Após 3 lavagens de 10 minutos cada com PBST, o anticorpo secundário anti-IgG humana conjugado com peroxidase (Amersham Biosciences®) foi adicionado na diluição 1:20.000 em PBST contendo leite 5% e incubado por 1 hora a temperatura ambiente. Ao final, após 3 lavagens com PBST, a reação foi revelada e as membranas expostas a filmes radiográficos *Hyperfilm™ ECL™* (Amersham Biosciences®).

Os ensaios foram realizados em duplicata e nos casos em que o resultado não foi concordante um terceiro *immunoblotting* foi realizado. Além disso, para cada ensaio também foi feito um controle positivo utilizando-se o anticorpo primário anti-*His-tag* (Amersham Biosciences®) na diluição 1:10.000.

### 3.5 OBTENÇÃO DA SEQUÊNCIA COMPLETA DOS CANDIDATOS A ANTÍGENOS CTS REPRESENTADOS EXCLUSIVAMENTE POR ESTS

A sequência completa dos novos antígenos CTs que não possuíam sequência completa de mRNA depositada em banco de dados públicos, CT704 e CT1001, foi gerada através da técnica de RACE (*Rapid Amplification of cDNA Ends*) que permite a amplificação das extremidades 5' e 3' de sequências transcritas. Para tanto, foi utilizado o *Marathon™ Amplification cDNA kit* (CLONTECH® n°K1802-1). A síntese da fita dupla de cDNA a partir de RNA poli A<sup>+</sup> e a ligação de adaptadores, importantes na obtenção dos fragmentos 5' e 3' dos transcritos às moléculas de cDNAs, foram realizadas segundo instruções do fornecedor.

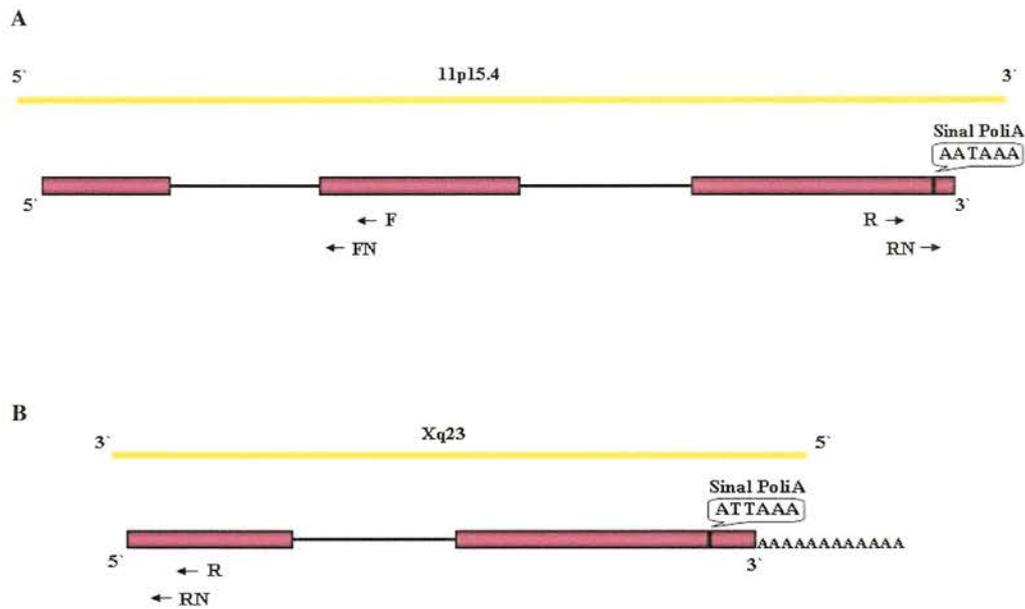
Primeiramente, foi feita uma PCR com iniciadores específicos para os adaptadores e para a porção interna dos transcritos de interesse (Tabela 3) em presença de 2,5µl de cDNA, 0,2mM dNTPs, 0,2µM de cada iniciador e 1U de *Advantage Taq DNA polimerase* (Clontech®), em tampão apropriado. A reação iniciou-se com 5 ciclos de 5 segundos a 94°C e 3 minutos a 72°C, seguida por 5 ciclos de 5 segundos a 94°C, 10 segundos a 70°C e 3 minutos a 72°C e, por último, 25 ciclos de 5 segundos a 94°C, 10 segundos a 68°C e 3 minutos a 72°C. Em seguida, foi feita uma *nested* PCR, na qual foi utilizada como molde 5µl do produto da primeira reação diluído 50 vezes. A reação foi feita nas mesmas condições, mas com iniciadores internos aos utilizados na primeira, tanto para o adaptador quanto para a sequência de interesse (Tabela 4). Além disso, a ciclagem consistiu de 20 ciclos de 5

segundos a 94°C, 10 segundos a 68°C e 3 minutos a 72°C. Os produtos de PCR foram analisados em gel de 1% agarose.

**Tabela 4** – Iniciadores utilizados nas reações de RACE para gerar a sequência completa dos transcritos CT704 e CT1001.

<i>Candidato</i>	<i>Iniciadores para a primeira reação</i>	<i>Iniciadores para a reação Nested</i>
CT704	F - 5'GATTTCCAGACCTTGTCCAAGCTCC3' R - 5'CACTCCTCTTCACTCCTTTCTATGGC3'	FN - 5'GTGGCCAACTGAGCTGCAGACTTCCC3' RN - 5'CTGCCTTGGATGGTGACATGAGTG3'
CT1001	R - 5'CACACAGAATGCCGCTCGCTAGGAG3'	RN - 5'GATGTTTCCTTTAGGCCTCTTCTCC3'

Vale ressaltar que para o candidato CT1001 não foram construídos iniciadores para a amplificação da extremidade 3' do transcrito uma vez que a sequência disponível em bancos de dados públicos já apresentava um sinal de poliA e a cauda poliA. A Figura 11 representa o alinhamento dos transcritos CT704 (Figura 11A) e CT1001 (Figura 11B) com a sequência genômica e a posição dos iniciadores na sequência dos mesmos.



**Legenda:** A linha amarela representa a seqüência genômica, os quadrados rosas os exons e as linhas pretas os introns. As flechas indicam as regiões onde foram construídos os iniciadores em cada seqüência. F: *Forward*, FN: *Forward* para reação *nested*, R: *Reverso* e RN: *Reverso* para a reação *nested*.

**Figura 11** – Representação esquemática do alinhamento entre a seqüência genômica e as seqüências disponíveis em bancos de dados públicos para os transcritos CT704 (A) e CT1001 (B).

Após a amplificação, os produtos foram clonados utilizando o *TA Cloning Kit* (Invitrogen<sup>®</sup>), segundo especificações do fornecedor, e conforme descrito no item 3.3.1. Em seguida, os plasmídeos obtidos na clonagem contendo os insertos de interesse foram seqüenciados como descrito anteriormente no item 3.2.8. As seqüências foram alinhadas contra a seqüência genômica humana através da ferramenta *BLAST* e a montagem das mesmas foi realizada manualmente através da busca por sobreposições. As seqüências geradas foram submetidas ao banco de dados de ESTs (dbEST) do GenBank tendo sido identificadas com os números de acesso de EF537578 à EF537582.

## ***RESULTADOS***



## **4 RESULTADOS**

### **4.1 SELEÇÃO DE GENES CANDIDATOS COM PADRÃO DE EXPRESSÃO CARACTERÍSTICO DE ANTÍGENOS CT NO BANCO DE DADOS DO TRANSCRIPTOMA HUMANO (MAP4)**

#### **4.1.1 Análise da distribuição de ESTs nos *clusters* correspondentes a antígenos CT conhecidos do Banco de Dados do Transcriptoma Humano (Map4)**

Para avaliar a possível utilização do banco Map4 na identificação de genes candidatos a novos antígenos CT, analisamos a distribuição de ESTs nos *clusters* correspondentes a 20 antígenos CT já descritos na literatura (Tabela 5). Esta análise serviu para definir os critérios de seleção dos *clusters* candidatos e também para avaliar possíveis limitações e artefatos de nossa estratégia.

**Tabela 5** - Lista dos 20 antígenos CT já descritos na literatura utilizados para a definição dos critérios para seleção dos *clusters* candidatos.

<i>Antígeno CT</i>	<i>Nº de acesso no GenBank</i>
MAGE-A1	NM_004988
MAGE-A3	NM_005362
MAGE-A10	NM_021048
MAGE-B1	NM_002364
CT7/MAGE-C1	NM_005462
CT10/MAGE-E1	NM_016249
NY-ESO-1	NM_001327
SSX-1	NM_005635
SSX-2	NM_003147
SSX-4	NM_005636
CT16/PAGE-5	NM_130467
HOM-TES-85	NM_016383
BRDT/CT9	NM_001726
CTp11/SPANX	NM_022661
OY-TES-1	NM_032489
CTAGE	NM_022663
CT15/Fert Beta	NM_001464
TRAG-3	NM_004909
IL-13RA2	NM_000640
HCA661/E2F-Like	NM_016521

Todos os 20 antígenos utilizados na análise estavam representados no banco de dados por pelo menos uma EST e 50% dos mesmos apresentavam em seus respectivos *clusters* um número de seqüências menor que 10, sugerindo um baixo nível de expressão destes genes. Notamos que as seqüências transcritas correspondentes a dois antígenos CT (NYESO-1 e BRDT/CT9) alinhavam com a mesma identidade na seqüência genômica em mais de uma localização sugerindo a ocorrência de duplicação gênica, já descrita na literatura para o antígeno NY-ESO-1

(ARADHYA et al. 2001). Desta forma, os 20 antígenos CT estavam representados no banco por 23 *clusters*.

Com relação à origem tecidual das seqüências expressas que compunham os *clusters* correspondentes aos antígenos CT conhecidos notamos que, ao contrário do esperado, apenas 10% apresentavam um perfil de distribuição de seqüências expressas característico, ou seja, *clusters* compostos por seqüências derivadas de testículo e tecidos tumorais. Isso porque somente 39,1% dos *clusters* possuíam ESTs provenientes de bibliotecas de cDNA de testículo e a grande maioria dos *clusters* (85%) possuía ESTs derivadas de bibliotecas de tecido normal. Notamos também que a grande maioria dos *clusters* apresentava ESTs cuja informação sobre o tecido de origem não constava em nosso banco de dados.

A ausência de ESTs provenientes de bibliotecas de cDNA de testículo provavelmente é um reflexo do reduzido número de ESTs originadas a partir desse tecido e depositadas em bancos de dados públicos. Das 3.422.614 ESTs catalogadas no banco Map4, 35.949 (1,05%) são derivadas de testículo. Assim sendo, a adoção da presença de ESTs de testículo nos *clusters* selecionados como candidatos a novos antígenos CT se tornaria um critério muito restritivo e a partir dessas informações a presença de ESTs de testículo passou a ser um critério de seleção facultativo.

A presença de ESTs cuja origem tecidual não constava em nosso banco de dados também poderia interferir diretamente na eficiência de seleção dos candidatos tornando a validação experimental um passo importante em nossa estratégia. Uma possível solução para esse problema seria realizar uma anotação manual de todas as bibliotecas catalogadas em nosso banco de dados. No entanto, devido ao elevado número de bibliotecas, um total de 7.330, esse processo seria muito demorado e nem

sempre eficiente uma vez que em muitos casos essas informações não poderiam ser extraídas da ficha descritiva de cada biblioteca de cDNA fornecida pelo GenBank.

Durante a análise do banco, KELSO et al. (2003) publicaram um trabalho propondo um novo sistema para a anotação das bibliotecas de cDNA disponibilizadas no domínio público, o eVOC. Assim sendo, entramos em contato com os autores do trabalho e adicionamos ao Map4 os dados de anotação gerados pelo eVOC. Para avaliar o impacto da introdução do sistema de classificação disponibilizado pelo eVOC, fizemos uma análise inicial comparando o nosso sistema de anotação com o eVOC. Para tanto, foram selecionados automaticamente *clusters* localizados no cromossomo X e que continham apenas ESTs derivadas de bibliotecas de testículo. Observamos que não houve diferença no número total de *clusters* selecionados, uma vez que ambos os sistemas selecionaram um total de 20 *clusters*. No entanto, apenas 7 *clusters* estavam presentes nas duas análises. Outra comparação realizada foi entre *clusters* localizados no cromossomo X que continham ESTs derivadas de bibliotecas de testículo e tecido tumoral. Para esta análise foram encontrados 28 *clusters* exclusivos de cada análise e 34 *clusters* em comum. Estes dados demonstraram que, apesar de haver diferença entre os *clusters* selecionados com base em cada uma das anotações, o número total de *clusters* não diferiu significativamente. Assim sendo, concluímos que os dois sistemas de classificação eram equivalentes e complementares e, desta forma, decidimos manter nossa seleção com base nos dados de anotação inicialmente disponíveis no Banco de Dados do Transcriptoma Humano.

Por fim, analisamos a presença de um número significativo de seqüências derivadas de bibliotecas de cDNA de tecidos normais nos *clusters* correspondentes a

antígenos CT conhecidos. Foi observado que a presença de seqüências de bibliotecas de tecido normal estava relacionada basicamente a dois fatores:

- 1 Falhas no processo de anotação das bibliotecas de cDNA, como discutido anteriormente. Neste caso, devido à ausência de uma nomenclatura padrão para a submissão de seqüências expressas no GenBank torna-se difícil determinar a origem normal ou tumoral de uma biblioteca. Assim, podem ocorrer erros na classificação da biblioteca, acarretando na presença em um determinado *cluster* de seqüências, erroneamente classificadas como oriundas de tecido normal.
- 2 Seqüências transcritas com múltiplos alinhamentos na seqüência genômica humana e fazendo parte de diferentes *clusters*. Alinhamentos múltiplos de ESTs na seqüência genômica são decorrentes da existência de famílias gênicas extensas com identidade em nível de nucleotídeos superior a 95% entre os membros. Desta forma, devido aos parâmetros de alinhamento estabelecidos para a construção do banco de dados, ESTs correspondentes a um determinado membro de uma família gênica também alinham nos *clusters* correspondentes aos demais membros. No caso específico dos antígenos CTs este fato foi de especial importância e interferiu diretamente no processo de seleção. Isso porque alguns membros das famílias dos CTs apresentam expressão ubíqua em vários tecidos normais e não correspondem a antígenos tumorais. Assim sendo, o alinhamento de ESTs correspondentes a membros de expressão ubíqua em *clusters* de antígenos CTs verdadeiros altera o perfil de expressão *in silico*, introduzindo de forma equívoca um número

significativo de ESTs derivadas de tecidos normais nos *clusters* de antígenos CTs.

Para solucionarmos o problema referente aos múltiplos alinhamentos de seqüências expressas no genoma humano foi criado um novo banco de dados, o Map4 *best hits* (Map4\_bh). Este banco continha apenas o melhor alinhamento de cada seqüência sendo que este foi escolhido com base na extensão do alinhamento em relação ao tamanho total da seqüência multiplicado pela identidade média dos alinhamentos no caso de seqüências com mais de um exon. Através de uma reanálise da lista de 20 CTs no Map4\_bh foi possível verificar que seqüências que alinhavam com diferentes membros de uma mesma família, no Map4\_bh passaram a alinhar em apenas um local na seqüência genômica, reduzindo o número de seqüências derivadas de tecidos normais nos *clusters* correspondentes a antígenos CTs conhecidos.

Assim sendo, frente às limitações apresentadas por nossa estratégia e a resolução parcial de alguns problemas encontrados, a implementação de um processo de inspeção manual dos clusters candidatos seguida por uma validação experimental do padrão de expressão tornou-se essencial e de grande importância para a correta identificação de novos antígenos CT.

#### **4.1.2 Seleção automática de candidatos a novos antígenos CT no Map4\_bh**

Para a procura por novos antígenos CT, um especialista em bioinformática de nosso laboratório desenvolveu um *pipeline* de seleção que realizou automaticamente buscas no banco Map4\_bh por *clusters* candidatos constituídos por ESTs derivadas de bibliotecas de testículo e/ou tecido tumoral. Também foi utilizado como critério

de seleção a presença de pelo menos uma EST com *splicing* nos *clusters* selecionados. Isso porque a presença de *splicing*, caracterizada por interrupções no alinhamento devido à presença de introns na sequência genômica, constitui uma evidência concreta de transcrição eliminando a hipótese de contaminação com DNA genômico. Desta forma, foram geradas três listas totalizando 1255 *clusters* selecionados (Tabela 6).

**Tabela 6** - Listas geradas pelo *pipeline* e respectivos critérios utilizados para seleção dos candidatos.

<i>Lista</i>	<i>Critério de seleção</i>	<i>Nº clusters</i>
1	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de testículo	889
2	<i>Clusters</i> compostos por ESTs derivadas de bibliotecas de testículo e tecido tumoral	70
3	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de tecido tumoral	296
<b>Total</b>	-	1255

#### 4.1.3 Inspeção manual e seleção de *clusters* candidatos para validação experimental

Depois de selecionados automaticamente pelo *pipeline*, os *clusters* candidatos passaram por um processo de inspeção manual, através do programa BLAT, para a confirmação da presença de *splicing* e de sítios conservados de *splicing*. No total, foram selecionados para validação experimental 93 *clusters* candidatos (Tabela 7).

**Tabela 7** - *Clusters* selecionados para o processo de validação experimental.

<i>Lista</i>	<i>Critério de seleção</i>	<i>Nº clusters selecionados</i>
1	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de testículo	46
2	<i>Clusters</i> compostos por ESTs derivadas de bibliotecas de testículo e tecido tumoral	23
3	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de tecido tumoral	24
<b>Total</b>	-	93

## 4.2 VALIDAÇÃO EXPERIMENTAL DOS CANDIDATOS SELECIONADOS

O processo de validação experimental do padrão de expressão dos *clusters* candidatos foi realizado através de RT-PCR e dividido em 4 fases. Em um primeiro momento, os candidatos foram analisados quanto à expressão em testículo, sendo que apenas os que apresentaram um resultado positivo seguiram em nossa análise. A segunda fase consistiu na análise de expressão no painel de 20 tecidos normais e os candidatos que não apresentaram expressão em nenhum dos tecidos normais analisados seguiram para a próxima fase. Na terceira fase analisamos o perfil de expressão em 17 linhagens tumorais e, por fim, na última fase analisamos a expressão dos mesmos candidatos em 160 amostras de diferentes tipos de tumores.

### 4.2.1 Análise da expressão dos candidatos selecionados em testículo

O padrão de expressão dos 93 candidatos selecionados foi, inicialmente, testado em testículo sendo que durante esta fase as reações de RT-PCR foram padronizadas visando otimizar a especificidade das mesmas como descrito no item 3.2.7. Os problemas de amplificação se resumiram a: (1) inespecificidade da reação,

(2) ausência de amplificação e (3) amplificação de um produto com tamanho diferente do esperado. Desta forma, o processo de padronização consistiu em alterações na concentração de Cloreto de Magnésio, temperatura de anelamento dos iniciadores e adição de 1M Betaína à reação de amplificação. Além disso, os fragmentos com tamanho diferente do esperado também foram seqüenciados.

Após o processo de padronização das reações de RT-PCR, do total de 37 candidatos que apresentaram problemas de amplificação, 14 candidatos foram padronizados e incluídos aos 56 que inicialmente já haviam apresentado expressão em testículo. Assim, no total, 70 candidatos seguiram para a análise de expressão em tecidos normais. A tabela 8 descreve o número de *clusters* com expressão em testículo referentes a cada lista de candidatos e a tabela 9 identifica os 70 candidatos que passaram para a segunda fase do processo de validação experimental.

**Tabela 8** - Número de *clusters* que apresentaram expressão em testículo.

<i>Lista</i>	<i>Critério de seleção</i>	<i>Expressão em testículo</i>
1	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de testículo	35 <i>clusters</i>
2	<i>Clusters</i> compostos por ESTs derivadas de bibliotecas de testículo e tecido tumoral	14 <i>clusters</i>
3	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de tecido tumoral	21 <i>clusters</i>
<b>Total</b>	-	70 <i>clusters</i>

**Tabela 9** – Candidatos que apresentaram expressão em testículo e que seguiram para a segunda fase do processo de validação experimental.

<i>Candidato</i>	<i>EST</i>	<i>Cluster</i>	<i>Lista</i>
CT101	BI458651	308498	1
CT102	BF059393	308666	1
CT105	AI825179	309320	1
CT108	AI371849	310984	1
CT111	BG716916	310732	1
CT112	BF979763	310994	1
CT113	BG720627	309442	1
CT115	BG720643	309558	1
CT201	BE858821	4575	2
CT202	AI360454	8033	2
CT206	AI366092	161285	2
CT207	BF887749	217341	2
CT209	BF979574	229203	2
CT210	BM556856	231893	2
CT214	BE843995	14477	2
CT215	BQ672145	63767	2
CT217	AL079669	110062	2
CT218	BG333817	119418	2
CT301	BE735233	199908	3
CT302	BM927640	1343	3
CT303	AI283993	47329	3
CT304	BE727460	198738	3
CT305	BG470816	282481	3
CT306	BG767408	50329	3
CT308	BF374746	221481	3
CT309	BM550300	219345	3
CT310	BE409638	310868	3
CT311	BE144331	266301	3
CT312	BG479526	311682	3
CT402	BE729130	310312	3
CT403	BG478724	268519	3
CT405	BI856424	113312	3
CT406	BM423379	268967	3
CT407	BM471826	1543	3
CT408	BF914944	273975	3

Cont/ Tabela 9

CT409	AI814314	13591	3
CT410	BF906533	163205	3
CT411	BF305093	271693	3
CT412	BI038406	276159	3
CT702	BI459819	115142	1
CT703	BG722032	163859	1
CT704	AI652043	191830	1
CT705	AI671000	229229	1
CT708	BG720994	157495	1
CT709	AI962439	199084	1
CT710	AI979310	230013	1
CT711	BF222867	278361	1
CT712	BM921607	84214	1
CT801	AL043076	226439	1
CT802	BG771896	115050	1
CT803	BI463953	83864	1
CT804	BI561956	115392	1
CT805	BG719982	218691	1
CT806	BI458986	220223	1
CT807	BI822069	231429	1
CT808	AI371744	277565	1
CT809	BG722950	309078	1
CT810	BG771667	311052	1
CT812	BI460168	65327	1
CT901	BI335051	4501	2
CT902	BG260350	8901	2
CT903	BG033516	85762	2
CT905	BG772361	234175	2
CT1001	AA451827	308330	1
CT1002	AA683332	308622	1
CT1003	AA401549	309308	1
CT1005	AA854001	310960	1
CT1006	BQ429641	311476	1
CT1007	AI209084	311774	1
CT1008	AI203928	311894	1

**Legenda:** Para cada candidato estão listados o número de acesso da seqüência referente (EST), o *cluster* correspondente no Map4\_bh e a lista de seleção à qual o mesmo pertence.

#### **4.2.2 Análise de expressão em tecidos normais dos candidatos com expressão positiva em testículo**

A análise de expressão dos 70 candidatos seguiu para um painel de 20 tecidos normais: cérebro, mama, pulmão, próstata, intestino delgado, cerebelo, colon, cérebro fetal, fígado fetal, coração, rim, placenta, glândula salivar, músculo esquelético, medula espinhal, baço, estômago, timo, traquéia e útero. Dentre estes, 58 candidatos apresentaram expressão em pelo menos um tecido normal e foram excluídos das análises subseqüentes (Tabela 10). Dos 12 candidatos restantes, sete (CT101, CT112, CT704, CT802, CT806, CT809 e CT1001) não apresentaram expressão em nenhum dos tecidos analisados sendo considerados fortes candidatos a novos antígenos CT. Vale destacar que o CT704 também apresentou expressão em placenta, mas como é comum a expressão de antígenos CT neste tecido o candidato não foi excluído. Além disso, identificamos cinco candidatos (CT115, CT712, CT808, CT1002 e CT1005) expressos somente nos tecidos normais do sistema nervoso (cérebro, cerebelo, cérebro fetal e medula espinhal) o que os caracteriza como possíveis novos antígenos de uma sub-classe de antígenos CT, os antígenos *Cancer/Testis/Brain* (CTB). Os membros desta classe apresentam uma expressão em tecidos normais em testículo e tecidos do sistema nervoso e, também, em uma grande variedade de tumores. Tais características tornam estes antígenos novos alvos no tratamento do câncer e por este motivo continuamos a caracterização dos mesmos juntamente com os sete candidatos a antígenos CT encontrados. Sendo assim, os 12 candidatos identificados passaram para a etapa seguinte da validação experimental, a análise de expressão em linhagens celulares tumorais.

**Tabela 10** - Número de *clusters* que apresentaram expressão exclusiva em testículo e em testículo e tecidos normais do sistema nervoso.

<i>Lista</i>	<i>Critério de seleção</i>	<i>Expressão exclusiva em testículo</i>	<i>Expressão em testículo e tecidos do sistema nervoso</i>
1	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de testículo	7 <i>clusters</i>	5 <i>clusters</i>
2	<i>Clusters</i> compostos por ESTs derivadas de bibliotecas de testículo e tecido tumoral	0 <i>clusters</i>	0 <i>clusters</i>
3	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de tecido tumoral	0 <i>clusters</i>	0 <i>clusters</i>
<b>Total</b>	-	7 <i>clusters</i>	5 <i>clusters</i>

#### 4.2.3 Análise de expressão em linhagens celulares de origem tumoral dos candidatos com expressão exclusiva em testículo e expressão em testículo e tecidos normais do sistema nervoso

A análise de expressão dos 12 candidatos seguiu para um painel de 17 linhagens celulares tumorais originadas de 11 diferentes tecidos (Tabela 11). Dentre os sete candidatos que apresentavam um perfil característico de um antígeno CT, apenas dois (CT112 e CT806) não apresentaram expressão nas linhagens celulares analisadas. Já com relação aos cinco candidatos que apresentaram um perfil característico de um antígeno CTB, quatro (CT115, CT808, CT1002, CT1005) não apresentaram expressão nas linhagens celulares analisadas. Desta forma, cinco candidatos (CT101, CT704, CT802, CT809 e CT1001) a antígenos CT e um candidato (CT712) a antígeno CTB apresentaram-se expressos em pelo menos uma das linhagens tumorais analisadas totalizando seis candidatos a novos antígenos tumorais (Tabela 12).

**Tabela 11** - Dados de expressão nas linhagens celulares tumorais analisadas dos 6 candidatos que apresentaram um perfil de expressão característico de um antígeno CT ou CTB.

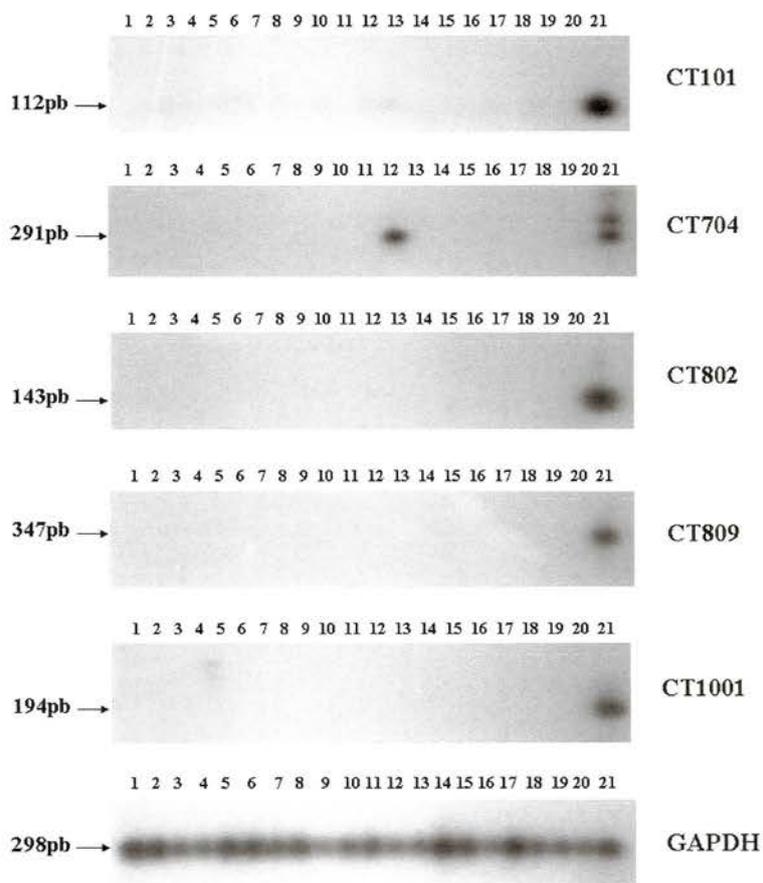
<i>Candidato</i>	<i>HL-60</i>	<i>SCABER</i>	<i>A172</i>	<i>T98G</i>	<i>MCF-7</i>	<i>MDA436</i>	<i>SW480</i>	<i>FADu</i>	<i>H1155</i>	<i>H358</i>	<i>A2058</i>	<i>SKmel-25</i>	<i>SAOS-2</i>	<i>DU145</i>	<i>PC3</i>	<i>Caski</i>	<i>Hela</i>
CT101											+						
CT704										+	+	+				+	
CT712							+	+			+						+
CT802		+			+					+							
CT809															+		
CT1001				+						+	+						+

**Tabela 12** - Número de *clusters* que apresentaram expressão em pelo menos uma linhagem celular analisada.

<i>Lista</i>	<i>Critério de seleção</i>	<i>Expressão exclusiva em testículo</i>	<i>Expressão em testículo e tecidos do sistema nervoso</i>
1	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de testículo	5 <i>clusters</i>	1 <i>cluster</i>
2	<i>Clusters</i> compostos por ESTs derivadas de bibliotecas de testículo e tecido tumoral	0 <i>clusters</i>	0 <i>clusters</i>
3	<i>Clusters</i> compostos por ESTs derivadas somente de bibliotecas de tecido tumoral	0 <i>clusters</i>	0 <i>clusters</i>
<b>Total</b>	-	5 <i>clusters</i>	1 <i>cluster</i>

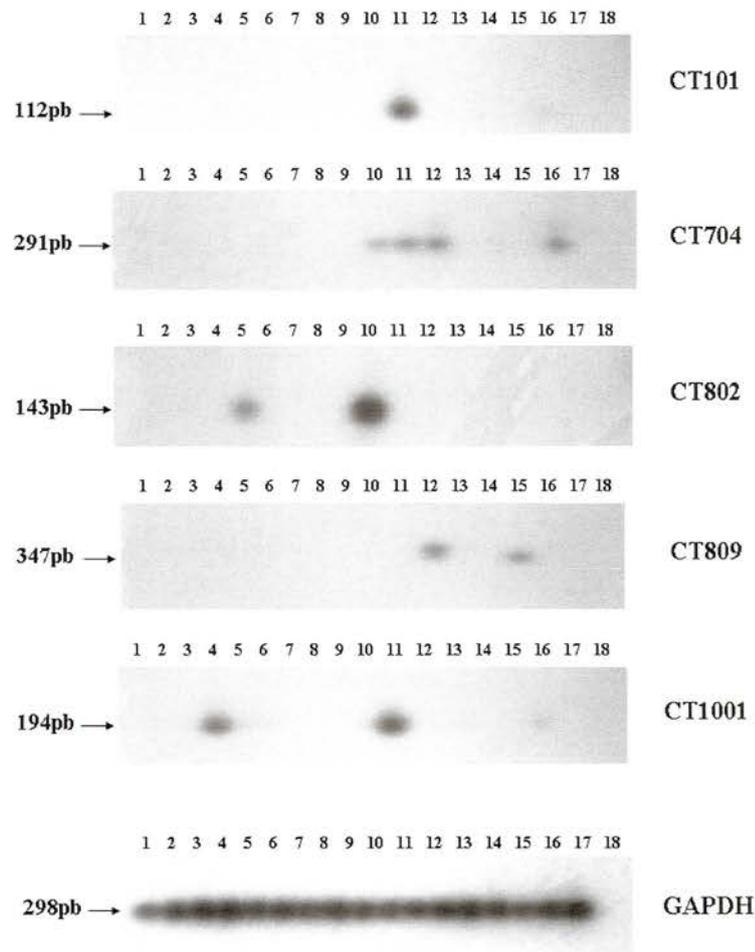
Os produtos de RT-PCR das amostras normais e de linhagens celulares de cada um dos candidatos a antígenos CT foram reanalisados por *Southern-blot* para se obter uma maior sensibilidade e especificidade em nossa análise, além de confirmar os dados já obtidos. Nas figuras 12 e 13, respectivamente, é possível visualizar a expressão única e específica destes candidatos em testículo dentre as amostras normais e confirmar os resultados de expressão obtidos através de RT-PCR nas linhagens celulares. Vale ressaltar que devido à alta sensibilidade da técnica foi possível encontrar expressão em um maior número de linhagens celulares tumorais em relação aos resultados anteriores avaliados apenas em gel de acrilamida. O

CT809, por exemplo, apresentou expressão na linhagem SKmel-25 de melanoma maligno enquanto o CT1001 também mostrou-se expresso na linhagem T98G de glioblastoma multiforme.



**Legenda:** 1- cérebro, 2- mama, 3- pulmão, 4- próstata, 5- intestino delgado, 6- cerebelo, 7- colon, 8- cérebro fetal, 9- fígado fetal, 10- coração, 11- rim, 12- placenta, 13- glândula salivar, 14- músculo esquelético, 15- medula espinhal, 16- baço, 17- estômago, 18- timo, 19- traquéia, 20- útero, 21- testículo.

**Figura 12** – Expressão dos candidatos CT101, CT704, CT802, CT809, CT1001 e do gene GAPDH, como controle da qualidade da síntese dos cDNAs, nas amostras de tecidos normais após 10 minutos de exposição.



**Legenda:** 1- HL-60 (leucemia aguda mielocítica), 2- SCABER (carcinoma de células escamosas da bexiga), 3- A172 (glioblastoma), 4- T98G (glioblastoma multiforme), 5- MCF-7 (adenocarcinoma de mama), 6- MDA436 (adenocarcinoma de mama), 7- SW480 (adenocarcinoma coloretal), 8- FADu (carcinoma de células escamosas da faringe), 9- H1155 (adenocarcinoma de pulmão), 10- H358 (adenocarcinoma de pulmão), 11- A2058 (melanoma), 12- SKmel-25 (melanoma maligno), 13- SAOS-2 (osteossarcoma), 14- DU145 (carcinoma de próstata), 15- PC3 (adenocarcinoma de próstata), 16- Caski (carcinoma epidermóide do cérvix uterino), 17- Hela (adenocarcinoma de útero), 18- controle negativo.

**Figura 13** – Expressão dos candidatos CT101, CT704, CT802, CT809, CT1001 e do gene GAPDH, como controle da qualidade da síntese dos cDNAs, nas linhagens celulares tumorais após 10 minutos de exposição.

FUNDAÇÃO ANTÔNIO PRUDENTE  
 Ana Maria Rodrigues Alves Kuninan  
 Diretora de Escola  
 RG 8.090.138

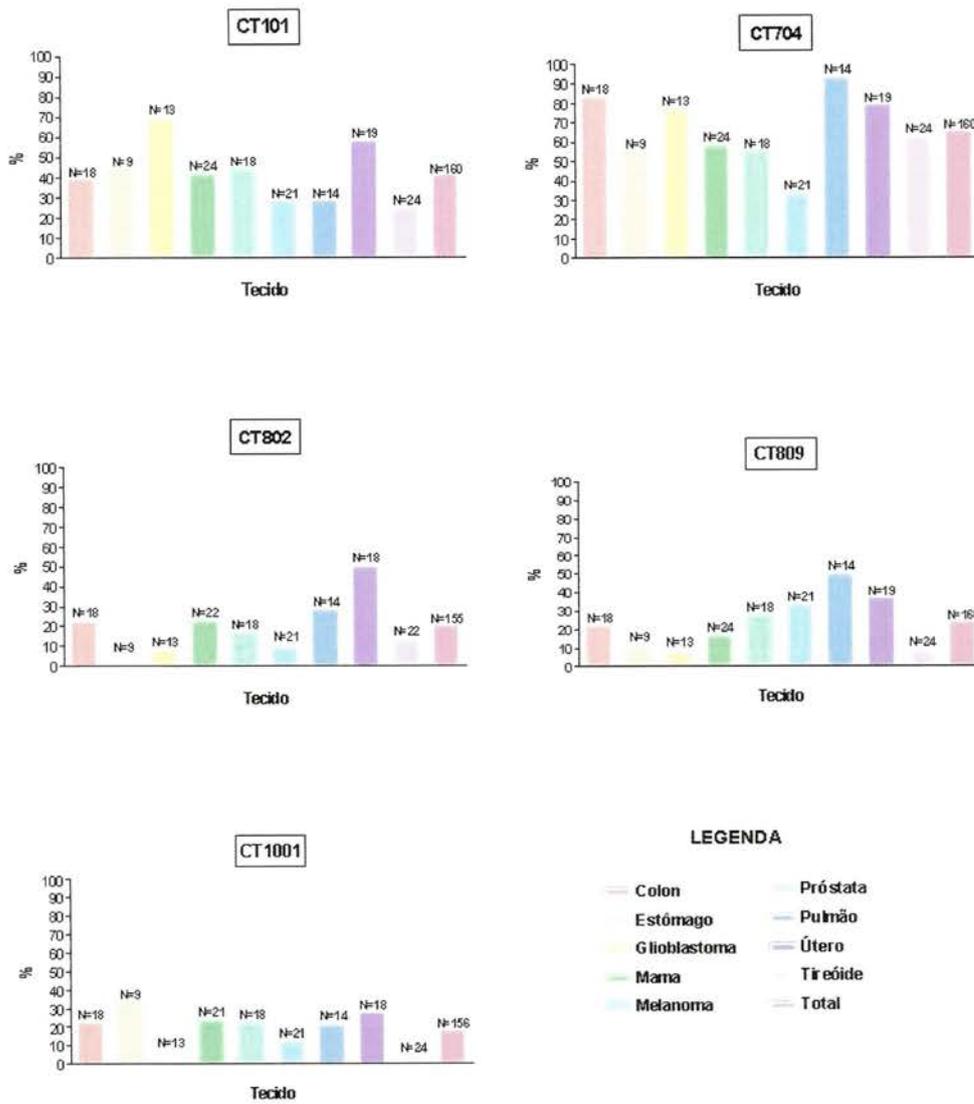
FUNDAÇÃO ANTÔNIO  
 PRUDENTE

#### 4.2.4 Análise de expressão dos candidatos em amostras de tumores de pacientes provenientes do Banco de Tumores do Hospital AC Camargo

A análise de expressão nas amostras de tumores foi realizada para os cinco candidatos (CT101, CT704, CT802, CT809 e CT1001) que apresentaram expressão exclusiva em testículo (Tabela 13 e Figura 14) e em pelo menos uma linhagem tumoral e também para os dois candidatos (CT112 e CT806) que não apresentaram expressão em nenhuma das linhagens tumorais. Acreditamos que por termos avaliado a expressão em um número restrito de linhagens tumorais, a ausência de expressão desses dois últimos candidatos (CT112 e CT806) nas mesmas não significaria que estes transcritos não estão expressos em tecidos tumorais. No entanto, os dois candidatos não apresentaram expressão nos tecidos tumorais analisados demonstrando serem possíveis genes com expressão exclusiva em testículo.

**Tabela 13** – Perfil de expressão dos cinco candidatos a antígenos CT nas amostras tumorais testadas por RT-PCR.

Amostras	CT101	CT704	CT802	CT809	CT1001
Colon	7/18 (39%)	15/18 (83,3%)	4/18 (22,2%)	4/18 (22,2%)	4/18 (22,2%)
Estômago	4/9 (44,5%)	5/9 (55,5%)	0/9 (0%)	1/9 (11,2%)	3/9 (33,3%)
Glioblastoma	9/13 (70%)	10/13 (77%)	1/13 (7,7%)	1/13 (7,7%)	0/13 (0%)
Mama	10/24 (41,7%)	14/24 (58,3%)	5/22 (22,7%)	4/24 (16,67%)	5/21 (23,8%)
Melanoma	8/18 (44,5%)	10/18 (55,5%)	3/18 (16,7%)	5/18 (27,8%)	4/18 (22,2%)
Próstata	6/21 (28,5%)	7/21 (33,3%)	2/21 (9,5%)	7/21 (33,3%)	3/21 (12,3%)
Pulmão	4/14 (28,5%)	13/14 (93%)	4/14 (28,5%)	7/14 (50%)	3/14 (21,4%)
Útero	11/19 (58%)	15/19 (79%)	9/18 (50%)	7/19 (37%)	5/18 (27,77%)
Tireóide	6/24 (25%)	15/24 (62,5%)	3/22 (13,7%)	2/24 (8,3%)	1/24 (4,16%)
Total	65/160 (41%)	104/160 (65%)	31/155 (20%)	38/160 (24%)	28/156 (18%)



**Figura 14** – Representação gráfica do perfil de expressão por RT-PCR dos cinco candidatos a antígenos CT nas amostras de tumores.

O candidato CT101 apresentou expressão em 41% (65/160) das amostras tumorais e uma alta frequência (70%) de expressão deste candidato foi observada nas amostras de glioblastoma. O candidato CT704 mostrou-se expresso em 65% (104/160) das amostras tumorais com alta frequência de expressão nas amostras de tumor de pulmão (93%) e colon (83,3%). O candidato CT802 apresentou expressão em 20% (31/155) das amostras tumorais e maior expressão nas amostras de tumor de útero (50%). O candidato CT809 mostrou uma expressão em 24% (38/160) das amostras tumorais e alta frequência de expressão nas amostras de tumor de pulmão (50%). E o CT1001 apresentou-se expresso em 18% (28/156) das amostras tumorais com maior expressão nos tumores de estômago (33,3%).

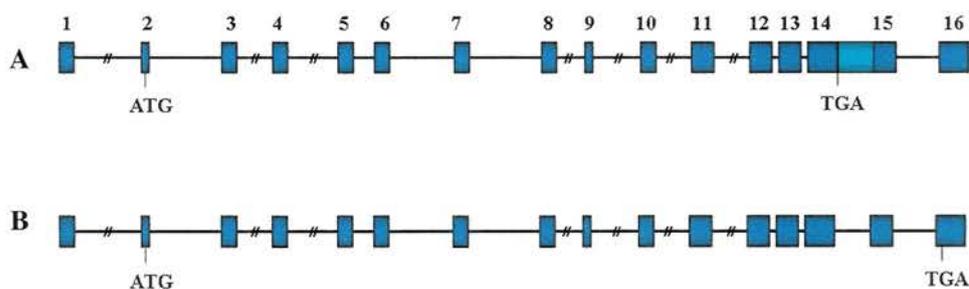
Uma vez validado o padrão de expressão de cinco candidatos a novos antígenos CT, partimos para a avaliação da resposta humoral contra as proteínas correspondentes aos candidatos que apresentavam uma sequência completa de mRNA depositada em banco de dados públicos. Desta forma, selecionamos os candidatos CT101, CT802 e CT809 cujas proteínas recombinantes foram utilizadas em ensaios de *Immunoblotting* para a detecção de anticorpos em plasma de pacientes com diferentes tipos de câncer.

### **4.3 ANÁLISE DA RESPOSTA HUMORAL CONTRA O CANDIDATO CT101 EM PACIENTES COM CÂNCER**

#### **4.3.1 Expressão e purificação da proteína recombinante**

O candidato CT101, representado em nosso banco de dados pela EST BI458651, corresponde ao gene PASD1 (*PAS domain containing protein 1*)

localizado na região cromossômica Xq28. A análise de sua estrutura gênica indicou que o mesmo contém 4120pb distribuídos em 15 exons e apresenta uma forma de *splicing* alternativo representando uma molécula de cDNA de 2850pb com 16 exons (Figura 15). Ambos os transcritos codificam proteínas com seqüência idêntica na extremidade N-terminal sendo que a variante maior (PASD1\_v1) apresenta retenção do intron 14 o qual introduz um códon de terminação após o aminoácido 639, enquanto a variante menor (PASD1\_v2) codifica 134 aminoácidos adicionais à extremidade C-terminal. Segundo dados de análise de expressão por RT-PCR, a variante PASD1\_v1 mostrou positividade em linhagens celulares derivadas de linfoma difuso de células B (DLBCL) enquanto a variante PASD1\_v2 apresentou expressão somente nas linhagens celulares relacionadas a um pior prognóstico. Desta forma, a expressão da proteína PASD1\_v2 pode ser considerada um marcador na identificação de pacientes com linfoma difuso de células B de alto risco (LIGGINS et al. 2004b).



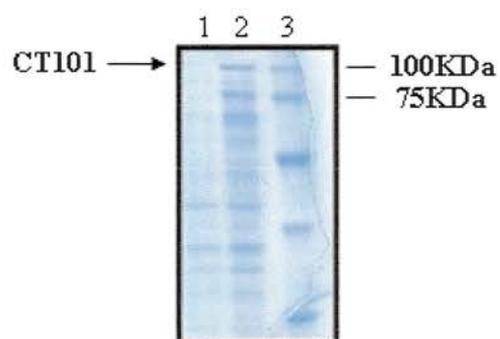
**Figura 15** – Figura esquemática da estrutura gênica das formas de *splicing* do transcrito PASD1. (A) PASD1\_v1, (B) PASD1\_v2.

A análise da seqüência da proteína PASD1 revelou a presença de dois domínios PAS sobrepostos na região N-terminal entre os aminoácidos 32-94 e 41-137 (LIGGINS et al. 2004b). Este domínio está envolvido no processo de transdução

de sinais e, além disso, proteínas que contém o domínio PAS estão associadas com a regulação do ritmo circadiano em eucariotos (PONTING e ARAVIND 2005). Vale ressaltar ainda a presença de um motivo de zipper de leucina (*leucine zipper motif*) e um sinal de localização nuclear sugerindo que a proteína PASD1 codifique um fator de transcrição. Estudos de *immunolabeling* utilizando anticorpos anti-PASD1 específicos para uma região comum entre as duas isoformas demonstraram que a expressão da proteína em tecidos normais está restrita ao núcleo de espermatogônias próximas à membrana basal dos túbulos seminíferos sendo, desta forma, consistente com a predição do sinal de localização nuclear em ambas as isoformas (COOPER et al. 2006).

Sendo assim, os iniciadores construídos para a amplificação da fase aberta de leitura deste transcrito tomaram como base a maior região codificante que corresponde a variante PASD1\_v2. A fase aberta de leitura foi amplificada por RT-PCR a partir de cDNA de testículo, clonada de forma direcionada no vetor de expressão pET-28a e, em seguida, os clones foram seqüenciados para confirmação da seqüência clonada.

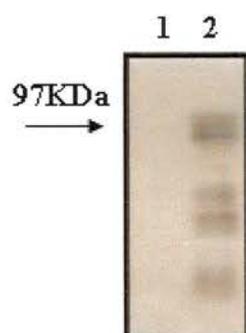
As colônias transformantes foram induzidas com 0,4mM de IPTG a 37°C por 4 horas. (Figura 16). Uma das bandas que aparece expressa apenas no extrato total de bactérias induzidas possui o tamanho esperado da proteína recombinante (97 KDa).



**Legenda:** (1) Extrato de bactéria não induzida. (2) Extrato de bactéria induzida com 0,4mM IPTG. (3) Marcador de peso molecular.

**Figura 16** – Extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT101.

Para a confirmação de que a banda predominante no extrato bruto de bactérias induzidas realmente correspondia à proteína recombinante CT101, o extrato foi analisado por *immunoblotting*, utilizando-se o anticorpo anti-*His-tag*. O anticorpo reconheceu várias bandas, mas somente no extrato de bactéria induzida confirmando a indução da expressão da proteína CT101 (Figura 17). Provavelmente as bandas de menor peso molecular reconhecidas pelo anticorpo correspondem a produtos de degradação da proteína recombinante.



**Legenda:** Os extratos de bactéria foram transferidos para uma membrana de nitrocelulose e, posteriormente, incubados com anticorpo anti-*His-tag*. (1) Extrato de bactéria não induzida. (2) Extrato de bactéria induzida por 4 horas a 37°C com 0,4mM IPTG.

**Figura 17** – *Immunoblotting* com extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT101.

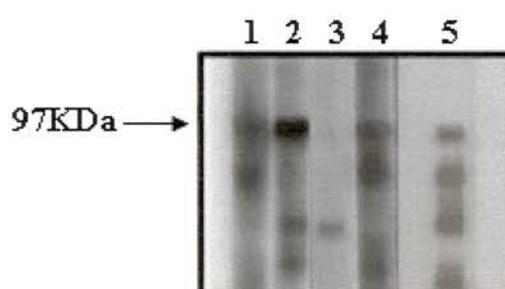
O teste de solubilidade da proteína CT101 indicou que a mesma encontra-se expressa nas formas solúvel e insolúvel. Assim, ambas as frações foram utilizadas na purificação da proteína sendo que esta foi eluída nas frações de 100 e 500mM de imidazol. Após a diálise, a proteína presente na fração solúvel foi armazenada em tampão de lise em concentração aproximada de 0,14µg/µL e a presente na fração insolúvel foi armazenada em tampão de lise contendo 2M de uréia em concentração aproximada de 0,2µg/µL.

#### 4.3.2 Detecção de anticorpos anti-CT101 em plasma de pacientes com câncer

Para completar a caracterização da proteína CT101 como um novo antígeno tumoral era necessário demonstrar a existência de resposta imune específica contra essa proteína em pacientes com câncer. Para tal, optamos por investigar a presença de anticorpos circulantes anti-CT101 em plasma de pacientes com câncer de útero por ter sido o tecido que apresentou maior expressão dentre os analisados por RT-

PCR. Assim, foram analisadas 22 amostras de plasma de pacientes com câncer de útero e plasmas de 30 indivíduos sadios como controle negativo.

A frequência de anticorpos anti-CT101 encontrada no plasma de pacientes com câncer de útero foi de 41% (9/22) (Figura 18). No entanto, sua frequência dentre os indivíduos sadios também foi significativa, sendo detectada a presença de anticorpos anti-CT101 em 43,3% (13/30) das amostras analisadas.



**Legenda:** A proteína recombinante CT101 foi fracionada e imobilizada na membrana, a qual foi incubada com: 1 e 2 - plasmas de pacientes com câncer de útero; 3 e 4 - plasmas de indivíduos sadios; 5 - Controle positivo incubado com anticorpo anti-*His-tag*.

**Figura 18** - Identificação de anticorpos anti-CT101 em plasma de pacientes com câncer de útero por *Immunoblotting*.

Problemas no processo de purificação da proteína recombinante ou até mesmo o reconhecimento de um epítipo apresentado devido à presença da cauda de histidina acoplada à proteína podem contribuir para a alta frequência de anticorpos nos indivíduos sadios. Desta maneira, para verificar se a alta positividade observada entre os plasmas de indivíduos normais poderia estar associada ao sistema de expressão utilizado decidimos avaliar a presença de anticorpos nos plasmas normais contra duas proteínas de cana-de-açúcar, CaneCPI-2 e CaneCPI-3. Estas proteínas foram expressas utilizando o mesmo protocolo descrito neste trabalho (GIANOTTI et al. 2006) e foram gentilmente cedidas pelo Dr. Flávio Henrique-Silva da

Universidade Federal de São Carlos. Por serem inibidores de cisteíno proteinases de planta, conhecidas como fitocistatinas, não esperávamos encontrar anticorpos contra essas proteínas nos plasmas humanos. No entanto, para cada uma das proteínas avaliadas encontramos dois plasmas com reatividade correspondendo a uma frequência de 6,7% (2/30) o que demonstra a existência de uma reação inespecífica. Isto nos mostra que parte da frequência de anticorpos anti-CT101 encontrada nos plasmas normais pode ser devido ao sistema de expressão utilizado.

#### **4.4 ANÁLISE DA RESPOSTA HUMORAL CONTRA O CANDIDATO CT802 EM PACIENTES COM CÂNCER**

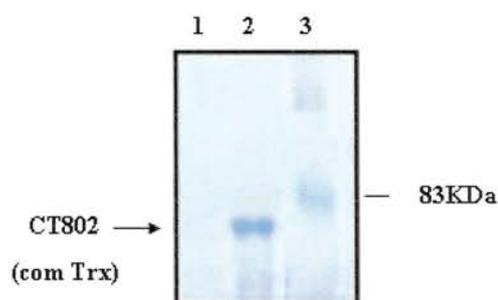
##### **4.4.1 Expressão e purificação da proteína recombinante**

O candidato CT802 representado em nosso banco de dados pela EST BG771896 corresponde ao gene ASZ1 ou GASZ (*Germ cell-specific ankyrin, SAM and basic leucine zipper domain containing protein 1*) localizado na região cromossômica 7q31.2. A análise de sua estrutura revelou um transcrito de 1831pb com 13 exons sendo que a região codificante compreende 1427pb correspondendo a 475aa.

A proteína codificada por este gene apresenta um domínio SAM (*Sterile alpha motif*) na posição 278–334, dois domínios de ankirina nas posições 77-202 e 176-234 e um domínio bZip (*basic leucine zipper*) na extremidade carboxi terminal. Por ser uma proteína citoplasmática, segundo dados de imunohistoquímica (YAN et al. 2002), pode agir através da interação com outras proteínas citoplasmáticas ou proteínas que podem alternar a presença no núcleo e no citoplasma. Assim, dada a

presença dos domínios SAM e de ankirina, sugere-se que a proteína GASZ apresente uma possível ação na regulação das vias de transdução de sinal citoplasmático, mediando interações proteína-proteína durante a maturação de células germinativas e na embriogênese.

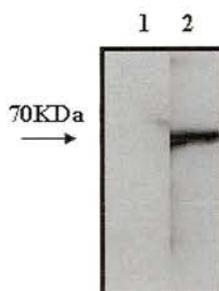
A fase aberta de leitura do transcrito CT802 foi amplificada por RT-PCR a partir de cDNA de testículo, clonada de forma direcionada no vetor de expressão pET-32a e os clones foram seqüenciados para confirmação da seqüência clonada. Em seguida, as colônias transformantes foram induzidas com 0,4mM de IPTG a 37°C por 4 horas (Figura 19). A banda que aparece fortemente expressa apenas no extrato total de bactérias induzidas corresponde ao tamanho esperado da proteína recombinante (57KDa) fusionada com a cauda de histidina e a proteína Tioredoxina (12,6KDa) presente no vetor utilizado.



**Legenda:** (1) Extrato de bactéria não induzida. (2) Extrato de bactéria induzida por 4 horas a 37°C com 0,4mM IPTG. (3) Marcador de peso molecular.

**Figura 19** – Extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT802 fusionada com Tioredoxina (Trx).

Como esperado, o anticorpo anti-*His-tag* reconheceu apenas a banda predominante do extrato bruto corado com *Coomassie Blue* (Figura 20).



**Legenda:** Os extratos de bactéria foram transferidos para uma membrana de nitrocelulose e, posteriormente, incubados com anticorpo anti-*His-tag*. (1) Extrato de bactéria não induzida. (2) Extrato de bactéria induzida por 4 horas a 37°C com 0,4mM IPTG.

**Figura 20** – *Immunoblotting* com extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT802 fusionada com Tioredoxina.

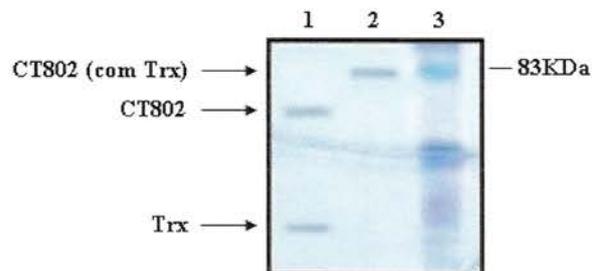
O teste de solubilidade da proteína recombinante CT802 indicou que a mesma encontra-se expressa na forma insolúvel. Assim, apenas a fração insolúvel foi utilizada para a purificação da proteína que foi eluída nas frações de 100 e 500mM de imidazol. Após a diálise, a proteína foi armazenada em tampão de lise contendo 1M de uréia em concentração aproximada de 0,36µg/µL.

#### 4.4.2 Detecção de anticorpos anti-CT802 em plasma de pacientes com câncer

Uma vez que a proteína recombinante CT802 foi expressa em conjunto com a proteína Tioredoxina e pessoas saudáveis podem apresentar anticorpos anti-Tioredoxina, decidimos retirar a Tioredoxina fusionada à proteína CT802 antes de avaliar a presença de anticorpos anti-CT802 nas amostras de plasma para evitar a ocorrência de reação cruzada.

A Tioredoxina foi retirada através de uma digestão enzimática com enterokinase como podemos ver na figura 21. Poderíamos, então, purificar a digestão e dessa forma retirar a Tioredoxina, mas optamos por não seguir este protocolo uma

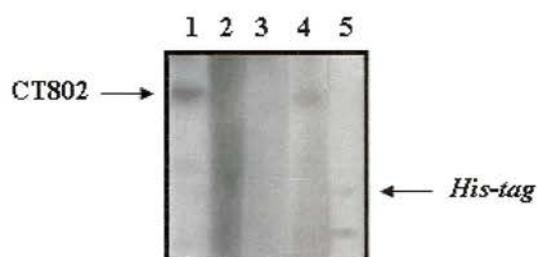
vez que a cauda de Histidina ficou acoplada à Tioredoxina e esta continuaria sendo o nosso controle positivo.



**Legenda:** (1) Proteína CT802 digerida. (2) Proteína CT802 fusionada com Tioredoxina. (3) Marcador de peso molecular.

**Figura 21** – Digestão enzimática da proteína CT802 com enterokinase.

Em seguida, partimos para a avaliação dos 30 plasmas normais e também a análise da presença de anticorpos anti-CT802 em plasma de 22 pacientes com câncer de útero por ter sido o tecido que apresentou maior expressão dentre os analisados por RT-PCR. A frequência de anticorpos anti-CT802 encontrada nos plasmas de pacientes com câncer de útero foi de 22,7% (5/22) enquanto dentre os indivíduos sadios a frequência foi de 6,7% (2/30), um valor aceitável quando comparado com os resultados encontrados para as proteínas de cana (Figura 22). Por termos encontrado frequências semelhantes entre a CT802 e as fitocistatinas acreditamos que este valor possa ser resultado de alguma reação inespecífica do sistema de expressão utilizado.



**Legenda:** A proteína recombinante CT802 foi fracionada e imobilizada na membrana, a qual foi incubada com: 1 e 2 - plasmas de pacientes com câncer de útero; 3 e 4 - plasmas de indivíduos saudios; 5 - Controle positivo incubado com anticorpo anti-*His-tag*.

**Figura 22** - Identificação de anticorpos anti-CT802 em plasma de pacientes com câncer de útero por *Immunoblotting*.

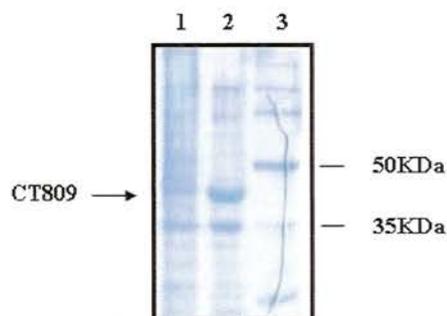
## 4.5 ANÁLISE DA RESPOSTA HUMORAL CONTRA O CANDIDATO CT809 EM PACIENTES COM CÂNCER

### 4.5.1 Expressão e purificação da proteína recombinante

O candidato CT809 representado em nosso banco de dados pela EST BG722950 corresponde ao gene FAM46D (*Family with sequence similarity 46, member D*) localizado na região cromossômica Xq21.1. Este transcrito contém 3008pb distribuídos em apenas 3 exons sendo que a região codificante compreende 1169pb correspondendo a 389aa (44KDa). A proteína hipotética codificada não apresenta nenhum domínio protéico conhecido.

A fase aberta de leitura do transcrito CT809 foi amplificada por RT-PCR a partir de cDNA de testículo e clonada no vetor de expressão pET-28a, já descrito anteriormente. Os clones foram seqüenciados para confirmação da seqüência clonada e as colônias transformantes foram induzidas com 0,4mM de IPTG a 37°C por 4 horas (Figura 23). A banda que aparece fortemente expressa apenas no extrato

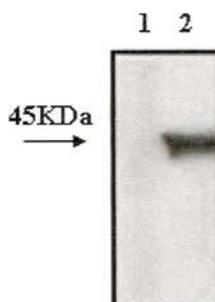
total de bactérias induzidas corresponde ao tamanho esperado da proteína recombinante (44KDa).



**Legenda:** (1) Extrato de bactéria não induzida. (2) Extrato de bactéria induzida por 4 horas a 37°C com 0,4mM IPTG. (3) Marcador de peso molecular.

**Figura 23** – Extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT809.

Como esperado, o anticorpo anti-*His-tag* reconheceu apenas a banda predominante do extrato bruto corado com *Coomassie Blue* (Figura 24).



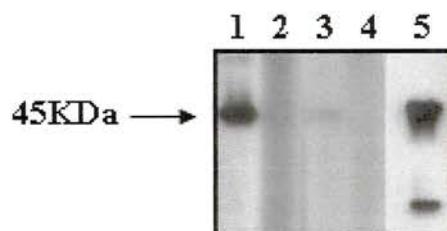
**Legenda:** Os extratos de bactéria foram transferidos para uma membrana de nitrocelulose e, posteriormente, incubados com anticorpo anti-*His-tag*. (1) Extrato de bactéria não induzida. (2) Extrato de bactéria induzida por 4 horas a 37°C com 0,4mM IPTG.

**Figura 24** – *Immunoblotting* com extrato bruto de bactéria BL21 Rosetta expressando a proteína recombinante CT809.

O teste de solubilidade da proteína recombinante CT809 indicou que a mesma encontra-se expressa na forma insolúvel. Assim, apenas a fração insolúvel foi utilizada para a purificação da proteína que foi eluída nas frações de 100 e 500mM de imidazol. Após a diálise, a proteína foi armazenada em tampão de lise contendo 2M de uréia em concentração aproximada de 0,45µg/µL.

#### **4.5.2 Detecção de anticorpos anti-CT809 em plasma de pacientes com câncer**

A presença de anticorpos anti-CT809 foi avaliada em plasma de pacientes com câncer de pulmão por ter sido o tecido que apresentou maior expressão dentre os analisados por RT-PCR. Assim, foram analisadas 13 amostras de plasma de pacientes com câncer de pulmão e plasmas de 30 indivíduos sadios como controle negativo. A frequência de anticorpos anti-CT809 encontrada no plasma de pacientes com câncer de pulmão foi de 7,7% (1/13) (Figura 25), valor encontrado para a maioria dos antígenos desta categoria já estudados. Já a frequência dentre os indivíduos sadios foi de 6,7% (2/30), novamente o mesmo valor encontrado para as proteínas de cana. Isto indica que possivelmente esta frequência corresponda a uma reação inespecífica representando uma limitação do método utilizado.



**Legenda:** A proteína recombinante CT809 foi fracionada e imobilizada na membrana, a qual foi incubada com: 1 e 2 - plasmas de pacientes com câncer de pulmão; 3 e 4 - plasmas de indivíduos saudáveis; 5 - Controle positivo incubado com anticorpo anti-*His-tag*.

**Figura 25** - Identificação de anticorpos específicos em plasma de pacientes com câncer de pulmão por *Immunoblotting*.

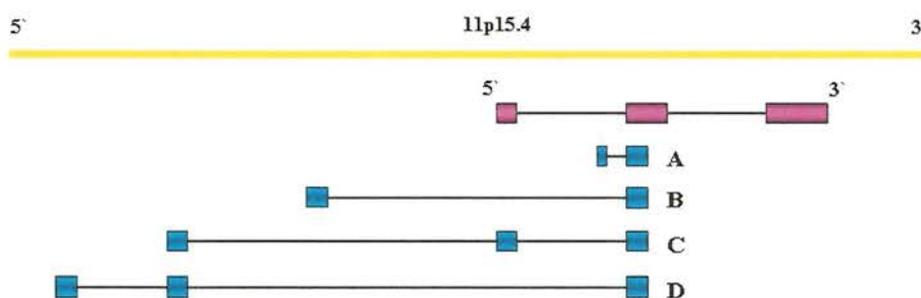
#### 4.6 OBTENÇÃO DA SEQUÊNCIA COMPLETA DOS CANDIDATOS CT704 E CT1001

A sequência completa dos novos antígenos CTs, CT704 e CT1001, foi gerada através da técnica de RACE (*Rapid Amplification of cDNA Ends*) que permite a amplificação das extremidades 5' e 3' de sequências expressas. Tendo conhecimento da sequência completa de cada candidato será possível expressar a proteína recombinante correspondente a cada transcrito e, desta forma, avaliar a resposta humoral contra as proteínas em pacientes com diferentes tipos de câncer.

##### 4.6.1 CT704

Para o candidato CT704, localizado na região cromossômica 11p15.4, foi possível amplificar quatro fragmentos específicos de 200pb, 300pb, 400pb e 600pb utilizando a reação de 5' RACE. Os produtos foram clonados conforme descrito no item 3.3.1 e, em seguida, os plasmídeos contendo os insertos de interesse foram sequenciados como descrito no item 3.2.8. As sequências foram alinhadas contra a

seqüência genômica humana através da ferramenta *BLAST* confirmando a especificidade das mesmas e a extensão da extremidade 5' do transcrito. Os quatro fragmentos gerados correspondem a formas alternativas de *splicing* (Figura 26) e as seqüências correspondentes às isoformas foram submetidas ao banco de dados de ESTs (dbEST) do GenBank. As seqüências foram identificadas com os números de acesso de EF537578 à EF537581. Para a extremidade 3' deste transcrito não foi possível gerar nenhum fragmento de extensão mesmo após diversas modificações nas condições da PCR.



**Legenda:** A linha amarela representa a seqüência genômica, os quadrados representam os exons e as linhas pretas os introns. A seqüência destacada em rosa representa a seqüência referente ao CT704 disponível em bancos de dados públicos. As seqüências em azul (A, B, C e D) representam as isoformas do mesmo transcrito identificadas pela técnica de RACE.

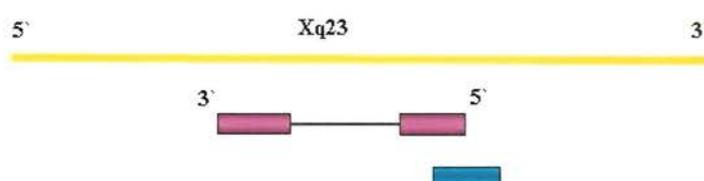
**Figura 26** – Representação esquemática do alinhamento entre a seqüência genômica e as seqüências geradas por RACE para extensão do transcrito CT704.

As seqüências consenso geradas pelos fragmentos A, B, C e D apresentam uma fase aberta de leitura de 324 nucleotídeos correspondendo a 107 aminoácidos. Esta fase aberta de leitura localiza-se nos dois exons em comum para as quatro isoformas e não apresenta nenhum domínio protéico tendo sido identificados apenas sítios de fosforilação.

Vale citar que novos iniciadores foram construídos na extremidade da seqüência para uma nova tentativa de extensão. Como não foi possível amplificar nenhum fragmento específico concluímos que havíamos identificado a seqüência completa referente ao transcrito.

#### 4.6.2 CT1001

Uma vez que o candidato CT1001, localizado na região cromossômica Xq23, já apresenta um sinal de poliA e cauda poliA em sua seqüência, realizamos apenas a reação de RACE para a extremidade 5' do transcrito. Foi possível amplificar um fragmento de 100pb que foi clonado, conforme descrito no item 3.3.1. Em seguida, o plasmídeo contendo o inserto de interesse foi seqüenciado como descrito no item 3.2.8. A seqüência foi alinhada contra a seqüência humana através da ferramenta BLAST confirmando a especificidade da mesma e a extensão da extremidade 5' do transcrito (Figura 27). A seqüência foi submetida ao banco de dados de ESTs (dbEST) do GenBank tendo sido identificada com o número de acesso EF537582.



**Legenda:** A linha amarela representa a seqüência genômica, os quadrados representam os exons e as linhas pretas os introns. A seqüência destacada em rosa representa a seqüência referente ao CT1001 disponível em bancos de dados públicos. A seqüência em azul representa a seqüência do mesmo transcrito identificada pela técnica de RACE.

**Figura 27** – Representação esquemática do alinhamento entre a seqüência genômica e a seqüência gerada por RACE para extensão do transcrito CT1001.

A seqüência consenso gerada apresenta uma maior fase aberta de leitura de 267 nucleotídeos entre as posições 287 e 553 correspondendo a 88 aminoácidos cuja seqüência apresenta apenas sítios de fosforilação. Novos iniciadores foram construídos na extremidade da seqüência para uma nova tentativa de extensão entretanto não foi possível amplificar nenhum fragmento específico demonstrando que havíamos identificado a seqüência completa referente ao transcrito.

## ***DISCUSSÃO***



## 5 DISCUSSÃO

Os antígenos *cancer/testis* são considerados candidatos ideais para o desenvolvimento de imunoterapia contra o câncer devido ao restrito padrão de expressão e alta imunogenicidade. No entanto, os antígenos CT apresentam um perfil de expressão bastante heterogêneo conforme o tipo de tumor representando um grande obstáculo para o desenvolvimento de uma imunoterapia eficiente para o tratamento do câncer.

Uma das possibilidades para contornar o problema da heterogeneidade de expressão destes antígenos é o desenvolvimento de vacinas polivalentes contendo diferentes epítomos derivados do maior número possível de antígenos CT. Desta forma, a identificação de novos antígenos CT capazes de induzir resposta imune em pacientes com câncer torna-se um importante passo para o desenvolvimento de vacinas e a aplicação da imunoterapia no tratamento de tumores.

Atualmente, com o aumento significativo no número de ESTs em bancos de dados públicos e considerando-se o perfil de expressão gênica dos CTs, a utilização de ferramentas de bioinformática para a análise *in silico* do padrão de expressão de qualquer transcrito tornou-se rápida, simples e de fácil acesso em relação as técnicas imunológicas utilizadas para identificação desses antígenos. Desta forma, com base na origem tecidual das ESTs correspondentes a um dado transcrito é possível selecionar candidatos a antígenos CTs cujas seqüências expressas sejam derivadas somente de bibliotecas de cDNA de testículo normal e de qualquer tecido tumoral. Esta análise pode ser, então, seguida por uma validação experimental através da

avaliação do padrão de expressão gênica dos candidatos selecionados e do perfil de imunogenicidade utilizando soros de paciente com câncer uma vez que os candidatos só poderão ser considerados novos antígenos *CT* após a identificação de resposta imune específica contra os mesmos em pacientes com câncer.

Para a identificação de novos antígenos *CTs* utilizamos o Banco de Dados do Transcriptoma Humano (Map4) construído pelo Laboratório de Biologia Computacional do Instituto Ludwig. No entanto, para avaliar a possível utilização do Map4 na identificação de genes candidatos a novos antígenos *CT*, antes da seleção dos candidatos analisamos a distribuição de ESTs nos *clusters* correspondentes a 20 antígenos *CT* já descritos na literatura (Tabela 4). Esta análise serviu para definir os critérios de seleção dos *clusters* candidatos e também nos indicou possíveis limitações e artefatos de nossa estratégia. Tais limitações incluíram, por exemplo, os múltiplos alinhamentos de seqüências expressas no genoma humano devido à presença de famílias gênicas e a presença de ESTs cuja origem tecidual não constava em nosso banco de dados.

Para solucionar pelo menos em parte os problemas encontrados, foi criado um novo banco de dados, o Map4\_bh, a partir do qual foram selecionados *clusters* candidatos constituídos por ESTs derivadas de bibliotecas de testículo e/ou tecido tumoral. Foram identificados um total de 1255 *clusters* candidatos sendo que após um processo de inspeção manual 93 candidatos foram selecionados para a validação experimental por RT-PCR. Para 23 candidatos não foi possível padronizar as reações de RT-PCR sendo que, desta forma, 70 candidatos passaram pelo processo de validação experimental. Foram identificados cinco candidatos (CT101, CT704, CT802, CT809 e CT1001) que apresentaram uma expressão exclusiva em testículo

dentre os tecidos normais e em pelo menos uma linhagem tumoral além de uma alta frequência de expressão dentre as amostras tumorais. Estes resultados indicam que a metodologia utilizada em nosso trabalho apresentou uma eficiência de validação de 7,1%.

Em um trabalho pioneiro utilizando dados de ESTs, SCANLAN et al. (2002b) foram capazes de identificar e validar experimentalmente três novos antígenos CT através da seleção de transcritos no banco de dados do Unigene representando uma eficiência de validação de 4,1%. Além disso, SCANLAN et al. (2002b) verificaram também que a análise executada falhou na identificação de nove antígenos CT conhecidos, todos eles membros de extensas famílias gênicas. É importante ressaltar que pelo menos 5 desses antígenos conhecidos apresentavam ESTs derivadas de bibliotecas de tecidos normais o que justifica o fato de não terem sido selecionados na análise.

Já no estudo desenvolvido por CHEN et al. (2005), que realizaram uma análise combinada de dados de MPSS e ESTs, foram selecionados, primeiramente a partir do banco de dados de MPSS, 1056 genes que apresentavam expressão pelo menos duas vezes maior em testículo quando comparada com outros tecidos normais sendo que destes genes, 15 correspondiam a antígenos CTs já conhecidos. Assim, 1041 genes foram, em seguida, analisados com base em bancos de dados públicos de ESTs sendo que foram considerados candidatos a antígenos CTs aqueles genes cujas ESTs não correspondiam a mais do que dois tecidos normais. Por fim, a partir da análise do perfil de expressão de 166 candidatos foi possível identificar cinco novos antígenos CTs (CHEN et al. 2005) representando uma eficiência de validação de 3%.

O fato de a nossa eficiência de validação ter sido maior que em outros trabalhos publicados deve-se primeiramente à análise prévia do banco de dados quanto a CTs já conhecidos que indicou possíveis limitações e artefatos em nossa estratégia favorecendo a identificação de novos antígenos, além do processo de inspeção manual e validação experimental que abrangeu um grande número de tipos de linhagens celulares e tumores quando comparados com outros trabalhos de identificação de CTs. Vale destacar que este é o único trabalho de identificação em larga escala que utiliza um número tão significativo de amostras tumorais e que, além disso, utiliza a metodologia de *Southern-blot* para confirmação dos dados de expressão, aumentando a sensibilidade e especificidade da análise.

É importante ressaltar que todos os candidatos validados foram selecionados a partir da lista 1 que corresponde a *clusters* compostos por ESTs derivadas somente de bibliotecas de testículo. Interessante notar que apesar do grande número de ESTs derivadas de bibliotecas tumorais disponíveis em bancos de dados públicos, nenhum candidato validado foi identificado a partir das listas 2 (*clusters* compostos por ESTs derivadas de bibliotecas de testículo e tecido tumoral) e 3 (*clusters* compostos por ESTs derivadas somente de bibliotecas de tecido tumoral). Desta forma, concluímos que a seleção de *clusters* compostos somente por ESTs de testículo foi responsável pela alta eficiência de validação que encontramos e pode ser considerada mais um diferencial de nosso trabalho. Enquanto SCANLAN et al. (2002b) selecionaram apenas transcritos cujas seqüências expressas correspondentes eram derivadas somente de bibliotecas de cDNA de testículo normal e de tecido tumoral, CHEN et al. (2005) ainda consideraram como candidatos transcritos cujas ESTs poderiam corresponder até duas bibliotecas de tecidos normais. Estes dados sugerem que o

nível de expressão desses genes é maior em testículo do que em tumores sustentando a hipótese de que o tumor recapitula o programa de expressão gênica das células germinativas (OLD 2001).

Quanto aos dados de expressão dos cinco candidatos, o candidato CT101 apresentou expressão em 41% (65/160) das amostras tumorais e uma alta frequência (70%) de expressão deste candidato foi observada nas amostras de glioblastoma. O candidato CT704 mostrou-se expresso em 65% (104/160) das amostras tumorais com alta frequência de expressão nas amostras de tumor de pulmão (93%) e colon (83,3%). O candidato CT802 apresentou expressão em 20% (31/155) das amostras tumorais e maior expressão nas amostras de tumor de útero (50%). O candidato CT809 mostrou uma expressão em 24% (38/160) das amostras tumorais e alta frequência de expressão nas amostras de tumor de pulmão (50%). E o CT1001 apresentou-se expresso em 18% (28/156) das amostras tumorais com maior expressão nos tumores de estômago (33,3%).

Vale ressaltar que as frequências de expressão dos candidatos CT802, CT809 e CT1001 enquadram-se nos resultados encontrados na literatura para outros antígenos da classe dos antígenos CT, cuja frequência de positividade varia de 10 a 40%, enquanto para os candidatos CT101 e CT704 a frequência de expressão pode ser considerada alta comparando-se com os mesmos resultados (DE PLAEN et al. 1994; SCANLAN et al. 2002a; ZENDMAN et al. 2002; PARMIGIANI et al. 2006). A alta frequência de expressão dos transcritos CT101 e CT704 os torna candidatos promissores para a imunoterapia do câncer.

Em um estudo desenvolvido por SCANLAN et al. (2004) foi feita a análise por RT-PCR de 43 antígenos CT conhecidos com o objetivo de padronizar a análise

do padrão de expressão dos antígenos *CT* em tecidos normais. Considerando-se apenas a expressão em tecidos normais, os antígenos puderam ser divididos em 4 grupos: 1) transcritos restritos a testículo; 2) transcritos expressos em 2 ou menos tecidos não gametogênicos (testículo, ovário e placenta); 3) transcritos expressos entre 3 e 6 tecidos não gametogênicos; e 4) transcritos com expressão ubíqua. Dos 43 antígenos estudados, apenas 19 apresentaram expressão restrita a testículo, correspondente ao padrão esperado para os antígenos *CT*. Os outros 24 antígenos apresentaram expressão em outros tecidos normais, sendo 5 deles expressos na maioria dos tecidos normais testados. Se a avaliação do padrão de expressão dos cinco antígenos identificados nesse trabalho fosse feita nas mesmas condições deste trabalho, certamente todos pertenceriam ao grupo dos antígenos restritos a testículo.

Além disso, SCANLAN et al. (2004) analisaram a expressão dos mesmos 43 antígenos *CT* em amostras de tumores segundo dados da literatura e revelaram que o perfil de expressão destes antígenos varia bastante conforme o tipo de tumor estudado. Embora os autores ressaltem que as condições de amplificação utilizadas nos trabalhos foram muito diferentes, algumas comparações generalizadas puderam ser feitas. Considerando-se o número de antígenos expressos e a sua frequência de expressão, os tumores puderam ser classificados em três grupos:

1. *High CT expressors* – tumores que expressam mais de 50% dos antígenos *CT* avaliados com uma frequência maior que 20%. Por exemplo, melanoma, câncer de pulmão e câncer de bexiga.
2. *Moderate CT expressors* – tumores que expressam entre 30 e 50% dos antígenos analisados com uma frequência maior que 20%. Por exemplo, câncer de mama e câncer de próstata.

3. *Low CT expressors* – tumores que expressam menos de 30% dos antígenos com frequência maior que 20%. Por exemplo, câncer renal e câncer de colon.

Desta maneira, considerando os tipos tumorais em que foi feita a avaliação da expressão dos nossos cinco candidatos, verifica-se que o perfil de expressão destes transcritos em melanoma e câncer de pulmão é condizente com os outros membros da classe dos antígenos CT. No entanto, nos chamou atenção a alta frequência de expressão destes candidatos em tumores de mama e próstata, considerados *moderate CT expressors*, e, especialmente, em tumores de colon, considerado um *low CT expressor*. Estes resultados novamente demonstram a alta frequência de expressão em amostras tumorais dos candidatos identificados quando comparados com dados da literatura. E, além disso, indicam que os cinco candidatos identificados podem representar alvos terapêuticos importantes para tipos de tumores ainda não cobertos por outros antígenos, como mama, próstata e colon, e, até então, sem possibilidade de imunoterapia.

Vale também destacar que três dos cinco transcritos identificados localizam-se no cromossomo X onde se encontram mapeados inúmeros antígenos tumorais da classe dos CT como MAGE-A, NY-ESO-1 e LAGE-14 (SCANLAN et al. 2002a). Além disso, é importante citar que, durante o desenvolvimento deste trabalho, LIGGINS et al. (2004a), utilizando a técnica de SEREX, identificaram 28 antígenos que são reconhecidos pela resposta imune humoral de pacientes com linfoma difuso de células B (DLBCL) incluindo o antígeno OX-TES-1. Este antígeno corresponde ao nosso candidato CT101, até então não descrito como um antígeno tumoral. Ao mesmo tempo, LIGGINS et al. (2004b) publicaram um estudo detalhado deste

transcrito avaliando a expressão do mesmo em linhagens celulares derivadas de DLBCL e *arrays* de tecidos normais e tumorais obtidos comercialmente, além de identificar a forma de *splicing* alternativo. É importante destacar que a identificação prévia de um dos candidatos como um antígeno tumoral vem novamente demonstrar a eficiência da técnica utilizada neste trabalho na identificação de novos antígenos tumorais.

Para esses candidatos serem considerados novos antígenos tumorais, além da confirmação da expressão dos mesmos em tumores primários, deve ser demonstrada a presença de anticorpos contra as proteínas codificadas por esses candidatos em soros de pacientes com câncer. Desta forma, avaliamos a resposta humoral contra as proteínas correspondentes aos candidatos CT101, CT802 e CT809 por apresentarem uma seqüência completa de mRNA depositada em banco de dados públicos. As proteínas recombinantes foram utilizadas para detecção de anticorpos em plasma de pacientes com diferentes tipos de câncer através de ensaios de *Immunoblotting*.

Com relação aos candidatos CT704 e CT1001, como correspondiam apenas a seqüências expressas, nosso objetivo foi gerar a seqüência completa destes transcritos. Assim, no futuro os mesmos poderão ser avaliados quanto à presença de anticorpos em plasmas de pacientes com diferentes tipos de tumores completando a caracterização destes transcritos como novos antígenos tumorais.

A freqüência de anticorpos anti-CT101 encontrada no plasma de pacientes com câncer de útero foi de 41% (9/22). No entanto, sua freqüência dentre os indivíduos sadios também foi significativa, sendo detectada a presença de anticorpos anti-CT101 em 43,3% (13/30) das amostras analisadas. Casos como este já foram descritos na literatura para outros antígenos *CT*, como o MAGE-A1 e outros

membros da família MAGE, mas em uma frequência menor e acredita-se que seja resultado de reação cruzada em indivíduos que têm Vitiligo e Lupus Eritematoso ou ainda resposta imune efetiva em indivíduos que se submeteram à vasectomia (HOON et al. 1995; LEA et al. 1997; MCCURDY et al. 1998; ROCHA et al. 2000). Além disso, não podemos descartar a possibilidade de termos identificado uma resposta imune a um tumor ainda não diagnosticado. Este seria um exemplo no qual a presença de anticorpos específicos poderia funcionar como um marcador para o diagnóstico do câncer.

Apesar de todas as explicações encontradas na literatura, a frequência de anticorpos anti-CT101 encontrada nas amostras normais foi muito elevada. Além disso, em um trabalho publicado por GUINN et al. (2005), este antígeno, através da técnica de SEREX, não foi reativo em nenhum dos 18 soros normais analisados, mas demonstrou reatividade em 35% (6/17) dos soros de pacientes com leucemia mielóide aguda, 6% (1/16) dos soros de pacientes com leucemia mielóide crônica e 10%(1/10) dos soros de pacientes com linfoma difuso de células B. Vale lembrar que esta reatividade foi encontrada contra a isoforma PASD1\_v2, a mesma analisada em nosso trabalho, demonstrando que a alta positividade que encontramos em soro normal pode ser decorrente de um artefato do sistema utilizado.

Neste mesmo estudo foi realizado um ensaio de apresentação de antígeno utilizando células dendríticas derivadas de monócitos que demonstrou que o antígeno CT101 é capaz de estimular a resposta de células T autólogas. Além disso, vale citar que este antígeno foi descrito recentemente como um candidato para imunoterapia de linfomas uma vez que a expressão da proteína foi detectada em linhagens celulares de linfoma (COOPER et al. 2006).

Todos estes dados sugerem que o CT101 pode se tornar um alvo promissor em futuros ensaios clínicos para a imunoterapia do câncer. Desta forma, para verificar se a alta positividade que observamos entre os plasmas de indivíduos normais poderia estar associada ao sistema de expressão utilizado, avaliamos a presença de anticorpos nos plasmas normais contra duas proteínas de cana-de-açúcar, CaneCPI-2 e CaneCPI-3 (GIANOTTI et al. 2006). Por serem inibidores de cisteíno proteinases de planta, conhecidas como fitocistatinas, não esperávamos encontrar anticorpos contra essas proteínas nos plasmas humanos. No entanto, para cada uma das proteínas avaliadas encontramos dois plasmas com reatividade correspondendo a uma frequência de 6,7% (2/30) o que demonstra a existência de uma reação inespecífica. Isto nos mostra que parte da frequência de anticorpos anti-CT101 encontrada nos plasmas normais pode ser devido ao sistema de expressão utilizado. Além disso, com relação aos outros dois candidatos, CT802 e CT809, a frequência de anticorpos nos plasmas de indivíduos normais também foi de 6,7% (2/30) confirmando a existência de uma reação inespecífica do sistema de expressão utilizado.

A frequência de anticorpos anti-CT802 encontrada nos plasmas de pacientes com câncer de útero foi de 22,7% (5/22) e a de anticorpos anti-CT809 encontrada no plasma de pacientes com câncer de pulmão foi de 7,7% (1/13). É importante destacar que a frequência de anticorpos anti-CT802 encontrada no plasma dos pacientes é extremamente significativa quando comparada aos resultados de imunogenicidade de outros antígenos *CT* (SCANLAN et al. 2002a). Até o momento, o CTSP-1 é o antígeno mais imunogênico da categoria dos antígenos *CT* tendo uma frequência de anticorpos específicos de 20% em pacientes com diferentes tipos de tumor com

destaque para os portadores de câncer de próstata que apresentaram a maior frequência de anticorpos (40%) (PARMIGIANI et al. 2006). Já para os outros antígenos *CT* a presença de anticorpos em pacientes com câncer é pouco frequente, nunca ultrapassando 10% dos casos, onde podemos incluir o CT809.

A maioria dos trabalhos que avaliam a presença de anticorpos contra antígenos tumorais da categoria dos antígenos *CT* utiliza a técnica de ELISA (JAGER et al. 1998, 1999; DONG et al. 2003; NAKADA et al. 2003; FOSSA et al. 2004). Nestes trabalhos são analisados soros de indivíduos sadios que servem de referência para determinar o *cut-off* de positividade das amostras dos pacientes com câncer. No entanto, com base em um trabalho anterior realizado em nosso laboratório foi demonstrado que esta técnica não apresenta bons resultados quando utilizamos o plasma como amostra ao invés de soro, uma vez que o plasma produz um alto *background* dificultando a análise dos resultados. Isto se deve à presença de uma grande quantidade de proteínas séricas no plasma, que interferem na especificidade do experimento. Desta forma, aplicamos a técnica de *Immunoblotting* que, embora mais trabalhosa, é bem mais específica do que o ELISA uma vez que a positividade se dá pelo reconhecimento da proteína recombinante no tamanho esperado após seu fracionamento em gel.

Apesar de apresentar uma frequência de anticorpos semelhante à encontrada para antígenos *CT* já descritos, para que o CT809 possa no futuro ser utilizado para aplicações na imunoterapia do câncer, novas análises abrangendo um número maior de amostras de diferentes tipos de tumor devem ser realizadas. No entanto, com relação ao CT802, a alta imunogenicidade encontrada dentre os plasmas de pacientes o torna um ótimo candidato para imunoterapia no tratamento do câncer. Para que

ensaios clínicos futuros sejam viabilizados outros estudos *in vitro* serão necessários para uma melhor compreensão da resposta imune espontânea anti-CT802 em pacientes com câncer.

Um importante passo neste sentido é a verificação da presença de células T anti-CT802 nestes pacientes. Para a maioria dos antígenos *CT* a presença de resposta imune celular está associada à presença da resposta humoral, o que pode sugerir a existência de tal resposta contra o CT802. Além da identificação de linfócitos T CD8, alguns trabalhos também têm investigado a presença de células T CD4 *helper* específicas contra estes antígenos, visto que a presença das mesmas é imprescindível para a manutenção da resposta citotóxica e da produção de anticorpos (CHEN et al. 2004; DAVIS et al. 2004).

Outro importante passo é a identificação de epítomos antigênicos do CT802 que são reconhecidos pelo sistema imune dos pacientes com câncer. Vários peptídeos antigênicos de diferentes antígenos *CT* reconhecidos por células T citotóxicas já foram identificados e alguns deles estão sendo testados como candidatos a imunoterapia em pacientes com câncer (GNJATIC et al. 2000, 2002; CHEN et al. 2004; DAVIS et al. 2004). Alguns ensaios clínicos em pacientes com melanoma têm sido feitos utilizando-se um peptídeo derivado do MAGE-3 e evidências de regressão do tumor têm sido verificadas em 20% dos casos (MARCHAND et al. 1999). Já com relação ao NY-ESO-1, pacientes com tumores que expressavam a proteína receberam uma vacina utilizando a proteína recombinante tendo sido detectadas células T CD4 e T CD8 circulantes específicas para uma grande variedade de epítomos do NY-ESO-1 (DAVIS et al. 2004).

Como podemos observar, este trabalho gera perspectivas para muitos outros, nos quais a função, a imunogenicidade, e o real potencial terapêutico destes candidatos podem ser avaliados.

## ***CONCLUSÕES***



## 6 CONCLUSÕES

1. A análise de antígenos CT conhecidos no Banco de Dados Map4 demonstrou que apenas um pequeno grupo de CTs apresentava o padrão de expressão esperado para esta classe de antígenos permitindo a identificação de artefatos e limitações da abordagem utilizada.
2. A análise computacional realizada neste trabalho permitiu a identificação de 1.255 genes candidatos com padrão de distribuição de ESTs característico de antígenos CT.
3. Após análise de expressão por RT-PCR em tecidos normais e tumorais, o padrão de expressão de cinco candidatos a antígenos CT (CT101, CT704, CT802, CT809 e CT1001) foi confirmado, validando a estratégia computacional utilizada.
4. Os cinco candidatos a antígenos CT apresentaram elevada frequência de expressão em diferentes tipos de tumores, variando entre 18% e 65%, o que os tornam alvos promissores para o desenvolvimento de vacinas.
5. Anticorpos contra 3 candidatos selecionados foram detectados em plasma de pacientes com câncer em frequências bastante elevadas (7-41%) validando-os como novos antígenos tumorais.

6. Anticorpos contra os 3 candidatos avaliados também foram encontrados em plasma de indivíduos sadios e a análise desses plasmas utilizando proteínas de cana de açúcar sugere que esses resultados possam estar associados ao sistema de expressão utilizado.
  
7. Utilizando a técnica de RACE foi possível gerar a seqüência de mRNA completa para os candidatos CT704 e CT1001 assim como identificar formas alternativas de *splicing* para o candidato CT704.

## ***REFERÊNCIAS BIBLIOGRÁFICAS***



## 7 REFERÊNCIAS BIBLIOGRÁFICAS

Abbas AK, Lichtman AH, Pober JS. **Celular and molecular immunology**. 4<sup>th</sup> ed. Philadelphia: W.B. Saunders; 2000. Immunity to tumors; p.384-403.

Abelev GI, Perva NI, Khramkova ZA. Production of embryonal alpha-globulin by transplantable mouse hepatomas. **Transplantation** 1963; 1:174-80.

Adams MD, Kerlavage AR, Fleischmann RD, et al. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. **Nature** 1995; Supl 377:3-174.

Aradhya S, Bardaro T, Galgoczy P, et al. Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the NEMO and LAGE2 genes. **Hum Mol Genet** 2001; 10:2557-67.

Baba T, Koizumi M, Suzuki T, Yamanaka I, Yamashita S, Kudo R. An acrosomal protein, sp 32, in mammalian sperm is a binding protein specific for two proacrosins and an acrosin intermediate. **J Biol Chem** 1994; 269:10133-40.

Barrow C, Browning J, MacGregor D, et al. Tumor antigen expression in melanoma varies according to antigen and stage. **Clin Cancer Res** 2006; 12:764-71.

Brass N, Heckel D, Sahin U, Pfreundschuh M, Sybrecht GW, Meese E. Translation initiation factor eIF-4gamma is encoded by an amplified gene and induces an immune response in squamous cell lung carcinoma. **Hum Mol Genet** 1997; 6:33-9.

Brasseur F, Rimoldi D, Lienard D, et al. Expression of mage genes in primary and metastatic cutaneous melanoma. **Int J Cancer** 1995; 63:375-80.

Brichard V, Vanpel A, Wolfel T, et al. The tyrosinase gene codes for an antigen recognized by autologous cytolytic T-lymphocytes on Hla-A2 melanomas. **J Exp Med** 1993; 178:489-95.

Camargo AA, Samaia HP, Dias-Neto E, et al. The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. **Proc Natl Acad Sci U.S.A.** 2001; 98:12103-8.

Carey TE, Takahashi T, Resnick LA, Oettgen HF, Old LJ. Cell surface antigens of human malignant melanoma: mixed hemadsorption assays for humoral immunity to cultured autologous melanoma cells. **Proc Natl Acad Sci U.S.A.** 1976; 73:3278-82.

Cheever MA, Disis ML, Bernhard H, et al. Immunity to oncogenic proteins. **Immunol Rev** 1995; 145:33-59.

Chen YT, Scanlan MJ, Sahin U, et al. A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. **Proc Natl Acad Sci U.S.A.** 1997; 94:1914-8.

Chen YT, Gure AO, Tsang S, et al. Identification of multiple cancer/testis antigens by allogeneic antibody screening of a melanoma cell line library. **Proc Natl Acad Sci U.S.A** 1998; 95:6919-23.

Chen QY, Jackson H, Parente P, et al. Immunodominant CD4(+) responses identified in a patient vaccinated with full-length NY-ESO-1 formulated with ISCOMATRIX adjuvant. **Proc Natl Acad Sci U.S.A** 2004; 101:9363-8.

Chen YT, Scanlan MJ, Venditti CA, et al. Identification of cancer/testis-antigen genes by massively parallel signature sequencing. **Proc Natl Acad Sci U.S.A.** 2005; 102:7940-5.

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. **Biochemistry** 1979; 18:5294-9.

Cooper CDO, Liggins AP, Ait-Tahar K, Roncador G, Banham AH, Pulford K. PASD1, a DLBCL-associated cancer testis antigen and candidate for lymphoma immunotherapy. **Leukemia** 2006; 20:2172-4.

Coulie PG, Lehmann F, Lethé B, et al. A mutated intron sequence codes for an antigenic peptide recognized by cytolytic T-lymphocytes on a human-melanoma. **Proc Natl Acad Sci U.S.A** 1995; 92:7976-80.

Crew AJ, Clark J, Fisher C, et al. Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with homology to the Kruppel-associated box in human synovial sarcoma. **EMBO J** 1995; 14:2333-40.

Dalmau J, Gultekin SH, Voltz R, et al. Ma1, a novel neuron- and testis-specific protein, is recognized by the serum of patients with paraneoplastic neurological disorders. **Brain** 1999; 122:27-39.

Davis ID, Chen W, Jackson H, et al. Recombinant NY-ESO-1 protein with ISCOMATRIX adjuvant induces broad integrated antibody and CD4+ and CD8+ T cell responses in humans. **Proc Natl Acad Sci U.S.A.** 2004; 101:10697-702.

Day EA. **The immunochemistry of cancer.** Springfield: C. C. Thomas; 1965.

de Plaen E, Arden K, Traversari C, et al. Structure, chromosomal localization, and expression of 12 genes of the MAGE family. **Immunogenetics** 1994; 40:360-9.

de Wit NJ, Weidle UH, Ruiter DJ, van Muijen GN. Expression profiling of MMA-1a and splice variant MMA-1b: new cancer/testis antigens identified in human melanoma. **Int J Cancer** 2002; 98:547-53.

Dong XY, Su YR, Qian XP, et al. Identification of two novel CT antigens and their capacity to elicit antibody response in hepatocellular carcinoma patients. **Br J Cancer** 2003; 89:291-7.

Dong XY, Yang XA, Wang YD, Chen WF. Zinc-finger protein ZNF165 is a novel cancer-testis antigen capable of eliciting antibody response in hepatocellular carcinoma patients. **Br J Cancer** 2004; 91:1566-70.

Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. **Nat Immunol** 2002; 3:991-8.

Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. **Annu Rev Immunol** 2004; 22:329-60.

Eckman BA, Aaronson JS, Borkowski JA, et al. The Merck Gene Index browser: an extensible data integration system for gene finding, gene characterization and EST data mining. **Bioinformatics** 1998; 14:2-13.

Ferrante K, Winograd B, Canetta R. Promising new developments in cancer chemotherapy. **Cancer Chemother Pharmacol** 1999; 43 Suppl:S61-S68.

Fisk B, Blevins TL, Wharton JT, Ioannides CG. Identification of an immunodominant peptide of her-2/neu protooncogene recognized by ovarian tumor-specific cytotoxic T-lymphocyte lines. **J Exp Med** 1995; 181:2109-17.

Fossa A, Berner A, Fossa SD, Hernes E, Gaudernack G, Smeland EB. NY-ESO-1 protein expression and humoral immune responses in prostate cancer. **Prostate** 2004; 59:440-7.

Galante PA, Sabake NJ, Kirschbaum-Slager N, Souza SJ. Detection and evaluation of intron retention events in the human transcriptome. **RNA** 2004; 10:757-65.

Gianotti A, Rios WM, Soares-Costa A, et al. Recombinant expression, purification, and functional analysis of two novel cystatins from sugarcane (*Saccharum officinarum*). **Protein Expr Purif** 2006; 47:483-9.

Gnjatic S, Nagata Y, Jager E, et al. Strategy for monitoring T cell responses to NY-ESO-1 in patients with any HLA class I allele. **Proc Natl Acad Sci U.S.A** 2000; 97:10917-22.

Gnjatic S, Jager E, Chen W, et al. CD8(+) T cell responses against a dominant cryptic HLA-A2 epitope after NY-ESO-1 peptide immunization of cancer patients. **Proc Natl Acad Sci U.S.A** 2002; 99:11813-8.

Gold P, Freedman SO. Specific carcinoembryonic antigens of the human digestive system. **J Exp Med** 1965; 122:467-81.

Goydos JS, Patel M, Shih WC. NY-ESO-1 and CTp11 expression may correlate with stage of progression in melanoma. **J Surg Res** 2001; 98:76-80.

Grunwald C, Koslowski M, Arsiray T, et al. Expression of multiple epigenetically regulated cancer/germline genes in nonsmall cell lung cancer. **Int J Cancer** 2006; 118:2522-8.

Guinn B-A, Bland EA, Lodi U, et al. Humoral detection of leukaemia-associated antigens in presentation acute myeloid leukaemia. **Biochem Biophys Res Commun** 2005; 335:1293-304.

Gure AO, Tureci O, Sahin U, et al. SSX: a multigene family with several members transcribed in normal testis and human cancer. **Int J Cancer** 1997; 72:965-71.

Hillier LD, Lennon G, Becker M, et al. Generation and analysis of 280,000 human expressed sequence tags. **Genome Res** 1996; 6:807-828.

Hoon DSB, Yuzuki D, Hayashida M, Morton DL. Melanoma patients immunized with melanoma cell vaccine induce antibody-responses to recombinant mage-1 antigen. **J Immunol** 1995; 154:730-7.

Huang X, Madan A. CAP3: A DNA sequence assembly program. **Genome Res** 1999; 9:868-77.

Jager E, Chen YT, Drijfhout JW, et al. Simultaneous humoral and cellular immune response against cancer-testis antigen NY-ESO-1: definition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes. **J Exp Med** 1998; 187:265-70.

Jager E, Stockert E, Zidianakis Z, et al. Humoral immune responses of cancer patients against 'Cancer-Testis' antigen NY-ESO-1: correlation with clinical events. **Eur J Cancer** 1999; 35:S353-S4

Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. **CA Cancer J Clin** 2007; 57:43-66.

Jungbluth AA, Busam KJ, Kolb D, et al. Expression of MAGE-antigens in normal tissues and cancer. **Int J Cancer** 2000; 85:460-5.

Jungbluth AA, Chen YT, Stockert E, et al. Immunohistochemical analysis of NY-ESO-1 antigen expression in normal and malignant human tissues. **Int J Cancer** 2001; 92:856-60.

Kawakami Y, Eliyahu S, Delgado CH, et al. Cloning of the gene coding for a shared human-melanoma antigen recognized by autologous t-cells infiltrating into tumor. **Proc Natl Acad Sci U.S.A.** 1994; 91:3515-9.

Kelso J, Visagie J, Theiler G, et al. eVOC: A controlled vocabulary for unifying gene expression data. **Genome Res** 2003; 13:1222-30.

Korenberg JR, Chen XN, Adams MD, Venter JC. Toward a cDNA map of the human genome. **Genomics** 1995; 29:364-70.

Lea IA, Adoyo P, O'Rand MG. Autoimmunogenicity of the human sperm protein Sp17 in vasectomized men and identification of linear B cell epitopes. **Fertil Steril** 1997; 67:355-61.

Lennette ET, Winberg G, Yadav M, Enblad G, Klein G. Antibodies to Lmp2A/2B in Ebv-Carrying Malignancies. **Eur J Cancer** 1995; 31A:1875-8.

Lethe B, Lucas S, Michaux L, et al. LAGE-1, a new gene with tumor specificity. **Int J Cancer** 1998; 76:903-8.

Liggins AP, Guinn BA, Hatton CS, Pulford K, Banham AH. Serologic detection of diffuse large b-cell lymphoma-associated antigens. **Int J Cancer** 2004a; 110:563-9.

Liggins AP, Brown PJ, Asker K, Pulford K., Banham AH. A novel diffuse large B-cell lymphoma-associated cancer-testis antigen encoding a PAS domain protein. **Br J Cancer** 2004b; 91:141-9.

Lucas S, Brasseur F, Boon T. A new MAGE gene with ubiquitous expression does not code for known MAGE antigens recognized by T cells. **Cancer Res** 1999; 59:4100-3.

Marchand M, van Baren N, Weynants P, et al. Tumor regressions observed in patients with metastatic melanoma treated with an antigenic peptide encoded by gene MAGE-3 and presented by HLA- A1. **Int J Cancer** 1999; 80:219-30.

McCurdy DK, Tai LQ, Nguyen J, et al. MAGE Xp-2: A member of the MAGE gene family isolated from an expression library using systemic lupus erythematosus sera. **Mol Genet Metabolism** 1998; 63:3-13.

Meuwissen RL, Offenberg HH, Dietrich AJ, Riesewijk A, van Iersel M, Heyting C. A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes. **EMBO J** 1992; 11:5091-100.

Ministerio da Saúde. Instituto Nacional de Câncer. **Estimativas/2006 incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2005.

Monte M, Simonatto M, Peche LY, et al. MAGE-A tumor antigens target p53 transactivation function through histone deacetylase recruitment and confer resistance to chemotherapeutic agents. **Proc Natl Acad Sci U.S.A.** 2006; 103:11160-5.

Nakada T, Noguchi Y, Satoh S. NY-ESO-1 mRNA expression and immunogenicity in advanced prostate cancer. **Cancer Immunol** 2003; 3:10.

Old LJ, Chen YT. New paths in human cancer serology. **J Exp Med** 1998; 187:1163-7.

Old LJ. Cancer/testis (CT) antigens – a new link between gametogenesis and cancer. **Cancer Immun** 2001; 1:1.

Parmigiani RB, Bettoni F, Vibranovski MD, et al. Characterization of a cancer/testis (CT) antigen gene family capable of eliciting humoral response in cancer patients. **Proc Natl Acad Sci U.S.A.** 2006; 103:18066-71.

Peng JR, Chen HS, Mou DC, et al. Expression of cancer/testis (CT) antigens in Chinese hepatocellular carcinoma and its correlation with clinical parameters. **Cancer Lett** 2005; 219:223-32.

Pfreundschuh M, Shiku H, Takahashi T, et al. Serological analysis of cell surface antigens of malignant human brain tumors. **Proc Natl Acad Sci U.S.A.** 1978; 75:5122-6.

Ponting CP, Aravind L. PAS: a multifunctional domain family comes to light. **Curr Biology** 2005; 7:R674-R677.

Quackenbush J, Liang F, Holt I, Pertea G, Upton JY. The TIGR gene indices: reconstruction and representation of expressed gene sequences. **Nucleic Acids Res** 2000; 28:141-5.

Reymond A, Camargo AA, Deutsch S, et al. Nineteen additional unpredicted transcripts from human chromosome 21. **Genomics** 2002; 79:824-32.

Rocha IM, Oliveira LJM, de Castro LCM, et al. Recognition of melanoma cell antigens with antibodies present in sera from patients with vitiligo. **Int J Dermatol** 2000; 39:840-3.

Sahin U, Tureci O, Schmitt H, et al. Human neoplasms elicit multiple specific immune responses in the autologous host. **Proc Natl Acad Sci U.S.A** 1995; 92:11810-3.

Sahin U, Tureci O, Pfreundschuh M. Serological identification of human tumor antigens. **Curr Opin Immunol** 1997; 9:709-16.

Scanlan MJ, Chen YT, Williamson B, et al. Characterization of human colon cancer antigens recognized by autologous antibodies. **Int J Cancer** 1998; 76:652-8.

Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, Chen YT. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. **Immunol Rev** 2002a; 188:22-32.

Scanlan MJ, Gordon CM, Williamson B, et al. Identification of Cancer/Testis Genes by Database Mining and mRNA Expression Analysis. **Int J Cancer** 2002b; 98:485-92.

Scanlan MJ, Simpson AJG, Old LJ. The cancer/testis genes: Review, standardization, and commentary. **Cancer Immun** 2004; 4:1-15.

Schuler GD. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. **J Mol Med** 1997; 75:694-8.

Segal NH, Blachere NE, Guevara-Patino JA, et al. Identification of cancer-testis genes expressed by melanoma and soft tissue sarcoma using bioinformatics. **Cancer Immunol** 2005; 5:2.

Shankaran V, Ikeda H, Bruce AT, et al. IFN gamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. **Nature** 2001; 410:1107-11.

Silva AP, Salim AC, Bulgarelli A, et al. Identification of 9 novel transcripts and two RGSL genes within the hereditary prostate cancer region (HPC1) at 1q25. **Gene** 2003; 310:49-57.

Simpson AJG, Caballero OL, Jungbluth A, Chen YT, Old LL. Cancer/testis antigens, gametogenesis and cancer. **Nat Rev Cancer** 2005; 5:615-25.

Strausberg RL, Buetow KH, Emmert-Buck MR, Klausner RD. The cancer genome anatomy project: building an annotated gene index. **Trends Genet** 2000; 16:103-6.

Suri A. Cancer testis antigens – their importance in immunotherapy and in the early detection of cancer. **Expert Opin Biol Ther** 2006; 6:379-89.

Tindle RW. Human papillomavirus vaccines for cervical cancer. **Curr Opin Immunol** 1996; 8:643-50.

Ueda R, Shiku H, Pfreundschuh M, et al. Cell surface antigens of human renal cancer defined by autologous typing. **J Exp Med** 1979; 150:564-79.

van der Bruggen P, Traversari C, Chomez P, et al. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. **Science** 1991; 254:1643-7.

Vidaeus CM, Kapp-Herr C, Golden WL, Eddy RL, Shows TB, Herr JC. Human fertilin beta: identification, characterization, and chromosomal mapping of an ADAM gene family member. **Mol Reprod Dev** 1997; 46:363-9.

Williamson AR. The Merck Gene Index Project. **Drug Discov Today** 1999; 4:115-22.

Wolfel T, Hauer M, Schneider J, et al. A P16(Ink4A)-insensitive cdk4 mutant targeted by cytolytic T-lymphocytes in a human-melanoma. **Science** 1995; 269:1281-4.

Yan W, Rajkovic A, Viveiros MM, Burns KH, Eppig JJ, Matzuk MM. Identification of *Gasz*, an evolutionary conserved gene expressed exclusively in germ cells and encoding a protein with four ankyrin repeats, a sterile- $\alpha$  motif, and a basic leucine zipper. **Mol Endocrinol** 2002; 16:1168-84.

Zendman AJ, Cornelissen IM, Weidle UH, Ruiter DJ, van Muijen GN. Ctp11, a novel member of the family of human cancer/testis antigens. **Cancer Res** 1999; 59:6223-9.

Zendman AJ, van Kraats AA, Weidle UH, Ruiter DJ, van Muijen GN. The XAGE family of cancer/testis-associated genes: alignment and expression profile in normal tissues, melanoma lesions and Ewing's sarcoma. **Int J Cancer** 2002; 99:361-9.

## ***ANEXOS***



***ANEXO 1***



Número de acesso da sequência referente a cada candidato selecionado e a sequência dos iniciadores referentes a cada um com o respectivo tamanho do produto amplificado.

Candidato	EST	Seqüência dos iniciadores	Tamanho do produto amplificado
CT101	BI458651	CT101F 5'TGCTGCATGTGTTCCCTCAG3' CT101R 5'AGCTTCATCACTGACTGCC3'	112pb
CT102	BF059393	CT102F 5'TGCCATATTCAGTCTGCTGCC3' CT102R 5'TATGAACGCTGCGAGCTG3'	157pb
CT103	BG773166	CT103F 5'TTGCTCTAAGATGGCGGAG3' CT103R 5'CAGAATCACTTGAGGTGGGAG3'	381pb
CT104	AA382712	CT104F 5'TGCTAGCACGTTTTCACCTC3' CT104R 5'CTTCCTACAGTAGTCGATA3'	200pb
CT105	AI825179	CT105F 5'GTCCACAGGCAACCAACC3' CT105R 5'TGCCGCAAAGAAAAATGAG3'	114pb
CT106	AI952718	CT106F 5'TGTATCCCTCCTGGCGTC3' CT106R 5'TAACAGATCCACCCACT3'	207pb
CT107	AI968049	CT107F 5'AGCCAAATGGGAACCCTG3' CT107R 5'GATTCTAAGGCGGCCAG3'	147pb
CT108	AI371849	CT108F 5'TGAGTGACAATGTGGAGCC3' CT108R 5'TTCCAGACCAACTGAGC3'	206pb
CT109	AI656983	CT109F 5'GCGGGATGGACACACAG3' CT109R 5'GATGGGGTGGTGTGAGG3'	252pb
CT110	AI352624	CT110F 5'CTGCCACAGAATTTGCTACG3' CT110R 5'AAAGCCCTTCGAGATGCC3'	197pb
CT111	BG716916	CT111F 5'GCGTTCACATCCTGTTTGG3' CT111R 5'TGTTGGTAAACCGCGCAC3'	205pb
CT112	BF979763	CT112F 5'TGTCTGAGCGATGATGCG3' CT112R 5'ATGCCCTTCTGGACTIONG3'	175pb
CT113	BG720627	CT113F 5'ATCTTCCACCGAATGCC3' CT113R 5'AGTTTGGAGGCTCCGGG3'	170pb
CT114	AI968777	CT114F 5'TCGCTCACTGGCTCCTC3' CT114R 5'TGAGTCCAAAGCCGAGAG3'	201pb
CT115	BG720643	CT115F 5'TGTCTTCTGGGACCACCG3' CT115R 5'ACGTGCCAATCCATCAGC3'	338pb

**Cont/ Anexo 1**

CT201	BE858821	CT201F 5'ACCGTGGACTTCTGGAGG3' CT201R 5'ACGGACAGAGGCTGCTC3'	283pb
CT202	AI360454	CT202F 5'GCCAAAAGGAACACAGAGG3' CT202R 5'TCACCCAGTGAAGGAGCC3'	281pb
CT203	AI653731	CT203F 5'AGCCTTGGCACCAGAAAC3' CT203R 5'AGCACAGAAGGGCAGTGG3'	218pb
CT204	BG772655	CT204F 5'AGACCCCCAAGGACAAGG3' CT204R 5'ATCTCAGGGTTCCGCTCC3'	589pb
CT205	AW613979	CT205F 5'AAGCCATGGTGTGCCTTC3' CT205R 5'ATCATGGTGTGGAGGGG3'	182pb
CT206	AI366092	CT206F 5'GTTTCGTGACCGGGAGATG3' CT206R 5'TGCTTGCCCTCTGGATTGG3'	64pb
CT207	BF887749	CT207F 5'TTGCTCAAGTTACTCTCCACG3' CT207R 5'GGATATTGTGTTGCCCTC3'	142pb
CT208	BF885090	CT208F 5'GATGAATGGTACACAACCTCTG3' CT208R 5'TTAGAAGATGATGATGATG3'	144pb
CT209	BF979574	CT209F 5'CCAGGCTGGGATTTGATG3' CT209R 5'ATGCAGCTGGGTCTTTCC3'	167pb
CT210	BM556856	CT210F 5'GGAAAAACAGACACCTGAAAG3' CT210R 5'TCCTCCCATCTCAGCCTC3'	130pb
CT211	AI655941	CT211F 5'TCTATGAGCCAGCCAG3' CT211R 5'GTACGACCTCGGTGCGG3'	212pb
CT212	BG910378	CT212F 5'CAGTTTGTCTCGACCCTCG3' CT212R 5'TTTGGCCAGGGTCAAGG3'	189pb
CT213	AI768847	CT213F 5'GCTCTGAATCTTCCTGGGG3' CT213R 5'ACCGGTCAGACTCCAGG3'	272pb
CT214	BE843995	CT214F 5'GATGGGGGCTCAGCAA3' CT214R 5'TGGAGCCACTGAGGAAGC3'	142pb
CT215	BQ672145	CT215F 5'AAGATGGAGGTGGCAAAGC3' CT215R 5'GATGACCACATGGGCAGAC3'	168pb
CT216	BI820625	CT216F 5'TGTCCTCAGCTGGGTGG3' CT216R 5'TGGGAAGGGTCTGGGAAC3'	283pb
CT217	AL079669	CT217F 5'CAGCTCTGAAGGCATGTTTAG3' CT217R 5'AGAAATTCCAGGAGCGGC3'	143pb
CT218	BG333817	CT218F 5'TCAATGGAAGACCCCG3' CT218R 5'TGATGCAGGTGTGATGGTC3'	215pb

**Cont/Anexo 1**

CT301	BE735233	CT301F 5'TACCGCCAGCTCTCCAC3' CT301R 5'GCCGGGTAAATGGAGAAC3'	275pb
CT302	BM927640	CT302F 5'AGGCGGTGTCTCTGTTCG3' CT302R 5'TCTGAAGGCGGGGAGAG3'	298pb
CT303	AI283993	CT303F 5' TACTGCGAGCAAATGGGG3' CT303R 5' TCATCCTAAAGCATAACGGGTC3'	320pb
CT304	BE727460	CT304F 5' CCGAGAGCTGTCCATTC3' CT304R 5' TTTCTCCGAAGCAGGTG 3'	134pb
CT305	BG470816	CT305F 5' AAGTGCTGCCCAAGTTGC 3' CT305R 5' CTGCCTTTCGTGGAGCTG 3'	111pb
CT306	BG767408	CT306F 5'TACATGTTTCGATGGGCCG3' CT306R 5'GTACACCATGCCACAGCG3'	189pb
CT307	BF683426	CT307F 5'CGAAAATGGCACAGAAATG3' CT307R 5'TTGGGCTTAACCTCCTTGG3'	167pb
CT308	BF374746	CT308F 5'CTGCTGCAACATCTTTTCAAC3' CT308R 5'CGAAATGTACACCATAGTCCC3'	184pb
CT309	BM550300	CT309F 5'ATCCACCCACAGAGCAC3' CT309R 5'ACCTGCAGCCACTGTGAG3'	93pb
CT310	BE409638	CT310F 5'TTCTGCTCTTTGCCACCTG3' CT310R 5'GGATTTTGACCAACTGTTTC3'	150pb
CT311	BE144331	CT311F 5'AAGGGCAGGACAGGAAGG3' CT311R 5'GATGAGGCTCTCTGCTCCC3'	298pb
CT312	BG479526	CT312F 5'TCATCTCTGCCCCCTCTG3' CT312R 5'GGATGACCTTGCCACAG3'	306pb
CT401	BG335088	CT401F 5'ACTCTGCCGTTCCGAGG3' CT401R 5'AAGACCCCGCTTTGAAGG3'	294pb
CT402	BE729130	CT402F 5'TTGGTCAGGAAGGTTCCGG3' CT402R 5'TGTCTGTCCAGCATCTGCC3'	309pb
CT403	BG478724	CT403F 5'TGTGAGGCACCCAGTGAC3' CT403R 5'GTCCACCCAAAGAAGGC3'	105pb
CT405	BI856424	CT405F 5'TGCTGATAGTCGGTGCCC3' CT405R 5'ACCAGGCCTTCCCACATC3'	187pb
CT406	BM423379	CT406F 5'ACGACCGCCGCTACTTC3' CT406R 5'AGTTCGGCGGTCCTAGG3'	271pb
CT407	BM471826	CT407F 5'ATGAGGAAGCCACGCATC3' CT407R 5'ACAAGCCTGTGGAATGTCCC3'	445pb

**Cont/Anexo 1**

CT408	BF914944	CT408F 5'GGAGGGTGCTTTGGATTTTC3' CT408R 5'ACCGACATGATCAGAAAGG3'	174pb
CT409	AI814314	CT409F 5'AAGACAAAGACCAAGCTGCC3' CT409R 5'AGAAGGTTGGAGAACAGCCC3'	280pb
CT410	BF906533	CT410F 5'CTTGATGCAGGACACACAGC3' CT410R 5'GATGTCTGGGAACTTGGGAC3'	189pb
CT411	BF305093	CT411F 5'GGTACCAGGATCAGGCAGTC3' CT411R 5'CTCGGGATTCTTGTAGGGC3'	277pb
CT412	BI038406	CT412F 5'GCTGGAGAACTTTGCACTTG3' CT412R 5'GGGTCACACATCTCACATGG3'	159pb
CT701	AW003075	CT701F 5'TTCTGCCAGGGGTCTTAGG3' CT701R 5'AGCATGCAGGACCTGACC3'	244pb
CT702	BI459819	CT702F 5'TTATTCCAGTGGGCCAGG3' CT702R 5'CTTGTACAATTCCTGTGGC3'	188pb
CT703	BG722032	CT703F 5'CCCGTGCCAAGAAGACTAGC3' CT703R 5'TGGGAGTGGGTGAGACG3'	211pb
CT704	AI652043	CT704F 5'GGAGTGAAGAGGAGTGGCTG3' CT704R 5'TTGGCCACCTCTGCATAAC3'	291pb
CT705	AI671000	CT705F 5'GATGCAAACATTTGGGGC3' CT705R 5'ATCCATCCAACGCTGCTG3'	113pb
CT706	AI962639	CT706F 5'ACGATTCACTACTATGCTC3' CT706R 5'TTAATGCCAGAAGAAATCC3'	134pb
CT707	BE504683	CT707F 5'GATGACAGGACTCCAAAGAAG3' CT707R 5'GGCCCCCATTTGTATCC3'	218pb
CT708	BG720994	CT708F 5'TGTCCTTAATTCCTGCCCC3' CT708R 5'TGAGGAAGTAGAGGCCAG3'	217pb
CT709	AI962439	CT709F 5'CAGGGTGACAGTGGCTTG3' CT709R 5'CATGTTCACTGCTGCGATG3'	180pb
CT710	AI979310	CT710F 5'GGGATCCCTTTGGCTCTC3' CT710R 5'CCAAATATGAGCCGCGAG3'	210pb
CT711	BF222867	CT711F 5'GCTGTGGACTGATGTGGG3' CT711R 5'CATGGGCACTGTCTGGC3'	225pb
CT712	BM921607	CT712F 5'AACCACCCCAAAAAGTGAG3' CT712R 5'GCTGCACAGTCCAAACCG3'	256pb
CT801	AL043076	CT801F 5'CCCGTGCCAAGAAGACTAGC3' CT801R 5'TGGGAGTGGGTGAGACG3'	211pb

**Cont/Anexo 1**

CT802	BG771896	CT802F 5'CGAGAGCGAGGATGATGG3' CT802R 5'GGAGCTCCTGGACCAATG3'	143pb
CT803	BI463953	CT803F 5'CGTCAACTAAAAGTCCGCAAC3' CT803R 5'CCAGCTGTTTCAGTCGAAGG3'	276pb
CT804	BI561956	CT804F 5'TTGTGGAACCATAGGGAGC3' CT804R 5'TGACCAAAGCCACTTCTGAG3'	263pb
CT805	BG719982	CT805F 5'AGAACGTTTCAGCCTCCACAG3' CT805R 5'CAGCAATGCTTAACTGTTTCG3'	131pb
CT806	BI458986	CT806F 5'GAGAAGTCGAGAAGCACAAGC3' CT806R 5'TCCTGGCTGGCTGTGAAG3'	151pb
CT807	BI822069	CT807F 5'AGTCCCCTTTGCTGCTCC3' CT807R 5'GCAGTCTTCTGACACTCCG3'	187pb
CT808	AI371744	CT808F 5'ACACGCTAGCTGATGGATG3' CT808R 5'AAGATGGTCTGGTTCCC3'	189pb
CT809	BG722950	CT809F 5'GATCAGGTGATGCAGAACTC3' CT809R 5'GGTTTTACCTCCATTGTG3'	347pb
CT810	BG771667	CT810F 5'AAAATCAGGTGGCAAAGC3' CT810R 5'ATCGCACCAAGCCCAAAC3'	166pb
CT811	AL041397	CT811F 5'GTCAATAAATAAACCCGAGG3' CT811R 5'GCTGAGGCTGGACGTGC3'	191pb
CT812	BI460168	CT812F 5'GCATCGCAGGTGTTACCC3' CT812R 5'CCTTTCCTTTTAGTGATTGCC3'	217pb
CT901	BI335051	CT901F 5'TGCGCTGGAGGAGAAAAG3' CT901R 5'TGGTACTGGGCTTTACAGAAC3'	209pb
CT902	BG260350	CT902F 5'AACCTGGTGGCTCTGTGC3' CT902R 5'AAACTTATAATCTCTGGCCCC3'	191pb
CT903	BG033516	CT903F 5'ACAAAAGTGCCTGGAGCG3' CT903R 5'GCCATGACCAGCTCACAG3'	232pb
CT904	BE798113	CT904F 5'TCCTGGAAGAGAATGGCG3' CT904R 5'ATAGTGGATGGAGCCCC3'	257pb
CT905	BG772361	CT905F 5'ACACTTTGCCAGGACCC3' CT905R 5'GCAAAGACTCCACCGAGC3'	231pb
CT1001	AA451827	CT1001F 5'CCGAAAAGGAGAAGAGGC3' CT1001R 5'GATCGTTGTCTGTATTGCTGC3'	194pb
CT1002	AA683332	CT1002F 5'ACCTCTCTCACTCCCAGC3' CT1002R 5'TGGTGACCTTGAAAGCC3'	207pb

**Cont/Anexo 1**

CT1003	AA401549	CT1003F 5'TGGCTTGAACAACCTCTGGC3' CT1003R 5'TCACTTCCTGCATGCGTC3'	242pb
CT1004	AA781439	CT1004F 5'TTGGATTTGATGCTGAGG3' CT1004R 5'GCCCAAAGTTACTTCTAGCTC3'	224pb
CT1005	AA854001	CT1005F 5'GATGTTCAGTGCGGGCTC3' CT1005R 5'GCCTGCAAGTTGTAGAATG3'	226pb
CT1006	BQ429641	CT1006F 5'GCCCACTGCTTCCCAAC3' CT1006R 5'AATTAACCCAGAAGAAGAG3'	216pb
CT1007	AI209084	CT1007F 5'CTTTGGACATGGTGCTGG3' CT1007R 5'TTGGGGAAAATGGAGGTG3'	176pb
CT1008	AI203928	CT1008F 5'GGTTCCTTGAAATAGCC3' CT1008R 5'TGCTTTTGGAAACCATCTG3'	166pb

***ANEXO 2***



Antibodies against the cancer-testis antigen CTSP-1 are frequently found in prostate cancer patients and are an independent prognostic factor for biochemical-recurrence after surgery

Raphael B. Parmigiani<sup>1&</sup>, Fabiana Bettoni<sup>1</sup>, Ademar Lopes<sup>2</sup>, André L. Carvalho<sup>3</sup>,  
Francisco Fonseca<sup>2</sup>, Anamaria A. Camargo<sup>1\*</sup>.

<sup>1</sup> Laboratory of Molecular Biology and Genomics. Ludwig Institute for Cancer Research. Hospital Alemão Oswaldo Cruz, São Paulo, SP, Brazil.

<sup>2</sup> Department of Pelvic Tumors, Hospital A.C. Camargo, São Paulo, SP, Brazil.

<sup>3</sup> Department of Head and Neck Tumors, Hospital A.C. Camargo, São Paulo, SP, Brazil.

<sup>&</sup> Present address: Cell Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021.

**Runninghead:** humoral response against CTSP-1 a prognostic marker for prostate cancer.

\*Corresponding author

Anamaria A. Camargo

Lab. of Molecular Biology and Genomics

Ludwig Institute for Cancer Research

Hospital Alemão Oswaldo Cruz

Rua João Julião 245

01323-930 São Paulo SP Brazil

Phone +55 11 33883248

Fax +55 11 32847821

[anamaria@compbio.ludwig.org.br](mailto:anamaria@compbio.ludwig.org.br)

**Keywords:** cancer/testis antigen, CTSP-1, humoral response, immunotherapy, prostate cancer.

## ABSTRACT

**Purpose:** CTSP-1 is a novel cancer-testis antigen frequently expressed in prostate tumors and capable of eliciting natural humoral immune response in prostate cancer patients. In this study, the prognostic significance of the presence of antibodies against CTSP-1 in the sera of prostate cancer patients was analyzed. **Patients and Methods:** Antibodies against CTSP-1 were detected by Western blot in plasma samples from 147 patients with clinically localized prostate cancer who were submitted to radical retropubic prostatectomy. Antibody response was associated with established clinico-pathological parameters. Biochemical-recurrence free survival was estimated using Kaplan-Meier followed by log rank test. Univariate and multivariate analysis were carried out using the Cox regression model. **Results:** Antibodies against CTSP-1 were observed in 37 (25%) patients, but no association was found between the presence of antibodies and the clinico-pathological variables analyzed. On univariate analysis, patients without antibodies against CTSP-1 had a shorter biochemical-recurrence free interval than did those with anti-CTSP-1 antibodies, although the difference between the groups was not statistically significant (57% versus 75%,  $p=0.075$ ). On multivariate analysis, antibodies against CTSP-1 showed a protective effect against biochemical-recurrence (Hazard ratio = 0.36, 95%IC 0.2-0.8,  $p=0.013$ ), being the second most powerful prognostic factor among Gleason score and preoperative PSA levels. Interestingly, none of the patients producing CTSP-1 antibodies developed clinical symptoms of disease progression. **Conclusions:** Our results suggest that CTSP-1 is a promising candidate for prostate cancer immunotherapy and that natural humoral response against CTSP-1 could be used as a complementary prognostic marker for biochemical-recurrence.

## INTRODUCTION

Prostate cancer is the most commonly diagnosed malignancy, and the second leading cause of cancer-related death in men in the United States.<sup>1</sup> Radical retropubic prostatectomy (RRP) is the most common and effective treatment for clinically localized prostate cancer. However, approximately 35% of prostate cancer patients submitting to RRP will have biochemical-recurrence of the disease within 10 years after surgery, as defined by an increase in Prostate Specific Antigen (PSA) level to >0.2 ng/mL.<sup>2</sup> Subsequently, these patients will develop clinical relapse of the disease typically with the development of bone metastases.

The ability to predict biochemical-recurrence and cancer progression after RRP using clinical and pathological variables has been extensively investigated. Recurrence of prostate cancer has been associated with multiple factors including preoperative PSA level and velocity, clinical stage, Gleason score, level of extracapsular extension, seminal vesicle invasion, pelvic lymph node status and surgical margin status.<sup>3,4</sup> However, due to the significant clinical heterogeneity of the disease, these factors are far from accurate and the identification of novel biomarkers that are able to identify patients at high risk for cancer progression is thus clearly needed.<sup>5,6</sup>

At present there is no cure for metastatic prostate cancer and the development of targetted therapies for prostate cancer is of considerable interest. Cancer vaccines represent one type of targetted therapy being investigated for prostate cancer.<sup>7,8</sup> Preclinical and early clinical data suggest that therapeutic cancer vaccines are feasible and safe and provide clinical benefits to patients with prostate cancer.<sup>7,8</sup>

Cancer/Testis (CT) antigens are proteins expressed in normal gametogenic tissues and in different types of tumors.<sup>9,10</sup> Most CT antigens are highly immunogenic, eliciting both humoral and cellular immune responses in cancer patients. Due to their restricted expression pattern and immunogenicity, CT antigens are considered to be ideal targets for cancer immunotherapy.<sup>9-11</sup>

We have recently described a novel CT antigen, named CTSP-1, which is expressed exclusively in normal testis and is aberrantly expressed in 47% of tumor cell lines and 44% of tumors from different histological types.<sup>12</sup> CTSP-1 was found to be aberrantly expressed in 58% of prostate tumors and capable of eliciting a

humoral immune response in approximately 20% of prostate cancer patients.<sup>12</sup> In the present study, we analyzed the prognostic value for biochemical-recurrence of the presence of antibodies against CTSP-1 in the sera of prostate cancer patients with clinically localized disease and submitted to RRP and bilateral pelvic lymphadenectomy.

## **PATIENTS AND METHODS**

*Patients and Surgical Specimens.* The study included 147 consecutive prostate cancer patients who underwent RRP with simultaneous bilateral pelvic lymphadenectomy at the Urology Division of the Pelvic Surgery Department of the Hospital do Cancer, São Paulo, Brazil, between November 1998 and July 2005. All cases had clinically localized disease as suggested by elevated PSA serum levels or palpable nodules on digital rectal examination, and then confirmed by transrectal ultrasound-guided needle biopsy of the prostate. Plasma samples were collected after explicit informed consent and the study had been reviewed and approved by the institution's ethical committee. Patients identifiers were removed and samples were coded to ensure confidentiality. Serum PSA levels were measured with Elecsys chemiluminescent assay on a Elecsys analyser 1010 (Roche). Patients with a second non-prostate primary tumour, or who were submitted to radiotherapy before the surgery, were excluded from this study. Twenty-seven patients received neoadjuvant androgen suppression. After RRP, patients were followed up with serum PSA tests every three months during the first year, twice a year between the second and the fifth year and then annually thereafter. Patients with an increase in PSA levels after RRP were submitted to rectal digital examination, pelvic and abdominal computerized tomography and transrectal ultrasound-guided needle biopsy to confirm local recurrence and bone scan to detect distant metastases. Surgical specimens were staged according to the 1997 American Joint Committee on Cancer System and graded using the Gleason system. Specimens were evaluated for surgical margin status and perineural, angiolymphatic, extracapsular and seminal vesicle invasion, as well as for the presence of metastasis in all removed lymph nodes. Clinicopathological data and follow-up information were obtained retrospectively from patient's medical records. Complete clinical follow-up, including disease

progression and mortality, was available on 143 patients included in this study. Patients lost to follow-up were followed for at least 2 years (range 29 to 64 months). The median follow up time of this cohort was 54.7 months (range 2.9 to 96.2) and the median follow-up time of censored cases was 74.6 months (range 35.1 to 96.2). Biochemical-recurrence was defined as a serum PSA level  $>0.2\text{ng/mL}$  on two consecutive PSA tests after RRP.

*Immunoblotting Analysis.* Plasma samples were collected after explicit informed consent and the study was reviewed and approved by the institution's ethical committee. Patient identifiers were removed and samples were coded to assure confidentiality. Serum antibody responses against the recombinant CTSP-1 protein were tested by standard Western blot analysis as previously described (22). Briefly, purified CTSP-1 recombinant protein was fractioned on SDS/PAGE gel electrophoresis and transferred to Hybond-P PVDF membranes (Amersham Biosciences). After blocking with phosphate buffer containing 5% low-fat milk, the membranes were incubated for 1 h at room temperature with plasma from prostate cancer patients. Plasma antibodies binding to CTSP-1 protein were detected by incubation with goat anti-human IgG HRP-conjugate (Amersham Biosciences) and visualized with ECL-TM Western Blotting Detection Reagents (Amersham Biosciences).

*Statistical Analysis.* Fisher exact test or Chi-square tests were used to evaluate the association between serum antibody response against CTSP-1 and categorical variables. For continuous variables, Mann-Whitney U test was used. The preoperative PSA level was considered a continuous variable, the pathological Gleason score was grouped in two categorical variables (4 to 6 and 7 to 8) and tumor volume was grouped in four categorical variables (1 to 10, 11-30, 31-50 and 51-100%). Biochemical-recurrence free survival (from the date of the RRP until the date of the first PSA measurement  $>0.2\text{ng/mL}$  or to the date of the last follow up) and clinical-recurrence free survival (from the date of the RRP until the date of detection of local or distant disease or until the date of the last follow up) curves were calculated with Kaplan-Meier method. A log-rank test was used to assess statistical differences between the curves. Multivariate analysis was carried out using

Cox proportional hazard regression model. All variables presenting p-value<0.20 on the univariate analysis were selected for building a multiple model (stepwise forward method), since this level of significance eliminates many insignificant variables and ensures that all potential explanatory variables are included. For all tests, type 1 error (alpha) was established as 0.05 and results were considered statistically significant when p<0.05. Patients with missing values for tumor volume, perineural and angiolymphatic invasion were excluded from the analysis of that particular variable. Statistical analyses were performed using SPSS Software 12.0 (SPSS Inc., Chicago, IL).

## **RESULTS**

*Humoral response against CTSP-1 and pathological variables.* The pathological information for the 147 patients analyzed in this study is summarized in Table 1. The median age of the patients at diagnosis was 62 years (range 41-75) and the median preoperative PSA level was 8.6 ng/mL (range 1.2 to 44, percentiles 5.7, 8.7, 14.4). Humoral response against CTSP-1 was detected by Western blot in 37 (25.2%) patients.

The distribution of the pathological variables according to the presence of humoral response against CTSP-1 is listed in Table 2. No statistically significant association was observed between the presence of antibodies against CTSP-1 and the following pathological parameters: preoperative PSA level (p=0.173), tumor stage (p=0.126), lymph node status (p=0.409), Gleason score (p=0.139), tumor volume (p=0.478), perineural invasion (p=0.427), angiolymphatic invasion (p=0.427), extracapsular invasion (p=0.319), seminal vesicle invasion (p=0.128) and surgical margins (p=0.344).

*Humoral response against CTSP-1 and Biochemical-recurrence.* Kaplan-Meier method was used to estimate the relationship between the presence of humoral response against CTSP-1 and the risk of Biochemical-recurrence of the disease after surgery. The probability of 5-year Biochemical-recurrence free survival among all 147 patients evaluated was 61.8%. Patients who did not produce antibodies against CTSP-1 had a shorter Biochemical-recurrence free interval than did those producing

anti-CTSP-1 antibodies, although the difference between the groups was not statistically significant (57.2 versus 75.6%,  $p=0.075$ , Figure 1A). Interestingly, none of the patients producing antibodies against CTSP-1 developed clinical symptoms of the disease progression (Figure 1B). Five-year biochemical-recurrence free survival in our patient population was also influenced by the following parameters: preoperative PSA level ( $p<0.001$ ), tumor stage ( $p<0.001$ ), Gleason score ( $p=0.016$ ), tumor volume ( $p<0.001$ ), seminal vesicle invasion ( $p<0.001$ ) and surgical margins ( $p<0.001$ ) (Table 3).

*Humoral response against CTSP-1 as an independent prognostic marker for biochemical-recurrence.* A multivariate analysis was performed to determine whether the presence of a natural humoral immune response against CTSP-1 was an independent factor in predicting biochemical-recurrence of the disease after surgery. The preoperative PSA level, the Gleason score and the presence of antibodies against CTSP-1 were considered as independent prognostic factors for biochemical-recurrence free survival (Table 4). Patients producing antibodies against CTSP-1 had a lower risk for biochemical-recurrence than patients without anti-CTSP-1 (HR 0.36, 95%CI 0.2-0.8,  $p=0.013$ ).

## DISCUSSION

Among various types of epithelial cancers, prostate cancer appears to be one of the best targets for specific immunotherapy. It is generally recognized that the human immune system is capable of mounting a significant immune response against the prostate tissue that can be both unremitting and destructive in nature.<sup>13</sup> In addition, it has been demonstrated that immunity against normal and cancerous prostate tissue can be induced by a variety of immunotherapies, including peptide, DNA, virus-based, dendritic cell and genetically-modified cancer vaccines.<sup>7,8</sup> Moreover, the generally slow progression of prostate cancer provides a relatively long interval of time for the induction and potentiation of T-cell mediated anti-tumoral immunity during immunotherapeutic treatment. Finally, given that prostate is a relatively dispensable organ, often removed by surgery as part of prostate cancer



treatment, immune responses against prostate-related antigens can be considered as cancer specific in patients with recurrent metastases.

The results of several cancer vaccine trials suggest a clinical benefit to patients with prostate cancer and phase III trials are current underway in patients with androgen-independent prostate cancer. The antigens included in these studies are mainly proteins with known prostate-restricted expression (e.g. PSA, PSMA, PAP) and most of them are not natural targets of an immune response in patients with cancer.<sup>7,8</sup> Identifying immunogenic proteins in patients with prostate cancer could thus lead to the development of other potential vaccine target antigens.

CT antigens represent a novel and expanding family of proteins that are expressed in normal gametogenic tissues and different types of tumors, and that are immunogenic in cancer patients.<sup>9-11</sup> These antigens are not immunogenic in normal tissues, as their expression is restricted to MHC class I-negative germ cells.<sup>11</sup> Thus, CT antigens are considered ideal targets for cancer immunotherapy.

To date, 89 individual CT antigen genes have been described, which are organized into 44 families.<sup>10</sup> Several CT antigens are expressed in prostate tumors including PAGE-4, NY-ESO-1, LAGE-1, XAGE-1 and members of the MAGE-A family, but their utility as targets for prostate cancer immunotherapy remains poorly explored.<sup>15-19</sup> Reports on the association between CT antigen expression and/or spontaneous serological responses with disease prognosis are limited to a few studies involving a small number of patients, with most studies focused only on the CT antigen, NY-ESO-1. In one of the more comprehensive studies, Nakada et al. examined NY-ESO-1 mRNA expression and the presence of NY-ESO-1 antibodies in the sera of patients with advanced prostate cancer.<sup>15</sup> NY-ESO-1 expression and antibodies were predominantly detected in stage D2 patients when compared to stage D1 patients.<sup>15</sup> In another study on prostate tumors, NY-ESO-1 protein was predominantly detected in hormone-refractory tumors (14.6%) compared to localized tumors (3%) and anti-NY-ESO-1 antibodies were also more frequently found in the serum of patients with advanced disease (18.9%). Spontaneous serological response against NY-ESO-1 was associated with poor survival in hormone-refractory prostate cancer patients.<sup>16</sup>

## **CONCLUSION**

In the present study, a natural humoral immune response against CTSP-1 was detected in 25% (37/147) of the patients with clinically localized prostate cancer. To our knowledge, CTSP-1 is the most immunogenic CT antigen described to date for prostate cancer, and thus should be considered a promising candidate for prostate cancer immunotherapy. We also demonstrated that the presence of antibodies against CTSP-1 is an independent predictor of biochemical-recurrence after RRP in patients with clinically localized disease. These results, together with the ones presented by others<sup>16</sup> suggest that the presence of a natural immune response against CT antigens could be used as prognostic factors in prostate cancer and deserves further evaluation.

## **REFERENCES**

1. Jemal, A.; Siegel, R.; Ward, E.; Murray, T.; Xu, J.; and Thun, M.J. Cancer statistics, 2007. *CA Cancer J Clin*, **57**:43, 2007.
2. Roehl, K.A.; Han, M.; Ramos; C.G.; Antenor, J.A. and Catalona, W.J.: Cancer progression and survival rates following anatomical radical retropubic prostatectomy in 3,478 consecutive patients: long-term results. *J Urol*, **172**:910, 2004.
3. Grossfeld, G.D.; Latini, D.M.; Lubeck, D.P.; Mehta, S.S. and Carroll, P.R.: Predicting recurrence after radical prostatectomy for patients with high risk prostate cancer. *J Urol*, **169**:157, 2003.
4. Buhmeida, A.; Pyrhonen, S.; Laato, M. and Collan, Y.: Prognostic factors in prostate cancer. *Diagn Pathol*, **1**:4, 2006.
5. Quinn, D.I.; Henshall, S.M. and Sutherland, R.L.: Molecular markers of prostate cancer outcome. *Eur J Cancer*, **41**:858, 2005.

6. Chin, J.L. and Reiter, R.E.: Molecular markers and prostate cancer prognosis. *Clin Prostate Cancer*, **3**:157, 2004.
7. Webster, W.S.; Small, E.J.; Rini, B.I. and Kwon, E.D.: Prostate cancer immunology: biology, therapeutics, and challenges. *J Clin Oncol*, **23**:8262, 2005.
8. Fong, L. and Small, E.J.: Immunotherapy for prostate cancer. *Curr Urol Rep*, **7**:239, 2006.
9. Scanlan, M.J.; Gure, A.O.; Jungbluth, A.A.; Old, L.J. and Chen, Y.T.: Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunol Rev*, **188**:22, 2002.
10. Simpson, A.J.; Caballero, O.L.; Jungbluth, A.; Chen, Y.T. and Old, L.J.: Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer*, **5**:615, 2005.
11. Bodey, B.: Cancer-testis antigens: promising targets for antigen directed antineoplastic immunotherapy. *Expert Opin Biol Ther*, **2**:577, 2002.
12. Parmigiani, R.B.; Bettoni, F.; Vibriantovsky, M.D.; Lopes, M.H.; Martins, W.K.; Cunha, I.W.; et al: Characterization of a Cancer/Testis (CT) antigen gene family capable of eliciting humoral response in cancer patients. *Proc Natl Acad Sci USA*, **103**:18066, 2006.
13. Nelson, W.G.; De Marzo, A.M.; DeWeese, T.L. and Isaacs, W.B.: The role of inflammation in the pathogenesis of prostate cancer. *J Urol*, **172**:S6, 2004.
14. Prikler, L.; Scandella, E.; Men, Y.; Engeler, D.S.; Diener, P.A.; Gillessen, S. et al.: Adaptive immunotherapy of the advanced prostate cancer - cancer testis antigen(CTA) as possible target antigens. *Aktuelle Urol*, **35**:326, 2004.

15. Nakada, T.; Noguchi, Y.; Satoh, S.; Ono, T.; Saika, T.; Kurashige, T. et al.: NY-ESO-1 mRNA expression and immunogenicity in advanced prostate cancer. *Cancer Immun*, **3**:10, 2003.
16. Fossa, A.; Berner, A.; Fossa, S.D.; Hernes, E.; Gaudernack, G. and Smeland, E.B.: NY-ESO-1 protein expression and humoral immune responses in prostate cancer. *Prostate*, **59**:440, 2004.
17. Lethe, B.; Lucas, S.; Michaux, L.; De Smet, C.; Godelaine, D.; Serrano, A. et al.: LAGE-1, a new gene with tumor specificity. *Int J Cancer*, **76**:903, 1998.
18. Egland, K.A.; Kumar, V.; Duray, P. and Pastan, I.: Characterization of overlapping XAGE-1 transcripts encoding a cancer testis antigen expressed in lung, breast, and other types of cancers. *Mol Cancer Ther*, **1**:441, 2002.
19. Kufer, P.; Zippelius, A.; Lutterbuse, R.; Mecklenburg, I.; Enzmann, T.; Montag, A. et al.: Heterogeneous expression of MAGE-A genes in occult disseminated tumor cells: a novel multimarker reverse transcription-polymerase chain reaction for diagnosis of micrometastatic disease. *Cancer Res*, **62**:251, 2002.

Reserach Funding: This study was supported by The Ludwig Institute for Cancer Research (LICR) and by the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). RBP and FB were sponsored by fellowships from FAPESP.

Figure 1. Survival curves according to the presence of an anti-CTSP-1 humoral immune response. (A) Biochemical-recurrence-free survival (B) Disease-free survival.

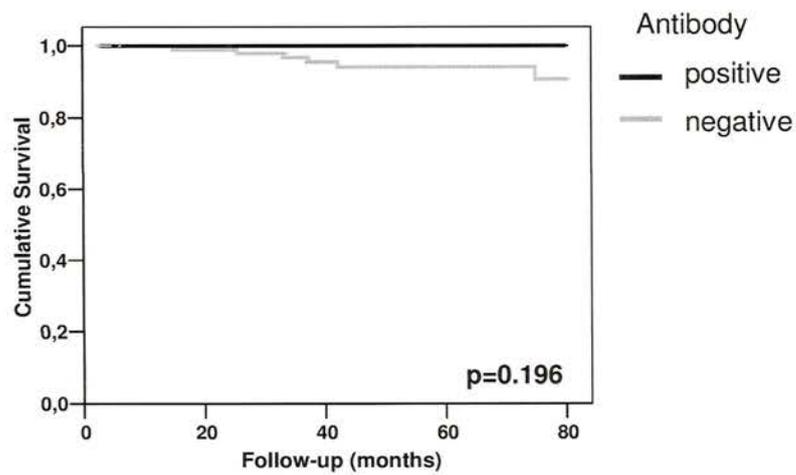
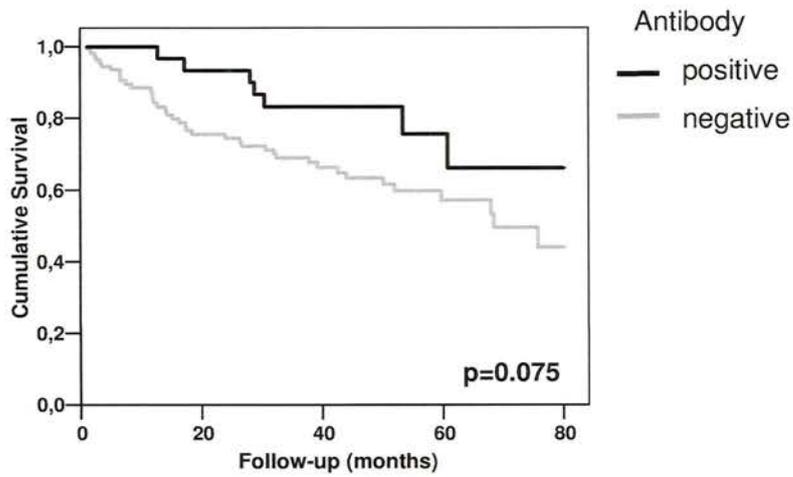


Table 1. Pathological characteristics of the prostate cancer patients.

Characteristic	Category	No. (%)
Pathologic T staging	T2a	
	T2b	16 (10.9)
	T3a	82 (55.8)
	T3b	27 (18.4)
	T4	14 (9.5)
Pathologic N staging	N0	8 (5.4)
	N1	145 (98.6)
Gleason score	4 - 6	2 (1.4)
	7 - 8	98 (66.7)
Tumour volume (%)*	1 - 10	49 (33.4)
	11 - 30	54 (38.8)
	31 - 50	59 (42.4)
	51 - 100	17 (12.2)
Perineural invasion*	No	9 (6.5)
	Yes	40 (27.4)
Angiolymphatic invasion*	No	106 (72.6)
	Yes	22 (15.1)
Extracapsular invasion	No	124 (84.9)
	Focal	70 (47.6)
	Diffuse	65 (44.2)
Margins	Negative	12 (8.2)
	Positive	115 (78.2)
Seminal vesicle invasion	No	32 (21.8)
	Yes	115 (78.2)

\* Tumor volume has eight patient data missing and perineural and angiolymphatic invasion have 1 patient data missing.

Table. 2 Patient's distribution according to the existence of humoral response against CTSP-1.

Characteristic	Humoral Response (%)		p-value
	No	Yes	
Pathologic T staging			
T2a	13 (81.3)	3 (18.7)	0.126*
T2b	60 (73.2)	22 (26.8)	
T3a	17 (63.0)	10 (37.0)	
T3b	14 (100.0)	0 (0.0)	
T4	6 (75.0)	2 (25.0)	
Pathologic N staging			
N0	108 (74.5)	37 (25.5)	0.409#
N1	2 (100.0)	0 (0)	
Gleason score			
4 - 6	77 (78.6)	21 (21.4)	0.139*
7 - 8	33 (67.3)	16 (32.7)	
Tumour volume (%)			
1 - 10	40 (74.1)	14 (25.9)	0.478*
11 - 30	41 (69.5)	18 (30.5)	
31 - 50	15 (88.2)	2 (11.8)	
51 - 100	7 (77.8)	2 (22.2)	
Perineural invasion			
No	28 (70.0)	12 (30.0)	0.427*
Yes	81 (76.4)	25 (23.6)	
Angiolymphatic invasion			
No	94 (75.8)	30 (24.2)	0.449*
Yes	15 (68.2)	7 (31.8)	
Extracapsular invasion			
No	55 (78.6)	15 (21.4)	0.319*
Focal	48 (73.8)	17 (26.2)	
Diffuse	7 (58.3)	5 (41.7)	
Margins			
Negative	84 (73.0)	31 (27.0)	0.344*
Positive	26 (81.3)	6 (18.7)	
Seminal vesicle invasion			
No	98 (73.1)	36 (26.9)	0.128#
Yes	12 (92.3)	1 (7.7)	

\* Chi square test

# Fisher exact test

Table 3. Univariate analysis of biochemical-recurrence free survival (BRFS).

Characteristic	Category	5-year BRFS (%)	p-value*
Pathologic T staging	T2a	0	<0.001
	T2b	66.6	
	T3a	80.3	
	T3b	23.6	
	T4	18.8	
Pathologic N staging	N0	62.0	0.341
	N1	50.0	
Gleason score	4 - 6	69.1	0.016
	7 - 8	49.9	
Tumour volume (%)	1 - 10	73.0	<0.001
	11 - 30	67.7	
	31 - 50	61.5	
	51 - 100	22.2	
Perineural invasion	No	67.3	0.134
	Yes	58.9	
Angiolymphatic invasion	No	64.0	0.093
	Yes	48.5	
Extracapsular invasion	No	73.1	0.161
	Focal	54.1	
	Diffuse	41.6	
Margins	Negative	68.4	<0.001
	Positive	40.5	
Seminal vesicle invasion	No	66.9	<0.001
	Yes	40.5	
	Continuous		
Preoperative PSA	Positive	75.6	0.075
	Negative	57.2	

\* Log-rank test

**Table 4. Multivariate analysis of biochemical-recurrence free survival.**

Variables	Category	HR	95%CI	p-value*
Preoperative PSA	Continuous	1.09	1.06 – 1.13	<0.001
Antibody	Negative	1.0	Ref.	0.013
	Positive	0.36	0.2 – 0.8	
Gleason score	4 – 6	1.0	Ref.	0.018
	7 – 8	2.0	1.3 – 3.7	

\* Cox regression

***ANEXO 3***



## Characterization of a cancer/testis (CT) antigen gene family capable of eliciting humoral response in cancer patients

Raphael B. Parmigiani, Fabiana Bettoni, Maria D. Vibranovski, Marilene H. Lopes, Waleska K. Martins, Isabela W. Cunha, Fernando A. Soares, Andrew J. G. Simpson, Sandro J. de Souza, and Anamaria A. Camargo

*PNAS* published online Nov 17, 2006;  
doi:10.1073/pnas.0608853103

**This information is current as of November 2006.**

### Supplementary Material

Supplementary material can be found at:  
[www.pnas.org/cgi/content/full/0608853103/DC1](http://www.pnas.org/cgi/content/full/0608853103/DC1)

This article has been cited by other articles:  
[www.pnas.org#otherarticles](http://www.pnas.org#otherarticles)

### E-mail Alerts

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

### Rights & Permissions

To reproduce this article in part (figures, tables) or in entirety, see:  
[www.pnas.org/misc/rightperm.shtml](http://www.pnas.org/misc/rightperm.shtml)

### Reprints

To order reprints, see:  
[www.pnas.org/misc/reprints.shtml](http://www.pnas.org/misc/reprints.shtml)

Notes:

# Characterization of a cancer/testis (CT) antigen gene family capable of eliciting humoral response in cancer patients

Phael B. Parmigiani\*, Fabiana Bettoni\*, Maria D. Vibranovski\*\*, Marilene H. Lopes\*, Waleska K. Martins\*, Bela W. Cunha‡, Fernando A. Soares‡, Andrew J. G. Simpson§, Sandro J. de Souza\*, and Anamaria A. Camargo\*¶

Ludwig Institute for Cancer Research, SP 01509-010, São Paulo, Brazil; \*Biochemistry Department, University of São Paulo, SP 05508-900, São Paulo, Brazil; †Immunology Department, Hospital A. C. Camargo, São Paulo, SP 01509-010, Brazil; and ‡Ludwig Institute for Cancer Research, Memorial Sloan-Kettering Cancer Center, New York, NY 10158

Communicated by Lloyd J. Old, Ludwig Institute for Cancer Research, New York, NY, October 6, 2006 (received for review August 9, 2006)

Cancer/testis (CT) antigens are immunogenic proteins expressed in normal gametogenic tissues and in different types of tumors. CT antigens are promising candidates for cancer immunotherapy, and identification of novel CT antigens is a prerequisite for the development of cancer vaccines. We have identified a CT antigen, named CTSP-1, with partial similarity to the breast differentiation antigen NY-BR-1. CTSP-1 presents several splicing and polyadenylation variants and has a very restricted expression pattern among normal tissues. CTSP-1 is exclusively expressed in normal testis and aberrantly expressed in 47.6% (10 of 21) of tumor cell lines and 44.4% (75 of 169) of tumors from different histological types. The highest percentages of positive expression were observed in melanomas (59.0%) followed by prostate (58.0%) and lung (57.0%) tumors. CTSP-1 is part of a highly conserved gene family, and members of this family also have a restricted expression pattern and similar protein structure. Antibodies against members of this gene family were detected in 10% (14 of 141) of plasma samples from patients with a wide spectrum of tumors. The highest percentages of antibody response were observed in patients with prostate (20.8%), thyroid (20.0%), and breast (16.6%) tumors. Because of its very restricted expression pattern in normal tissues and immunogenicity in different types of tumors, CTSP-1 should be considered a promising candidate for cancer immunotherapy.

Immunotherapy | tumor antigen

Cancer/testis (CT) antigens are predominantly expressed in normal gametogenic tissues as well as in different histological types of tumors (1–4). In testis, CT antigens are expressed exclusively in cells of the germ cell lineage, although there is a marked variation in the protein expression pattern during different stages of sperm development. Likewise, a heterogeneous expression is observed in tumors (1, 4). The methylation status of the promoter region seems to be the main, but not the only, regulator of their specific expression pattern (1, 3, 4). Most CT antigens have no defined biological function, but their involvement in signaling, transcription, translation, and chromosomal recombination has been proposed (3, 4). It has also been proposed that the aberrant expression of CT antigens in tumors recapitulates the germ-line gene expression program, and that it is related to some characteristics of the neoplastic phenotype such as immortality, invasiveness, immune evasion, and metastatic capacity (3, 5). Because of their restricted expression pattern, CT antigens are considered ideal targets for cancer immunotherapy (1, 6). Indeed, a small subset of patients immunized with the known CT antigens MAGE-A and NY-ESO-1 have shown clinical benefits after immunization (7–10). However, because CT antigens are expressed in only a small subset of human tumors and in only a fraction of cases of a given tumor type, the identification of additional CT antigens is crucial for improving current immunotherapy protocols. Presently, 44 distinct CT-antigen families

have been described, of which several have multiple members resulting in a total of 89 transcripts (3).

During the process of identification of genes located on human chromosome 21 (HC21) (11), we found a gene (C21ORF99) that was predominantly expressed in testis and had partial similarity to the 5' region of the breast differentiation antigen NY-BR-1 gene (12). In this work, we refined the characterization of C21ORF99 gene structure and renamed it CTSP-1. We found that CTSP-1 has a restricted expression pattern in normal tissues characteristic of CT antigens and is expressed at a high frequency in tumor cell lines and tumor samples. We also found that CTSP-1 is part of a highly conserved gene family, and that some members of this family also have an expression pattern characteristic of CT antigens. Antibodies against members of this gene family were detected in plasma samples from patients with a wide spectrum of tumors.

## Results and Discussion

**CTSP-1 Gene Structure and Polyadenylation Variants.** Using the information generated by alignments between ESTs and the sequence of the HC21 (11), we identified a gene, named C21ORF99 (AF427490), which was predominantly expressed in normal testis and showed partial similarity to the 5' region of the breast differentiation tumor antigen NY-BR-1 gene (AF269087) and its paralog NY-BR-1.1 (AF269088), located at chromosomes 10 and 18, respectively (12).

C21ORF99 sequence was obtained by sequencing the full insert of cDNA clones from which the ESTs aligning to HC21 were derived (11). However, based on the mapping of NY-BR-1 and NY-BR-1.1 to HC21 sequence, we suspected that the initial sequence obtained for C21ORF99 (derived from cDNA clones containing both polyA signal and tail) corresponded to a shorter polyadenylation variant (Fig. 1).

To extend the initial C21ORF99 sequence and verify the existence of additional polyadenylation variants, we used a combination of RT-PCR and 3'-RACE experiments. First, we designed two pairs of primers specific for HC21 sequence in regions that showed high similarity to NY-BR-1 and NY-BR-1.1. RT-PCR products obtained with these primers were cloned and

Author contributions: F.B. and M.D.V. contributed equally to this work; R.B.P., M.D.V., F.A.S., A.J.G.S., S.J.d.S., and A.A.C. designed research; R.B.P., F.B., M.D.V., M.H.L., W.K.M., I.W.C., and A.A.C. performed research; W.K.M. contributed new reagents/analytic tools; R.B.P., F.B., M.D.V., M.H.L., I.W.C., F.A.S., A.J.G.S., S.J.d.S., and A.A.C. analyzed data; and R.B.P., M.D.V., A.J.G.S., S.J.d.S., and A.A.C. wrote the paper.

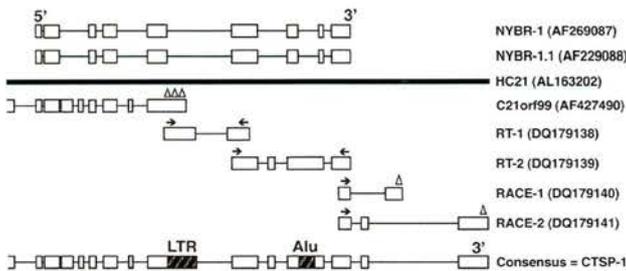
The authors declare no conflict of interest.

Abbreviations: CT, cancer/testis; HC21, human chromosome 21.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ179138–DQ179141).

¶To whom correspondence should be addressed. E-mail: anamaria@ludwig.org.br.

© 2006 by The National Academy of Sciences of the USA



**Fig. 1.** Schematic representation of the strategies used in the generation of CTSP-1 consensus sequence. NY-BR-1, NY-BR-1.1, and C21ORF99 sequences are aligned to the genomic sequence of the human chromosome 21 (HC21). Exons are represented as boxes, and lines correspond to introns. Primers are represented with arrows, and triangles correspond to polyadenylation signals. Repetitive elements (LTR and Alu) are also represented.

sequenced, and the original C21ORF99 sequence was extended and illustrated in Fig. 1.

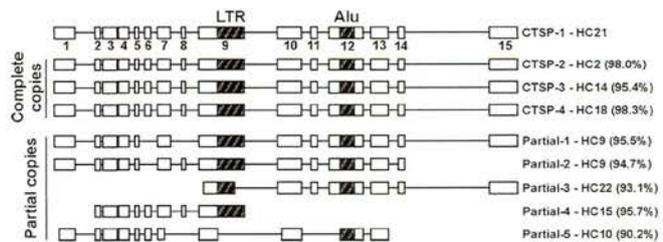
Because the 3' end of NY-BR-1 (from nucleotide 3,199 to 58) and NY-BR-1.1 (from nucleotide 3,220 to 3,673) did not show significant similarity to the HC21 sequence, we have also performed 3'-RACE experiments to generate a full-length sequence. Two distinct 3'-RACE fragments corresponding to two additional polyadenylation variants were amplified and sequenced (Fig. 1). All 3'-ESTs available in public databases and corresponding to CTSP-1 are derived from the shortest variant, suggesting it is the most abundantly expressed variant (data not shown).

A consensus mRNA sequence of 3,915 bp, organized in 15 exons, was obtained by assembling the C21ORF99 sequence, the RT-PCR products, and the longest 3'-RACE fragment (Fig. 1). It was named CTSP-1. This sequence is provided as Data Set which is published as supporting information on the PNAS web site. The alignment between the CTSP-1 consensus sequence and NY-BR-1 nucleotide sequence is not continuous but presents seven regions of high similarity (average identity of 6%) that encompass 1,223 of the 3,915 nucleotides (31%) of CTSP-1 sequence.

Careful analysis of the CTSP-1 mRNA consensus sequence revealed the presence of two transcribed repetitive elements located within exons 9 (LTR repeat) and 12 (Alu repeat), which could not be found in the NY-BR-1 and NY-BR-1.1 sequences (Fig. 1). Because the polyadenylation sites associated with the shorter polyadenylation variant are contained within the LTR repeat, it is tempting to speculate that the presence of this repetitive element within exon 9 is associated with the occurrence of alternative polyadenylation. Supporting this hypothesis, as recently been reported that the insertion of retroelements in the human genome can modulate gene expression by introducing intragenic polyadenylation signals as well as by disrupting the structure (13).

**P-1 Gene Family.** Recent duplication events have been described for many CT antigens such as NY-ESO-1 and SSX, resulting in two or more identical copies with identical coding sequences (14, 15). To identify additional CTSP family members,

CTSP-1 consensus sequence was aligned to the human genome. Three complete copies of highly similar genes were found on chromosomes 2 (CTSP-2, 98.0% identity), 14 (CTSP-3, 95.4% identity), and 18 (CTSP-4, 98.3% identity). The exon-intron structure seems highly conserved among these four members, including the presence of both LTR and Alu repeats, suggesting they are products of recent gene-duplication events (Fig. 2). However, because the exon-intron structure for these genes were predicted based on the alignment of CTSP-1 to the



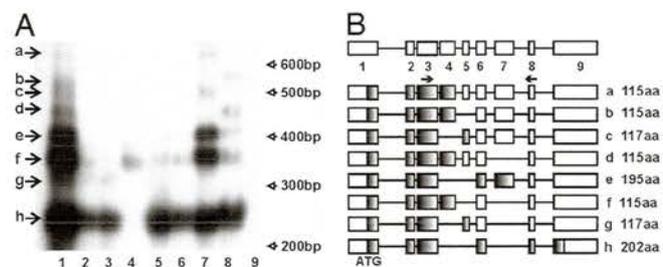
**Fig. 2.** Schematic representation of the exon/intron structure of the CTSP-1 family members. Complete and partial copies are represented with their corresponding chromosomal location and similarity to HC21 genomic sequence.

genomic sequence, confirmation of this structure will depend on the generation of expressed sequences covering the full extension of all these genes. ESTs and/or cDNA clones corresponding to all these family members can be found in public databases, suggesting that all of them are transcribed. As for CTSP-1, all ESTs and cDNA clones corresponding to these family members represent the first nine exons and are probably derived from the shorter polyadenylation variant.

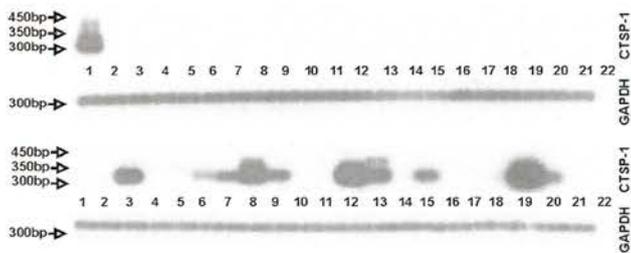
In addition to these complete copies, two partial copies containing the LTR and Alu repeat were identified on chromosome 9. However, their exon-intron structure is slightly different, and exons 6, 8, and 15 are missing when compared with the complete copies (Fig. 2). Additionally, partial copies covering <58% of the CTSP-1 mRNA length and with a lower similarity (<96%) were found on chromosomes 22, 15, and 10 (Fig. 2), and they probably result from failed duplication events, as has also been observed for other CT-gene families, such as SSX. An extended analysis of the evolution and conservation of this gene family among primates will be presented elsewhere.

**Expression Profile and Alternative Splicing.** Because one of the criteria for identifying CT-antigen genes is their specific expression in tumors, but not in normal tissues except in testis, the mRNA expression pattern of CTSP-1 was determined by RT-PCR. Primers used in the expression analysis were manually designed to ensure specific amplification, and RT-PCR products were sequenced to confirm their specificity.

CTSP-1 mRNA expression was first examined in normal testis. A pair of primers located in exons 3 and 8 was used in RT-PCR with 40 cycles of amplification. Three predominant fragments of 239, 376, and 402 bp were obtained (Fig. 3A), and the whole amplification product was cloned without prior gel purification.



**Fig. 3.** CTSP-1 splicing variants. (A) Southern blot of RT-PCR products amplified with CTSP-1-specific primers. Splicing variants (a-h) are indicated with arrows. cDNA samples used were as follows: normal testis tissue (lane 1), tumor cell lines A172 (lane 2), A2058 (lane 3), H1155 (lane 4), breast tumor samples (lanes 5 and 6), prostate tumor samples (lanes 7 and 8), and no cDNA-negative control (lane 9). (B) Schematic representation of CTSP-1 splicing variants (a-h). The coding exons are represented as gray boxes. Primers used in RT-PCR are represented as arrows, and the size of the putative ORF for each splicing variant is provided in amino acids (aa).



**4.** CTSP-1 mRNA expression pattern in normal tissues and tumor cell lines. Southern blot of RT-PCR products amplified with CTSP-1-specific primers. (A) Normal cDNA samples used were as follows. Lanes: 1, testis; 2, lung; 3, prostate; 4, small intestine; 5, breast; 6, brain; 7, heart; 8, uterus; 9, bone marrow; 10, placenta; 11, colon; 12, fetal brain; 13, liver; 14, fetal liver; 15, muscle; 16, salivary gland; 17, spinal cord; 18, kidney; 19, spleen; 20, skeletal muscle; 21, adrenal gland; and 22, no cDNA-negative control. (B) cDNA samples from tumor cell lines used were: 1, Caski; 2, HeLa; 3, A172; 4, T98G; 5, 60; 6, K562; 7, H358; 8, H1155; 9, Du145; 10, PC3; 11, SCABER; 12, IM9; 13, DU; 14, MCF-7; 15, MDA-MB-436; 16, MDA-MB-231; 17, SW-480; 18, SAOS-2; 19, A2058; 20, SKmel-25; 21, HEPG2; and 22, no cDNA-negative control. GAPDH amplification was used as positive control for cDNA synthesis.

ones with different insert sizes were selected for sequencing. Using this strategy, we were able to identify at least eight splicing variants in which exons 4, 5, and 7 were alternatively spliced, as illustrated in Fig. 3B. No alternative splicing variants involving exons 1, 2, and 9 were observed when RT-PCR analysis was tried out using primers located on exons 1 and 9.

CTSP-1 mRNA expression was then analyzed in cDNA samples derived from 20 additional normal tissues and 21 tumor cell lines. To favor the detection of all splicing variants, RT-PCR products were transferred to nylon membranes and hybridized with a cDNA probe corresponding to exon 3, which is present in all splicing variants. Among the 21 normal tissues, CTSP-1 expression was restricted to testis and could also be detected in 6% (10 of 21) of all analyzed tumor cell lines (Fig. 4). Interestingly, although the expression of several splicing variants was detected in normal testis cDNA, a less complex pattern of alternative splicing could be visualized in cDNA derived from tumor cell lines (Fig. 3A). The shortest variant, in which exons 5, and 7 are skipped, appears to be the most abundant and frequently expressed variant. Although we do not have a biological explanation for such diversity in testis, our results are in agreement with the general view that a higher proportion of splicing variants are present in testis, and that alternative splicing is particularly important in testis development and spermatogenesis (16).

The expression profile of CTSP-1 was then analyzed in 169 tumor samples derived from 10 different histological types of tumors. Positive expression was detected in 44.4% (75 of 169) of samples (Table 1). The highest percentages of positive expression were observed in melanomas (59.0%) followed by prostate (58.0%) and lung (57%) tumors. As observed in the tumor cell lines, the shorter variant was the most abundant and frequently expressed (data not shown). This expression pattern is partially in agreement with the classification proposed by Melan et al. (2). In the proposed classification, melanomas and lung tumors were defined as tissues with higher CT expression; prostate, prostate, and esophagus were classified as having moderate and stomach and colon as having low CT expression. Melanomas, thyroid, and uterus tumors were not classified in this analysis due to the insufficient number of samples analyzed.

The expression pattern of the other three family members was also analyzed by RT-PCR. Primers used in the expression analysis were manually designed to ensure specific amplification. All RT-PCR products were sequenced to confirm their speci-

**Table 1.** Frequency of mRNA expression of CTSP family members in tumor samples

Tumor	CTSP-1 (%)	CTSP-2 (%)	CTSP-4 (%)
Breast	9/25 (36.0)	1/6 (16.5)	0/6 (0.0)
Colon	7/18 (39.0)	0/14 (0.0)	0/14 (0.0)
Esophagus	2/5 (40.0)	2/4 (50.0)	0/4 (0.0)
Glioblastoma	6/13 (46.0)	Not done	Not done
Lung	8/14 (57.0)	2/11 (18.8)	0/11 (0.0)
Melanoma	10/17 (59.0)	1/9 (11.0)	0/9 (0.0)
Prostate	14/24 (58.0)	8/17 (47.0)	0/17 (0.0)
Stomach	4/9 (44.0)	1/3 (33.3)	0/3 (0.0)
Thyroid	7/24 (29.0)	0/6 (0.0)	0/6 (0.0)
Uterus	8/20 (40.0)	4/14 (28.5)	0/14 (0.0)
Total	75/169 (44.4)	19/84 (22.6)	0/84 (0.0)

ficity. Although the expression of the other CTSP-1 family members was also restricted among normal tissues, they were not frequently expressed in tumor cell lines and tumor samples. This fact has also been observed for other CT-antigen members of the MAGE and SSX families (17–19). A positive expression in normal testis was detected for CTSP-2 and -4, the first also being expressed in brain, fetal brain, and spinal cord among the 21 normal tissues analyzed (data not shown). Expression of CTSP-2 and -4 was not detected in any of the tumor cell lines (data not shown), although CTSP-2 expression was detected in tumor samples at a lower frequency when compared with CTSP-1 (Table 1). CTSP-3 expression was not detected in any of the samples analyzed. Splicing variants were also detected for CTSP-2 and -4.

An important element in the induction of CT-antigen gene expression is promoter demethylation. It has been shown that demethylation of CT gene promoters with 5-Aza-2'-deoxycytidine induces antigen expression in cells that do not normally produce them (20–22). CTSP-1 gene expression was induced in the MCF-7 breast tumor cell line after treatment with 5-Aza-2'-deoxycytidine, suggesting that methylation indeed plays an important role in CTSP-1 gene expression regulation (Fig. 7, which is published as supporting information on the PNAS web site).

**CTSP-1 Protein.** The CTSP-1 sequence was translated to amino acid sequence, and the presence of a stop codon was detected in exon 4, generating a 115-aa-long putative protein, which is much shorter than the NY-BR-1 (1,341 aa) and NY-BR-1.1 (1,011 aa) proteins. The presence of “premature” stop codons was also detected in all splicing variants, generating small “truncated” proteins of size ranging from 115 to 202 aa (Fig. 3B). Interestingly, the longest ORF was predicted from the shortest splicing variant (without exons 4, 5, and 7), which is the most frequently expressed variant in testis, tumor cell lines, and tumor samples (Fig. 3). Although restricted to the N-terminal region of the NY-BR-1, protein similarity between the CTSP-1 longest ORF and the NY-BR-1 protein is high [identities = 159 of 249 (63%), positives = 180 of 249 (72%), and gaps = 46 of 249 (18%)]. All other members of the family seem to have “premature” stop codons as evaluated by aligning the CTSP-1 consensus sequence to the genomic sequence of chromosomes 2, 14, and 18.

Motif analysis of the amino acid sequence corresponding to the longest ORF of CTSP-1 identified four tandem ankyrin repeats (14–47, 48–80, 81–103, and 114–146 aa). However, the bZIP site and the tandem repetitive elements located at the C-terminal region of the NY-BR-1 protein could not be found in the CTSP-1 sequence. A similar protein structure is present in the other CTSP-1 family members.

The existence of the CTSP-1 protein was confirmed by

Western blot by using a mouse polyclonal antibody against the CTSP-1 recombinant protein. As shown in Fig. 5A, a single band with the expected molecular mass ( $\approx 23$  kDa) was detected in all protein extracts from normal testis. No higher molecular mass bands were detected. Preimmune serum and serum raised against CTSP-1 but depleted of anti-CTSP-1 Igs were used as controls of serum specificity.

CTSP-1 protein was also analyzed by immunohistochemistry in normal testis. Again, preimmune serum and serum raised against CTSP-1 but depleted of anti-CTSP-1 Igs were used as controls of serum specificity. CTSP-1 exhibited an intense staining in testis, and the staining was restricted to germ cells. Along the germ cells, an intense staining was detected in both the nucleus and the cytoplasm of the cells in later stages of spermatogenesis, such as spermatocytes and spermatids (Fig. 5B). It should be noted, however, that spermatogonia were mostly negative. A similar expression pattern has been observed for other non-X CT antigens (CT antigens mapped outside chromosome X; ref. 3).

CTSP-1 protein expression was then analyzed in prostate tissues. Prostate tumor samples, some of them expressing CTSP-1 mRNA, and normal prostate samples, were analyzed by C-immunohistochemistry. In the CTSP-1 mRNA-positive prostate tumors, staining was restricted to the tumor cells (Fig. 5C). No reactivity was observed in the CTSP-1 mRNA-negative prostate tumors, as well as in normal prostate samples, confirming the specificity of the anti-CTSP-1 antibody (Fig. 5B). However, because of the high similarity observed among all family members, we cannot rule out that the Western blot and immunohistochemistry results represent the accumulated expression of all CTSP members.

**Humoral Response Against CTSP-1 in Cancer Patients.** To evaluate the presence of antibodies against the CTSP proteins in plasma from cancer patients, we established an immunoblotting assay by using recombinant his-tagged CTSP-1 ( $\approx 32$  kDa). Another his-tagged unrelated recombinant protein, purified under the same conditions, served as internal negative control.

A total of 141 plasma samples from cancer patients with different histological types of tumors were used in the immunoblotting assay (Table 2, Fig. 6). Among the 141 plasmas analyzed, 14 (10.0%) were reactive to the CTSP-1 recombinant protein. Plasma samples from 50 healthy donors were used as negative controls. Reactivity against CTSP-1 recombinant protein was detected in only one sample from this control group (data not shown). Antibodies against CT antigens have been detected in patients with autoimmune disorders such as lupus

**Table 2. Frequency of anti-CTSP-1 antibodies in plasma samples from cancer patients**

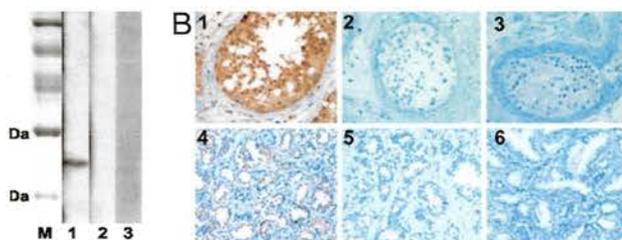
Tumor	Antibody response (%)	
Breast	3/18	(16.6)
Colon	0/20	(0.0)
Esophagus	0/4	(0.0)
Lung	1/13	(8.0)
Melanoma	1/22	(4.5)
Prostate	5/24	(20.8)
Stomach	1/8	(12.5)
Thyroid	2/10	(20.0)
Uterus	1/22	(4.5)
Total	14/141	(10.0)

erythematosus (23) and vitiligo (24) and in men subjected to vasectomy (25). In addition, we cannot exclude the possibility that we have identified a not yet diagnosed cancer patient within the healthy control group.

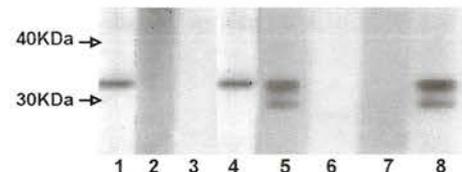
The highest percentages of antibody response were observed in patients with prostate (20.8%), thyroid (20.0%), and breast (16.6%) tumors (Table 2). With the exception of NY-ESO-1, humoral immune responses to CT antigens are relatively rare, occurring in <10% of the cancer patients (26–29).

For 121 of the 141 patients included in this study, a correlation between CTSP-1 mRNA expression and the presence of a humoral immune response could be analyzed because both RNA and plasma samples were available (Table 3, which is published as supporting information on the PNAS web site). A significant percentage (15.5%) of patients bearing CTSP-1-expressing tumors developed a humoral immune response against the protein. However, antibodies against CTSP-1 were also detected in 6.3% of patients with CTSP-1-negative tumors. A possible explanation for this apparent discrepancy would be the presence of an immune response against other CTSP family members. The expression of CTSP-2 and -4 in these CTSP-1-negative tumors was then assayed by RT-PCR. Of the CTSP-1-negative tumors analyzed, only one showed expression of CTSP-2, and none of them showed expression of CTSP-4. We have also analyzed CTSP-2 and -4 mRNA expression in 84 of the 121 samples from which the corresponding plasma was available. The small percentage of positive expression of CTSP-2 (22.6%) and CTSP-4 (0%) in these samples suggests that the immune response observed in these patients is likely to be directed to CTSP-1. Alternative explanations for the discrepancy between CTSP-1 expression and antibody response would be the presence of tumor heterogeneity not represented in the small tissue fragment analyzed by RT-PCR; the presence of undetected metastases expressing CTSP-1 or the clearance of CTSP-1-positive tumor cells by the immune system.

**Concluding Remarks.** The identification of genes that are selectively expressed in tumors and code for proteins inducing a specific



**Fig. 5.** CTSP-1 protein expression in normal testis and in prostate tissue. (A) CTSP-1 protein was detected by Western blot in protein extract from normal testis by using a CTSP-1 polyclonal antibody (1). Preimmune serum (2) and CTSP-1 polyclonal antibody depleted of anti-CTSP-1 Igs (3) were used as negative controls. Molecular weight markers are indicated (M). (B) Immunohistochemistry staining of CTSP-1 protein using the anti-CTSP-1 polyclonal antibody in normal testis (1), CTSP-1 mRNA-positive prostate tumor (4), normal prostate tissue (5), and CTSP-1 mRNA-negative prostate tumor (6). Preimmune sera (2) and CTSP-1 polyclonal antibody depleted of anti-CTSP-1 Igs (3) were used as negative controls.



**Fig. 6.** Detection of antibodies against CTSP-1 recombinant protein in plasma samples from cancer patients. A Western blot using CTSP-1 recombinant protein and plasma samples from prostate cancer patients (lanes 1–7) is shown. An anti-HisTag antibody was used as positive control (lane 8). Molecular weight markers are indicated.

nune response in cancer patients is a prerequisite for the development of efficient immunotherapeutic approaches to cancer. Antigenes coded by these antigens could be used to develop polypeptide cancer vaccines, overcoming obstacles associated with tumor heterogeneity and immune escape and increasing the number of cancer patients eligible for vaccination (30, 31).

We have identified a CT antigen, which is frequently expressed in different types of tumors and is capable of eliciting a humoral immune response in cancer patients. Because antibodies detected against CTSP-1 are of the IgG class, the activation of CD4<sup>+</sup> T cell response to CTSP-1 could also be present in these patients. Further experiments are ongoing to test whether the CTSP-1 protein is able to induce CD4<sup>+</sup> and CD8<sup>+</sup> T cell responses in cancer patients.

## Materials and Methods

**Cloning of CTSP-1 Full-Length Sequence. cDNA sequencing.** C21ORF99 sequence was obtained by sequencing the full insert cDNA clones from which the ESTs aligning to HC21 were derived. cDNA clones (IMAGE 1461135 and 2909444) were obtained from Research Genetics (Invitrogen, Carlsbad, CA) and were sequenced directly by using the vector's primers.

**PCR.** NY-BR-1 and NY-BR-1.1 sequences were aligned to C21 genomic sequence by using the BLASTN program. Primers for RT-PCR were designed based on the HC21 genomic sequence in regions that showed similarity to NY-BR-1 and NY-BR-1.1 and that were not covered by the C21ORF99 sequence. RT-PCRs were carried out in 25  $\mu$ l containing 1  $\mu$ l of total testis cDNA, 1 $\times$  TaqDNA polymerase buffer (Invitrogen), 0.1 mM dNTP, 2 mM MgCl<sub>2</sub>, 1 unit of TaqDNA polymerase (Invitrogen), and 6 pmol of the following primers: SP-F1 (5'-CTGAAAGCTTGGTGGAAAG-3') and SP-R1 (5'-GTTCTTCTTCCAAACTTC-3') or CTSP-F2 (5'-AAGACTGAATGAGTGGCAG-3') and CTSP-R2 (5'-GATTCAAATTACTTCTTACAG-3'). Amplification conditions were: initial denaturation for 4 min at 94°C followed by cycles of 45 sec at 94°C, 45 sec at 58°C, and 1 min at 72°C. PCR products were analyzed on 8% silver-stained polyacrylamide gels and cloned for sequencing.

**5'-3'-RACE** was performed on normal testis poly(A)<sup>+</sup> RNA using the Marathon cDNA Amplification Kit (Clontech, Mountain View, CA). Amplification reactions were performed in 25  $\mu$ l by using 5  $\mu$ l of cDNA, 0.2 mM dNTPs, 0.2  $\mu$ M SP-1-specific primer (SP-RACE, 5'-CCATGGCTCACACGTAATCTCATCAC-3'), 0.2  $\mu$ M adaptor primers, 1 unit of Advantage TaqDNA polymerase. PCR conditions were: 1 min at 94°C followed by 5 cycles of 5 sec at 94°C and 4 min at 70°C, 5 cycles of 5 sec at 94°C and 4 min at 68°C, and 25 cycles of 5 sec at 94°C, 30 sec at 65°C with a final extension step of 4 min at 68°C. RT-PCR was carried out by using 1  $\mu$ l of the first reaction product and the following primers: SP-RACE3N (5'-CAAAAAGTTAGAAGTGAAGCAACTTGAG-3') and nested adaptor primers. PCR products were cloned and sequenced as described above.

**Computational Analysis.** The CTSP-1 consensus sequence was obtained by assembling the C21ORF99 sequence the two RT-PCR fragments and 3'-RACE fragments using Phred/Phrap/Consed (32). Repetitive elements along the CTSP-1 sequence were identified with RepeatMasker (A. F. A. Smit, R. Hubley, P. Green, RepeatMasker, <http://repeatmasker.org>). CTSP-1 family members and partial copies were identified by aligning SP-1 consensus sequence against the human genome (Mar 2006) using BLAT (33). Consensus sequences including all predicted exons were translated in all six reading frames and active proteins were aligned by using ClustalW (34). Protein BLAST searches were performed with PROSITE (35) and Pfam databases (36).

**Expression Analysis Samples.** Human tumor cell lines A172 and T98G (glioblastoma); FaDu (squamous cell carcinoma); SW480 (colorectal adenocarcinoma); Skmel-28 and A2058 (malignant melanoma); DU145 and PC3 (prostate carcinoma); HeLa and CasKi (cervix adenocarcinoma); MDA-MB-231, MCF-7, and MDA-MB-436 (breast ductal carcinoma); IM9 (B transformed lymphoblasts); HL60 and K562 (lymphocytes); Hep2 (hepatocarcinoma); H1155 and H358 (lung carcinoma); SCABER (urinary bladder carcinoma); and SAOS-2 (osteosarcoma) were obtained from American Type Culture Collection (Manassas, VA). Tumor samples were collected from patients treated at Hospital A. C. Camargo. All samples were collected after explicit informed consent and with local ethical committee approval. Total RNA derived from 21 normal human tissues (testis, lung, prostate, small intestine, breast, brain, heart, uterus, bone marrow, placenta, colon, fetal brain, liver, fetal liver, thymus, salivary gland, spinal cord, kidney, spleen, skeletal muscle, and adrenal gland) was purchased from Clontech (Mountain View, CA).

**RNA extraction, cDNA synthesis, and RT-PCR.** Total RNA was extracted by the conventional CsCl-guanidine thiocyanate method (37). RNA samples were checked for integrity by agarose gel electrophoresis, and 2  $\mu$ g was used for cDNA synthesis. Reverse transcription was performed with DNA-free RNA by using SUPERScript II Reverse Transcriptase (Invitrogen) and oligo(dT). Primers used for expression analysis of CTSP-1, -2, -3, and -4 were CT1F (5'-GCTGTCCATTATGCTGTTAAC-3'), CT1R (5'-TTTTGAGAATTTTAGATATC-3'), CT2F (5'-CCTGTGGTGCAGACATCG-3'), CT2R (5'-ATTCCAAAA GTTGTGTGATGAAC-3'), CT3F (5'-CTGTCCATTATGCTGTTTATG-3'), CT3R (5'-AATTTTATAGATATCTTTT GTTTG-3'), CT4F (5'-TGCTGATCCAAATATTGTAGG-3'), and CT4R (5'-CATTTAAACTTATCAACTGCAA-3'). PCR fragments were analyzed either on agarose gels followed by Southern blot analysis or on 8% silver-stained polyacrylamide gels. PCR fragments were cloned and sequenced as mentioned above to confirm their specificity.

**5-aza 2'-deoxycytidine treatment.** MCF-7 breast tumor cell line was treated with 30  $\mu$ M 5-Aza 2'-deoxycytidine for 48 h. After treatment, the cells were harvested and used for RNA extraction and cDNA synthesis as described above. RT-PCR amplifications were carried out by using the same CTSP-1 primers as for gene expression analysis.

**CTSP-1 Protein Detection. CTSP-1 recombinant protein.** CTSP-1 longest ORF (202 amino acids) was amplified from normal testis cDNA by using the following primers: CTSP1RecF (5'-CGG-GATCCATGAAGA AGACGACAATG-3') and CTSP-1 RecR (5'-CGGAATTCCTATTCGTCAGGTGTTCT-3'). The PCR product was digested with BamHI and EcoRI and cloned into the expression vector pET28a (Stratagene, La Jolla, CA). After sequencing, the recombinant plasmid pET28a/CTSP-1 was transformed into *Escherichia coli* BL-21 A494. After induction with 1 mM isopropyl  $\beta$ -D-thiogalactoside at 37°C for 4 h, the CTSP-1 recombinant protein fused with a 6His-tag ( $\approx$ 32 kDa) was purified by Ni<sup>2+</sup> affinity chromatography by using the NiNTA agarose resin (Invitrogen). The purified protein was analyzed by Western blot by using an anti-His-tag monoclonal antibody (Amersham Biosciences, Piscataway, NJ) to confirm the purification specificity.

**CTSP-1 polyclonal antibody.** C57 mice (3-month-old females) were immunized s.c. with 25  $\mu$ g of CTSP-1 recombinant protein by using Freund's Complete Adjuvant (Sigma, St. Louis, MO) and Mg(OH)<sub>2</sub>. Three weeks later, the mice were boosted with an s.c. injection of 25  $\mu$ g of the recombinant protein by using Freund's Incomplete Adjuvant, and a final boost was repeated 3 weeks later. Two weeks after the final boost, the sera were collected and analyzed by ELISA to determine antibody titer. Mouse serum raised against CTSP-1 but depleted of anti-CTSP-1 Ig was used

a control of serum specificity. For depletion, 100  $\mu\text{g}$  of the  $\gamma$ -tagged CTSP-1 recombinant protein diluted in PBS was nubilized on Ni-NTA agarose beads (Qiagen, Valencia, CA) 3 h at room temperature. Beads were then incubated for 16 h at 4°C with mouse serum raised against CTSP-1. Depleted serum was recovered from beads by centrifugation at 500  $\times g$  for 10 min and used in Western blot (1:1,000 dilution) and IHC (1:10 dilution) experiments.

**Western blot.** Normal testis tissue was homogenized in sample buffer (240 mM Tris, pH 6.8/0.8% SDS/200 mM 2-mercaptoethanol/40% glycerol/0.2% bromophenol blue). A total of 100  $\mu\text{g}$  of the lysate was loaded on 12% SDS/PAGE gels and after electrophoresis transferred to Hybond-P PVDF membranes (Amersham Biosciences). After blocking with PBS solution containing 5% low-fat milk, membranes were incubated with sera at a 1:1,000 dilution for 1 h at room temperature. Anti-CTSP-1 antibodies were detected by rabbit anti-mouse IgG HRP conjugate (Amersham Biosciences) and visualized with ECL Western Blotting Detection Reagents (Amersham Biosciences).

**Immunohistochemistry.** Normal testis and prostate specimens were fixed with buffered formalin and embedded in paraffin. Briefly, sections were heated at 60°C for 20 min, deparaffinized in xylene, followed by a graded series of alcohols (100–75%), and rehydrated in water. Slides were placed three times in 3% hydrogen peroxide for 5 min and washed in water rinses for 5 min before incubation for 24 h with the primary antibody diluted 1:100. Slides were washed with PBS and incubated with biotinylated goat anti-rabbit IgG for 20 min and then with the streptavidin–biotin peroxidase LSAB kit (Pierce and Warriner, Carpinteria, CA). Immunostaining was performed by incubating slides in diaminobenzidine (Dako) solution containing 1  $\mu\text{g}$  of chromogen for 50  $\mu\text{l}$  of buffer substrate during 5 min. After color development, slides were washed, dehydrated with

alcohol and xylene, and mounted with coverslips by using a permanent mounting medium.

**Antibody Response in Cancer Patients.** Plasmas were obtained from patients treated at Hospital A. C. Camargo. All samples were collected after explicit informed consent and with local ethical committee approval. In addition, plasma samples from 50 healthy individuals were collected from blood donors at the Hospital A. C. Camargo Blood Center. Antibodies against CTSP-1 recombinant protein were detected by Western blot. Five hundred nanograms of purified CTSP-1 recombinant protein ( $\approx 32$  kDa) were fractionated on 12% SDS/PAGE and transferred to Hybond-P PVDF membranes. After blocking with PBS solution containing 5% low-fat milk, membranes were incubated for 90 min at room temperature with plasma from healthy individuals or cancer patients at a 1:500 dilution. Plasma antibodies binding to CTSP-1 protein were detected by incubation with goat anti-human IgG HRP conjugate (Amersham Biosciences) and visualized with ECL Western Blotting Detection Reagents.

We thank Drs. Alexandre Barbuto, Marcello Barcinsky, Erney Camargo, and Natanja Kirschbaum-Slager for critically reading this manuscript and Dr. Gerd Ritter for helpful discussions and experimental support. Valéria Paixão, Anna Christina Salim, Maria Cristina Costa, Sueli Nonogaki, Carlos Nascimento, Miyuki da Silva, and Severino Ferreira provided outstanding technical assistance. We also thank Dr. Flávio Henrique Silva for assistance with protein expression and Dr. Mônica C. Poli (Blood Center Hospital A. C. Camargo) for providing the plasma samples from the blood donors. This work was supported by the Ludwig Institute for Cancer Research, São Paulo, Brazil, and by Conselho Nacional de Desenvolvimento Científico e Tecnológico Grant 506250/04-0. R.B.P. was sponsored by a fellowship from Cordenção de Aperfeiçoamento de Pessoal de Nível Superior; F.B., M.D.V., M.H.L. and W.K.M. were sponsored by fellowships from Fundação de Amparo à Pesquisa do Estado de São Paulo.

Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, Chen YT (2002) *Immunol Rev* 188:22–32.

Scanlan MJ, Simpson AJ, Old LJ (2004) *Cancer Immunol* 4:1.

Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ (2005) *Nat Rev Cancer* 5:615–625.

Zendman AJ, Ruitter DJ, Van Muijen GN (2003) *J Cell Physiol* 194:272–288.

Old LJ (2001) *Cancer Immunol* 1:1.

Bodey B (2002) *Exp Opin Biol Ther* 2:577–584.

Chen Q, Jackson H, Parente P, Luke T, Rizkalla M, Tai TY, Zhu HC, Mifsud NA, Dimopoulos N, Masterman, KA, et al. (2004) *Proc Natl Acad Sci USA* 101:9363–9368.

Jager E, Chen W, Jackson H, Parente P, Shackleton M, Hopkins W, Chen Q, Dimopoulos N, Luke T, Murphy R, et al. (2004) *Proc Natl Acad Sci USA* 101:10697–10702.

Jager E, Gnjjatic S, Nagata Y, Stockert E, Jager D, Karbach J, Neumann A, Rieckenberg J, Chen YT, Ritter G, et al. (2000) *Proc Natl Acad Sci USA* 97:12198–12203.

Marchand M, van Baren N, Weynants P, Brichard V, Dreno B, Tessier MH, Rankin Z, Parmiani G, Arienti F, Humblet Y, et al. (1999) *Int J Cancer* 80:219–230.

Reymond A, Camargo AA, Deutsch S, Stevenson BJ, Parmigiani RB, UCLA C, Bettoni F, Rossier C, Lyle R, Guipponi M, et al. (2002) *Genomics* 79:824–832.

Jager D, Stockert E, Gure AO, Scanlan MJ, Karbach J, Jager E, Knuth A, Old LJ, Chen YT (2001) *Cancer Res* 61:2055–2061.

Roy-Engel AM, El Sawy M, Farooq L, Odom GL, Perepelitsa-Belancio V, Bruch H, Jyeniran OO, Deininger PL (2005) *Cytogenet Genome Res* 110:365–371.

Aradhya S, Bardaro T, Galgoczy P, Yamagata T, Esposito T, Patlan H, Accodico A, Munnich A, Kenwick S, Platzer M, et al. (2001) *Hum Mol Genet* 10:2557–2567.

Jure AO, Wei JJ, Old LJ, Chen YT (2002) *Int J Cancer* 101:448–453.

Huang X, Li J, Lu L, Xu M, Xiao J, Yin L, Zhu H, Zhou Z, Sha J (2005) *J Androl* 26:189–196.

Jure AO, Tureci O, Sahin U, Tsang S, Scanlan MJ, Jager E, Knuth A, Freunschuh M, Old LJ, Chen YT (1997) *Int J Cancer* 72:965–971.

Lucas S, Brasseur F, Boon T (1999) *Cancer Res* 59:4100–4103.

Old M, Zhou J, Chen GL, Hall JM, Vescio RA, Berenson JR (1999) *Genomics* 9:161–167.

Weber J, Salgaller M, Samid D, Johnson B, Herlyn M, Lassam N, Treisman J, Rosenberg SA (1994) *Cancer Res* 54:1766–1771.

De Smet C, Lurquin C, Lethe B, Martelange V, Boon T (1999) *Mol Cell Biol* 19:7327–7335.

Coral S, Sigalotti L, Altomonte M, Engelsberg A, Colizzi F, Cattarossi I, Maraskovsky E, Jager E, Seliger B, Maio M (2002) *Clin Cancer Res* 8:2690–2695.

McCurdy DK, Tai LQ, Nguyen J, Wang Z, Yang HM, Udarn N, Naiem F, Concannon P, Gatti RA (1998) *Mol Genet Metab* 63:3–13.

Rocha IM, Oliveira LJ, De Castro LC, Araujo Pereira LI, Chaul A, Guerra JG, Silvestre MC, Batista KM, Pereira FA, et al. (2000) *Int J Dermatol* 39:840–843.

Lea IA, Adoyo P, O’Rand MG (1997) *Fertil Steril* 67:355–361.

Stockert E, Jager E, Chen YT, Scanlan MJ, Gout I, Karbach J, Arand M, Knuth A, Old LJ (1998) *J Exp Med* 187:1349–1354.

Yakirevich E, Sabo E, Lavie O, Mazareb S, Spagnoli GC, Resnick MB (2003) *Clin Cancer Res* 9:6453–6460.

Sugita Y, Wada H, Fujita S, Nakata T, Sato S, Noguchi Y, Jungbluth AA, Yamaguchi M, Chen YT, Stockert E, et al. (2004) *Cancer Res* 64:2199–2204.

Jager E, Stockert E, Zidianakis Z, Chen YT, Karbach J, Jager D, Arand M, Ritter G, Old LJ, Knuth A (1999) *Int J Cancer* 84:506–510.

Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD (2002) *Nat Immunol* 3:991–998.

Dunn GP, Old LJ, Schreiber RD (2004) *Annu Rev Immunol* 22:329–360.

Gordon D, Abajian C, Green P (1998) *Genome Res* 8:195–202.

Kent WJ (2002) *Genome Res* 12:656–664.

Higgins DG, Thompson JD, Gibson TJ (1996) *Methods Enzymol* 266:383–402.

Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P (2002) *Brief Bioinform* 3:265–274.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al. (2004) *Nucleic Acids Res* 32:D138–D141.

MacDonald RJ, Swift GH, Przybyla AE, Chirgwin JM (1987) *Methods Enzymol* 152:219–227.