Identification of the Transcription Factor Single-Minded Homologue 2 as a Potential Biomarker and Immunotherapy Target in Prostate Cancer

Mohamed S. Arredouani,1 Bin Lu,1 Manoj Bhasin,2 Miriam Eljanne,1 Wen Yue,1 Juan-Miguel Mosquera,4 Glenn J. Bubley,3 Vivian Li,1 Mark A. Rubin,4 Towia A. Libermann,2 and Martin G. Sanda1

Abstract

Purpose: Identification of novel biomarkers and immunotherapy targets for prostate cancer (PCa) is crucial to better diagnosis and therapy. We sought to identify novel PCa tumor-associated antigens (TAA) that are expressed in PCa, absent in nonprostate human tissue, and immunogenic for immune responses restricted by human HLA.

Experimental Design and Results: Using microarray analysis of normal and cancerous human prostate tissues, we identified 1,063 genes overexpressed in PCa. After validating 195 transcripts in publicly available array data sets, we interrogated expression of these TAAs in normal human tissues to identify genes that are not expressed at detectable levels in normal, nonprostate adult human tissue. We identified 23 PCa TAA candidates. Real-time PCR confirmed that 15 of these genes were overexpressed in PCa (P < 0.05 for each). The most frequently overexpressed gene, single-minded homologue 2 (SIM2), was selected for further evaluation as a potential target for immunotherapy. ELISA assay revealed that a fraction of PCa patients exhibited immune responsiveness to SIM2 as evidenced by the presence of autoantibodies to SIM2 in their sera. We next showed binding of putative HLA-A2.1–restricted SIM2 epitopes to human A2.1, and immunization of transgenic HLA-A2.1 mice showed induction of SIM2-specific CTL responses in vivo.

Conclusions: Our findings that SIM2 is selectively expressed in PCa, that human HLA-A2.1–restricted SIM2 epitopes induce specific T cells in vivo, and that anti-SIM2 antibodies are detectable in PCa patients’ sera implicate SIM2 as a PCa-associated antigen that is a suitable potential target for PCa immunotherapy. (Clin Cancer Res 2009;15(18):5794–802)

Gene expression profiling of prostate cancer (PCa) has proven effective in identifying genes and molecular pathways associated with PCa. Profiling of RNA transcripts has been widely used to dissect molecular aspects of tumor cell biology as well as to project disease outcome that can be of high prognostic value (1–5). For example, the determination by several such array studies that ERG is commonly overexpressed in PCa led to identification of novel gene arrangements between TMPRSS2 and ETS transcription factors in PCa (6).

Identifying PCa-associated genes (those with higher levels in PCa than benign prostate) that are concurrently not expressed at abundant levels in normal human adult extraprostatic tissues would potentially identify PCa tumor-associated antigens (TAA) with greater specificity as therapy targets than that of cancer-associated genes identified without consideration of their extraprostatic expression patterns. However, prior genome-wide expression array studies to identify genes that are overexpressed in PCa have usually focused on interrogating...
we sought to identify novel targets for prostate cancer (PCa) detection and therapy. Toward this objective, we used fresh-frozen prostatectomy specimens to generate new PCa gene expression arrays and interrogated the expressed gene profile against gene expression of normal human adult tissue in silico to identify genes expressed in PCa but not in nonprostatic normal human tissues. This novel strategy identified 15 genes that are abundant in PCa and not in other adult human male tissue. Proof of principle that these are rational targets for PCa detection or therapy was shown for the transcription factor single-minded homologue 2 (SIM2), for which we showed that some PCa patients have intrinsic immune response as evidenced by autoantibodies to SIM2 in patient sera and that human HLA-A2.1–restricted, cytotoxic T-cell responses can be induced against SIM2 epitopes in vivo in HLA-A2.1 transgenic mice. Our findings identify peptide epitopes of SIM2 that may serve as PCa immunotherapy targets in future clinical trials.

relative levels of gene expression in cancerous prostate tissue compared with normal prostate tissue, and the relationship of prostatic gene expression to expression of such genes in adult tissues outside of the prostate has previously only been explored with genes expressed in normal prostate and not PCa (7).

One approach for translating newly discovered TAAs to a new direction for cancer therapy is to interrogate newly discovered TAA sequences for immunogenic peptide sequences that are predicted to bind human class I MHC and that are therefore putative targets for T-cell–mediated immunotherapy (8). However, such a strategy for identifying putative targets for PCa immunotherapy has not yet been linked directly to the interrogation of prostate TAAs discovered through concurrent interrogation of cancer and normal human tissue expression arrays. Instead, recent clinical trials of PCa immunotherapy have targeted PCa TAAs that had been identified before the era of genome-wide gene expression profiling. Limitations of this “first generation” of prostate TAA targets for immunotherapy have included limited cancer specificity of the target [as in prostate-specific antigen (PSA)] or limited tissue specificity [as with prostate-specific membrane antigen (PSMA) or prostate stem cell antigen (PSCA)], but despite these limitations, reduction of tumor activity has been observed in several PCa immunotherapy studies, and modest survival benefit was noted in two such trials (9, 10). The targeting of immunogenic peptides in PCa TAA identified from genome-wide expression profiling is an untested but promising direction for improving on the early foundations of PCa immunotherapy.

We sought to identify such immunogenic peptide targets (for immunotherapy) first by interrogating new PCa and normal prostate expression arrays against existing human expression arrays to identify PCa-specific TAA, then by ascertaining the immunogenicity of the TAA target through detecting autoantibody responses in PCa patients, next by evaluating binding to human HLA-A2.1 of potentially immunogenic peptide sequences from the lead TAA, and finally by ascertaining the ability of these epitopes to induce cellular immune responses in HLA-A2.1 transgenic mice.

Materials and Methods

Animals

The HHD mice were received from Dr. Francois Lemonnier (Unité d’Immunité Cellulaire Antivirale, Institut Pasteur, Paris, France). These mice are β2 m–/–, Dα2–/– double knockout and express an HLA-A*0201 monochain composed of a chimeric heavy chain (α1 and α2 domains of HLA-A*0201 allele and the α3 and intracellular domains of Δα1 allele) linked by its NH2 terminus to the COOH terminus of the human β2 m by a 15–amino acid peptide arm (11). Mice were housed in pathogen-free conditions, and all experimental procedures involving animals were approved by the Institutional Animal Care and Use Committee at Beth Israel Deaconess Medical Center.

Cell lines

The human TAP-deficient T2 cell line was purchased from the American Type Culture Collection and cultured as per the American Type Culture Collection’s protocol.

Peptides

All peptides used in this study were purchased from the Biopolymers Laboratory at Harvard Medical School. Peptides were >90% pure and high-performance liquid chromatography tested. Peptides were dissolved in DMSO and stored in aliquots at -20°C until use.

Sample selection and RNA purification.

Radical prostatectomy tissue samples were obtained from the Hershley Foundation Prostate Cancer Serum and Tumor Bank at our institution. Morphologic diagnosis was done by a pathologist. OCT blocks containing >30% of PCa tissue (with Gleason score of 6 or 7) were selected for RNA purification. A biopsy punch was used to select the PCa tissues from the OCT sample blocks. Benign or PCa tissues were homogenized using a TissueLyser (Qiagen) at 28 Hz for 5 min. Total RNA was isolated using Trizol reagent. RNA was quantified by NanoDrop ND-1000 spectrophotometer, and quality was evaluated with Agilent RNA 6000 NanoChip and the 2100 Bioanalyzer, with 28S/18S ratios and RIN determined by 2100 Expert software.

Gene expression microarrays and analysis

Total RNA (250 ng) was amplified using Ambion MessageAmp II mRNA Amplification kit. Biotin-LITP was incorporated during the overnight in vitro transcription step according to the manufacturer’s protocol. Gene expression was assessed using Affymetrix GeneChip U133 array (Plus 2.0 chip) consisting of >52,000 transcripts from whole human genome transcripts. cRNA (15 μg) was fragmented and hybridized to arrays according to the manufacturer’s protocols. The quality of scanned array images was determined based on background values, percent present calls, scaling factors, and 3’-5’ ratio of β-actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) using the BioConductor R packages. The signal value for each transcript was summarized using perfect matched only–based signal modeling algorithm described in dchip. The perfect matched only–based modeling algorithm yields less number of false positives compared with the perfect matched-mismatched model. In this way, the signal value corresponds to the absolute level of expression of a transcript (12). These normalized and modeled signal values for each transcript were used for further high-level bioinformatic analysis. During the calculation of model-based expression signal values, array and probe outliers are interrogated and image spikes are treated as signal outliers. The outlier detection was carried out using dchip outlier detection algorithm. A chip is considered as an outlier if the probe, single, or array outlier percentage exceeds a default threshold of 5%. When comparing two groups of samples to identify genes enriched in a given phenotype, if 90% lower confidence bound (LCB) of the fold change
between the two groups was above 1.2, the corresponding gene was considered to be differentially expressed. LCB is a stringent estimate of fold change and is the better ranking statistic. It has been suggested that a criterion of selecting genes that have a LCB above 1.2 most likely corresponds to genes with an "actual" fold change of at least 2 in gene expression (13).

**Analysis of the Stanford prostate data.** The raw gene expression data from 62 PCa and 41 normal prostate published by Lapointe et al. (3) were obtained from the BRB-Array archived data sets. The preprocessing data were normalized using the Z transformation (M1, ONCOMINE: a cancer microarray database and integrated data-mining platform). The differentially expressed genes were identified based on fold change (≥0.5) and Q value of <0.05. The analysis identified 510 genes that were differentially expressed genes were identified based on fold change (≥0.5) and Q value of <0.05) in 7 PCa compared with 8 normal prostate specimens (9). Statistical analysis showed that these genes are overexpressed in PCa over normal prostate tissue.

**Biomarker analysis.** To prioritize the biomarker and immunotherapy targets, we need to identify the genes that are not ubiquitously expressed in all normal tissues. The gene expression data for the various human normal tissues were obtained from gene expression atlas of the Genomics Institute of the Novartis Research Foundation. Using this database, MAS5 normalized expression data along with present (P) absent (A), and marginal (M) calls for each transcript were obtained. Based on present and absent calls for each transcript, a priority value is calculated (Eq. A). The gene that is absent in all tissues was given highest priority (rank 1), and the gene that is present in all of tissues is given a least priority. In other words, number of present and absent calls in different tissues was used to find out the genes having restrictive expression level. To further extend the list of genes, we have also obtained a list of prostate-specific genes by analyzing the Novartis gene expression data (11). The genes that are annotated absent based on MAS5 calls in all the normal tissues except prostate were considered as prostate-specific genes.

**Quantitative real-time PCR**
Validation of overexpression of the selected prostate TAAs in seven PCa tumors and eight control tissue specimens was done by quantitative real-time PCR (qRT-PCR). High-quality RNA samples (50 ng; RIN > 6.0 by Agilent 2100 Bioanalyzer) were reverse transcribed to first-strand cDNA and 1 μL cDNA was used for each well RT-PCR. Samples were done in triplicates. SYBR Green PCR Master Mix (Applied Biosystems) was used for two-step RT-PCR analysis on SA Biosciences 7900HT Prism instruments. PCR primer sequences for targeted genes are shown in Table 1. The sequences for GAPDH are 5'-TGGCATGGACCTTGTAGAG-3' (forward) and 5'-GGCATGGACCTTGTAGAG-3' (reverse). Expression value of the targeted gene in a given sample was normalized to the corresponding expression of GAPDH. The 2−ΔΔCt method was used to calculate relative expression of the targeted genes (14).

**Table 1. Validation of 15 novel prostate TAAs by qRT-PCR**

<table>
<thead>
<tr>
<th>Gene accession number</th>
<th>Gene name</th>
<th>Primer sequences (5′–3′)</th>
<th>P (t test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMACR (NM_014324)</td>
<td>α-Methylacyl-CoA racemase</td>
<td>F: ttatgtgtcattggtgatt</td>
<td>0.0196</td>
</tr>
<tr>
<td>BICD1 (NM_001714)</td>
<td>Bicaudal D homologue 1</td>
<td>R: tggggtctcattgtcaccac</td>
<td>0.033</td>
</tr>
<tr>
<td>C10orf137 (NM_015608)</td>
<td>Chromosome 10 open reading frame 137</td>
<td>F: aggccagcttgctgatta</td>
<td>0.013</td>
</tr>
<tr>
<td>CDC2L6 (NM_015076)</td>
<td>Cell division cycle 2–like 6</td>
<td>R: tttacagggagttgaggctg</td>
<td>0.0065</td>
</tr>
<tr>
<td>ICA1 (NM_022307)</td>
<td>Islet cell autoantigen 1</td>
<td>F: aagaacagacccagaccaac</td>
<td>0.0065</td>
</tr>
<tr>
<td>KIAA1661 (AB051448)</td>
<td>KIAA1661 protein</td>
<td>R: acagctctcagtcagttgctc</td>
<td>0.03</td>
</tr>
<tr>
<td>MAP7 (NM_003980)</td>
<td>Microtubule-associated protein 7</td>
<td>F: ttggacgaagacccgataaag</td>
<td>0.005</td>
</tr>
<tr>
<td>MY06 (NM_004999)</td>
<td>Myosin VI</td>
<td>R: ggtggtgacccaagagaac</td>
<td>0.003</td>
</tr>
<tr>
<td>OR51E2 (NM_030774)</td>
<td>Olfactory receptor, family S1, subfamily E, member 2</td>
<td>R: tggcactgctctcattgagc</td>
<td>0.0467</td>
</tr>
<tr>
<td>PAICS (NM_006452)</td>
<td>Phosphoribosylaminomimidazole carboxylase</td>
<td>F: aggcagcagcagtagagtgt</td>
<td>0.003</td>
</tr>
<tr>
<td>PCSK6 (NM_002570)</td>
<td>Proprotein convertase subtilisin/kexin type 6</td>
<td>R: cagctctcagtcagttgctc</td>
<td>0.03</td>
</tr>
<tr>
<td>PVT1 (NR_003367)</td>
<td>PVT1 oncogene (nonprotein coding)</td>
<td>F: gccaactctgtagactgatt</td>
<td>0.0045</td>
</tr>
<tr>
<td>RGS10 (NM_002925)</td>
<td>Regulator of G protein signaling 10</td>
<td>R: gggttctctcactgctcagc</td>
<td>0.01</td>
</tr>
<tr>
<td>SGEF (NM_015595)</td>
<td>Src homology 3 domain-containing guanine nucleotide exchange factor</td>
<td>R: ggggcatagatgaataagta</td>
<td>0.02</td>
</tr>
<tr>
<td>SIM2 (NM_005069)</td>
<td>Single-minded homologue 2</td>
<td>R: cctttctctgactctcagc</td>
<td>0.0082</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: aggtgtggctcagcttgatt</td>
<td>0.0045</td>
</tr>
</tbody>
</table>

NOTE: qRT-PCR validation of mRNA expression levels of individual genes (AMACR, BICD1, C10orf137, CDC2L6, ICA1, KIAA1661, MAP7, MYO6, ORS1E2, PAICS, PCSK6, PVT1, RGS10, SGEF, and SIM2) was done using the Taqman gene expression assay. From a total of 23 genes tested, only the 15 genes that were significantly overexpressed (P < 0.05) in 7 PCa compared with 8 normal prostate specimens are shown. Statistical analysis showed that these genes are overexpressed in PCa over normal prostate tissue.

---

**5** http://symatlas.gnf.org

**Clinc Cancer Res 2009;15(18) September 15, 2009 5796 www.aacrjournals.org**

Downloaded from clincancerres.aacrjournals.org on April 15, 2017. © 2009 American Association for Cancer Research.
Detection of autoantibodies to single-minded homologue 2 in sera from PCa patients

Patients and sera. Serum samples were collected at Harvard University and University of Michigan patient accrual sites. All patients were over the age of 40 y and were seen at the clinic because of PSA value exceeding 2.5 ng/mL, abnormal digital rectal exam, rising PSA, or lower PSA with risk factors such as family history. The study also includes men who have had previous biopsies that have not been positive for cancer. After enrollment and blood collection, all patients get a prostate biopsy to determine the presence or absence of cancer.

Cloning and in vitro cell-free expression of single-minded homologue 2. Full-length human single-minded homologue 2 (SIM2) cDNA in a pCR-BLUNT2-topo plasmid was amplified using two rounds of PCR. The PCR product was cloned into pDONR plasmid to produce entry clones of each cDNA. Entry clones (130 ng) were used to produce expression clones using pCITE-glutathione S-transferase (GST) expression vector (130 ng), LR clonase II enzyme mix (2 μL), and TE with a total volume of 10 μL. The SIM2 protein was produced as GST recombinant proteins with GST at the COOH terminus. A GST control vector from which only GST is expressed served as a negative control for serum antibody binding. The Kozak sequence was introduced into the original pDEST15 5′ of GST so that it can be used in the mammalian cell-free system.

Detection of serum autoanti-SIM2 antibodies by ELISA. GST prestained ELISA plates (GE Biosciences) were blocked overnight with 5% milk and 0.2% Tween 20. The SIM2 protein was expressed using rabbit reticulocyte lysate cell-free expression system (Promega) according to the manufacturer’s instructions. The expressed protein was transferred to the ELISA plate and bound overnight at 4°C. The plates were washed and incubated with serum diluted 1:300 in blocking buffer for 1 h followed by incubation for 1 h with horseradish peroxidase-linked anti-human antibodies. The substrate (100 μL; SuperSignal ELISA Femto Maximum Sensitivity Substrate, Pierce) was added to each well, and the luminescence signal was read using the Vector3 ELISA reader. Each serum sample was screened in duplicates. The plate also included a secondary antibody negative control and a GST control.

Prediction of HLA-A2.1–restricted epitopes from SIM2

To predict potential nonamer epitopes that bind HLA-A*0201, the most frequent haplotype in Caucasians, SIM2 protein sequence (SIM2_HUMAN, Q14190) was processed using BIMAS6 and SYFPEITHI7 as well as MHCPred8, RankPep9, NetMHC,10,11 ProPred-I,12,13 and MAPPP.13 Only epitopes that were predicted by most algorithms were selected for further testing.

Measurement of peptide/HLA-A*0201 binding and stability

MH C stabilization assay using T2 cells was used to assess binding of peptides to the HLA-A2.1 complex. Briefly, T2 cells were cultured for 6 h in serum-free Iscove’s modified Dulbecco’s medium (American Type Culture Collection) before the addition of candidate peptide at a concentration of 50 μg/mL. Cells were washed with PBS and saturated with RPMI 1640/10% FCS for 1 h at 37°C. A total of 2.5 × 105 splenocytes were seeded in each well in four replicates, and 5 × 104 peptide-loaded (10 μg peptide/mL, for 2 h at 37°C) spleenocytes pretreated with 50 μg/mL mitomycin C for 1 h were added to each well. Plates were incubated for 1 to 2 d at 37°C in 5% CO2, washed five times with PBS, and then incubated with 1 μg/mL of biotinylated rat anti-mouse IFN-γ mAb (R4-6A2; Mabtech) for 24 h at 4°C or room temperature for 2 h. The wells were washed and 100 μL of diluted alkaline phosphatase–conjugated streptavidin were added for 1 h at room temperature. Spots were developed by adding peroxidase substrates (5-bromo-4-indolyl phosphate and nitroblue tetrazolium) and counted using the ELR04 AID ELISPOT Reader System (Autoimmune Diagnostika GmbH).

Statistical analysis

Gene expression array data were analyzed as described in Materials and Methods. Group differences for gene expression (RT-PCR), autoantibody (ELISA), and IFN-γ (ELISPOT) data were analyzed using the Student’s t test. P values of ≤0.05 were considered significant.

Results

Identification of novel prostate TAAs using gene expression profiling. In an effort to identify novel putative PCA TAAs with expression specificity for PCA over normal prostate or normal nonprostate tissue, we did a genome-wide gene expression analysis of a PCA and normal prostate microarray generated in our laboratory; validated the candidate TAAs in an external, published PCA tissue array data set; and then excluded those with detectable expression in nonprostatic adult tissues (Fig. 1). First, we used the Affymetrix U133 array (Plus 2.0 chip) to evaluate gene expression in cancer and normal fresh-frozen prostate tissue specimens from our tissue repository. The class comparison analysis based on LCB(1.2) and mean difference in absolute intensity of >40 identified 1,063 genes overexpressed in PCAs compared with normal prostate. The heat map of top 100 genes is shown in Fig. 1A. Examples of the top 100 genes include AMACR, ERG, MMP26, THBS4, and FOXD1 (Supplementary Table S1). Next, we validated the 1,063 putative TAA and conducted a comprehensive
analysis of microarray data from a previously published data set, which included 41 normal and 62 neoplastic prostate tissues (3). We looked at the genes that are significantly overexpressed in PCa for their potential to be used as biomarkers or targets for immunotherapy. A list of 426 PCa up-regulated genes was obtained based on the fold change (>0.5) and false discovery rate value of <0.05 after preprocessing and normalizing data (Z transformation). Validation of genes that were overexpressed in PCa in our data set by comparison with the Stanford PCa array data set implicated 195 transcripts with concordant overexpression between the array data sets. To identify PCa TAA with the greatest specificity for PCa, we then sought to exclude, by in silico analysis, those genes that are detectable in nonprostate normal human adult male tissues. For this purpose, gene expression data for various human tissues were obtained from the two studies conducted by Su et al. (18) and Ge et al. (19), and genes that were annotated absent based on MAS5 calls in all the normal tissues except prostate were considered as prostate-specific genes. The comprehensive analysis led the identification of 26 transcripts that are overexpressed in the PCa and are highly tissue restricted (Fig. 1B). These transcripts correspond to 23 genes (listed in Supplementary Table S2) that include SIM2. The analysis also identified 17 more genes that are present in the prostate and absent in the rest of the normal tissues (Supplementary Table S3).

We then did qRT-PCR targeting each of the 23 candidate antigens and confirmed that 15 (AMACR, BICD1, C10orf137, CDC21c, ICA1, KIAA1661, MAP7, MYO6, OR51E2, PAICS, PCSK6, PVT1, RGS10, SGEF, and SIM2) were overexpressed in PCa (Table 1; Supplementary Fig. S1). Frequency of overexpression in PCa for these antigens ranged from 57% to 86%.
SIM2 as a Prostate Tumor–Associated Antigen

In a previous study, we identified autoantibodies reactive with the prostate tumor-associated antigen SIM2 from healthy adult tissues, its overexpression in PCa, and its ability to induce humoral responses in PCa patients, we sought to identify potential HLA-A2.1–restricted, SIM2-derived epitopes that could be used as vaccines to generate SIM2-specific cytotoxic lymphocytes directed against prostate tumors.

To predict potential SIM2-derived, HLA-A*0201–binding epitopes, SIM2 full protein sequence was screened using the eight algorithms. Out of all possible nonamer motifs, 15 epitopes that had the highest cumulative prediction scores were selected for further evaluation in vitro and in vivo (Table 2).

After predictions were done, we ascertained epitope homology between mouse and human to ensure that we are testing tolerant motifs relevant in both hosts. A BLAST sequence analysis revealed that these five human immunogenic epitopes are 100% identical to their corresponding murine orthologs and are not present in any other known protein sequences in humans or mice, with the exception of SIM2(241), which is also present in SIM1.

Binding to HLA-A2.1 molecules was assessed using T2 assembly assay, which is based on the ability to stabilize MHC class I molecules from the T2 cell line by the addition of suitable peptides. This peptide-HLA binding screen revealed nine SIM2 peptides that were able to stabilize HLA-A2.1 molecules, resulting in increased detection of surface A2.1 molecules with a specific mAb (Fig. 3A). Peptide-HLA dissociation rate correlated with time and showed weak stabilizing epitopes (epitopes 84, 199, 237, and 430) and strong stabilizing epitopes (epitopes 87, 205, 241, and 244). However, epitopes with a high dissociation rate (weak stabilizers) still showed a slight binding that was above the nonbinding control epitope even after 8 hours of incubation.

These nine binding epitopes were then tested for their capacity to elicit in vivo CTL responses in transgenic HHD mice. Mice were immunized with a mixture of candidate epitopes and a

![Fig. 2. SIM2 elicits spontaneous humoral responses in PCa patients. Sera from nine PCa patients and five healthy donors were subjected to an in vitro, cell-free protein expression–based ELISA to detect autoantibodies to SIM2. SIM2 was expressed as a GST-tagged protein and an anti–GST-coated plate was used in the assay. Serum antibodies that bound to immobilized SIM2 were detected using a labeled anti-human antibody. In each assay, wells containing a GST-expressing vector were used as a negative control. Signals obtained from GST wells were subtracted from those obtained from wells that contained GST-SIM2. Three experiments were done, with triplicate wells for each serum sample per experiment. Statistics were done in triplicate, and antibody amounts were plotted as the difference of absorbance (OD) signals produced by GST-SIM2 and GST alone. Columns, mean; **, \( P < 0.01 \).](image)

---

Table 2. Identification of SIM2-derived HLA-A2.1–binding epitopes

<table>
<thead>
<tr>
<th>Epitope</th>
<th>Sequence</th>
<th>SYFPEITHI score</th>
<th>Binding (FI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIM2(25)</td>
<td>KLLPLPSAI</td>
<td>25</td>
<td>3.22</td>
</tr>
<tr>
<td>SIM2(84)</td>
<td>LLOTLDFGV</td>
<td>23</td>
<td>1.88</td>
</tr>
<tr>
<td>SIM2(87)</td>
<td>TLDGFVFV</td>
<td>27</td>
<td>5.43</td>
</tr>
<tr>
<td>SIM2(167)</td>
<td>VLAKRAGNL</td>
<td>23</td>
<td>0.82</td>
</tr>
<tr>
<td>SIM2(174)</td>
<td>GLTCGSKYKV</td>
<td>23</td>
<td>1.10</td>
</tr>
<tr>
<td>SIM2(199)</td>
<td>SLDSCYQ</td>
<td>23</td>
<td>4.39</td>
</tr>
<tr>
<td>SIM2(205)</td>
<td>YQVLQVAV</td>
<td>23</td>
<td>3.82</td>
</tr>
<tr>
<td>SIM2(237)</td>
<td>SLDLKFFL</td>
<td>27</td>
<td>1.70</td>
</tr>
<tr>
<td>SIM2(241)</td>
<td>KLFLDLSRV</td>
<td>23</td>
<td>2.00</td>
</tr>
<tr>
<td>SIM2(244)</td>
<td>FDLRSRTVE</td>
<td>27</td>
<td>4.75</td>
</tr>
<tr>
<td>SIM2(339)</td>
<td>ELQLSLEQV</td>
<td>21</td>
<td>1.28</td>
</tr>
<tr>
<td>SIM2(430)</td>
<td>LLTPSYSL</td>
<td>27</td>
<td>2.61</td>
</tr>
<tr>
<td>SIM2(527)</td>
<td>GSGLGVKGK</td>
<td>18</td>
<td>0.82</td>
</tr>
<tr>
<td>SIM2(530)</td>
<td>LVLGKVGGL</td>
<td>30</td>
<td>0.97</td>
</tr>
<tr>
<td>SIM2(558)</td>
<td>SRFQGTCPL</td>
<td>17</td>
<td>0.88</td>
</tr>
<tr>
<td>flu-M1</td>
<td>GILGFVFTL</td>
<td>30</td>
<td>3.17</td>
</tr>
<tr>
<td>NEG</td>
<td>IAGNSAYEY</td>
<td>9</td>
<td>1.00</td>
</tr>
</tbody>
</table>

NOTE: Prediction algorithms were used to predict SIM2-derived, HLA-A2.1–binding epitopes. Also shown are peptide scores predicted by the algorithm SYFPEITHI. Binding of predicted peptides to HLA-A2.1 was assessed using the assembly assay on T2 cells in vitro. Of the 15 peptides we tested, 9 showed binding ability compared with a nonbinding peptide (NEG).
helper MHC II–restricted peptide, and splenocytes were restimulated 12 days later in vitro with peptide-loaded APC and subjected to an IFN-γ ELISPOT assay (Fig. 3B). This provides evidence that tolerance to SIM2 is circumvented through immunization of mice to these epitopes because SIM2 (and SIM1) is also expressed in other nonprostatic tissues in mice. Interestingly, SIM2(25) and SIM2(199) were not immunogenic despite their ability to strongly bind to A2.1.

Discussion

Previous genome-wide expression profiling studies for PCA aimed at identifying genes that are overexpressed in prostate tumors regardless of their levels of gene expression in normal prostate and other adult tissues. The relationship of prostatic gene expression to expression of such genes in nonprostatic tissues has previously only been addressed in normal prostate and not PCA (7).

In the present study, we undertook a multistep strategy that combined gene expression profiling of malignant and benign human prostate tissues and in silico analysis of microarray data sets to identify novel prostate TAA. Our analysis focused on antigens that are overexpressed in PCA but are not or are weakly expressed in nonprostatic healthy adult male tissues. This innovative approach is undertaken to minimize the possibility of unwanted collateral autoimmune responses against nonprostate normal tissues and to optimize the possibility of attenuating autoantigen tolerance by prostatic manipulation (e.g., hormonal or ablative). This strategy has led to the identification of 23 potential TAA, among which 15 were validated by qRT-PCR. Whereas 4 of these TAAs [AMACR (28), MYO6 (29), OR51E2 (PSGR; ref. 30), and SIM2 (20)] have been previously reported to be associated with PCA, the remaining 11 (BICD1, C10orf137, CDC2L6, ICA1, KIAA1661, MAP7, PAICS, PCSK6, PVT1, RGS10, and SGEF) have not. These 11 novel TAAs represent a significant addition to the prostate TAA repertoire and warrant further investigation of their implication in PCA cancer biology.

Interestingly, our data revealed the presence of anti-SIM2 autoantibodies in sera from a fraction of PCA patients. It is well documented that PCA patients' immune system can mount antibody responses to prostate TAA (26) as well as to ubiquitous antigens such as the androgen receptor (31) and cellular proteins p90 and p62 (32). The presence of antibodies to TAA in PCA patients' sera is indicative of a humoral immune response against these TAA and results from a well-orchestrated response where an antigen-specific CD4 response is indispensable. This same CD4 response is a prerequisite for the immune system to mount a proper CTL response to a given antigen. In fact, previous studies have shown a close correlation between serum antibodies to TAA and both CD4 and CD8 T-cell responses in vivo (33, 34). Therefore, we assumed that a SIM2-derived, MHC I–restricted, peptide-based vaccine would lead to an optimal response in PCA patients.

SIM2 is a member of the basic helix-loop-helix per-Anti-Sim (bHLH-PAS) family of transcription factors (35). It is mainly known as a contributing factor to Down's syndrome (36). SIM2 expression persists through adulthood in muscle and kidney (37) where its function remains to be elucidated. A prior in silico approach using the Cancer Genome Anatomy Project database of the National Cancer Institute identified SIM2 as associated with colon cancer, pancreatic cancer, and PCA, whereas absent in the corresponding normal tissues (38). Both spliced

Fig. 3. SIM2 harbors HLA-A2.1–restricted immunogenic epitopes. A, the binding to and rate of dissociation of peptides from HLA-A2.1 were determined by monitoring the decrease in HLA-A2.1 expression over time after incubation with binder peptides. B, immunization of A2.1 transgenic HHD mice with the nine binding peptides revealed five immunogenic SIM2 peptides as shown by an IFN-γ ELISPOT assay. **, P < 0.01; ***, P < 0.001.
isoforms of SIM2 transcript, SIM2-long and SIM2-short, have been reported to be overexpressed in cancer (38). However, a biological role of SIM2 in cancer has not emerged yet, as it has been attributed suppressive and oncogenic properties that depend on the type of cancer (39, 40).

Vaccines designed to eradicate tumors by triggering immune responses against TAAs represent a tempting therapeutic modality. However, developing successful vaccines is hampered by the lack of highly specific tumor-derived antigens, immune tolerance, and undesired autoimmune responses.

Peptide-based cancer vaccines were among the first vaccines showing both protective and therapeutic efficacy in animal models and currently represent the majority of clinical trials of cancer immunotherapy (41, 42). This is attributed to the well-recognized requirement of T lymphocytes, especially CTLs, for the eradication of solid tumors as they represent the primary effector cells involved in tumor-specific immunity. Peptide-based vaccines offer considerable advantages over other vaccine formulations, namely, the absence of infectious material, the easy characterization and purification, the absence of risk of restored virulence or genetic integration, possibility of sequence modification, better storage, and lower cost (reviewed in ref. 41). In PCa, several TAA-derived, HLA-restricted, peptide-based vaccines have been tested clinically (reviewed in ref. 43). Hence, peptides derived from multiple TAAs were administered to hormone-refractory PCa patients and induced increased numbers of specific CD8 T cells (44). Similarly, promising clinical trials were done that tested a vaccine formulation based on dendritic cells loaded with peptides derived from PSMA (45, 46), from hTERT (47), or from a combination of two or more TAAs (PSA, PSMA, survivin, prostein, and trp-p8, PAP, and PSCA; refs. 48–50).

Because both SIM2-long and SIM2-short isoforms are expressed in PCa tumors, we processed both protein sequences using various algorithms to predict potential HLA-A2.1-binding nonamer epitopes. Interestingly, all epitopes that were predicted with high scores are common to both isoforms, suggesting that these epitopes could theoretically be used as targets on prostate tumors regardless of the type of isoform they express. We used a panel of 8 prediction algorithms that culminated in a selection of 15 candidate epitopes, among which 9 showed significant binding to A2.1 molecules in the T2 cell assay and 5 of these induced specific in vivo CD8 responses in A2.1 transgenic HHD mice as evidenced by their ability to trigger IFN-γ release by CD8 T cells on restimulation. Although prior studies have shown that SIM2 expression can be detected by RT-PCR in some normal tissues defined such as normal adult kidney, the level of SIM2 expression in such noncancer tissues is apparently low enough such that tolerance to SIM2 can be readily overcome via immunization with class I MHC–restricted SIM2 peptides as we have shown here and below the level of detection of microarray platforms that have been used to catalogue normal human tissue gene expression (18, 19). Accordingly, human trials would need to carefully monitor subjects for possible nephritis or other autoimmune responses. Whether the other 14 putative prostatic TAAs have lower levels of normal tissue expression than SIM2 will require validation assays other than the microarray platform results (18, 19) that we interrogated to identify genes having low or absent expression in nonprostate adult tissue.

Collectively, our data showing overexpression of SIM2 in malignant prostate tissue, combined to the identification of humoral responses to SIM2 in PCa patients’ sera and the ability of SIM2-derived peptides to induce HLA-A2.1–restricted cellular immune responses in humanized A2.1 transgenic mice, implicate SIM2 as a potential novel TAA target for PCa immunotherapy.

Disclosure of Potential Conflicts of Interest

M. Sanda received a commercial research grant from Beckman-Coulter.

References


Identification of the Transcription Factor Single-Minded Homologue 2 as a Potential Biomarker and Immunotherapy Target in Prostate Cancer

Mohamed S. Arredouani, Bin Lu, Manoj Bhasin, et al.