An Integrated Genome-Wide Approach to Discover Tumor-Specific Antigens as Potential Immunologic and Clinical Targets in Cancer

Qing-Wen Xu1, Wei Zhao1, Yue Wang6,9, Maureen A. Sartor11, Dong-Mei Han2, Jixin Deng10, Rakesh Ponnala8,9, Jiang-Ying Yang3, Guo-Qing Liao4, Yi-Mei Qu4, Lu Li5, Fang-Fang Liu6, Hong-Mei Zhao7, Yan-Hui Yin1, Wei-Feng Chen1,1, Yu Zhang1, and Xiao-Song Wang8,9

Abstract

Tumor-specific antigens (TSA) are central elements in the immune control of cancers. To systematically explore the TSA genome, we developed a computational technology called heterogeneous expression profile analysis (HEPA), which can identify genes relatively uniquely expressed in cancer cells in contrast to normal somatic tissues. Rating human genes by their HEPA score enriched for clinically useful TSA genes, nominating candidate targets whose tumor-specific expression was verified by reverse transcription PCR (RT-PCR). Coupled with HEPA, we designed a novel assay termed protein A/G-based reverse serological evaluation (PARSE) for quick detection of serum autoantibodies against an array of putative TSA genes. Remarkably, highly tumor-specific autoantibody responses against seven candidate targets were detected in 4% to 11% of patients, resulting in distinctive autoantibody signatures in lung and stomach cancers. Interrogation of a larger cohort of 149 patients and 123 healthy individuals validated the predictive value of the autoantibody signature for lung cancer. Together, our results establish an integrated technology to uncover a cancer-specific antigen genome offering a reservoir of novel immunologic and clinical targets. Cancer Res; 72(24); 6351–61. ©2012 AACR.

Introduction

Unlike microorganisms, whose molecular characteristics are distinct from human cells, cancer cells greatly resemble normal somatic cells, making them difficult to target. For this reason, a central aim of cancer research is to identify molecular targets distinguishing cancer cells from normal somatic cells. Tumor-specific antigens (TSA) are normal human gene products that are relatively uniquely expressed in cancer cells in contrast to normal somatic tissues, and thus often evoke specific antitumor immune responses. TSAs often have important functions in embryonic or germ line cells but are silenced in most somatic tissues. In contrast to “overexpression” in cancer, this “tumor-specific” expression pattern implies crucial clinical significance of these targets for cancer management.

Indeed, TSAs are pivotal targets in the management of human cancers. Assays detecting TSAs are commonly used in the clinic for the early detection or monitoring of human cancers, such as tests for prostate-specific antigen (PSA; ref. 1), α-fetoprotein (AFP; ref. 2), and CA19-9 (3). In recent years, autoantibody signatures against TSAs have been adopted as a new type of immunologic biomarker (4, 5). By combining the responses against multiple immunogenic TSAs, autoantibody signatures provide considerably higher sensitivity and specificity than a single biomarker, implying the high potential for early cancer detection (6–8). In the therapeutic perspective, substantial progresses are being made in the development of active and adoptive immunotherapy against TSAs for human cancers (9–14), and a prostatic acid phosphatase (or ACPP) vaccine has recently been approved by the U.S. Food and Drug Administration (FDA; ref. 15).

In this study, we designed a novel, integrated, computational, and experimental approach for the discovery and validation of tumor-specific antigens (TSA) as potential clinical targets in human cancers.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Q.-W. Xu and W. Zhao share first authorship.

Y. Zhang and X.-S. Wang share senior authorship.

†Deceased.

Corresponding Authors: Yu Zhang, Department of Immunology, Peking University Health Science Center, 38 Xue Yuan Road, Beijing 100191, China. Phone: 86-10-8280-5055; Fax: 86-10-8280-1436; E-mail: zhangyu007@hsc.pku.edu.cn; and Xiao-Song Wang, Lester & Sue Smith Breast Center, Dan L. Duncan Cancer Center, Department of Medicine, Baylor College of Medicine, One Baylor Plaza, MSB200, Houston, TX 77030. Phone: 713-798-1624; Fax: 713-798-1642; E-mail: xiaosong@bcm.edu

doi: 10.1158/0008-5472.CAN-12-1656

©2012 American Association for Cancer Research.

www.aacrjournals.org
of TSA genes, as well as translation toward cancer detection using autoantibody signatures against candidate TSAs.

Materials and Methods

Meta-analysis of Affymetrix microarray gene expression data

We compiled the publicly available Affymetrix U133 plus 2.0 microarray data for 34 normal tissue types from the human body index dataset (HBI, GSE7307), and 28 cancer types profiled by the Expression Project of Oncology (EXPO, GSE2109; Supplementary Table S1). To increase the coverage of cancer types poorly represented by EXPO, we integrated a hepatocellular carcinoma dataset (GSE6764; ref. 16), a lymphoma dataset (GSE6338; ref. 17), and a melanoma dataset (GSE7127; ref. 18). Gene expression values were extracted with the MAS5 algorithm and were scaled to a reference sample, using a housekeeping gene probe set provided by Affymetrix. These normalized expression signals are directly applied to heterogeneous expression profile analysis (HEPA) analysis, which represent "absolute expression level" as opposed to "relative expression level" in mean- or median-centered data. To depict gene expressions in a similar scale in heat map, the gene expression values were median-centered, and then divided by their median absolute deviation (MAD).

Heterogeneous expression profile analysis

HEPA analysis first calibrates the outlier expression of a gene X in a specific cancer k by taking the adjusted upper quartile mean of its expression signals in cancer k. Let

$$\bar{x} = \text{Mean}\{P_{75} \sim P_{95}\}$$

Then, the Beneficial Bonus (BB) of gene X in cancer type k is given as

$$BB_k = \log_2(\bar{x})$$

Next, a Depreciatory Penalty (DP) score for gene X is calculated on the basis of its expression signals across different types of normal human tissues. Let

$$y_j = \text{Mean}\{(x_i - \bar{x}) \times I(x_i - \bar{x})/(\kappa - \bar{x})\}$$

Here, $x_i$ is the expression signal of ith sample in tissue type j, $y_j$ is the tissue j-specific ratio. The Depreciatory Penalty score was then estimated from the weighted power sum of tissue-specific ratios $y_j$ across all normal tissues:

$$\text{DP} = \sum_{j=1}^{n} \omega_j y_j$$

Here, $\omega_j$ is the weight of tissue type j, n is the total number of normal tissue types. The power r penalizes the expression signals in somatic tissues.

The HEPA score for gene X in cancer k was then calculated as

$$\text{HEPA}_k = BB_k - \text{DP}$$

to highlight genes with the most remarkable heterogeneous expression profiles (Supplementary Fig. S1A and S1B). The parameters of HEPA score are detailed in Supplementary Methods and Fig. S2. The cutoff for a meaningful HEPA score is determined on the basis of optimal detection of known TSAs from http://www.ca.lnlcc.br/ In our compendia of datasets, a cutoff of 6 is found to be the highest cutoff offering optimal detection of known TSAs (Supplementary Fig. S2C). In addition, a HEPA score of more than 8 is empirically considered as high and more than 10 as very high.

The HEPA analysis ranks the putative antigens using the individual HEPA scores for different cancer types. The representative probe generating the highest average HEPA score for each gene is selected in this ranking. For each probe, the average HEPA score is calculated from the different HEPA scores computed for different tumor entities. Lead protein-coding gene candidates (according to the Consensus Coding Sequence Database; ref. 19) were selected and subjected to a manual inspection of expression profiles. The putative TSA of 16 major cancer types are supplied in the following website: http://hepa.cagenome.org.

Kolmogorov–Smirnov test for HEPA score

To determine the optimal range of the power r in the calculation of Depreciatory Penalty score, we tested the enrichment of the 8 prototype antigens (AFP, CTAG1B, MAGEA3, ACPP, PSA, MLANA, PMEL, and TYR) in top-scoring genes using Kolmogorov–Smirnov rank statistics (20). The resulting P value was plotted against the power r ranging between 1 and 2 (Supplementary Fig. S2A).

Tissue and serum samples

Total RNA from human normal tissues was purchased from Clontech. Tumor tissues and paired normal tissues were obtained from People's Hospital, First Hospital, and Third Hospital of Peking University (Beijing, China). The diagnosis was confirmed by independent pathologic review. Serum samples from patients with bladder, renal, gastric, colon, liver, and lung cancer were collected at People's Hospital, First Hospital, Oncology School of Peking University, PLA 306 Hospital, and 309 Hospital. Healthy donor sera with matched age and sex were collected at Third Hospital of Peking University from volunteers without known disease. All sera were stored in aliquots at -80°C until use. This study was approved by the Peking University's Ethics Committee. All clinical samples were collected with informed consent of patients.

Reverse transcription PCR and semiquantification

Reverse transcription was conducted using the Reverse Transcription System (Promega). The expression profile of putative TSA genes was examined using a 35-cycle endpoint PCR with primers listed in Supplementary Table S4. For semiquantification, band intensities were quantitated using ImageJ software (NIH) and normalized to respective glyceraldehyde-3-phosphate dehydrogenase (GAPDH) controls. A signal value less than 10% of the saturated signal was considered as no expression, between 10% and 30% as median/weak expression, and more than 30% as strong expression.
Gene cloning, protein expression, and in vitro translation

The open reading frames (ORF) of the TSA genes were cloned into the pGEM-T easy vector (Promega) and subcloned into pET-28a (Novagen). For a gene encoding a large protein product (IQGAP3, CLCA2, and IGF2BP3), the ORF was cloned as fragments. Preparation of TSPY and IGF2BP3 proteins was conducted as previously described (21). 35S-labeled proteins were prepared using the TnT T7 Coupled Transcription/Translation System (Promega) and 35S-methionine (PerkinElmer). Unincorporated 35S-methionine was removed by a desalting column (Pierce).

ELISA and Western blot analysis

ELISA and Western blot analysis to detect serum autoantibodies were conducted as previously described (21, 22).

PARSE assay

To conduct the protein A/G–based reverse serological evaluation (PARSE) assay, the black-wall high-binding 96-well plate (Greiner) was used to minimize the cross-talk between wells during scintillation counting. The plate was coated with protein A/G (Sigma, 5 μg/mL each), diluted in Na2CO3/NaHCO3 buffer (pH 9.5), and incubated at 4°C overnight. After 3 washes with PBS-0.05% NP-40 (PBSN), the plate was blocked with PBS containing 1% bovine serum albumin (BSA) for 1 hour. Then the plate was incubated with 1:100 diluted serum samples, or with anti-His (Sigma, 1:1,000) or anti-TSPY (1:1,000) antibody (21) in PBS/BSA at 4°C overnight. After 3 washes, the plate was incubated with in vitro translated protein (2 μL/well diluted in PBS/BSA) for 1 hour at room temperature. After washing, radioactivity was read on a liquid scintillation counter (PerkinElmer). Serum samples with a radioactivity value exceeding the median detected value (μ) by 3 MADs (σ) were considered to be autoantibody positive. Each radioactivity value was therefore normalized with the equation: \( x_i = \frac{t_i - \mu}{\sigma} - 3 \), in which \( x_i \) is the readout of the \( i \)th well. Negative resulting values were set to 0.

Immunoprecipitation

For each sample, 4 μL serum was diluted in 400 μL PBSN and incubated with protein A/G Sepharose CL-4B (GE Healthcare) on a rotator at 4°C overnight. After 3 washes, 5 μL in vitro–translated protein was added. After incubation for 1 hour at room temperature, the beads were washed with PBSN, eluted by sampling buffer, separated on SDS-PAGE gel, and analyzed by autoradiograph.

Statistical analysis

Receiver-operating curves (ROC) were generated and the areas under the ROC curves (AUC) were calculated to assess the predictive value of each antigen as well as the combination of the 4 selected TSAs for detection of lung cancer.

Results

Developing the computational model for genome-wide detection of tumor-specific antigens

To analyze the expression profiles of human genes under normal and cancer status, we compiled Affymetrix U133 plus 2 microarray datasets for 34 normal tissues and 28 cancer types (Supplementary Table S1). By analysis of expression profiles for 8 well-established TSA genes widely accepted as clinical targets, AFP (23), CTAG1B (13), MAGEA3 (12), ACPP (24, 25), PSA (26), MLANA (27), PMEL (28) and TTY (29), we observed that these prototype TSA genes usually exhibit distinctive expression profiles (Fig. 1 and Supplementary Fig. S1A). First, instead of being generally upregulated in cancer cells, an exceptional overexpression profile is most often noted in a relatively small subset of tumors, whereas the rest subset of tumors maintains expression silence. The level of this outlier overexpression may increase the likelihood of clinical benefits (beneficial outliers). Conventional statistical tests (e.g., Student t test) that search for ubiquitously upregulated genes across a panel of cancer samples would fail to accentuate this profile, whereas rank-based nonparametric tests, such as the Mann–Whitney U test, would ignore the strength of overexpression in a small subset of patients. Second, the expressions of most TSA genes in normal tissues present exclusively in embryonic and germ-line tissues (exemplified by fetal, placenta, and testis tissues), or are restricted to specific lineages. Such a restricted expression pattern discriminates normal somatic cells from cancer cells, implying potential clinical applications. Furthermore, this expression privilege may presumably decrease the likelihood
that recirculating lymphocytes will access these antigens. Lymphocytes responsive to them, therefore, may be well preserved in the mature repertoire (30). These observations lay the foundation for discovering TSA genes as immunologic and clinical targets by gene expression profile analysis. Toward this end, we developed a novel analysis called the HEPA, incorporating the key expression features of clinically adopted TSA genes described earlier.

The scoring scheme of HEPA takes into account both the "beneficial" outlier profile in cancers and the "depreciatory" expression profile in normal tissues (Supplementary Fig. 5LB). In this approach, a positive score is given for a signal measure of increased beneficial outliers in a specific cancer (Beneficial Bonus), and a negative penalty score is given in proportion to increased expression signals across normal tissues (Depreciatory Penalty). The scoring scheme of Beneficial Bonus takes advantage of the adjusted upper quartile mean to feature the beneficial outliers in cancer (see Materials and Methods). This function highlights the genes overexpressed in only a subset of cancer samples, while avoiding the misrepresentation caused by extremely high outlier expression signals (Supplementary Fig. 5C).

The algorithm for the Depreciatory Penalty penalizes the expression signals that are above an expression silence threshold $\psi$ in normal tissues, so that differences in the expression signals lower than this threshold will not affect the Depreciatory Penalty score (Supplementary Fig. 5B). The average of these adjusted expression signals in each normal tissue is then compared with the estimated threshold of expression signals, below which the gene products would not be able to educate the developing immune system for central tolerance specifically below which the gene products would not be able to educate the developing immune system for central tolerance specifically. These adjusted expression signals in each normal tissue is then weighted and summarized on a power scale to generate the Depreciatory Penalty score, such that the moderately to highly weighted and summarized on a power scale to generate the Depreciatory Penalty score, such that the moderately to highly expression signals in normal tissues (Depreciatory Penalty). The scoring scheme of Beneficial Bonus takes advantage of the adjusted upper quartile mean to feature the beneficial outliers in cancer (see Materials and Methods). This function highlights the genes overexpressed in only a subset of cancer samples, while avoiding the misrepresentation caused by extremely high outlier expression signals (Supplementary Fig. 5C).

The final HEPA score is then obtained by subtracting the Depreciatory Penalty from the Beneficial Bonus score to estimate the potential clinical value of human genes.

**Detecting tumor-specific antigen genes as potential clinical targets**

We first tested the feasibility of HEPA by determining its performance in detecting known clinically important TSAs. Rating the previously documented TSAs by HEPA score resulted in a substantial separation of TSAs active in cancer diagnostics and therapeutics from those with less clinical applicability based on literature reports (Fig. 2A and Supplementary Table S3A). We then ranked all human genes in our compendium of datasets by their HEPA score to highlight those showing the best heterogeneous expression profiles (Fig. 2B and Supplementary Table S3B). It is notable that we are able to isolate the clinically useful TSA genes including the 8 prototype TSAs directly from the human genome, which comprises tens of thousands of genes, simply by this single parameter. This shows the strength of the HEPA score in prioritizing TSA genes with immunologic and clinical value.

**Confirming the cancer-specific expression profile of candidate targets**

HEPA analysis also revealed a number of highly heterogeneously expressed genes in a variety of cancer types whose implication in cancer immunology and management has not yet been well established (Fig. 3A). Reverse transcription PCR (RT-PCR) detecting these gene products were conducted using 16 normal tissues and 154 paired cancer and noncancerous tissues covering 8 common human cancers. A 35-cycle endpoint RT-PCR assay was applied as a strict test to assess expression silence in normal somatic tissues. Among the 60 genes examined (Supplementary Table S4), 19 were found to be highly expressed in a variety of human cancers while silent in most normal somatic tissues (Fig. 3B, Supplementary Fig. 5A, and Table 1). The expression of each of these TSA genes was examined against a panel of tumor tissues from 154 patients with cancer (Supplementary Fig. 5B).

This validation leads to several interesting new findings. VGLL1, a poorly characterized X chromosome gene, was found to exhibit a dramatic tumor-specific expression pattern. Its expression was strictly restricted to the placenta in normal tissues but highly represented in bladder transitional cell carcinoma (TCC). Up to 50% of patients with TCC have strong expression of VGLL1, whereas it is completely absent in normal bladder and adult soma (Fig. 3B and Supplementary Fig. 5B). VGLL1 encodes a transcriptional coactivator that binds to transcriptional enhancer factor domain containing transcription factors (31). Our result calls for future studies of the functional and translational significance of VGLL1 in human cancers. Like VGLL1, SNX31 is upregulated in more than 50% of TCCs and absent in normal tissues (Fig. 3B and Supplementary Fig. 5B). SNX31 belongs to the sorting nexins family (32): its role in cancer is poorly understood. KRT8J, whose normal expression was restricted to placenta and testis, was found to be overexpressed in 67% of renal cell carcinoma (RCC), suggesting that KRT8J may be an attractive immunologic target. SLC6A3, a multipass transmembrane dopamine transporter (33), was absent in adult soma but overexpressed in 67% of patients with RCC (Fig. 3B and Supplementary Fig. 5B), which could be a potential target for monoclonal antibody therapy. Although further investigation of the literature revealed recent reports of some of the validated TSA genes as tumor biomarkers (CDH17 (34), IGF2BP3 (35), RGS4 (36), CLCA2 (37), CST1 (38), UPK2 (39), and TNFSF7 (40)), their substantial tumor specificity revealed in this study calls for further studies about their role in tumor immunology or serologic detection of human cancers. In addition, IGF2BP3 was reported as a TSA in colorectal cancers during the course of this study (41).

**Detecting autoantibody responses against the candidate TSAs in cancer patients**

Next, we sought to evaluate the immunologic significance of the candidate tumor-specific targets in patients with cancer. The presence of specific autoantibodies was taken as a surrogate marker of immunogenicity. Conventionally, serum autoantibodies are detected by Western blot analysis or ELISA, which require recombinant proteins with high purity and thus pose a great challenge for antigen preparation when evaluating...
a large number of candidates. To accelerate the process, we developed a novel assay called PARSE for rapid and easy detection of the autoantibodies against putative TSAs. Small amounts of radiolabeled protein were prepared by in vitro translation, and used as a probe for detection of autoantibodies in patients’ sera. To remove the complex repertoire of immunoglobulins to protein A/G, most of the serum proteins and other components are presumably removed in subsequent washes. Moreover, protein A/G binds to the Fc region of antibodies, therefore retaining antigen-binding activity, which may be attenuated when antibodies are directly immobilized on plates by passive adsorption (Fig. 4A, left).

Using a TSPY antigen we previously identified (21), the feasibility of PARSE was evaluated with serum samples from patients with hepatocellular carcinoma, which yielded a result comparable with ELISA (Fig. 4B and C). To verify the specificity of the PARSE assay, unlabeled TSPY protein was added in 50-fold excess to compete with 35S-labeled TSPY protein and the positive signal was completely abolished (Fig. 4C). The positive serum from the PARSE assay was further validated using a specially designed immunoprecipitation assay (Fig. 4A, right).

Figure 2. HEPA analysis preferentially identifies clinically important tumor-specific targets. A, the heat map of HEPA scores for reported TSA genes across 9 tumor entities. TSA genes reported to have clinical use are marked with colored asterisks. B, separation of TSA genes active in cancer diagnostics and therapeutics from the rest of the human genome by HEPA scoring. Left, all 16,435 coding genes included in the Affymetrix U133 plus 2 arrays are ranked by their maximum HEPA score. Right, the HEPA scores of the top 100 TSA genes in distinct tumor entities. Each point represents the HEPA score of 1 gene in 1 cancer type. Different cancer types are depicted by different colors. The 8 prototype TSAs are highlighted in red. The complete list of related literature is summarized in Supplementary Table S3.
in which immune complexes were allowed to form between the 35S-labeled antigen and the specific autoantibodies, and then pulled down by protein A/G beads. The result from this immunoprecipitation assay correlated well with Western blot analysis but with a clearer background (Fig. 4D and E). Immunoprecipitation of radioactive 35S-labeled TSPY protein was also abolished by excess unlabeled TSPY protein (Fig. 4E). This modified immunoprecipitation was thus used in subsequent studies as a confirmation test for the PARSE assay. The performance of PARSE in comparison with ELISA was further evaluated using another TSA, IGF2BP3 (Supplementary Fig. S4A and S4B), which yielded similar results.

To apply the PARSE assay to the newly identified TSA genes, 12 putative antigen proteins were successfully expressed using in vitro translation (Table 1). To show the general applicability of HEPA–PARSE approach in cancer, we have attempted to...
include multiple cancer types with available serum samples in our laboratory. These include lung cancer (n = 72), gastric cancer (n = 48), colon cancer (n = 48), RCC (n = 48), and TCC (n = 42) patients. We therefore interrogated each antigen against 1 or 2 of these 5 cancer types with the highest expression level and positive rate for that antigen (according to RT-PCR results). Serum samples from 72 healthy donors are used as control. Spontaneously arising antibodies against 7 of these TSAs were detected in patients with cancer (Fig. 5A and Table 1). The results obtained from the PARSE assay were subsequently verified by the specially designed immunoprecipitation assay noted earlier (Fig. 5B). The radiolabeled tumor antigens were immunoprecipitated by PARSE-positive sera from patients with cancer, but not by sera from PARSE-negative patients or healthy donors.

Translation toward diagnostic autoantibody signatures in cancer

The overall positivity of autoantibodies specific for each of the TSAs varies from 4.2% to 11.1% in distinct cancer subtypes (Table 1), and in the same tumor entity there was limited overlap in the spectrum of antibodies against different antigens (Fig. 5A). This raised the possibility of combining these TSA targets into diagnostic autoantibody signatures in specific cancers. Among 72 patients with lung cancer, for example, the numbers of patients showing positive responses against each of the 4 individual antigens, IGF2BP3, KRT23, IQGAP3, and CLCA2, were 5, 8, 5, and 5, respectively, but the total number reached 21 when results for the 4 antigens were combined. Similarly, 11 of 48 patients with gastric adenocarcinoma contained autoantibodies against at least 1 of the 3 tumor antigens, KRT23, IQGAP3, and REG3A (Fig. 5A). To further assess the diagnostic value of the antibody signature in lung cancer, ROC curves were generated on the basis of the normalized PARSE scores. In a cohort of 72 patients with lung cancer and 70 healthy donors, the AUC ranged from 0.614 to 0.667 for each of the 4 TSAs alone, and increased to 0.711 when combined (Fig. 5C). The same analysis was conducted against an independent, but larger cohort containing 149 cases of patients with lung cancer and 123 healthy individuals (Supplementary Fig. S4C), which yielded an AUC of 0.768 when the 4 antigens were combined (Fig. 5D). In both cohorts, the age and sex distributions are comparable between cancer and healthy donor groups (Supplementary Table S5A). Furthermore, the positive rates of autoantibodies are similar between patients with early- and late-stage lung cancer (Supplementary Table S5B). These

### Table 1. Summary of mRNA and autoantibody profile of TSAs validated in this study

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Subcellular localization</th>
<th>Cancer type</th>
<th>mRNA expression (%)</th>
<th>Cancer patients</th>
<th>Healthy donors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mRNA expression (%)</td>
<td>PARSE+/total (%)</td>
<td>IP+/total tested</td>
</tr>
<tr>
<td>HAS3</td>
<td>Multipass membrane</td>
<td>LC</td>
<td>8.3</td>
<td>5/72 (6.9)</td>
<td>0/70 (0)</td>
</tr>
<tr>
<td>IGF2BP3</td>
<td>Cytoplasm; nucleus</td>
<td>LC</td>
<td>25</td>
<td>5/72 (6.9)</td>
<td>0/70 (0)</td>
</tr>
<tr>
<td>IQGAP3</td>
<td>Cytoplasm</td>
<td>LC</td>
<td>25</td>
<td>5/72 (6.9)</td>
<td>0/70 (0)</td>
</tr>
<tr>
<td>KRT23</td>
<td>Cytoplasm</td>
<td>GA</td>
<td>26.7</td>
<td>2/48 (4.2)</td>
<td>2/9</td>
</tr>
<tr>
<td>KRT81</td>
<td>Cytoplasm</td>
<td>RCC</td>
<td>33.3</td>
<td>3/48 (6.3)</td>
<td>1/6</td>
</tr>
<tr>
<td>VGLL1</td>
<td>Nucleus</td>
<td>RCC</td>
<td>11.1</td>
<td>3/42 (7.1)</td>
<td>1/6</td>
</tr>
<tr>
<td>CDH17</td>
<td>Type I membrane</td>
<td>CA</td>
<td>16.1</td>
<td>1/46 (2.2)</td>
<td>0/18</td>
</tr>
<tr>
<td>KRT20</td>
<td>Cytoplasm</td>
<td>CA</td>
<td>12.9</td>
<td>0/48 (0)</td>
<td>0/46 (0)</td>
</tr>
<tr>
<td>MMP3</td>
<td>Secreted</td>
<td>CA</td>
<td>6.5</td>
<td>0/48 (0)</td>
<td>0/46 (0)</td>
</tr>
<tr>
<td>REG4</td>
<td>Secreted</td>
<td>CA</td>
<td>3.2</td>
<td>0/48 (0)</td>
<td>0/46 (0)</td>
</tr>
<tr>
<td>SPINK4</td>
<td>Secreted</td>
<td>CA</td>
<td>6.5</td>
<td>0/48 (0)</td>
<td>0/46 (0)</td>
</tr>
<tr>
<td>REG3A</td>
<td>Cytoplasm; secreted</td>
<td>GA</td>
<td>6.7</td>
<td>2/48 (4.2)</td>
<td>1/6</td>
</tr>
<tr>
<td>CST1</td>
<td>Secreted</td>
<td>GA</td>
<td>20</td>
<td>6.7</td>
<td></td>
</tr>
<tr>
<td>CLCA2</td>
<td>Type II membrane</td>
<td>LC</td>
<td>8.3</td>
<td>5/72 (6.9)</td>
<td>1/48 (2.1)</td>
</tr>
<tr>
<td>NTS</td>
<td>Secreted</td>
<td>LC</td>
<td>8.3</td>
<td>0/70 (0)</td>
<td>0/70 (0)</td>
</tr>
<tr>
<td>SLC6A3</td>
<td>Multipass membrane</td>
<td>RCC</td>
<td>16.7</td>
<td>0/70 (0)</td>
<td>0/70 (0)</td>
</tr>
<tr>
<td>TNFSF7</td>
<td>Type II membrane</td>
<td>RCC</td>
<td>16.7</td>
<td>0/70 (0)</td>
<td>0/70 (0)</td>
</tr>
<tr>
<td>SNX31</td>
<td>Soluble, nonsecreted</td>
<td>TCC</td>
<td>18.5</td>
<td>0/42 (0)</td>
<td>0/42 (0)</td>
</tr>
<tr>
<td>LIPK2</td>
<td>Type I membrane</td>
<td>TCC</td>
<td>25.9</td>
<td>0/40 (0)</td>
<td>0/48 (0)</td>
</tr>
</tbody>
</table>

*Predicted by the protein subcellular localization database—LOCATE (http://locate.imb.uq.edu.au/).

bIVT, protein that successfully in vitro translated; Y, yes.
cIP, immunoprecipitation.
The principle and feasibility of PARSE and immunoprecipitation assays. A, schematic depiction of the principle of PARSE (left) and immunoprecipitation assays (right). B, detection of autoantibodies against a known cancer-testis antigen, TSPY, by ELISA. Representative results are shown with serum samples from healthy individuals (N21, N22) and patients with HCC (H89, H90, and H91). The anti-TSPY antibody served as a positive control and PBS as a negative control. *, positive serum. C, detection of TSPY-specific autoantibodies using the PARSE assay. D, confirmation of TSPY-specific autoantibodies by Western blot analysis. The specific band for TSPY is indicated with arrowhead. *, nonspecific bands. E, confirmation test for TSPY-specific autoantibodies by immunoprecipitation. Input, radio-labeled TSPY. B and C, error bars, SD; n = 3; C and E, to verify the binding specificity, unlabeled TSPY protein was added in 50-fold excess to compete with 35S-labeled TSPY (H90 + TSPY). OD, optical density. CPM, counts per minute.

Discussion

Here, we have reported the development and application of a mathematical scheme for genome-wide detection of TSA genes as immunologic and clinical targets. Rather than relying solely on traditional statistical tests based on distribution estimation or ranks, our novel approach incorporates the distinctive biologic features within a panel of established TSA genes widely adopted in clinics. Applying HEPA analysis to large gene expression datasets for multiple cancer types and a spectrum of normal tissues revealed arrays of TSA genes with diagnostic and therapeutic potential. Profound enrichment of known clinically useful TSA genes was observed in these lists. Endpoint RT-PCR validation revealed that this approach has greatly improved predictive power in terms of tumor specificity, over the statistical methods that we previously used in a putative tumor antigen target database (ref. 42; data not shown). Of note, the parameters of the HEPA analysis applied in this study have been optimized to discover TSAs whose expressions are absent in most somatic tissues, as shown by strict endpoint RT-PCR assays. However, some of the over-expressed antigens may also be important in the therapeutic setting, such as EGF receptor (EGFR) and HER-2. Interestingly, we observed that such therapeutically relevant overexpressed antigens also possess some degree of tumor specificity (Supplementary Fig. S5A). With adjusted parameters, these therapeutically relevant overexpressed antigens can also be detected by HEPA analysis (Supplementary Fig. S5B). The list of putative overexpressed antigens is provided in the website: http://hepa.cagenome.org.

To quickly evaluate the immunologic significance of the large number of candidate TSA genes in patients with cancer, we developed a novel experimental approach called the PARSE assay. PARSE leverages the mammalian in vitro translation system as a quick and efficient way to produce a small amount of TSAs with natural conformations and posttranslational modifications. Thus, it presumably can generate a precise picture of autoantibody responses in patients with cancer by detecting the full spectrum of epitopes on these antigen proteins. Furthermore, the in vitro translation system enables the expression of toxic proteins not producible in live cells and avoids aggregation of proteins in inclusion bodies as occurs in bacteria. Our data suggest that the output of PARSE is comparable with conventional ELISA (Fig. 4B and C and Supplementary Fig. S4A and S4B), with greatly improved throughput in detecting autoantibodies against new TSAs.

Identifying gene targets highly specific for a malignant condition is a great challenge for developing a serologic cancer detection assay. Careful choices are required when considering which antigen genes should be selected for assembling an autoantibody signature. The HEPA–PARSE approach showed improved specificity and greater speed in the discovery of TSAs and validation of autoantibody signatures. In contrast to frequent autoantibody responses observed in patients with cancer, specific autoantibodies against these TSAs were found to be extremely rare in a comparable number of healthy individuals (Fig. 5A and B). This remarkable tumor specificity may be attributable to the highly restricted expression of the TSAs in normal somatic tissues and the improved background readout of the PARSE assay. Of note, 5 of 12 candidate antigens tested do not elicit autoantibody response in patients of selected cancer types. Interestingly, 2 of 3 membrane proteins, 2 of 3 secreted proteins, and 2 of 2 colon-specific proteins are found to be negative. Further studies will be needed to elucidate the impact of subcellular localization, lineage specificity, and other factors on the natural choice of an antigen. Furthermore, we observed a high degree of mutual exclusivity in the patient autoantibody spectrum for different antigens. For example, while specific autoantibodies against IGF2BP3, KRT23, IQGAP3, and CLCA2 were detected in 6.9%, 11.1%, 6.9%, and 6.9% of patients with lung cancer, respectively, their combination covered up to 29.2% of patients (21 of 72), resulting in a distinctive autoantibody signature (Fig. 5A).
To further evaluate the value of this autoantibody signature in the detection of lung cancer, we conducted the PARSE assays using sera from an independent and larger cohort of 149 patients with lung cancer and 123 healthy individuals (Supplementary Fig. S4C). Comparable AUC of ROC curves were obtained from these 2 independent cohorts, confirming the diagnostic value of this autoantibody signature (Fig. 5C and D). Future studies will be needed to elucidate the translational
significance of this autoantibody signature and to compare with other related but noncancer conditions. More TSAs should be identified by the HEPA–PARSE approach and combined with the current antigen panel to increase sensitivity. Furthermore, the level of mutual exclusiveness in the gene expression profiles of candidate antigens can be analyzed using microarray datasets to identify the best combination of TSAs in the detection of specific tumors.

It is notable that most of the TSAs validated in this study are related to multiple tumor entities except SLC6A3 (exclusive to RCC), or UPK2 and SNX31 (exclusive to TCC). Those antigens expressed in multiple tumor entities may be used for the detection of cancer in general, whereas others exclusive to certain cancer types can be used to identify the specific tumor entity. This could be important for determining treatment options, which is often difficult when the original tumor is undetectable. The beauty of HEPA analysis is that it provides a quantitative value for each gene in each tumor entity (as shown in Fig. 2 and Supplementary Fig. S3A), so that the preference of expression for each TSA in different cancers can be apparent at a glance.

Together, the HEPA–PARSE technology provides a generally applicable approach for genome-wide dissection of the cancer-specific antigen genome, and for the quick development of autoantibody signatures in the detection of epithelial tumors. By optimizing the parameters, HEPA technology can be readily adapted to many specific applications, such as the identification of autoantibody signatures, biomarkers, tumor vaccine targets, or membrane antigen targets. Its performance could be further enhanced by combining it with other datasets, such as cancer plasma or membrane proteomics datasets. With the development of deep transcriptome sequencing and digital gene expression technology, we expect that interest in HEPA analysis will increase as an approach to interrogate these new datasets.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: Q.-W. Xu, W.-F. Chen, Y. Zhang, X.-S. Wang
Development of methodology: Q.-W. Xu, W. Zhao, M.A. Sartor, X.-S. Wang
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): Q.-W. Xu, W. Zhao, D.-M. Han
Analysis and interpretation of data (e.g., statistical analysis, bios-statistics, computational analysis): Q.-W. Xu, W. Zhao, Y. Wang, J. Deng, Y. Zhang, X.-S. Wang
Writing, review, and/or revision of the manuscript: Q.-W. Xu, M.A. Sartor, Y. Zhang, X.-S. Wang
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Y. Wang, R. Ponnala, J.-Y. Yang, Q.-Y. Zhang, G.-Q. Liao, Y.-M. Qu, L. Li, F.-F. Liu, H.-M. Zhao, Y.-H. Yin
Study supervision: Y.-H. Yin, W.-F. Chen, Y. Zhang, X.-S. Wang

Acknowledgments
The authors thank Dr. Jun Zhang for helpful discussion and critical reading of the article and Xuewen Pang, Xiao-Ping Qian, and Yan Li for technical assistance.

Grant Support
This study was supported by grants from National Natural Science Foundation of China (nos. 30525044 and 30800901), National High-tech R&D Program of China (863 Program, no. 2007AA021103), National Basic Research Program of China (973 Program, no. 2011CB940600), Congressionally Directed Medical Research Programs (W81XWH-12-1-0166), NIH grants P30-D15123-06, and NCRR S10RR02950.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 7, 2012; revised August 29, 2012; accepted September 25, 2012; published OnlineFirst November 7, 2012.

References
An Integrated Genome-Wide Approach to Discover Tumor-Specific Antigens as Potential Immunologic and Clinical Targets in Cancer

Qing-Wen Xu, Wei Zhao, Yue Wang, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-12-1656

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2012/11/05/0008-5472.CAN-12-1656.DC1

Cited articles
This article cites 41 articles, 17 of which you can access for free at:
http://cancerres.aacrjournals.org/content/72/24/6351.full.html#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
/content/72/24/6351.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.