Technical validation of an autoantibody test for lung cancer

A. Murray¹, C. J. Chapman², G. Healey¹, L. J. Peek³, G. Parsons⁴, D. Baldwin⁵, A. Barnes³, H. F. Sewell⁶, H. A. Fritsche⁷ & J. F. R. Robertson²*

¹Oncimmune Ltd; ²Division of Breast Surgery, University of Nottingham, Clinical Sciences Building, Nottingham City Hospital, Nottingham, UK; ³Oncimmune LLC, De Soto, KS; ⁴Parsons Group LLC, Arlington, MA, USA; ⁵Department of Respiratory Medicine, Nottingham City Hospital; ⁶Division of Immunology, School of Molecular Medical Sciences, Medical School, Queens Medical Centre, Nottingham, UK and ⁷Department of Laboratory Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA

Received 9 October 2009; accepted 11 December 2009

Background: Publications on autoantibodies to tumour-associated antigens (TAAs) have failed to show either calibration or reproducibility data. The validation of a panel of six TAAs to which autoantibodies have been described is reported here.

Materials and methods: Three separate groups of patients with newly diagnosed lung cancer were identified, along with control individuals, and their samples used to validate an enzyme-linked immunosorbant assay. Precision, linearity, assay reproducibility and antigen batch reproducibility were all assessed.

Results: For between-replicate error, samples with higher signals gave coefficients of variation (CVs) in the range 7%–15%. CVs for between-plate variation were only 1%–2% higher. For between-run error, CVs were in the range 15%–28%. In linearity studies, the slope was close to 1.0 and correlation coefficient values were generally >0.8. The sensitivity and specificity of individual batches of antigen varied slightly between groups of patients; however, the sensitivity and specificity of the panel of antigens as a whole remained constant. The validity of the calibration system was demonstrated.

Conclusions: A calibrated six-panel assay of TAAs has been validated for identifying nearly 40% of primary lung cancers via a peripheral blood test. Levels of reproducibility, precision and linearity would be acceptable for an assay used in a regulated clinical setting.

Key words: autoantibodies, diagnostic test, lung cancer, tumour-associated antigens

Introduction

Lung cancer is the highest cause of death from cancer worldwide, responsible for the deaths of >1 million men and women every year [1]. Lung cancer is often detected on chest X-ray, but by this time, the cancer is usually advanced and few patients are cured by treatment. However, if diagnosed early, the chances of cure are ~90%. At present, there is no early detection test or acceptable screening method for this disease; therefore, there is an urgent need to produce a screening test that can identify the cancer in its early curable stage, especially in high-risk individuals.

Circulating antibodies which react with tumour-associated antigens (TAAs) have been found in serum samples from patients with a variety of cancers, including lung cancer [2–10], and may represent an early indicator of the presence of cancer. It has been hypothesised that the heterogeneity of antigen expression will mean that a panel of assays for autoantibodies of various TAA specificities will be needed for effective detection of lung cancer [11]. Recent publications have confirmed that measuring autoantibodies to a panel of antigens gives a significantly greater level of sensitivity compared with that for a single antigen [5–10].

Autoantibodies have been found in the blood of patients who develop lung cancer up to 5 years before screening spiral computed tomography scans were able to detect the tumour [2]. Consequently, monitoring people at increased risk of lung cancer for the presence of serum autoantibodies may enable earlier detection of the disease, allowing earlier therapeutic intervention.

This article reports the laboratory validation data and performance characteristics for a serum autoantibody test panel consisting of six TAAs to which autoantibodies have been described. The antigens are p53, NY-ESO-1, cancer-associated antigen (CAGE), GBU4-5, Annexin 1 and SOX2, and the samples are from patients with or without newly diagnosed lung cancer. Specifically, we address the development of quality assurance in reagent preparation, analyte calibration and quality-control (QC)
protocols, which are required of a serum autoantibody panel test carried out in a clinical laboratory setting.

**materials and methods**

**production of recombinant antigens for autoantibody assays**

**vector construct.** Specific complementary DNA (cDNA) for p53, NY-ESO-1, CAGE, GBU-4 and Annexin 1 were sub-cloned into the pET21b expression vector (which had previously been engineered to also express a BirA tag). The specific cDNA for SOX2 was sub-cloned into the pET44b expression vector (expressing a NusA tag) (Novagen, Darmstadt, Germany).

**expression of recombinant proteins/antigens.** The recombinant proteins were expressed in BL21 (DE3) bacteria (Novagen), grown in terrific broth (TB) or autoinduction TB media (Novagen), and purified using HisTrap affinity columns (GE Healthcare, Uppsala, Sweden) according to manufacturers’ protocols. Negative control proteins consisting of either BirA or NusA alone were also produced. Antigens were produced by one of the two external manufacturers, except in the case of one study for which SOX2 was produced by Oncimmune Ltd. Details of the antigens used for each study are given in the supplemental Table S1 (available at Annals of Oncology online).

**patients**

Three separate groups of patients with newly diagnosed lung cancer were identified (supplemental Table S1, available at Annals of Oncology online).

Group 1 contained 145 lung cancer patients (median age 66; range 41–87) and 146 normal controls (median age 66; range 41–87). Similarly, group 2 had 241 (63; 28–87) and 240 (63; 28–87), respectively, while group 3 had 269 (65; 38–87) and 269 (65; 38–86). All patients with lung cancer were as far as possible individually matched by sex, age and smoking history to a control individual with no previous history of malignant disease. In patients with lung cancer, blood samples were obtained after diagnosis but before receiving any anticancer treatment. Samples were obtained, with full informed consent, from the enrolment sites.

**assay procedure**

A semi-automated indirect enzyme-linked immunosorbent assay was utilised (all liquid-handling steps were carried out using an automated liquid-handling system). Purified recombinant antigens were diluted to provide a semi-log titration series for each antigen ranging from 160 to 1.6 nM. Control antigens (BirA and NusA) were also included to allow subtraction of the signal due to nonspecific binding to bacterial contaminants. Antigen dilutions were passively adsorbed to the surface of microtitre plate wells in high phosphate buffer overnight at room temperature. After washing in phosphate-buffered saline containing 0.1% Tween 20 (pH 7.6), microtitre plates were blocked with a gelatine-based blocking buffer. Coated plates were found to be stable for at least 48 h after coating if washed and stored at 4°C in the presence of blocking buffer (Oncimmune Ltd, data on file). Serum samples (diluted 1 in 110 in a blocking buffer) were then added to the plates and allowed to incubate at room temperature with shaking for 90 min. Following incubation, plates were washed and horseradish peroxidase (HRP)-conjugated rabbit anti-human IgG (Dako, Glostrup, Denmark) was added. After a 60-min incubation with shaking, the plates were washed and 3,3’ 5,5’-tetramethylbenzidine was added. The optical density (OD) of each well was determined spectrophotometrically at 650 nm after a 15-min incubation. Control plates were run on the same day (between plate). Interassay precision (intermediate ‘between-run’ reproducibility) for each antigen in the assay was based on the variation over days within the same study (between run).

Variance components and coefficients of variation (CVs) were estimated using standard analysis of variance methods. Separate calculations were carried out on antilogged RU values for samples with high, medium and low signals. The high and medium signals generally spanned the diagnostic test cut-point values. Since independent reference materials of known concentration were not available for these assays, it was not possible to assess accuracy.

**linearity.** Linearity was assessed by a serum dilution study using two serum samples for each of the six antigens under the assumption that the top concentration gave 100% recovery. A doubling dilution series was prepared for each sample. The panel of autoantibody assays was carried out on each dilution and the OD measured. The dilution was estimated from the relevant calibration curve and a plot of the estimated versus the actual (known) dilution was constructed. The slope and intercept were estimated using linear regression and the goodness of fit assessed using the linear correlation coefficient. A slope of between 0.8 and 1.2, an intercept within 0.1 of zero and a correlation coefficient of at least 0.8 were required for satisfactory linearity.

**QC monitoring.** Six high-signal QC serum samples, one for each antigen, were interspersed amongst test samples in a series of studies carried out over a 14-week period. The results were compared with their respective expected values as established in preliminary studies. This enabled the QC sera to be monitored over time to evaluate their reproducibility using Levey–Jennings plots. The usual mean ± 3 SD chart limits were adjusted to mean ± 3.5 SD for the QC of six separate antigens.

**assay reproducibility.** The sensitivity and specificity for lung cancer of each of the individual autoantibody assays as well as the panel were assessed for each of
the patient sample groups. Values for groups 1 and 2 were derived from background-corrected OD data, whereas values for group 3 were derived from data to which calibration had been applied in order to obtain measurements in RUs. For each assay, a test cut-point of mean ± 3 SD was used.

Between-run reproducibility was also assessed using concordance rates, i.e. the percentage of samples where the calibrated result, measured in RUs, was the same (i.e. positive or negative) on two runs carried out on separate days. Calibrated Annexin 1 results were not available for one of the two runs in this particular study, so between-run reproducibility is reported for the other five assays in group 3.

antigen batch reproducibility. The capture antigen is the most critical reagent in the autoantibody assay and, as such, it is essential that reproducibility between batches be demonstrated.

A subset of the group 3 samples was run in an assay constructed with four different batches of CAGE as the capture antigen. These four different batches had been obtained from the same supplier over a 3-month period and had been produced and purified from four different fermentation runs. The mean calibrated signal for cancer and normal groups, and the sensitivity/specificity, was compared across batches.

results
precision
For measurement of within-plate variation, samples with high signals gave CVs in the range 7%–15% (Table 1).

Table 1. Between-replicate precision and intermediate reproducibility estimates

<table>
<thead>
<tr>
<th>Medium signal</th>
<th>High signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean OD</td>
<td>CVe (%)</td>
</tr>
<tr>
<td>p53</td>
<td>0.17</td>
</tr>
<tr>
<td>SOX2</td>
<td>0.31</td>
</tr>
<tr>
<td>NY-ESO-1</td>
<td>0.35</td>
</tr>
<tr>
<td>CAGE</td>
<td>0.23</td>
</tr>
<tr>
<td>GBU4-5</td>
<td>0.10</td>
</tr>
<tr>
<td>Annexin 1</td>
<td>0.13</td>
</tr>
<tr>
<td>Average</td>
<td>0.15</td>
</tr>
</tbody>
</table>

OD means on the basis of ~216 observations and 12 runs each; CVs on the basis of mean of two replicates of antilogged RU values.

OD, optical density; CVe, between-replicate CV; CVr, between-run CV (including intra-assay component); CVs, coefficients of variation; RU, reference unit.

As expected, for samples with medium signals, CVs were higher, in the range 9%–23%. Samples with low signals (not presented) generally gave CVs > 20%, increasing as the mean OD approached zero.

For between-plate variation (not presented), CVs were only 1%–2% higher than for between-replicate error, indicating very little additional variability due to plate-to-plate differences.

For measurement of between-run variation, high-signal samples gave an average CV of 19% (range 15%–28%) (Table 1). Again as expected, for samples with medium signals, the average CV was higher at 28% (range 22%–42%). Samples with low signals generally gave CVs > 30%. These figures indicate significant additional variation due to run-to-run effects, increasing the CV by roughly 10% (Table 1). This amount is typical for assay work.

linearity
For all samples, the slope estimate was close to 1.0 and the intercept close to zero (Table 2; Figure 1), thereby indicating a linear assay. Correlation coefficients (r) were >0.83 in all but one case (Annexin 1; sample C6; r = 0.77), with a median of 0.98, thereby demonstrating satisfactory goodness of fit.

QC monitoring
The plots of calibrated results (RUs) versus time (Figure 2) showed that all QC serum control values for each of the six antigens fell within the standard deviation limits, demonstrating that the calibration system was effective in producing stable day-to-day QC results.

assay reproducibility
The sensitivity and specificity of each autoantibody assay as well as the panel for each of the sample groups are summarised in the supplemental Table S2 (available at Annals of Oncology online). The overall panel sensitivities and specificities were very similar between groups, demonstrating the validity of the calibration system and the robustness of the assay.

Using concordance data, the reproducibility of the calibrated panel (group 3) was confirmed as >95%. The number of samples that changed status (i.e. positive or negative) ranged from 0.5% to 2.4% per antigen. When categorised by cancer or...
normal, the figures were 0%–2.6% and 0.4%–2.2%, respectively.

antigen batch reproducibility

The characteristics of the assays run on a subset of group 3 samples using four different batches of CAGE antigen are summarised in Table 3. The mean calibrated signals for cancers and normal groups were similar across batches, demonstrating the effectiveness of the calibration system. The sensitivity and specificity of individual batches of antigen was also very similar, as was the sensitivity and specificity of the panel of antigens as a whole, demonstrating the robustness of an assay on the basis of a panel approach.

discussion

Assays used to measure serum antibodies for the diagnosis and management of autoimmune disease often employ World Health Organization (WHO) standards as calibration materials [14, 15]. These are usually derived from human serum or plasma (WHO website: http://www.who.int/bloodproducts/ref_materials/en/). As such, issues of longevity arise with new
standard materials needing to be sourced, validated and introduced on a regular basis.

The fluids used as calibrator material in our study were drained from the pleural cavities of patients suffering from lung cancer as part of the normal course of their disease management. Use of these fluids provided a long-term source of calibrator material so that the inevitable difficulties encountered with sourcing and validating new materials could be avoided for as long as possible and the reproducibility of the assay result over months or even years could be assured.

Although every effort was made to ensure that the fluids were stored under conditions that would not allow their reactivity to deteriorate, a study is underway to investigate the stability of pleural fluids as calibrator materials under long-term storage.

The optimised dilution of sera used (Oncimmune Ltd, data on file), together with the use of an anti-human IgG-HRP conjugate detection system, is considered important in the development of this test. They optimise detection of relevant autoantibodies while minimising detection of low-titre nonspecific antibodies (known to accumulate with advancing age) and poly-reactive IgM antibodies [16, 17].

Linearity is often difficult to demonstrate in serology assays and can show wide variability between assays [18]. However, we report good assay linearity with a slope close to 1 and a high correlation coefficient for all samples. We have also shown that the assay has CVs generally <15% and <25% for intra- and interassay precision, respectively. This is in line with that reported for measurement of serum autoantibodies in patients with benign autoimmune disease [19, 20]. Three discrete sets of clinical samples (totalling 655 cancers and 655 normals) were used for clinical validation of the assay. The sensitivity and specificity was shown to be in the region of 40% and 90%, respectively, for all three groups, demonstrating the reproducibility and robustness of this assay system. The fact

Figure 2. Levey–Jennings plots of control sera for each antigen over a 14-week period.
The performance characteristics of an assay for the measurement of autoantibodies against a panel of antigens known to be expressed in lung cancer have confirmed its reproducibility for identifying nearly 40% of primary lung cancers through a peripheral blood test. The levels of reproducibility, precision and linearity would be acceptable for an assay used in a regulated clinical setting. A simple blood test that is both reliable and reproducible, such as that described here, represents a potential aid to imaging modalities. This assay has now been commercialised and will be used as a ‘platform technology’ that is technically valid for identifying all types of solid tumours.

### Table 3. Antigen batch reproducibility

<table>
<thead>
<tr>
<th>CAGE batch</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean RU at 160 nM</td>
<td>3.18</td>
<td>3.00</td>
<td>3.17</td>
<td>3.15</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>97</td>
<td>98</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>Panel sensitivity (%)</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>Panel specificity (%)</td>
<td>87</td>
<td>87</td>
<td>86</td>
<td>87</td>
</tr>
</tbody>
</table>

Four different batches of CAGE were used in the assays run against a subset of group 3 samples. RU, reference unit.

aMeans on the basis of ~250 samples each.

### funding

Oncimmune Ltd. Funding to pay the Open Access publication charges for this article was provided by The University of Nottingham and Oncimmune Ltd.

### acknowledgements

P. Maddison is gratefully acknowledged for the provision of samples. J. Halsey is gratefully acknowledged for advice, valuable technical suggestions and comments on the manuscript. The manuscript was drafted by the authors. However, the authors would acknowledge and thank Sandra Cusco, from Complete Medical Communications, who provided final editorial assistance funded by Oncimmune Ltd.

### disclosure

CC and JFRR are consultants to Oncimmune Ltd, a University of Nottingham spinout company. HAF, HFS and GP are members of the Oncimmune Scientific Advisory Board.

### references