Human Proteome Arrays for Autoantibody Identification in Clinical Cancer Studies

Mikayla Higgins, John Kenten, Christopher Frye, Sudeep Kumar, John Joern, George B. Sigal, Eli N. Glezer, and Jacob N. Wohlstadter

Meso Scale Diagnostics, LLC, Rockville, Maryland, USA

1 Abstract

Early detection of cancer is a critical factor for successful treatment of cancer patients. Autoantibody signatures have value in the diagnosis and management of autoimmune disorders and may also be valuable for detection of cancers. Previous groups have applied several techniques to identify cancer-specific autoimmune responses with varying degrees of success. An important bottleneck in the development of autoantibodies as biomarkers for cancer diagnosis has been a lack of high-throughput methods to rapidly screen large numbers of patient samples and validate multiple antigens of interest. We have developed a novel, flexible assay platform to address this issue.

We tested normal and cancer patient sera against protein arrays generated using a gene library capable of expressing full-length human proteins. Autoantibodies against proteins in these arrays were identified using electrochemiluminescence detection technology from MSD®, a proven technology for high-throughput array-based measurements. This approach offers key advantages over existing array-based approaches by providing a high-throughput, sensitive, and specific assay platform, with large-scale protein arrays. This approach also enables a rapid transition from antigen discovery to clinical validation using small focused arrays on the same diagnostic platform.

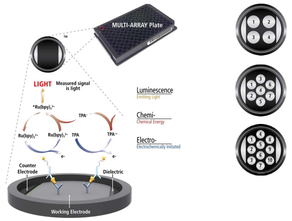
Protein arrays containing ~7,000 non-redundant full-length human proteins were generated using a eukaryotic expression system. This set of proteins represents a significant portion (28%–35%) of the human proteome. We estimate that greater than 80% of autoantibodies recognize unmodified epitopes on single protein antigens. The protein arrays should therefore be highly representative of the vast majority of autoantigens. These "proteome-scale" arrays were used to screen for cancer-specific autoantibodies in patient serum samples. Patient samples were initially screened as pools, followed by screening of individual samples.

Screening was done using samples from patients with breast, lung, ovarian, prostate, colon, kidney, melanoma, and pancreatic cancers. Control samples included samples from normal patients as well as those with benign breast disease, bronchial dysplasia, benign ovarian disease, and benign prostate hyperplasia. Samples from patients with breast, lung, ovarian, and prostate cancers were sub-divided into additional groups based on clinical findings. In total we screened 22 different groups.

The approach was validated by the identification of well-established cancer autoantigens, including tumor antigen p53 (TP53). Ovarian cancer yielded the highest percentage of positive autoantibody responses with 21% for TP53, 8% for DEK, and 6% for GMPR2. The discovery of diagnostic autoantigens in cancer using this approach may be improved by further evaluation using sets of molecularly classified patient samples, especially in combination with patient-specific protein arrays determined from tumor whole-genome sequencing data.

2 Methods

MSD's electrochemiluminescence detection technology uses SULFO-TAG™ labels that emit light upon electrochemical stimulation initiated at the electrode surfaces of MULTI-ARRAY® and MULTI-SPOT® microplates.



Electrochemiluminescence Technology

- Minimal non-specific background and strong responses to analyte yield high signal-to-background ratios.
- The stimulation mechanism (electricity) is decoupled from the response (light signal), minimizing matrix interference.
- Only labels bound near the electrode surface are excited, enabling non-washed assays.
- Labels are stable, non-radioactive, and directly conjugated to biological molecules.
- Emission at ~620 nm eliminates problems with color quenching.
- Multiple rounds of label excitation and emission enhance light levels and improve sensitivity.
- Carbon electrode surface has 10X greater binding capacity than polystyrene wells.
- Surface coatings can be customized.

3 Strategy

To create an assay platform with an enhanced ability to discover autoantibodies, we developed a highly systematic method for expressing and immobilizing proteins from the mammalian genome. To achieve this, we made use of our curated mammalian gene collection in combination with an efficient eukaryotic transcription and translation system that enables the production of native proteins on a genome-wide scale. This gene collection had previously been characterized in-silico based on gene function and by SDS-PAGE for protein molecular weight (Figure 1). We were able to produce and simultaneously label individual proteins, immobilizing them on MSD plates in a multiplexed format via a proprietary, solution-addressable linker-based approach. These immobilized proteins could then be used for the detection and direct identification of auto-reactive antibodies using the uniquely simple and sensitive multiplexed detection technology developed by MSD.

Figure 2 is a schematic of MSD's proprietary linker-based method for generation of antigen arrays using proteins produced in transcription and translation (TnT) reactions. Individual clones are expressed in a TnT reaction, generating biotinylated proteins that are subsequently individually bound to unique streptavidin-proprietary linker conjugates. These protein-linker complexes are then pooled into mixtures. Corresponding linkers are immobilized on spatially distinct spots in wells on MSD microplates to create an array of unique linking sites for the specific capture of the protein-linker complexes. The pooled protein-linker complexes are dispensed into the wells and specific binding of corresponding linker pairs generates the protein arrays. The overall workflow from cloned DNA to sample screening is illustrated in Figure 3.

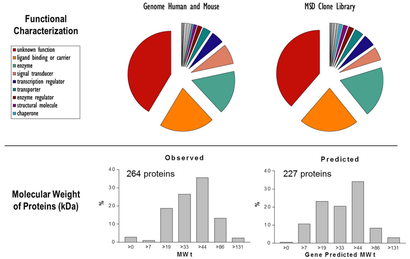


Figure 1. Characterization of MSD's clone library

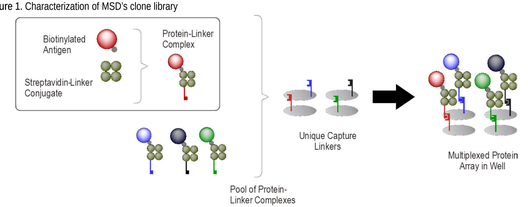


Figure 2. Schematic of Linker-Based Generation of Proteome Arrays

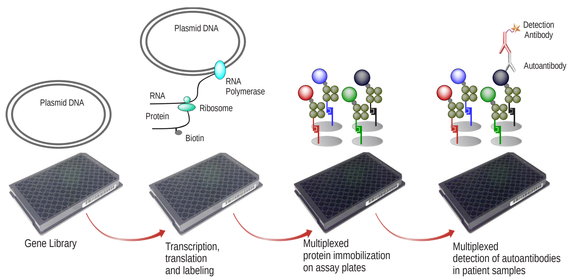


Figure 3. Screening Workflow Schematic

4 Automated Array Production

In order to produce large scale, reproducible antigen arrays, we developed a largely automated production protocol. We selected 6,912 human clones from our clone library of human and mouse cDNAs for the initial primary screen. The library consists of full-length cDNA clones in a plasmid vector suitable for in vitro protein production using a high-yield TnT system. Purified plasmid DNA had been previously prepared from these clones, an aliquot of which was used for the TnT reaction.

4 Automated Array Production (continued)

A batch process was used for array production. In brief, TnT reactions were carried out using 3 plasmid source plates at a time by adding TnT reaction mix containing biotin-lysyl-tRNA to the plate followed by a 2 hr incubation. Then a set of 24 streptavidin-linker conjugates was added to form specific protein-linker complexes. This was followed by pooling 24 individual protein-linker complexes. Thus, we generated 288 proteins from 3 plates of TnT reactions which were combined into 12 pools of 24 proteins coupled to unique linkers. These 12 protein-linker pools were each added to a column of wells to facilitate dispensing into array plates. Many of these steps were automated using an epMotion® automated pipetting system (Eppendorf®).

Next, the protein mixtures were distributed from the antigen source plate to 19 corresponding linker capture array plates using a CyBi®-Well dispenser (CyBio), generating 152 copies of each protein. The 25-spot linker array plates had previously been coated with 24 corresponding linkers and human IgG. The protein-linker complexes were allowed to bind to their corresponding linker on the plate for 2 hrs, generating the immobilized protein arrays. Following this binding step, the plates were washed, dried, and stored with desiccant. The schematic in Figure 4 illustrates the antigen arrays generated on 25 spots within the wells of a 96-well microtitre plate.

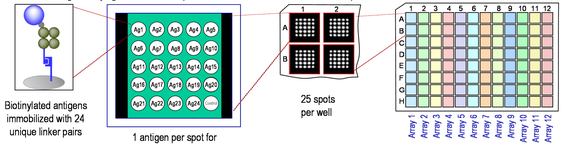


Figure 4. Schematic of Arrays

5 Screening

The primary screening of our protein arrays looked at a range of cancer serum and plasma samples. Samples were selected from our patient sample collection, which includes the following cancers: breast (benign breast disease, 40; *ER positive*, 39; DCIS, 23; invasive, 39), lung (bronchial dysplasia, 20; *squamous carcinoma, 40; adenocarcinoma, 40;* BAC, 24; SCLC, 20) , ovarian (benign, 40; *late stage premenopausal, 40; late stage postmenopausal, 40; early stage postmenopausal, 39;* mucinous, 38), prostate (BPH, 40; *low grade [Gleason 3–6] 20; high grade [Gleason 7–10] 16);* colon (20); kidney (30); melanoma (10); and pancreatic (10). Each major group was split into 2 sets for the initial screening. Sample set 1 included cancer subtypes in italics and normal (male, 40, and female, 60) serum and plasma samples. Samples were pooled, using 5 individual patients per pool and 2 pools per subtype. Prior to pooling, all patient samples were pre-screened to ensure that we used patient samples with low nonspecific reactivity to the arrays. We excluded from the pools any patient samples that produced elevated signals across all proteins in the arrays. Removal of these patient samples improved sensitivity in the primary screen, eliminating the possibility of a sample masking specific signals from other samples in the pool. Excluded samples were tested individually in a secondary screen.

To select candidate antigens for additional testing, primary screen data were analyzed as follows: after log transformation, the minimum ECL value of duplicate measurements was used. Minimum ECL values less than 10 were not significant compared to instrument noise; therefore, these values were replaced with 10. The log-transformed ECL values were normalized to remove any systematic biases associated with each measurement, taking into account signal variation among capture antigens and array batches. This was done separately for each of the 2 sample sets. Antigens were selected based on signals that exceeded specific thresholds according to any of the following criteria:

- Normalized signal of at least 1 disease sample pool above a set threshold.
- Normalized signal of all sample pools from the same disease group above a lower set threshold. This criterion was implemented to include antigens with weaker but consistent responses across sample pools.
- Analysis of an antigen's response profile across samples:
  - Sample set 1: Ratio between response in a disease pool and median response across all tested normal samples. This criterion was implemented to include antigens with response profiles that were highly selective for patients with specific cancers as compared to controls.
  - Sample set 2: Normal samples were not tested with this sample set. Instead, antigens with a wide response profile across different disease states were selected.

We selected 188 antigens from our primary screens (96 from sample set 1 and 92 from sample set 2) for secondary screening using individual patient samples.

The antigens selected from the primary screen were recovered from our source library of bacterial clones and cultured, and then new preparations of plasmid DNA were purified. These were added to new TnT reactions, and then new antigen array plates were generated using the methods described earlier. The array layout was altered compared to the primary screen to make it more conducive to screening large numbers of individual samples. This layout differed from Figure 4 in that 4 batches of arrays were produced with the same 24 proteins arrayed in all 96 wells of each plate within a batch. Nineteen plates of each batch were generated producing 1824 copies of each protein.

Individual patient samples from our collection, 716 in total, were screened. Since we had split the primary screen into 2 sections, we generated 2 sets of candidate antigens for screening against the patient sample collection. Each antigen set was only tested against cancer subtypes from the same section. From these results, candidate autoantigens were identified using a normalization approach. Median signals of triplicate measurements were normalized to account for variations between serum samples and days:

Normalized ECL = median ECL(sample, antigen)/median ECL(sample, array)*median ECL(all data)

6 Results

The ovarian cancer samples produced the best results, with the well-known cancer autoantigen TP53 producing the highest hit rate (Figure 5). This value agreed well with published data on autoantibodies to TP53. We were also able to confirm other previously reported autoantigens, such as DEK (Figure 6). In addition, we identified some new cancer-associated autoantigens such as GMPR2 (Figure 7). Unfortunately, most of the selected antigens had hit rates that were less than 20%, too low to be of clinical value. Our low observed autoantibody response hit rates were similar to those reported by other investigators. The presence of elevated autoreactive antibodies for many of these antigens in normal and benign samples further compromised their potential diagnostic value. Combining antigens did not provide a significant improvement in our ability to distinguish cancer patients from normal or benign conditions.

A few antigens appeared in the arrays as multiple distinct clones. Comparison of data from these clones gave an indication of the reproducibility of our antigen array approach. Individual sera that were reactive to an antigen typically had similar signals on both clones (e.g. GMPR2, Figure 7), suggesting good overall reproducibility.

Autoantigens selected for the secondary screen were verified by sequencing. The sequencing of 92 clones identified 5 incorrect sequences and 1 mixed (containing 1 correct and 1 unidentified sequence). This suggests that the clone library has an error rate of ~6%. The incorrect sequences matched clones expected to be elsewhere in the library, indicating that these errors were generated during its commercial production (Incyte Corp., CA).

To further evaluate our linker array format relative to direct immobilization of pure autoantigens, we obtained and compared 8 of our candidate antigens as recombinant proteins. This comparison demonstrated that the linker array approach had equivalent or better performance overall compared to directly coated purified antigens. In the case of TP53, a correlation analysis of the data demonstrated comparable performance of the approaches (Figure 8). This result was surprising, in part since we anticipated significantly higher signals with pure antigen and better resolution of positive reactivity from unreactive patient samples.

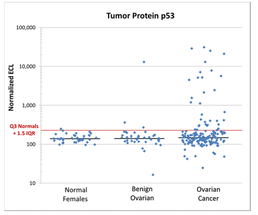


Figure 5. Autoreactive antibody responses in ovarian cancer patients to TP53 (p53) cancer autoantigen.

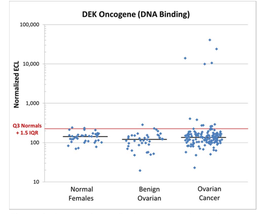


Figure 6. Autoreactive antibody responses in ovarian cancer patients to TP53 (p53) cancer autoantigen.

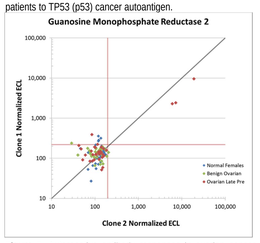


Figure 7. Autoreactive antibody responses in ovarian cancer patients to the GMPR2 cancer autoantigen (2 clones from the clone library).

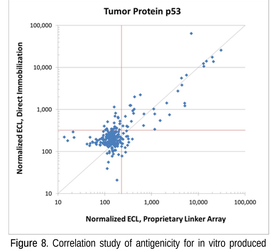


Figure 8. Correlation study of antigenicity for in vitro produced TP53 vs. recombinant TP53 using cancer and normal patient samples. Selected reactive and unreactive patient samples were used to evaluate relative antigenicity of TP53 produced in an in vitro TnT reaction and immobilized via our linker array approach vs. direct adsorption of recombinant TP53.

7 Conclusion

In conclusion, we have demonstrated the utility of MSD's MULTI-ARRAY technology in the discovery of autoreactive antibodies from cancer patient samples. This was achieved using an array of human proteins generated using a novel solution-addressable linker multiplexing approach and a proprietary human gene expression library.

With this system we were able to identify TP53, a previously reported autoantigen, and additional autoantigens in ovarian cancer. Unfortunately, the autoantibody responses in cancer patients were relatively infrequent. This low frequency has been observed in other studies and may be a reflection of the inherent heterogeneity of tumors. Better results might be obtained using a more homogenous selection of patient samples with molecularly classified tumors. The success of this approach might also be improved by the use of patient-tumor specific protein arrays which would include relevant mutated proteins.

Support: This work was supported in part by NCI through SBIR grant #R44CA130391.

Acknowledgments: We would like to thank Harvey I. Pass, MD, New York University School of Medicine, and Michael Tainsky, PhD, Wayne State University School of Medicine, for providing some of the samples used in our screen.

Meso Scale Discovery®
A division of Meso Scale Diagnostics, LLC.
www.mesoscale.com®

DOWNLOAD POSTER