Blood Test for Early Detection of Lung Cancer

Martin Stengelin,[1] Harvey Pass,[2] William Rom,[2] Sudeep Kumar,[1] Srikant Vaithilingam,[1] Anahit Aghvanyan,[1] Daisy Roy,[1] Gelu Dobrescu,[1] Ramou Sivakamasundari,[1] Eli N. Glezer,[1] and Jacob N. Wohlstadter[1]

[1]Meso Scale Diagnostics, LLC., Rockville, Maryland, USA, [2]New York University, School of Medicine, New York, NY, USA

1 Abstract

Lung cancer is the largest single cause of death from cancer worldwide. Even though lung cancer often can be treated successfully when detected early, approximately 90% of patients diagnosed with lung cancer ultimately succumb to the disease. Screening with low dose computerized tomography (CT) can reduce mortality, but the positive predictive value of this test is low, leading to a large number of suspicious but ultimately non-malignant results that nevertheless require follow-up. Our objective was to develop a simple blood test to risk-stratify patients at high risk of lung cancer.

We developed multiplexed, serum/plasma immunoassay panels to measure more than 40 lung cancer-related biomarkers using a 96-well, 7-spot format and electrochemiluminescence detection. Due to the high sensitivity of MSD's MULTI-ARRAY® technology, these panels were run with diluted serum or plasma, bringing the total sample volume required to run all 40 assays down to approximately 40 µL per replicate. This enabled us to measure all markers simultaneously in precious, high-quality serum and EDTA plasma samples. We used samples from early-stage lung cancer patients (drawn before lung cancer surgery) and from a lung-cancer screening cohort of age-matched heavy smokers who did not have lung cancer at the time of the blood draw.

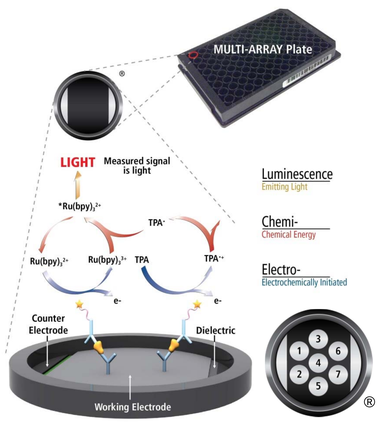
In a training set of 300 samples, 12 serum and 6 plasma markers had areas under an ROC curve (ROC areas) of 0.7 or higher. We used a logistic regression model with 100x cross-validation to develop a multi-marker panel. One serum panel (Flt-3L, EGFR, MMP-3, and NME-2) and one plasma panel (Flt-3L, cytokeratin-19, MMP-3, Flt-1, KGF, and PlGF) were selected and tested using approximately 250 additional samples from the same cohort. For the serum panel, the ROC area dropped to 0.85 (vs. 0.95 for the training set); for the plasma panel, the ROC area dropped to 0.81 (vs. 0.93). The ROC area of 0.85 for the serum panel with clinical sensitivity and specificity of 81% and 84%, respectively, and the ROC area of 0.81 for the plasma panel (with clinical sensitivity and specificity of 76% and 78%, respectively), are expected to be clinically useful despite this ROC area decrease.

Analysis of the combined training and test sets with 100x cross-validation resulted in a 4-marker serum panel (Flt-3L, EGFR, MMP-3, and NME-2) with an ROC area of 0.91 and clinical sensitivity and specificity of 88% and 82%, respectively, and a 5-marker plasma panel (Flt-3L, cytokeratin-19, Flt-1, KGF, and HGF) with an ROC area of 0.91 and clinical sensitivity and specificity of 84% and 83%, respectively.

Using MULTI-ARRAY technology and high quality clinical samples, we were able to identify promising biomarker panels for early detection of lung cancer in high-risk individuals.

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.

Assay Protocol

1. Add 200 µL MSD® Blocker A to each well. Incubate for 30 min at room temperature (RT).
2. Wash with PBS-T. Add 25 µL of assay diluent and 25 µL of diluted sample or calibrator. Incubate for 2 hours at RT with shaking. (Note: For the last three panels in the bottom left table [Ca 125, …, uPA], add 40 µL of assay diluent and 10 µL of non-diluted sample.) Wash with PBS-T. Add 25 µL of detection antibody. Incubate for 1 hour at RT.
3. Wash with PBS-T. Add 150 µL of Read Buffer T. Read on MSD imager. (Note: For the 3rd panel [Ca 15.3, …, OPN], use Read Buffer P.)

MULTI-ARRAY Panels for 47 Lung Cancer Serum/Plasma Biomarkers			
Assays	Sample Volume & Sample Dilution	Detection Limit (LOD)	Total CV
Adiponectin	25 µL 30 000x diluted	0.4 pg/mL	11%
CRP (C-reactive Protein)		6 pg/mL	23%
SAA (Serum Amyloid A)		6 pg/mL	10%
cMet (HGFR)	25 µL 50x diluted	8 pg/mL	20%
EGFR (Epidermal Growth Factor Receptor)		70 pg/mL	12%
Flt-3L (Flt-3 Ligand)		0.2 pg/mL	14%
Mesothelin		3 pg/mL	10%
S100A6		80 pg/mL	16%
TNFR-2 (Soluble Tumor Necrosis Factor Receptor 2)		0.5 pg/mL	13%
Ca 15.3 (Cancer Antigen 15.3)	25 µL 50x diluted	0.04 mU/mL	12%
Ca 50 (Cancer Antigen 50)		4 mU/mL	17%
cKit (SCF Soluble Receptor)		40 pg/mL	19%
E-Cadherin		1 pg/mL	12%
MMP-3 (Matrix Metalloproteinase 3)		3pg/mL	12%
OPN (Osteopontin)		10 pg/mL	15%
Ca 19-9 (Cancer Antigen 19-9)	25 µL 10x diluted	0.1 U/mL	9%
ErbB2 (Her2, Neu)		0.3 pg/mL	7%
HGF (Hepatocyte Growth Factor)		0.6 pg/mL	12%
M-CSF (Macrophage Colony-Stimulating Factor)		1 pg/mL	12%
MMP-2 (Matrix Metalloproteinase 2)		200 pg/mL	8%
MMP-9 (Matrix Metalloproteinase 9)		100 pg/mL	21%
AFP (Alpha Feto-Protein)	25 µL 5x diluted	20 pg/mL	11%
B7-H3 (B7 Homolog 3)		2 pg/mL	12%
IL-2R (Soluble IL-2 Receptor 2)		6 pg/mL	13%
MCP-1 (CCL2)		1 pg/mL	9%
Nectin-4		0.5 pg/mL	10%
NME-2 (NME/NM23 Nucleoside Diphosphate Kinase 2)		200 pg/mL	14%
VEGF-D (Vascular Endothelial Growth Factor D)		2 pg/mL	10%
CA125 (Cancer Antigen 125)	10 µL neat	0.1 U/mL	13%
Cytokeratin-8		3 ng/mL	15%
Dkk-1 (Dickkopf-related 1)		50 pg/mL	14%
Flt-1 (VEGF Receptor 1)		1 pg/mL	12%
PlGF (Placental Growth Factor)		1 pg/mL	19%
VEGF (Vascular Endothelial Growth Factor A)		6 pg/mL	13%
Cytokeratin-19 (Cyfra 21.1)	10 µL neat	80 pg/mL	16%
IL-6 (Interleukin 6)		0.5 pg/mL	16%
ITAC (CXCL11)		7 pg/mL	28%
KGF (Keratinocyte Growth Factor; FGF-7)		0.3 pg/mL	11%
NSE (Neuron-Specific Enolase)		30 pg/mL	23%
OPG (Osteoprotegerin)		8 pg/mL	13%
SCF (Stem Cell Factor)		1 pg/mL	18%
CEA (Carcinoembryonic Antigen)	10 µL neat	20 pg/mL	14%
GPI (Glucose Phosphate Isomerase)		10 ng/mL	17%
MDC (CCL22)		20 pg/mL	22%
uPA (Urokinase Plasminogen Activator)		30 mU/mL	13%

Lung Cancer Panel Development Process	
1	Select 46 serum/plasma markers expected to be relevant for lung cancer.
2	Test training set: ~300 samples; equal number of serum, EDTA plasma, cases, controls.
3	Develop a primary and a backup serum and plasma panel.
4	Validation set: ~250 samples (from the same cohort, not used for training).
5	Calculate ROC of validation set using cut-offs from training set.
6	Calculate ROC of best serum and plasma panels using the combined data set.

Clinical Samples	Cases (n=200)	Controls (n=200)
Selection criteria	Early stage lung cancer patients. Samples were drawn just before lung cancer surgery.	Cohort of individals at high risk of lung cancer: >20 pack years of smoking, >50years, no current or past cancer.
Age (Average ± SD)	69 ± 10	69 ± 10
Smoking status	18% current; 82% ex-smoker	17% current; 82% ex-smoker
Pack years (median ± IQR)	35 ± 41	35 ± 45
Gender	44% male, 56% female	44% male, 56% female
Ethnicity	84% caucasian, 7% asian, 3% hispanic	84% caucasian, 7% asian, 4% hispanic
Cancer Stage	49% IA, 28% IB, 4% IIA, 7% IIB, 9% IIIA, 3% IIIB, 1% IV	n/a

One 150 µL serum and one 150 µL EDTA plasma aliquot was available from each of the 400 individuals (200 cases, 200 matched controls). This was sufficient to test all 47 biomarkers in duplicates. Each plate had an approximately equal number of cases and controls, and samples were run and analyzed blinded. Serum and plasma samples were selected and analyzed independently.

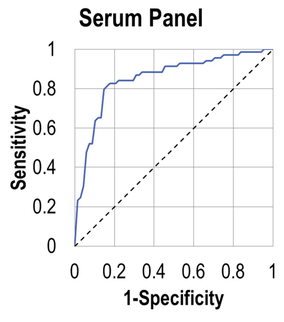
Data Analysis

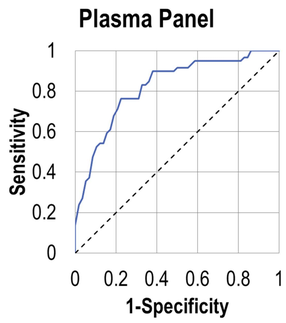
- Use **logistic regression** to model combinations of assays

$$\Pr(Y_i = Diseased) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_N X_{iN})}}$$

  - Y_i is the observed response in the i^{th} trial
  - X_{i1}, X_{i2}, X_{i3}, … , X_{iN} are the known Biomarker concentrations in the i^{th} trial
  - 1,2,3, …, N represent the N biomarkers in the model
- Solve for parameters B_0 , β_1 , β_2 , β_3 , …, β_N using **Training** data
- Test created model on **Validation** data
- Construct ROC (Receiver Operating Characteristic) curve
- **For each model, perform 100-fold cross-validation:**
  - Randomly chose 2/3 of cases and 2/3 of controls
  - Develop algorithm
  - Calculate ROC with remaining 1/3 of data
  - Repeat 100x, calculate confidence intervals

3 ROC Analysis of Training and Validation Set

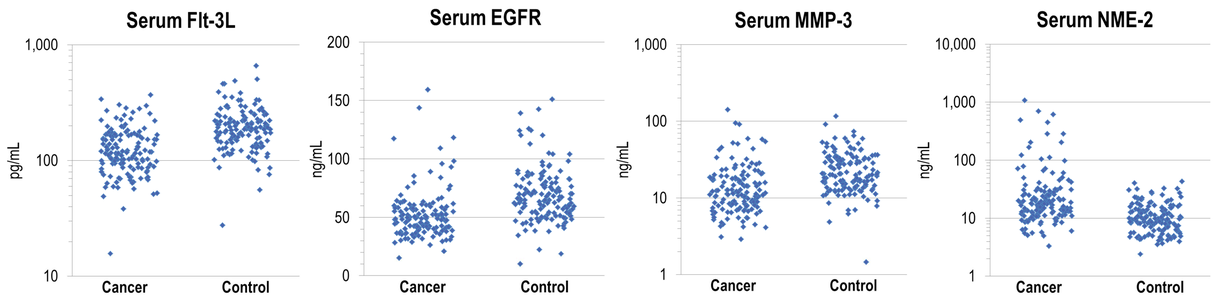


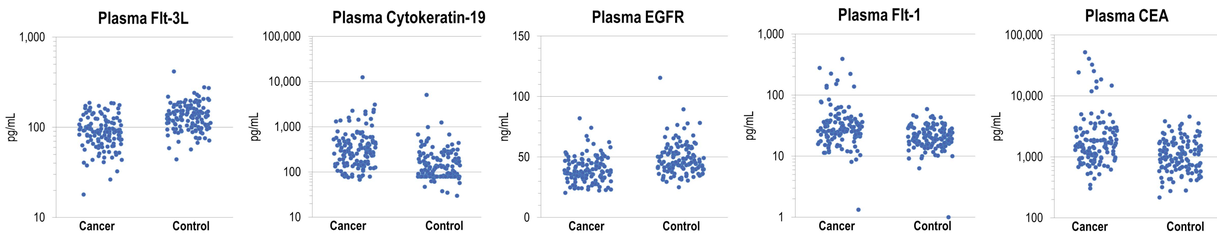


Panel	Markers	Training ROC auc	Training Clinical Sensitivity, Specificity	Validation ROC auc	Validation Clinical Sensitivity, Specificity
Primary Serum	Flt-3L, EGFR, MMP-3, NME-2	0.945	95%, 85%	0.846	81%, 84%
Back-up Serum	Flt-3L, MMP-3, NME-2, Osteopontin	0.954	93%, 87%	0.847	77%, 82%
Primary Plasma	Flt-3L, Cytokeratin-19, MMP-3, Flt-1, PlGF	0.929	83%, 94%	0.814	76%, 78%
Back-up Plasma	Flt-3L, Cytokeratin-19, EGFR, MMP-3	0.889	73%, 93%	0.792	81%, 69%

Serum biomarker concentrations of a training set (73 cases, 73 controls) were analyzed as described previously. Two panels were selected: a primary 4-marker panel and a 5-marker panel as back-up. Next, serum biomarker concentrations of a validation set of 69 cases and 68 controls were measured. An ROC analysis was performed for these two panels using the algorithm determined with the training set. The table shows Clinical sensitivity and specificity of the serum training and validation set, and an analogous analysis for plasma samples (73/73 samples for training, 59/58 for validation).

4 Best Markers and Panels from Combined Data Set



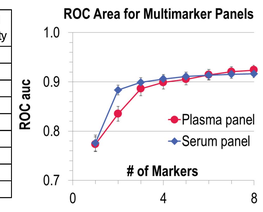


Single Marker; Serum	ROC auc
Flt-3L	0.776
EGFR	0.765
NME-2	0.754
MMP-3	0.733
ErbB2	0.727
cKit	0.689
MMP-2	0.687
CytoKeratin-19	0.679
S100A6	0.676

4-Marker Panel; Serum	ROC auc	Clinical Specificity	Clinical Sensitivity
Flt-3L + EGFR + NME-2 + MMP-3	0.906	82%	88%
Flt-3L + NME-2 + MMP-3 + ErbB2	0.905	75%	94%
Flt-3L + NME-2 + MMP-3 + IL-6	0.904	76%	92%
Flt-3L + NME-2 + MMP-3 + CytoKeratin-19	0.904	77%	90%
Flt-3L + NME-2 + MMP-3 + AFP	0.903	81%	87%
Flt-3L + NME-2 + MMP-3 + KGF	0.903	76%	93%
Flt-3L + NME-2 + CytoKeratin-19 + KGF	0.902	79%	91%
Flt-3L + NME-2 + ErbB2 + CytoKeratin-19	0.902	81%	87%
Flt-3L + NME-2 + MMP-3 + NSE	0.902	76%	92%

Single Marker, Plasma	ROC auc
Flt-3L	0.773
CytoKeratin-19	0.757
EGFR	0.702
Flt-1	0.689
CEA	0.668
MMP-3	0.667
Cytokeratin-8	0.661
cKit	0.661
HGF	0.648

5-Marker panel, Plasma	ROC auc	Clinical Specificity	Clinical Sensitivity
Flt-3L + CytoKeratin-19 + Flt-1 + HGF + KGF	0.905	83%	84%
CytoKeratin-19 + EGFR + Flt-1 + CEA + KGF	0.905	90%	76%
Flt-3L + CytoKeratin-19 + Flt-1 + CEA + KGF	0.904	90%	79%
Flt-3L + CytoKeratin-19 + Flt-1 + KGF + IL2-R	0.903	84%	85%
CytoKeratin-19 + EGFR + Flt-1 + HGF + KGF	0.902	82%	86%
Flt-3L + CytoKeratin-19 + Flt-1 + Cytokeratin-8 + KGF	0.902	86%	84%
Flt-3L + CytoKeratin-19 + EGFR + Flt-1 + KGF	0.901	85%	83%
Flt-3L + CytoKeratin-19 + Flt-1 + KGF + MMP-9	0.900	81%	86%
Flt-3L + CytoKeratin-19 + Flt-1 + KGF + Osteopontin	0.900	86%	83%



Data from the training and validation sets were pooled and analyzed as described earlier (with 100-fold cross-validation). The table shows ROC areas for the best individual serum and plasma markers. The ROC area for multimarker panels was 0.9 or higher for 4-marker serum panels and for 5-marker plasma panels. The table illustrates that several panels have excellent ROC areas. Adding additional markers to the panels improves the ROC area only slightly.

5 Conclusion

- MULTI-ARRAY panels containing 45 lung cancer-related markers were developed and analytically validated.
- The high sensitivity, dynamic range, and multiplexing of MSD assays allowed all 45 markers to be measured in a 40 µL sample volume.
- Approximately 550 serum and EDTA-plasma samples from early stage lung cancer patients and from heavy smokers who do not have lung cancer were run on these panels in several batches over a period of nine months.
- An initial training set of ~150 serum samples and ~150 EDTA plasma samples was used to develop a serum algorithm and a plasma algorithm to separate cases from controls. These algorithms were then validated using an independent validation set (~250 samples).
- The ROC area for the serum validation set was 0.85, and for the plasma validation set 0.81.
- Using the entire data set for both training and validation (including cross-validation), the ROC area for the best 4-marker serum panel (containing Flt-3L, EGFR, NME-2, and MMP-3) was 0.91.
- Using the entire data set for both training and validation (including cross-validation), the ROC area for the best 5-marker plasma panel (containing Flt-3L, Cytokeratin-19, Flt-1, HGF & KGF) was 0.91.

Meso Scale Diagnostics, LLC.
www.mesoscale.com®

This project has been funded in whole or in part with Federal funds from the National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN261201000018C.