Using Targeted Mass Spectrometry to Identify a Plasma-based Protein Marker Panel for Colorectal Cancer Risk Determination
A Novel Approach to Proteomics and Cancer

OUTLOOK PHYSICAL SCIENTISTS TAKE ON CANCER

The odd couple

An unlikely duo is trying to make sense of the avalanche of data that confronts cancer scientists, pointing the way towards a new era of research.

BY ERIK VANCE

In March 2004, oncologist David Agus was leaving his hospital in West Hollywood, California, at the end of a long day. On his way to his car, he passed by a bookstore. The cover picture was of an ominous, translucent cancer cell overlaid by the words: “Why We’re Losing the War on Cancer.” The story took the entire cancer research community by storm. Now, three years later, Agus and his team at the University of Southern California in Los Angeles are changing the game. They’ve discovered a way to detect cancer at an early stage, using a combination of genetic and proteomic analysis.

API was founded by David Agus, MD, a medical oncologist, and W. Daniel Hillis, PhD, a computer scientist.

“We are obligated to do things differently because the current method hasn’t worked.”

“There is this notion that this is the century of biology — well, that’s poppycock. This is the century of the convergence of the sciences.”
The Conversation of the Body as an Early Warning System

Wait until you are sick

Diagnose and treat early
This is What the Conversation Looks Like to API

Plasma LC - Mass Spectrometry Data

Retention Time

Mass/Charge Ratio
The Evolving Dx Paradigm – From Silo to System

**Past**

- Discovery
- Test

**Present**

- Discovery
- Test

**Future**

- Discovery
- Test

*The Discovery IS the Test*
The Image of Healthcare – Total System Viewing

Observe All of the Differences
All of the Time
API’s Dual Track Platform – \textit{de novo} and Targeted Mass Spec

**Novel Content Identification**

- **Wide-angle View**
  - >300,000 measures of 100’s of samples

**Targeted Content Testing**

- **Laser-focused View**
  - 10’s – 100’s of measures on 1000’s of samples
Discovery Platform – Lab and Informatics Fully Integrated

Sample → Delipidation Protein Assay → Abundant Protein Immunodepletion → Intact Protein Fractionation → TFE Trypsin Digestion

LCMS Quantification → Feature Detection → Classifier Analysis → Feature Identification → Biomarker Panel
Proprietary process tracking software follows all factors upstream of a measurement (e.g., reagents, protocols, instruments, QC, etc.)
Exceptional Reproducibility Achieved

Two LCMS runs
Same sample
5 days apart
Red and green overlap to yellow

Position and intensities overlap

Log2(Area): Day 1
Log2(Area): Day 5
-4 -2 0 2 4 6
-4 -2 0 2 4 6
Processing and Analysis Overview: Sample Data to Biomarkers

Lab

Single Injection Processing

Cross Injection Processing

Classification Analysis

Feature Identification

Process Raw Data, Store Results

Database

Network File System

Data Matrix

Cluster

FIB

Normalize

10 x 10-fold Cross Validation

ROC Curve

Discriminating Features

Verification, Test Development, …
Platform Breadth – Discovery to Implementation

**Applied Proteomics**

- **Marker Discovery**
  - *de novo* MS
  - Targeted MS
  - Proteins
  - Metabolites

- **Commercial Test Development**
  - Targeted MS

- **Clinical Validation**
  - Prospective Samples
  - Retrospective Samples

- **Test Implementation**
  - CLIA/LDT (IVD)
Colorectal Cancer: A Huge Opportunity via Prevention

A deadly, yet preventable cancer with significant treatment costs attached to treatment.

CRC deaths reduced by 53% with colonoscopy screening and removal of adenomatous polyps
End of An Illness – Early Detection and Treatment

<table>
<thead>
<tr>
<th>Disease Development Timeline</th>
<th>Year 0-2</th>
<th>Year 2-4</th>
<th>Year 4-6</th>
<th>Year 6-8</th>
<th>Year 8-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Stages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**API Early Risk Assessment**
- Pin prick sample
- Results in minutes
- Reduces treatment cost
- Saves lives

**Current Dx**
**Higher Risk/More Deaths**
- Current diagnostics rely on existing disease
- Detect cancer in later stages which means higher risk of death and higher treatment costs

Emphasis on Prevention, Rather Than Just Detection, of Colorectal Cancer
48% of people avoid screening, thereby increasing their cancer risk.

![Diagram showing PAYORS and cost breakdowns]

- **Eligible for Screening**: 88M
- **Non-Compliant**: 42M
- **Undiagnosed**: 13M

**API Early Assessment**
- **Prevented Cases**: 390,000
- **Lives Saved**: 250,000
- **Costs Saved**: $27B
- **Costs Avoided**: $7B
- **Treatment Costs**: $27B
- **Cases Prevented**: 100,000
- **Lives Lost**: 70,000

**PAYORS** in population of 1M over 5 years:
- **API**: $679M (Treatment Cost: $622M, Diagnostic Cost: $17M, Screening Cost: $2M)
- **Unscreened**: $652M (Treatment Cost: $600M, Diagnostic Cost: $42M, Screening Cost: $10M)

**Notes**:
- Assumes 30% compliance
- Treatment Cost: $200M, Diagnostic Cost: $42M, Screening Cost: $2M
Colonoscopy-Based Sample Collection – “All Comers”

**Project Plan**

**Inclusion and Exclusion Criteria**

The following inclusion and exclusion criteria are to be followed for all subjects enrolled in the Colonoscopy Clinical Specimen Collection Study (Project # 039-10-01).

**Enrollment Target:** 300 Subjects

**Inclusion Criteria:**
1. Subjects must be at least 18 years of age.
2. Subjects must be undergoing a colonoscopy.
3. Subjects must be willing to participate in this study by signing the Informed Consent Form.

**Exclusion Criteria:**
1. Those subjects not meeting the inclusion criteria.

**Study Documentation:**

A Colonoscopy Case Report Form must be completed for each subject enrolled.

The following source documentation should be provided with the completed Case Report Form:

- Laboratory Report(s) for diagnostic testing results recorded in CRF
- Colonoscopy Report
- Biopsy report (if applicable)

Ensure that the patient’s name does not appear on any source documentation provided with the Case Report Form. Write the corresponding Subject ID # on all source documentation provided with the Case Report Form.

All subjects enrolled must be recorded in the Enrollment Log and Shipping and Accountability Log.

**Specimen Collection / Processing**

**NOTE:** Specimens need to be collected on the day of the colonoscopy procedure, before any anesthesia is administered.

1. Assign the enrolled subject a Subject ID #.
2. Affix the appropriate label on each of the following:
   - Consent Form – Affix a Subject ID # label
   - Colonoscopy Case Report Form – Affix a Subject ID # label on each page
   - 10 ml Polypropylene Storage Tube – Affix a Subject ID # - Plasma label
   - PAXgene Blood DNA Tube – Affix a Subject ID # - PAXgene label
   - BD P100 Collection Tube – Affix a Subject ID# label
3. QC all Labeled material to ensure that the same SUBJECT ID # appears on all elements.
4. Draw whole blood into each of the following tubes utilizing a butterfly needle provided. Collect in the order listed. Ensure that the PAXgene tube is held completely upright while collecting (in order to eliminate backflow).
   1 – BD P100 Collection Tube (8.5 ml)
   1 – PAXgene Blood DNA Tube (8.5 ml)
5. Mix each tube by gently inverting the tube 8 – 10 times (DO NOT SHAKE).

**BD 100 Collection Tube**

1. Centrifuge tubes for 30 minutes +/- 5 minutes at 3500 RPM (1318g).
2. Transfer Plasma from the BD P100 collection tube into 10 ml Polypropylene Storage Tube labeled with the corresponding Subject ID # and Plasma.

   **NOTE:** The minimum volume requirement is 3 ml of Plasma.

3. Place storage tube in the appropriate storage box provided.
4. Place Storage Boxes in Freezer (≤ -70°C) within 4 hours from the time of collection.

**PAXgene Blood DNA Tube**

1. Place PAXgene Blood DNA collection tube into -20°C storage following collection.
2. PAXgene tube must remain in -20°C storage for AT LEAST 24 hours, then transferred into -70°C storage.

Record the following information in the CASE REPORT FORM:

- Specimen Matrix
- Centrifugation times
- Freezer ID# and Temperature
- Collection date and time
- Date and times placed in storage
- Initials for collection and processing
6 mm polyp found in cecum

Biopsy identifies as tubular adenoma
From Interesting Proteins to Transition Measurement

Chemistry

Protein

Peptides

Precursor Ion

Fragment Ions

Transitions

Tryptic Digestion

MS1 Ion Selection

Collision Induced Dissociation

MS2 Ion Selection

QTOF MS

QQQ MS

DPTFIP APIQAK ANGT_HUMAN crc 188
ms1 649.36 - ms2 461.199

From Interesting Proteins to Transition Measurement

Chemsitry

Protein

Peptides

Precursor Ion

Fragment Ions

Transitions

Tryptic Digestion

MS1 Ion Selection

Collision Induced Dissociation

MS2 Ion Selection

QTOF MS

QQQ MS

DPTFIP APIQAK ANGT_HUMAN crc 188
ms1 649.36 - ms2 461.199

From Interesting Proteins to Transition Measurement

Chemsitry

Protein

Peptides

Precursor Ion

Fragment Ions

Transitions

Tryptic Digestion

MS1 Ion Selection

Collision Induced Dissociation

MS2 Ion Selection

QTOF MS

QQQ MS

DPTFIP APIQAK ANGT_HUMAN crc 188
ms1 649.36 - ms2 461.199
Developing Simultaneous Assays for 188 Proteins

- Proteins with Good Evidence of CRC Connection – Human, Plasma/Serum
- Isoforms
- Tryptic Peptides

Proteotypic Peptide Selection
- Homology
- Post-translational Modifications

Synthesize Best Peptides for Evaluation
- Model Predicted Signal Observed in Literature

Evaluation in QTOF MS for Fragmentation

Evaluation in QQQ MS for Quantitation

Goal -
- 2 Peptides/Protein
- 2 Transitions/Peptide

Covers 188 Proteins and 309 Isoforms
- 1,348 Transitions
- Heavy and Light

30 Minute LCMS Run
Quantitative Data are Obtained: MS-based “ELISAs”

APOA1 Peptide - VSFLSALEEYTK

![Graph showing elution times and signal intensities for APOA1_HUMAN VSFLSALEEYTK y8 peptide with C13 Reference Peptide and Endogenous Peptide.](graph.png)

**Heavy, Isotope-Labeled Reference Peptide**

**Light, Endogenous Peptide**

**Signal Intensity**

**Elution Time (min)**
Current Advanced Adenoma Classifier Performance – AUC 0.81

- Ten-times 10-fold cross-validation
- 128 prospectively collected Advanced Adenoma samples and controls
  - ≥ 1 cm Adenoma
  - Villous or Tubulovillous Adenoma
- 4 proteins comprise classifier

![ROC curve with AUC 0.81 and annotations for stool tests and random predictions.](image-url)
Study Analytical Design for Initial CRC Classifier Definition

- **Discovery (138 Samples)**
  - Proteogenex
  - Asterand
  - Cap. Bio

- **Validation (136 Samples)**
  - Proteogenex
  - Asterand
  - Cap. Bio

- All Data (274 Samples)
  - Proteogenex
  - Asterand
  - Cap. Bio

- **CRC samples and controls obtained from three different sources**
- **Samples merged to dilute individual set biases**
- **Total sample pool divided into Discovery and Validation sets**
- **Discovery set used in 10-fold cross-validation to define classifier**
- **Validation set held out for final test**
CRC Discovery Performance and Validation Confirmation

138 Samples
Training AUC = 0.82

136 Samples
Validation AUC = 0.91

15 MRM transitions measured from 13 proteins
Random Forest Model
Current CRC Classifier Performance – AUC 0.91

- Good performance across all CRC stages for typical classifiers:

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Misclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>12 of 68 (18%)</td>
</tr>
<tr>
<td>CRC Stage I</td>
<td>2 of 15 (13%)</td>
</tr>
<tr>
<td>CRC Stage II</td>
<td>5 of 33 (15%)</td>
</tr>
<tr>
<td>CRC Stage III</td>
<td>5 of 13 (38%)</td>
</tr>
<tr>
<td>CRC Stage IV</td>
<td>0 of 2 (0%)</td>
</tr>
<tr>
<td>CRC, stage unknown</td>
<td>2 of 5 (40%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>26 of 136 (19%)</strong></td>
</tr>
</tbody>
</table>

- Good performance at all sites for typical classifiers:

<table>
<thead>
<tr>
<th>Sample Set</th>
<th>Misclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asterand</td>
<td>7 of 48 (15%)</td>
</tr>
<tr>
<td>Capital Bio</td>
<td>10 of 40 (25%)</td>
</tr>
<tr>
<td>Proteogenex</td>
<td>9 of 48 (19%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>26 of 136 (19%)</strong></td>
</tr>
</tbody>
</table>
Not All Protein Combinations are Equal

- Measuring all proteins in all samples allows brute-force testing of all combinations
- Plot is test vs. train AUC for all possible “Choose 2” combinations
- Non-random distribution indicates that some proteins and combinations are significantly better

350 features, 2 features/classifier = 61,075 possibilities
Next Steps on the Path to Implementation

1. CRC classifier definition
   - Retrospective samples; Asterand, Proteogenex, CapitalBio

2. Advanced adenoma classifier definition
   - Described
   - Prospective samples; ProMedDx

3. Commercial assay format development
   - SISCAPA or ELISA panel/multiplex
     - Ongoing

4. Verification

5. Validation
   - Prospective samples; DKFZ, Health Decisions (API study), EDRN

6. Launch
   - CLIA/LDT ➔ IVD
Parting Reminder…and Thank You!

**Discovery**

**Past**

**Present**

**Future**

The Discovery IS the Test