CLINICAL PROTEOMIC BIOMARKER DISCOVERY AND TRANSLATION FOR PATHOLOGY

Ed Nice

Head, Clinical Biomarker Discovery and Validation
Department of Biochemistry and Molecular Biology
Monash University
AUSTRALIA

and

Visiting Professor, Sichuan University and West China Hospital Chengdu
CHINA
(ed.nice@monash.edu)
Format of Main Presentation

1. The Human Proteome Organisation (HUPO) and the role of Pathology

2. Faecal proteomics: biomarkers for colorectal cancer

3. The microbiome in health and disease

4. Omics technologies: towards personalised/precision medicine
The Human Proteome Organisation (HUPO)

(HUPO) is an international scientific organization established in 2001 representing and promoting proteomics through international cooperation and collaborations and fostering the development of new technologies, techniques and training. It is designed to systematically map the entire human proteome using currently available and emerging techniques. Completion of the human proteome will enhance understanding of human biology at the cellular level and lay a foundation for development of diagnostic, prognostic, therapeutic, and preventive medical applications.

Human Proteome Project Mission Statement
By characterizing all 20,230 genes of the known genome, the Human Proteome Project will generate the map of the protein based molecular architecture of the human body and become a resource to help elucidate biological and molecular function and advance diagnosis and treatment of diseases.
HUPO Human Proteome Metrics 2018

17,470 PE1/2 proteins. 86% Completed
The Missing Proteins

Journal of Proteome Research

neXtProt PE2-4 Missing Proteins

- Olfactory (GPCR) Receptors
- Transmembrane Proteins incl. GPCRs, Taste Receptors & Solute Carriers
- Zinc Finger Proteins
- Homeobox Proteins
- Keratin-Associated Proteins
- Coiled-Coil Domain Proteins

2013: Blue
2016: Brown
2017: Green
2018: Red
Interfacing Pathology into the HPP

Future knowledge on proteins: structured

2015

2010

Present knowledge on proteins: segmented

Disease or Biology driven projects

Brain trauma...
Leukemia
Stoke

Stem cells
Neurological disorders
Secreted proteins
Cardiovascular disorders

Metabolic pathways
Membrane receptors
Diabetes
Cardiac diseases
Cancer

Cell signaling

“Adopt-a-chromosome” groups

(C-HPP)

(BD-HPP)
The BD-HPP Initiatives

- Cancer
- Cardiovascular
- Diabetes
- Extreme conditions
- EyeOme
- Food and Nutrition
- Glycoproteomics
- Immunepeptidome
- Infectious diseases
- Kidney and urine
- Liver
- Mitochondria
- Model organisms
- Musculoskeletal
- Protein aggregation
- Plasma
- Rheumatic disorders
- Brain

Biology and Disease-driven Human Proteome Project
Pathology and “Omics” Are Central to Advancing Personalized Medicine

Clinical Outcomes
- Diagnosis
- Prognosis
- Surveillance
- Treatment

Personalized Medicine
- Right Drug
- Right Dose
- Right Patient

Big Data
- Collect
- Store
- Analyze
- Integrate
- Interpret
- Disseminate

Histopathology
Immunohistochemistry
Imaging technologies
Molecular techniques
Clinical Data
Family History
Proteomics
Genomics
Epigenomics
Transcriptomics
Metabolomics

Pathology and the Proteomics Pipeline
The Bidirectional Interaction Between Proteomics and Pathology

- Discover and validate candidate biomarkers to assist in diagnosis, prognosis, surveillance and individualized patient medication
- Assay Development

- Assist with specific tissue identification and selection (e.g. laser capture microdissection, tissue sections for MS imaging)
- Validation of emerging automated histopathology techniques
- Classifying relevant patient medical information
- Optimization of experimental design for clinical trials
- Biobanking
Faecal Proteomics and Colorectal Cancer
Unmet CRC Clinical Needs

Sensitive and specific biomarkers for early detection at a time when the disease is curable by simple surgical resection

Prediction of metastastic potential in Stage B CRC

Prediction of drug response

Faecal Proteomics: The Hypothesis

Stool samples will contain signature proteins and peptides relating to GI Tract pathology (e.g., haemoglobin peptides due to bleeding, CEA). These peptides may be relatively more abundant in stools than in blood. Stool samples will be easier to assay than blood where proteome is dominated by a number of abundant “housekeeping” proteins.

<table>
<thead>
<tr>
<th></th>
<th>CEA in Faeces (ng/mg stool)</th>
<th>CEA in Serum (ng/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>45.2 +/- 63.8</td>
<td>8.9 +/- 13.3</td>
</tr>
<tr>
<td>Normal</td>
<td>3.7 +/- 3.5</td>
<td>1.2 +/- 1.0</td>
</tr>
</tbody>
</table>


May be present due to leakage, secretion or exfoliation.
Advantages of Stool Testing

Non-invasive

Does not require trained staff

Can be performed at home

No bowel preparation required

Large biomass: not sample limited

“Samples” whole length of the colon

Reluctance to handle faecal material results in low compliance: education is needed

Normalisation is required for variable sample size
A New Paradigm for Faecal Biomarker Discovery and Validation

Faecal Sample

Multidimensional Fractionation

AIM: Mine as deeply as possible into the proteome

nLC-MS/MS Analysis

AIM: Obtain quality ms/ms data for biomarker discovery

Peptide MS/MS Faecal Library

AIM: Establish library of PROTEOTYPIC FAECAL PEPTIDES for MRM transitions or DIA

Analytical MRM

AIM: Confirm/optimise biomarker panel on small clinical test set

Quantitative MRM or DIA Validation

AIM: Take panel of validated markers for detailed clinical analysis
LICR Faecal MS/MS Library

Initial library based on 10 CRC patients and 5 normals. Currently contains MS/MS data from faecal samples for 650 proteins (from both normals and CRC patients)

>3500 proteotypic peptides

Initial data generated on LCQ Deca and LTQ-Orbitrap. Library has been regularly updated from both our own studies and the literature (eg Verberkmoes et al, ISME, 2008, 1 – 11, HUPO BDD Initiative, Bosch et al. Ann Intern Med. 2017 Nov 21)

These data have been used to identify specific peptide precursors and MS/MS transitions for directed quantitative MRM analysis

Now also using SWATH Data Independent Analysis
Many Colon Cancer Associated Proteins (CCAP) Are Present in the MS/MS Library

More than 120 CCAP have already been identified in our library. These include a number of blood-related proteins (e.g. hemoglobin, haptoglobin, lactoferrin), low abundance proteins such as CEA, MMP9 and putative adenoma markers including lipocalin, MMP 9, Cadherin 17.
Stool-Based Protein Biomarkers for Improved Colorectal Cancer Screening

• Large case-controlled study
• 834 human proteins were identified
• 29 were statistically significantly enriched in CRC versus control
• Combinations of 4 proteins reached sensitivities of 80% and 45% for detecting CRC and advanced adenomas at 95% specificity
• Many proteins common to our data set

Quantitative MRM Analysis of CCAP in Unfractionated Faecal Samples from a Patient with CRC
Optimised Sample Preparation

1) 100mg of stool extracted with 5 volume of 0.15% TFA

2) Manual disruption plus sonication (2x 30sec), centrifuge 16000g for 10min and retain supernatant

3) Boil supernatant (3 min at 95C) followed by acetone precipitation overnight (20C)

4) Resuspend in 8M urea, 50mM Tris HCl, pH 8.0

5) Reduce with 2.5mM DTT, 30min 37C

6) Alkylate with 10mM iodoacetamide, 30min dark, RT

7) Dilute to 1M Urea with ultra pure water

8) Digest O/N with sequence grade modified trypsin (1:20)

7) SPE clean up (Waters Sep-Pak C18 cartridges)

8) Lyophilize and redissolve in injection buffer
Multiplex MRM Screening on 9 CRC and 7 Normals

- Using **unfractionated samples**

- Screening of 73 proteins in a single assay (min 3 transition each)

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Protein Name</th>
<th>Protein Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid sphingomyelinase-like phosphodiesterase 3b</td>
<td>Elastase-3A</td>
<td>Neutrophil defensin 3</td>
</tr>
<tr>
<td>Actin, beta (Fragment)</td>
<td>Fibrinogen gamma</td>
<td>Neutrophil gelatinase-associated lipocalin</td>
</tr>
<tr>
<td>Albumin</td>
<td>Galectin-3-binding protein</td>
<td>Pancreatic amylase</td>
</tr>
<tr>
<td>alpha 1-antitrypsin</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>Pancreatic secretory granule membrane</td>
</tr>
<tr>
<td>Alpha-1-acid glycoprotein 1</td>
<td>Haptoglobin</td>
<td>major glycoprotein GP2</td>
</tr>
<tr>
<td>Alpha-1-antichymotrypsin</td>
<td>Hemoglobin alpha</td>
<td>Peroxiredoxin-2</td>
</tr>
<tr>
<td>Alpha-2-HS-glycoprotein</td>
<td>Hemoglobin beta</td>
<td>Plasmin-2</td>
</tr>
<tr>
<td>Alpha-2-macroglobulin amylase</td>
<td>Hemoglobin subunit gamma</td>
<td>Polymeric immunoglobulin receptor</td>
</tr>
<tr>
<td>B-RAF</td>
<td>hemopexin</td>
<td>Profilin-1</td>
</tr>
<tr>
<td>Cadherin 17</td>
<td>Ig alpha-1 chain C</td>
<td>Protease, serine, 2</td>
</tr>
<tr>
<td>Carbonic anhydrase 1</td>
<td>IgGFc-binding protein</td>
<td>Protein S100-A11</td>
</tr>
<tr>
<td>Carboxypeptidase A1</td>
<td>IGKC protein</td>
<td>Protein S100-A6</td>
</tr>
<tr>
<td>Cartilage oligomeric matrix protein catalase</td>
<td>Immunoglobulin J chain</td>
<td>Protein S100A8</td>
</tr>
<tr>
<td>CEACAM1</td>
<td>Intelectin 1</td>
<td>Protein S100-A9</td>
</tr>
<tr>
<td>CEACAM5</td>
<td>Kallikrein-1</td>
<td>Protein S100-P</td>
</tr>
<tr>
<td>CEACAM6</td>
<td>Kininogen-1</td>
<td>Protocadherin LKC (24)</td>
</tr>
<tr>
<td>CEACAM7</td>
<td>Lactotrans</td>
<td>Selenium-binding protein 1</td>
</tr>
<tr>
<td>Ceruloplasmin</td>
<td>Lactotransferrin</td>
<td>Transaldolase</td>
</tr>
<tr>
<td>Chloride channel, calcium activated, family member 1</td>
<td>Leukocyte elastase inhibitor</td>
<td>transthyrin</td>
</tr>
<tr>
<td>Complement C3</td>
<td>Leukotirole A-4 hydrolase</td>
<td>Vitamin D binding protein</td>
</tr>
<tr>
<td>Complement component 4A</td>
<td>Maltase-glucoamylase, intestinal Mucin and cadherin-like protein</td>
<td>Vitronectin</td>
</tr>
<tr>
<td>Complement factor H</td>
<td>Myeloblastin</td>
<td></td>
</tr>
<tr>
<td>Dipeptidase 1</td>
<td>Myeloperoxidase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N-acetylated-alpha-linked acidic dipeptidase-like protein</td>
<td></td>
</tr>
<tr>
<td>Dipeptidase 4</td>
<td>Neutrophil defensin 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zinc-alpha-2-glycoprotein</td>
<td></td>
</tr>
</tbody>
</table>
Proteins Consistently Found in CRC and Normals

- 16 proteins were consistently found in faeces of both normal and CRC patients (Found at least 80% of all times ie $\geq 6$ for normal and $\geq 8$ for CRC)
Proteins found only in CRC

8 proteins were found only in faeces of CRC

Interlectin, myeloblastin also validated
### Absolute MRM Quantitation of CEACAM5 (CRC)

<table>
<thead>
<tr>
<th></th>
<th>CEACAM5</th>
<th>RT (L)</th>
<th>Area Light</th>
<th>RT (H)</th>
<th>Area Heavy</th>
<th>ng/mg stool</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRC 1</td>
<td>(Light)</td>
<td>28.467</td>
<td>28525</td>
<td>28.388</td>
<td>142879</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>733.3 - 1051.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRC 2</td>
<td>(Heavy)</td>
<td>28.246</td>
<td>28314</td>
<td>28.310</td>
<td>143672</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>738.3 – 1061.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRC 3</td>
<td></td>
<td>28.246</td>
<td>28792</td>
<td>28.384</td>
<td>149191</td>
<td>11.7</td>
</tr>
<tr>
<td>Ave</td>
<td></td>
<td>28.32</td>
<td>28544</td>
<td>28.36</td>
<td>145247</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>0.13</td>
<td>240</td>
<td>0.04</td>
<td>3438</td>
<td></td>
</tr>
</tbody>
</table>

**CEACAM5**

SDLVNEATGQFR

**CEACAM5**

SDLVNEATGQF(\textsuperscript{13}C\textsubscript{15}N)R
The Microbiome
The Human Microbiome

• There are about $10^{14}$ microbes in our body, comprising bacteria, fungi and viruses
• The microbiota function in tandem with the host’s defence and immune systems to protect against pathogen colonisation and invasion
• They perform an essential metabolic function, acting as a source of essential nutrients and vitamins and aid in the extraction of energy and nutrients from food
• Around 30% of the human microbiota reside in the GI tract, existing in a complex but balanced homeostasis that is highly individual
• They can also be found in the skin, airways, the urogenital tract and other organs
• There are “good” (probiotics) and “bad” (e.g. C. difficile) microbiota
The Microbiome in Disease.

Changes extend to metastases

Science. 2017 Nov 23. pii: eaal5240
FDA approved commercial instrumentation: Bruker's MALDI Biotyper and BioMérieux Vitek MS. Both TOF-based. Spectral library-based identification (e.g. SpectraBank and mMass and the online BacteriaMS system developed by Fudan University)
FMT is gaining increased acceptance, especially for the treatment of *C. difficile* infection (CDI).

A number of Biobanks have been established.

Success rate of around 90%.

Obesity, metabolic syndrome and diabetes mellitus could be potential candidates for FMT.


The Role of Faecal Proteomics in FMT

• Faecal proteomics can clearly play an important role in treatment, monitoring and surveillance of FMT

• Can help in diagnosis and the understanding of disease mechanisms using suitable biomarkers

• Can help determine the type of transplant material and the method of administration by analysing the structure, function and composition of the microbiota in stool samples

• Metaproteomics can help to identify an optimal donor by comparative studies between recipients and donors. Importantly proteomics can efficiently monitor the efficacy of the procedure

• Proteomics coupled with genomic methods such as DNA sequencing can be used to survey changes in microbiota diversity, function and structure before and after FMT, in order to evaluate the therapeutic effect
Faecal Proteomics and Personalised Medicine
Faecal Proteomics in Personalised Medicine


Human Personalized Omics Profiling (hPOP)).
Extending the iPOP concept, 1000 volunteers are being enrolled at the annual HUPO meetings. Volunteers are consented and blood, urine and stool samples collected at the meeting under SOP, and health details (including food habits, personal health, physical activity and stress levels) taken.
HUPO meetings rotate between the US, Europe and Asia allowing both longitudinal studies and analysis of the variance of molecular markers across a large number of participants from different ethnic backgrounds.
What Are The Hurdles?

• High quality standardised faecal biobanks compatible with proteomics analysis are required.
• There is a need to normalise faecal samples for their varying size and consistency (analogous to use of creatinine for urine analysis) or colon related content. A33 cell surface antigen that is expressed > 95% of human colon cancers and normal human colon (Chr1)?
• Chromatographic reproducibility is still rate limiting in several areas of proteomics.
• Robust high throughput automated sample workup protocols need to be developed.
• While MS is becoming accepted for clinical applications, throughput is still not yet compatible with population-based clinical screening.
• It is important that all raw MS data be deposited to save data being lost, and enable it to be globally accessible for use and reanalysis.
• Education is needed to improve peoples’ willingness to handle their own stool samples.
Perspectives

Recent advances in rapid, sensitive and specific high throughput omics methods are paving the way for personalised medicine.

As proteomics moves from a discovery to a translational phase, a strong bidirectional synergy with the pathology community is vital for the development and optimisation of validated clinical assays. Cross training efforts will be essential to facilitate this.

The proteomics community will develop assays that will inform the pathology community on detection and management of disease. It will also assist in the education of the medical/pathology community in the role of omics technologies in personalised medicine.

The pathology community will provide expert advice on automated IHC analysis, sample selection, biobanking and big data analysis.
Acknowledgements

Ping Ji, Kui Wang, Canhua Huang

Rob Goode, Zon Weng Lai

Julie Rothacker, Helen Patsiouras, Peter Gibbs, Tony Burgess

Ching Seng Ang

Mark Baker

Ping Ji, Kui Wang, Canhua Huang
"The Omics Revolution: Beyond Genomics"

The Oaks Resort, Port Douglas, Queensland, Australia
Sun 8th Sept, 2019

This one-day satellite will focus on the key role that proteomic and other omics technologies will play in the development of personalised/precision medicine. The meeting will be held immediately prior to the 13th Australian Peptide Conference (apc2019.org). Previous successful meetings in this series were held in conjunction with the 2013 APC meeting in Penang, the 2015 meeting in Kingscliff, and the 2017 meeting in Noosa. HUPO 2019 (www.hupo2019.org) will be held in Adelaide the following week (15th – 18th Sept 2019). For more details contact ed.nice@monash.edu
Save the date

DATE
September 15–18, 2019

VENUE
Adelaide Convention Centre
North Terrace, Adelaide, South Australia, Australia

www.hupo2019.org