Proteomics: Understanding Proteins...

Sylvie LaBoissiere, Ph.D.
Director, Proteomics Platform
McGill University and Genome Québec Innovation Centre (MUGQIC)
Proteomics is the Logical Complement to Genomics

DNA → RNA → Proteins → Post-Translationally Modified Proteins

~ 25,000 Genes → > 1,000,000 Proteins

“A caterpillar and a butterfly have the same genome but completely different proteomes” - Ron Orlando
Support large-scale Genome Canada-funded projects with state-of-the-art technology platforms.

- Provide specialized genomics/proteomics services to the research community.
Creation of the MUGQIC Proteomics Platform: The CellMap Project

The CellMap Project aimed at the identification, the localization and the abundance determination of all proteins at the organelle level.

- The proteomics platform was designed to accommodate large-scale projects
- Results published in high impact journals
What Does Proteomics Do?

- Identify proteins within samples.
- Provide information on their PTM.
- Provide information on their abundance.
Challenges in Proteomics

- Only a fraction of the proteome can be identified and quantified.
  - Results depend on samples, sample preparation, instruments, databases and bioinformatics tools
  - Different results obtained with the same samples in different labs


- Proteomics Standardization Project on Platform
  - Come up with standardized protocols and statistical analysis
Classical Proteomics Workflow

Protein identification is based on probabilities
MUGQIC Proteomics Platform
Interaction with Users

- Give advice and suggestions on experimental design
- Give advice and suggestions on sample preparation
- Provide protocols
- Provide support for data interpretation
MUGQIC Proteomics Platform
Protein Separation

- Gel imaging system
- Gel cutting robot
- Automated in-gel tryptic digestion
- Depletion columns, SCX fractionation, in-solution digestion
**MUGQIC Proteomics Platform**

**Mass Spectrometry**

- Protein identification
- PTM (phosphorylation)
- Protein quantification
  - Redundant peptide counting, iTRAQ
- Capacity
  - 2 LC-QToF, 1 LC-QTRAP, 1 MALDI QToF
  - Up to **12,000** LC-MS analyses per year
MUGQIC Proteomics Platform
Bioinformatics

- Interaction with Users
- Protein Separation
- Mass Spectrometry
- Bioinformatics

- All raw MS data always preserved
- CellMapBase (Oracle relationale database)
- Biologically relevant information
- Custom reports for comparing samples and/or experiments
- Elimination of redundancy
MUGQIC Proteomics Platform Overview

Input from Client

- Sample
- Gel

Protein Separation

- Gel
- Imaging
- Excision
- Digestion

Mass Spectrometry

- Sample
- Peaklists
- Raw data

Biinformatics

- Peaklists
- Mascot Results
- Custom reports

Output to Client

- Gel image

Proteomics Platform
Mascot reports
Proteomics Platform

Custom Reports

- Custom reports are used for data mining and protein abundance determination.

Short Report (data mining)

Single Peptide Report (useful for publication purposes)

Shared Peptides Group Report (helps user with redundancy)
Of mice and men...

- Prototype Species:
  - D. megalonaster
  - C. elegans
  - A. thaliana

- Classical Species:
  - Mouse
  - Rat
  - Human

- More Exotic Species:
  - Chicken
  - Cow
  - Bee
  - Mollusca
  - HSV
  - Phaseolus
  - C. difficile

- Databases provided by users
Of mice and men…and more!

- Prototype Species:
  - *D. megalonaster*
  - *C. elegans*
  - *A. thaliana*

- Classical Species:
  - Mouse
  - Rat
  - Human

- More Exotic Species:
  - Chicken
  - Cow
  - Bee
  - Mollusca
  - HSV
  - *Phaseolus*
  - *C. difficile*

- Databases provided by users

Sample Types:
- Whole Organisms
- Cells in Culture
- Whole Organs
- Purified Organelles
- Milk
- CSF
- Plasma
- Honey
- Saliva
Protein Identification

- Identifying binding partners to proteins of interest.
- e.g. from immunoprecipitations or pulldowns (Torchia; UWO)

![Protein Identification Diagram]

- Protein of interest
- Putative binding partner #1
- Putative binding partner #2
- Putative binding partner #3
**Protein Identification**

- Identify unknown binding partners by SPR-MS
- e.g. identify novel Progranulin-1 binding partners (Bennett; RVH/McGill)

http://www.mcgill.ca/sheldon

PROGRANULIN

Growth factor in health and disease

Potential target for FTD / ALS therapeutics

SPR-MS

Biacore 3000 + ABI QTRAP
Protein Identification

- Identifying proteins associated to a structure or an organelle.
- e.g. host-pathogen interactions (in vitro purified HSV particles; Lippe, UdM).
- Creation of hybrid databases (human + HSV sequences) for protein searches.

Remillard-Labrosse et al., manuscript in preparation

129 human proteins
7 viral proteins*

* Proteins have to contain at least one unique peptide and > 1 peptide
**Protein Identification**

- Gene annotation of bacterial genomes:
  - Confirm expression and localization of predicted proteins
  - Identify proteins missed by gene annotation tools
  - Study host-pathogen interactions (cell-wall bound proteins)
  - User provided his database for protein identification (Dewar, McGill)

- Prep enriched cell wall proteins
- 1D gel
- LC-MS/MS

*Clostridium difficile*, pathogenic isolate QCD-66c26; 3745 predicted proteins

<table>
<thead>
<tr>
<th>McGill</th>
<th>24 Cell Wall (Ob.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>32 Cell Wall (Pr.)</td>
</tr>
</tbody>
</table>
Post-Translational Modifications

- **Phosphorylation:**
  - Enrichment on affinity column
  - Specific scanning method

- **Glycosylation:**
  - Scanning method specific to the glyco moiety

- **All other modifications:**
  - On the condition that the modification gives a unique chemical signature that can be monitored by the mass spectrometer
  - e.g. acetylation, methylation, etc.
Post-Translational Modifications
Phosphorylation

- plays in vital role in cells (apoptosis, oncogenesis, etc.)
- identification of phosphopeptides/phosphoproteins difficult:
  - Only a portion of proteins phosphorylated at any given time
  - Only a portion of the protein is phosphorylated

Proof of principle: HeLa cell lysate

1078 proteins identified
939 proteins with at least one phosphate group (87%)
Post-Translational Modifications
Phosphorylation

MAAHGSAASSALKGLIQQFTTITGASESVGKHMLEACNNNLEMAVTMFLDGGLGIAEEPSTSSASVSTVR
PHTEEEEVRAPIPQKEIIVEPEPLGAPKRRPARSIFDGFRDFQTETIRQEQLRNNGAIDKPLTLAD
LFRPFPIDLMHKSGFEETAKECGMIONKLMINIQNDVDFACQCLNRDVSNENAEVKNIIREHFIFQVYHDSEEGQRYIQFSGFYKLGLDPFYVSIILDPRTQKLVEWHQLDVSSFDQVTGFLGEGHQLQDSLSSPPKCARSES
LIDASEDSQLEAIRASLGETHDSTQKQDSRDESESELFSGSEGFIIVCSGSEDEEEERVNAKRSPHS
HKDLGHRKEENRRFLTEPVVRTPGTATTNHQGLPADVSEILMPEKADGVVGEIDVNGPKALMLRYP
DGKREQITLPEQAKLLALVKHIVQSKGPNFEELTNFPFRKLHLDYDITLQEAGLCFETVQVERN
Quantitative Proteomics

- **Redundant peptide counting (label-free; MS/MS)**
  - Developed on this platform
  - Based on accounting all MS/MS spectra matched to peptides

- **iTRAQ (labeled peptides; MS/MS)**
  - Based on the use of mass tag specific signature ions
  - Ratio of signature ions related to protein abundance

- **DIGE (labeled proteins; MS/MS)**
  - Service offered in collaboration with CIAN (McGill)
  - Gel image quantitation approach
Comparison of Protein Contents Using Redundant Peptide Counting

- Determination of protein contents of seeds from 2 cultivars.
  - Identification of proteins in both types of samples (n=3/type).
  - Comparison of abundance of proteins found in samples (Marsolais; AC).

- Only proteins with at least 1 unique peptide above id score seen in at least 2 replicates were considered.

- 17 proteins down-regulated in wild type vs mutant
- 12 proteins up-regulated in wild type vs mutant
- 139 proteins not significantly modulated

**1D gel**

**LC-MS/MS**

- wild type (140)
- mutant (149)

- 19 proteins
- 121 proteins
- 28 proteins
Analysis of Differentially Expressed Proteins Using DIGE

- Identify and quantify proteins from complex mixtures

Proteins labeled with dyes on same 2D gel + Image analysis

Robotic spot picking
In-gel digestion
Mass spectrometry
Bioinformatics

http://biology.mcgill.ca/CIAN/
cian@mcgill.ca

Proteomics platform
Depletion of Abundant Proteins from CSF for Biomarker Discovery

- Identify and quantify proteins from CSF
  - Study on Huntingdon Disease (Sponsored by HUPO)
  - 6 participating labs
  - 30 samples (10 samples/group)
  - Development of method to deplete abundant proteins from CSF

Normal Subjects

Early Stage

Diseased

Depletion of abundant proteins
1D gels
LC-MS/MS

424 proteins identified across 30 samples
Multiple Reaction Monitoring (MRM)

- Most sensitive and highly specific scanning method
- Ideal for:
  - PTM analysis
  - Absolute quantitation
  - Monitoring of known low abundance proteins (FMRP; Corbin, USherbrooke)

Shotgun LC-MSMS

- 137 peptides > ID score
- 0 peptide specific to FRMP

MRM LC-MSMS

- 8 peptides > ID
- 4 FMRP peptides

FMRP peptide 1
FMRP peptide 2
Launching New Activities on the MUGQIC Proteomics Platform

- What: self-service activities:
  - Sample preparation
  - Gel electrophoresis
  - Manual band/spot excision
- When: May 1st onwards
- Where: PSU laboratory
- For whom: regular or semi-regular users of the platform

proteomicservices@genomequebec.com
http://www.genomequebecplatforms.com/mcgill