Integration of Glycotyper into OncoMX; a web resource for exploring cancer biomarker data

Dr. Raja Mazumder  George Washington University
Dr. Anand Mehta  Medical University of South Carolina
Dr. Daniel Crichton  Jet Propulsion Laboratory, CIT
Dr. Brian Haab  Medical University of South Carolina
mazumder@gwu.edu  |  ITCR-IMAT presentation February 7, 2020
Specific Aims, OncoMX

**AIM 1:** Create with close collaboration a list of serum proteins with standard nomenclature, accessions and annotations that can act as biomarkers of liver cancer

- **Aim 1a)** Extend existing OncoMX data model to include glycan-protein biomarker data, and
- **Aim 1b)** Retrieve and integrate glycan-protein data from Glycotyper and link to other gene and protein centric molecular information for liver cancer

**AIM 2:** Data visualization. Users will be able to retrieve such biomarker information using glycan or protein accession numbers

- **Aim 2a)** Create a method (view) to visualize integrated glycan-specific biomarkers with existing protein-specific biomarker data
- **Aim 2b)** Build a side-by-side view to include relevant mutation, expression, literature mining, and pathway information
Specific Aim, Glycotyper

Aim 1. MALDI-MSI glycan analysis of antibody-captured (array) serum glycoproteins

Use Glycotyper method\(^1,2\)

1. Spot antibodies (glycosylation removed) on nitrocellulose-coated slides
2. Immuno-capture serum proteins
3. Release N-glycans from immobilized proteins by PNGase F
4. Coat with MALDI matrix to prepare for imaging
5. MALDI FT-ICR mass spectrometry imaging for detection of N-glycans
6. Identify N-glycans present on each captured protein

Glycotyper data generation

**Blood samples from HCC, cirrhotic, transplant patients**

**Glycotyper analysis**
- Immuno-captured A1AT, Haptoglobin, Hemoglobin, IgG, LMWK, Serotransferrin, Orosomucoid

A.) **Array Fabrication**
- Antibody Spotting

B.) **Glycoprotein Capture**

C.) **MALDI Preparation**
- N-glycan Enzymatic Release (In situ PNGase F)

D.) **Matrix Application**
- (Glycan Immobilization)

E.) **MALDI Mass Spectrometry Imaging**

![ MALDI Mass Spectrometry Imaging Image ]
A web-accessible cancer genomics biomarker knowledgebase

- Overcomes challenges integrating, analyzing, interpreting large international datasets
- FAIR data principles, harmonization, ontology-based data integration
- Defined biomarker data model, extensible data types
- User community focus, provenance transparency (BCOs)

Focus

- Integrated cancer biomarker data
  - Mutation, differential/normal expression, literature mining
  - Biomarker exploration in evolutionary or pathway context
  - Identify new cancer biomarker datasets and datatypes for integration

Currently contains little glycoconjugate or glycosylation-linked cancer biomarkers data
OncMX

WHAT WE DO
- Mutation
- Curation
- Orthology
- Clinical Data
- Text Mining

Compare Expression Profiles
Annotation
Ontology
Pathways

WHAT WE PROVIDE
- Custom Portal
- Search
- Browse
- EDRN Integration

HOW WE DO IT
- Ontology
- Unification
- User Driven Development
- Data Infrastructure
- Gene-centric Search

Uniform Reporting of Cancer Biomarkers

CLINICAL TRANSLATION

Existing Data Resources

Value-added Data

Community Tools

DNA-Seq, RNA-Seq, VCF, OBO, xml, OWL, etc.
Significance of glycobiology

A growing area of cancer research

Glycans/Glycosylation

- One of the most abundant and complex PTMs
- Critical in
  - Cell signaling; Cell-matrix interactions; Immune modulation; Development
  - Tumor angiogenesis; Tumor cell invasion; Metastasis

Hepatocellular Carcinoma (HCC)

- Top six cancer type; Fourth highest mortality; Aggressive
- Potential glycoprotein biomarkers
  - Peroxiredoxin 3, Osteopontin, Alpha Fetoprotein (AFP), Transferrin, Alpha-1-antitrypsin, GP73
# Use-case for Glycotyper in OncoMX

<table>
<thead>
<tr>
<th>As a &lt;role&gt;</th>
<th>I want &lt;what&gt;</th>
<th>So that &lt;why&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research investigator</td>
<td>To identify a list of genes for N-glycosylated serum proteins</td>
<td>I can find biomarkers for Hepatocellular carcinoma</td>
</tr>
<tr>
<td>Research associate</td>
<td>To identify a list of genes expressed in lung cancer</td>
<td>I can identify common pathways implicated in lung cancer</td>
</tr>
<tr>
<td>Clinical researcher</td>
<td>To evaluate a list of biomarkers available in OncoMX and also in the EDRN portal</td>
<td>I can compare biomarkers available in OncoMX as well as EDRN portal</td>
</tr>
<tr>
<td>PhD Student</td>
<td>To identify mutations in all forms of breast cancer</td>
<td>I can see which important mutations are conserved in breast cancer</td>
</tr>
<tr>
<td>Post-doc in bioinformatics</td>
<td>To observe differential expression reported in literature</td>
<td>I can compare experimentally identified genes with existing literature</td>
</tr>
</tbody>
</table>
Integration of Glycotyper data

Identify glycoprotein data for addition to OncoMX model

- Extend OncoMX data model to glycan data
- Map Glycotyper data to OncoMX model

Adapt OncoMX pipeline to glycan data

Create views of glycan data
1. The XLS file of raw GlycoTyper data (MALDI reads) is converted to CSV and formatted for the OncoMX knowledgebase.

2. M/Z designations (identities) of assayed N-glycans (50) are mapped to corresponding glycan structures recorded in a “Master glycan list” (MUSC).

3. Clarification.

<table>
<thead>
<tr>
<th>M/Z</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1257.4226</td>
<td>Hex5HexNAC2</td>
</tr>
<tr>
<td>1298.4492</td>
<td>Hex4HexNAC3</td>
</tr>
<tr>
<td>1419.4754</td>
<td>Hex3HexNAC2</td>
</tr>
<tr>
<td>1460.5020</td>
<td>Hex2HexNAC3</td>
</tr>
<tr>
<td>1485.5337</td>
<td>Hex3Hex1HexNAC4</td>
</tr>
</tbody>
</table>
Glycan data integration workflow (2)

2. Normalize, means

3. For each immuno-captured protein, MALDI reads are normalized across all N-glycans for each patient and mean values are calculated for each glycan in each patient group (HCC, Cirrhosis, Transplant) in each protein.

4. Standard protein names, gene names, and unique identifiers (ID) are mapped from UniProt and HGNC, for each immuno-captured protein.

5. Integrated genomic, proteomic, normalized, and statistical data are loaded into the OncoMX data tables and verified.

Map gene, protein data

OncoMX data table
OncoMX Data Model (Partial)
Visualization of Glycotyper data

- Integrated views of glycoprotein data in OncoMX
- N-glycans linked to specific serum protein biomarkers
- Cross-linkage to GlyTouCan and GlyGen
- SNFG-based N-glycan images
Outcomes

- Diverse biomarker datatypes modeled, standardized
- Extension of data model to glycan-protein biomarker data
- Glycotyper data integrated with other data types, standardized
Future directions

Development plans
• Expanding the data model

Collaboration
• U24 submitted – includes new data types such as glycans and imaging data
• Continue to add connections with other ITCR funded resources

Upcoming outreach
• ISB (submitted two talks and posters)
• JCO manuscript accepted
• EDRN-Biomarker workshop; ISMB (planning to submit)
Acknowledgements

- Raja Mazumder (Multi-PI)
- Robel Kahsay
- Amanda Bell
- Evan Holmes
- Stephanie Singleton
- Charles Hadley King
- Jonathon Keeney
- Jeet Vora
- Rahi Navelkar
- Janisha Patel
- Naila Gulzar
- Nakhita Gogate
- Hayley Dingerdissen
- Edmund Cauley
- Jiuge Yang
- Daniel Lyman

- Dr. Anand Mehta (Multi-PI)
- Dr. Brian Haab (Multi-PI)
- Alyson Black
- Daniel Crichton (Multi-PI)
- Heather Kincaid
- David Liu
- Frédéric Bastian (Sub-Awardee)
- Marc Robinson-Rechavi
- Tarcisio Mendes

- K. Vijay-Shanker (Sub-Awardee)
- Samir Gupta

- W. Evan Johnson (ITCR Collaborator)

Previous contributions
- Ashique Mahmood
- Lama Elzohary
- Yu Hu
- Ting-Chia Chang
- Prince Birring
- Julien Wollbrett

Current funding: NCI U01CA2015010
Past funding: NCI EDRN award 156620
Integration of Glycotyper into OncoMX; a web resource for exploring cancer biomarker data

Dr. Raja Mazumder  George Washington University
Dr. Anand Mehta  Medical University of South Carolina
Dr. Daniel Crichton  Jet Propulsion Laboratory, CIT
Dr. Brian Haab  Medical University of South Carolina

February 7, 2020