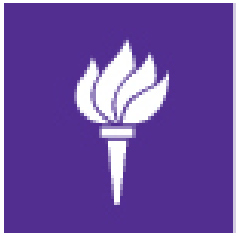
Spring 3-2012

Microarray Analysis as a Strategy to Identify and Characterize Glycome Regulatory

John F. Rakus
Marshall University, rakus@marshall.edu



NEW YORK UNIVERSITY

Microarray analysis as a strategy to identify and characterize glycome regulatory mechanisms

John Rakus, Ph.D.

Department of Chemistry
Advisor: Prof. Lara Mahal
March 16, 2012

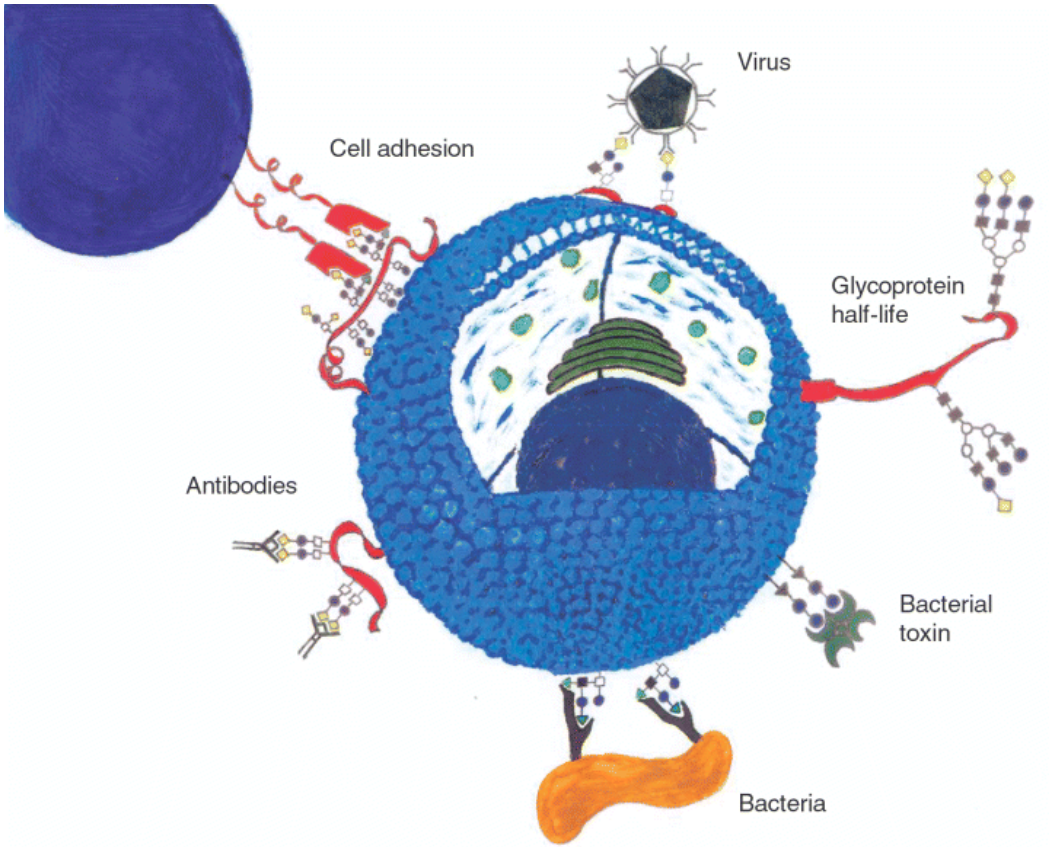
Agenda

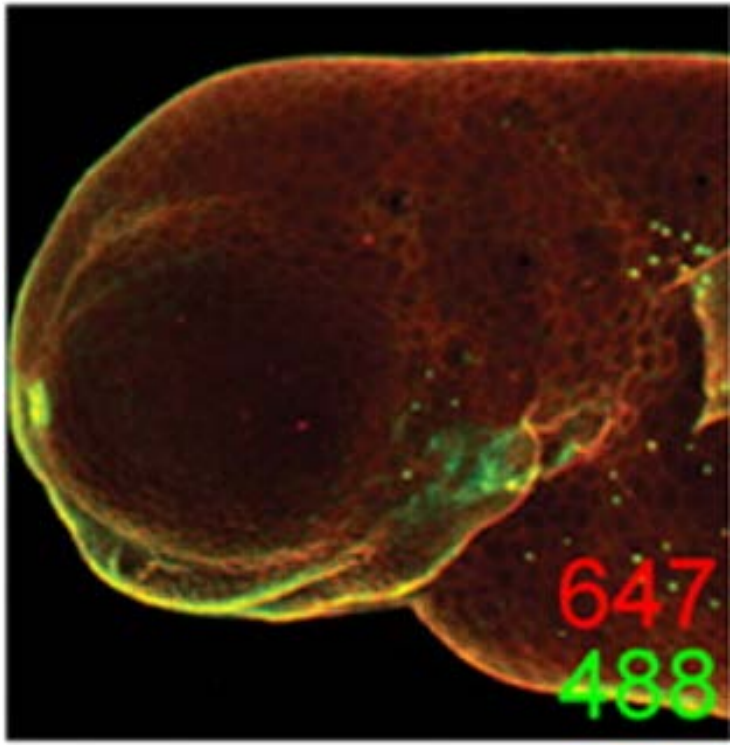
- **Carbohydrates are a diverse and critically important class of biological macromolecules**
- **Microarray strategy to identify glycome regulation**
- **Biochemistry and biological relevance of *C*-linked glycosylation**

Agenda

- **Carbohydrates are a diverse and critically important class of biological macromolecules**
- **Microarray strategy to identify glycome regulation**
- **Biochemistry and biological relevance of *C*-linked glycosylation**

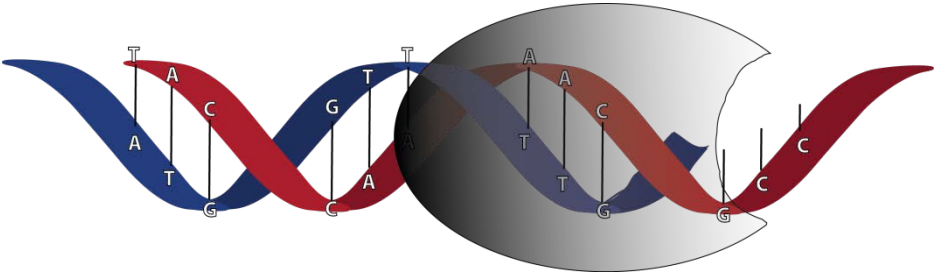
Carbohydrates are pervasive and involved in many cellular interactions

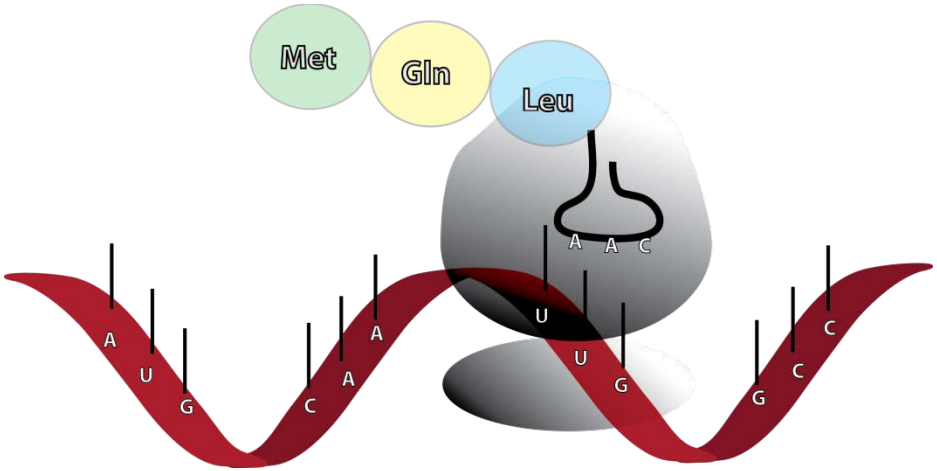




Holgersson et al, *Immuno Cell Biol*, 2005
Baskin et al, *PNAS*, 2007

Nucleic acids and proteins are synthesized with a defined template and dedicated polymerases





Macromolecule: Nucleic acid

Polymerase: DNA Pol or RNA Pol

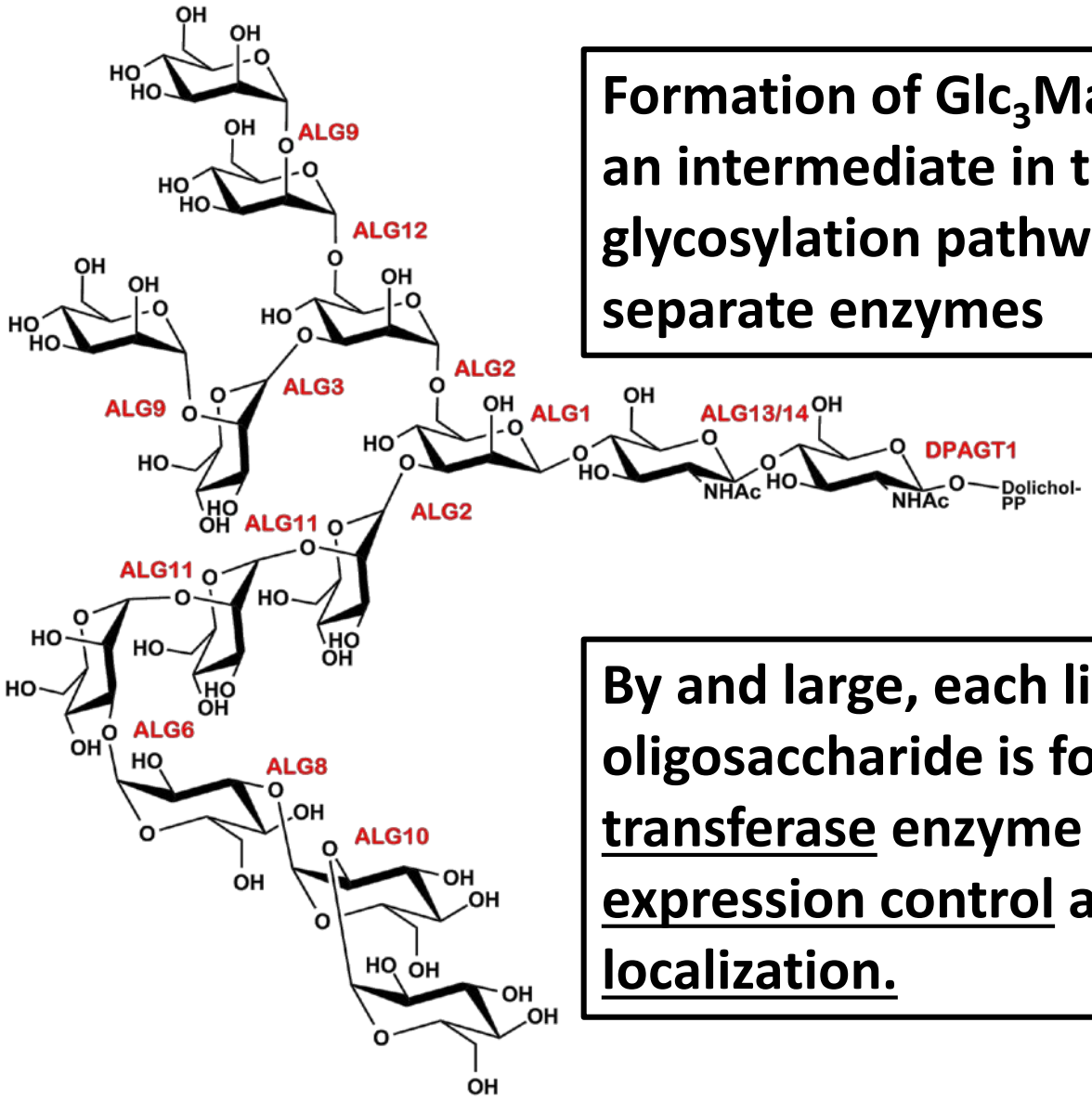
Template: DNA strand

Macromolecule: polypeptide

Polymerase: Ribosome

Template: mRNA strand

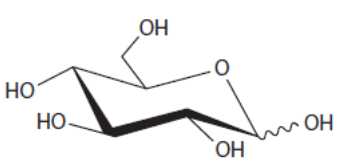
Glycan biosynthesis lacks a dedicated polymerase and genetic template

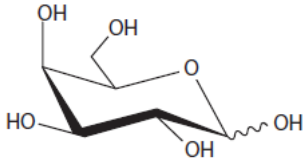


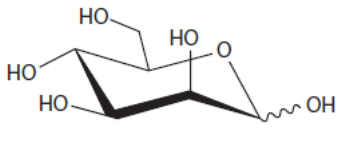
Formation of $Glc_3Man_9GlcNAc_2$-DolPP, an intermediate in the *N*-linked glycosylation pathway, requires 12 separate enzymes

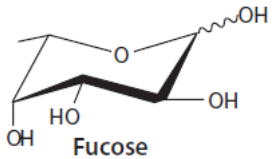
By and large, each linkage in an oligosaccharide is formed by a specific transferase enzyme with specific expression control and subcellular localization.

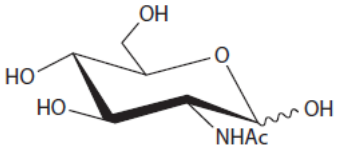
Consortium for Functional Glycomics (CFG) Notation

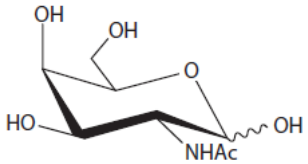


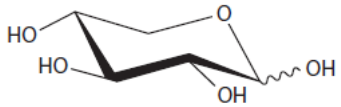


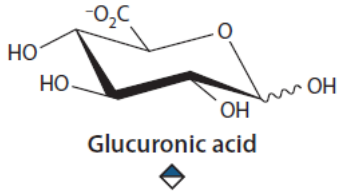


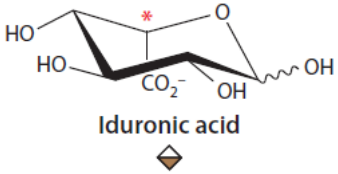


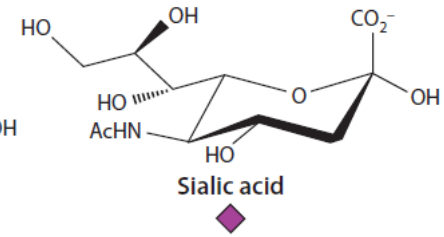


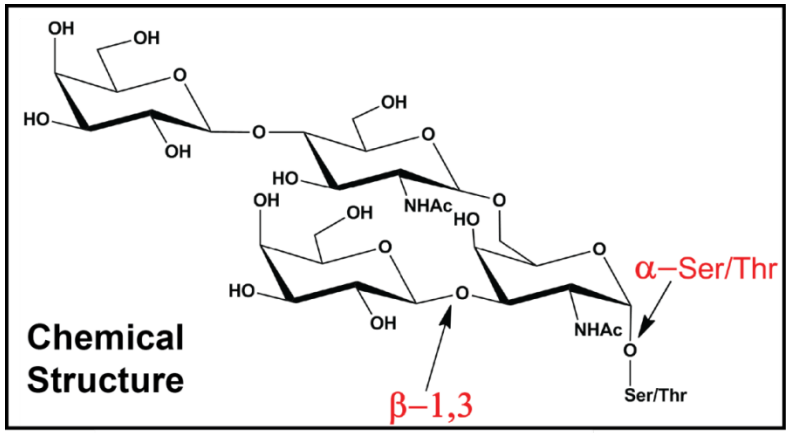


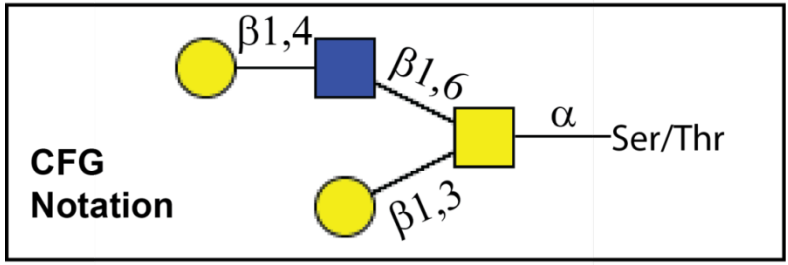






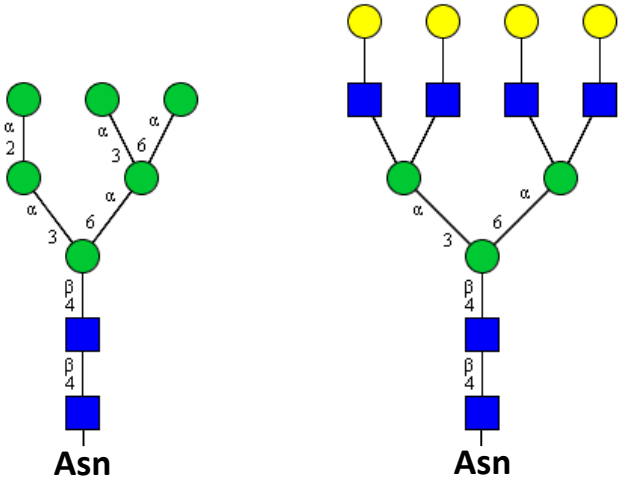






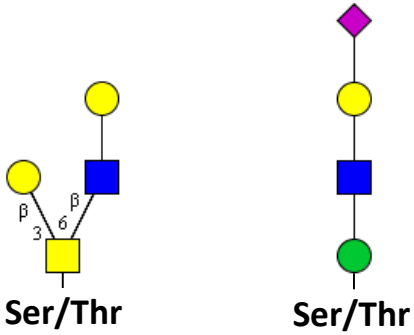
Rakus and Mahal, *Ann Rev Anal Chem*, 2011

N-linked glycosylation



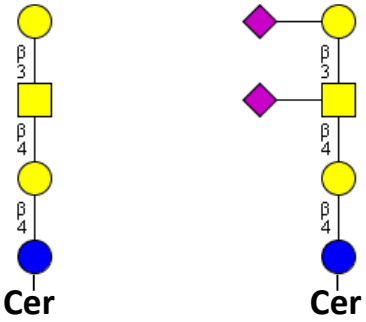
Functions: protein folding and trafficking; signaling ligand

O-linked glycosylation



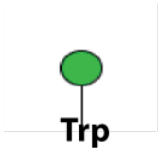
Functions: protein folding and trafficking; ECM composition

Glycolipids



Functions: Cell surface receptors; membrane composition

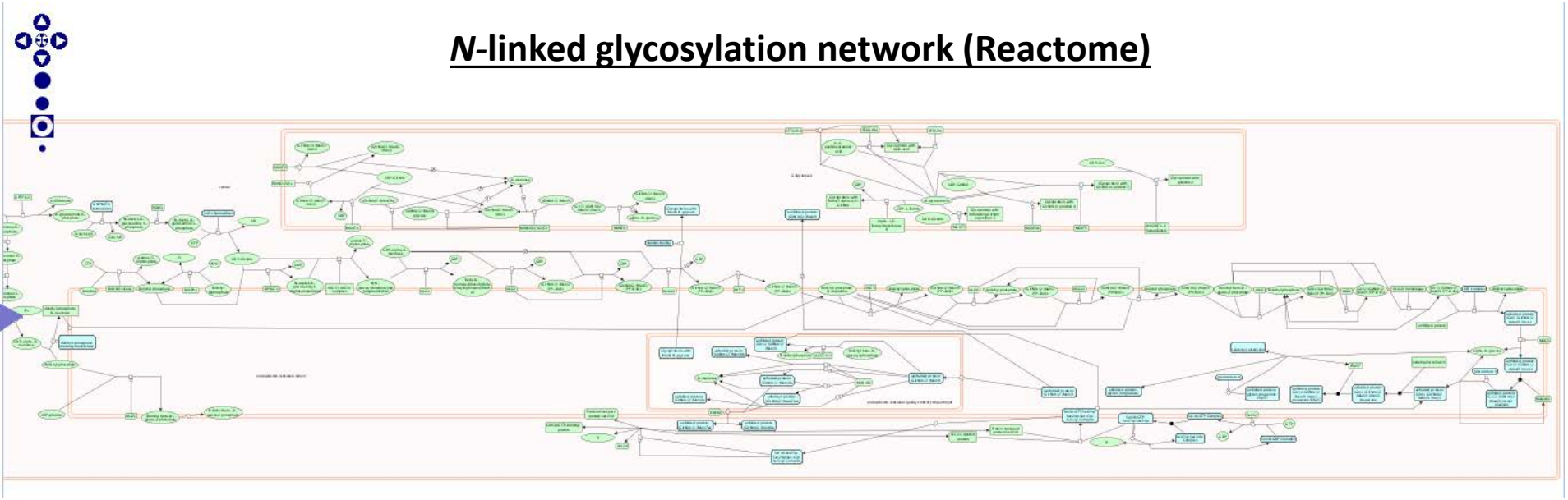
C-linked glycosylation



Functions: Unknown (stabilization?)

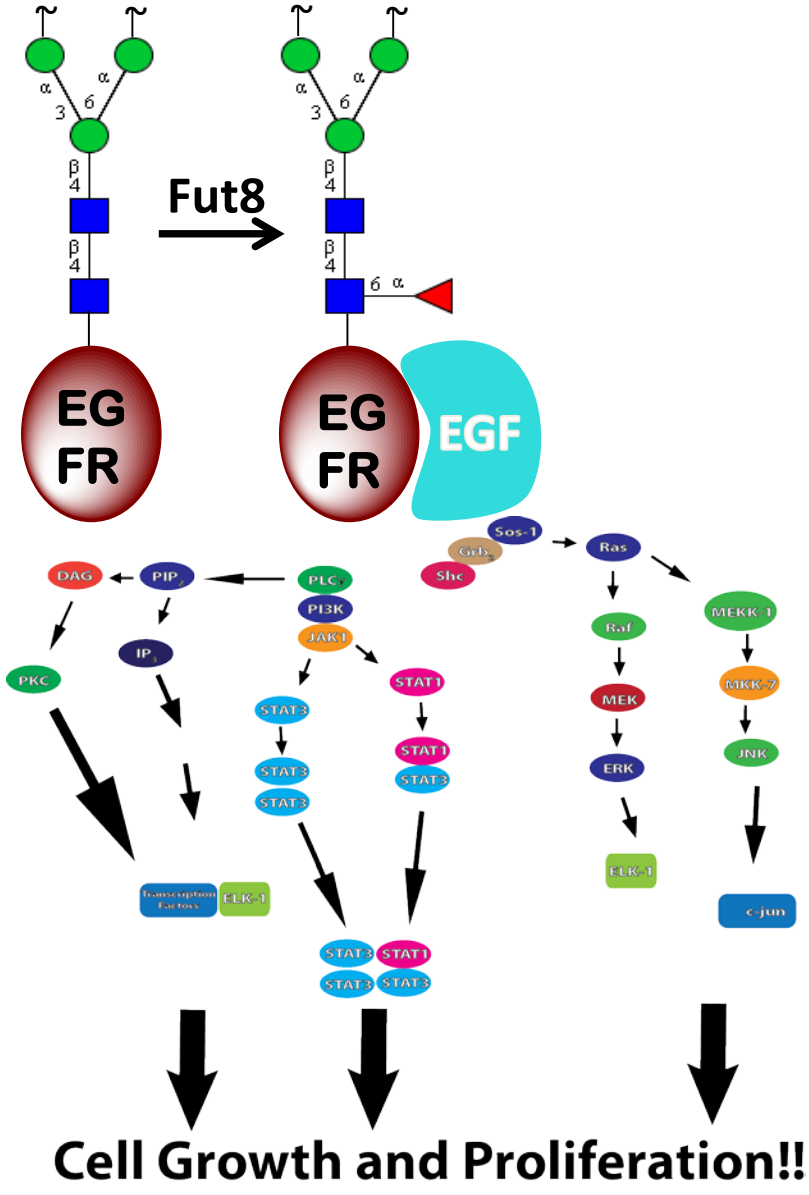
Carbohydrate synthetic regulation

***N*-linked glycosylation network (Reactome)**



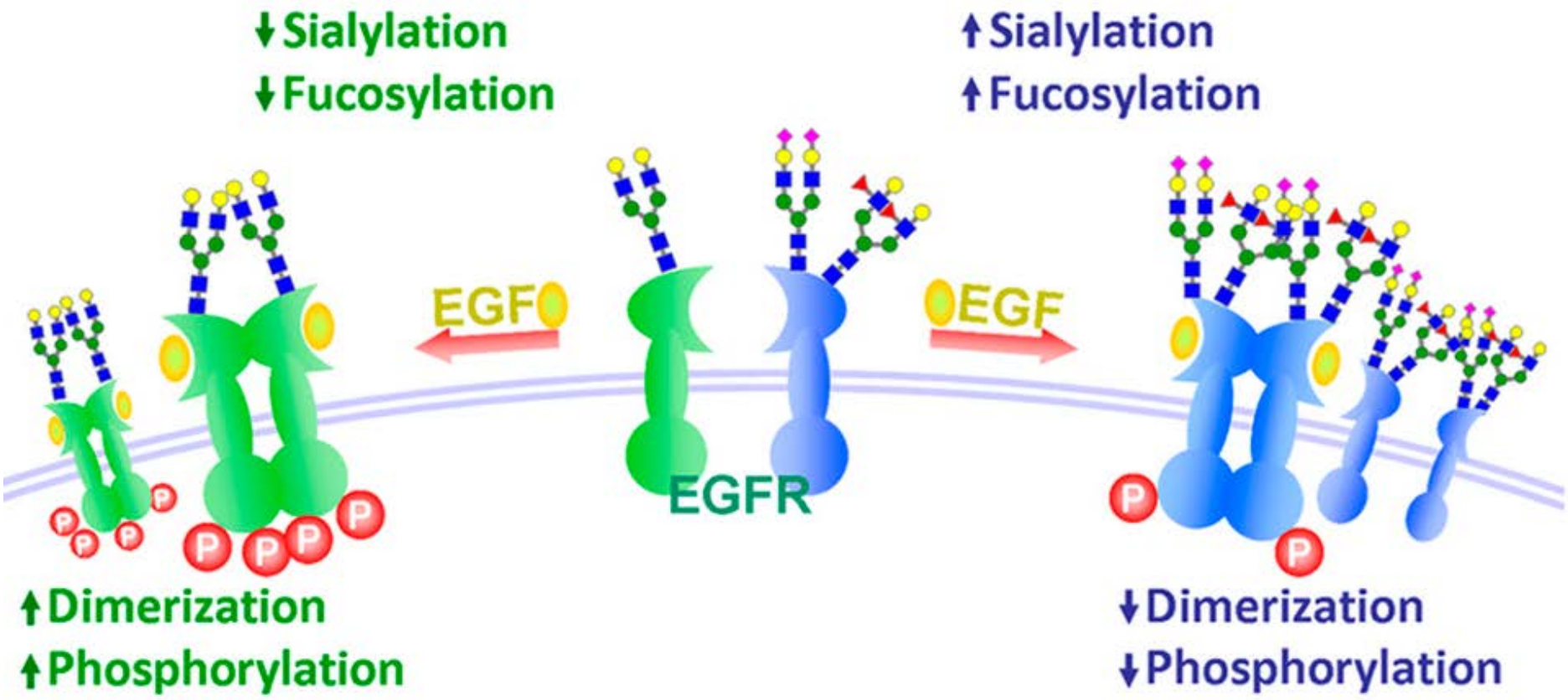
- **Synthesizing a glycome requires a large commitment of cellular resources**
- **Many glycosylation enzymes (glycosyltransferases and glycosidases), sugar transporter and metabolic proteins, and regulation elements (over 120 identified as of 2011)**
- **Can we study the genome-wide regulation of this process?**

EGFR: core α1,6-fucosylation activiates cell growth pathways through EGF binding



Wang et al, *J Biol Chem*, 2006
Matsumoto et al, *Cancer Sci*, 2008

…but terminal α1,3 fucosylation inhibits signaling activation by preventing EGFR dimerization



Liu *et al, PNAS*, 2011

Differential core fucosylation in cancer

Increase in core fucosylation has been observed in…

- **Prostate cancer (Saldova et al, *Glycobiology*, 2011)**
- **Pancreatic cancer (Sarrats et al, *Proteomics Clin Appl*, 2010)**
- **Stomach cancer (Bones et al, *J Proteome Res*, 2011)**

Decrease in core fucosylation has been observed in…

- **Lung cancer (Arnold et al, *J Proteome Res*, 2011)**

Agenda

- Carbohydrates are a diverse and critically important class of biological macromolecules

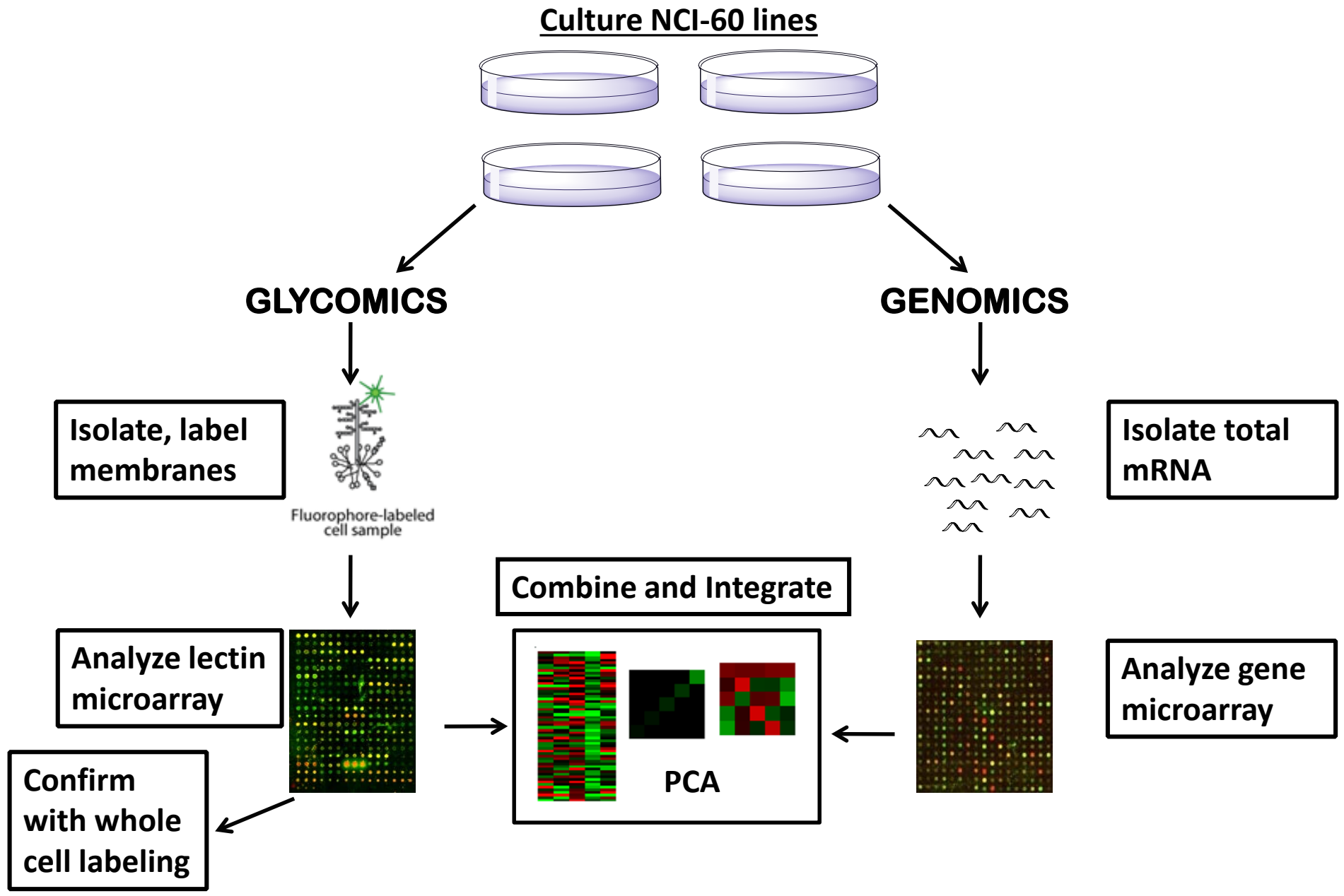
- **Microarray strategy to identify glycome regulation**

- Biochemistry and biological relevance of *C*-linked glycosylation

Model System: The NCI-60 Cell Panel

- **NCI-60: 60 cell lines for screening of potential cancer therapeutics**
- **Vary in tissue type, metastasis, individual of origin**
- **CellMiner.org: open source database containing mRNA, miRNA and protein array data, genetic mapping, pharmacological and mutational analysis**

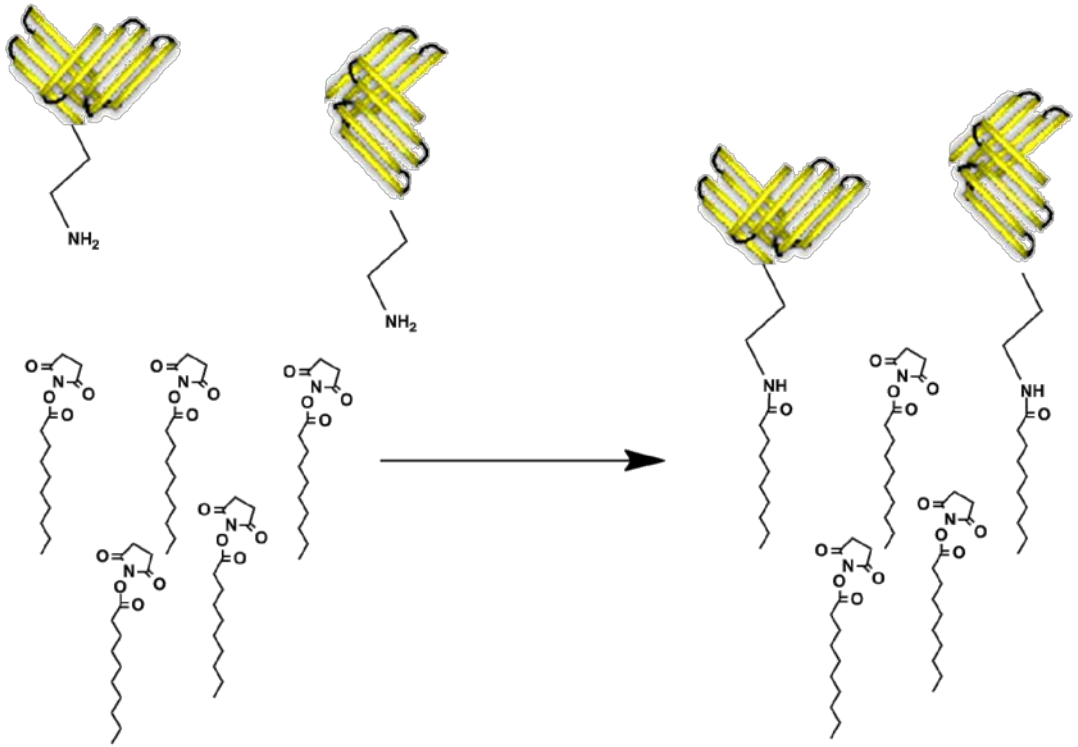
Experimental Strategy



Lectins

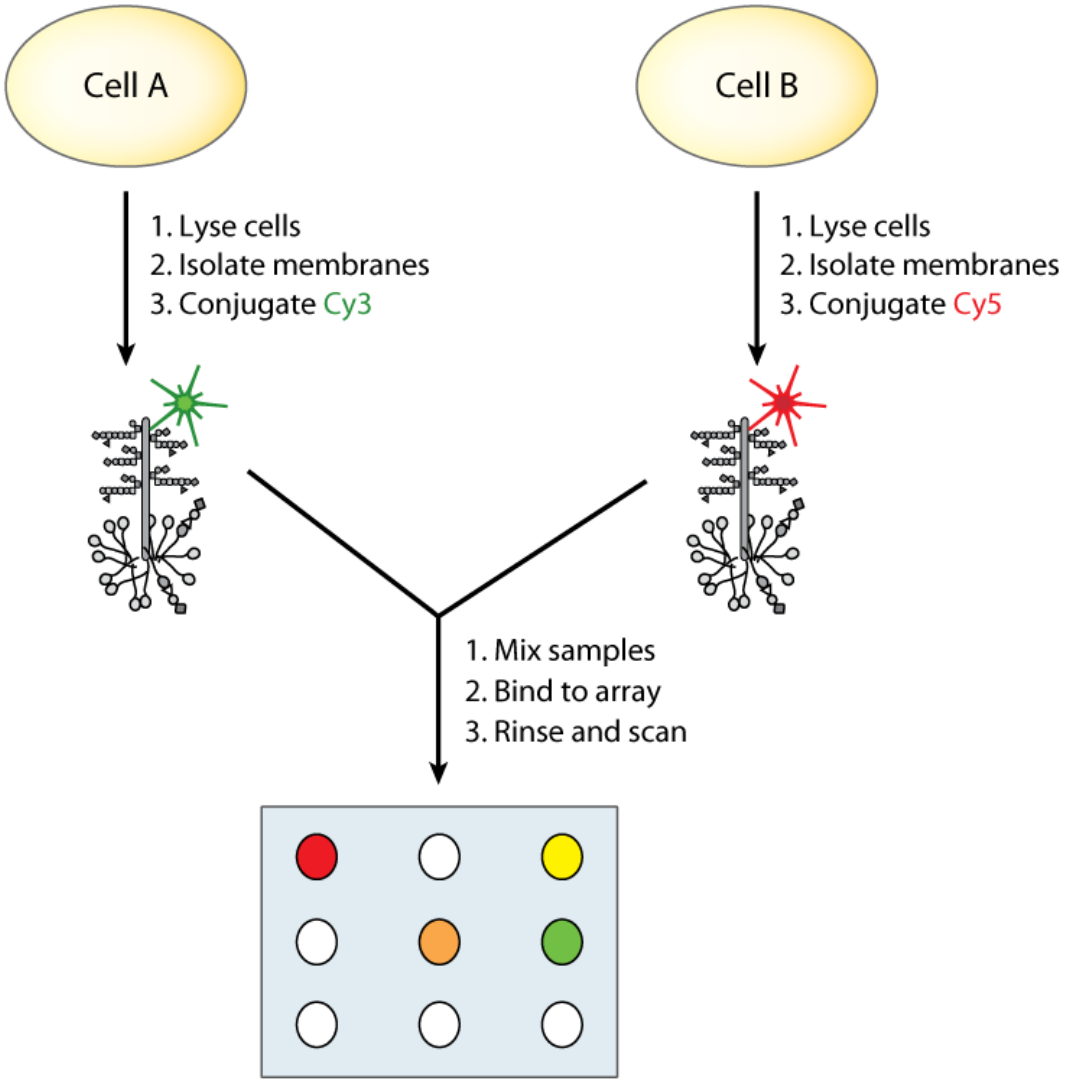
- **Lectins: non-enzymatic, non-immunological carbohydrate-binding proteins**
- **Found in all domains of life and demonstrate a wide range of structural design and biological function**
- **Lectins are often modified post-translationally (particularly in higher organisms)**
- **Many are readily available commercially**
- **Some have very broad specificities (WGA) some are extremely restricted (PSA)**

Generation of Lectin Microarrays



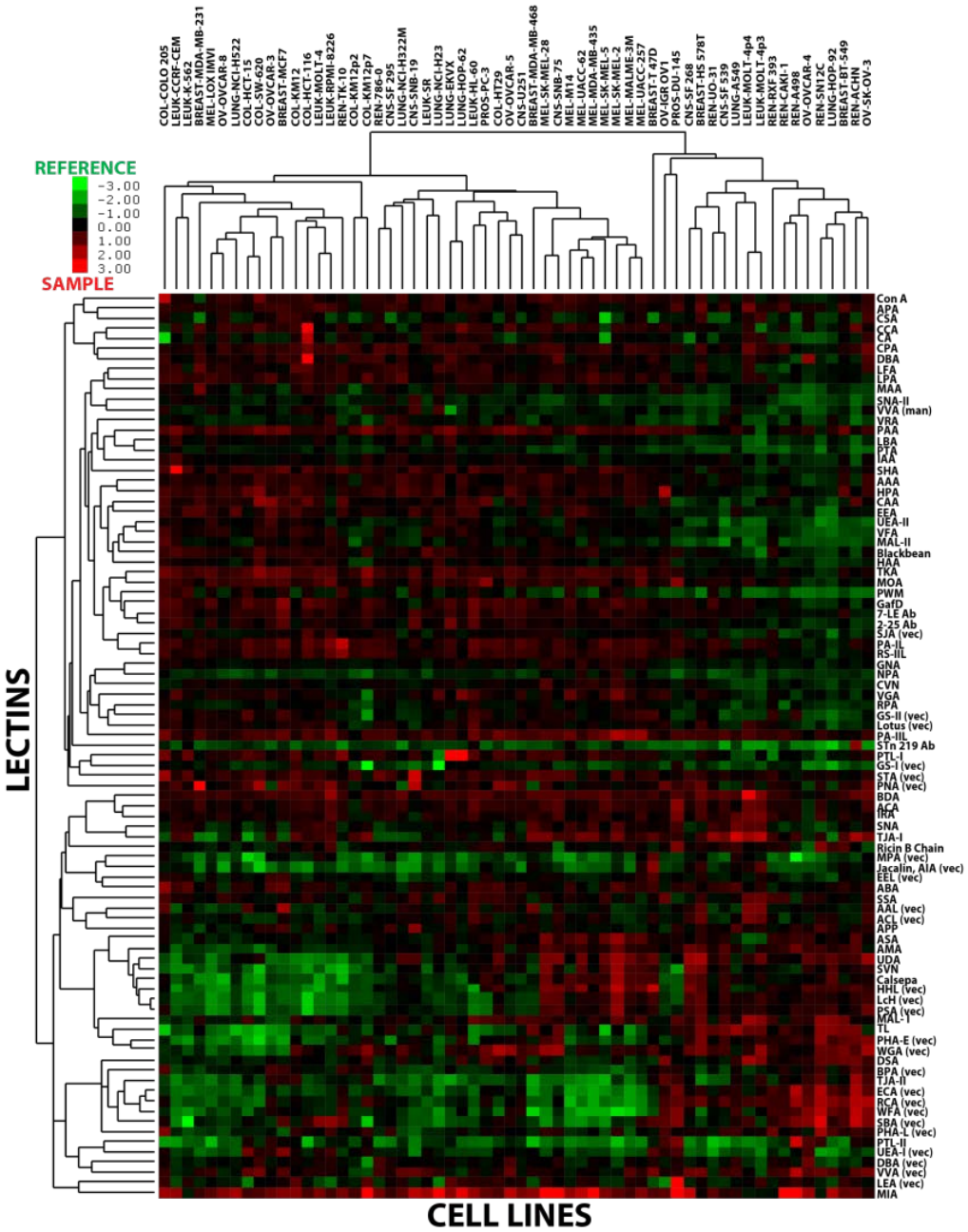
- **Lectins are printed on NHS-ester coated glass slides in high spatial density at 10°C and ambient humidity**
- **Protein lysine residues react with esters to form amide-bound conjugates**
- **Unreacted esters are blocked with ethanolamine**
- **Slides can be stored for up to two months**

Ratiometric lectin microarray analysis for semi-quantitative analysis of the dynamic glycome



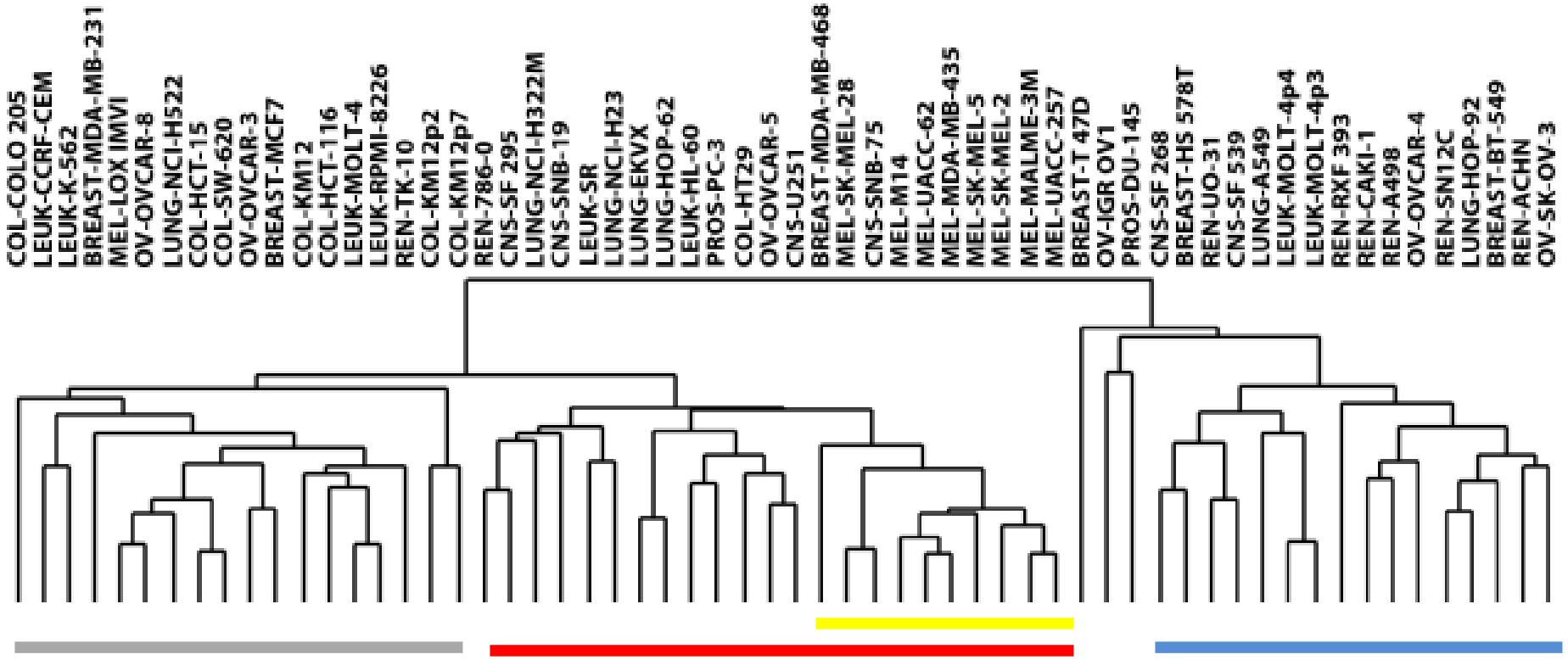
Pilobello et al, *PNAS*, 2007
Rakus and Mahal, *Ann Rev Anal Chem*, 2011

Lectin Microarray Analysis: NCI-60



- 90 usable lectin probes (two were misprinted)
- Two lung cell lines were discarded due to technical error (NCI-H226, NCI-H460)
- One colon line was not included due to inability to culture (HCC 2998)
- Biological replicates of KM12 and MOLT-4 were included

Cell lines cluster based on their glycosylation



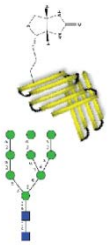
COLON (5/6, p=0.39)
LEUKEMIA (4/6, p=0.39)

CNS (4/5, p=0.44)
LUNG (4/7, p=0.44)

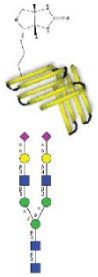
MELANOMA (8/9, p=0.60)

RENAL (6/8, p=0.39)

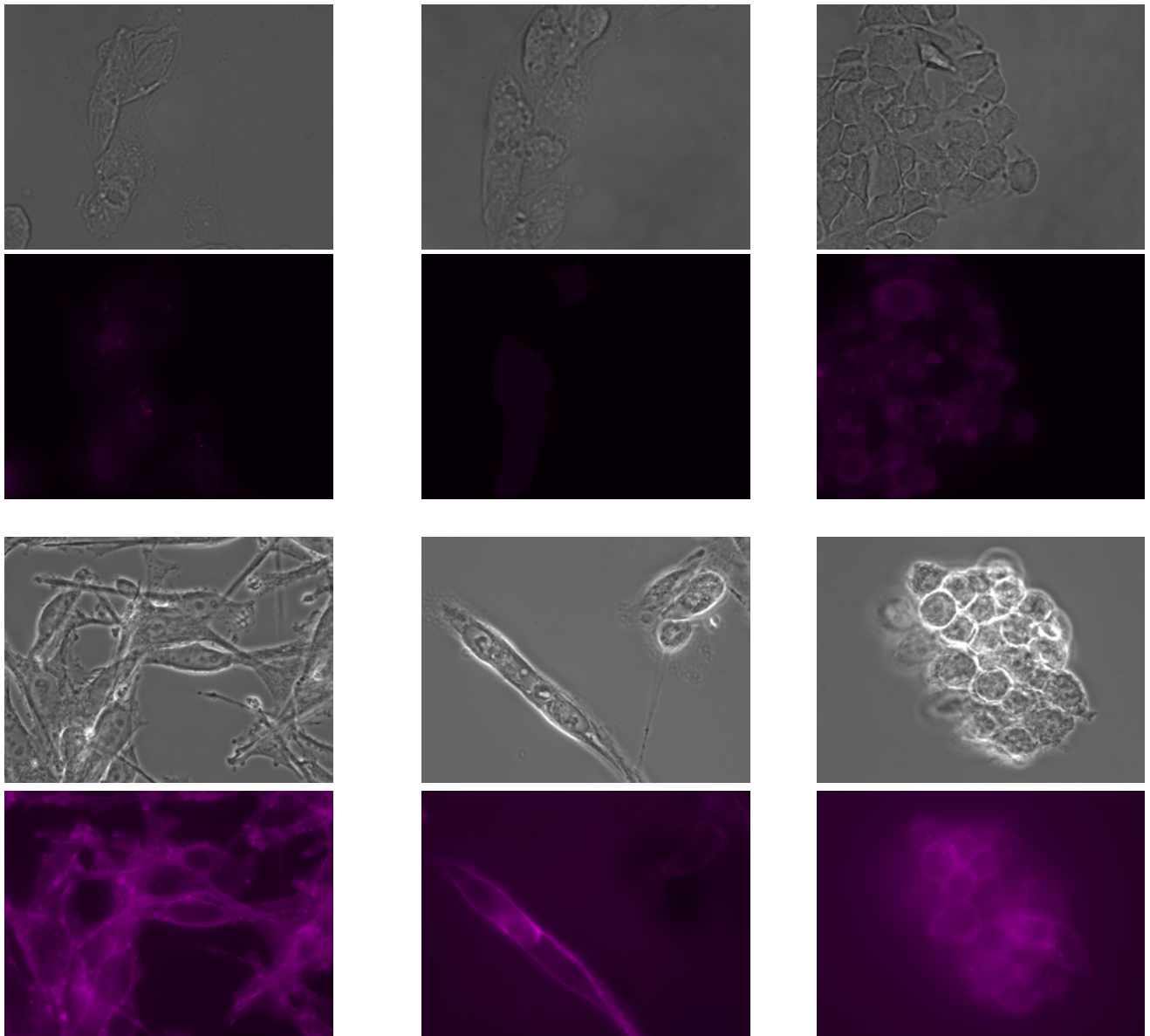
Lectin histology confirms glycan presence



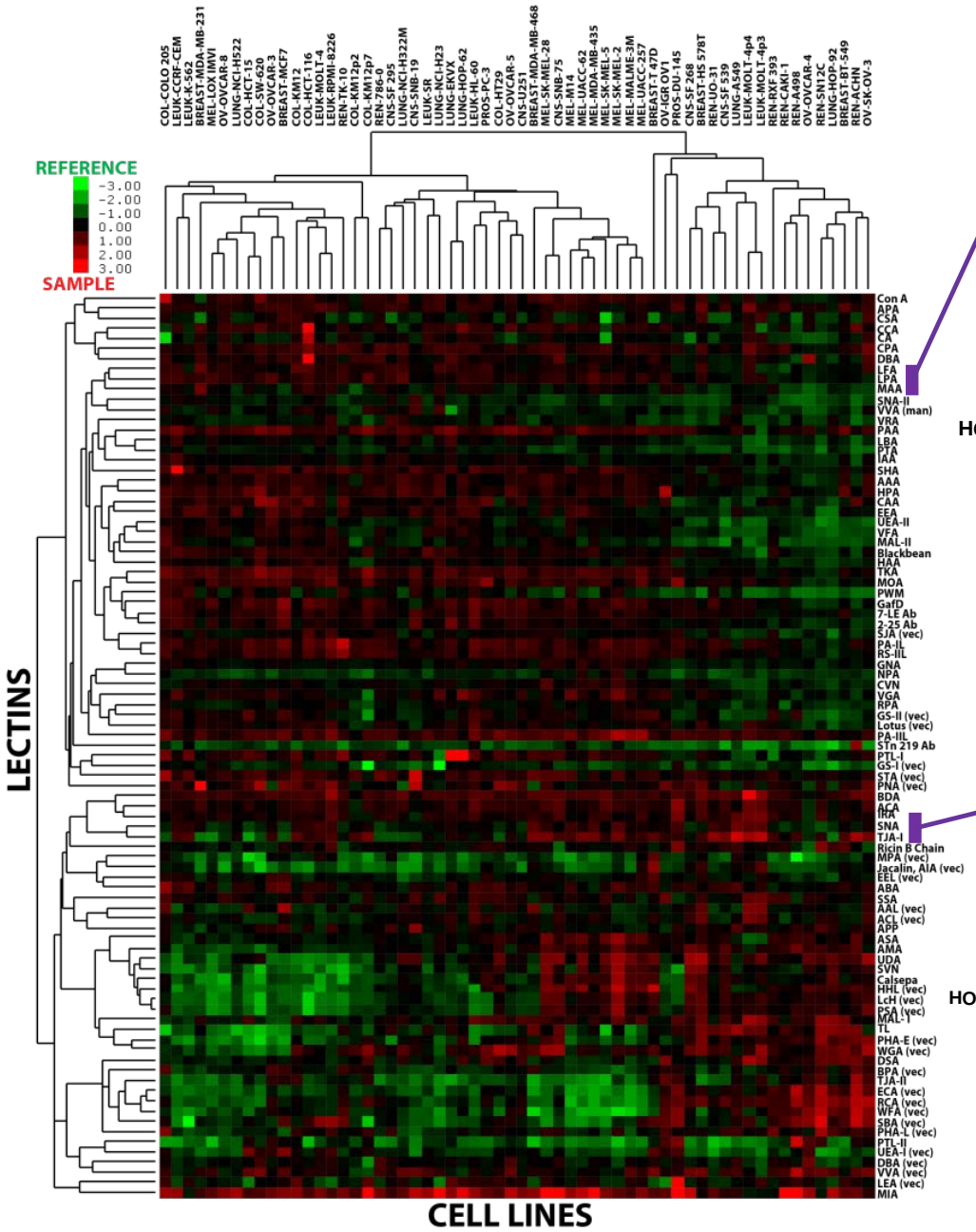
GNA



SNA



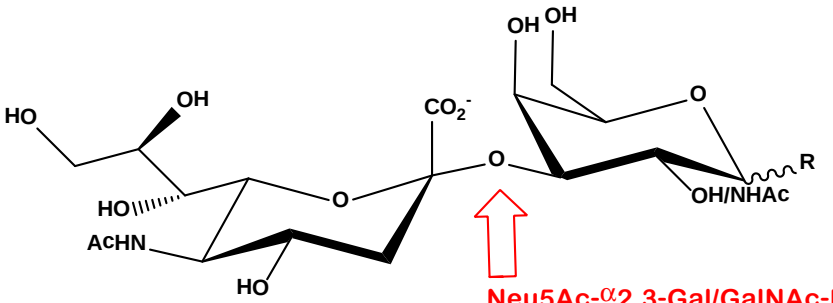
Glycosylation signatures: Sialic acid



α2,3 sialic acid recognizing lectins

LFA, LPA, MAA

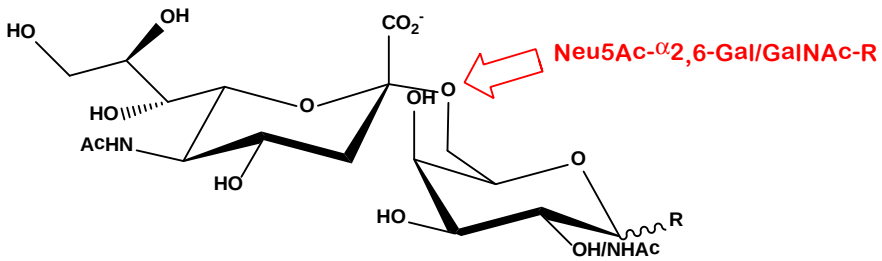
in RENAL cell lines



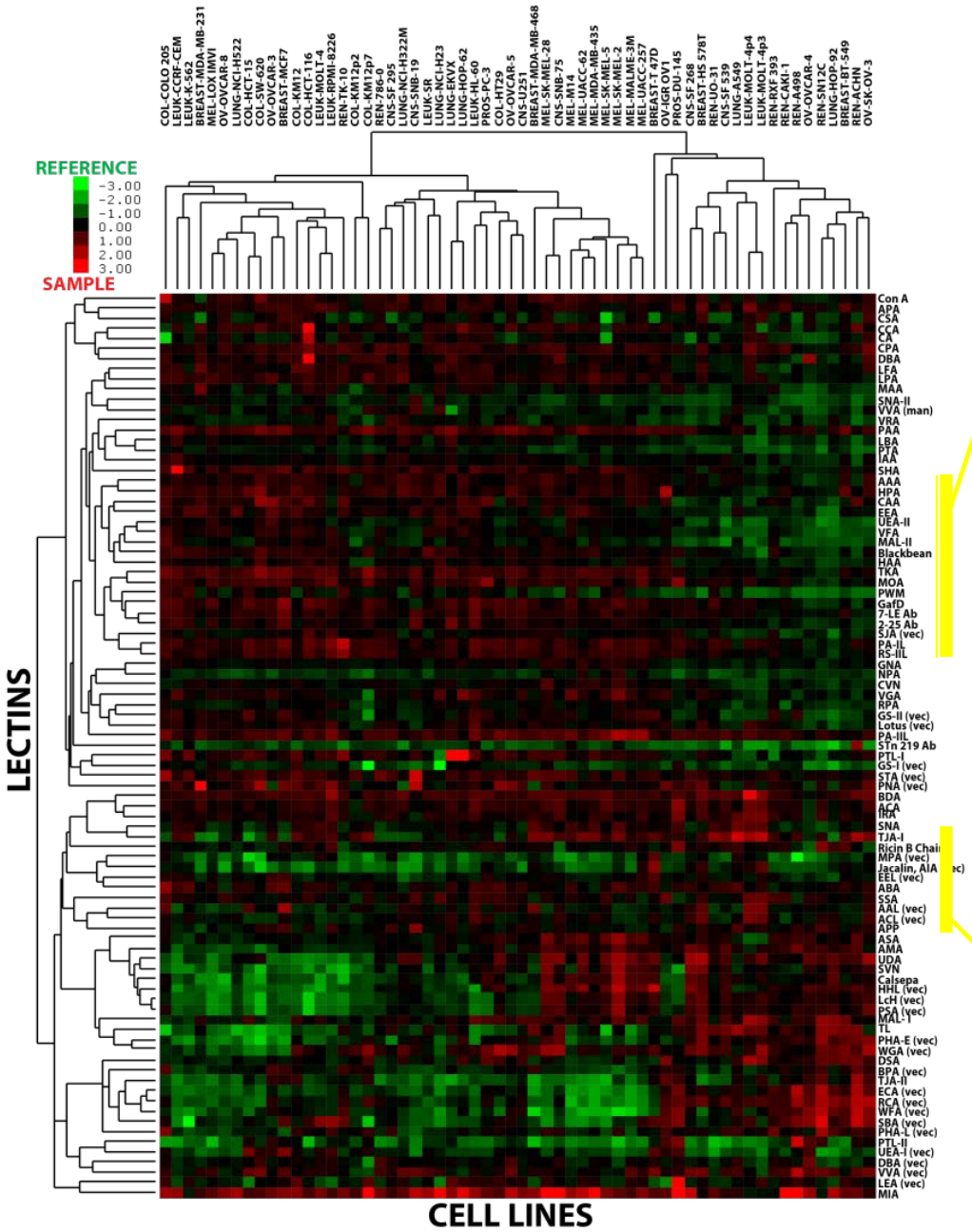
α2,6 sialic acid recognizing lectins

SNA, TJA-I

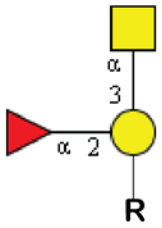
in MELANOMA and LUNG cell lines

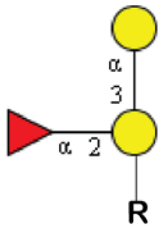


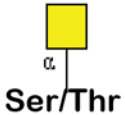
Glycosylation signatures: Gal/GalNAc



Blood groups/Gal/GalNAc
MOA, EEA, LBA, VRA, IAA, PTA, HPA

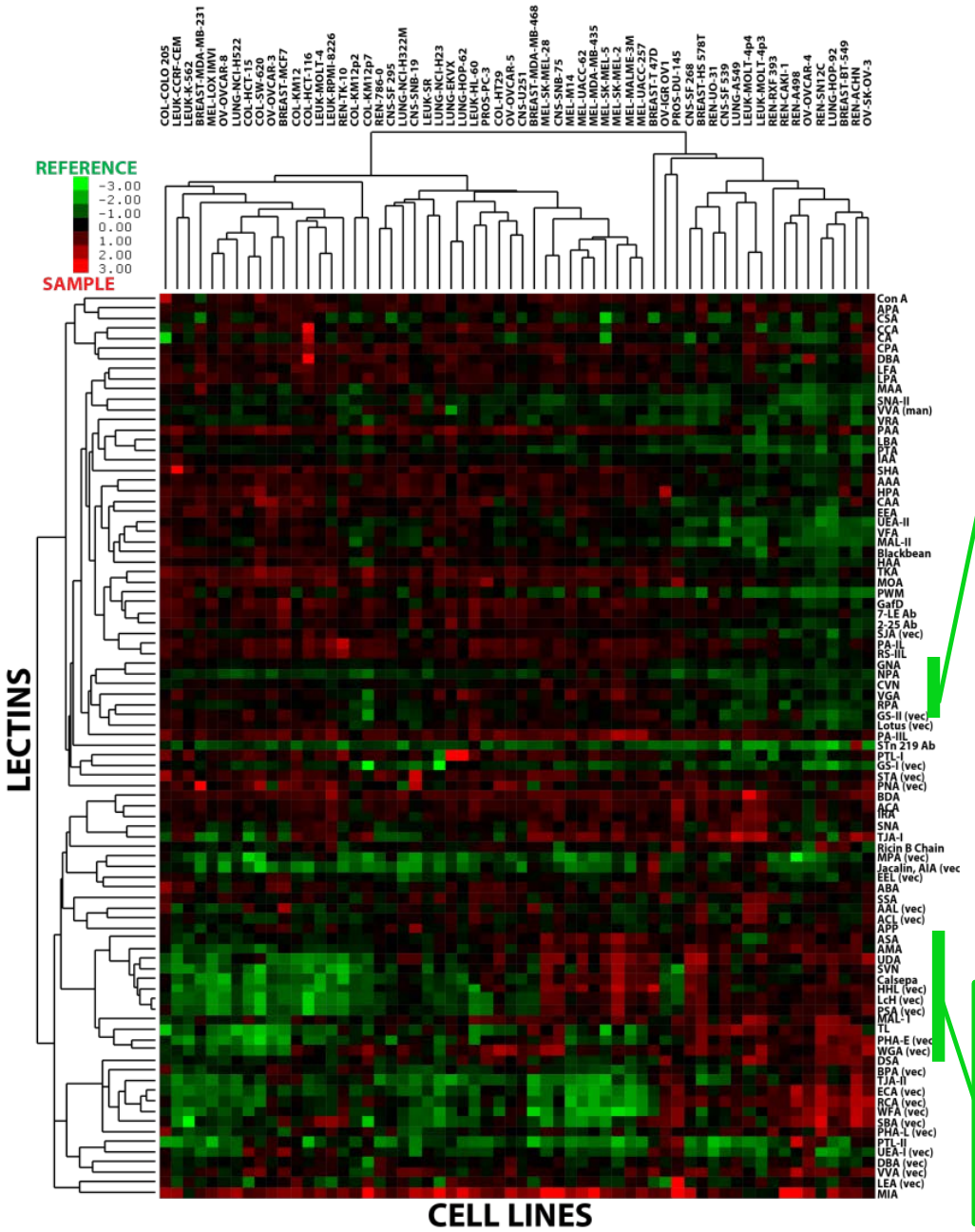


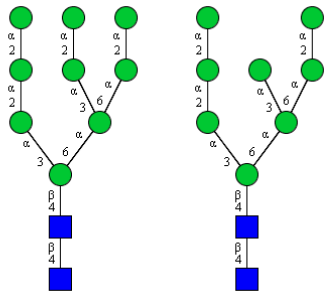




Blood groups/Gal/GalNAc
BDA, RICIN, ABA, AIA, AAL

Glycosylation signatures: Mannose

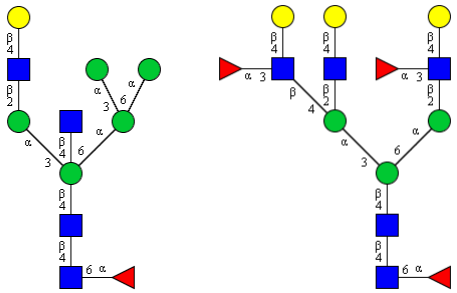




RENAL

MELANOMA
LUNG,COLON

Mannose (high)
GNA, NPA, CVN, RPA

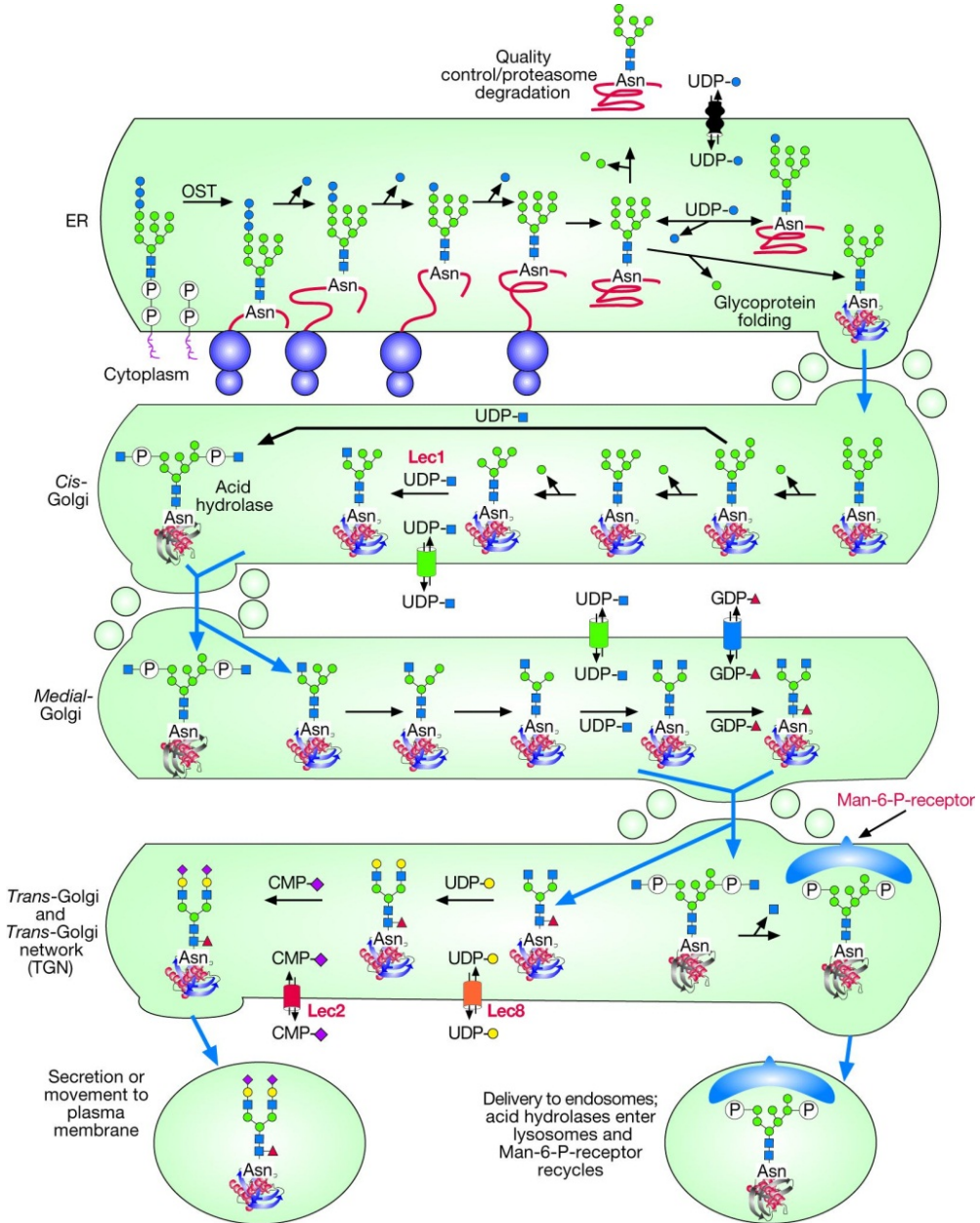


Mannose (complex)
UDA, SVN, Calsepa, HHL, LcH, PHA-E, MAL-I, PSA

COLON, LUNG

MELANOMA, RENAL

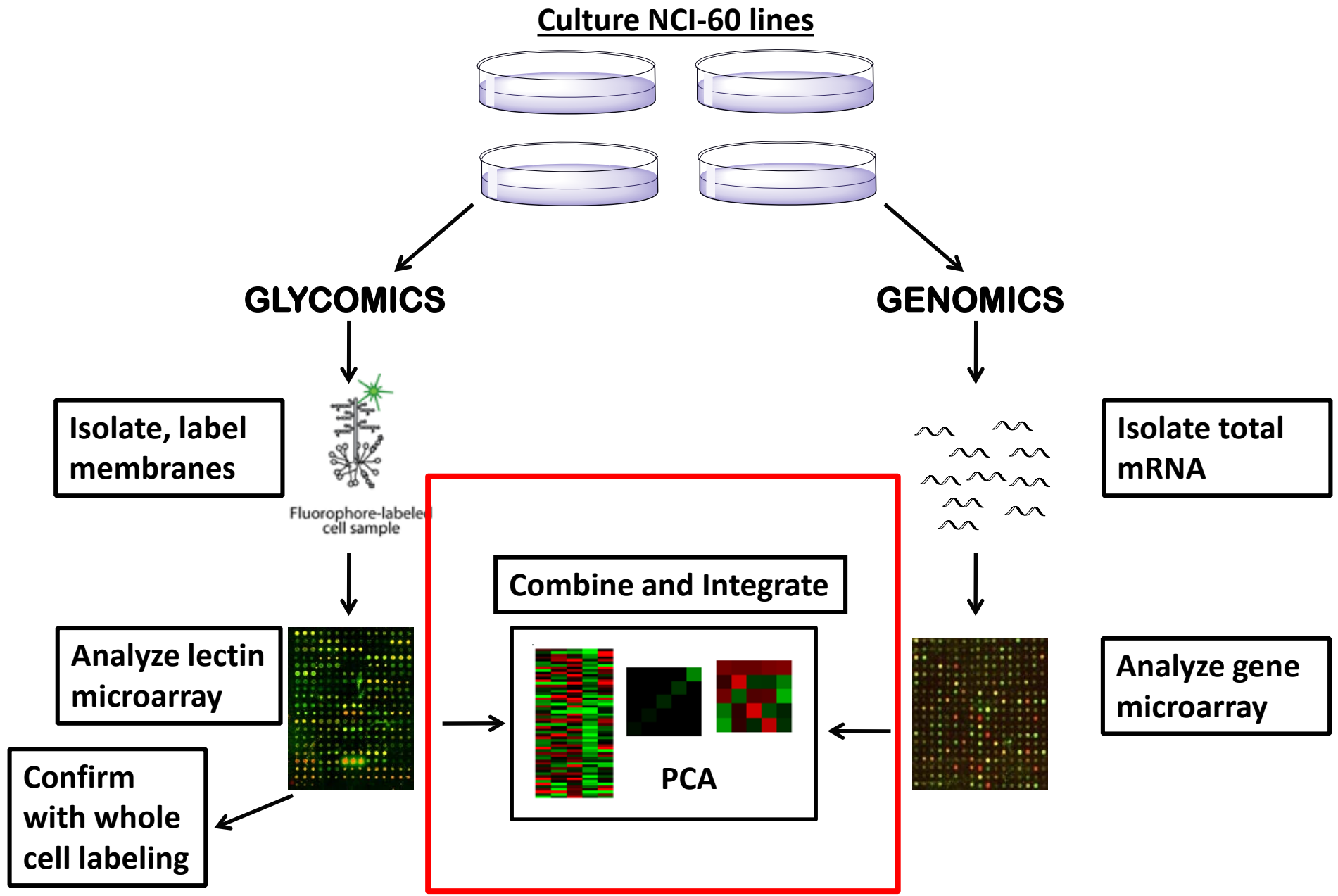
Maturation of *N*-linked glycans



- **$Glc_3Man_9GlcNAc_2$ precursor is transferred to nascent polypeptide in ER**
- **Upon proper folding, glycan is trimmed to high-mannose and, potentially further modified to hybrid/complex**

Essentials of Glycobiology

Experimental Strategy



Principal Component Analysis

- **Method to decompose a matrix into an orthogonally transformed set of variables**
- **This method can be used to identify significant, co-varying patterns in large data sets**
- **Can be applied to either lectin or mRNA array data individually or both data sets combined for SYSTEMS BIOLOGY examination of glycosylation pathways**

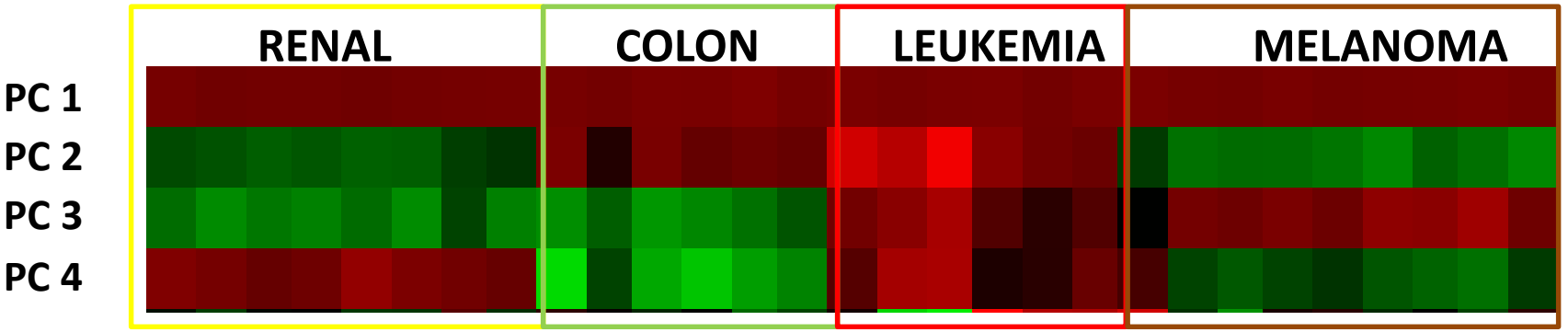
PCA Matrix

$$M = U\,V^T$$

Eigenarrays vs genes/lectins

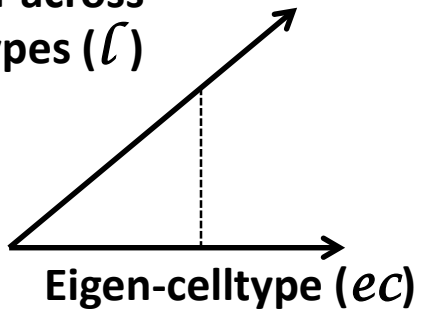
Arrays vs eigengenes/eigenlectins

PCA Analysis of the glycome



Projection correlation = $\ell \,||\, ec$

Gene/lectin vector across cell types (ℓ)



Glycomic Signatures

Component 2	APA, Con A, DSA , GNA, LcH, LPA, NPA, PAA ,SNA (vector), UDA, HHL, MAL-II, CVN, SVN, GRFT, ACA, PSL, TJA-I, AMA, Calsepa, IRA, RS-IIL	High mannose, α2,6 sialic acid (*N*-linked)
Component 3	VRA, PTL-II, HAA	Terminal GalNAc (*O*-linked)
Component 4	TKA, PapGII	Glycolipid

PCA Matrix

$$M = U\,\mathrm{V}^T$$

Lani Pilobello

Patterns of NCI-60 Gene Expression Regulation

- **mRNA and miRNA expression patterns investigated across entire NCI-60 (Liu et al, *Mol Biol Cell,* 2010)**

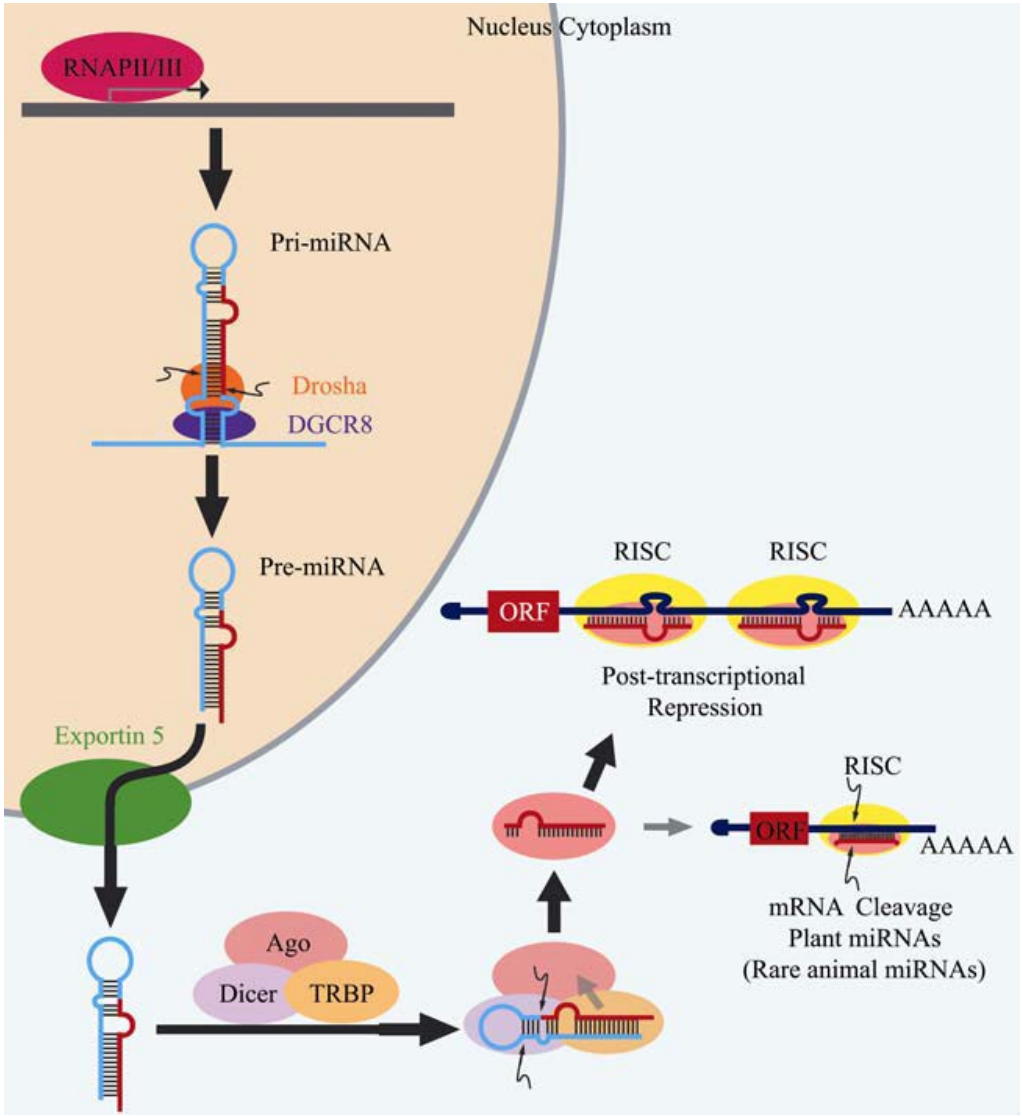
OBSERVATION OF HIGHLY EXPRESSED AND DIVERSE PROBES

-Melanoma (8/9 mRNA, 9/9 miRNA) and leukemia (6/6 mRNA and miRNA) lines cluster together based on expression patterns

-Renal (8/8 mRNA, 6/8 miRNA) and colon (6/7 mRNA and miRNA) also show strong correlation

-CNS and lung cancer show little correlation

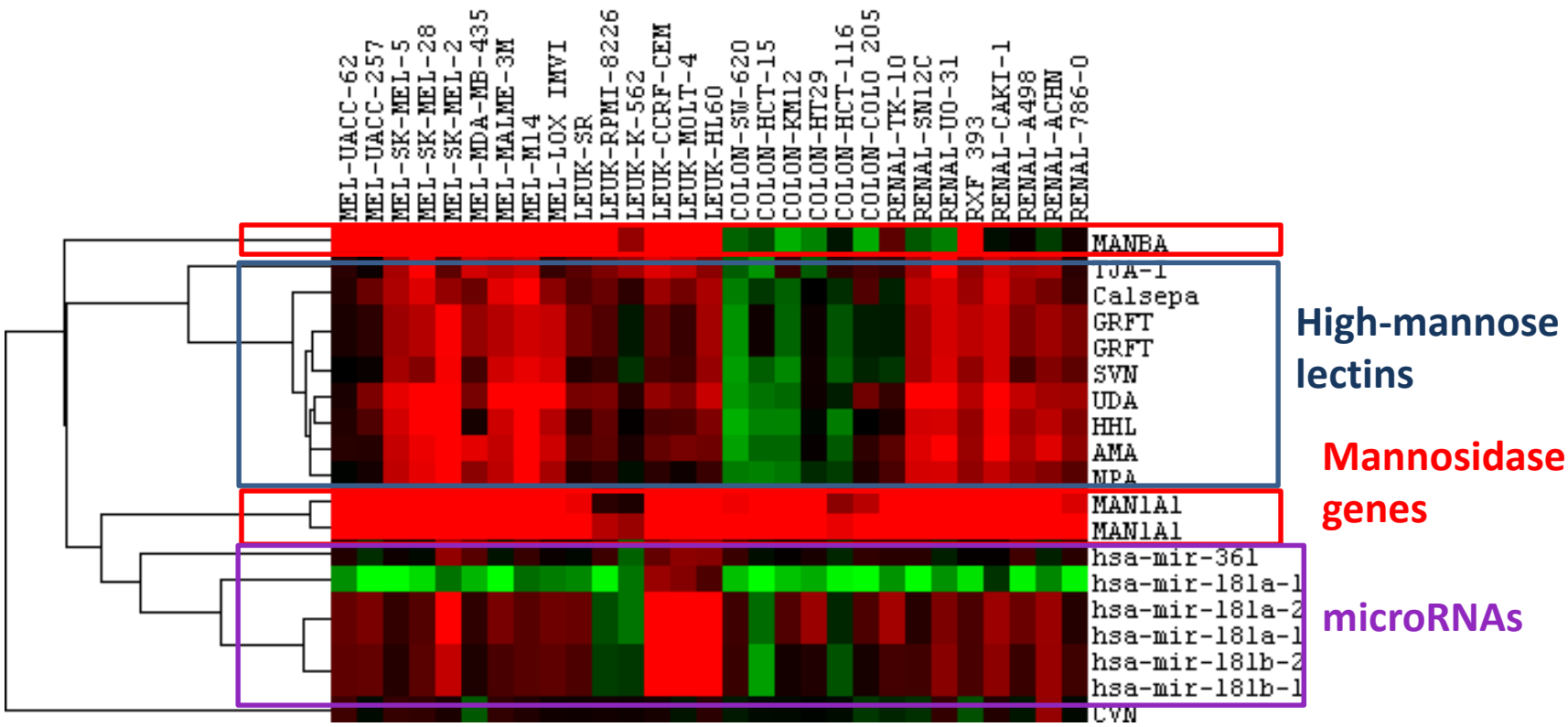
microRNAs (miRNA)



- **miRNA are genomically encoded short (~22 nt) molecules involved in repressing expression of mRNA**
- **Upon processing, miRNAs recruit a post-transcriptional silencing complex to inhibit expression**
- **This mechanism regulates a myriad of cellular functions, including glycosylation**

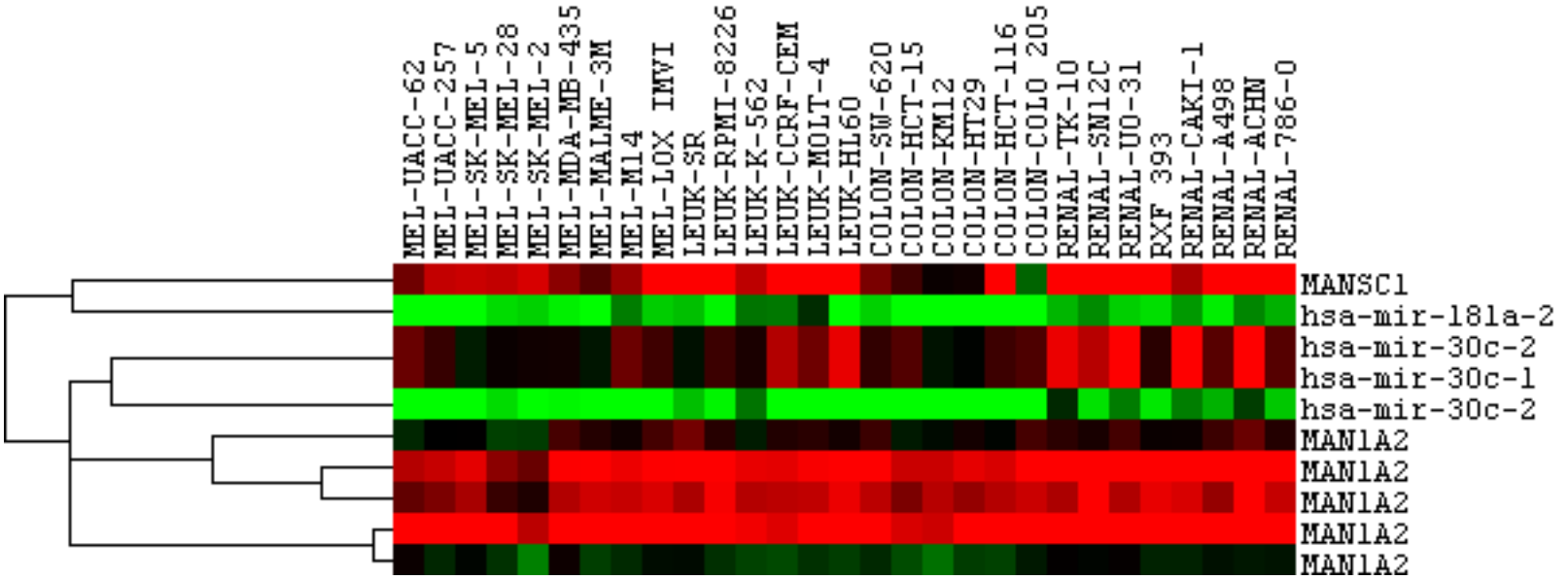
Cong et al, Springer Press, 2009

Integration of miRNA and mRNA data



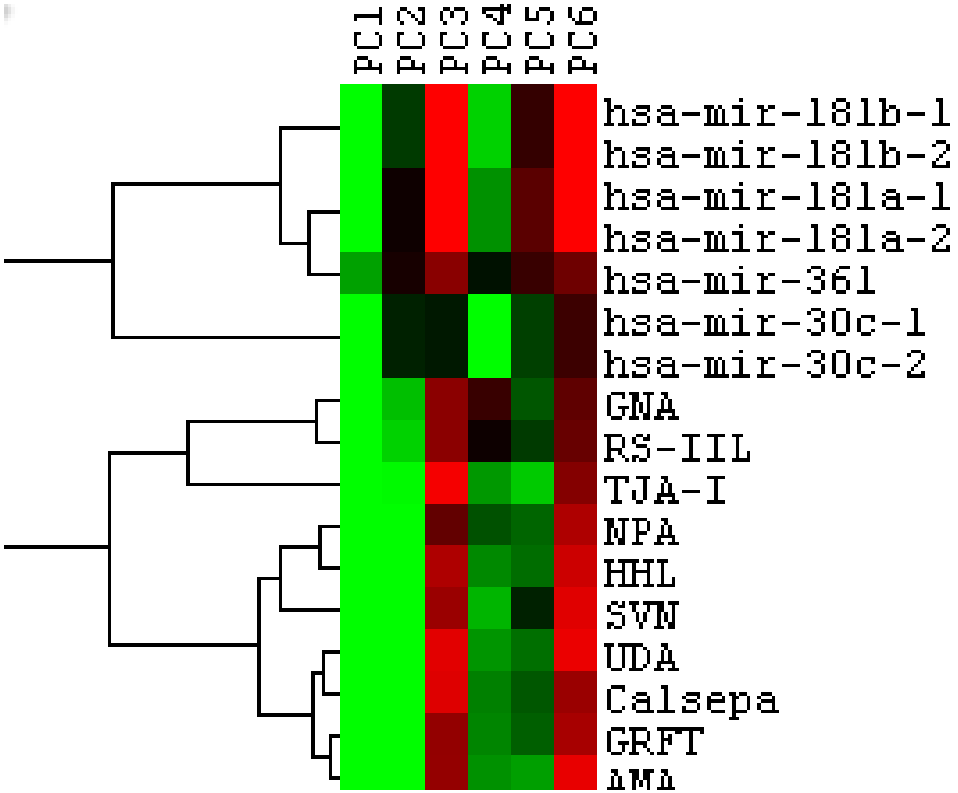
- Lectin and miRNA/mRNA array data sets integrated and clustered without PCA analysis

Analysis of miRNA and mRNA data



Clustering of transcript data shows additional mannosidase targets

PCA integration of glycome and miRNA



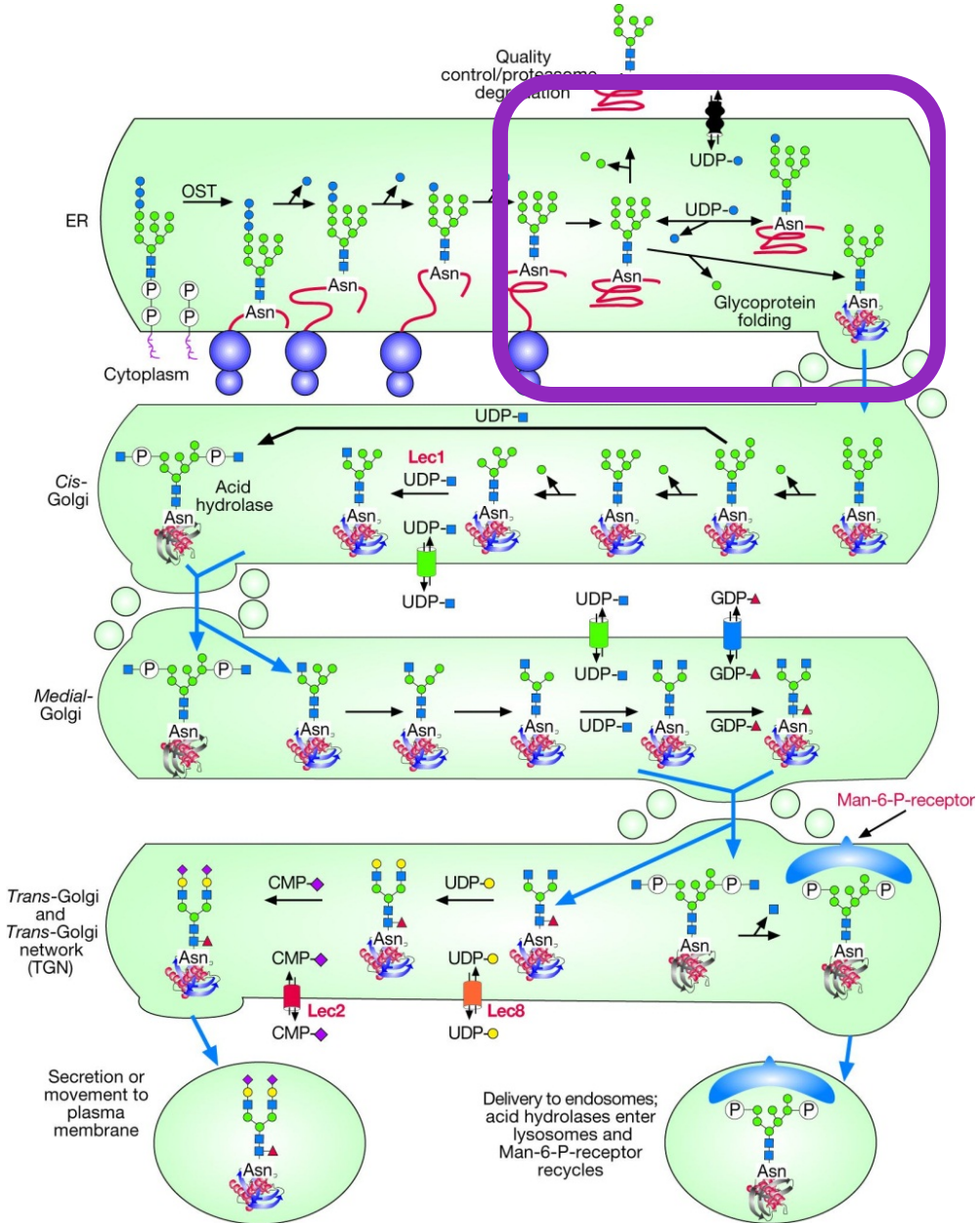
-Integration of lectin and miRNA arrays

-high mannose lectins cluster with and co-vary with four miRNAs across six principal components

-miRNAs co-vary with α-mannosidase I isoforms

-MAN1A1, MAN1A2, MAN1B1 predicted targets of these miRNAs

Maturation of *N*-linked glycans



- $Glc_3Man_9GlcNAc_2$ precursor is transferred to nascent polypeptide in ER
- Upon proper folding, glycan is trimmed to high-mannose and, potentially further modified to hybrid/complex
- α-mannosidase I controls all hybrid/complex maturation steps

Essentials of Glycobiology

PCA Analysis: Implications for regulation of *N*-linked pathway

- **The lectin microarray data shows distinct separation of high mannose and complex/hybrid structures are cell-type dependent**

- **The mRNA array data shows that α-mannosidase I expression co-varies with mannose lectins**

- **The miRNA array data shows that four miRNAs also co-vary with mannose lectins and α-mannosidase I**

- **Hypothesis: *N*-linked maturation pathway is regulated by miRNA transcriptional control of α-mannosidase I**

miRNA Expression	Increased	Decreased
MAN1 expression	Decreased	Increased
N-linked glycome	High Mannose	Complex/Hybrid

Strategy

Use high mannose (mel, ren) and complex (mel, col) lines

Modulate miRNA expression

Isolate total RNA

Analyze mannosidase gene expression (RT-PCR)

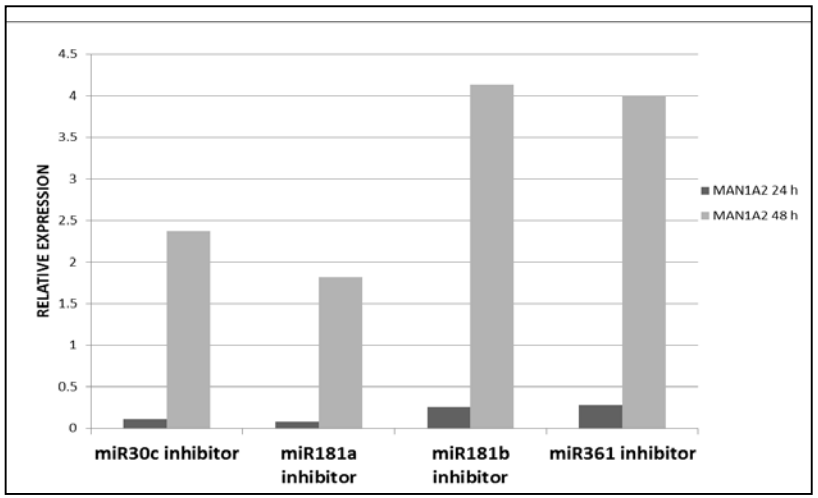
Isolate total protein

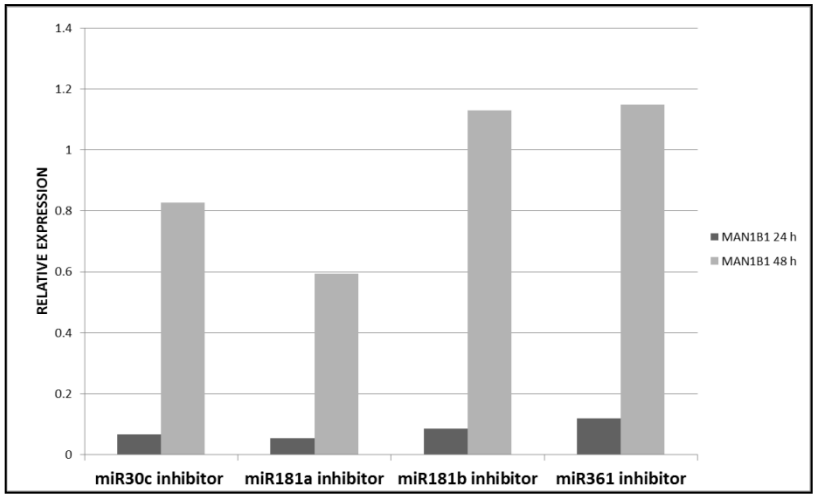
Analyze mannosidase protein expression (Western)

Analyze glycome (lectin microarray)

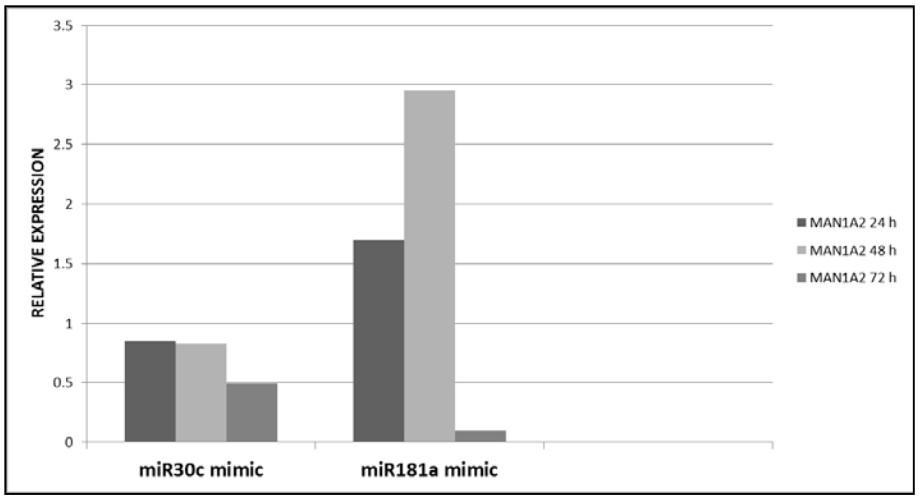
miRNAs affect expression of α-mannosidase I in complex-expressing renal cancer line SN12C

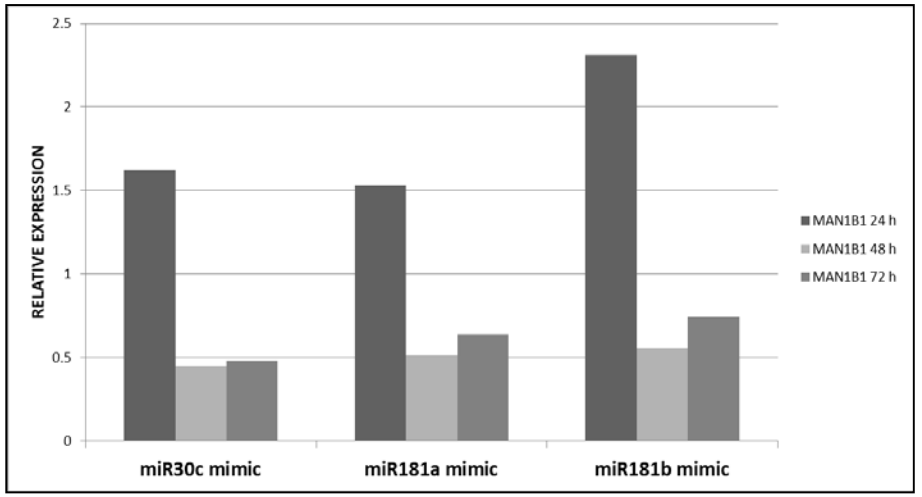
miRNA inhibitors





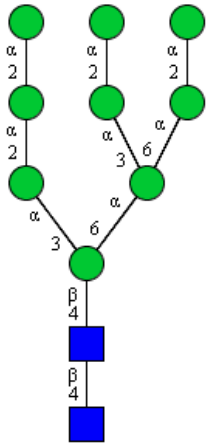
miRNA overexpression



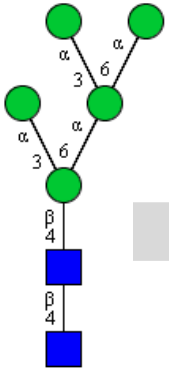


Praveen Agrawal

A model for cell-type specific carbohydrate expression?



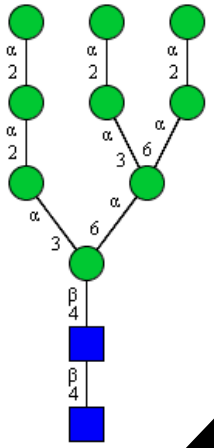
α-mannosidase I



miR-30c
miR-181a
miR-181b
miR-361

Expression of hybrid/complex glycans (renal)

A model for cell-type specific carbohydrate expression?



α-mannosidase I

miR-30c
miR-181a
miR-181b
miR-361

Expression of high mannose (colon)

Expression of hybrid/complex glycans (renal)

Future Directions

- **Fucose and galactose primary component analysis suggests further subcategories of glycome regulation**
- **Evidence suggests Golgi transport structural proteins also map to our lectin microarray observations**
- **Investigate the glycome's role in melanocyte to melanoma cancer transition**

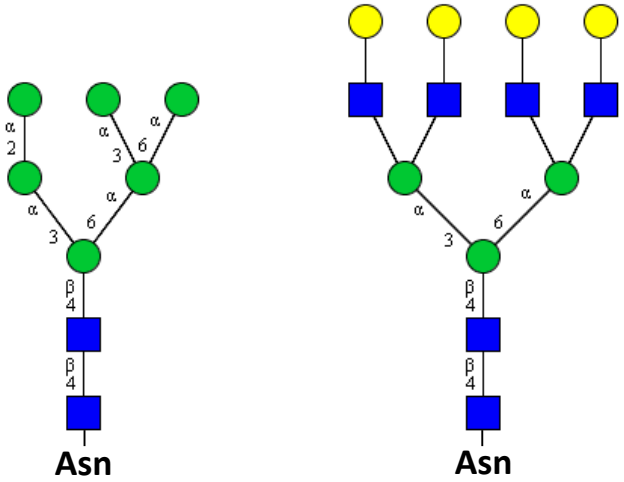
Agenda

- Carbohydrates are a diverse and critically important class of biological macromolecules

- Microarray strategy to identify glycome regulation

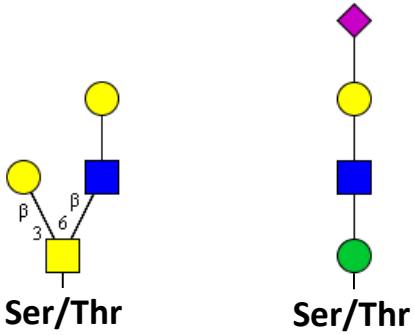
- **Biochemistry and biological relevance of *C*-linked glycosylation**

N-linked glycosylation



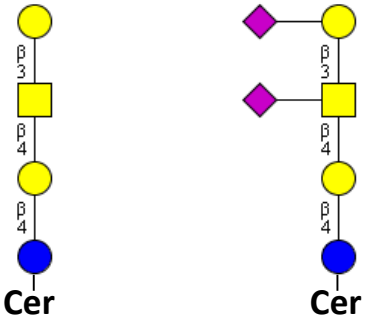
Functions: protein folding and trafficking; signaling ligand

O-linked glycosylation



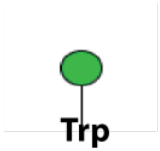
Functions: protein folding and trafficking; ECM composition

Glycolipids



Functions: Cell surface receptors; membrane composition

C-linked glycosylation



Functions: Unknown

Background: *C*-linked glycosylation

- **First identified in human RNaseB by Edman degradation, NMR and MS (Hofsteenge et al, *Biochemistry*, 1994)**

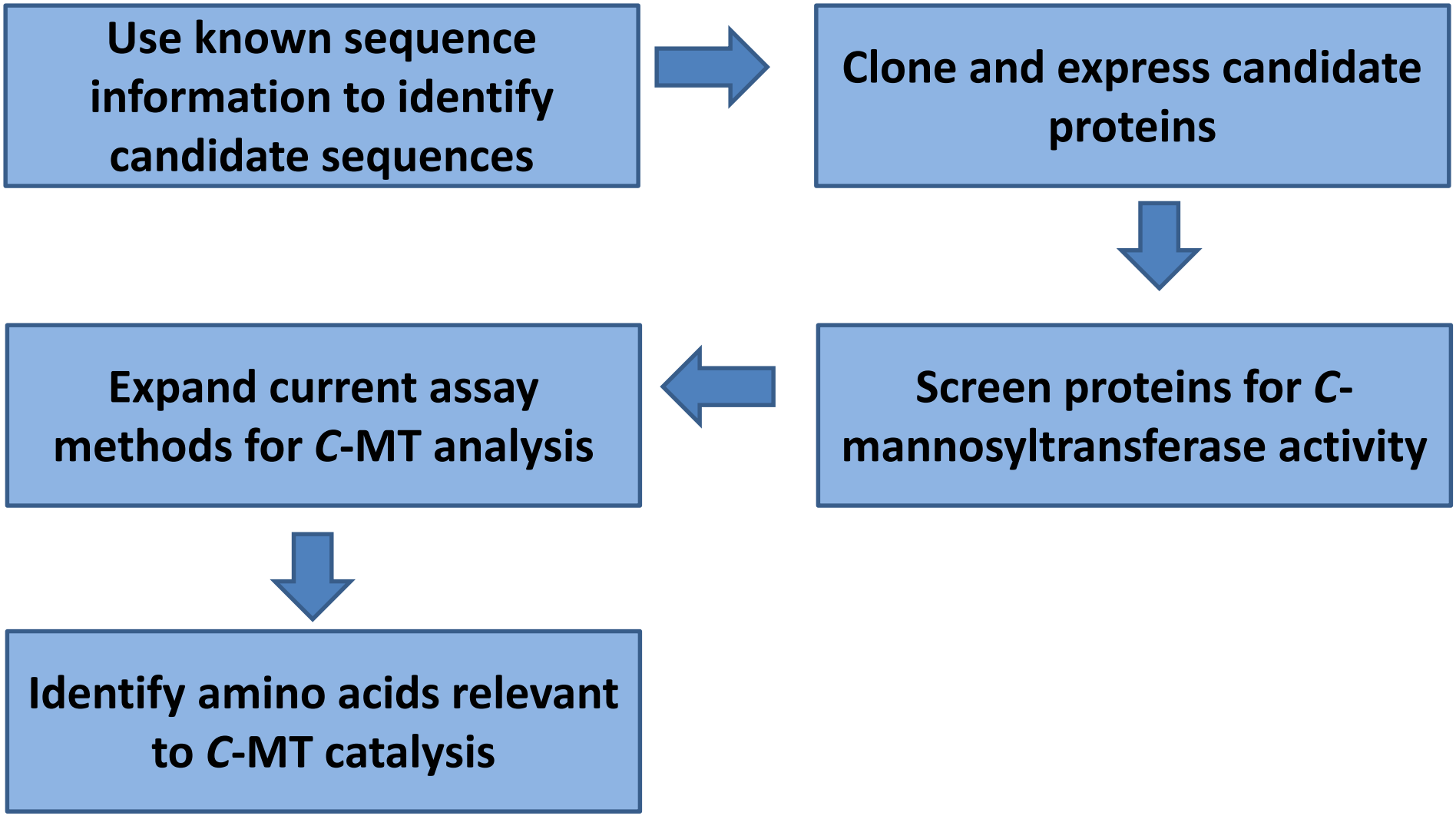
- **Mannose is transferred from dolichyl-phosphate-mannose (not GDP-mannose) by an endoplasmic reticulum-associated protein (Doucey et al, *Mol Biol Cell*, 1998)**

- **Several dozen human proteins confirmed to be *C*-mannosylated; over 2500 candidates contain W-x-x-W consensus sequence (Julenius, *Glycobiology*, 2007)**

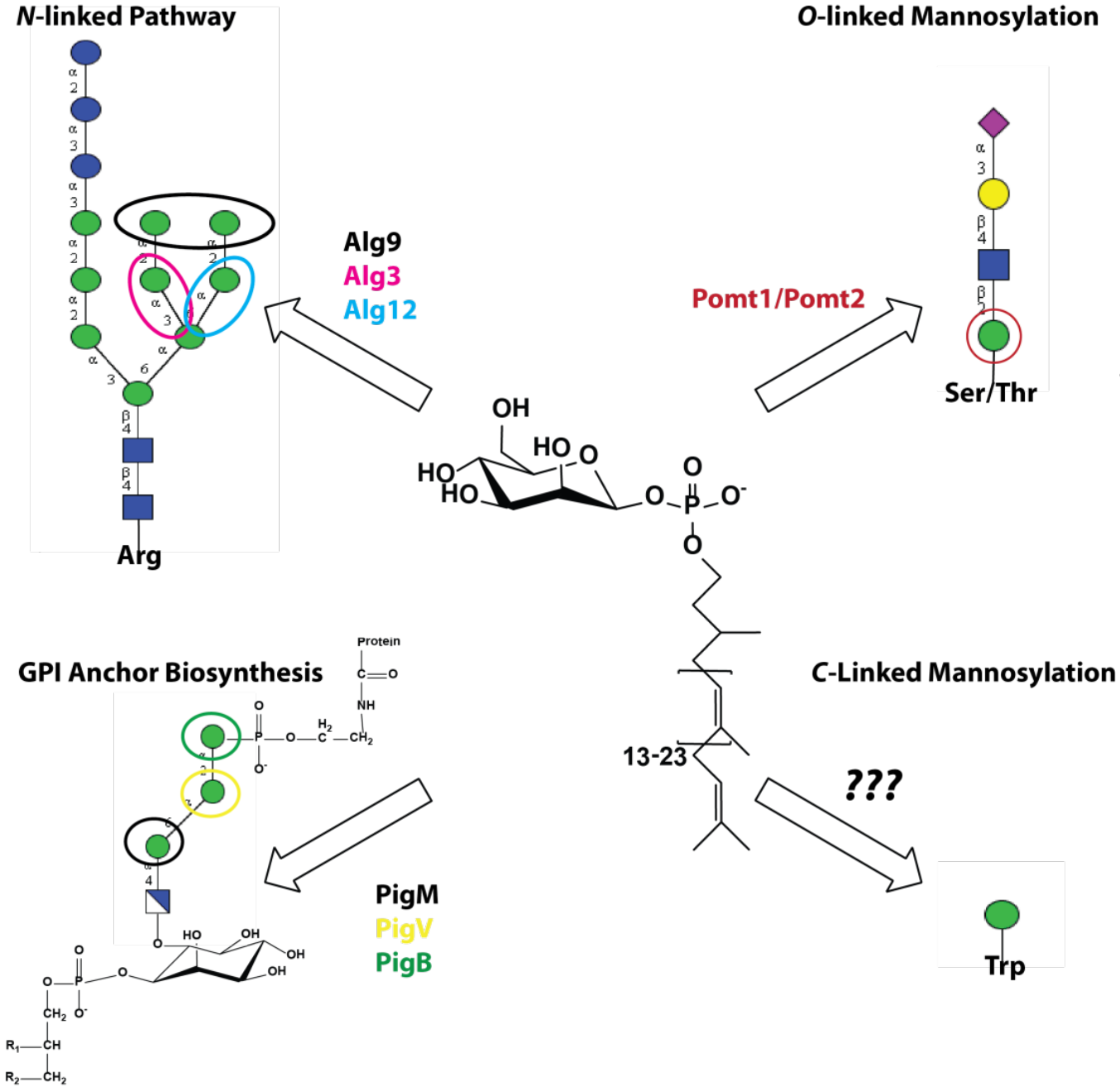
- **Present in insects and *C. elegans*, not in yeast or *E. coli* (Krieg et al, *J Biol Chem*, 1997)**

- **The glycosyltransferase responsible for this modification is not known**

Identification and characterization strategy



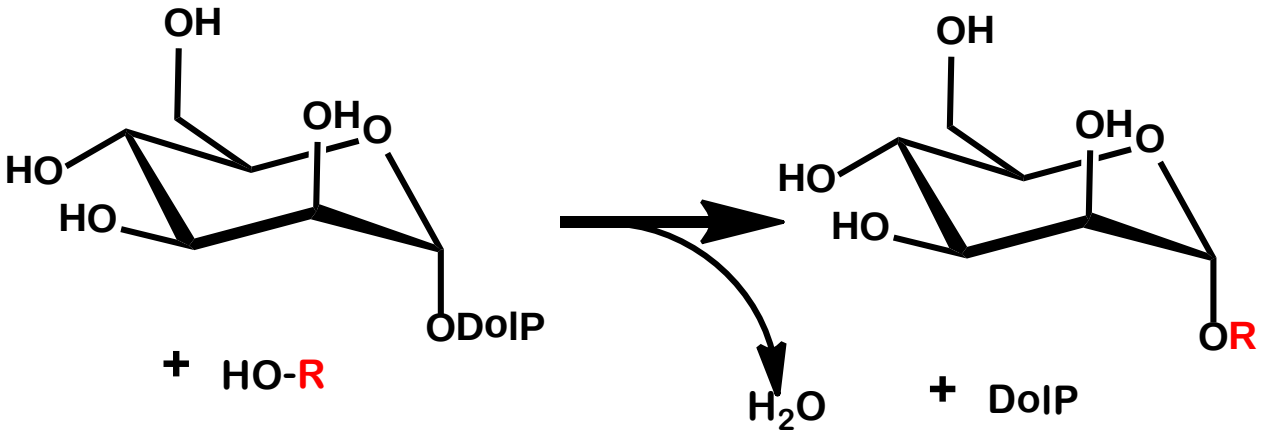
Non-Leloir Transferases utilize dolichyl-phosphate mannose



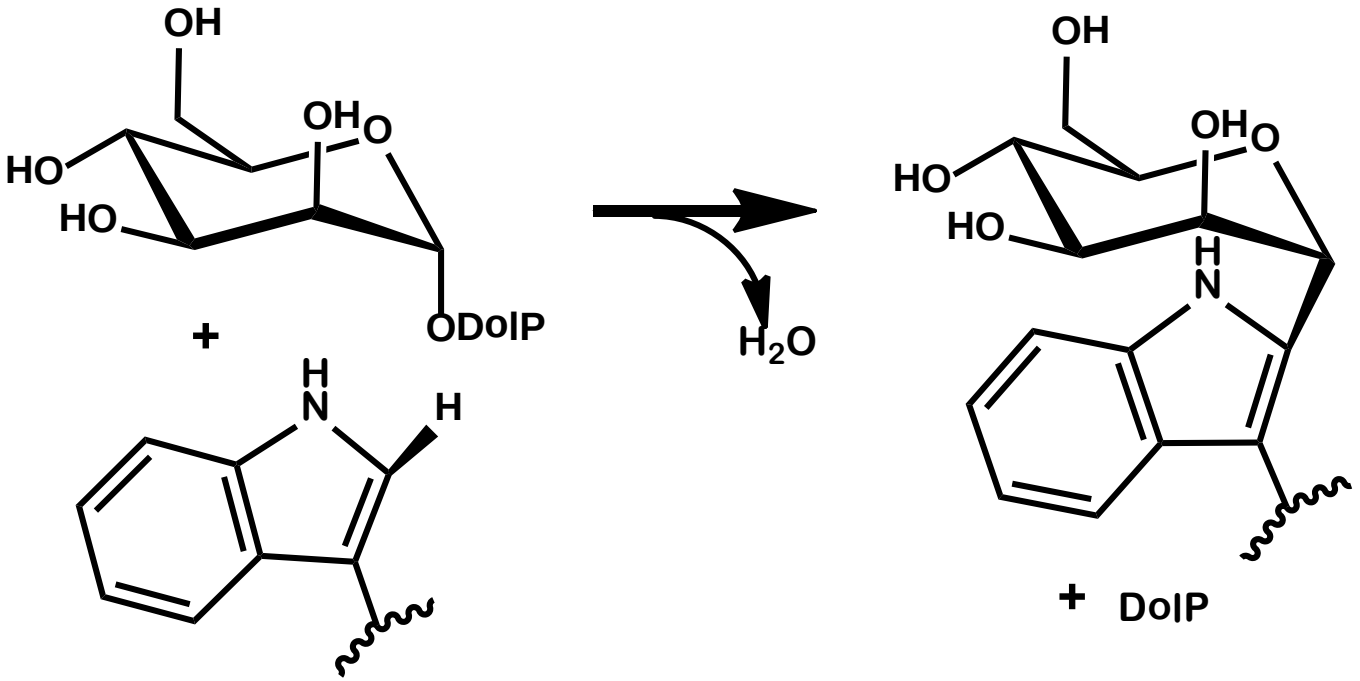
Non-Leloir transferases share a common ancestor and are evolutionarily related in two distinct sequence regions

(Oriol et al, *Mol Biol Evol*, 2002)

Transfer of α–Mannose from Dolichyl-Phosphate-Mannose to Hydroxyl Acceptor

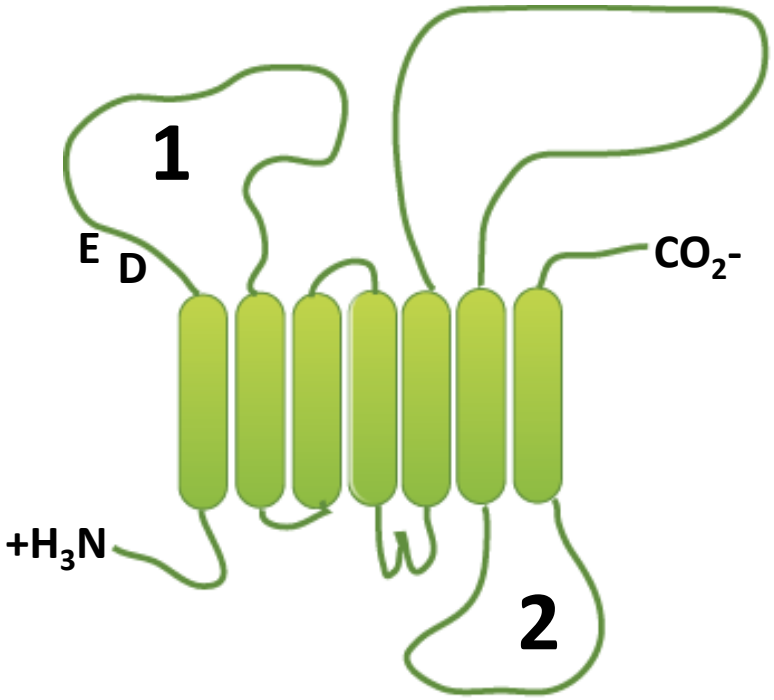


Transfer of α–Mannose from Dolichyl-Phosphate-Mannose to Tryptophan

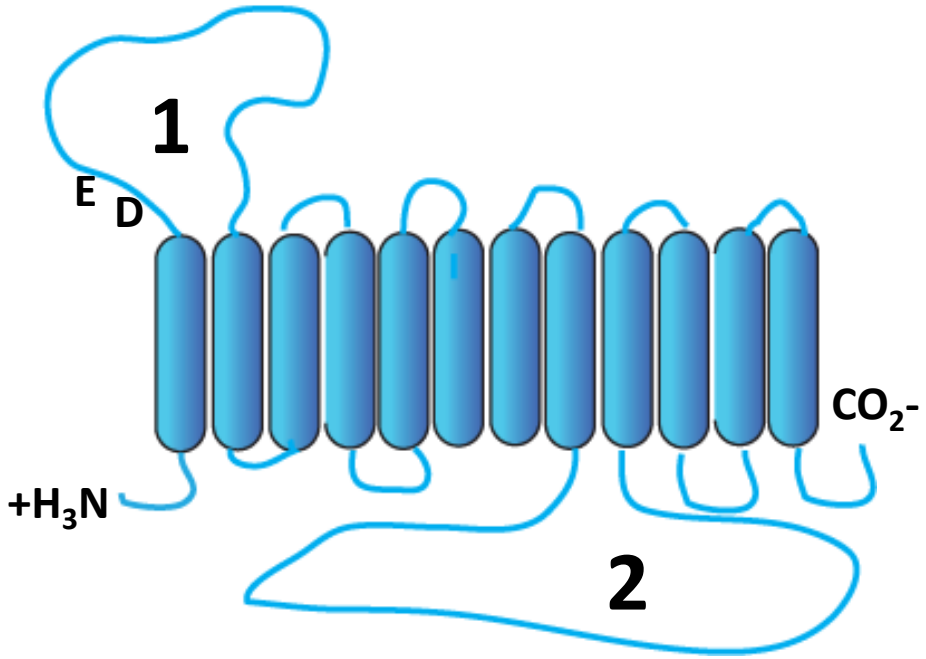


Topology of Non-Leloir Mannosyltransferases

***O*-mannosyltransferases**



***N*-linked and GPI anchor**

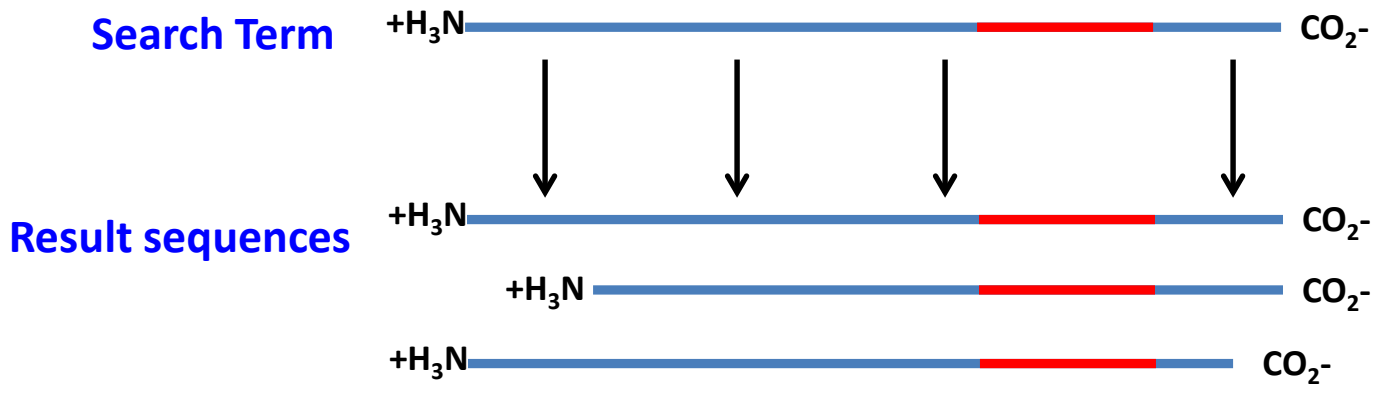


Loop 1: Contains N-terminal acidic domain (DE/EE/DD). Work on POMT1/2 family suggests this is the catalytic domain

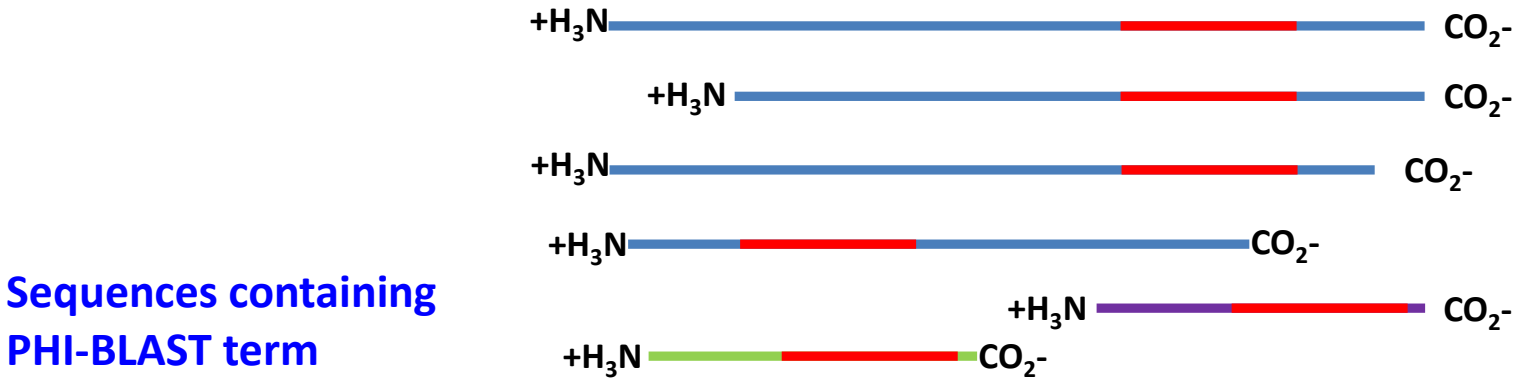
Loop 2: Homologous throughout dolichyl-phosphate-carbohydrate utilizing enzymes

PHI-BLAST Method to Identify Potential *C*-MT Sequences

Step 1: Query human non-redundant protein database with Alg3 sequence, PHI-BLAST loop 2



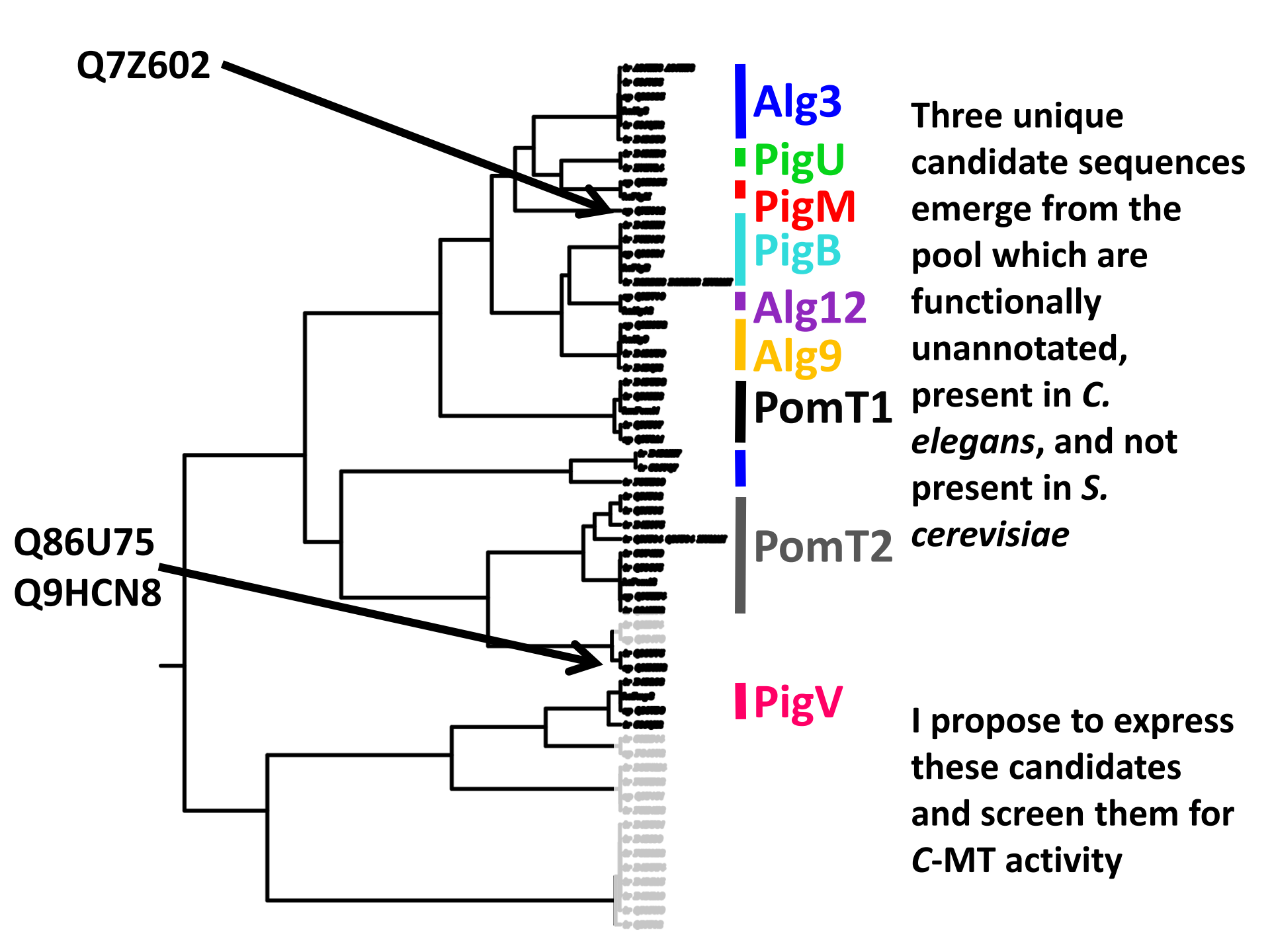
Step 2: Iterate PHI-BLAST search to find "loop 2" sequences in distantly-related sequences



Step 3: a.) Repeat search with 7 other human DPM transferase sequences

b.) Pool and align search results

c.) Query *C. elegans* and *S. cerevisiae* for presence of candidate


Q7Z602
Q86U75
Q9HCN8
Alg3
PigU
PigM
PigB
Alg12
Alg9
PomT1
PomT2
PigV
Three unique candidate sequences emerge from the pool which are functionally unannotated, present in C. elegans, and not present in S. cerevisiae
I propose to express these candidates and screen them for C-MT activity

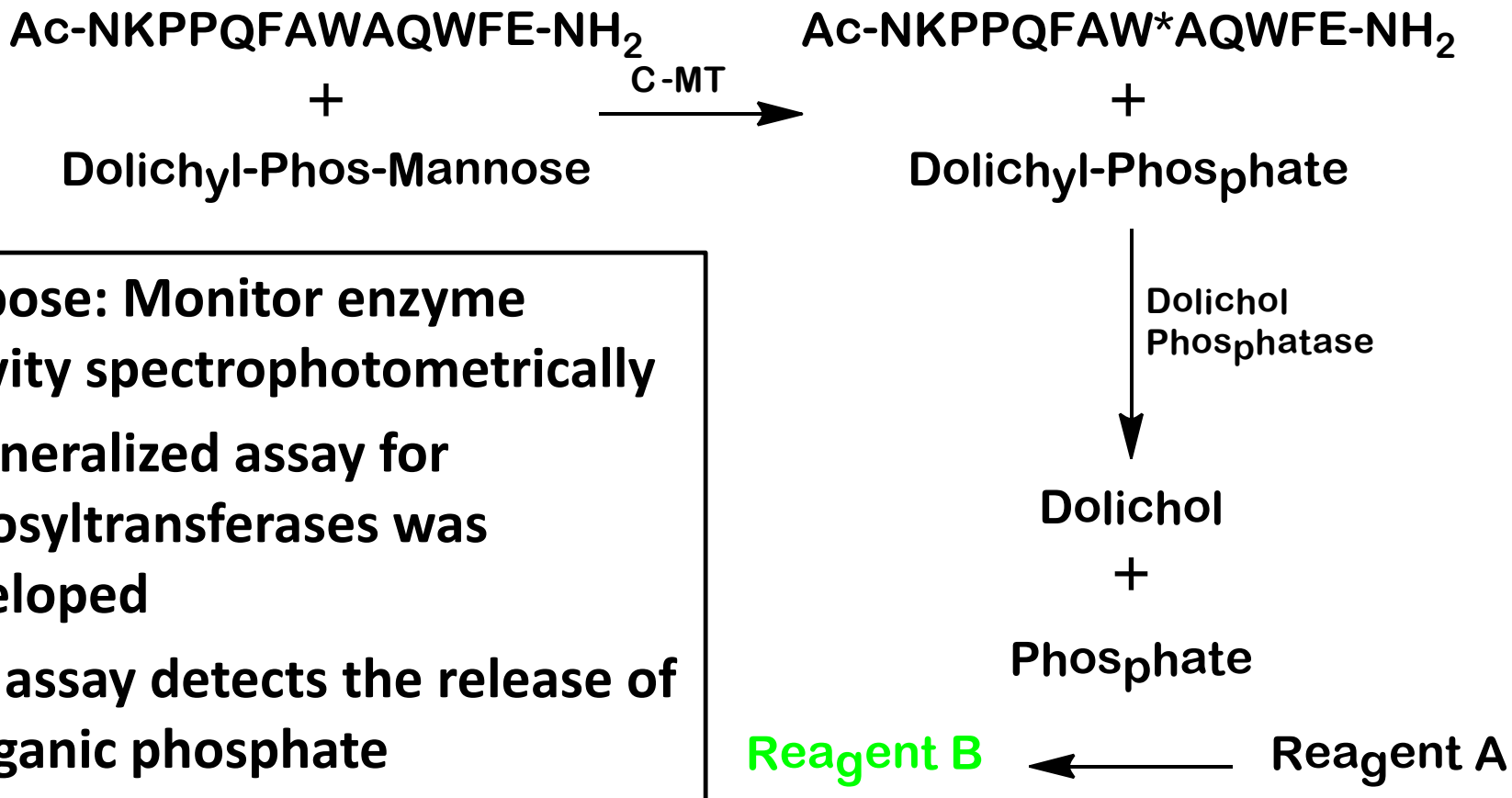
Methods to identify *C*-man-Trp

- **GC-MS (Zanetta et al, *Anal Biochem*, 2004)**
  - **Can identify modified Trp residue from peptide or protein samples**
  - **Cannot identify site of modified tryptophan or be used for kinetic analysis**

- **End point assay (Doucey et al, *Mol Biol Cell*, 1998)**
  - **Can be used to determine reaction kinetic information**
  - **Requires radioactive substrates (^{3}H or ^{14}C)**
  - **Non-continuous assay increases measurement error**

Proposed Development of Coupled-Enzyme Assay for *C*-MT activity

Coupled Assay: C-MT Activity

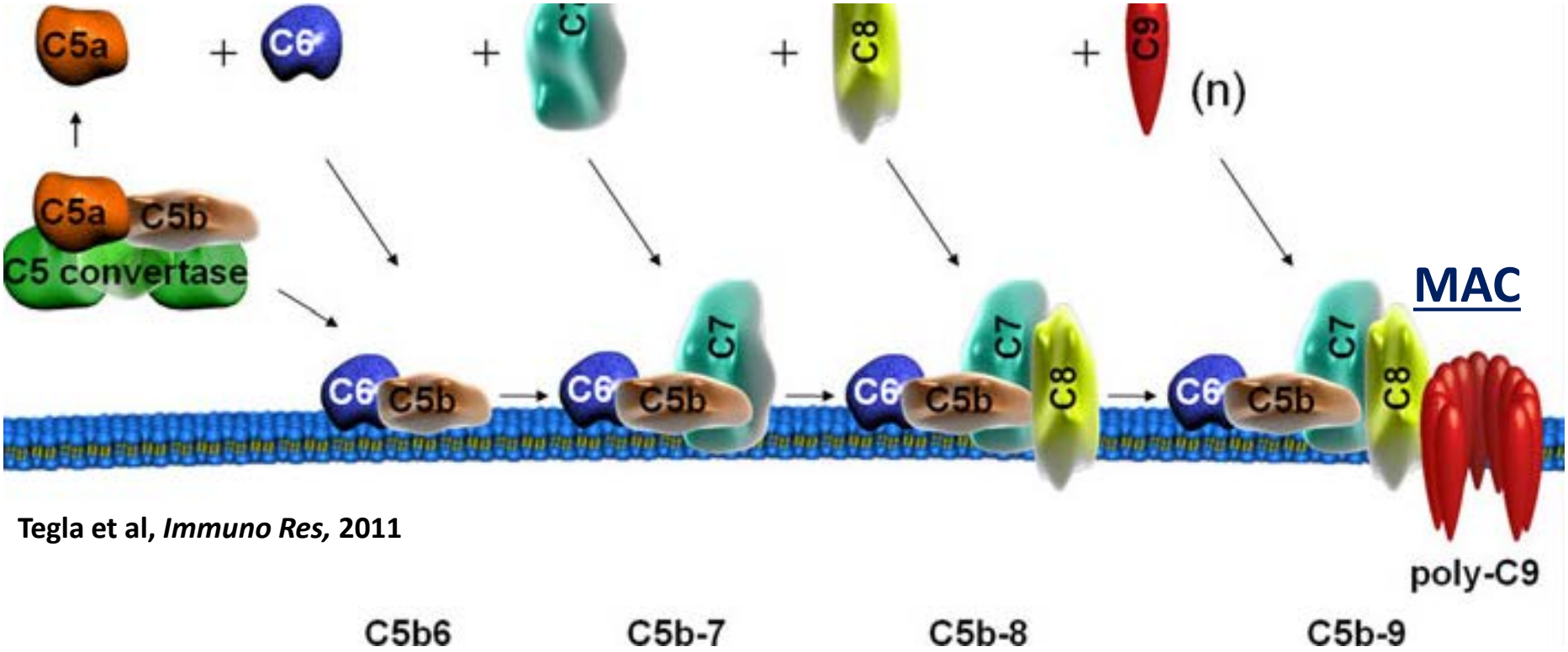


- **Purpose: Monitor enzyme activity spectrophotometrically**
- **A generalized assay for glycosyltransferases was developed**
- **This assay detects the release of inorganic phosphate**

Proposal: Investigation of *C*-mannosylation in innate immune responses

- **Complement Cascade: An encoded response to extracellular stimulus**
- **Stimuli include bacterial cell surface, antigen molecules**
- **Effectors: pathogen opsonization, release of pore-forming toxins**
- **Activators (properdin) and effectors (membrane attack complex) are *C*-mannosylated**

Membrane Attack Complex (MAC): Pore-forming Effector of Innate Immunity

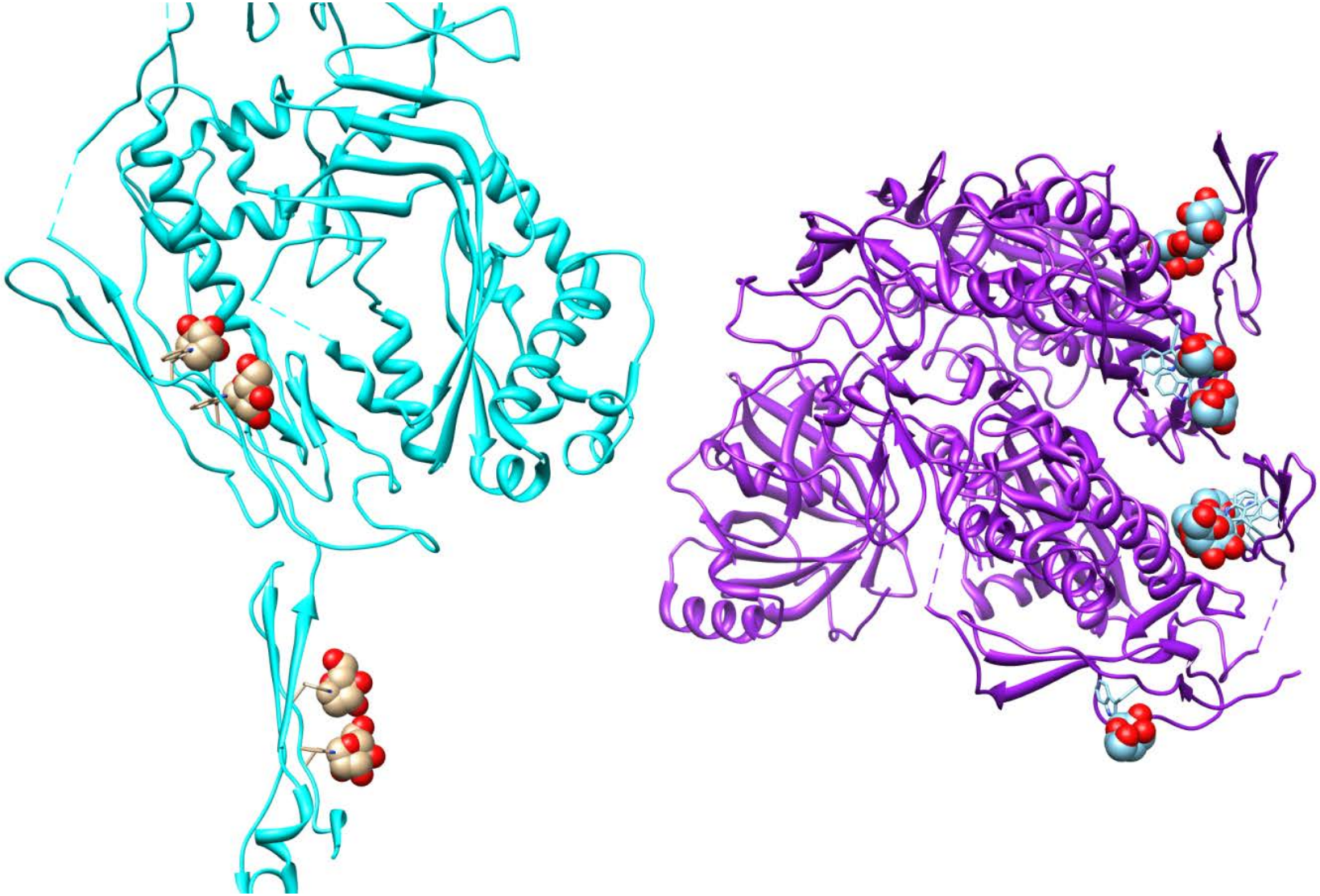


Tegla et al, *Immuno Res,* 2011

- Complement system is activated by external stimulus, such as bacteria

- C6-C9 contain at least one *C*-mannosylated tryptophan residue (Hofsteenge et al, *J Biol. Chem.,* 1999)

C-mannosylated TSR domains in C6 and C8



C6 (cyan), Aleshin *et al*, *J. Biol. Chem.*, 2012 (pdb: 3t5o)
C8 (purple), Lovelace *et al.*, *J. Biol. Chem.*, 2011 (pdb: 3ojy)

Question: What role does *C*-mannosylation have in MAC function?

- **Hypothesis:** ***C*-man-Trp on complement proteins is involved in MAC formation and function**
- **Strategy: Determine what contribution *C*-man-Trp plays by making non-modified mutants and assaying interactions**
- **Approach: Determine protein-protein binding affinities using known assay strategies (surface plasmon resonance, native gel electrophoresis)**

Proposal: *in vivo* identification of *C*-mannosylated proteins

- Over 2500 mammalian proteins may be *C*-mannosylated
- Nearly all remain unconfirmed
- What stimuli enhance *C*-mannosylation?
- Innate immune system activation results in several *C*-mannosylated proteins

Proposal: Role for *C*-linked glycosylation in lipo-polysaccharide-induced TNFα activation

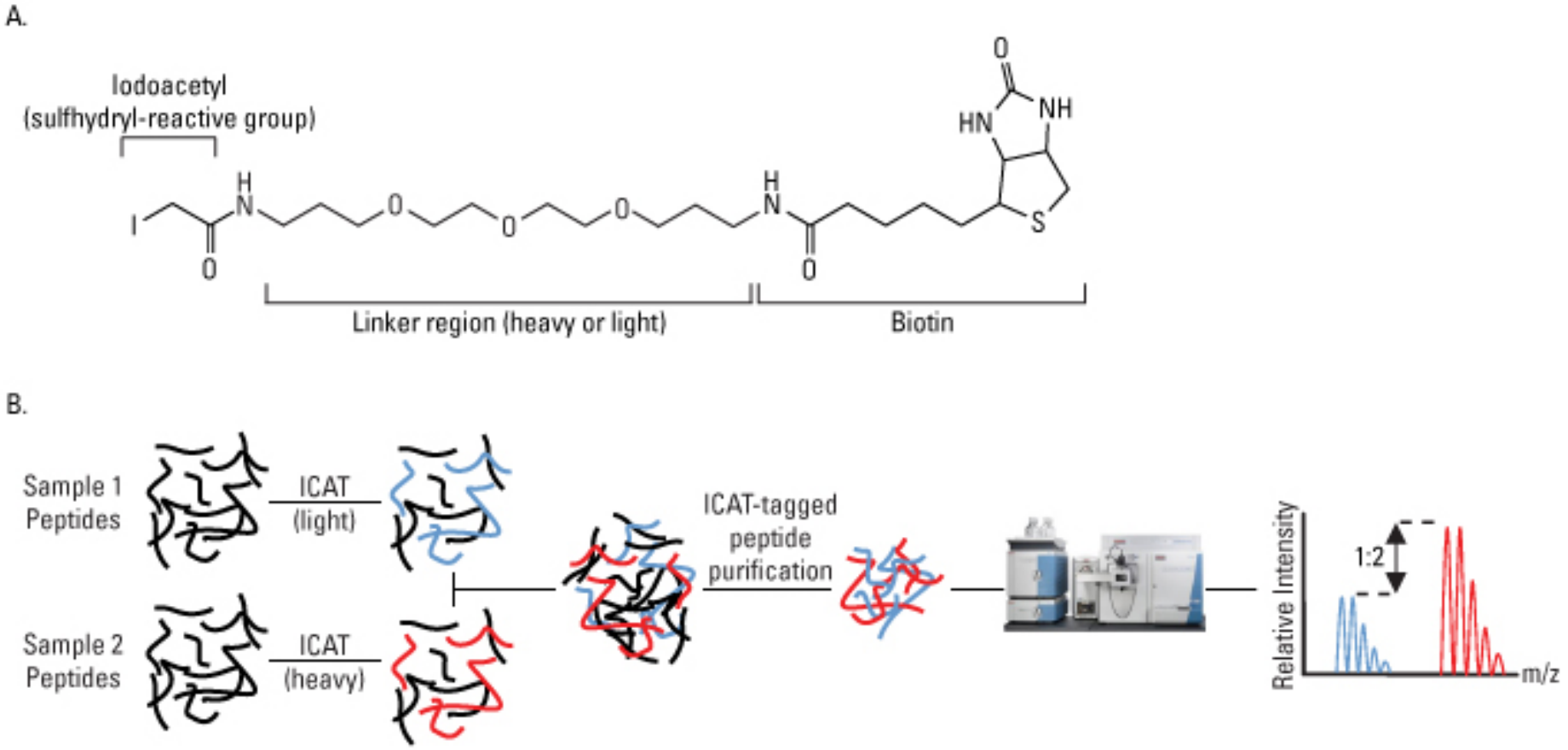
KNOWN: Stimulation of macrophages with LPS and *C*-mannosylated peptides induces TNFα signaling (Muroi et al, *Glycobiology*, 2007) via interaction with Hsc70 (Ihara et al, *Glycobiology*, 2010)

Question 1: What role does *C*-mannosylation have in co-activating this response?

Question 2: What proteins are involved?

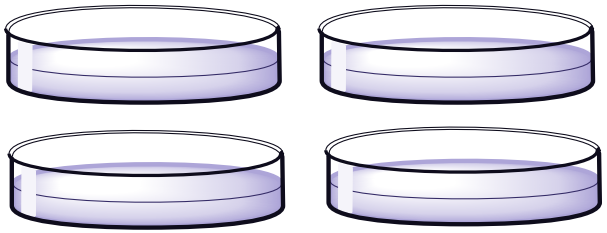
Proposal: Utilize metabolic labeling strategy to identify response of *C*-mannosylated proteins upon immune challenge.

Isotopically Coded Affinity Tags (ICAT): A strategy to identify and quantify peptides



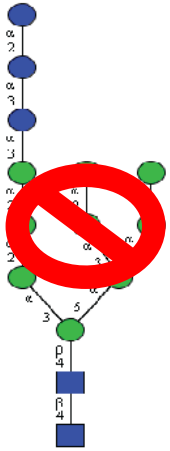
Strategy to identify *C*-mannosylated proteins *in vivo*

Culture hepatocyte cell line



+ tunicamycin

Inhibit *N*-glycan synthesis

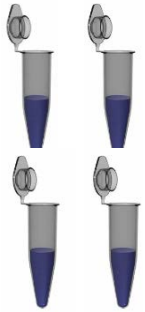


+ labeled DPM

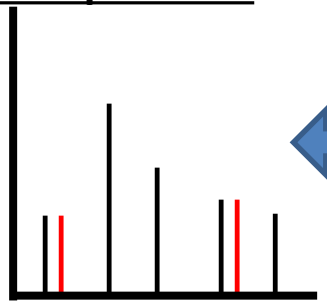
+ unlabeled DPM

Stimulate with LPS

Prepare Samples



Identify *C*-man incorporation



Requirements for Innate Immunity Project

1. Culturable hepatocyte cell line (Gibco)
2. Confirm ICAT-like strategy can differentially identify induced proteins
3. Access to LC/MS-MS (either on site, or in collaboration with CCRC or academic lab)
4. Access to deuterated DPM
   - Commercially available (Omicron)
   - Enzymatically synthesized from GDP-Man

Acknowledgements

Prof. Lara Mahal

Current group members

-Dr. Praveen Agrawal*

-Dr. Bianca Batista*

-Dr. Kanoelani Pilobello*

-Dr. João Ribeiro

-Linlin Wang

-Yaxuan Liang*

-Tomasz Kurcon

-Zhongyin Liu

-Chris Vaiana

-Lauren Zhang

-Sarah Abbassi (undergraduate)

Former group members

-Dr. Ku-Lung Hsu*

-Dr. Lakshmi Krishnamoorthy*

-Will Eng*

-Prof. Eva Hernando (NYU Med)

-Prof. Iman Osman (NYU Med)

-Avital Gaziel-Sovran (NYU Med)

-Dr. Erica Friedman, M.D. (NYU Med)

NIH DIRECTOR'S NEW INNOVATOR AWARD

NATIONAL INSTITUTES OF HEALTH