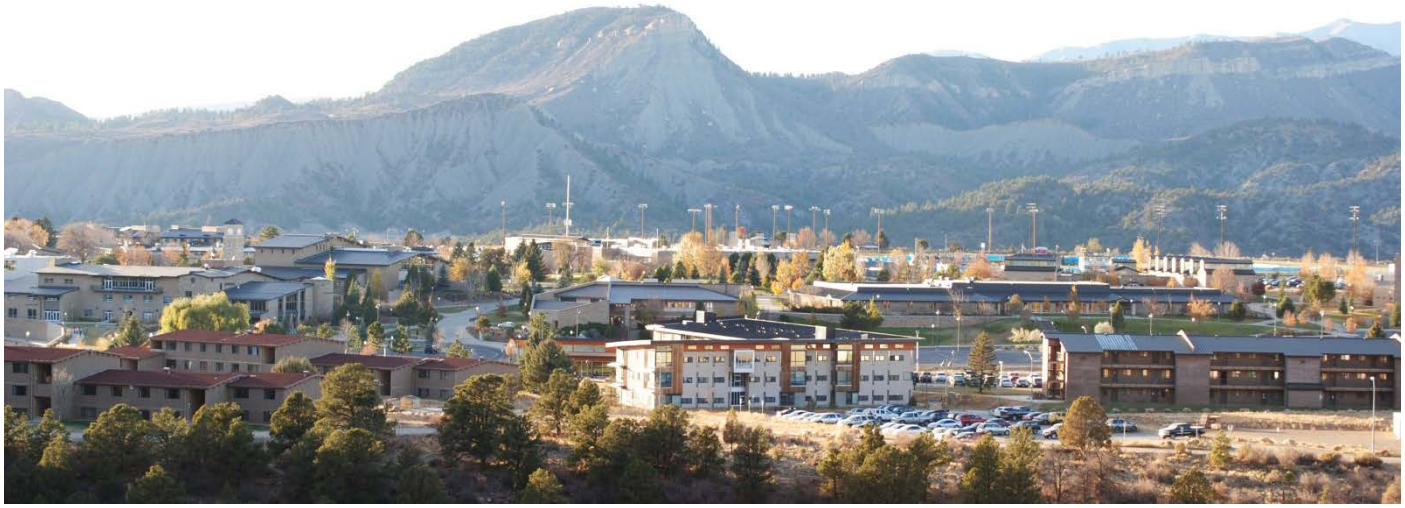
Winter 12-2012

Designing Tools for Studying the Dynamic Glycome

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

FORT LEWIS COLLEGE



Designing tools for studying the dynamic glycome

John Rakus, Ph.D.
Department of Chemistry
Fort Lewis College
December 5, 2012

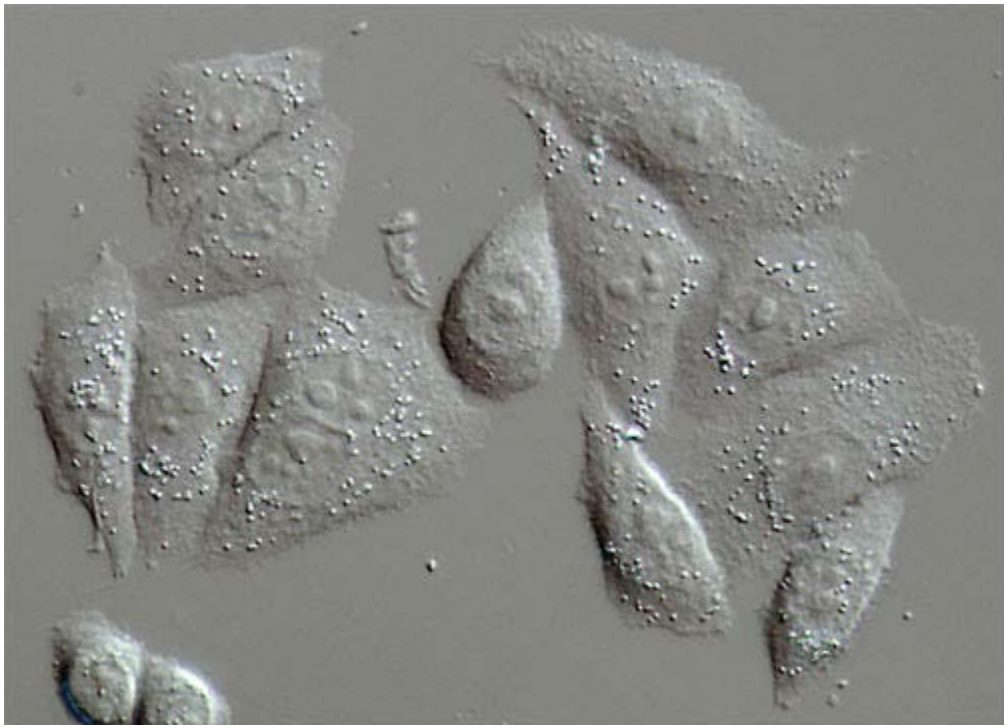
Designing tools for studying the dynamic glycome

John Rakus, Ph.D.
Advisor: Dr. Lara Mahal

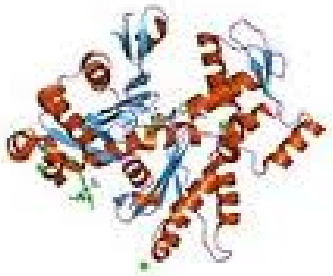

HELVETICA

Cells are primarily compose of three types of biomolecules

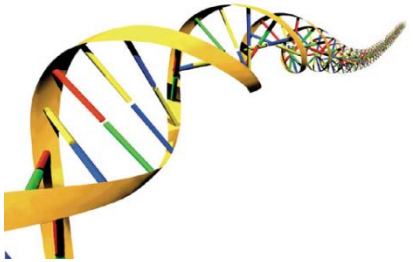
HeLa cell



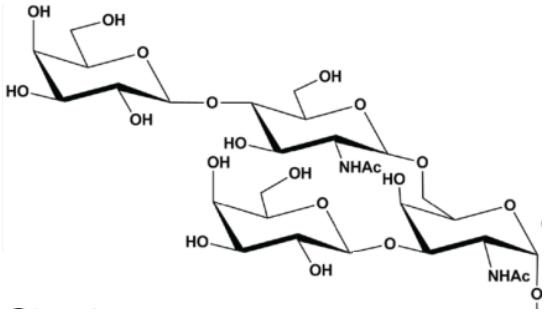
Protein (50% dry weight)



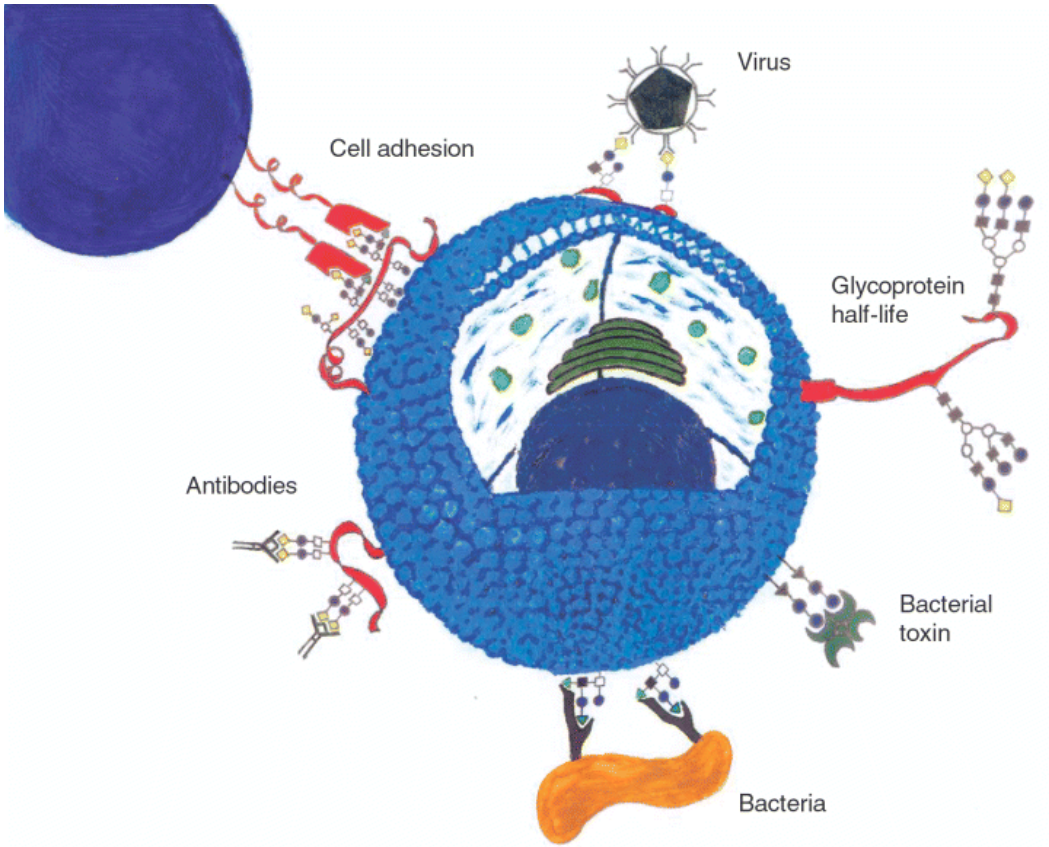
Nucleic acid (25% dry weight)

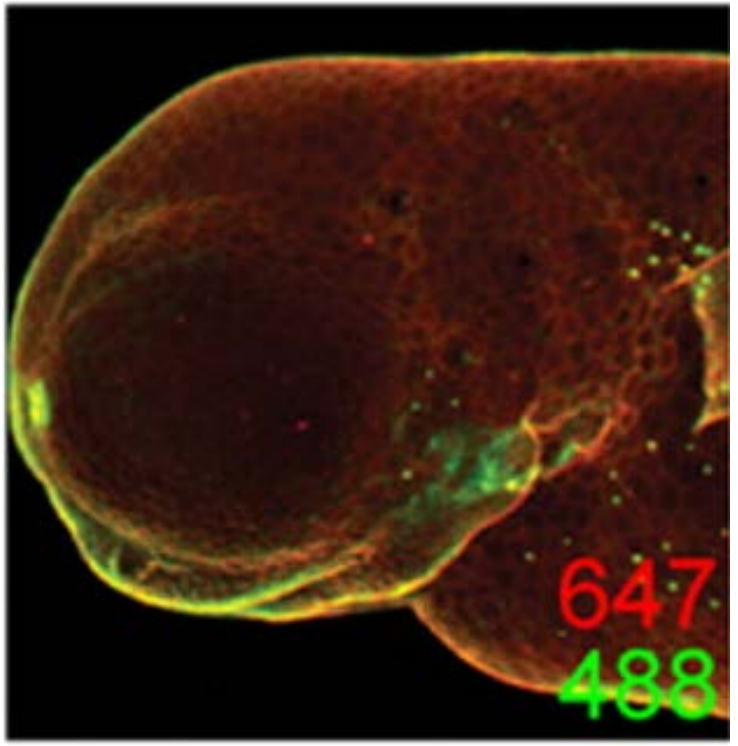


Carbohydrate (10% dry weight)



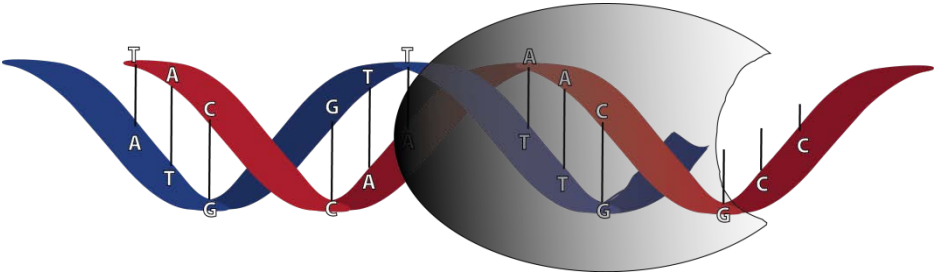
Carbohydrates are pervasive and involved in many cellular interactions

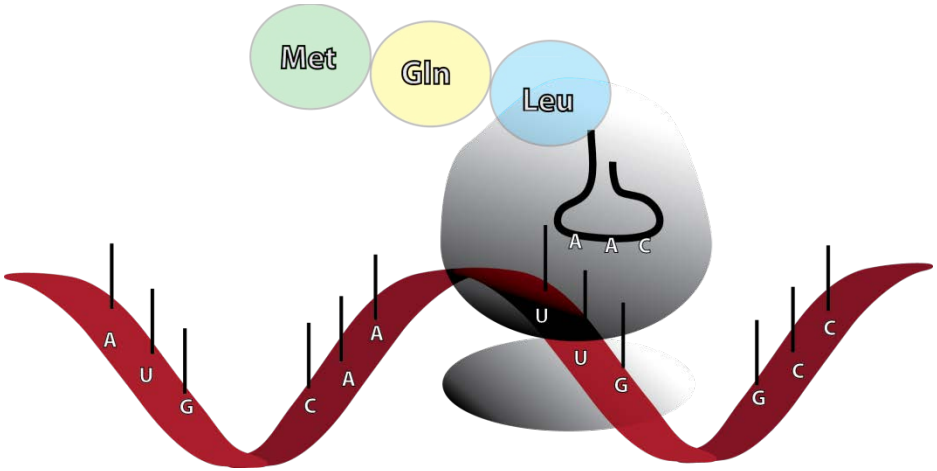




Holgersson et al, *Immuno Cell Biol*, 2005
Laughlin et al, *Science*, 2008

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





Macromolecule: Nucleic acid

Polymerase: DNA Pol or RNA Pol

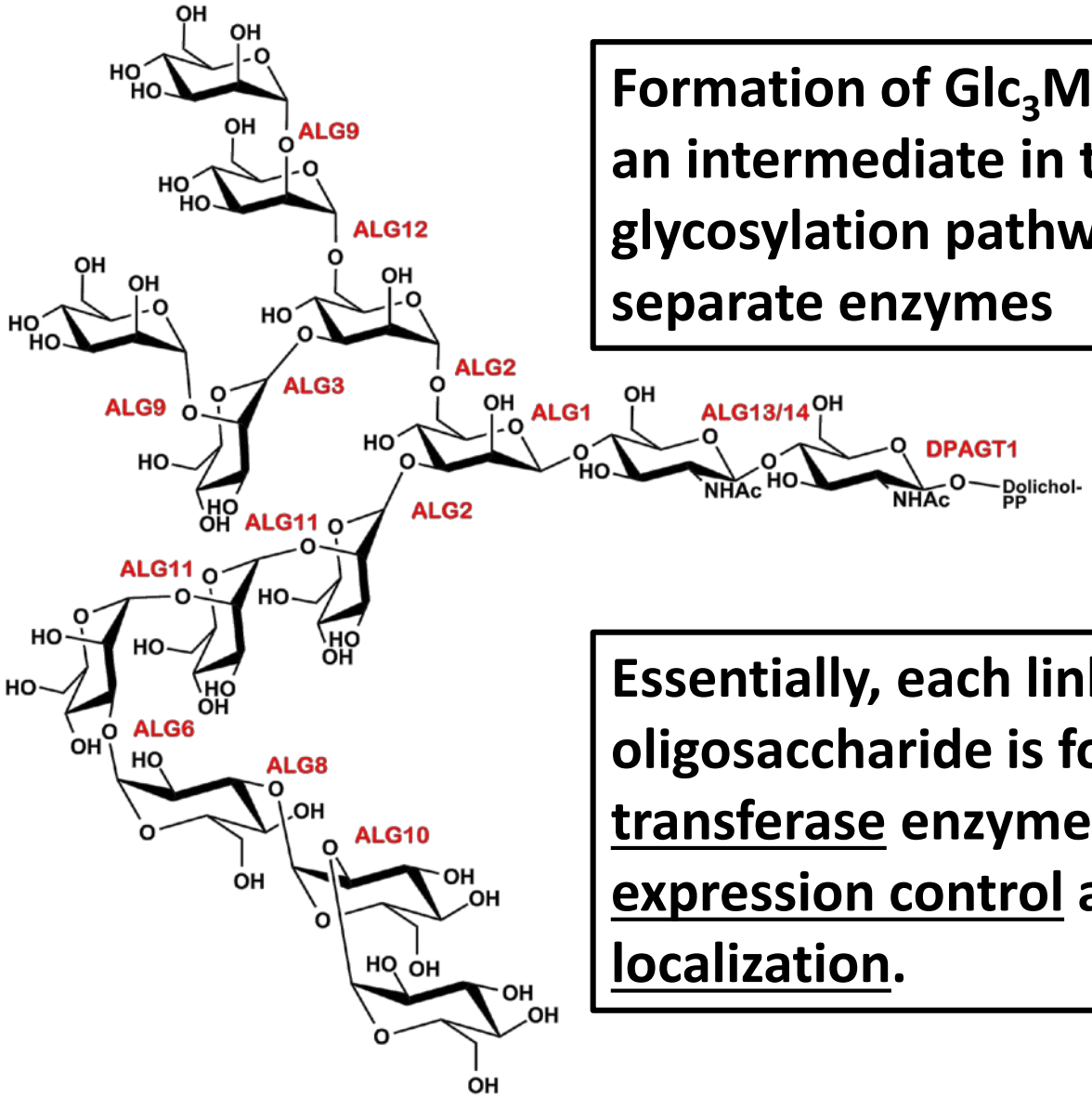
Template: DNA strand

Macromolecule: polypeptide

Polymerase: Ribosome

Template: mRNA strand

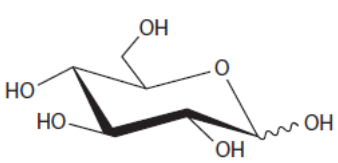
Glycan biosynthesis lacks a dedicated polymerase and genetic template

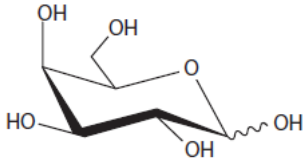


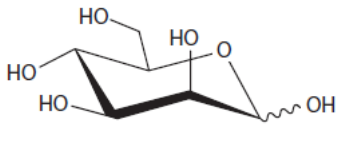
Formation of $Glc_3Man_9GlcNAc_2$-DolPP, an intermediate in the *N*-linked glycosylation pathway, requires 12 separate enzymes

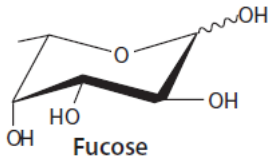
Essentially, each linkage in an oligosaccharide is formed by a specific transferase enzyme with specific expression control and subcellular localization.

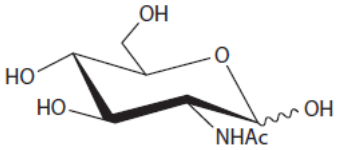
Consortium for Functional Glycomics (CFG) Notation

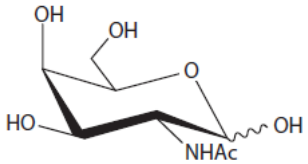


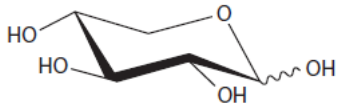


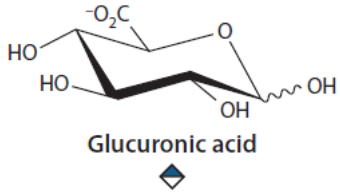


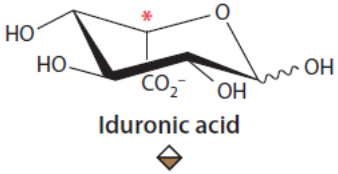


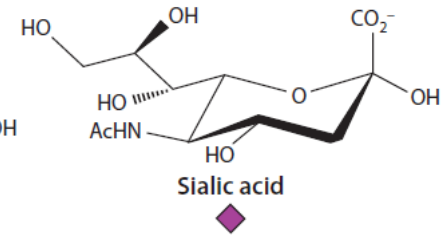


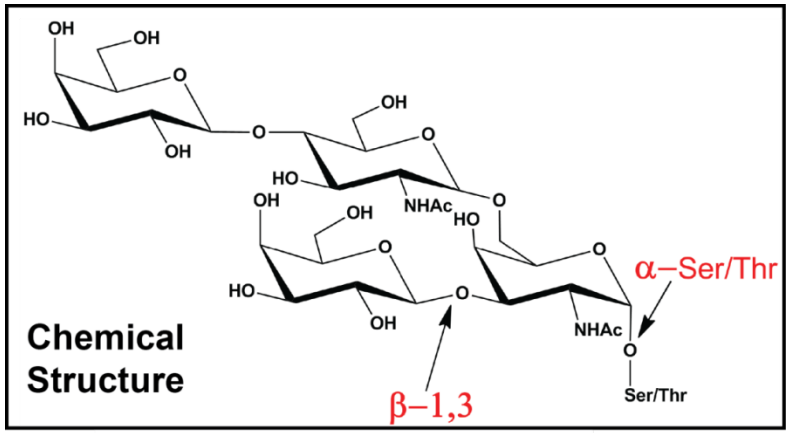


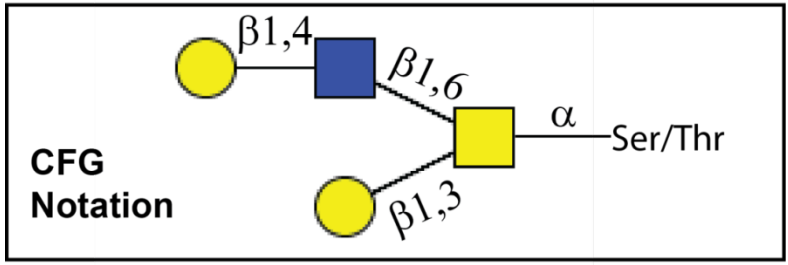








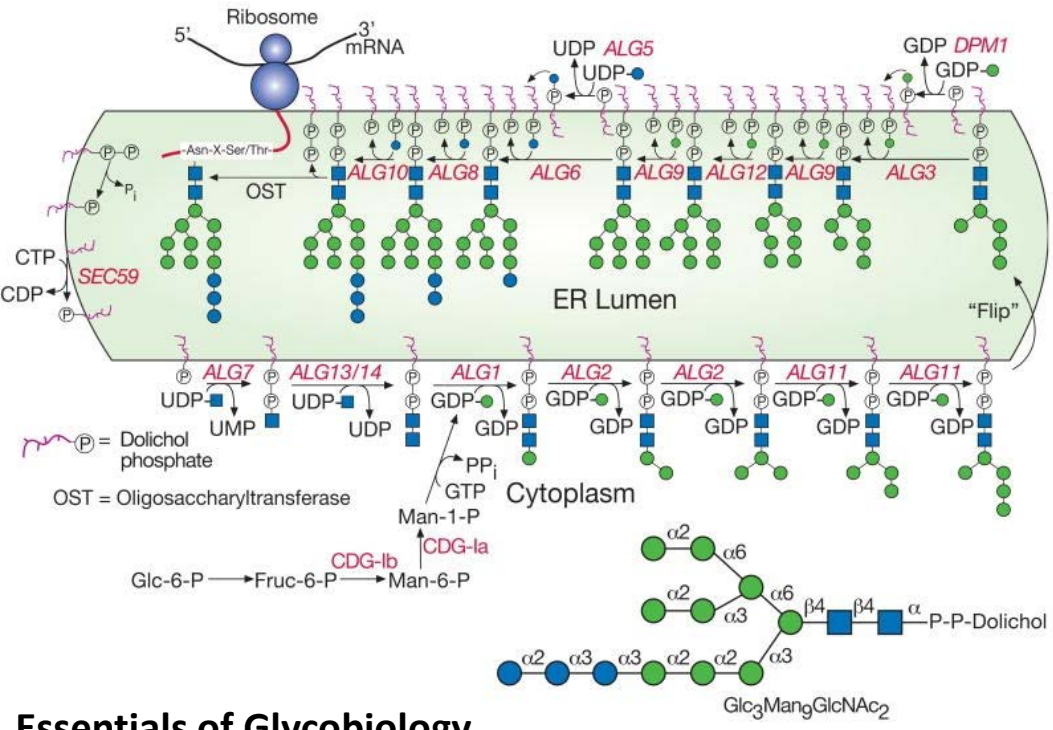




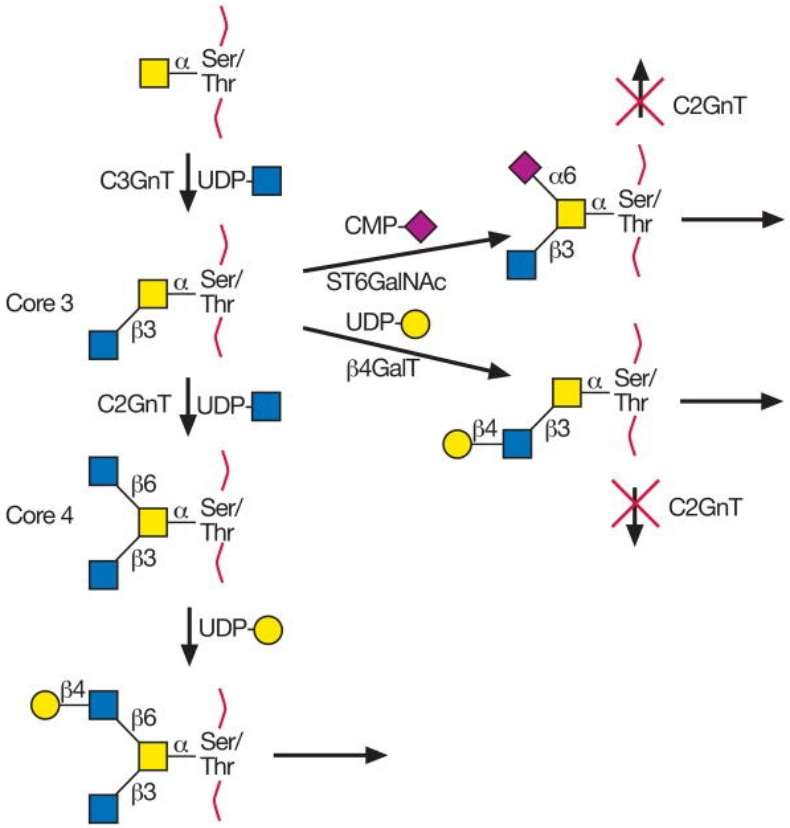
Rakus and Mahal, *Ann Rev Anal Chem*, 2011

There are two primary glycosylation pathways

***N*-linked glycosylation** occurs in the ER and Golgi and involves construction of a lipid-linked 14-mer precursor before being transferred to an Asn residue and further modified to form the final structure. Modified proteins have N-x-S/T consensus sequence



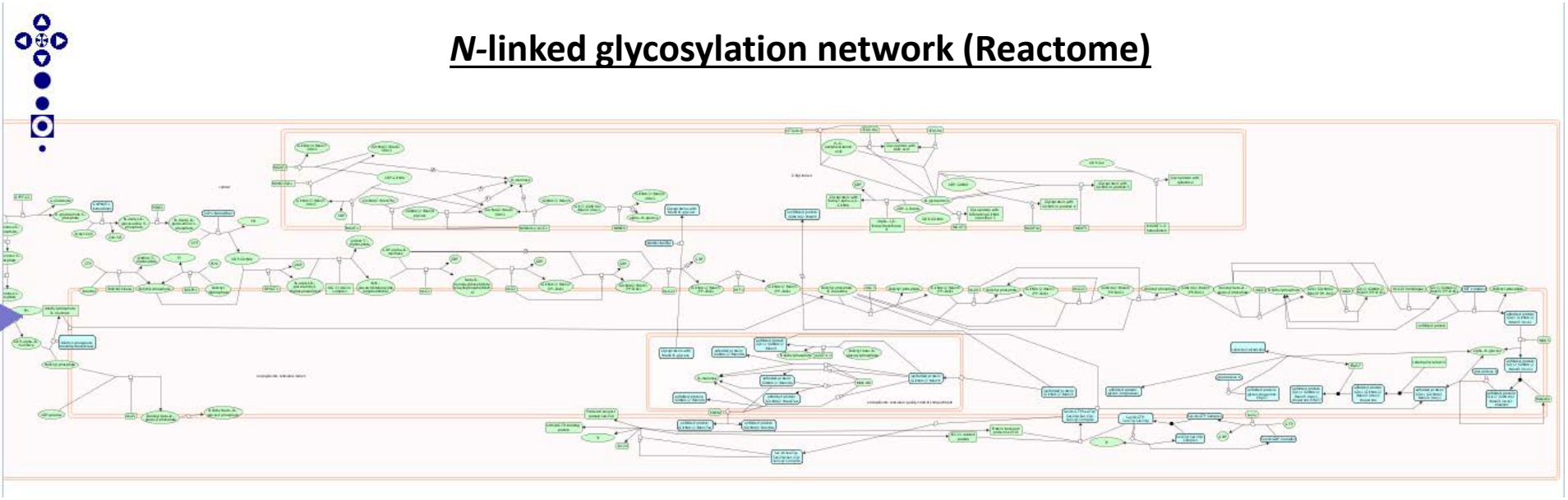
***O*-linked glycosylation** occurs in the Golgi apparatus and involves transfer of a monosaccharide directly to a Ser/Thr residue by a specific ppGalNacT followed by further elaboration. No known consensus sequence



Essentials of Glycobiology

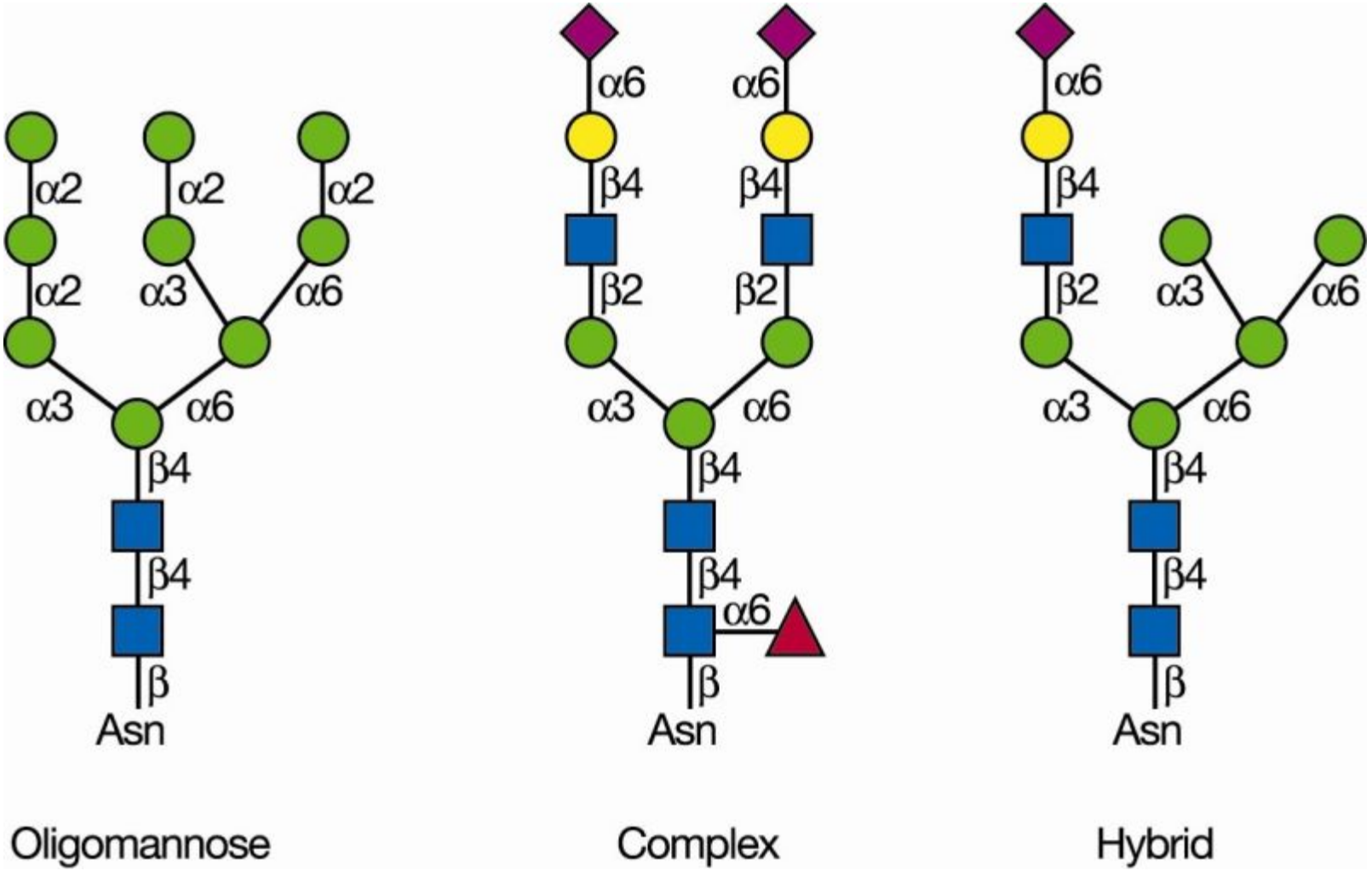
Carbohydrate synthetic regulation

***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)**

There are three types of *N*-linked glycans



Glycosylation is complicated

- **The structures are hard to discern**
- **The pathways are often redundant and overlapping**
- **The effects are subtle and, occasionally, conflicting**

Systems Biology

“Systems biology” is a strategy to look at complex phenomena and try to break it down into discernible pathways

The experiment

1. Look at the cell surface (glycome)

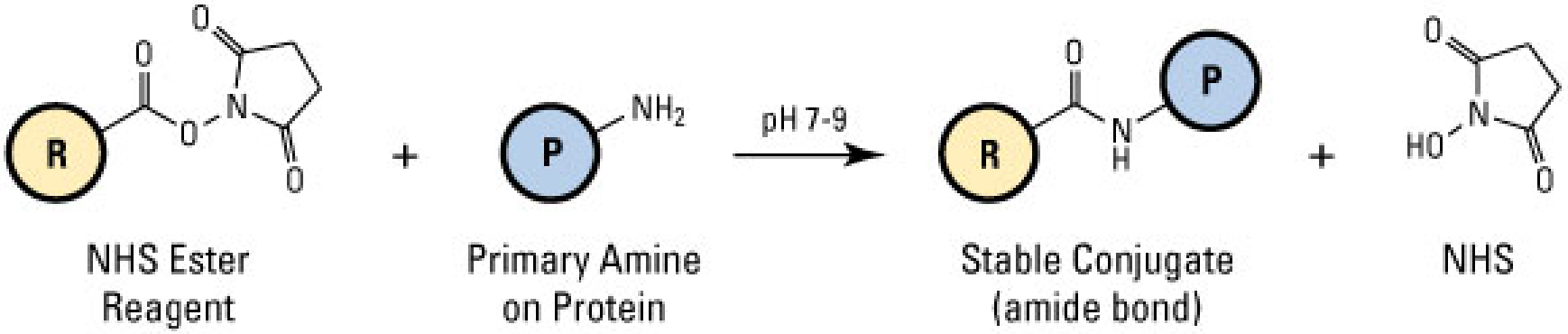
2. See if different cells express different carbohydrates

3. If so, try to explain why

Model System: The NCI-60 Cell Panel

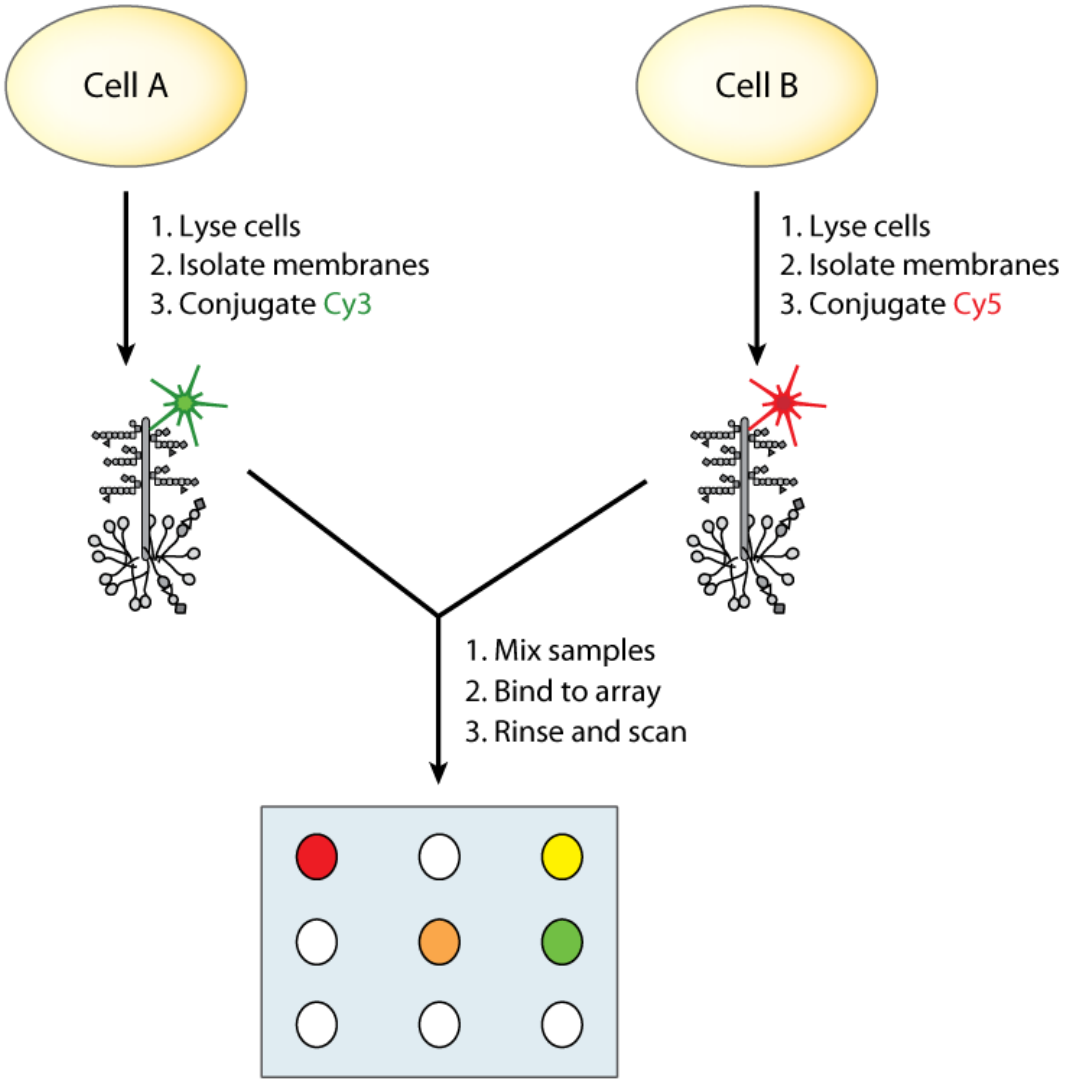
- **NCI-60: 60 cell lines for screening of potential cancer therapeutics**
- **Vary in tissue type, metastasis, individual of origin**
- **CellMiner.org: open source database containing mRNA, miRNA and protein array data, genetic mapping, pharmacological and mutational analysis**

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**

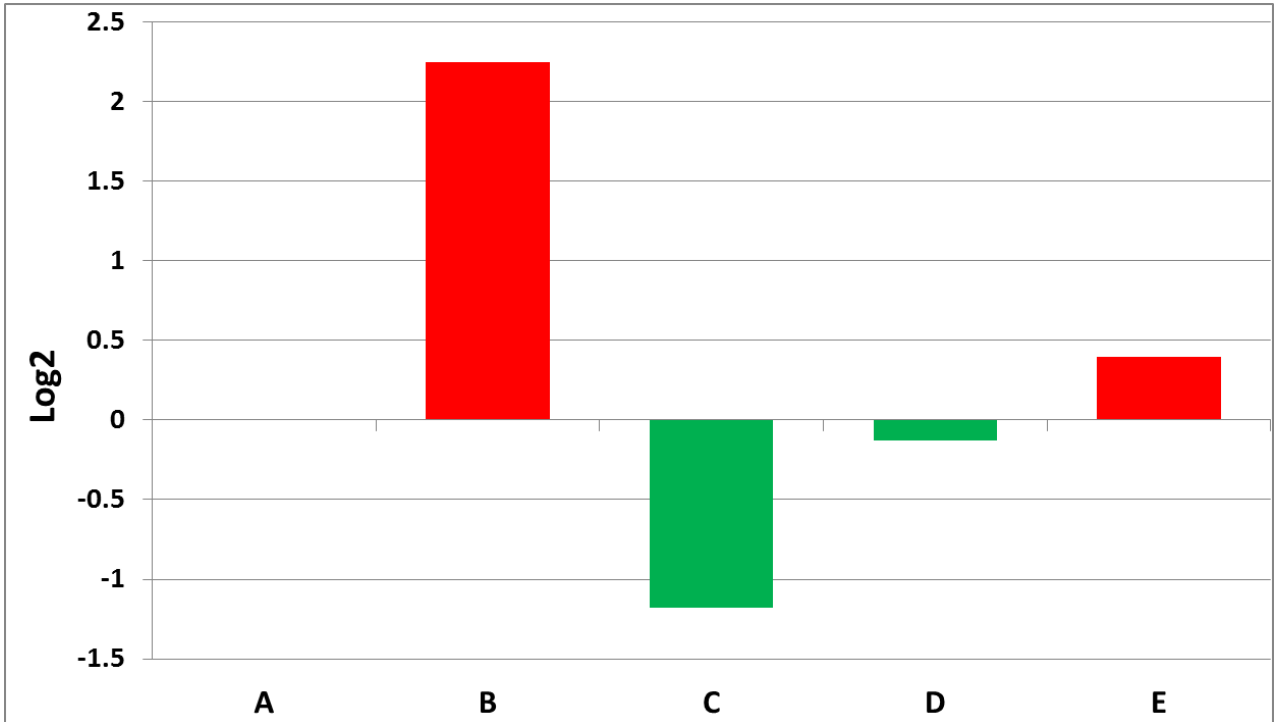
Ratiometric lectin microarray analysis for semi-quantitative analysis of the dynamic glycome



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

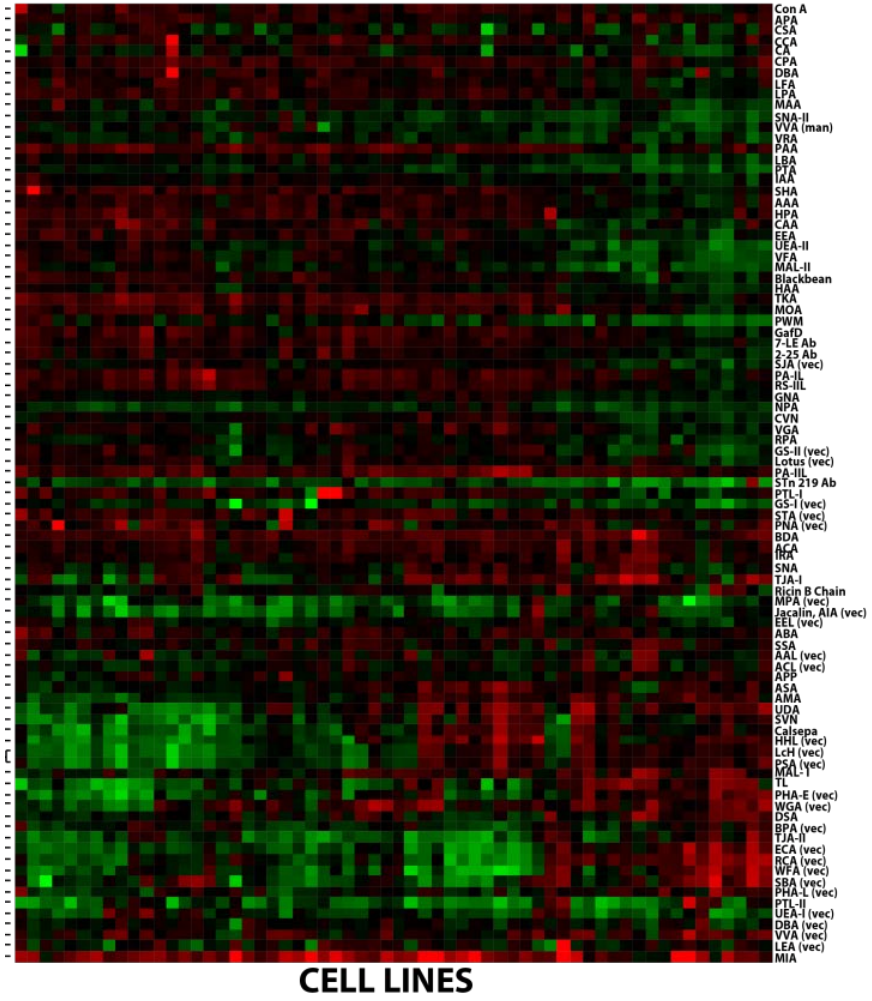
Interpreting a two-color experiment

Probe	Sample Fluorescence	Reference Fluorescence	Sample/Reference	$Log_2(S/R)$
A	500.00	500.00	1	0.00
B	2497.0	525.00	4.76	2.25
C	5500.0	12485	0.44	-1.18
D	125.20	137.23	0.91	-0.13
E	2545.0	1928.0	1.32	0.40

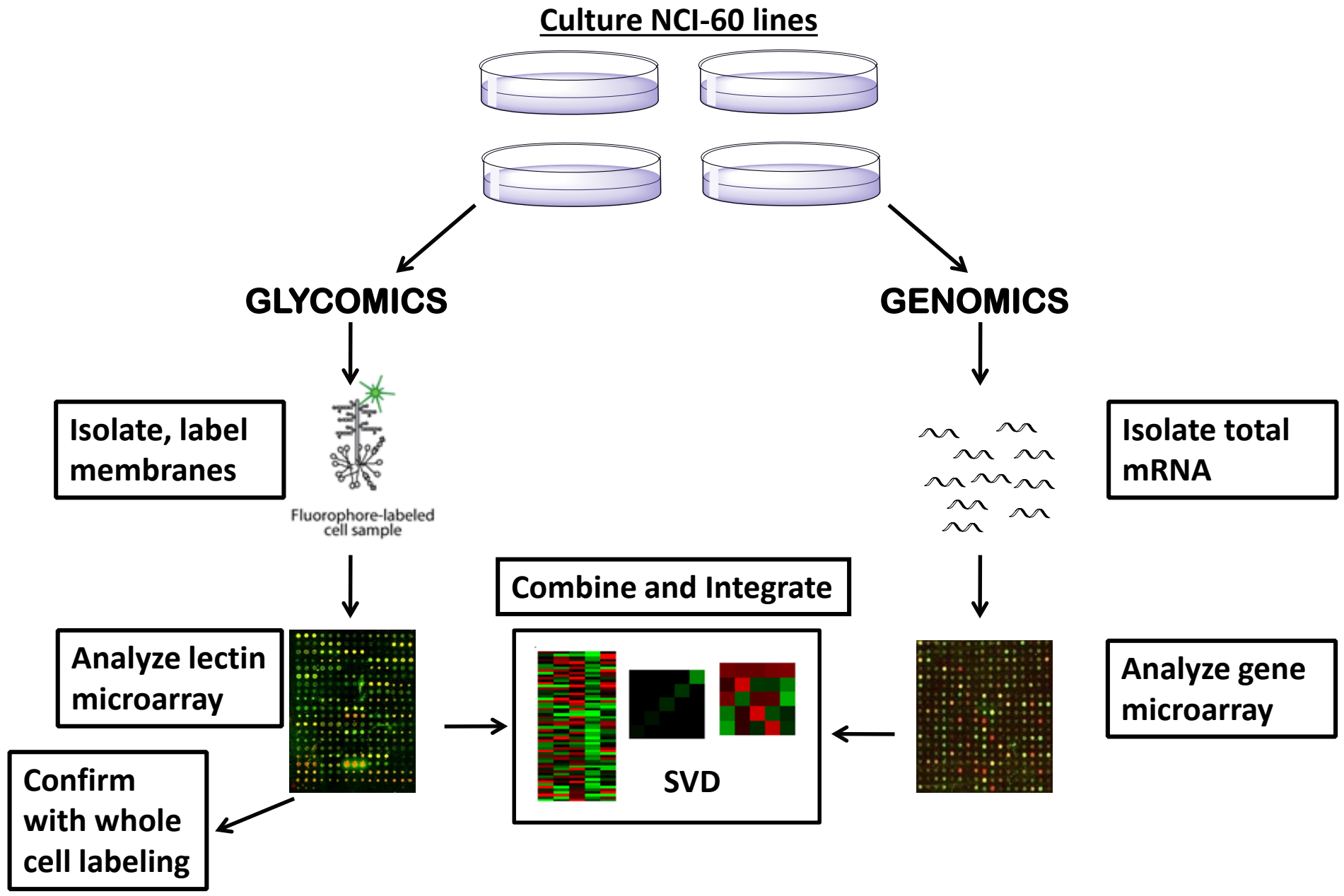


Interpreting a two-color experiment

When you have many samples and many probes, the data gets represented as a heat map

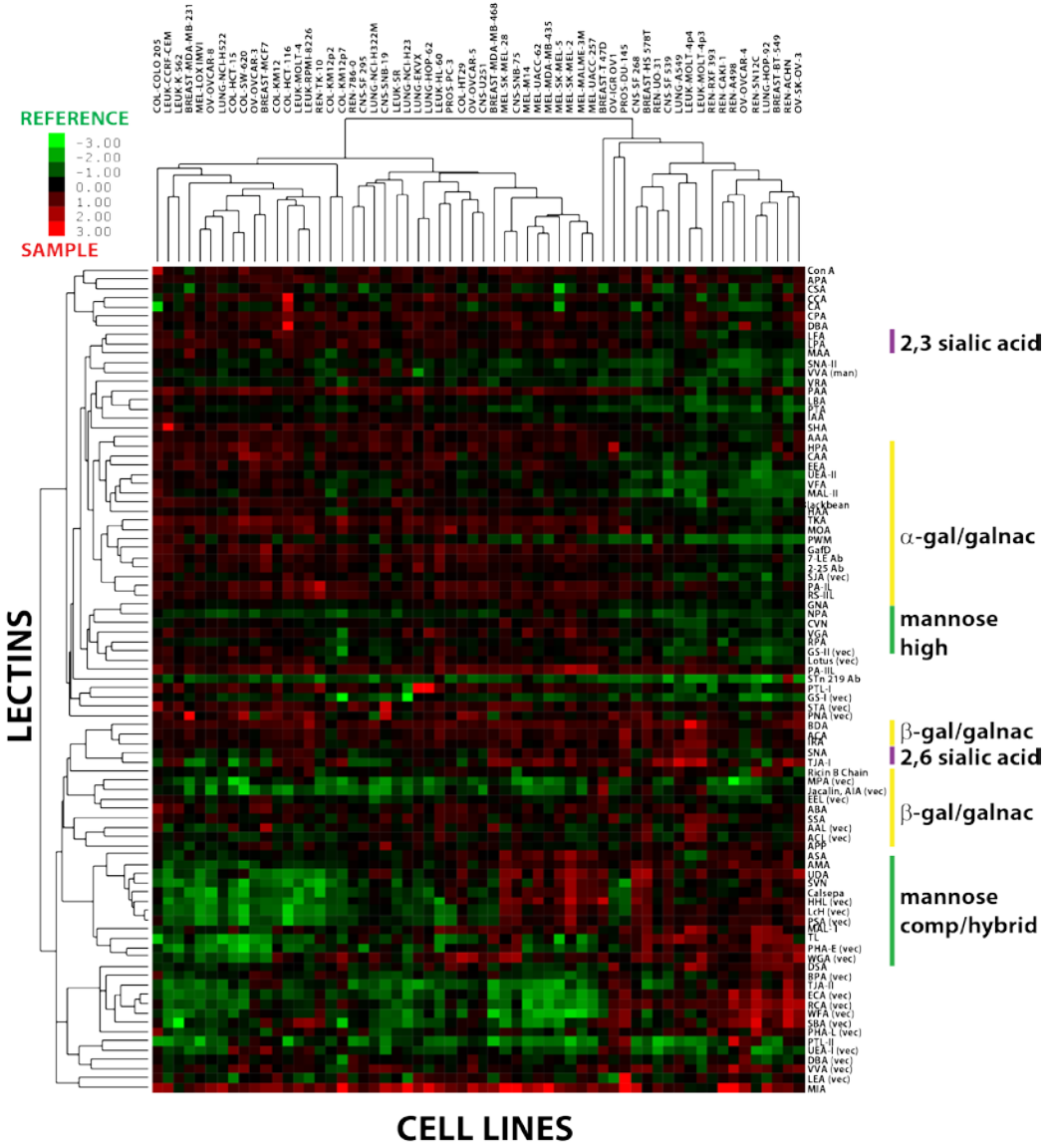


Experimental Strategy



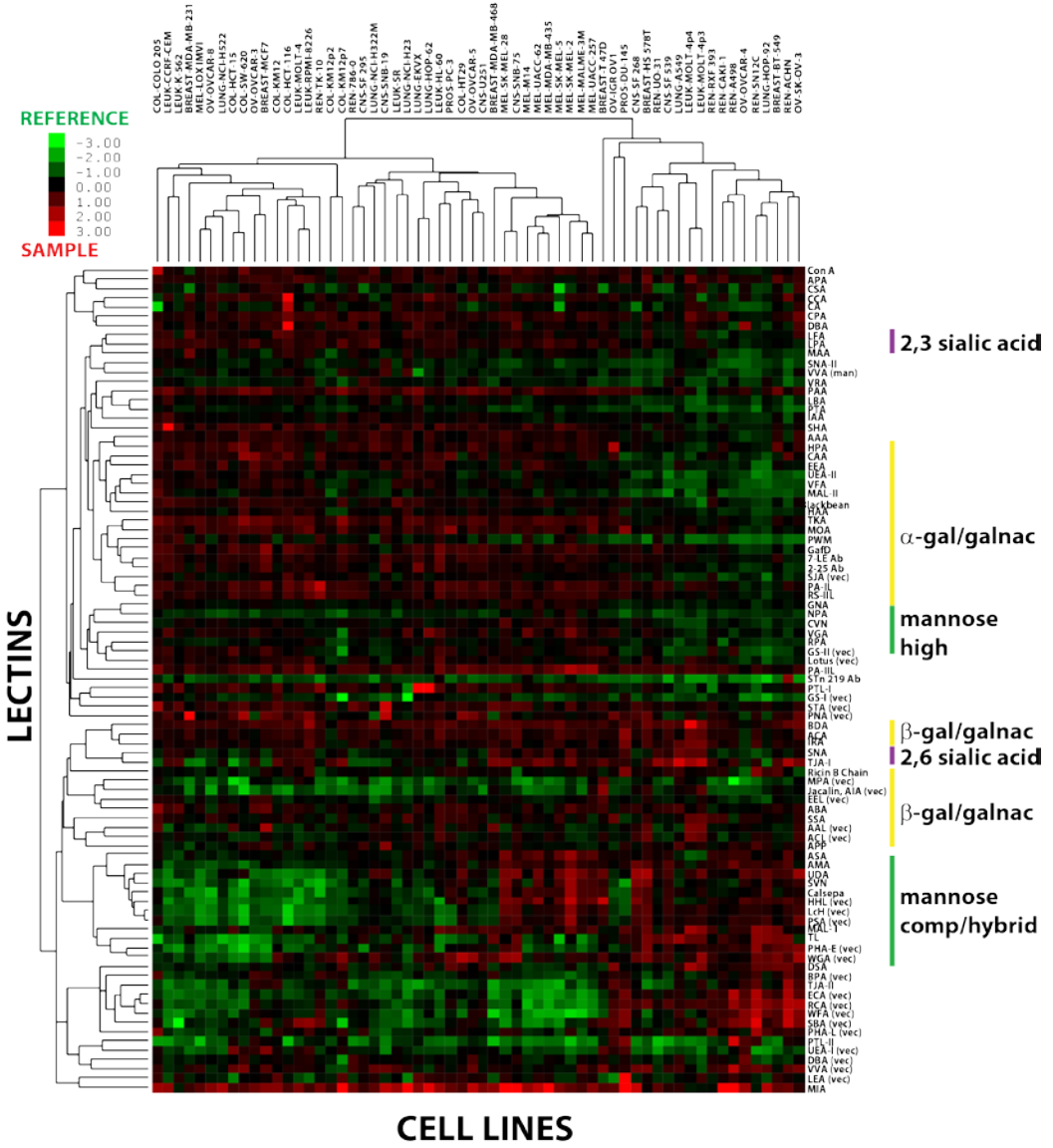
Lectin microarray analysis shows meaningful patterns

- 90 lectins
- 56 cell lines
- Biological replicates of two cell lines

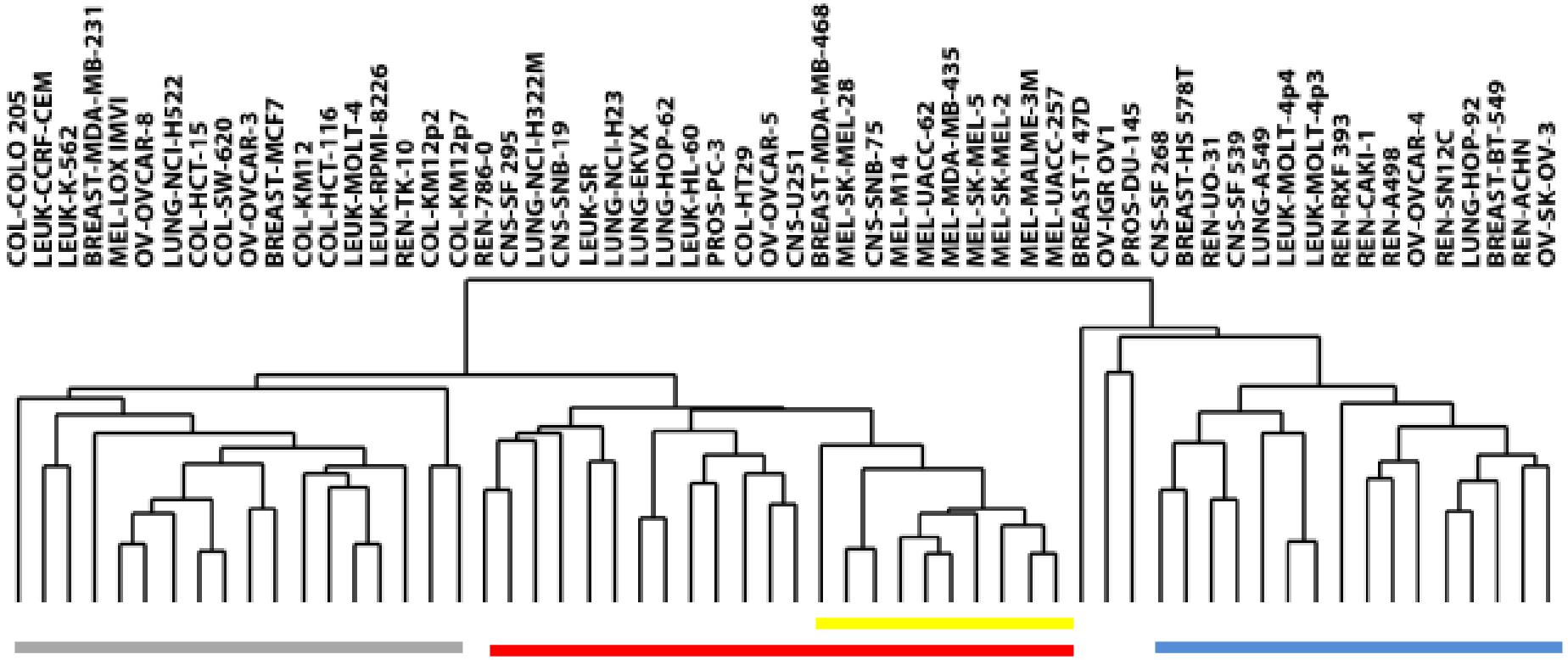


Lectin microarray analysis shows meaningful patterns

- Several clusters of lectins by carbohydrate specificity
- Cell lines cluster by tissue of origin



Cell lines cluster based on their glycosylation



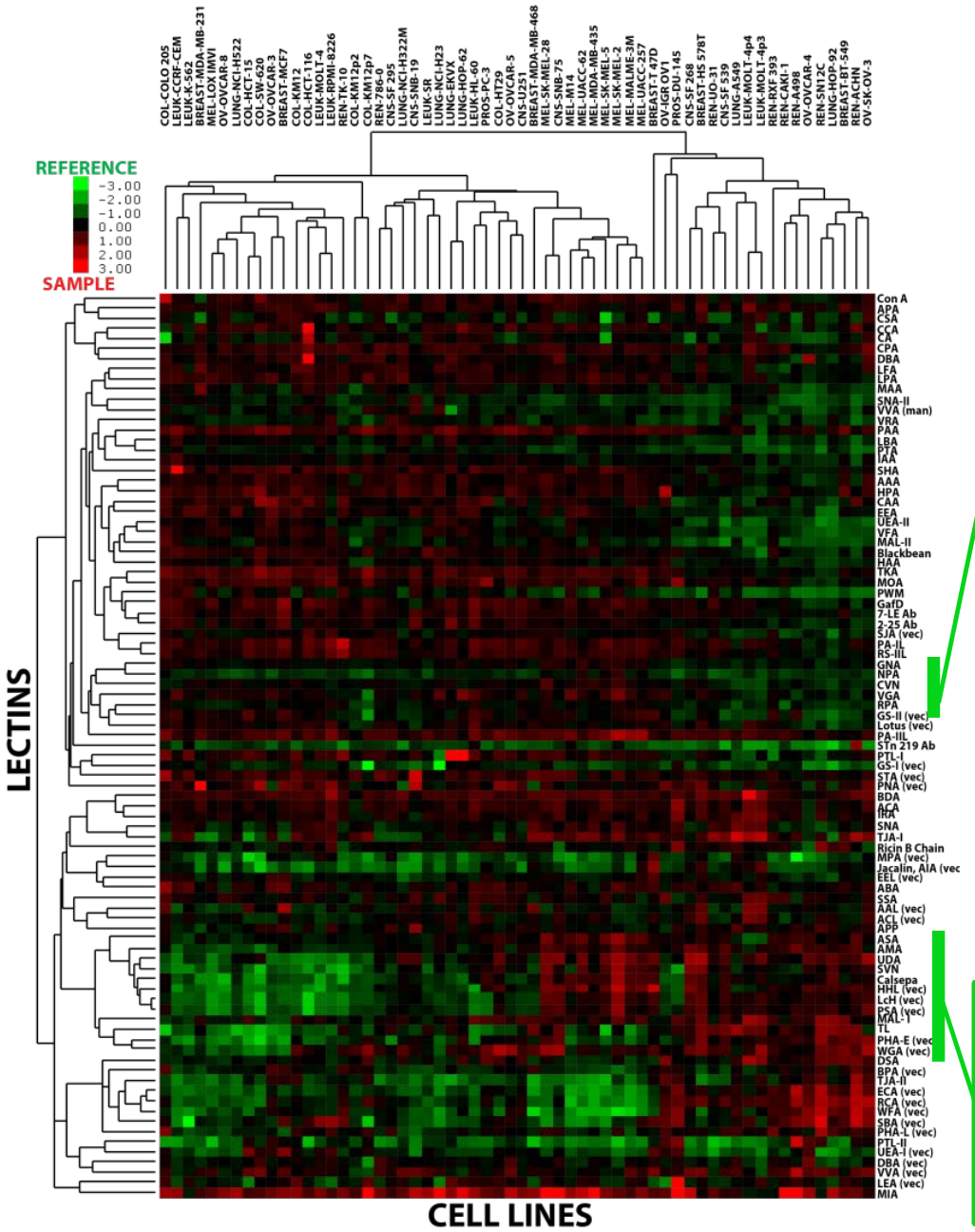
COLON (5/6, p=0.39)
LEUKEMIA (4/6, p=0.39)

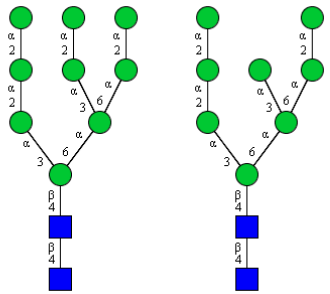
CNS (4/5, p=0.44)
LUNG (4/7, p=0.44)

MELANOMA (8/9, p=0.60)

RENAL (6/8, p=0.39)

Glycosylation signatures: Mannose

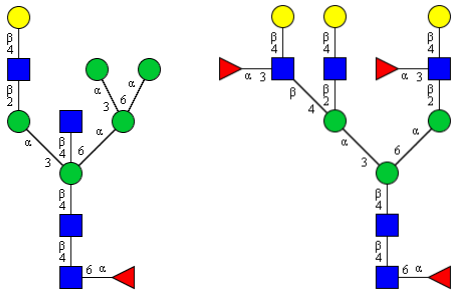




RENAL

MELANOMA
LUNG,COLON

Mannose (high)
GNA, NPA, CVN, RPA

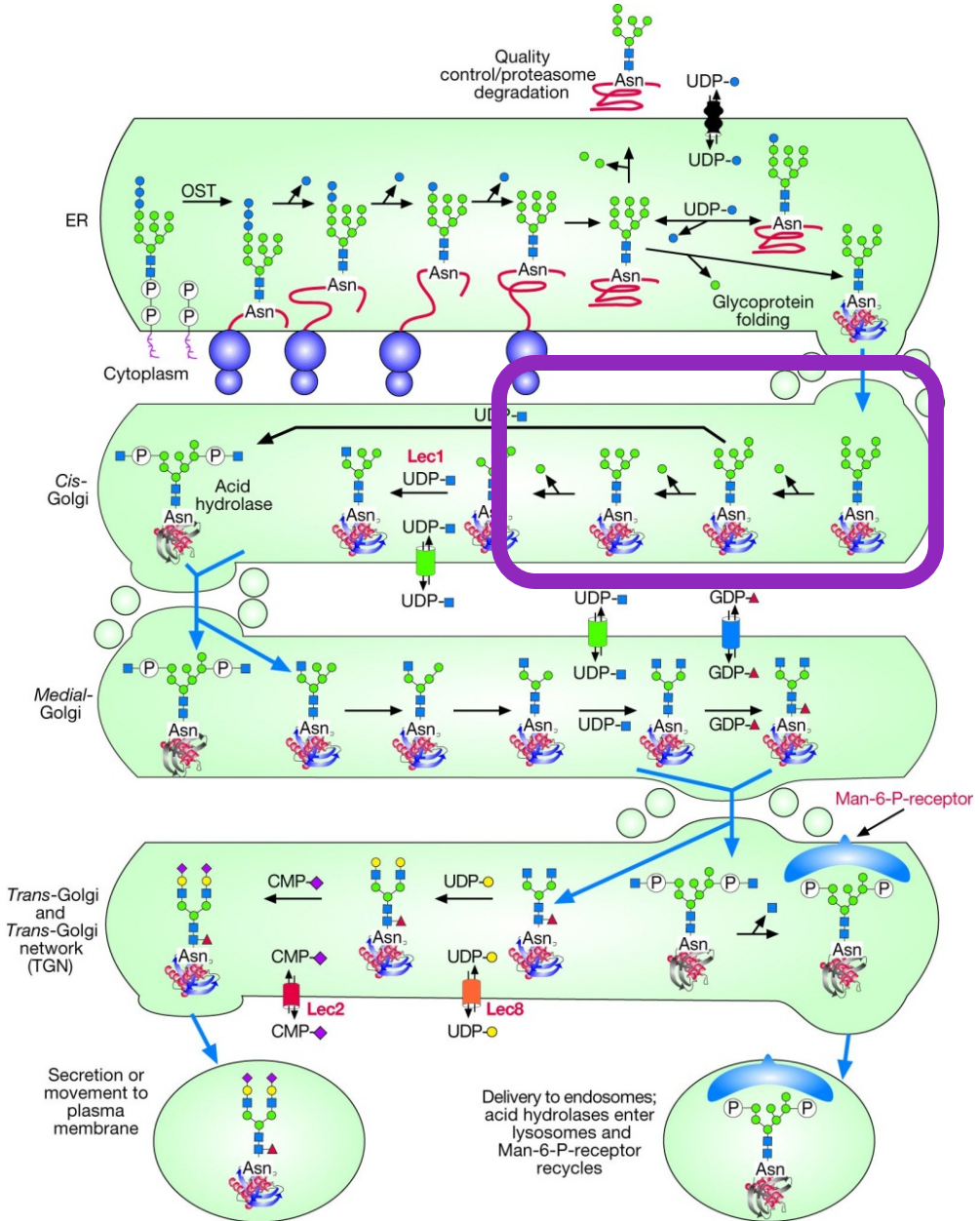


Mannose (complex)
UDA, SVN, Calsepa, HHL, LcH, PHA-E, MAL-I, PSA

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
- <u>α-mannosidase I</u> controls all hybrid/complex maturation steps

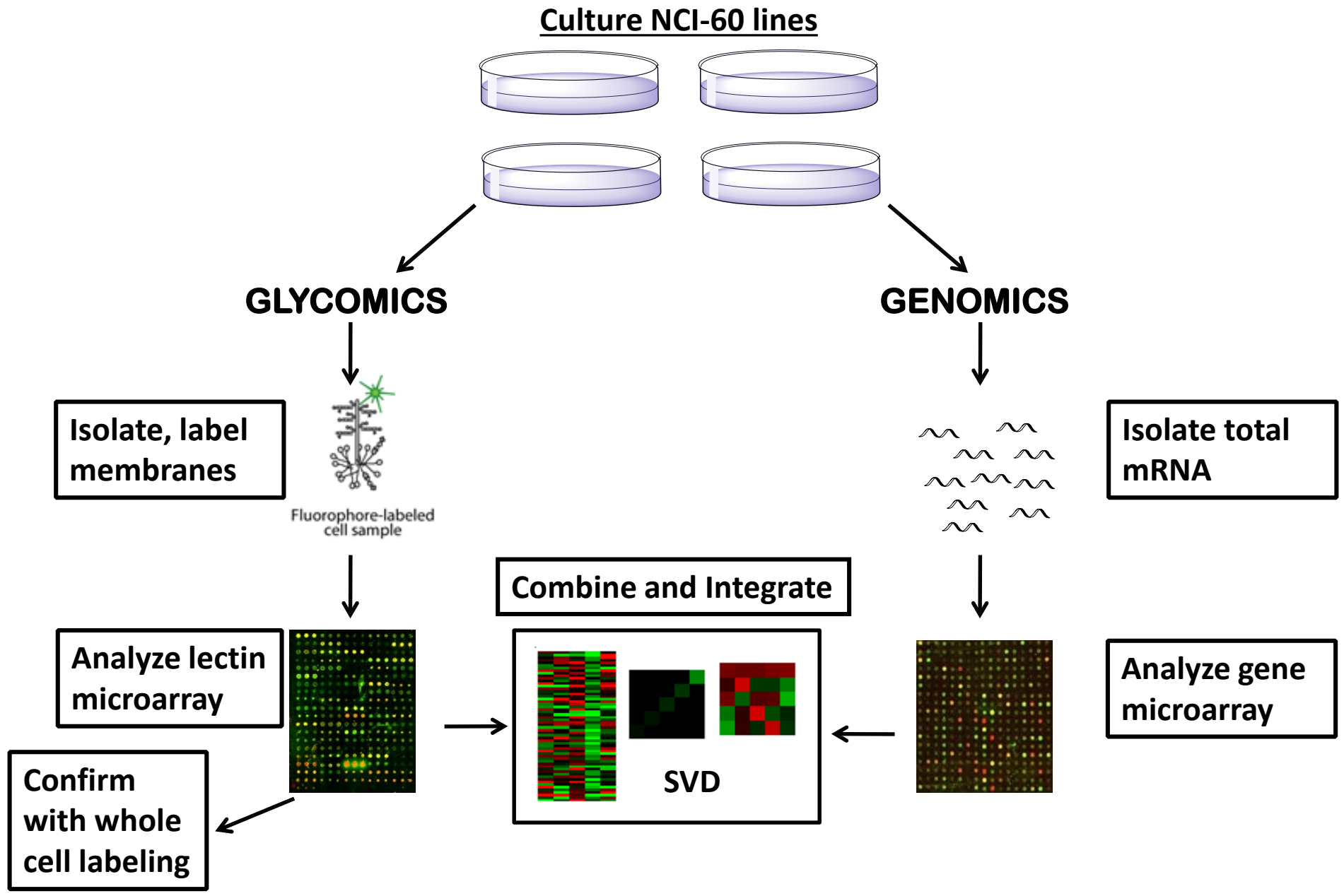
Essentials of Glycobiology

Conclusions (some)!

- **Human cells express different, specific carbohydrate patterns on their surfaces, based on their histological function**

- **What controls the glycome variation? (what does the interior of the house look like?)**

Experimental Strategy



Patterns of NCI-60 Gene Expression Regulation

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

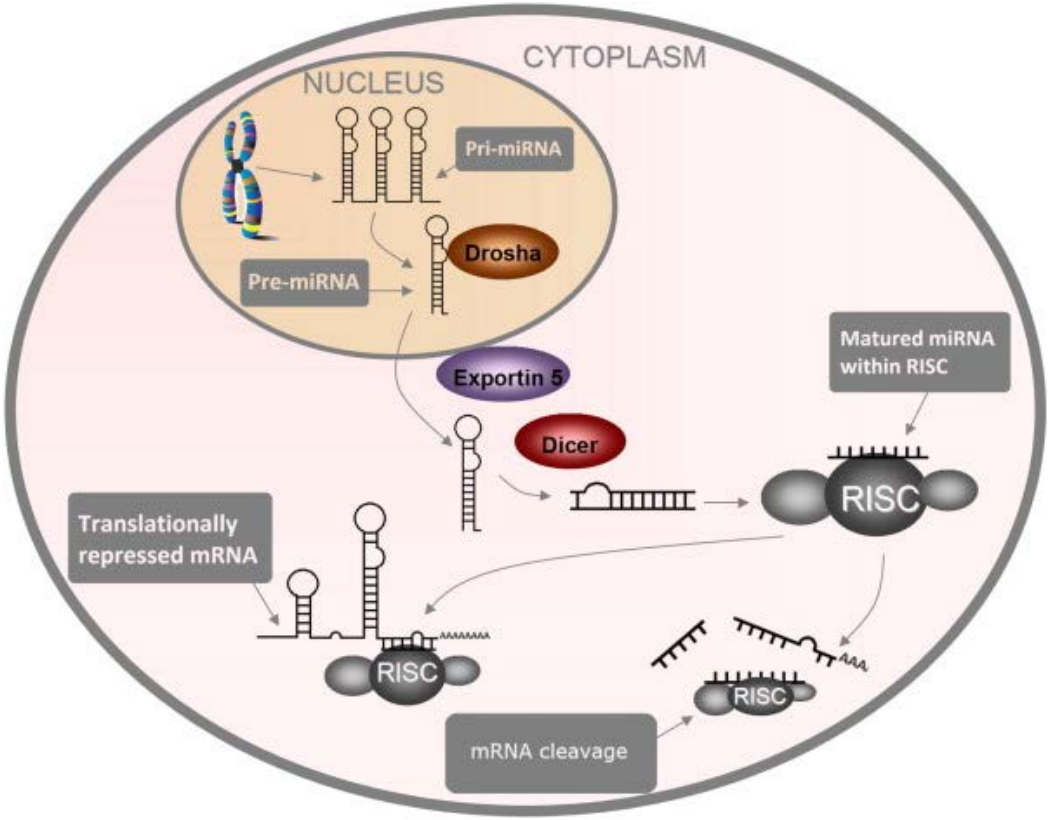
OBSERVATION OF HIGHLY EXPRESSED AND DIVERSE PROBES

-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 (miRNAs) regulate gene expression



- **miRNA are genomically encoded short (~22 nt) molecules involved in repressing expression of mRNA**
- **Bind target miRNAs in a 6-10 bp seed region**
- **Upon processing, miRNAs recruit a post-transcriptional silencing complex to inhibit expression**

Thamilarasan et al, *Autoimmun Rev*, 2012

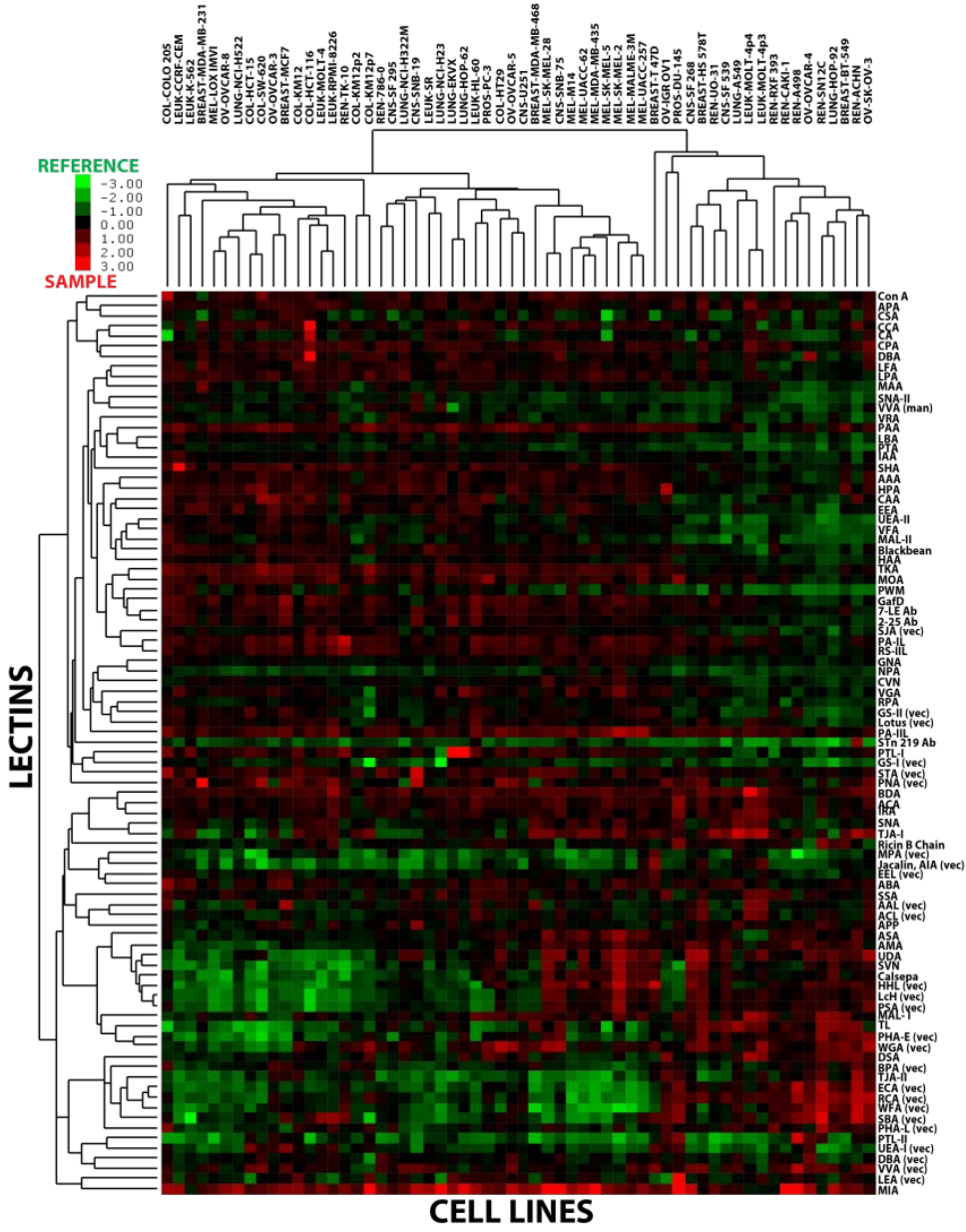
How we make a cluster

$$r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\sigma_x}\right)\left(\frac{y_i - \overline{y}}{\sigma_y}\right)$$

$\overline{x}$ = average across all set of numbers

σ = standard deviation against all set of numbers

- **This is a fancy way of saying I am working with vectors**
- **Probes vs cell line (horizontal) and cell line vs probe (vertical)**
- **The intensity for each point on a heat map has a distance from every other data point**



Singular Value Decomposition

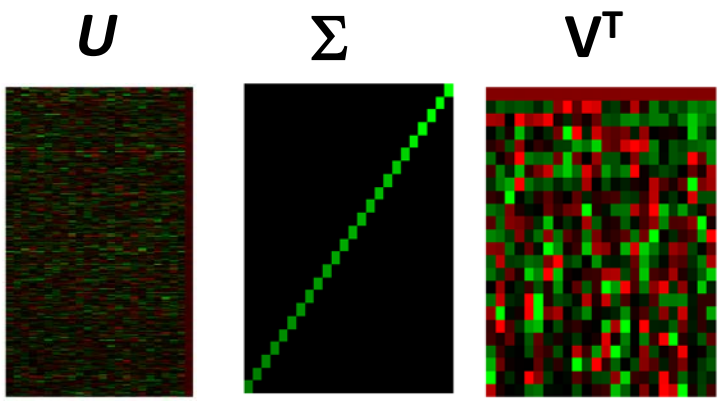
- **Method to decompose a matrix into an orthogonally transformed set of variables**
- **Used to identify significant, co-varying patterns in large data sets. I.E. What signals account for the variation?**
- **Can be applied to any array individually or combined for SYSTEMS BIOLOGY examination of glycosylation pathways**

SVD Matrix

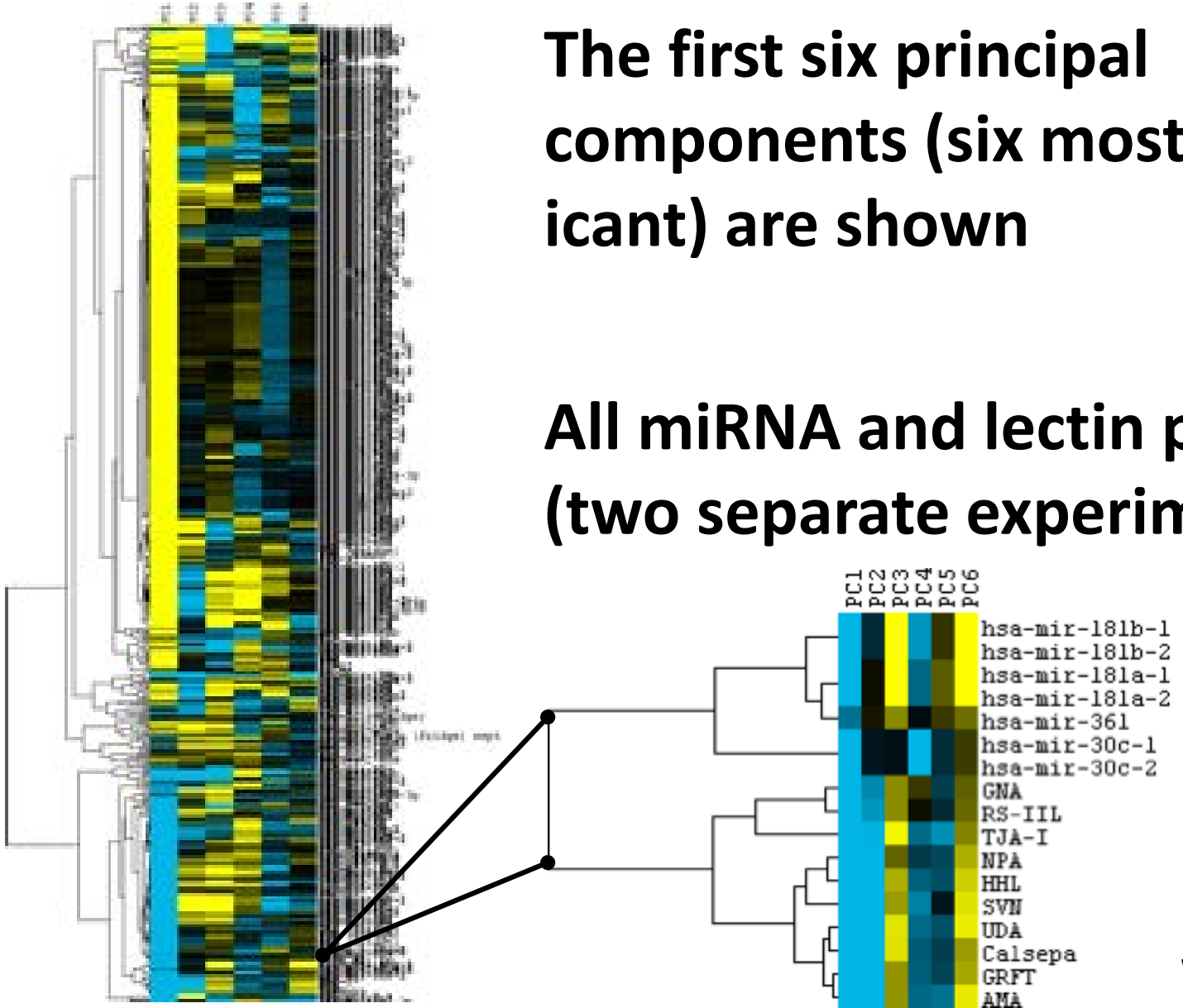
$$M = U \Sigma V^T$$

Eigenarrays vs genes/lectins

Arrays vs eigengenes/eigenlectins



Lectin plus miRNA SVD Results



The first six principal components (six most signif-icant) are shown

All miRNA and lectin probes (two separate experiments)

With Lani Pilobello

We now have a hypothesis

- **miRNA's regulate the *N*-linked maturation pathway**
- **miRNA expression controls *N*-linked glycome by repressing (or activating) α-mannosidase I**
- **High mannose cells have high miRNA expression which means low α-mannosidase I**

We need evidence

Use high mannose (SN12C, SK-MEL-5) and complex (HCT-116, HT29) lines

Modulate miRNA expression in cells

Analyze mannosidase gene expression (RT-PCR)

Isolate total RNA

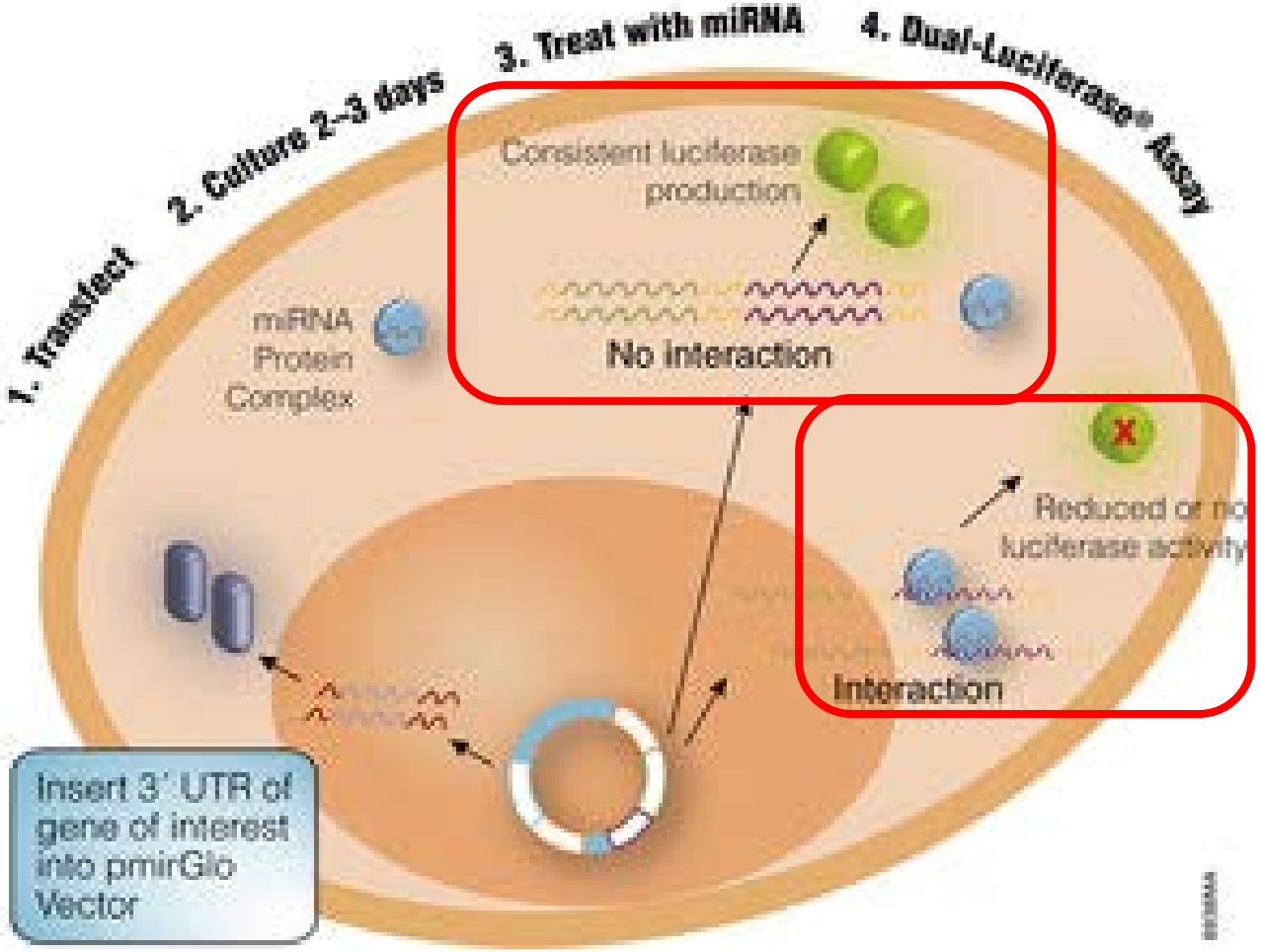
Isolate total protein

Analyze mannosidase protein expression (Western blot)

Analyze glycome (lectin microarray)

Can the miRNAs bind the mannosidase transcript?

Luciferase read-through assay

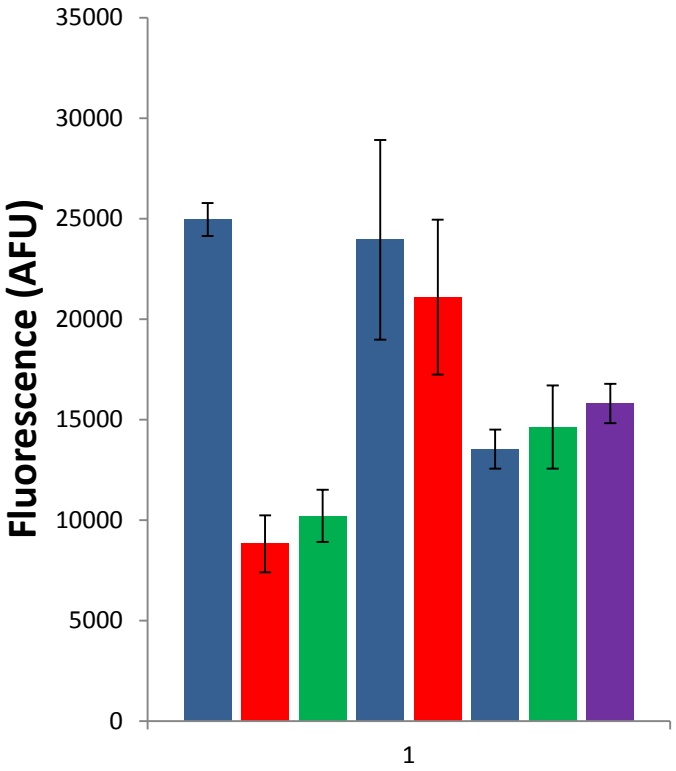


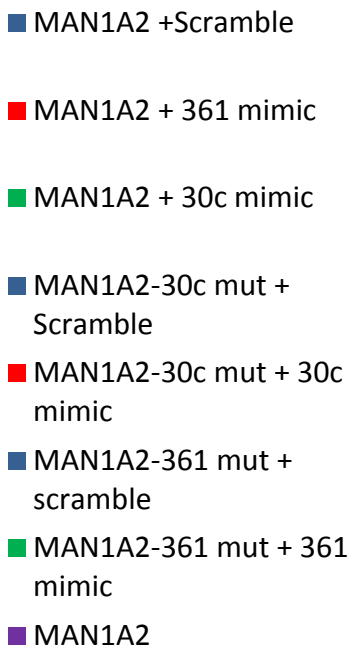
Promega.com

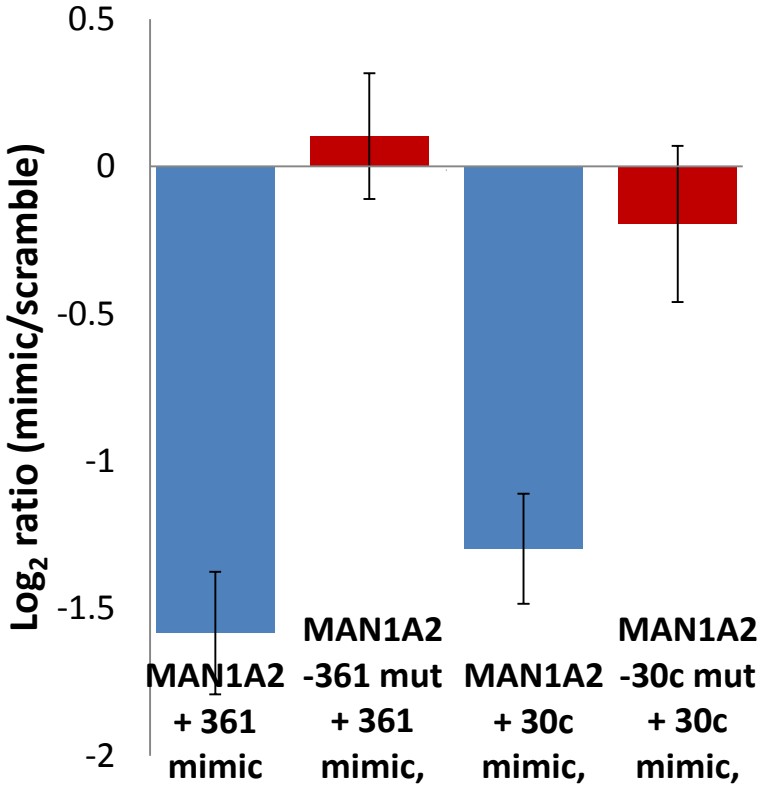
miR-30c and -361 bind the MAN1A2 3’ UTR

```
MAN1A2.30cMut.stk<      CCTGTGTTTCGCTCATATGGACCACTACAGAAATTAGTTTGAAGGGGCGGCTTTTGAAAA
MAN1A2.3UTR<            CCTGTGTTTTGTTTACATGGACCACTACAGAAATTAGTTTGAAGGGGCGGCTTTTGAAAA
                        ********* * * * ****************************************

MAN1A2.361Mut.stk<      AAATTTCTTGGTTAAGTTTAATTTTCTATAGAGACATGATCTCACAAAGAATACTCAAGT
MAN1A2.3UTR<            AAATTTCTTGGTTAAGTTTAATTTTCTCTGGGGGCATGATCTCACAAAGAATACTCAAGT
                        ***************************.*.*.***************************
```



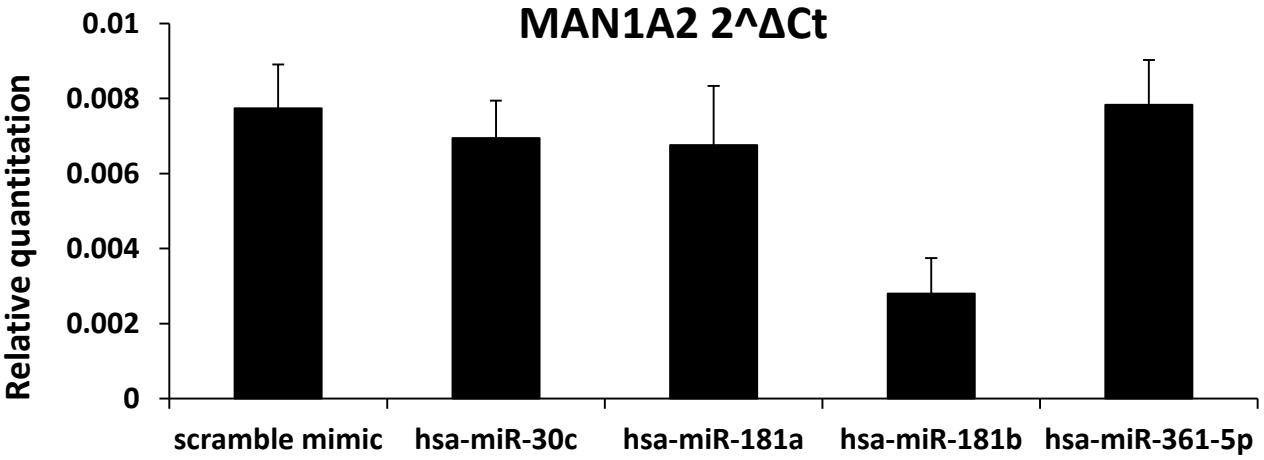




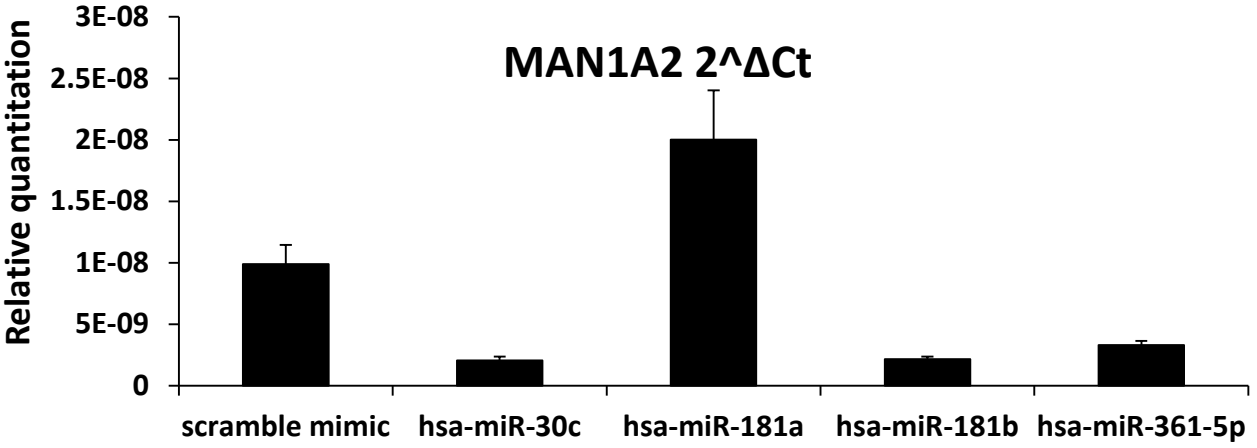
With Praveen Agrawal

Do the miRNAs affect mannosidase transcription?

Transfection of cell lines with miRNA repress MAN1A2



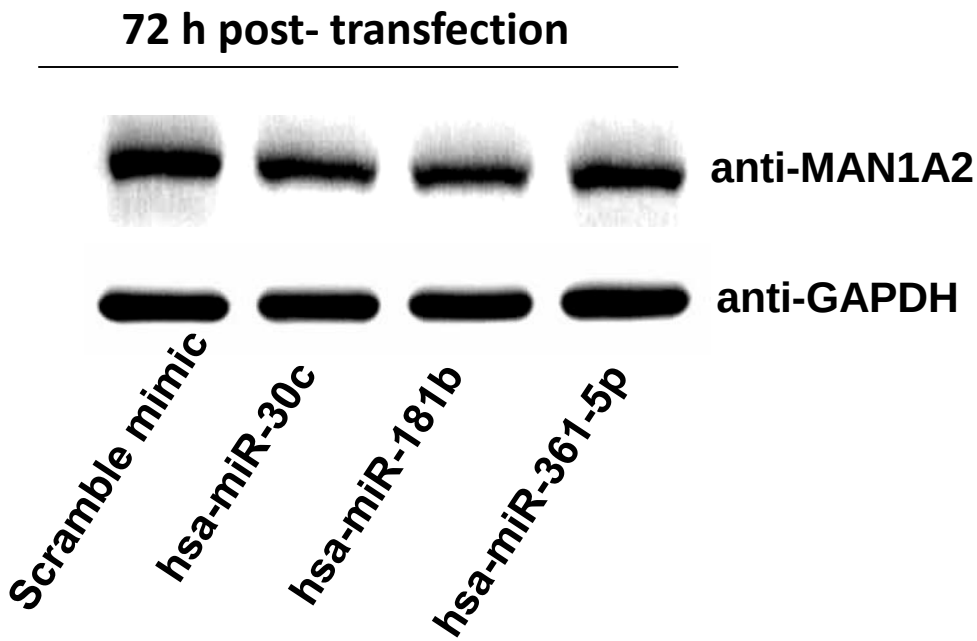
24 h post- transfection



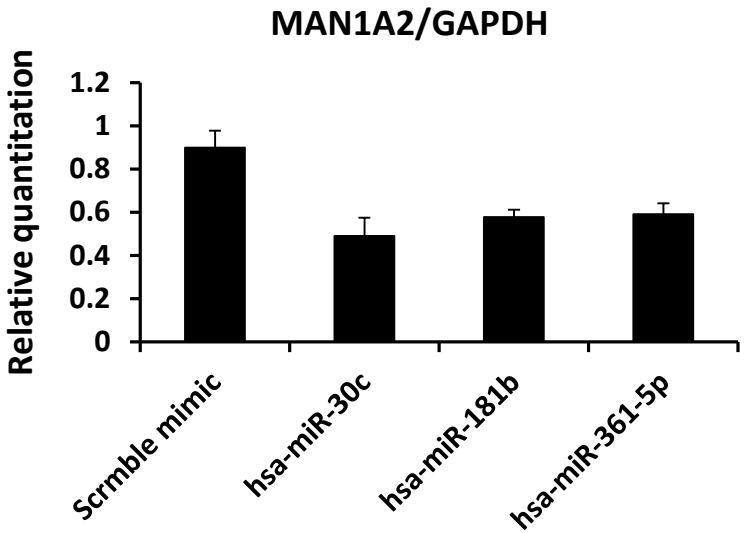
72 h post- transfection

With Praveen Agrawal

Do the miRNAs affect mannosidase protein expression?

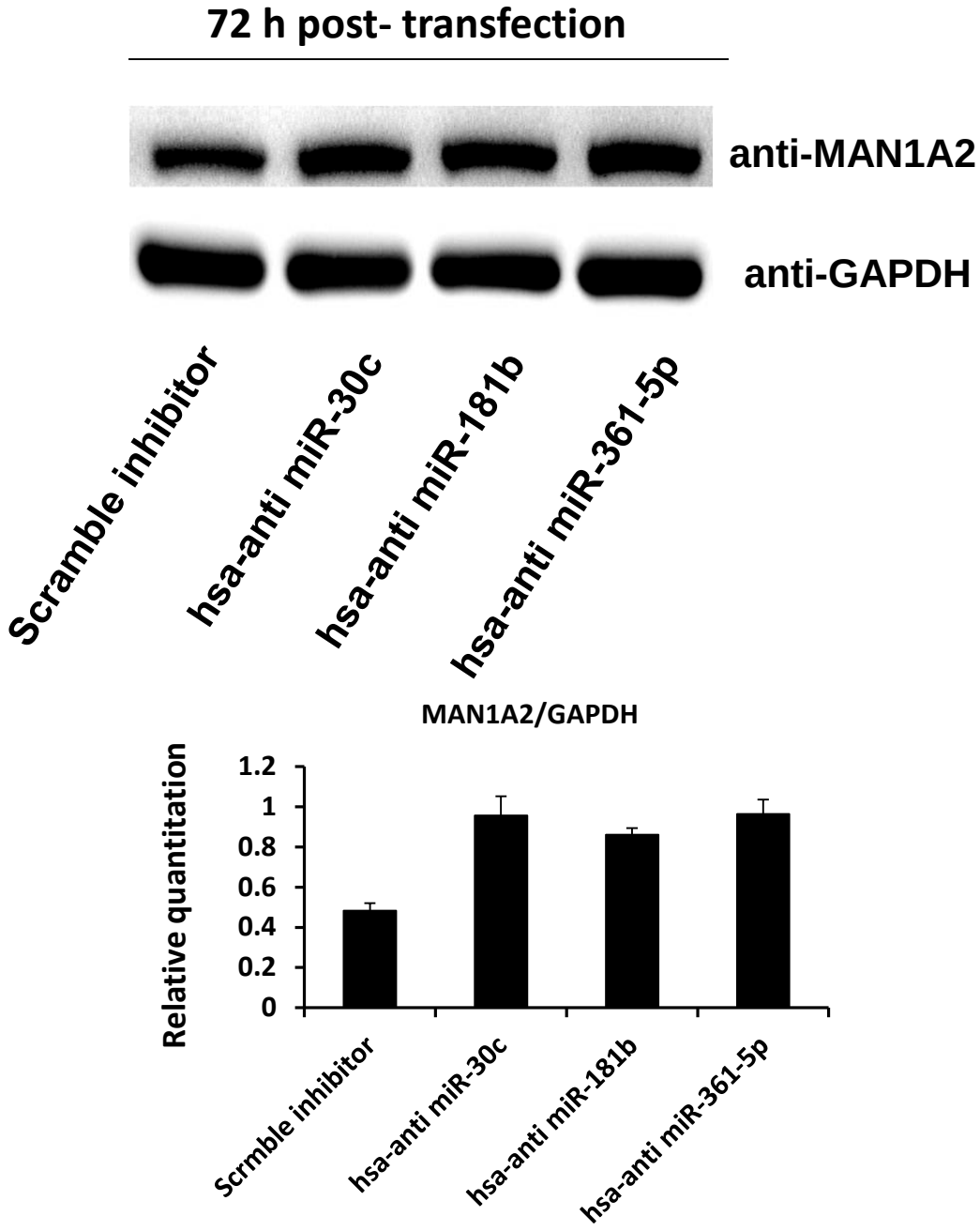


Western Blot of MAN1A2 72 hrs after transfection with miRNA



MAN1A2 expression decreases ~50%

With Praveen Agrawal



Western Blot of MAN1A2 72 hrs after transfection with miRNA inhibitor

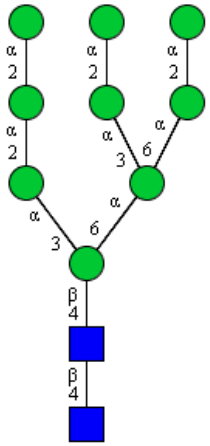
MAN1A2 expression increases ~50%

With Praveen Agrawal

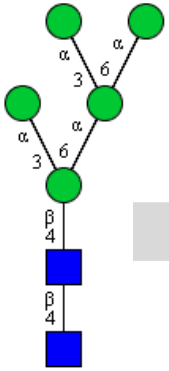
Seeing inside the house

- **miRNA-30c and -361 affect expression of the enzyme α-mannosidase I (maybe -181b as well)**
- **Expression of α-mannosidase I determines the type of *N*-linked glycan expresses**
- **Cells differentially express miRNAs resulting in different glycomes**

A model for cell-type specific carbohydrate expression?



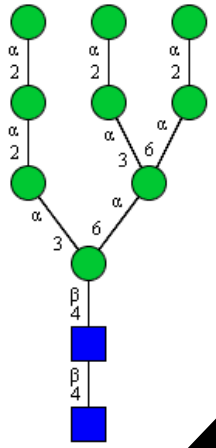
α-mannosidase I



~~miR-30c~~
~~miR-361~~

Expression of hybrid/complex glycans (renal)

A model for cell-type specific carbohydrate expression?



α-mannosidase I

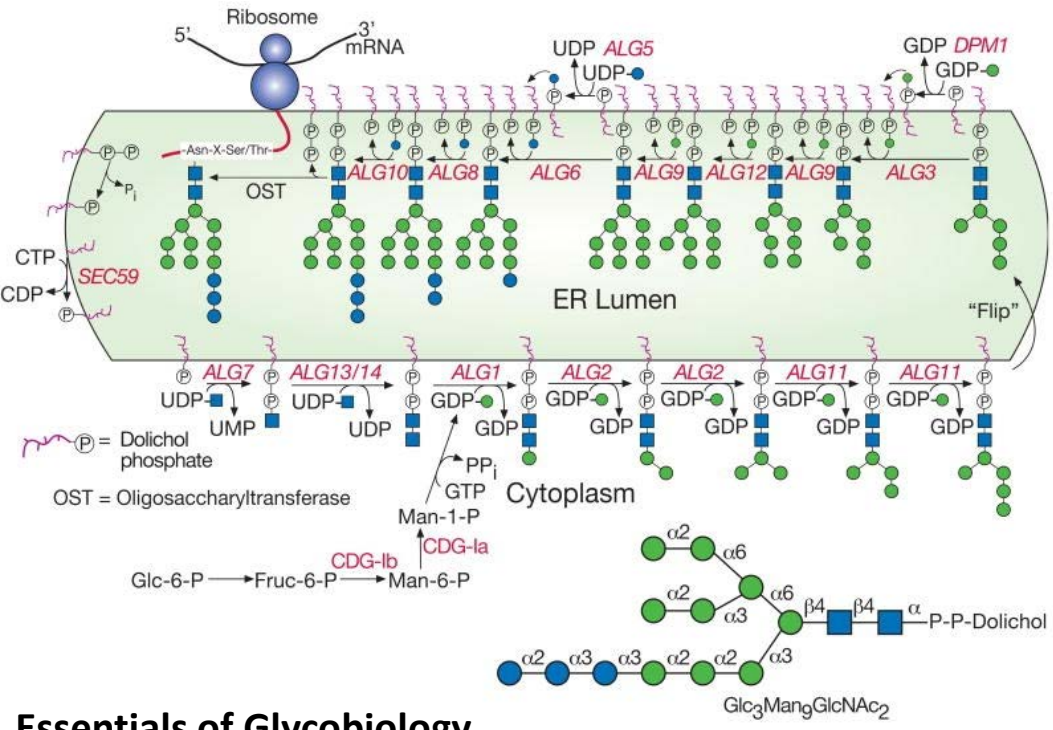
miR-30c
miR-361

Expression of high mannose (colon)

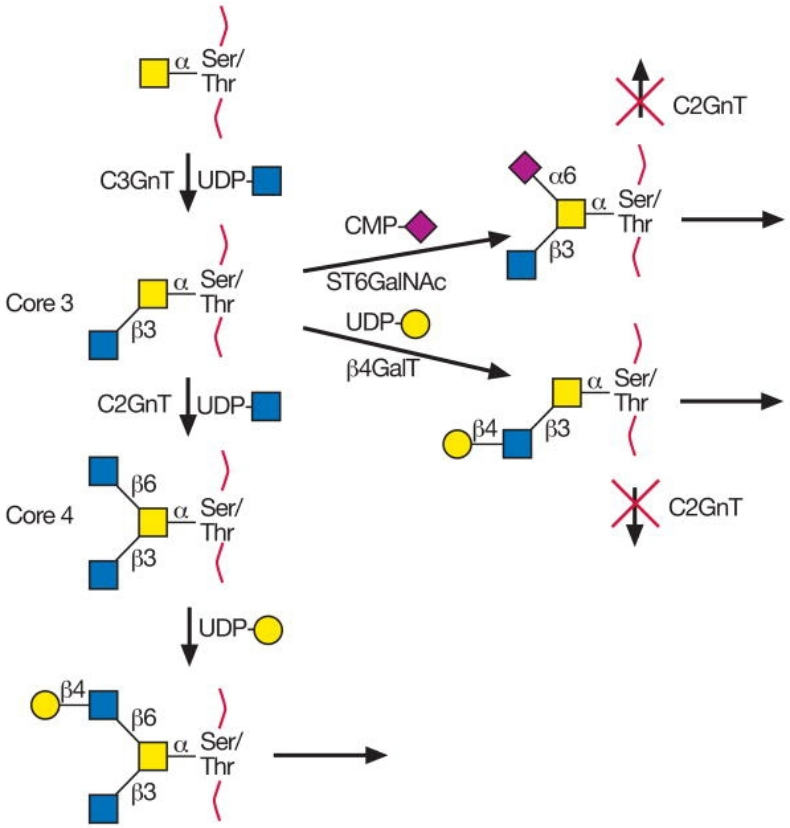
Expression of hybrid/complex glycans (renal)

There are two primary glycosylation pathways

***N*-linked glycosylation** occurs in the ER and Golgi and involves construction of a lipid-linked 14-mer precursor before being transferred to an Asn residue and further modified to form the final structure. Modified proteins have N-x-S/T consensus sequence

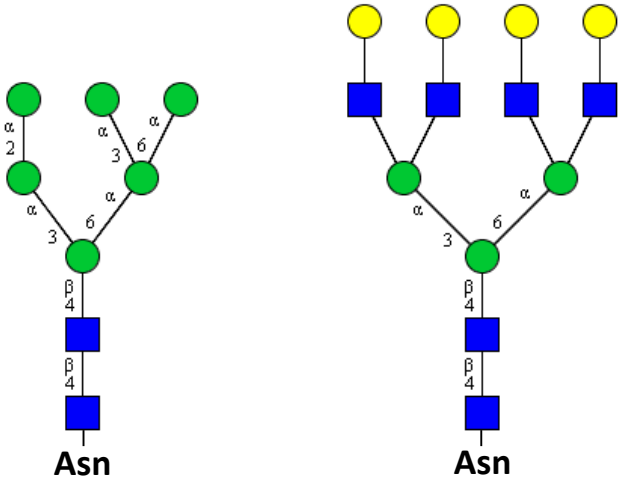


***O*-linked glycosylation** occurs in the Golgi apparatus and involves transfer of a monosaccharide directly to a Ser/Thr residue by a specific ppGalNacT followed by further elaboration. No known consensus sequence



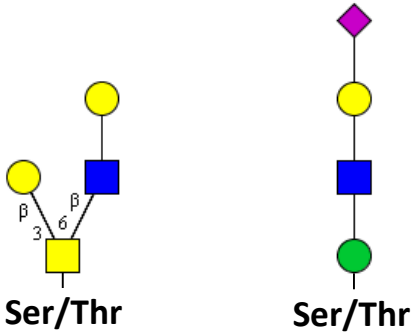
Essentials of Glycobiology

N-linked glycosylation



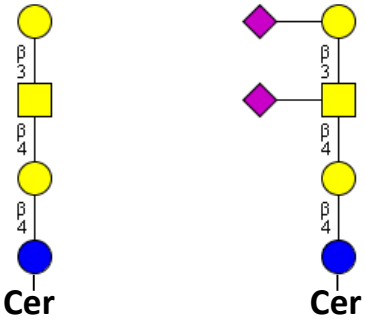
Functions: protein folding and trafficking; signaling ligand

O-linked glycosylation



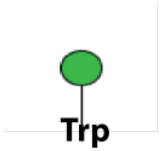
Functions: protein folding and trafficking; ECM composition

Glycolipids



Functions: Cell surface receptors; membrane composition

C-linked glycosylation



Functions: Unknown (stabilization?)

Background: *C*-linked glycosylation

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

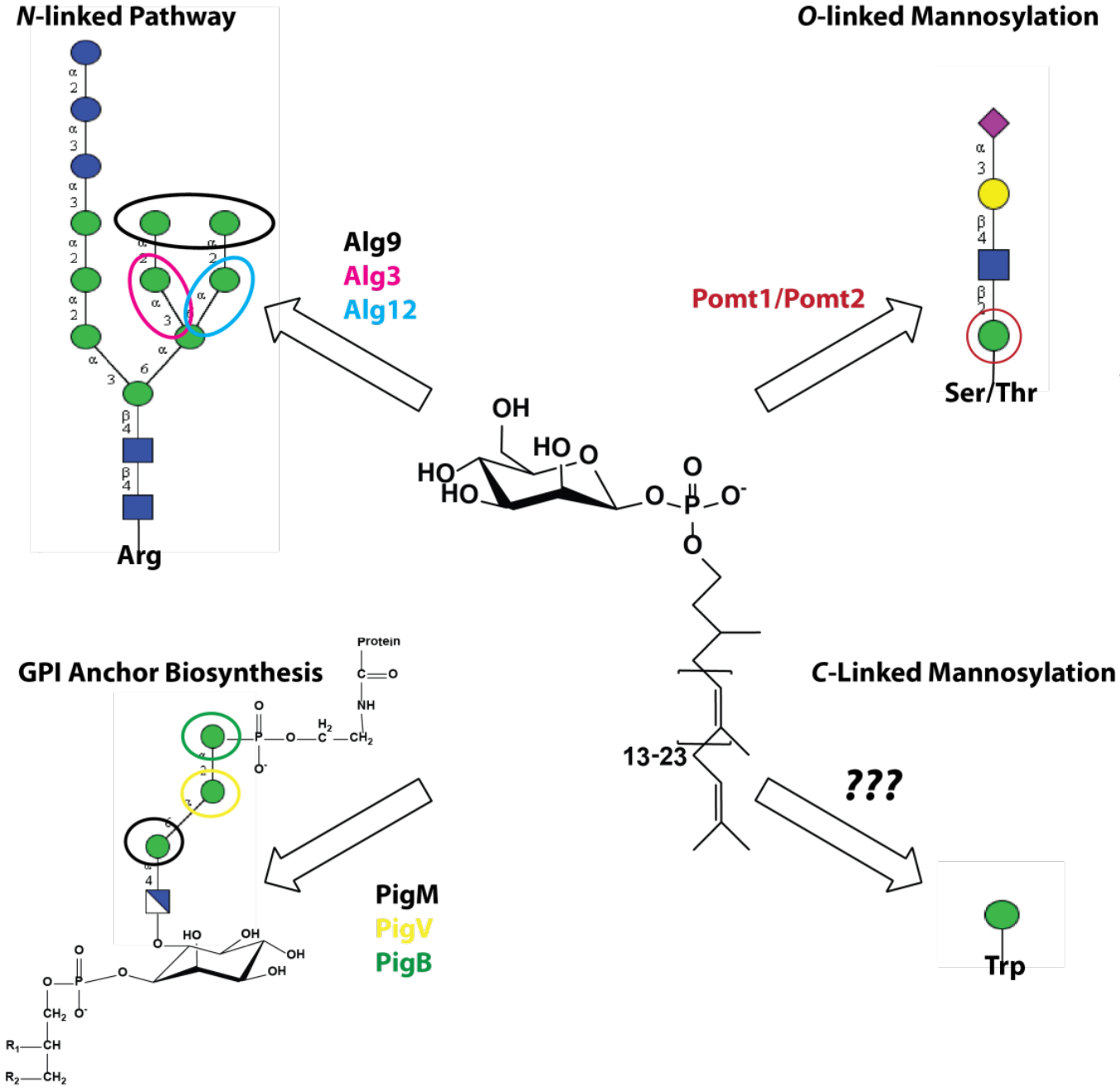
- **Mannose is transferred from dolichyl-phosphate-mannose (not GDP-mannose) by an endoplasmic reticulum-associated protein (Doucey et al, *Mol Biol Cell*, 1998)**

- **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**

Non-Leloir Transferases utilize dolichyl-phosphate mannose

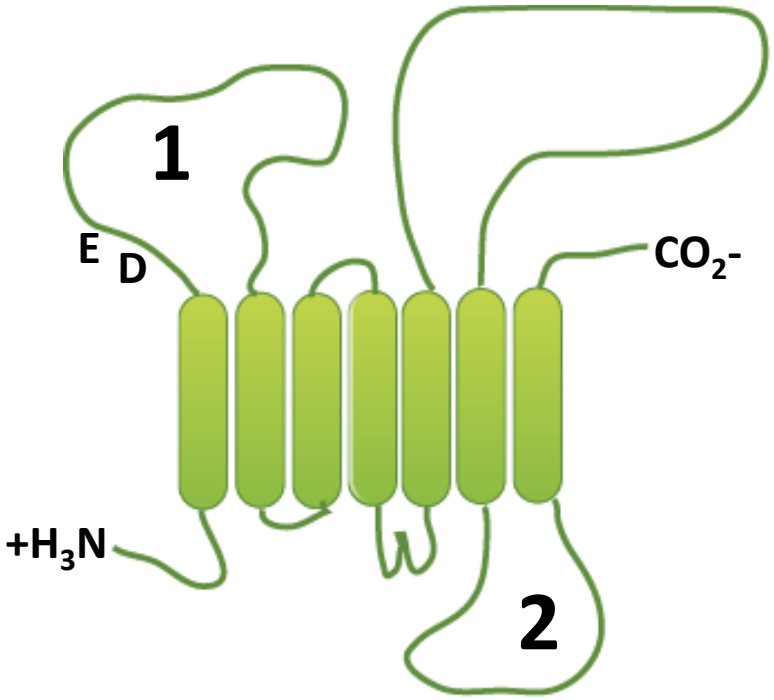


Non-Leloir transferases share a common ancestor and are evolutionarily related in two distinct sequence regions

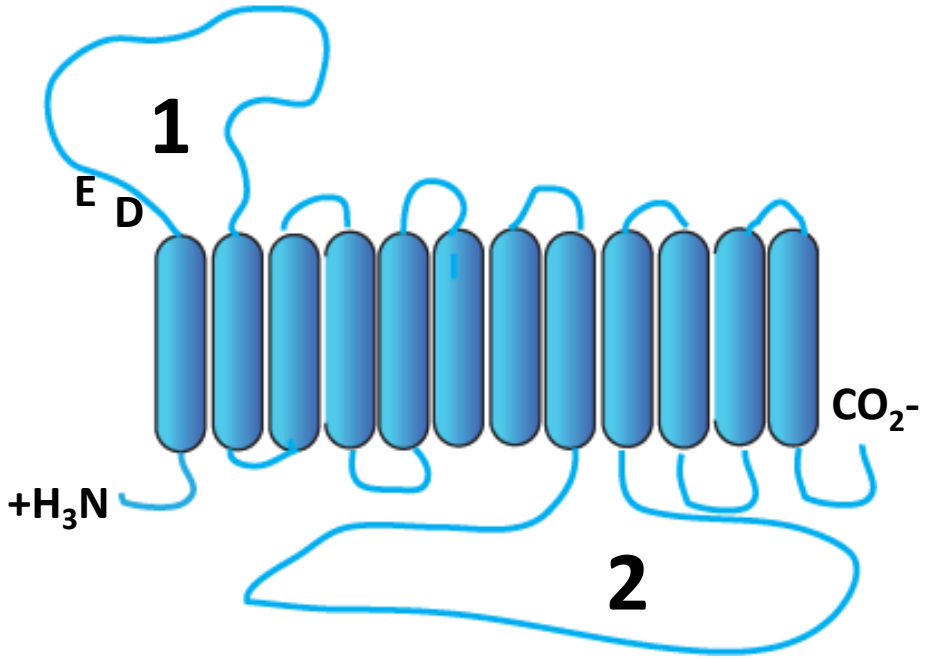
(Oriol et al, *Mol Biol Evol*, 2002)

Topology of Non-Leloir Mannosyltransferases

***O*-mannosyltransferases**



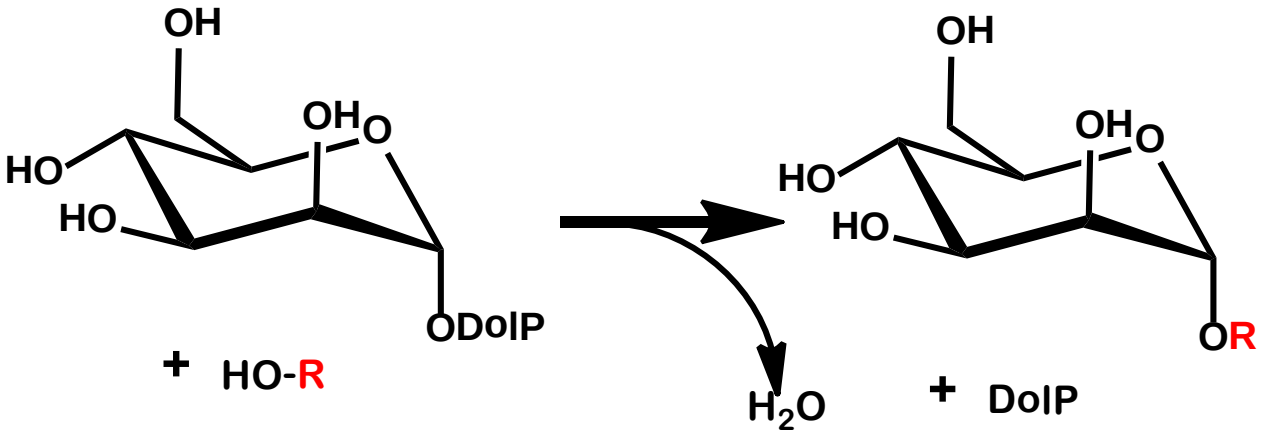
***N*-linked and GPI anchor**



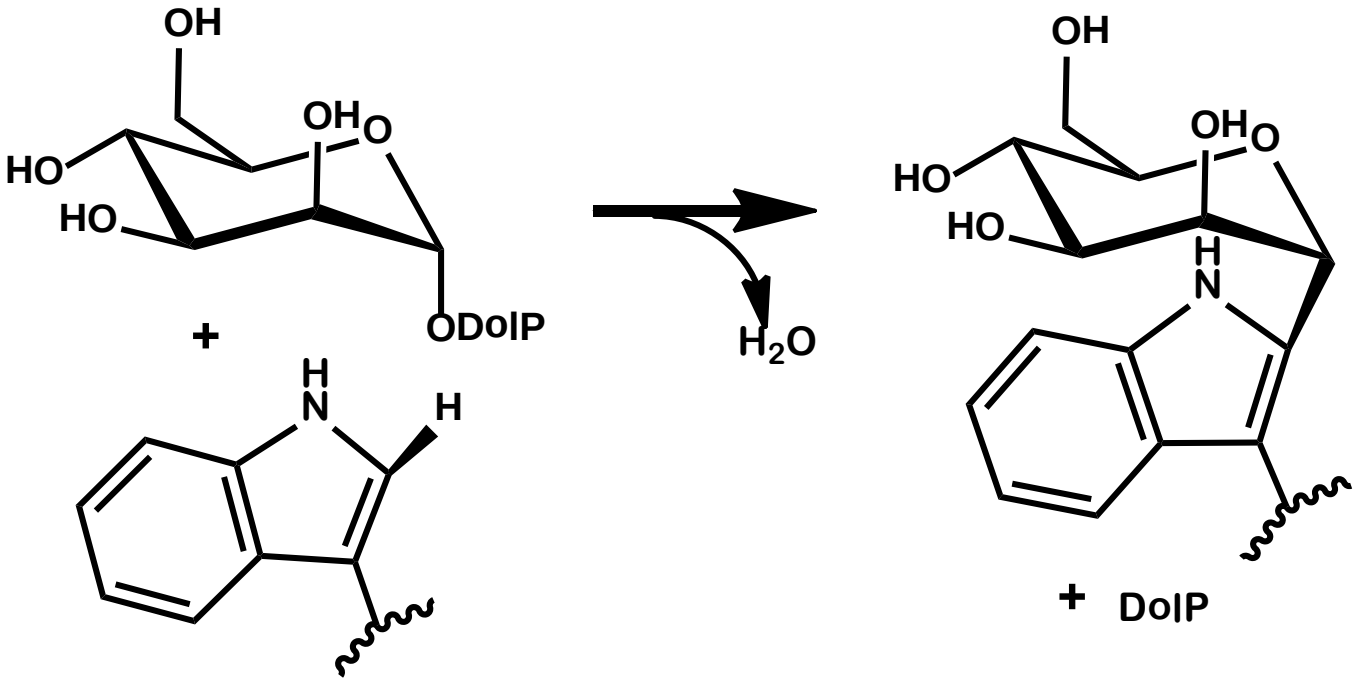
Loop 1: Contains N-terminal acidic domain (DE/EE/DD). Work on POMT1/2 family suggests this is the catalytic domain

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

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

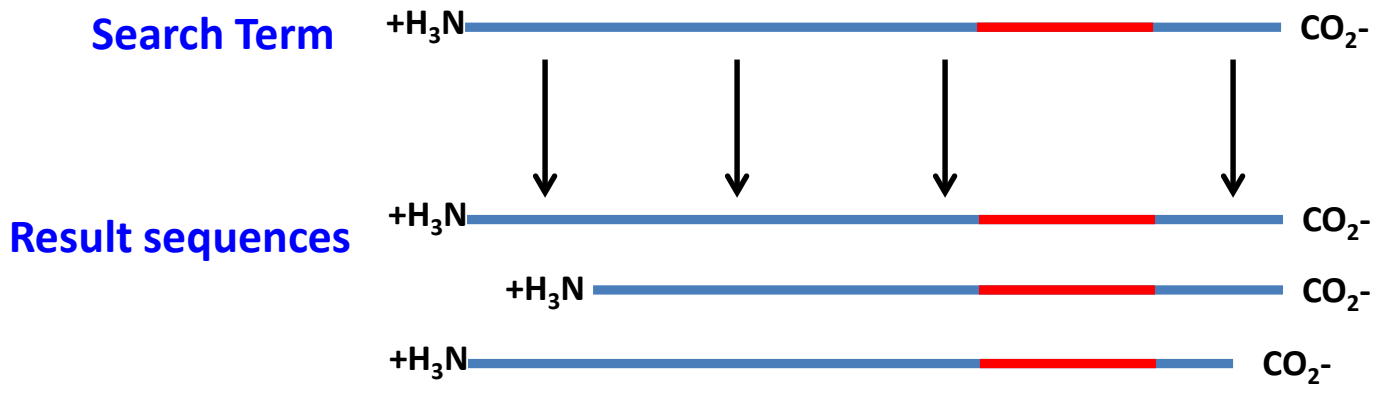


Transfer of α–Mannose from Dolichyl-Phosphate-Mannose to Tryptophan

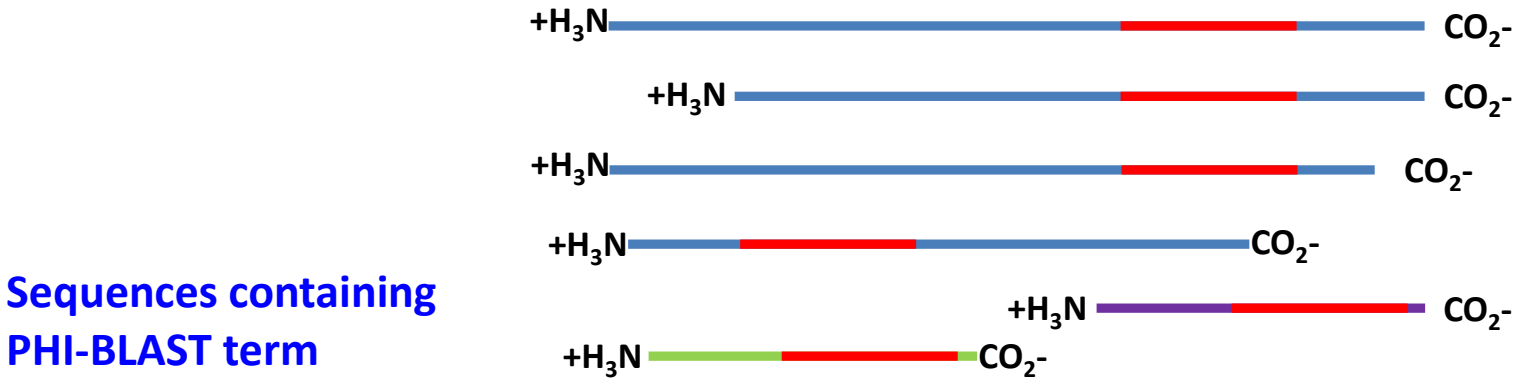


PHI-BLAST Method to Identify Potential *C*-MT Sequences

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



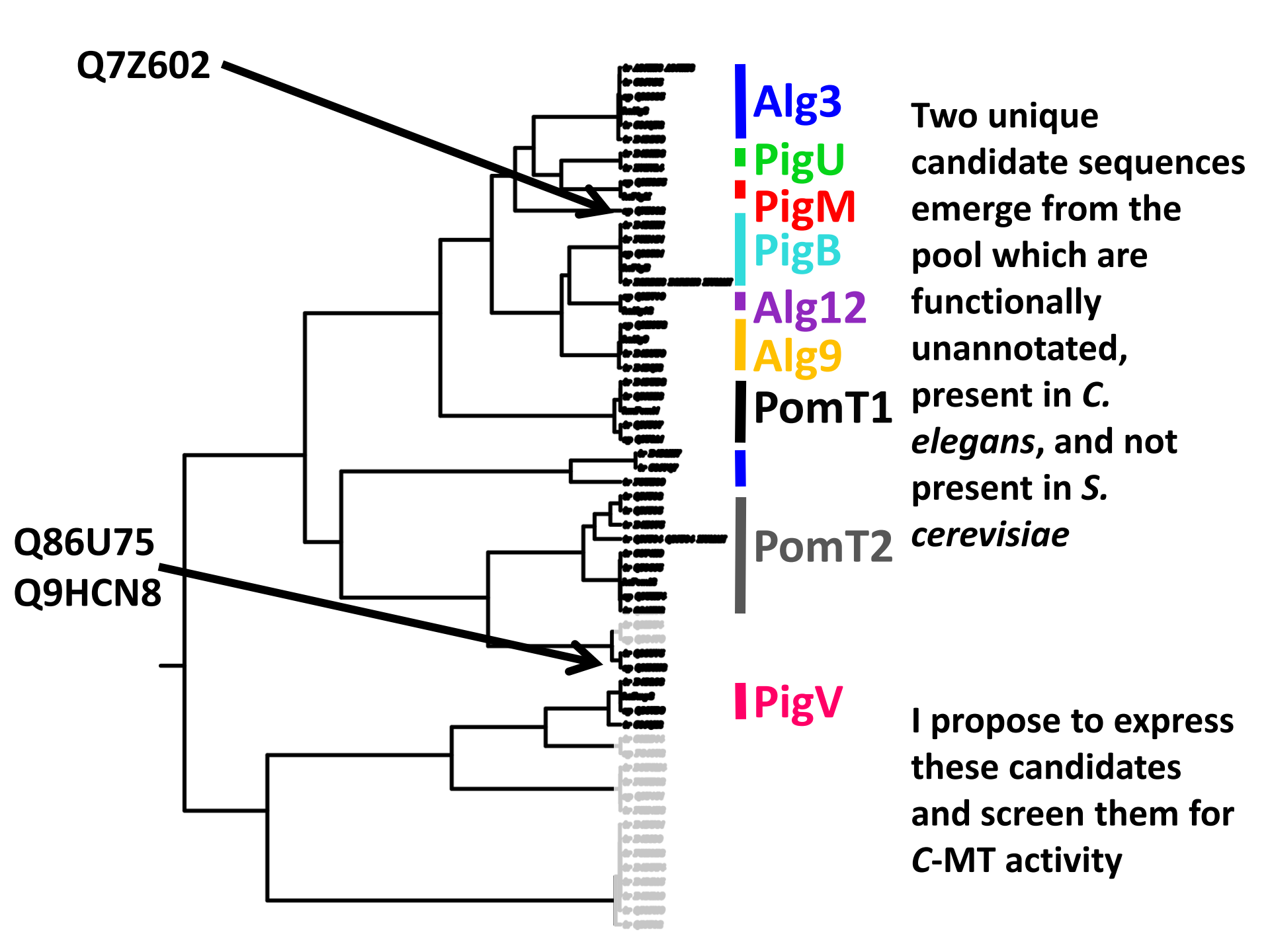
Step 2: Iterate PHI-BLAST search to find "loop 2" sequences in distantly-related sequences



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
Two 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

C-MT, what are my goals?

- **Identify C-MT**
- **Determine if acidic motif is responsible for catalysis**
- **Identify C-MT in other organisms**

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 glycomic strategy to identify response of *C*-mannosylated proteins.

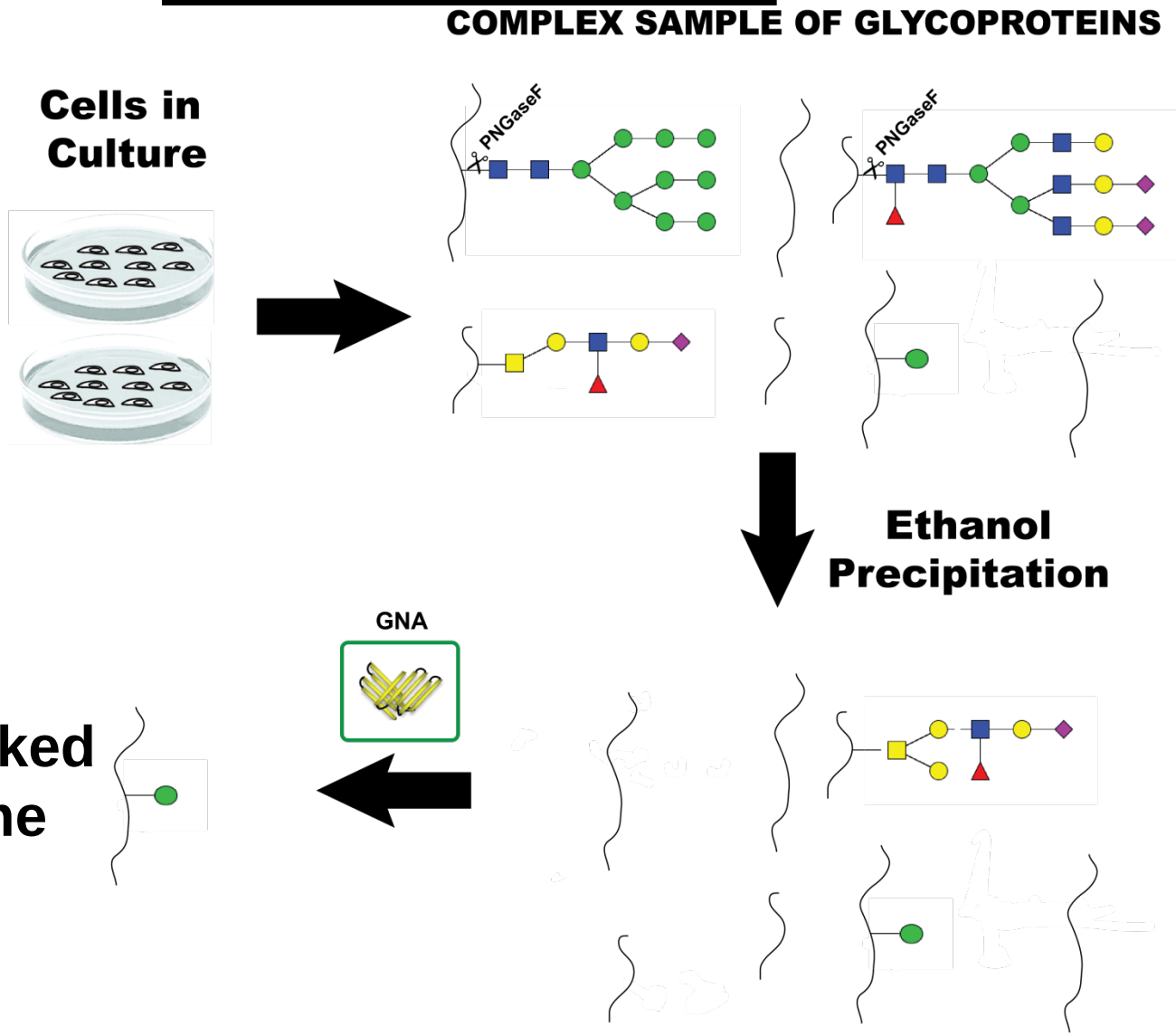
Glycoproteomics

- **Galanthus nivalis agglutinin (GNA) binds terminally exposed mannose, included mannosyl-tryptophan (Perez-Villar *et al*, *Glycobiology*, 2004)**

- **PNGaseF can remove mannose-containing *N*-linked glycans**

- **Cellular proteins can be deglycosylated with PNGaseF, the only remaining terminal mannoses will be mannosyl-tryptophan, which can be recognized by GNA-pulldown**

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



Goals

- **Get from 150 to 2500 C-mannosylated proteins**
- **Standardized proteomic method, identification of C-mannosylation under stimulated conditions (LPS-induced macrophage response)**
- **Identification of protein-protein interactions and activation pathways**

Acknowledgements

Prof. Lara Mahal

-Dr. Praveen Agrawal*

-Dr. Bianca Batista*

-Dr. João Ribeiro

-Linlin Wang

-Yaxuan Liang*

-Tomasz Kurcon*

-Zhongyin Liu

-Chris Vaiana

-Lauren Zhang

-Sarah Abbassi (undergraduate)

-Dr. Ku-Lung Hsu*

-Dr. Lakshmi Krishnamoorthy*

-Dr. Kanoelani Pilobello*

-Will Eng*

-Prof. Eva Hernando (NYU Med)

-Avital Gaziel-Sovran

-Prof. Iman Osman (NYU Med)

-Dr. Erica Friedman, M.D.

NIH DIRECTOR'S NEW INNOVATOR AWARD

NATIONAL INSTITUTES OF HEALTH