GlyCompare

Bokan, Ben

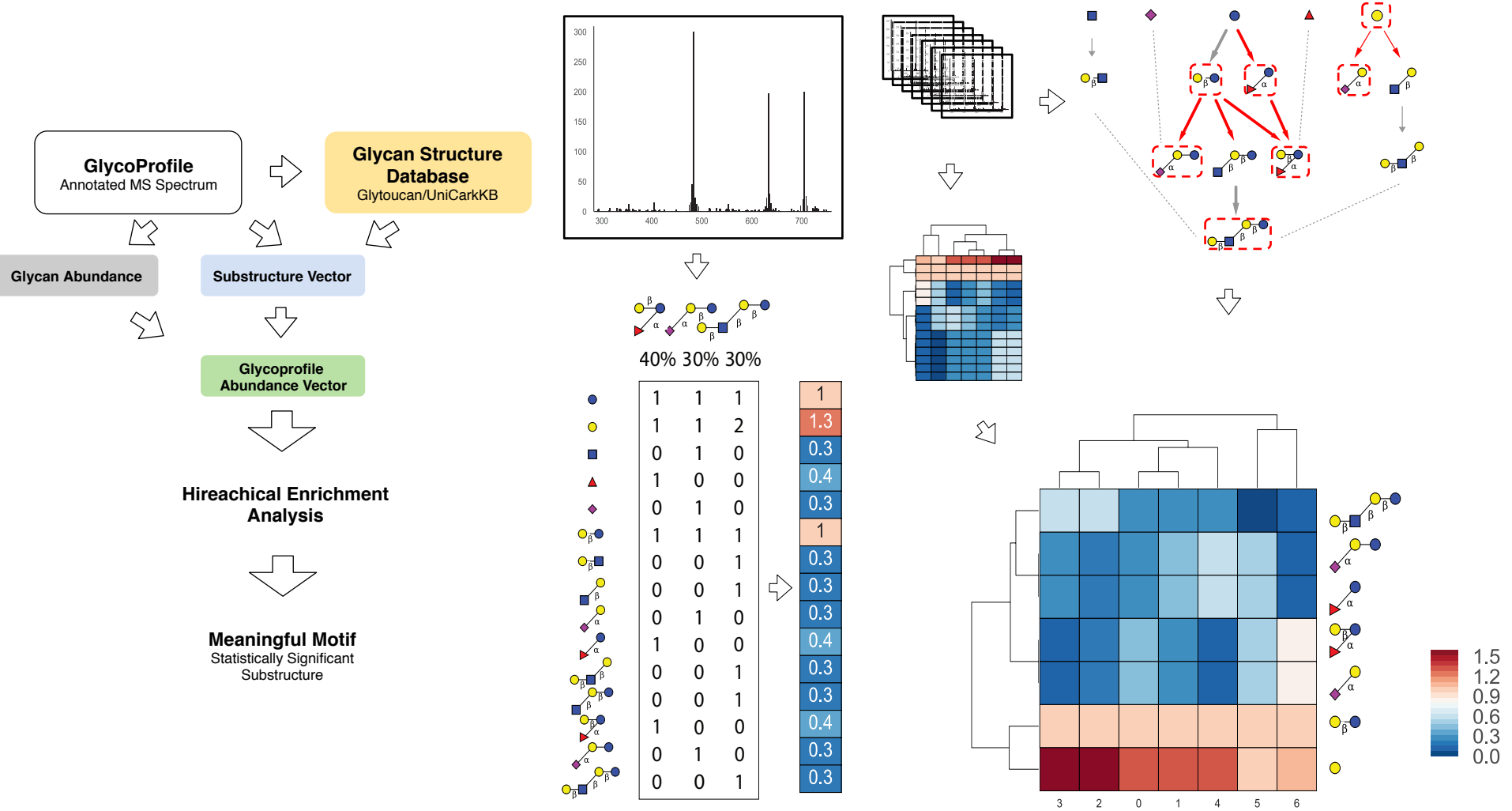
Compare the glycoprofile

- Comparing the glycoprofile is important. It tells us how glycan synthesis process is regulated or influenced by the genome modification. Difference, substructre(motif), synthesis perturbation.
- Reading glycoprofile is hard. Each glycoprofile has multiple glycan structures and each glycan structure has overlapping chemical structure.
- Comparing glycoprofile is even harder. Because each glycoprofile has different glycans, you cannot directly compare them.

GlyCompare

- Comparing large set of glycoprofiles
- Tracking the substructure(motif) abundance change
- Correlating the phenotypes with the motif’s abundance.
- Illustrating the perturbation of the synthesis netowork.

Figure1 Working pipeline



2 Result CHO

- Engineered CHO cell-line are capable to produce various forms of N-glycan. 16 glycoprofiles of digested glycan from EPO are curated (Yang's 2016), the engineered CHO cell lines includes the single or joint knockout targeting galactosylation (b4galt1/2/3/4), sialyation capping (st3gal4/6), N-glycan branching (mgat1/2/4A/4B/5) and core \alpha6-fucosylation. We generate our cluster and compare the result with two clustermap, the clustering by full glycan structure and the clustering by native mass spectrum.

Result 2.1 Clustering of 16 glycoprofiles

Figure 2A

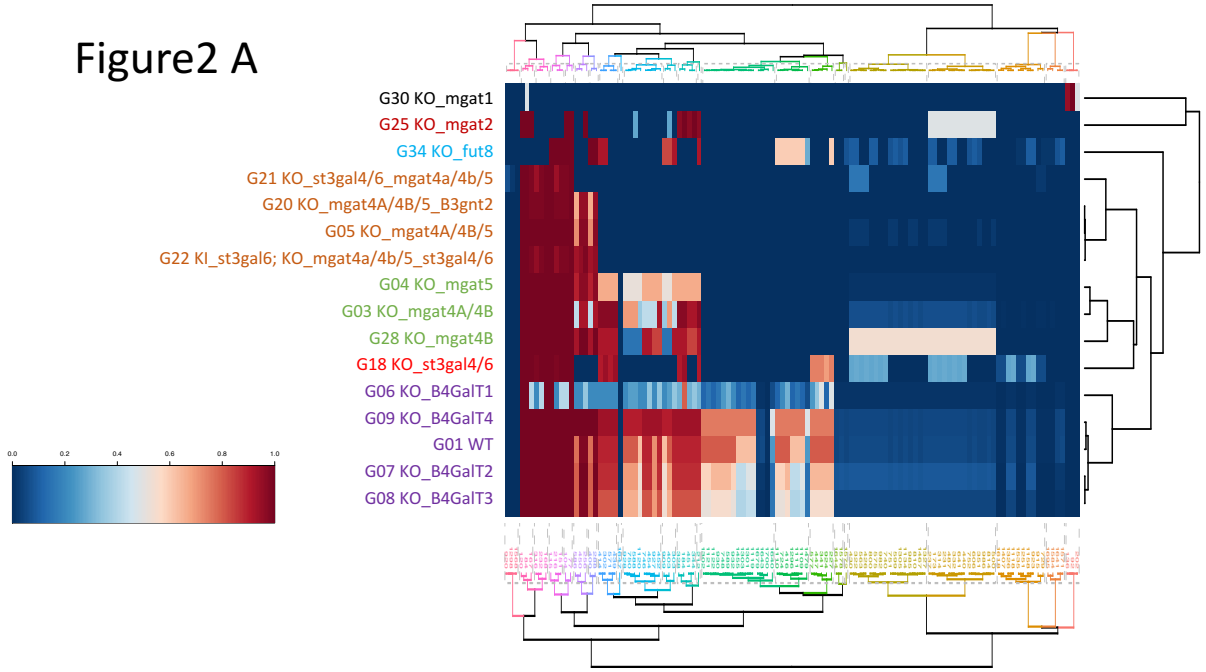
A glycoprofile vector with 722 glycan substructure are generated and then reduced to 118 glycan substructures. The cluster map (Figure2 A) shows the glycoprofile are clustered based on the severity comparing with the WT. Each cluster has distinguished glycan structure patterns. The group contains WT also has b4galt1/2/3/4/ knockout. The mid severity group contains mainly magt4b/4a&4a/5 knockout and st3gal4/6 knockout. The group that contains joint knockout has KO_st3gal4/6_mgat4a/4b/5, KO_mgat4A/4B/5_B3gnt2, KO_mgat4A/4B/5, KI_st3gal6; KO_mgat4a/4b/5_st3gal4/6. The most different glycoprofiles, that falls on the other side of the clusters are glycoprofiles with Fuc 8, mgat2, mgat 1 knockout. It means these three glycoprofiles have major different glycol-substructures.

The clustering with annotated glycan is bad because there are only a few glycans.

Figure 2B/Figure2C

- First, the glycans abundant table are also used for clustering as a comparison. (Figure2 C) It is worse than the clustering with the glyco-substructure. The glycan abundant table used for clustering is very sparse. The cluster is mainly based on the existence of the glycans and the distances between profiles are not well linearly characterized. The cluster is not consistent with the NCM and our result.
- Second, the grouping result is very consistent with the clustering by using the native mass spectrum data. The disagreement comes with the rearrange of the KO mgat2 and KO fu8. Since our method takes care of the structure difference across isomers. We can distinguish the structure that has same mass and the structure change that effect lots of substructures. Thus, we have a different way to differentiate the glycoprofile KO mgat2 and KO fut8.

Figure2 A



Left side is the glycoprofiles in the NCM paer, the right side is the shared glycoprofile from our pool. The major difference comes from the position of the KOfu8.

77 glycan structure

Figure2 B

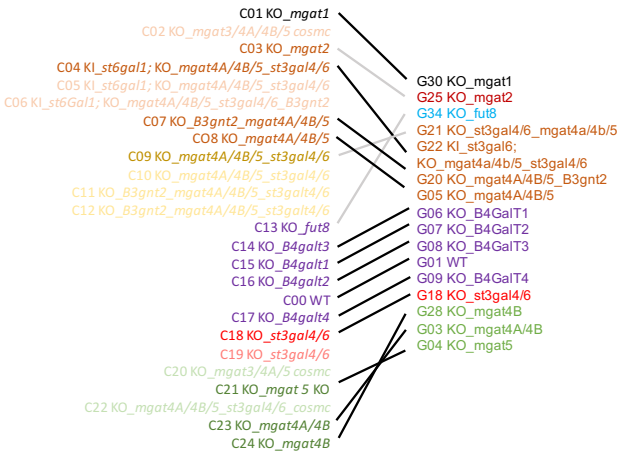
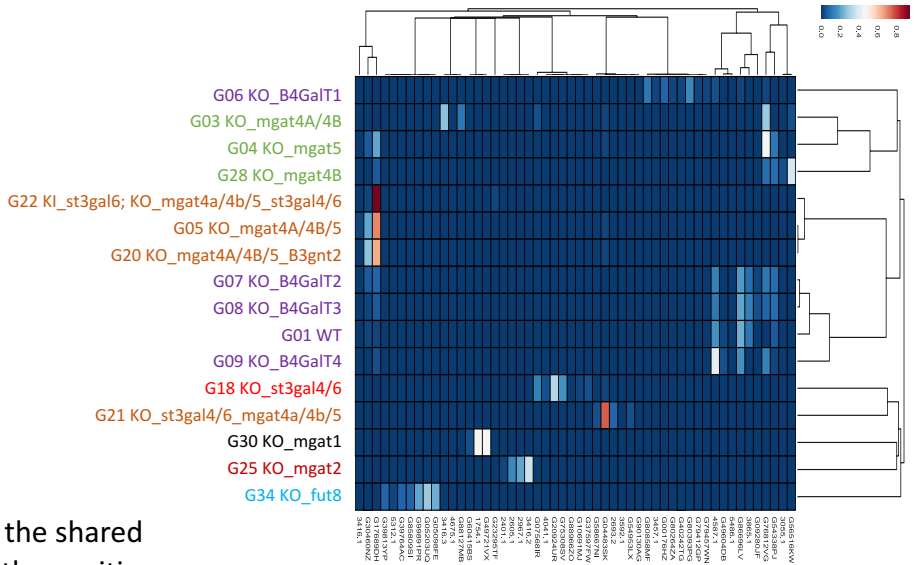


Figure2 C



Left side is the glycoprofiles in the NCM paer, the right side is the shared glycoprofile from our pool. The major difference comes from the position of the KOfu8.

Result 1.2 Characterization of substructure abundance change

- By setting the threshold for cutting the motif cluster. 24 clusters of substructure are picked and the representative substructure for each cluster are generated(see supplementary). The mean abundance of a cluster is used to represent the abundance of the representative substructure. This is the first time to have a comprehensive, automatic quantification of the abundance variation of substructure across multiple glycoprofile.

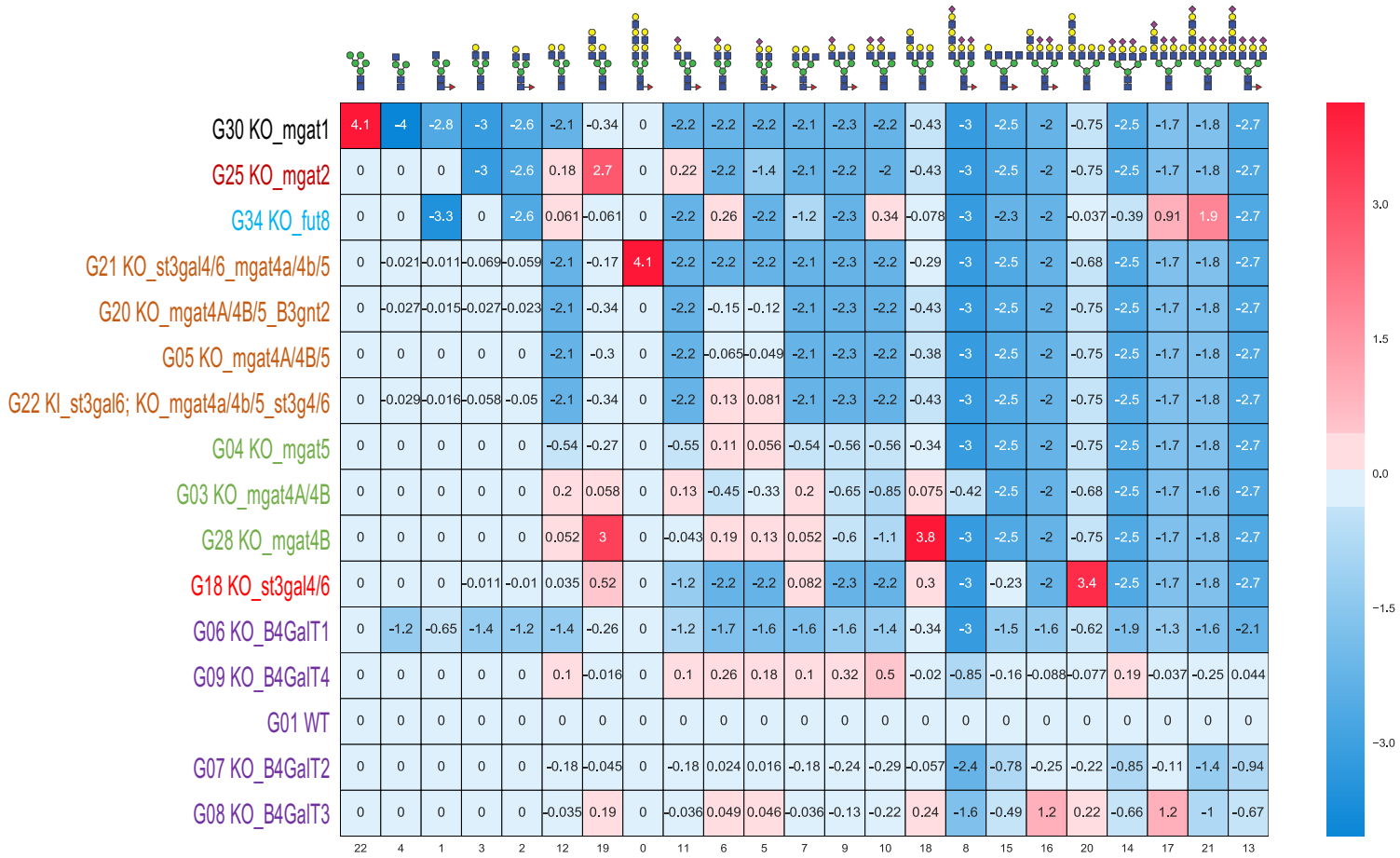


Figure2 D: We are able to quantify the relative abundance of the representative substructure across 16 glycoprofiles. Each column is rescaled with z score. And adjusted with the WT abundance(the abundance of the WT is 0). The substructures from left to right are reordered from single-antennary to tetra-antennary. It shows the decreasing of the complex glycan structure when the knockout becomes more complicate, from bottom to the top. The substructure in deep red implies the significantly increase and several of them have not been noticed before.

Figure 2 EFG

- Figure2 E shows in the KO_fu8 profile, the relative abundance of structures without fucose have significant decrease. But the tetra-antennary polyLacNac elongated N-glycan with no fucose increases as well as the tri-antennary one.
- Figure2 F shows in the KO_stgal4/6 profile, the relative abundance of structures with sialyation have significant decrease. But the tetra-antennary and tri-antennary polyLacNac elongated N-glycan with no sialyation increase.
- Figure2 G shows the knockout in mgat 4B, mgat 4A/4B and mgat 5. Most of the tetra-antennary structures decrease. The Mgat 4B and mgat 5 both have significantly decrease in tri-antennary polyLacNac elongated N-glycan. While mgat 4B has significantly increase in tri-antennary LacNac elongated N-glycan.

Figure2 E

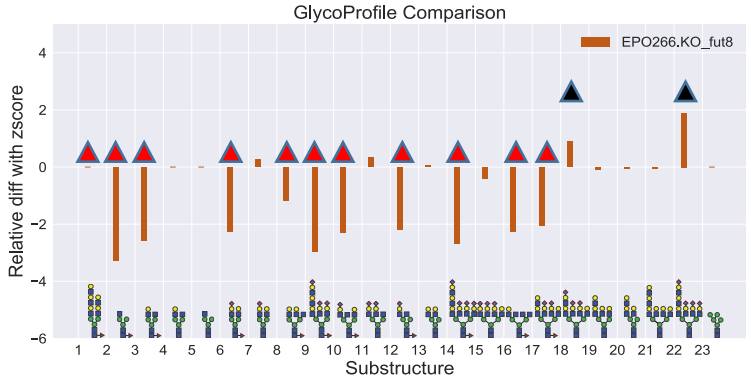


Figure2 F

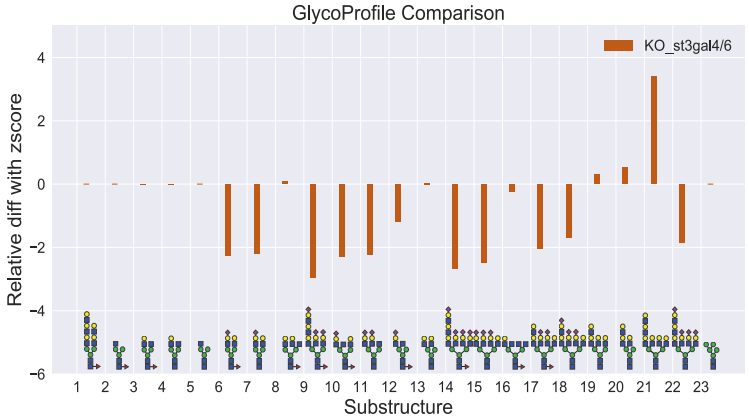
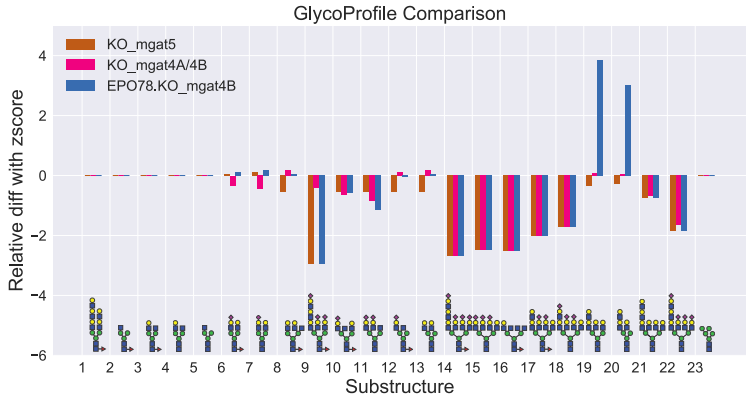
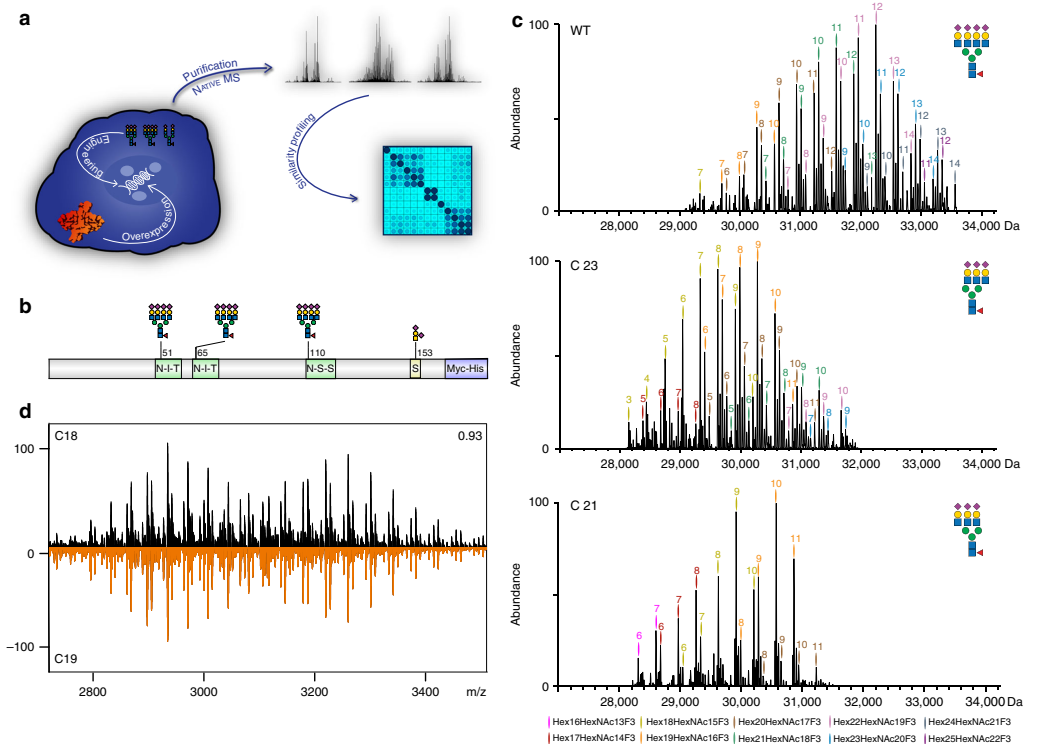


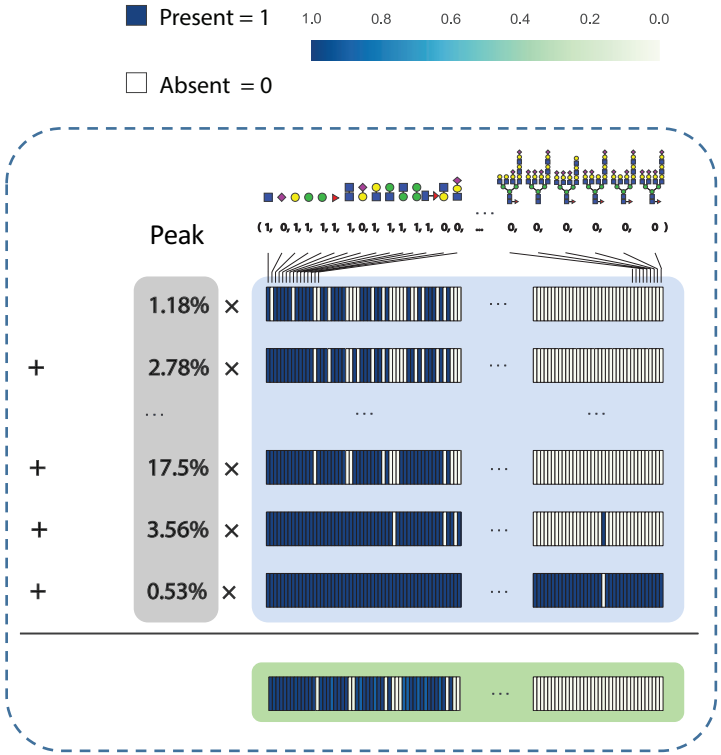
Figure2 G



Data 1 Part 3 Inferring the glycoprofile abundance vector from native mass spectrum (ongoing)

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-05536-3 ARTICLE





The plot on the left side comes from the **Nature Communication** paper, which shows the native mass spectrum profile for WT, KO mgat4A, KO mgat4B profile. The plot on the right shows it can be transferred to motif vector.

Green box : glycoprofile abundance vector
Grey box: glycan abundance
Blue box: motif structure for each signal peak
This plot shows that how the native mass spectrum data is transformed to glycan **Substructure Vector** and **Glycoprofile Abundance Vector**

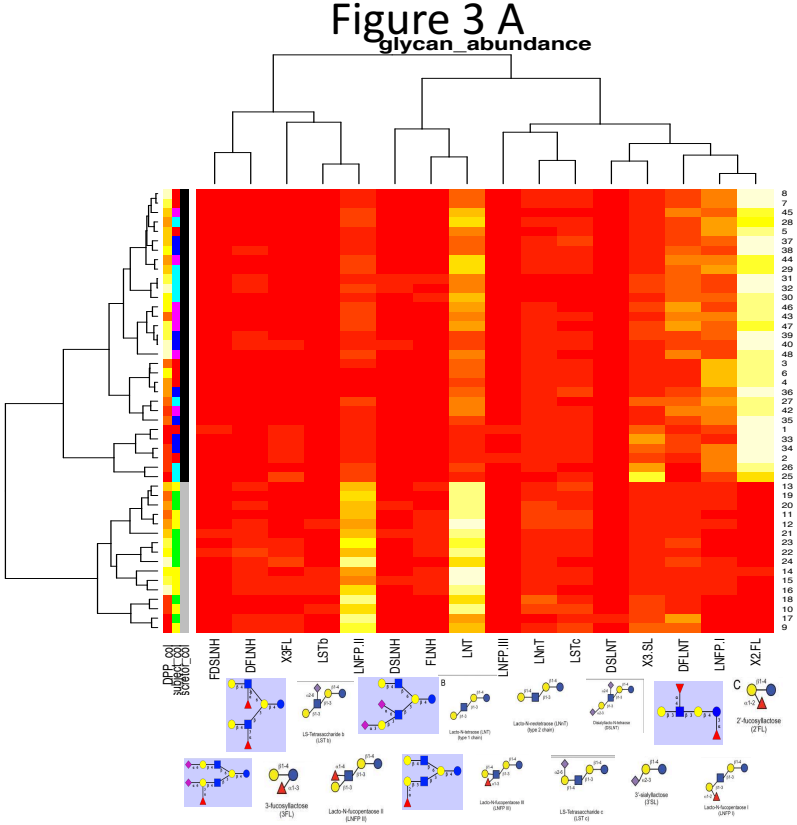
GlyCompare

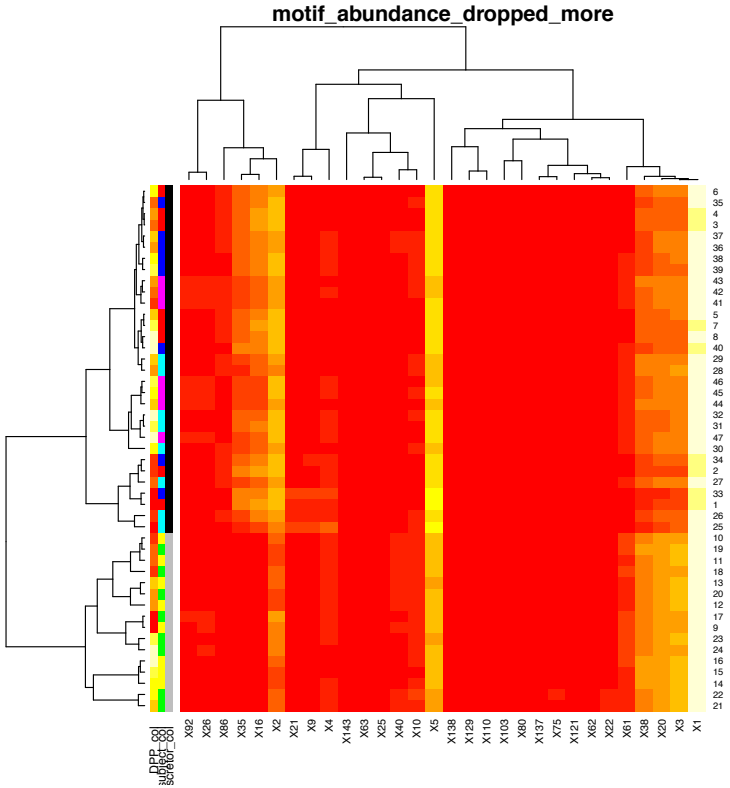
- Comparing large set of glycoprofiles
- Tracking the substructure(motif) abundance change
- Correlating the phenotypes with the motif’s abundance.
- Illustrating the perturbation of the synthesis netowork.

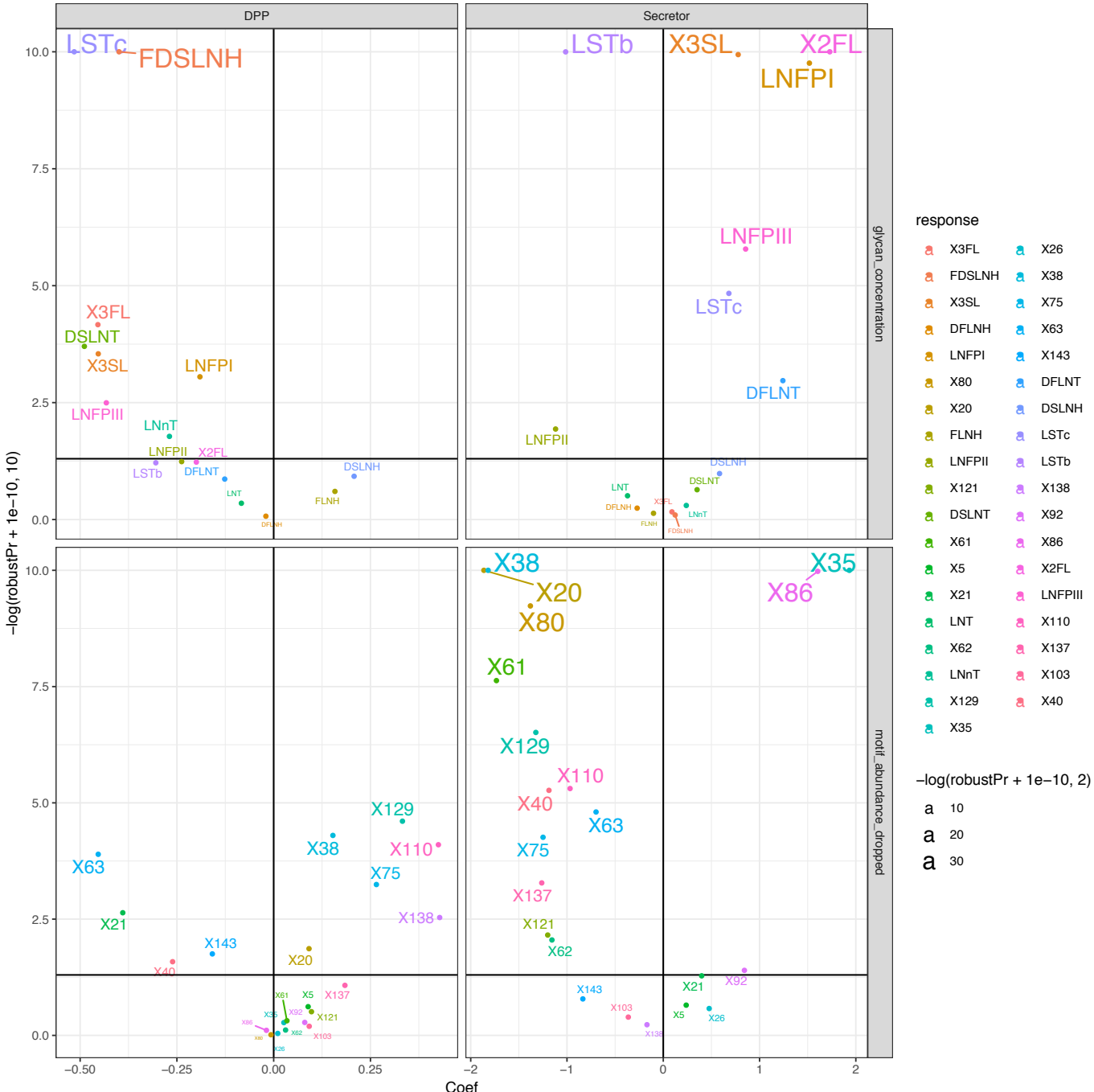
Data 2: HMO data 16 structure, 47 (48-1) glycoprofile

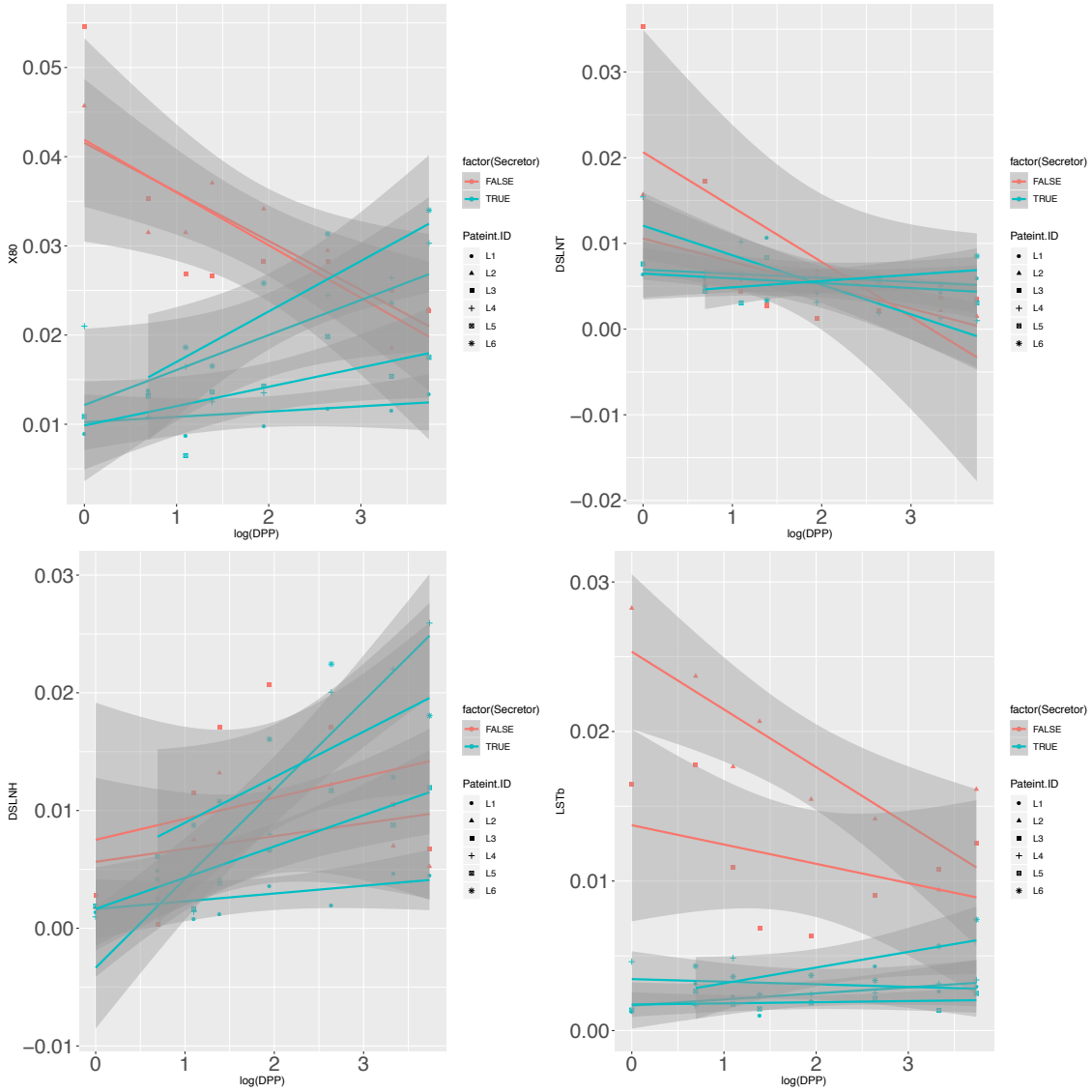
- Results
- Heatmaps
  - Motifs and glycans segregate secretor vs non-secretor
  - Second major organizing trend is time
  - Syalaition decreases over time
  - secretors enriched & non secretors for: 26, 92, 2, 16, 35, 86
- Regression:
  - X80 is anticorrelated with secretor status (log odds, -.015, p=1e-2), X21 is positively correlated with secretor status.
  - X129 & X110 are positively associated with time while syaliated motifs (X?, X??, and X???) are negatively associated with time.
- Scatterplots
  - X80 abundance is anticorrelated with time in non-secretors and positively associated with log time in secretors.
  - This trend is not retrievable in DSLNH, DSLNH and LSTb, the X80 containing HMOs.
- Secretor/non-secretor spearman correlations
  - X?, X??... are positively correlated in secretors and negatively correlated in non-secretors
  - X?, X??... are positively correlated in non-secretors and negatively correlated in secretors
  - motifs positvely correlated in secretors are (fucosylated, fuc & syalaited, galactosylated…?)
  - glycan leve observations
    - LSTb is correlated to LSTc in nonsecretors but not secretors suggesting a secretor-status dependent competition

Figure 3 A


glycan_abundance
FDSLNH
DFLNH
X3FL
LSTb
LNFP.II
DSLNH
FLNH
LNT
LNFP.III
LNnT
LSTc
DSLNT
X3.SL
DFLNT
LNFP.I
X2.FL
B
C
3-fucosyllactose (3FL)
Lacto-N-fucopentaose II (LNFP II)
3'-sialyllactose (3'SL)
2'-fucosyllactose (2'FL)


motif_abundance_dropped_more


DPP
Secretor
glycan_concentration
motif_abundance_dropped
−log(robustPr + 1e−10, 10)
Coef
response
X3FL
FDSLNH
X3SL
DFLNH
LNFPI
X80
X20
FLNH
LNFPII
X121
DSLNT
X61
X5
X21
LNT
X62
LNnT
X129
X35
X26
X38
X75
X63
X143
X137
X103
X2FL
LNFPIII
LSTc
LSTb
X138
X86
X92
X40
X110
−log(robustPr + 1e−10, 2)
10
20
30


X80
0.05
0.04
0.03
0.02
0.01
0
1
2
3
log(DPP)
DSLNT
0.03
0.02
0.01
0.00
−0.01
−0.02
DSLNH
LSTb
factor(Secretor)
FALSE
TRUE
Pateint.ID
L1
L2
L3
L4
L5
L6