



Supporting Information

for

A systems-based framework to computationally describe putative transcription factors and signaling pathways regulating glycan biosynthesis

Theodore Groth, Rudiyanto Gunawan and Sriram Neelamegham

Beilstein J. Org. Chem. **2021**, *17*, 1712–1724. [doi:10.3762/bjoc.17.119](https://doi.org/10.3762/bjoc.17.119)

Comparison of wet-lab studies and entries in DBs as well as Alluvial plots for all cancer types

Level of agreement between wet-lab studies and corresponding entries in the databases.

Proposition: We determined if findings reported in literature (see Introduction section) have corresponding entries in bioinformatics databases that present TF–(glyco)gene relationships. Such comparisons were made with Cistrome DB, GTRD, and Regulatory Circuits DB (see table below). The footnote explains the metrics used for this comparison.

Conclusion: The analysis concludes that experimental data reported in literature only have partial support in currently established databases. The lack of agreement was due to several reasons: i) Sometimes, the TF was not included in the curation of the database; ii) sometimes, the TF in question was filtered out due to the database criteria for establishing TF–gene relationships; and iii) the putative TF–glycogene relationship is present in some cell systems, but not the cell system reported in literature. This suggests that the same TF–glycogene relationship may not hold for all cell types.

Table S1: To what degree are interactions reported in literature observed in various databases?

literature reports ^a		GTRD ^b		Cistrome database ^c	Regulatory Circuits ^d	
glycogene	TF	expt. system	cell	cell system where TF–glycogene relationship was found in database		
B4GALT1	SP1	A549		A375 (malignant melanoma); GM12878 (female B-cells lymphoblastoid cell line); HCT-116 (colon carcinoma); HEK293 (embryonic kidney); HUES64 (embryonic stem cells); liver	SP1 not present in DB	107 tissue-specific relationships with conf ≥ 0.1 , significant
ST6Gal1	HNF1a	HepG2		not found (N.F.)	HNF1A not present in DB	2 with conf ≥ 0.1 , (hepatocyte, spinal cord)
MGAT5	ETS2	NIH3T3		N.F.	N.F.	N.F.
ST6Gal2	HNF1a	HepG2		N.F.	HNF1A not in DB	N.F.
B3GNT8	c-JUN	Gastric Carcinoma		A549 (lung carcinoma); Dexamethasone-treated HUVEC-C. Metaclusters are absent in gastric carcinoma	N.F.	N.F.
HS3ST1	ZNF263	HeLa		HEK293 (embryonic kidney) K562 (myelogenous leukemia). Metaclusters are absent in HeLa.	ZNF263 not in DB	ZNF263 not in DB
HS3ST3A1	ZNF263	HeLa		K562 (myelogenous leukemia). Metaclusters are absent in HeLa	ZNF263 not in DB	ZNF263 not in DB

^aReported in literature (see Introduction section in the main manuscript). ^bGTRD: The GTRD curates TF binding sites on the genome by employing 4 different ChIP-Seq peak calling algorithms. Any binding region identified by more than one ChIP-Seq peak finding algorithm is designated as a metacluster. Thus, metaclusters represent TF binding regions with support from multiple ChIP-Seq peak calling algorithms. In our analysis, we parsed glycogene transcriptional start sites (TSSs) from Ensembl BioMart. Only transcripts with an Ensembl support level of 1 were used. To identify regulating TFs, the distance between the center of a metacluster and a glycogene TSS was computed. TF metaclusters identified 2 kb upstream from any glycogene TSS constituted a TF–glycogene relationship. This yielded a total 5,190 TF–glycogene interactions across 302 biological systems. ^cCistrome DB: TF–glycogene relationships were parsed from Cistrome Cancer DB by picking regulatory potential ≥ 0.5 and correlation values ≥ 0.4 . This yielded 22,654 TF–glycogene interactions in Cistrome Cancer DB. These thresholds were selected to find long- and short-range TF–glycogene interactions. ^dRegulatory Circuits: TF–glycogene relationships with a confidence score ≥ 0.1 were parsed from each of the 394 tissue-specific networks from Regulatory Circuits. This yielded 66,071 relationships across all 394 tissue systems.

Details: The following is a detailed explanation of our findings for each interaction and cell system:

SP1→B4GALT1 in A549 cell line:

- i) Cistrome DB: This relationship was not found in Cistrome DB since this TF was not included in the database. It is unclear if SP1 has been filtered from the database due to low tumor expression specificity, or if it was not included in the analysis.
- ii) GTRD: SP1 was found in the promoter region of 6 different cell systems: A375, GM12878, HCT-116, HEK293, HUES64, and the “liver” system. In the A549 data (wet-lab data are reported in literature for this system), however, the SP1→B4GALT1 relation was not supported by the multiple ChIP-Seq algorithm used by GTRD. In this regard, while the SP1→B4GALT1 relation was reported in A549 using the MACS algorithm in one study. Three other algorithms used in GTRD (GEM, PICS, SISRIS) do not support a SP1→B4GALT1 relationship in A549. Thus, the SP1→B4GALT1 metacluster for A549 is not reported in the GTRD.
- iii) Regulatory Circuits: This relation was observed in large cell lung carcinoma cell line network, which is derived from NCI-H460 and IA-LM cell lines (confidence threshold = 0.105). Since this network does not contain the A549 relationship, we conclude that SP1→B4GALT1 relation does not exist in A549 in Regulatory Circuits.

HNF1A→ST6Gal1/2 in HepG2 cell line:

- i) Cistrome DB: HNF1A was not found in the downloaded Cistrome DB dataset. It was also not found upon searching the online database portal (<http://cistrome.org/CistromeCancer/CancerTarget/>). This may be due to low expression of this TF across TCGA cancers.
- ii) GTRD: HNF1A data were available for 6 different cell systems across 9 distinct experiments. None of these systems supported a HNF1A→ST6GAL1 or HNF1A→ST6GAL1 metacluster relationship. Peaks called by GEM, MACS2, PICS, and SISRIS were loaded into the UCSC genome browser to see if individual algorithms supported the HNF1→ST6GAL1/2 relationship. No peak calling algorithm found HNF1A binding in the ST6GAL1 promoter.
- iii) Regulatory Circuits: The HNF1A→ST6GAL1 relationship exists in the “hepatocyte” (confidence = 0.110) and “spinal cord adult” (confidence = 0.103) networks. However, the HNF1A→ST6GAL1 confidence threshold for three HepG2 CAGE sequencing replicates, called “hepatocellular carcinoma cell line”, was low and below our cutoff criterion (confidence value = 0.013).

ETS2→MGAT5 in NIH3T3 cell line:

- i) Cistrome DB: ETS2 was not found in the downloaded Cistrome DB, nor was it searchable from the database portal (<http://cistrome.org/CistromeCancer/CancerTarget/>). This could be due to the lack of expression in the TCGA dataset.
- ii) GTRD: No metaclusters were identified supporting the ETS2→MGAT5 relationship.

iii) Regulatory Circuits: No ETS2→MGAT5 relationships were found in any of the tissue-specific networks in Regulatory Circuits

c-JUN—B3GNT8 in gastric carcinoma:

i) Cistrome DB: None of the correlation data or regulatory potential data present in Cistrome DB lists B3GNT8 being regulated by c-JUN. It is unclear for what reason this relationship was omitted.

ii) GTRD: dexamethasone-treated HUVEC-C cells and A549 cells support the c-JUN→B3GNT8 relationship. But this relationship is missing in gastric carcinoma systems, tissues and cell lines, and thus there is no evidence supporting this relationship.

iii) Regulatory Circuits: The B3GNT8 gene is absent in the regulatory networks of this database (see the other supplementary data files). Thus, we cannot assess validity of this relationship in Regulatory Circuits.

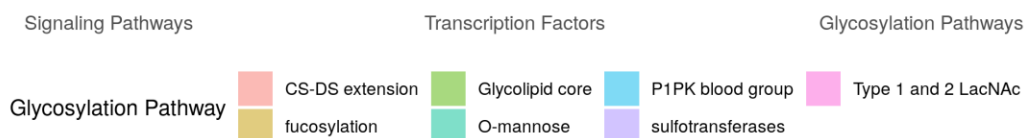
HS3ST1/HS3ST3A1→ZNF263 in HeLa cells:

i) Cistrome DB: ZNF263 was not found in the downloaded Cistrome DB, nor was it searchable from their database portal (<http://cistrome.org/CistromeCancer/CancerTarget/>). Low expression of this TF across TCGA cancers could potentially explain its absence.

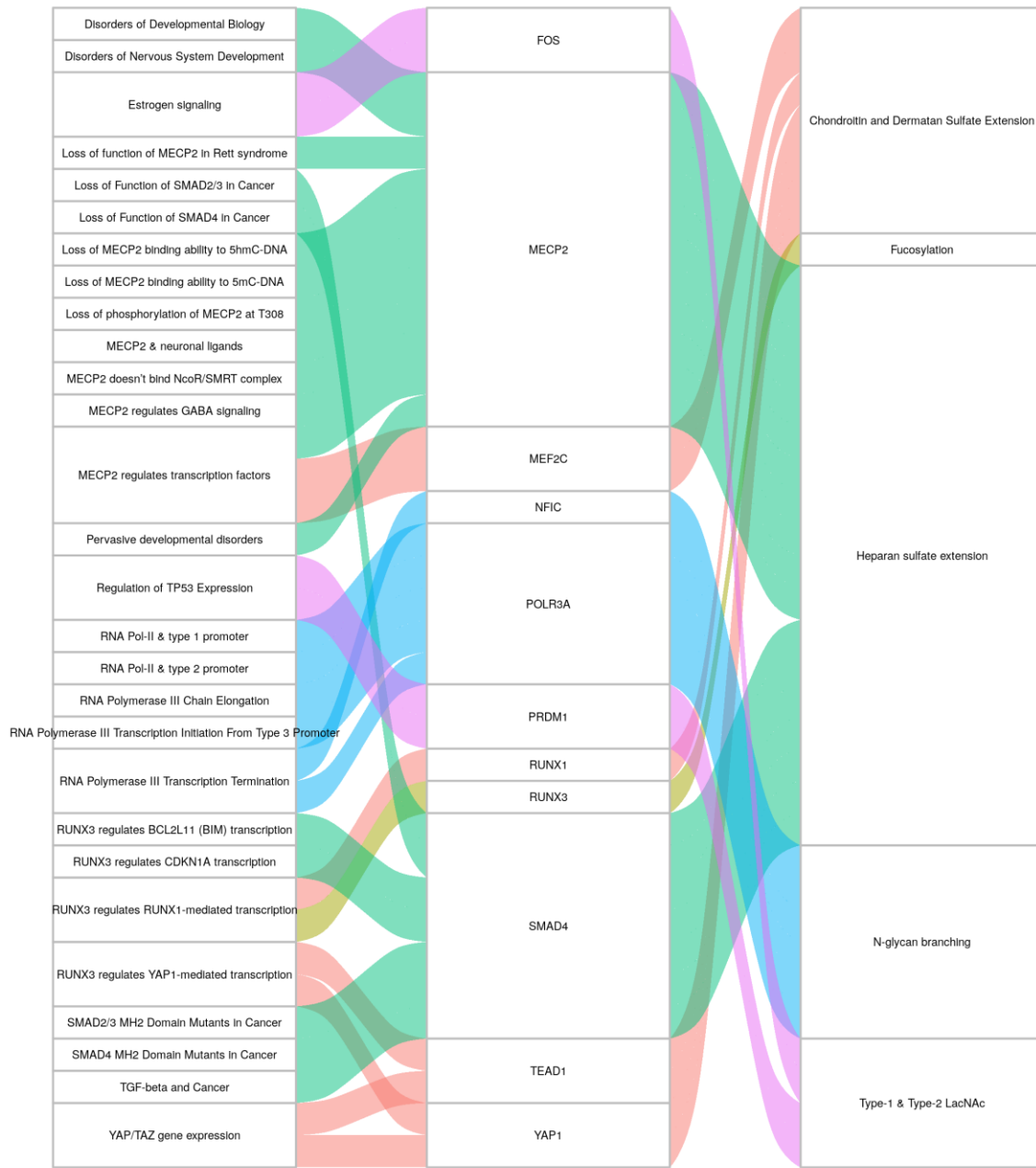
ii) GTRD: ZNF263 has metaclusters in the promoter region of HS3ST1 in both HEK293, as well as K562 cells. ZNF263 also has metaclusters in the promoter region of HS3ST3A1 in K562. However, this relationship was absent in HeLa cells.

iii) Regulatory Circuits: ZNF263 was not considered in the construction of tissue-specific networks. There is no position weight matrix (PWM) information present for ZNF263 (see Regulatory Circuits supplemental material).

BLCA Signaling to Glycosylation Pathway Matching



BRCA_1 Signaling to Glycosylation Pathway Matching



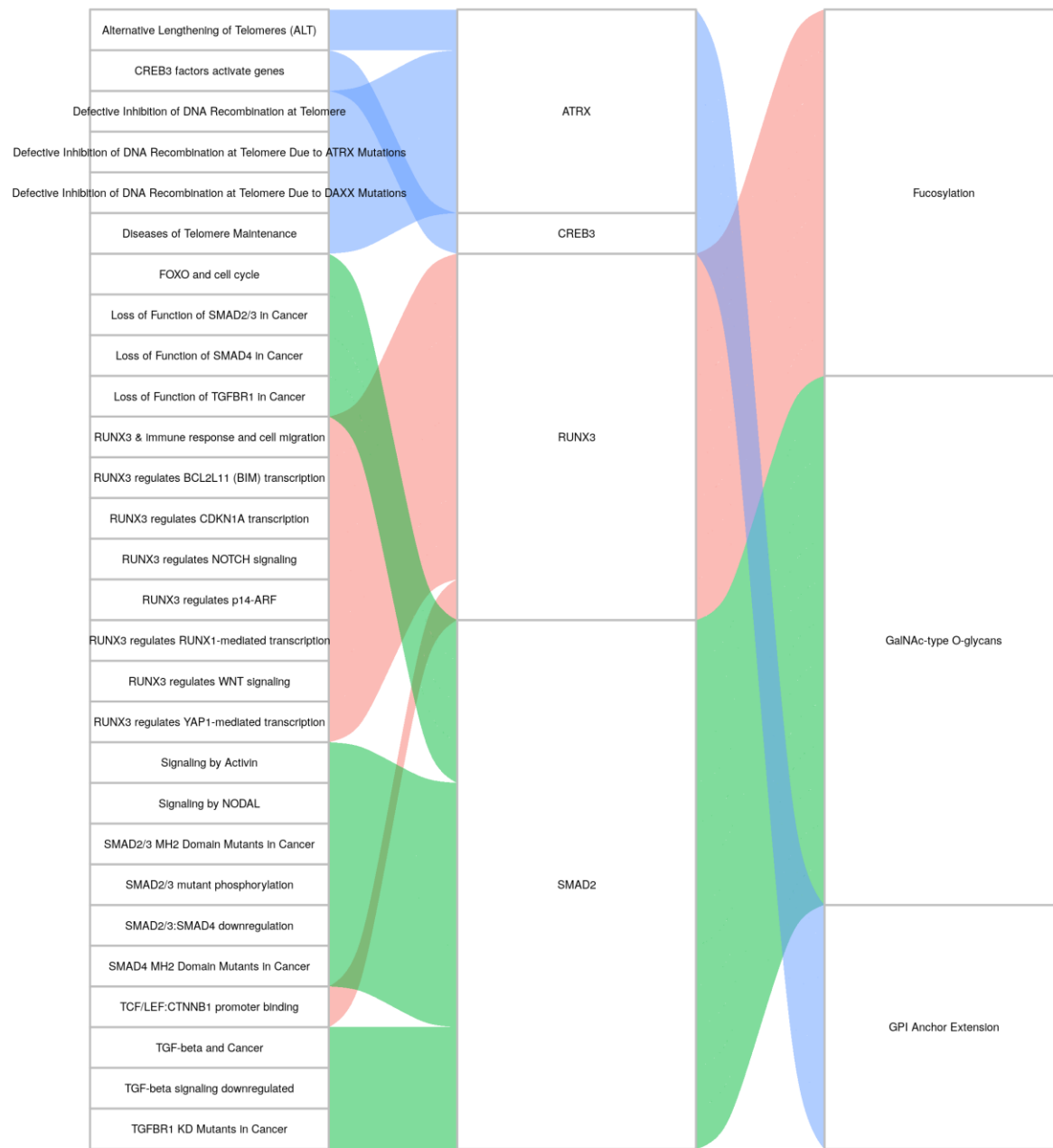
Signaling Pathways

Transcription Factors

Glycosylation Pathways

ation Pathway ■ Chondroitin and Dermatan Sulfate Extension ■ Fucosylation ■ Heparan sulfate extension ■ N-glycan branching ■ Type-1 & Typ

BRCA_2 Signaling to Glycosylation Pathway Matching



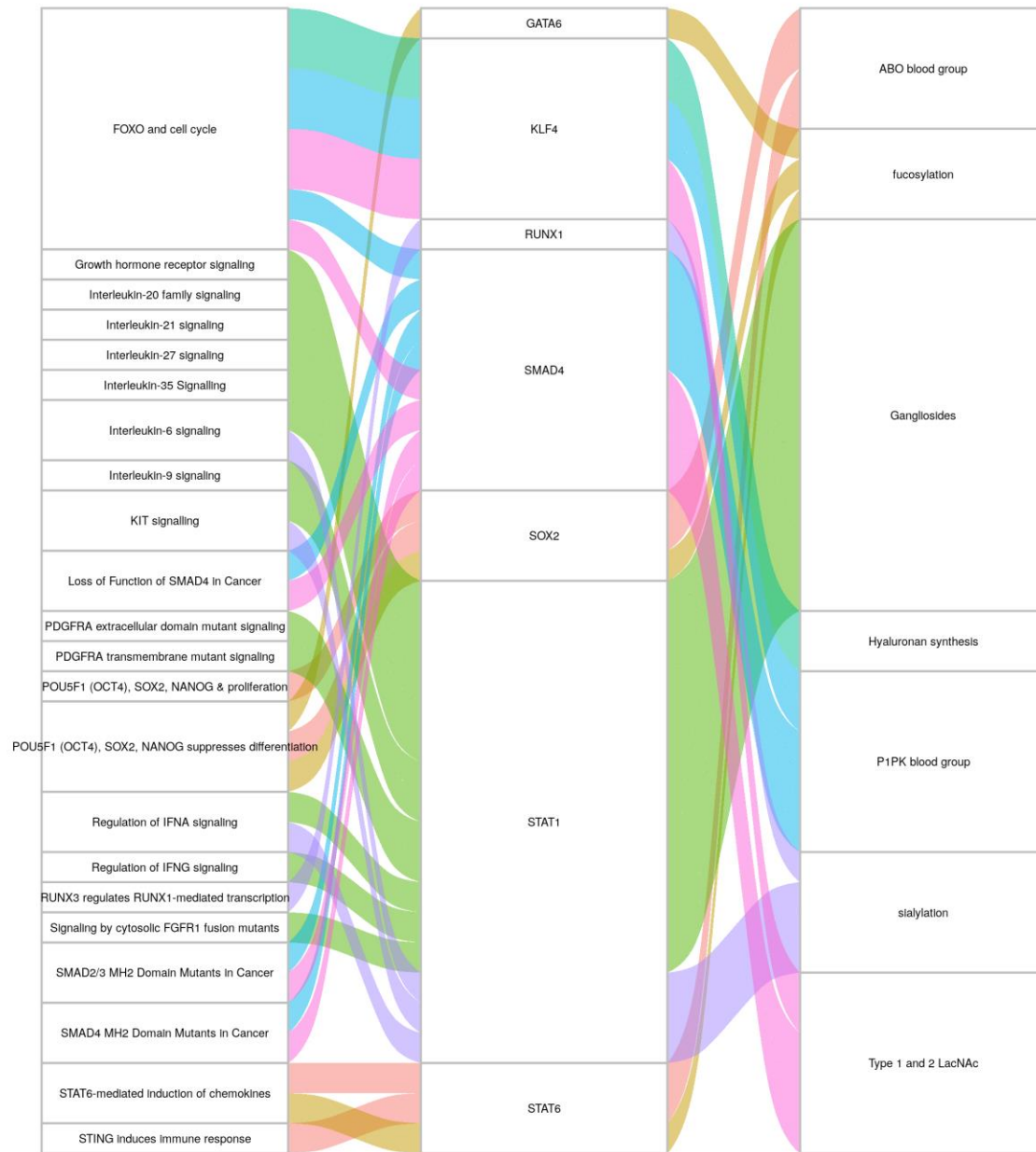
Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway ■ Fucosylation ■ GalNAc-type O-glycans ■ GPI Anchor Extension

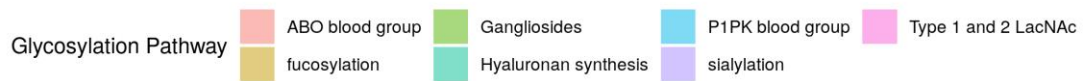
CESC Signaling to Glycosylation Pathway Matching



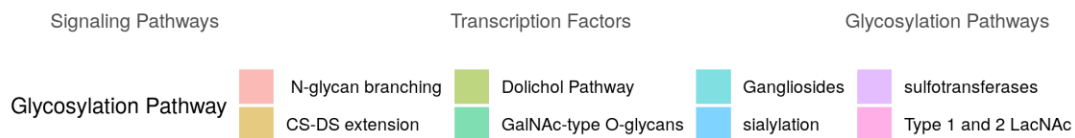
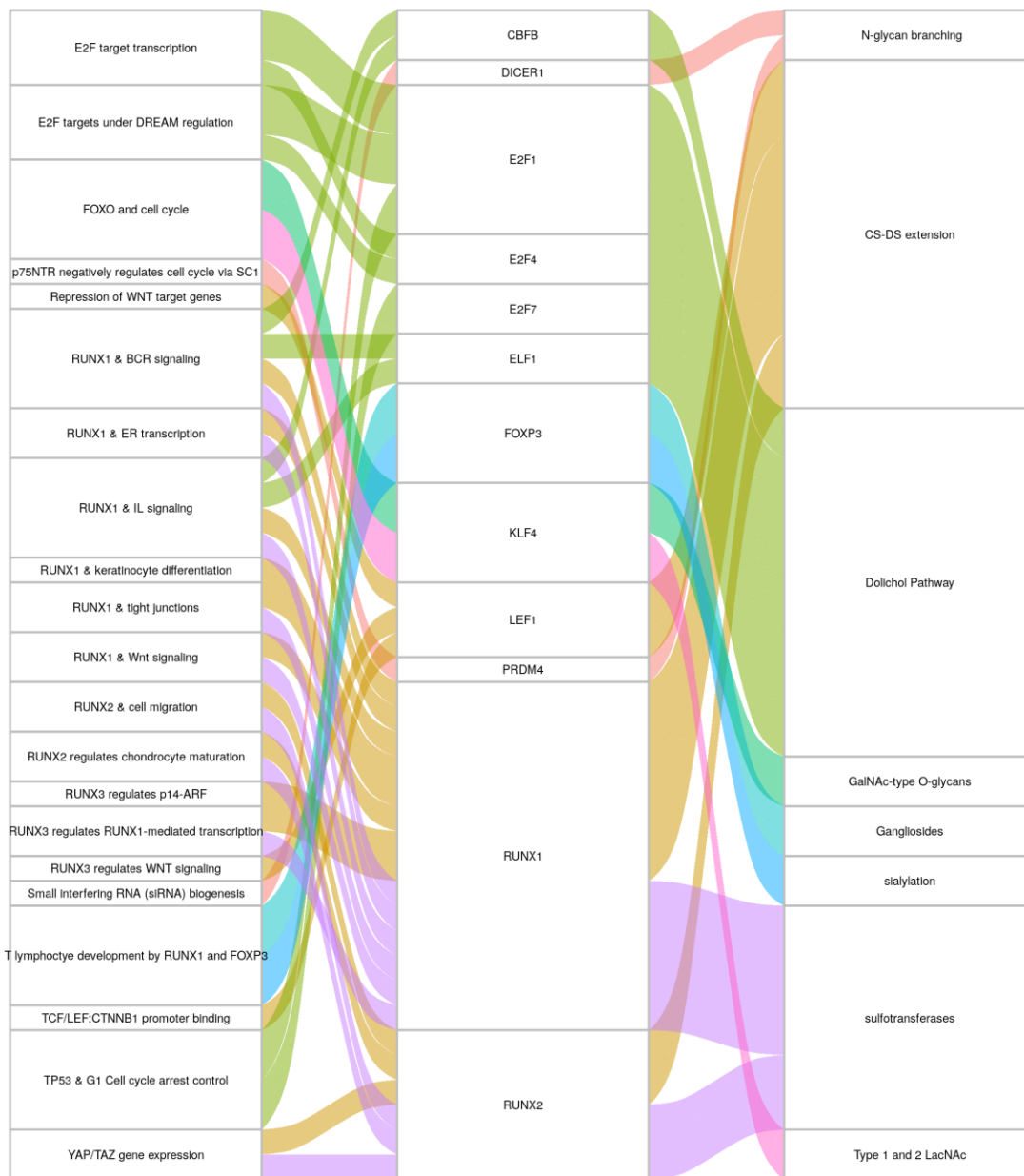
Signaling Pathways

Transcription Factors

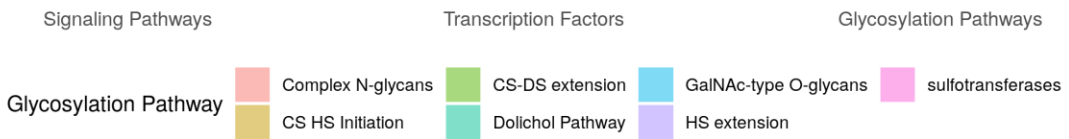
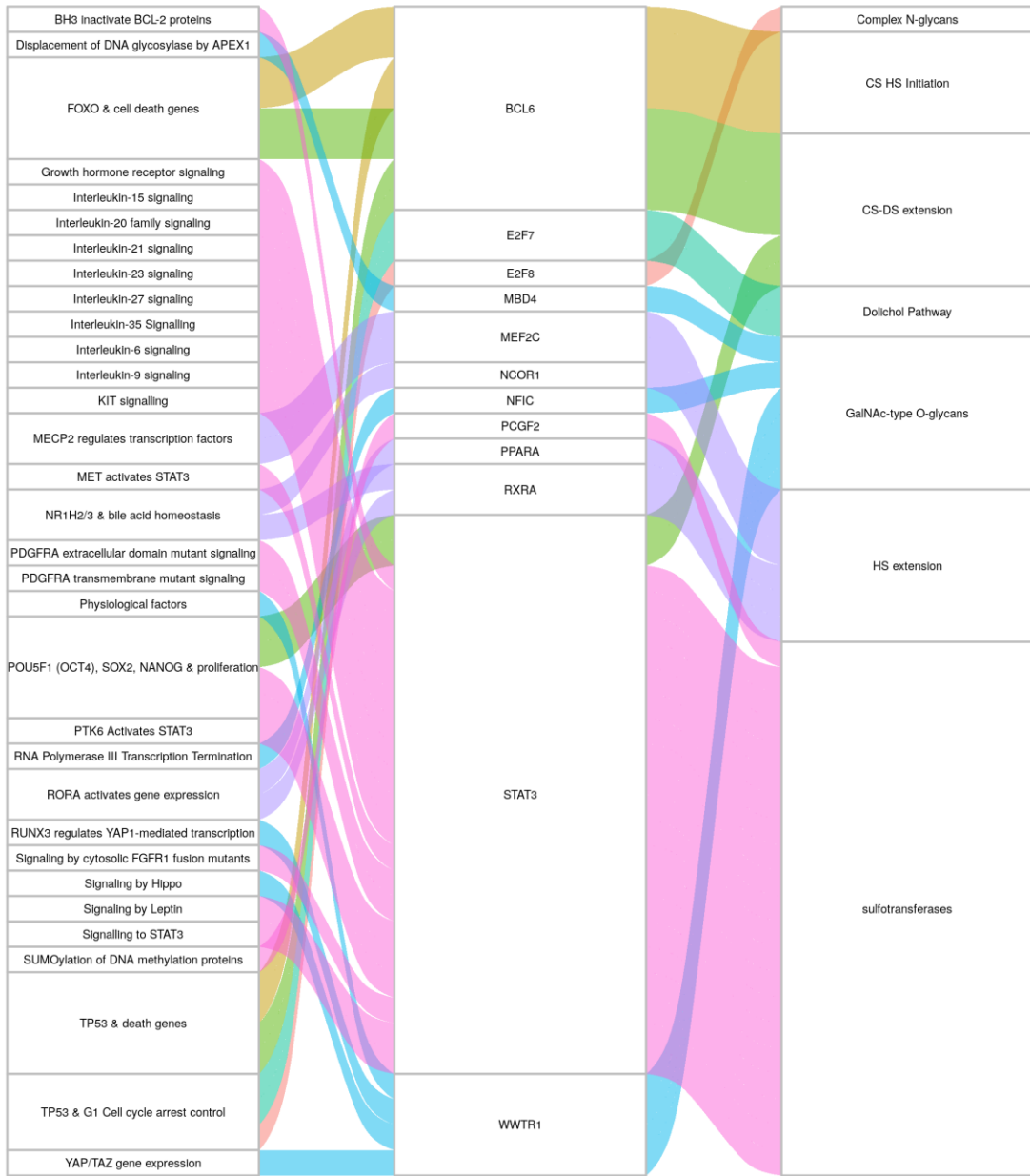
Glycosylation Pathways



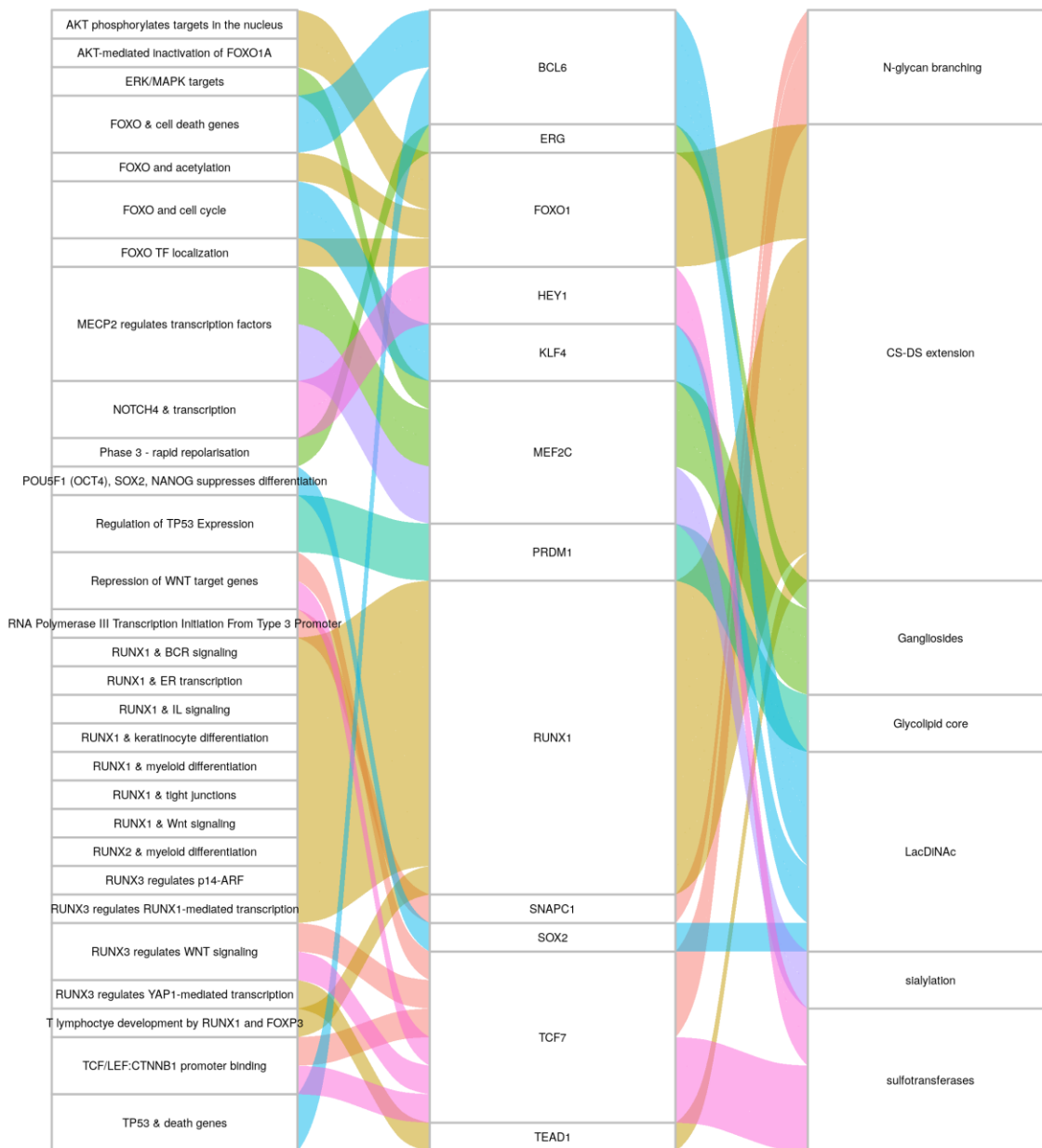
COAD_READ. Signaling to Glycosylation Pathway Matching



GBM. Signaling to Glycosylation Pathway Matching



HNSC Signaling to Glycosylation Pathway Matching



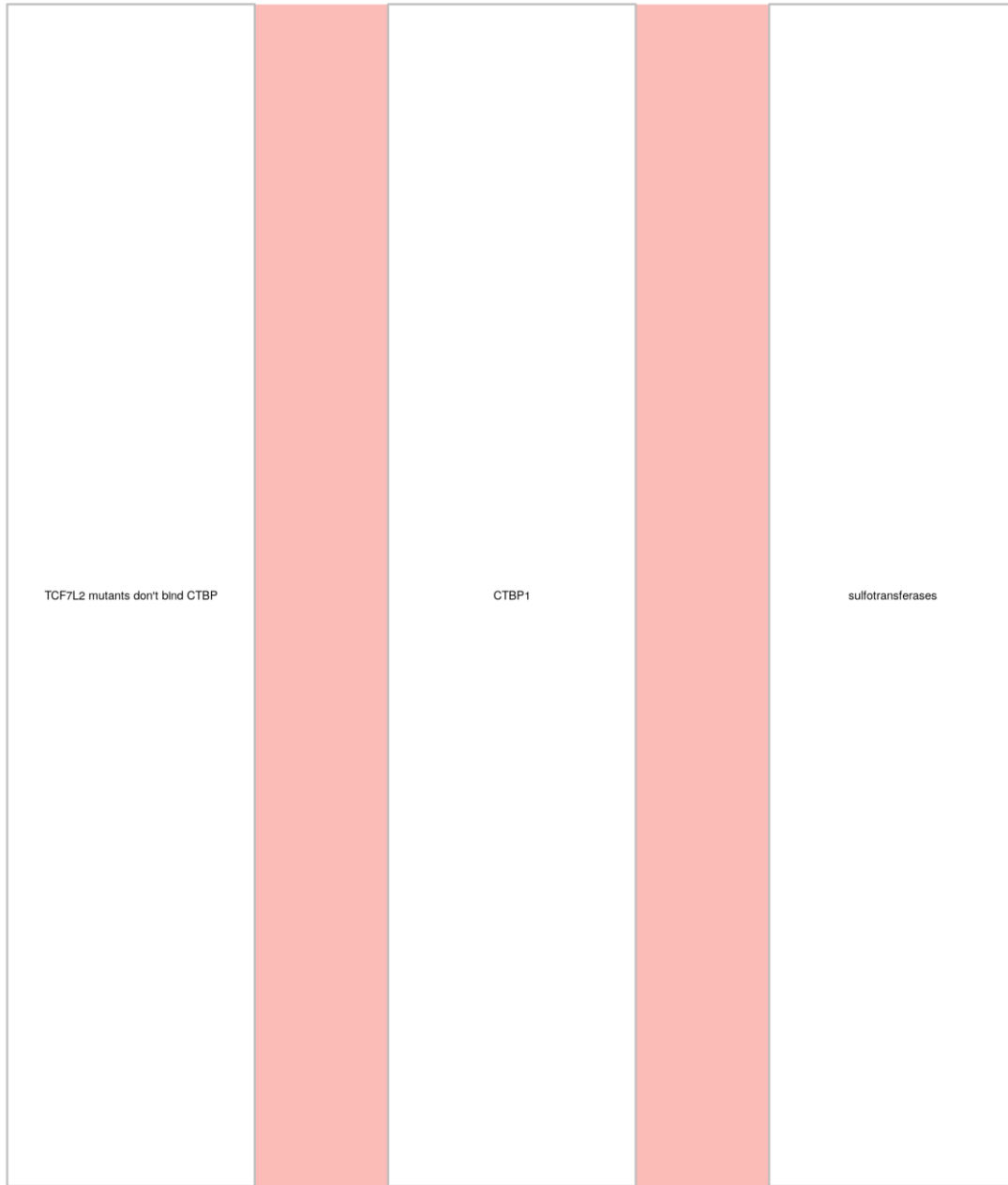
Signaling Pathways

Transcription Factors

Glycosylation Pathways



KICH Signaling to Glycosylation Pathway Matching



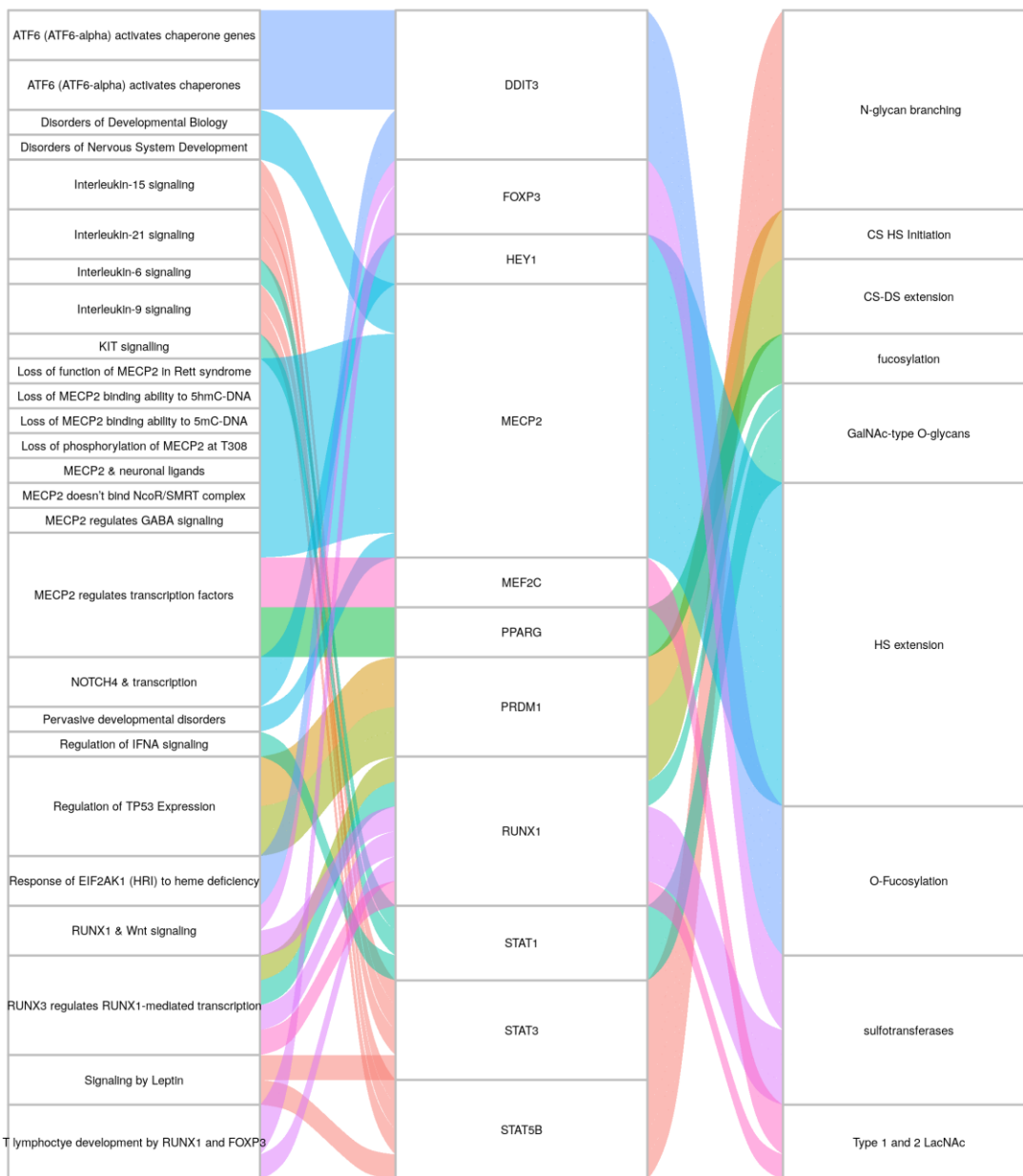
Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway  sulfotransferases

KIRC Signaling to Glycosylation Pathway Matching

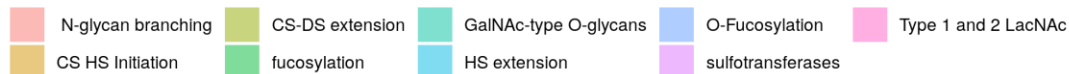


Signaling Pathways

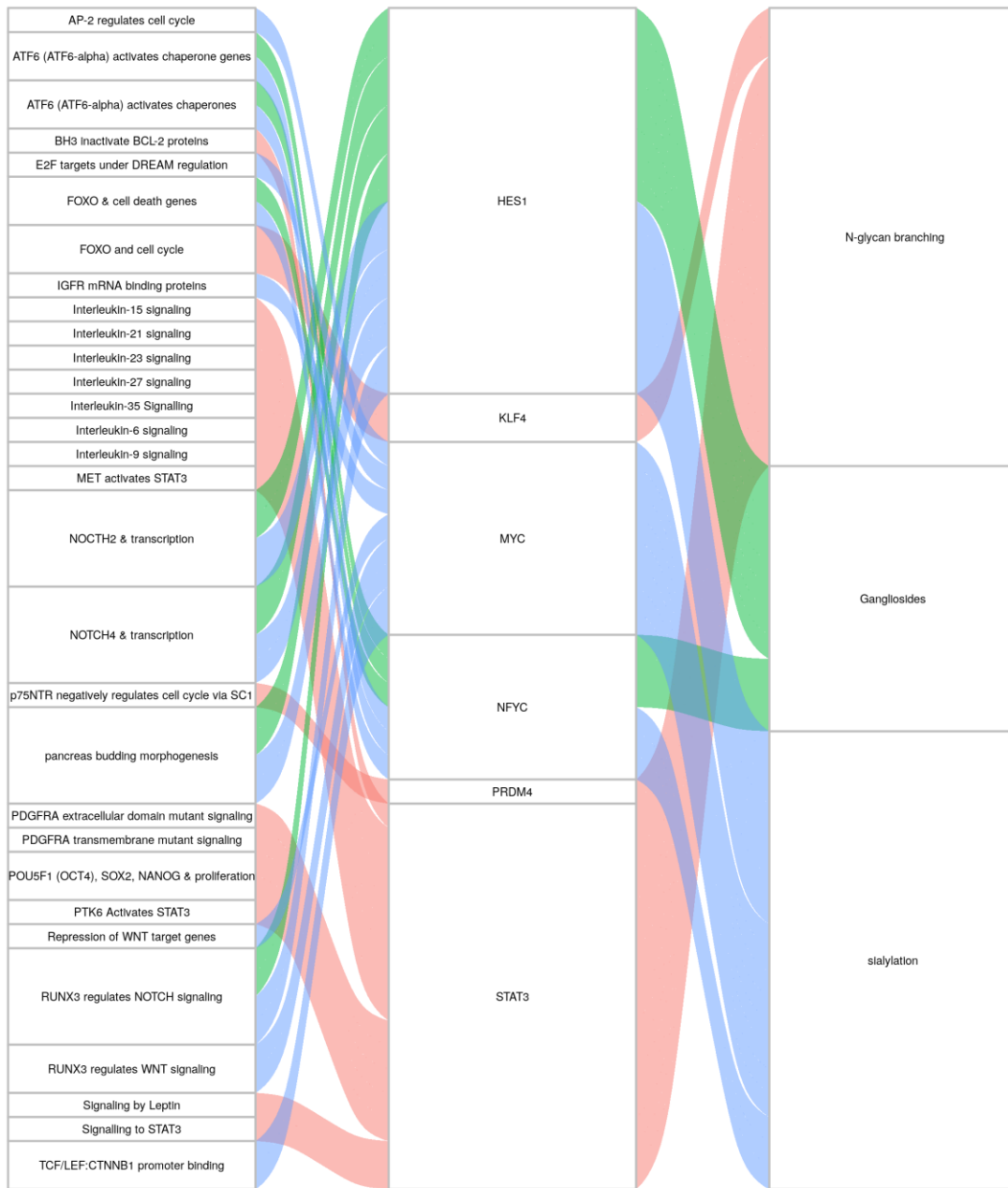
Transcription Factors

Glycosylation Pathways

Glycosylation Pathway



KIRP Signaling to Glycosylation Pathway Matching



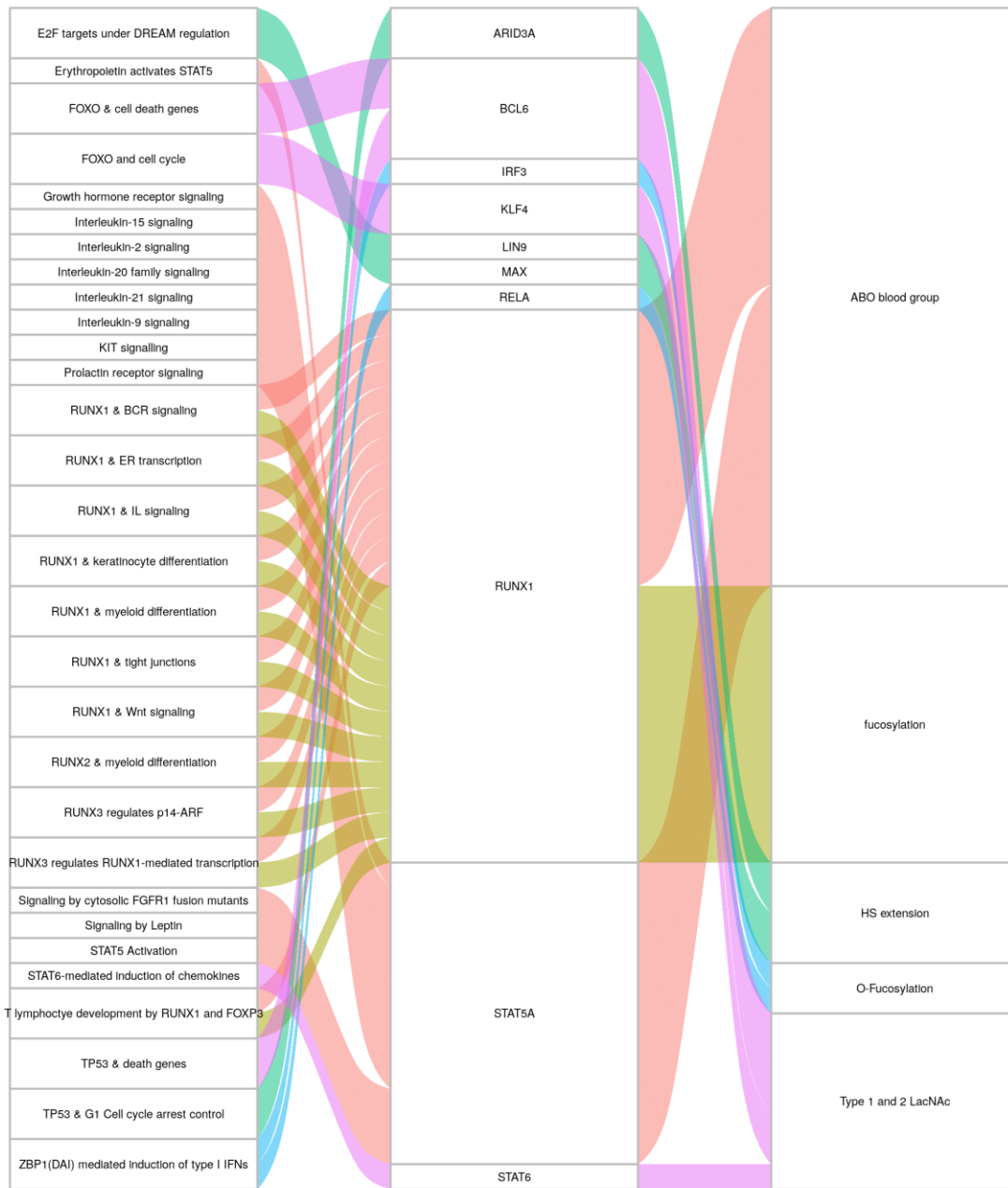
Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway ■ N-glycan branching ■ Gangliosides ■ sialylation

LAML Signaling to Glycosylation Pathway Matching



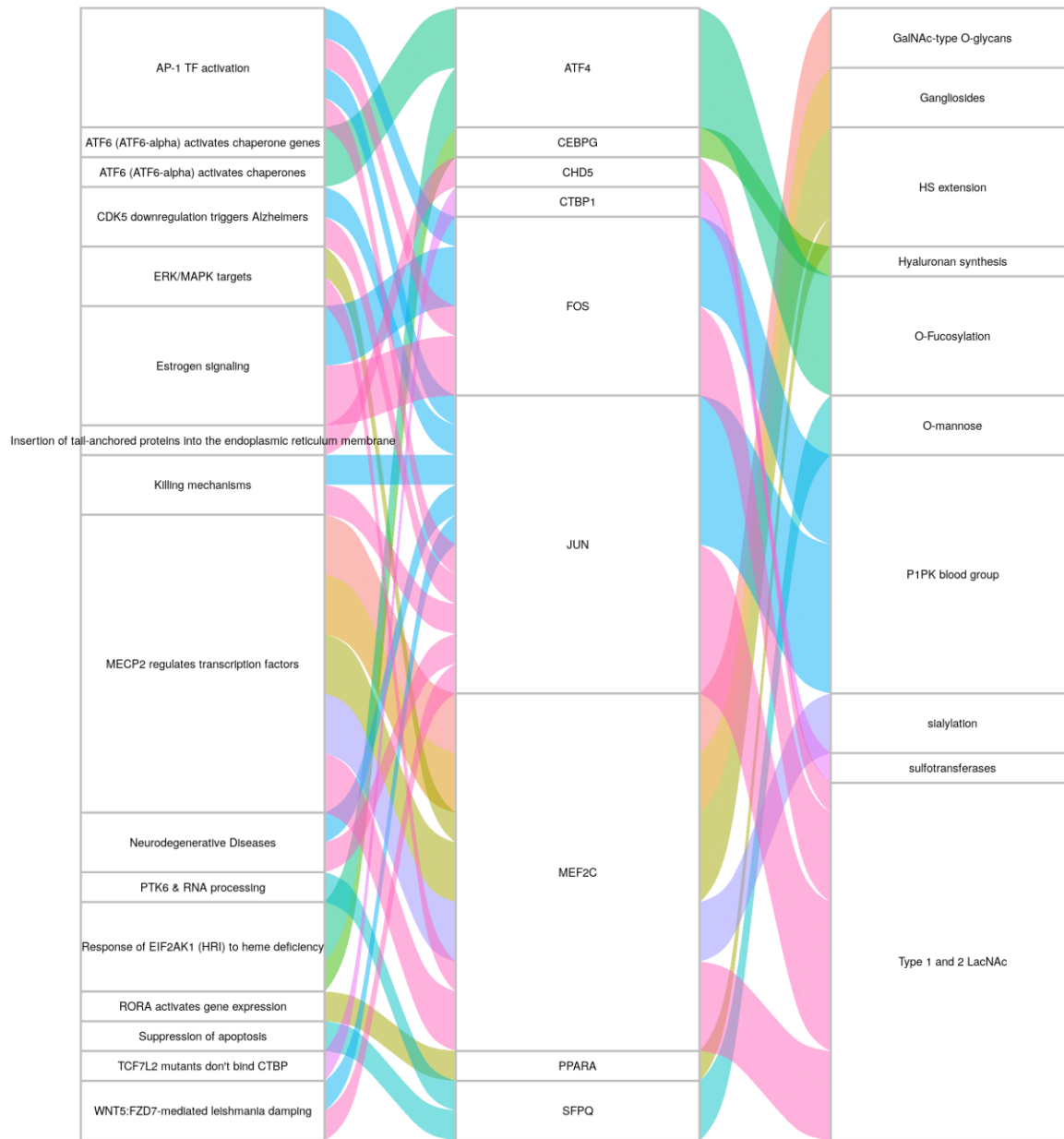
Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway ■ ABO blood group ■ fucosylation ■ HS extension ■ O-Fucosylation ■ Type 1 and 2 LacNAc

LGG Signaling to Glycosylation Pathway Matching



LIHC Signaling to Glycosylation Pathway Matching



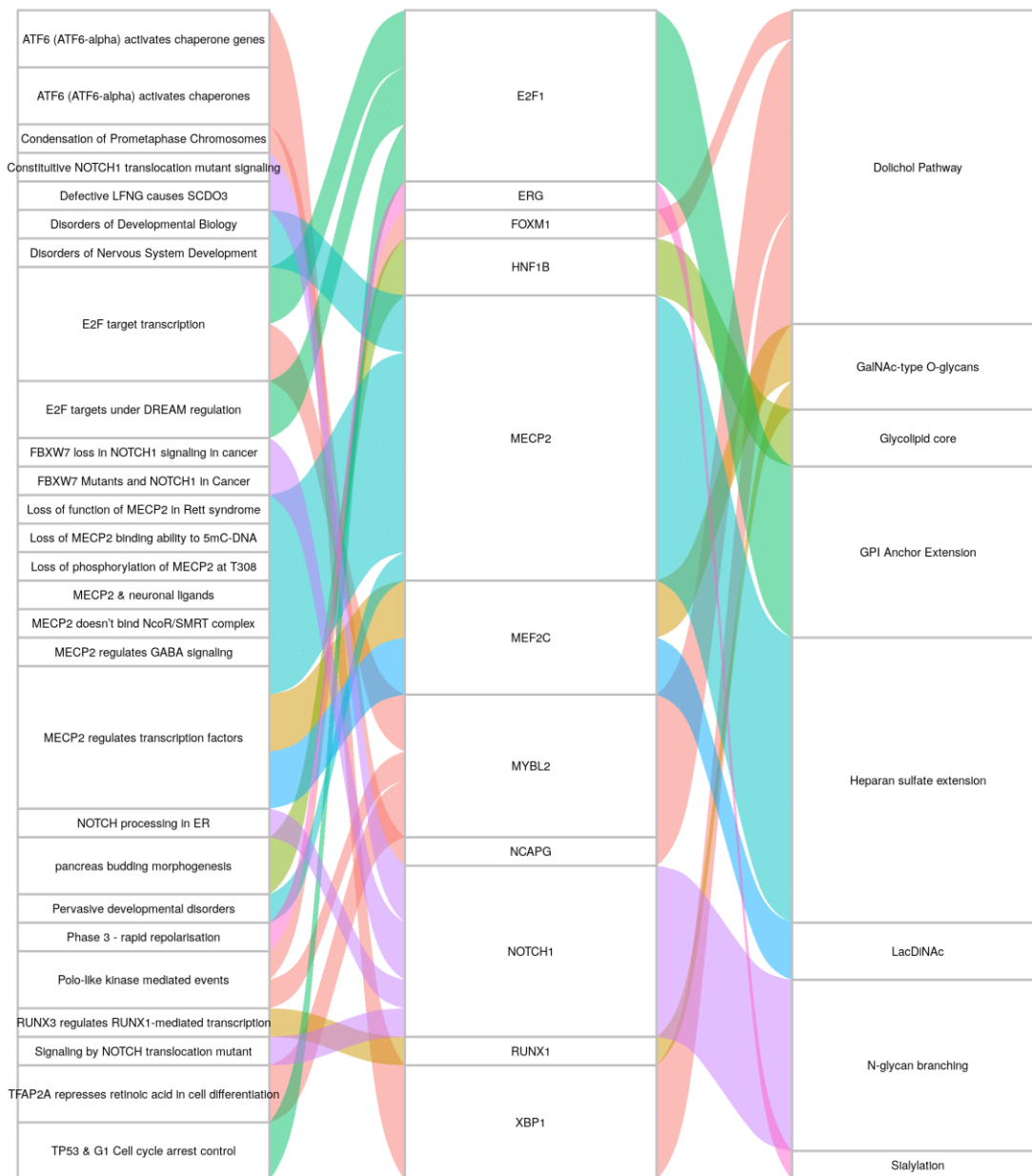
Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway ■ ABO blood group ■ CS-DS extension ■ fucosylation ■ sulfotransferases

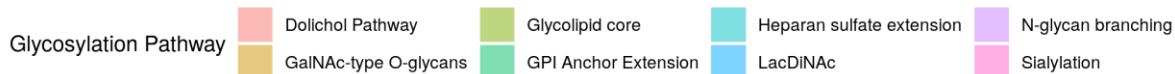
LUAD Signaling to Glycosylation Pathway Matching



Signaling Pathways

Transcription Factors

Glycosylation Pathways



LUSC Signaling to Glycosylation Pathway Matching



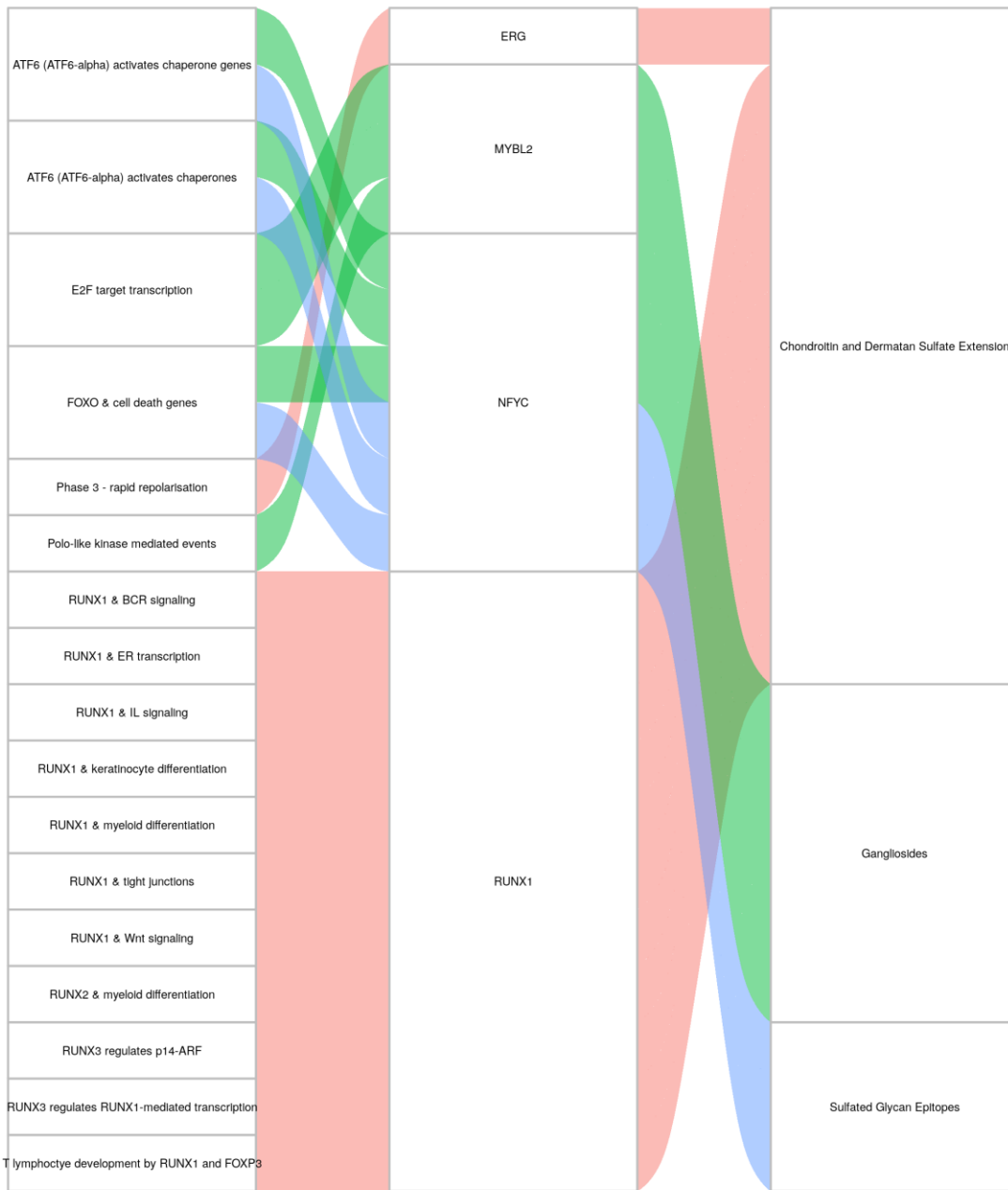
Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway ■ LacDINAc

MESO Signaling to Glycosylation Pathway Matching



Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway ■ Chondroitin and Dermatan Sulfate Extension ■ Gangliosides ■ Sulfated Glycan Epitopes

OV Signaling to Glycosylation Pathway Matching



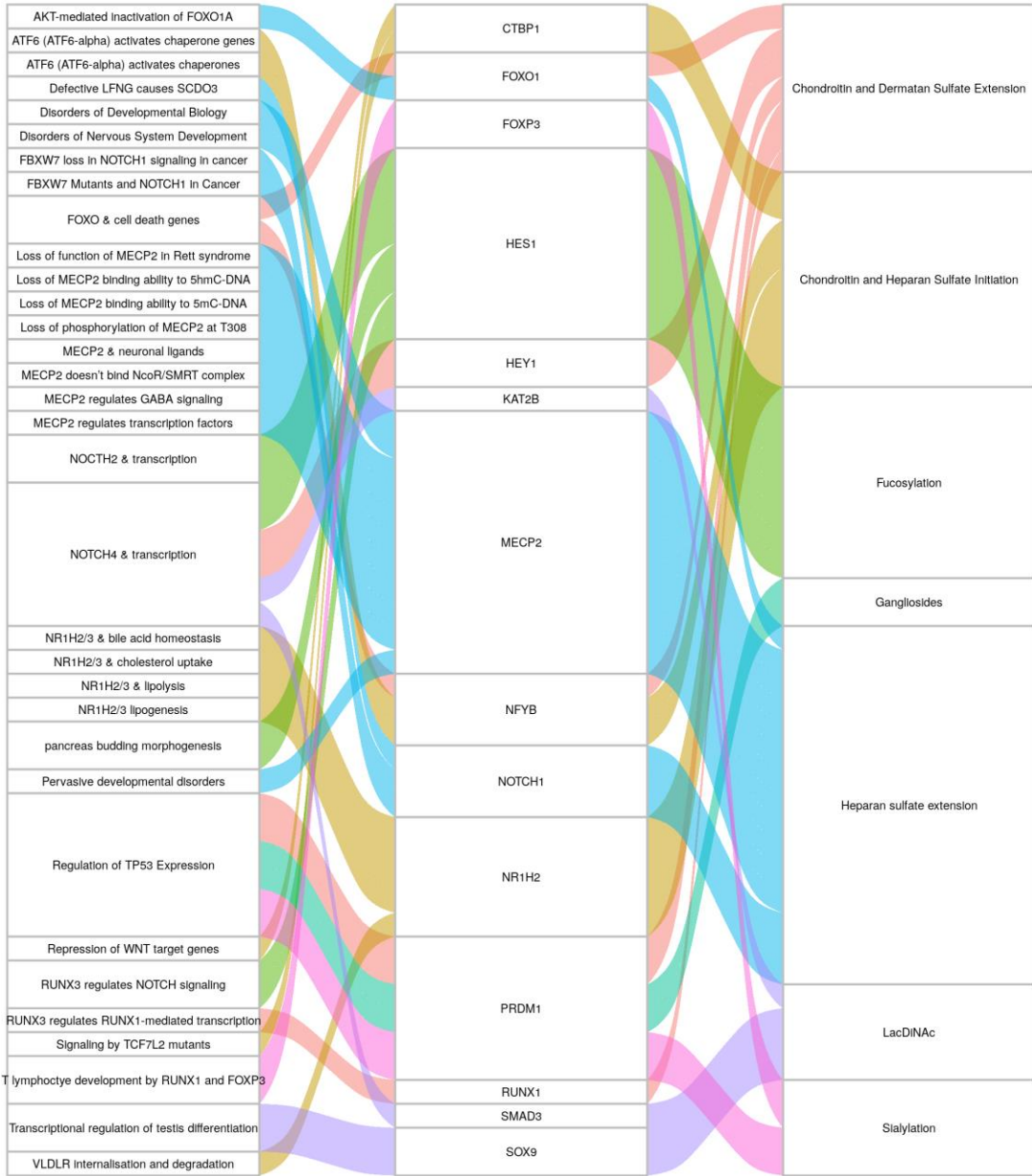
Signaling Pathways

Transcription Factors

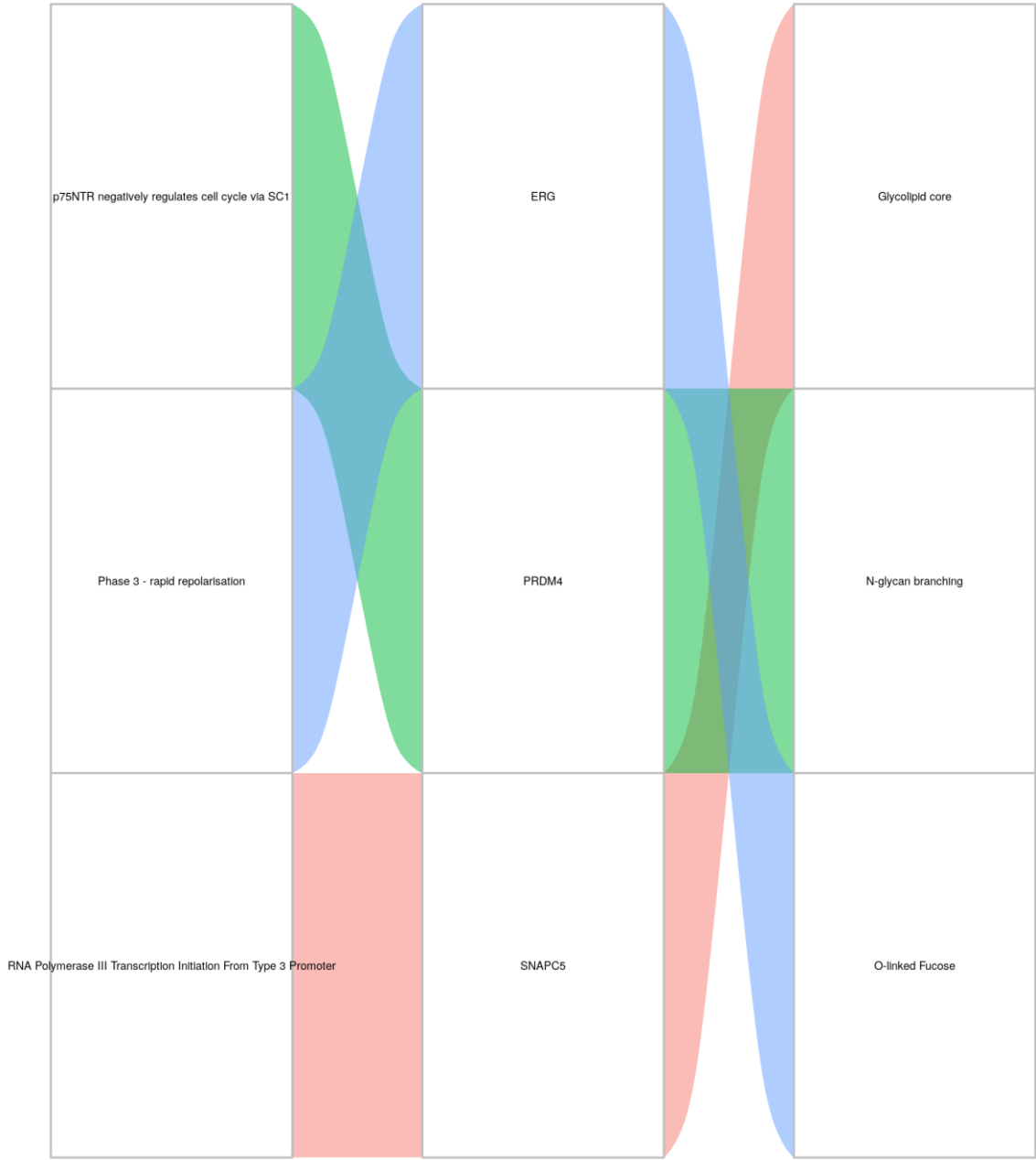
Glycosylation Pathways

Glycosylation Pathway ABO blood Group Synthesis

PAAD Signaling to Glycosylation Pathway Matching



PCPG Signaling to Glycosylation Pathway Matching



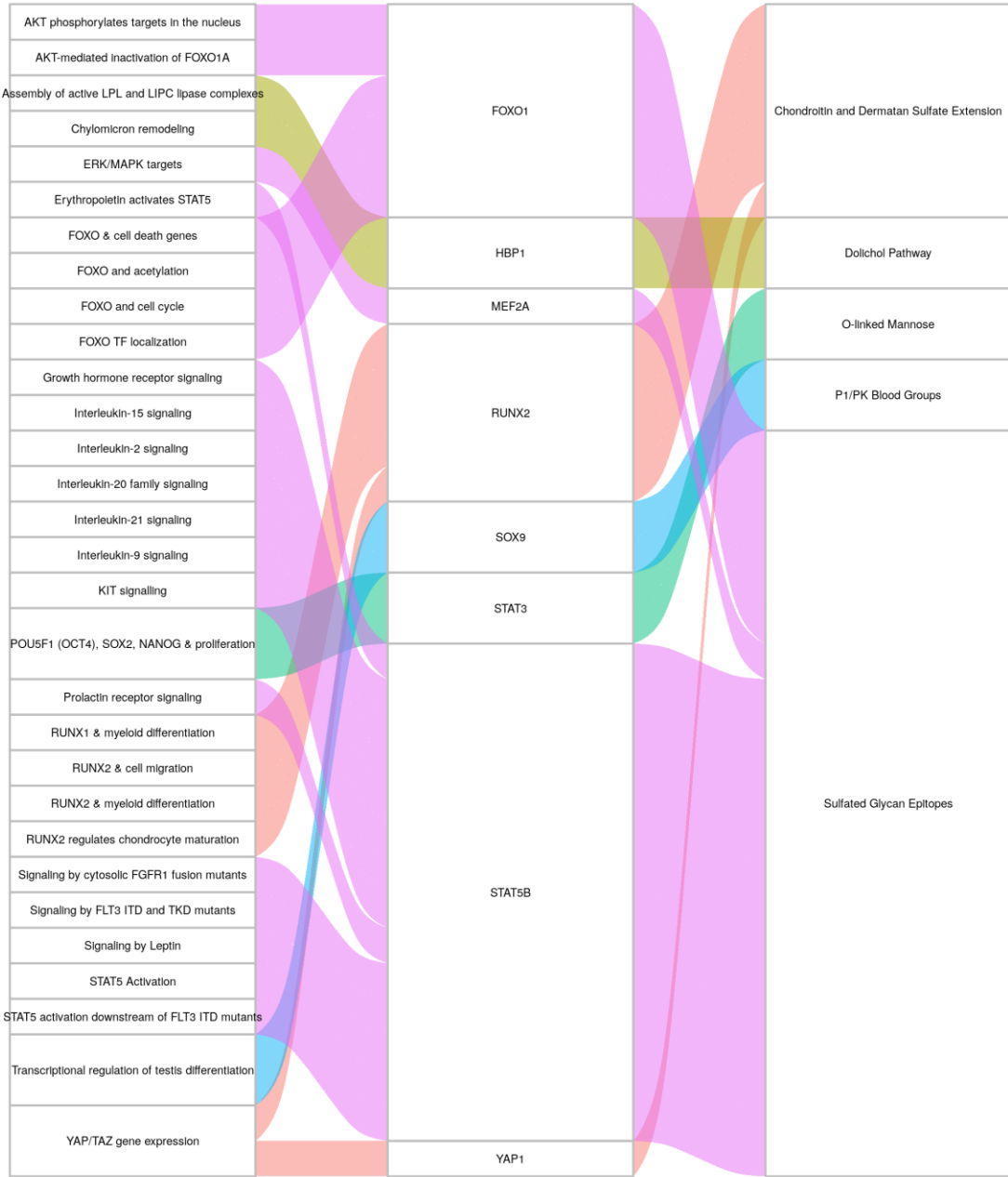
Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway ■ Glycolipid core ■ N-glycan branching ■ O-linked Fucose

PRAD Signaling to Glycosylation Pathway Matching



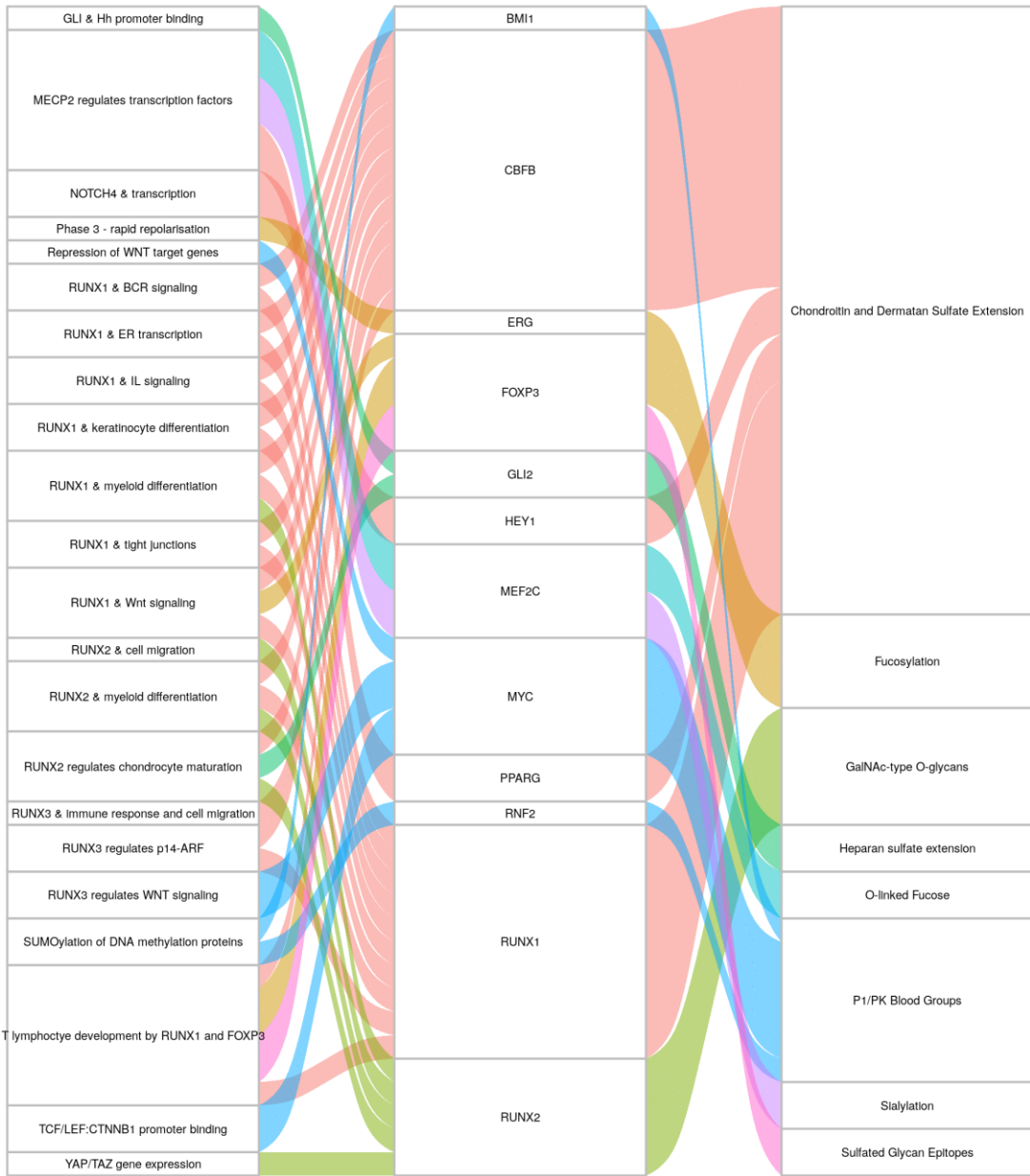
Signaling Pathways

Transcription Factors

Glycosylation Pathways

Chondroitin and Dermatan Sulfate Extension Dolichol Pathway O-linked Mannose P1/PK Blood Groups Sulfated Glyca

SARC Signaling to Glycosylation Pathway Matching



Signaling Pathways

Transcription Factors

Glycosylation Pathways



SKCM Signaling to Glycosylation Pathway Matching



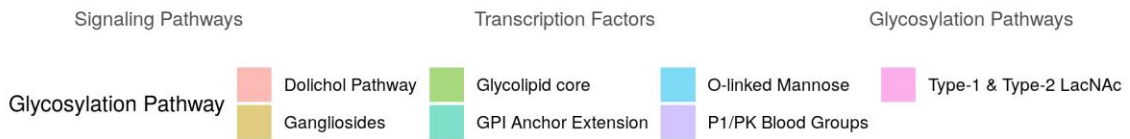
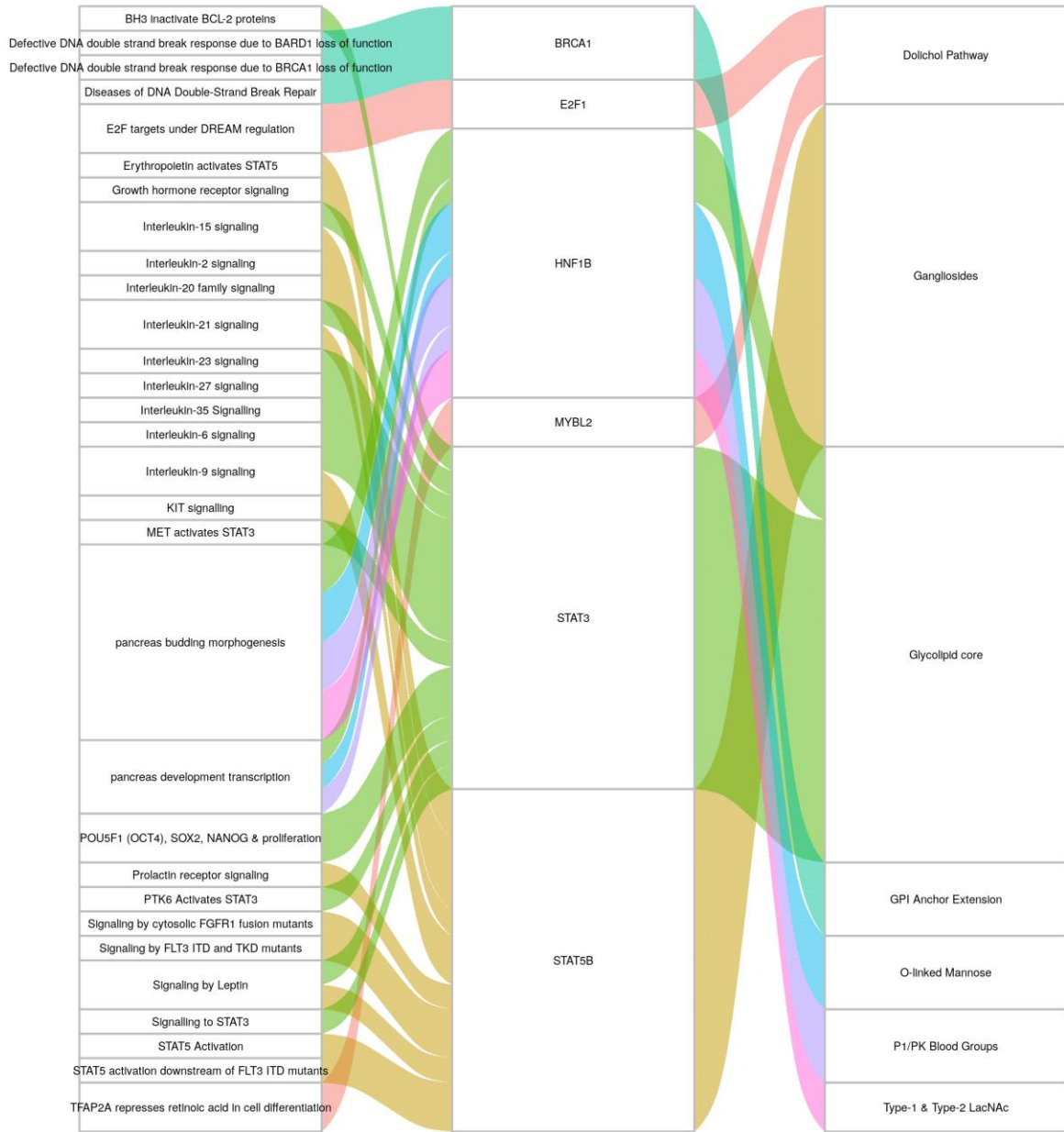
Signaling Pathways

Transcription Factors

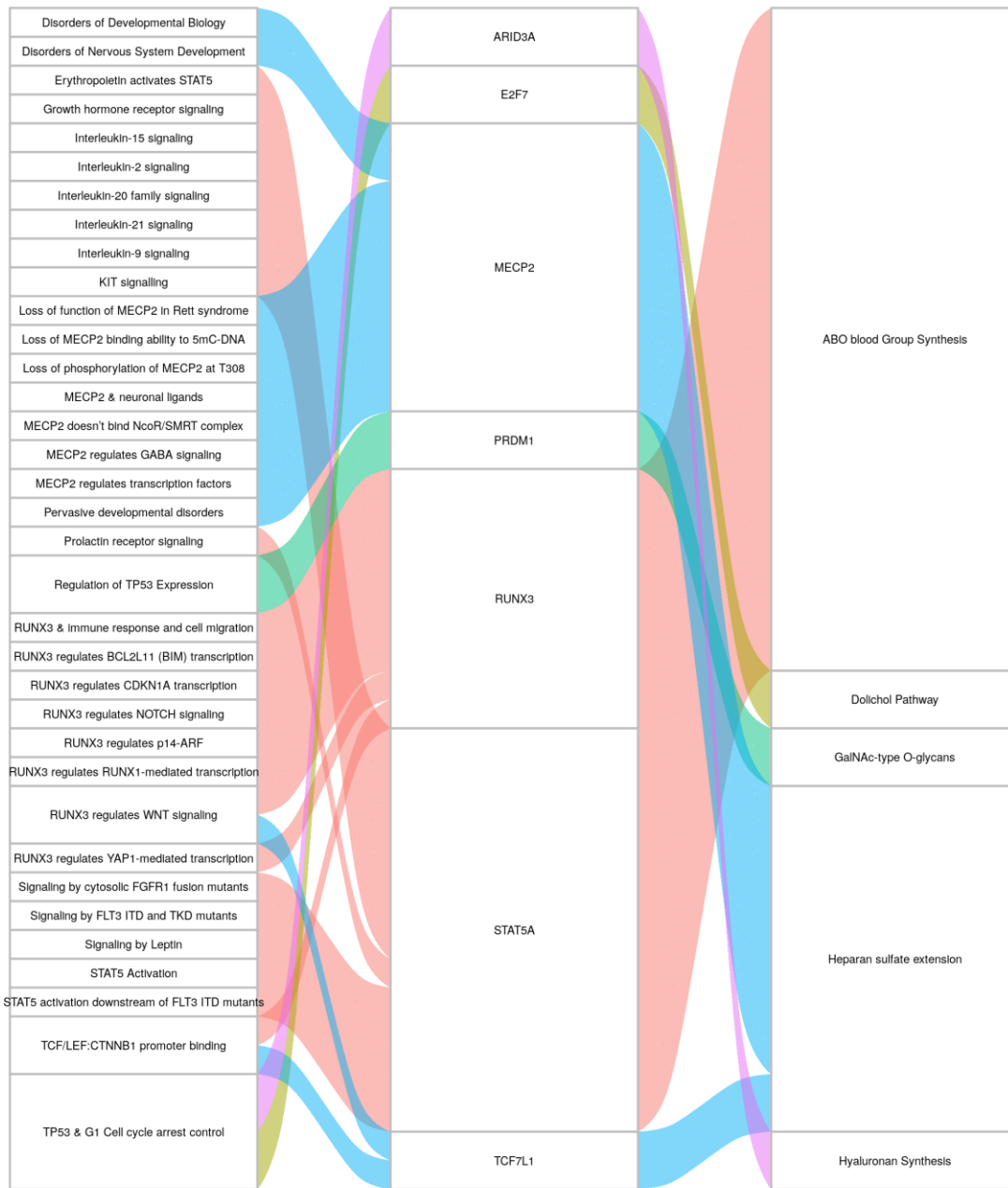
Glycosylation Pathways

Glycosylation Pathway ■ Gangliosides ■ GPI Anchor Extension

STES. Signaling to Glycosylation Pathway Matching



TGCT Signaling to Glycosylation Pathway Matching



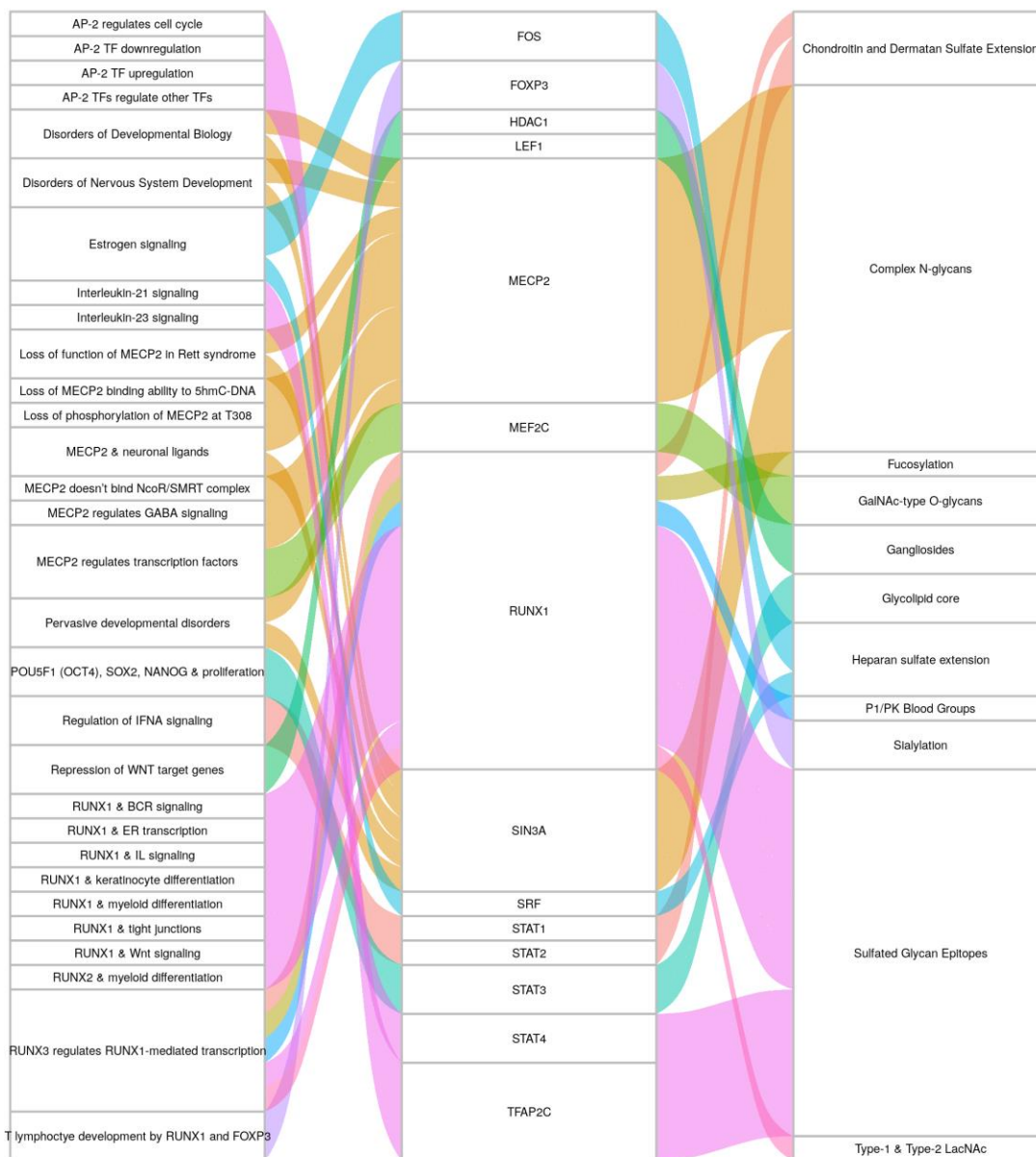
Signaling Pathways

Transcription Factors

Glycosylation Pathways

osylation Pathway ■ ABO blood Group Synthesis ■ Dolichol Pathway ■ GalNAc-type O-glycans ■ Heparan sulfate extension ■ Hyaluronan Syn

THCA Signaling to Glycosylation Pathway Matching



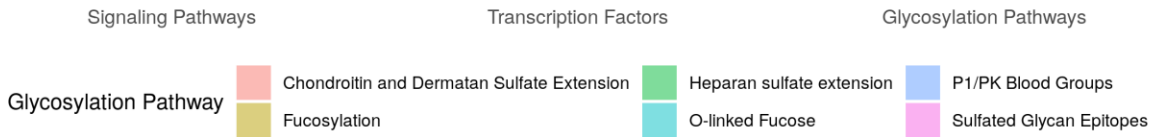
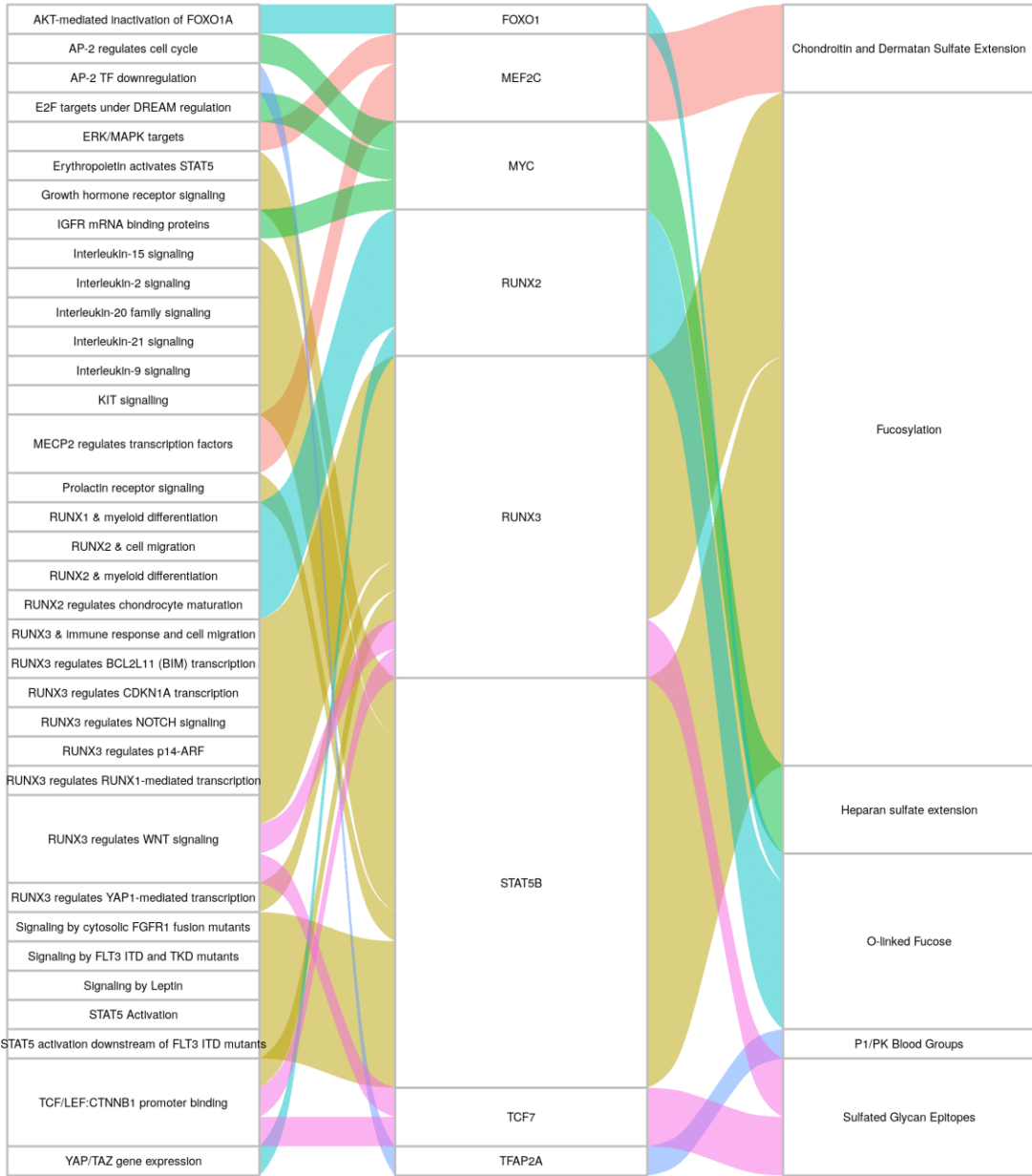
Signaling Pathways

Transcription Factors

Glycosylation Pathways



THYM Signaling to Glycosylation Pathway Matching



UCEC Signaling to Glycosylation Pathway Matching



Signaling Pathways

Transcription Factors

Glycosylation Pathways



UVM Signaling to Glycosylation Pathway Matching



Signaling Pathways

Transcription Factors

Glycosylation Pathways

Glycosylation Pathway ■ Glycolipid core ■ N-glycan branching ■ Type-1 & Type-2 LacNAc