Long-read RNA-seq transcriptomic analysis with TALON and Swan
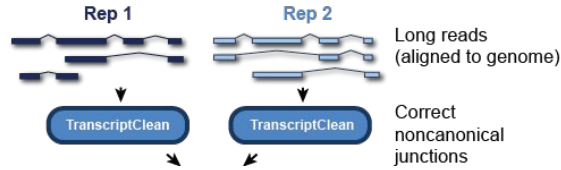
Fairlie Reese
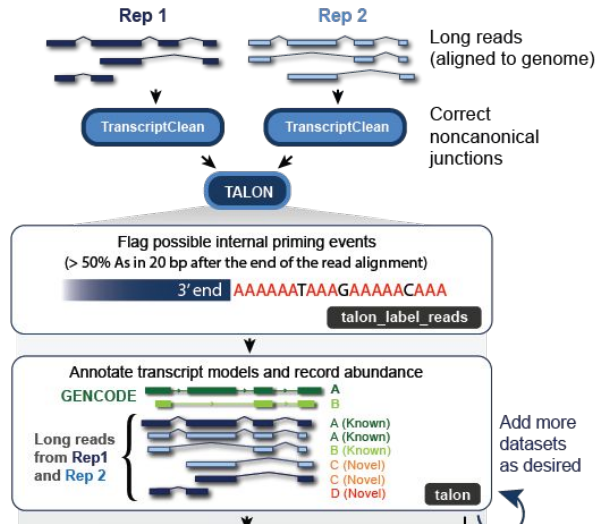UC Davis IsoSeq Workshop
September 1, 2021

TALON identifies, quantifies, and filters full-length transcript isoforms

TALON identifies, quantifies, and filters full-length transcript isoforms
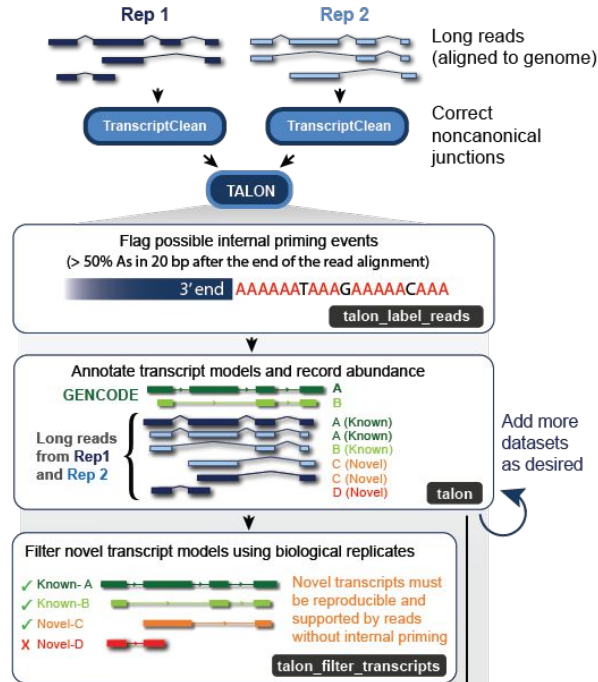

- Preprocessing with TranscriptClean corrects long-read sequencing artifacts

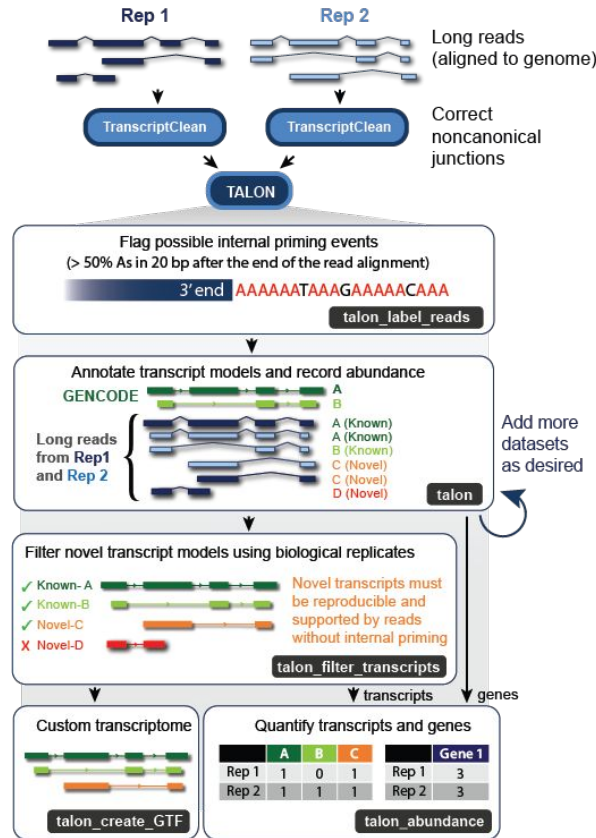# TALON identifies, quantifies, and filters full-length transcript isoforms

- Preprocessing with TranscriptClean corrects long-read sequencing artifacts

- TALON annotates reads to transcript models

- Assigns novel models an informative novelty type

Wyman, Balderrama-Gutierrez and Reese et. al., bioRxiv, 2020

TALON identifies, quantifies, and filters full-length transcript isoforms
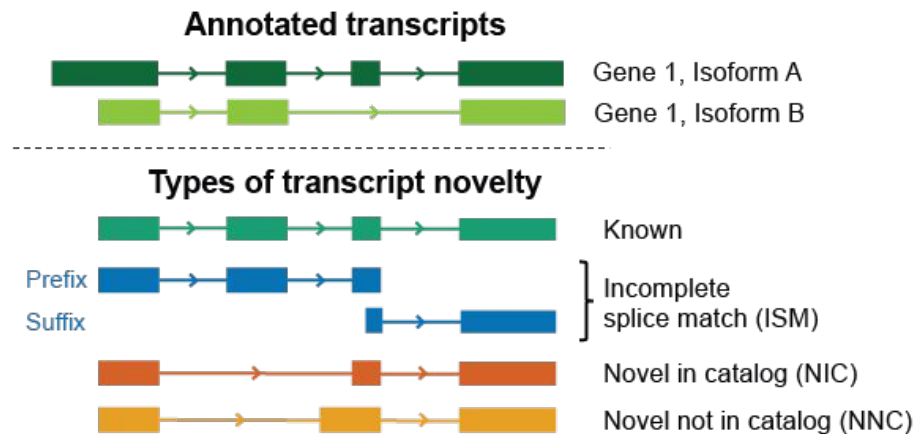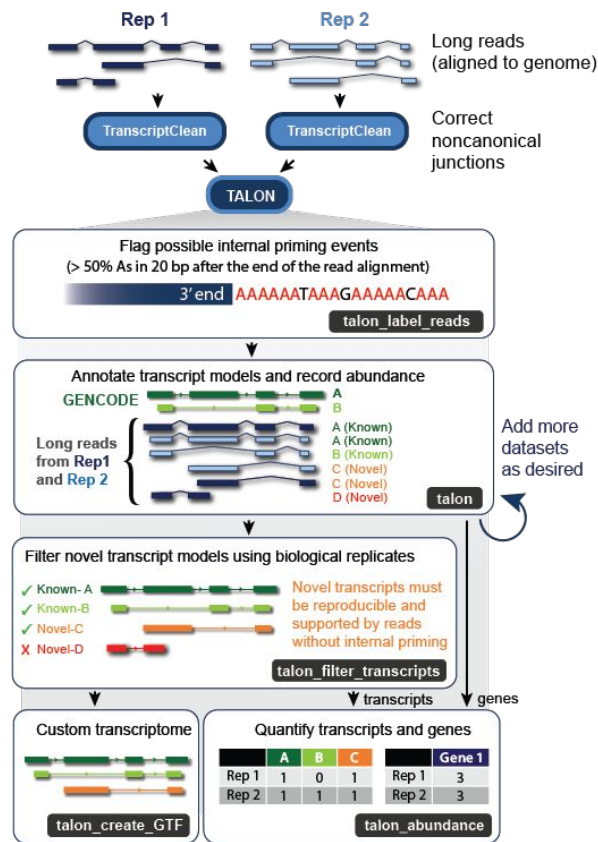

- Preprocessing with TranscriptClean corrects long-read sequencing artifacts

- TALON annotates reads to transcript models

- Assigns novel models an informative novelty type

- Filters novel transcript models for reproducibility and internal priming

Wyman, Balderrama-Gutierrez and Reese et. al., bioRxiv, 2020

TALON identifies, quantifies, and filters full-length transcript isoforms


- Preprocessing with TranscriptClean corrects long-read sequencing artifacts

- TALON annotates reads to transcript models

- Assigns novel models an informative novelty type

- Filters novel transcript models for reproducibility and internal priming

- Quantifies transcript abundances

Wyman, Balderrama-Gutierrez and Reese et. al., bioRxiv, 2020

TALON identifies, quantifies, and filters full-length transcript isoforms

Wyman, Balderrama-Gutierrez and Reese et. al., bioRxiv, 2020

Transcript isoforms differences are difficult to visually interpret and analyze

- Difficult to tell what the difference between any two transcript models is

- Current transcriptome analysis platforms do not prioritize isoform-level differences from long-read data

- How do we compare long-read transcriptomes between samples?

Mouse *Mapt* transcripts
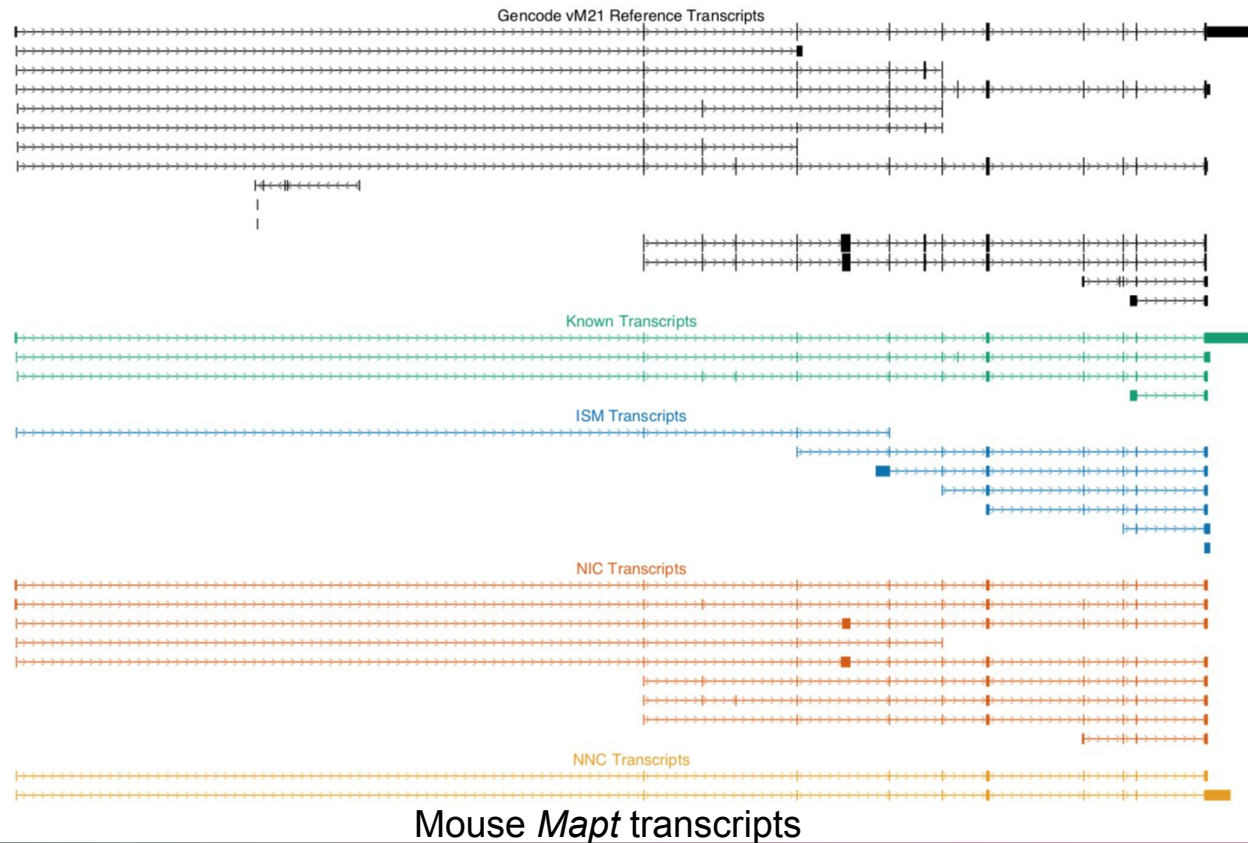
Swan is a Python library designed for the analysis and visualization of full-length transcriptomes

- Differential expression tests on the gene and transcript level

- Isoform/TSS/TES switching detection

- Novel exon skipping and intron retention event detection

- Unique and interpretable visualizations and reports with a robust set of options

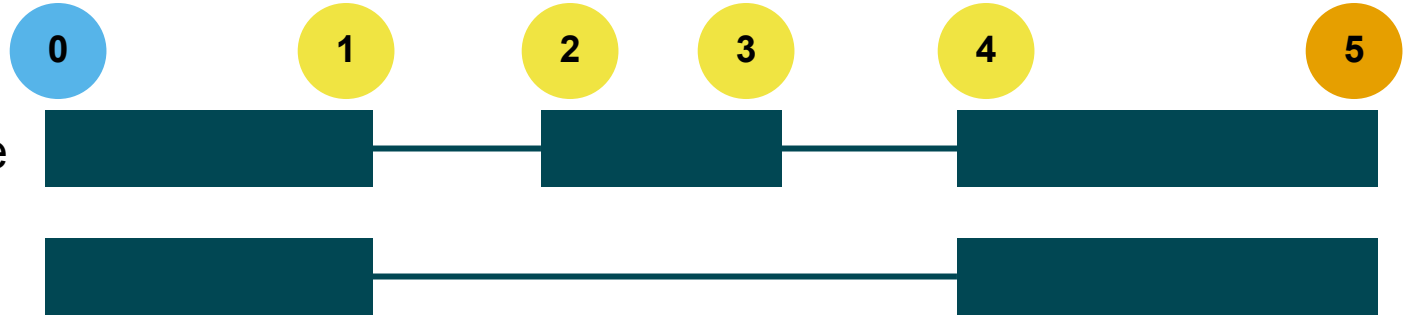- Compatibility with single-cell data
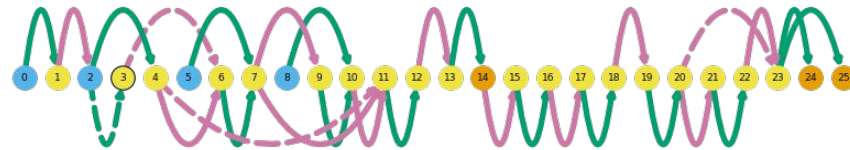
Swan represents transcript models as a series of splice sites and intronic or exonic connections between them
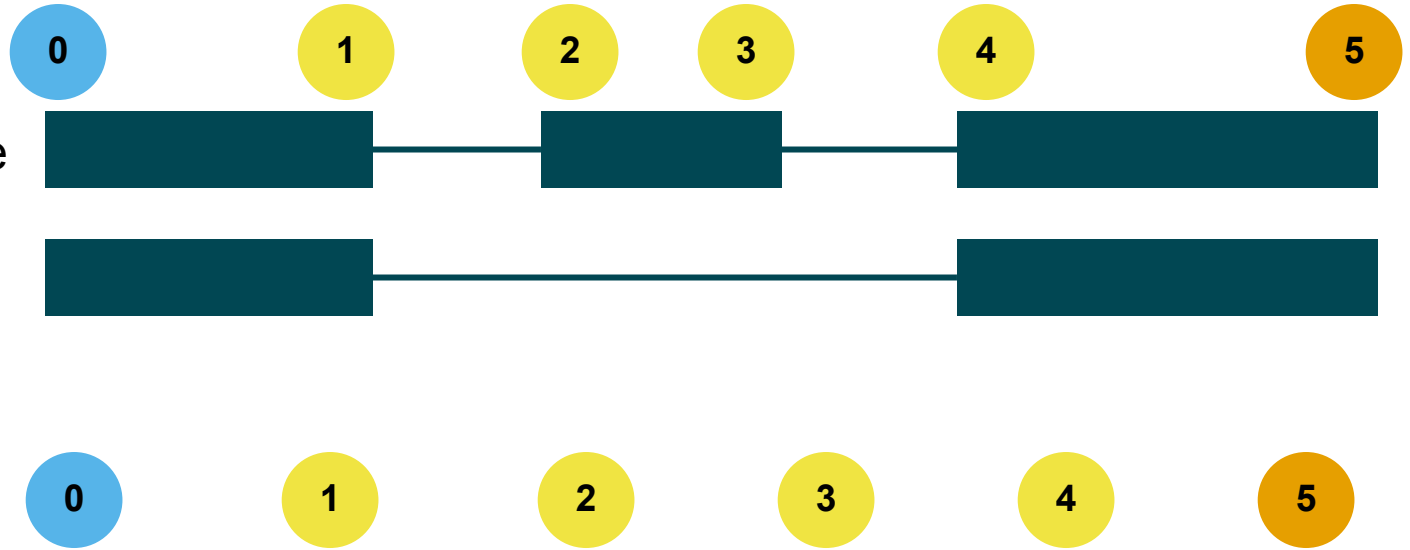

TALON DB or GTF

Swan represents transcript models as a series of splice sites and intronic or exonic connections between them

- Assign each unique splice site to a node

Swan represents transcript models as a series of splice sites and intronic or exonic connections between them
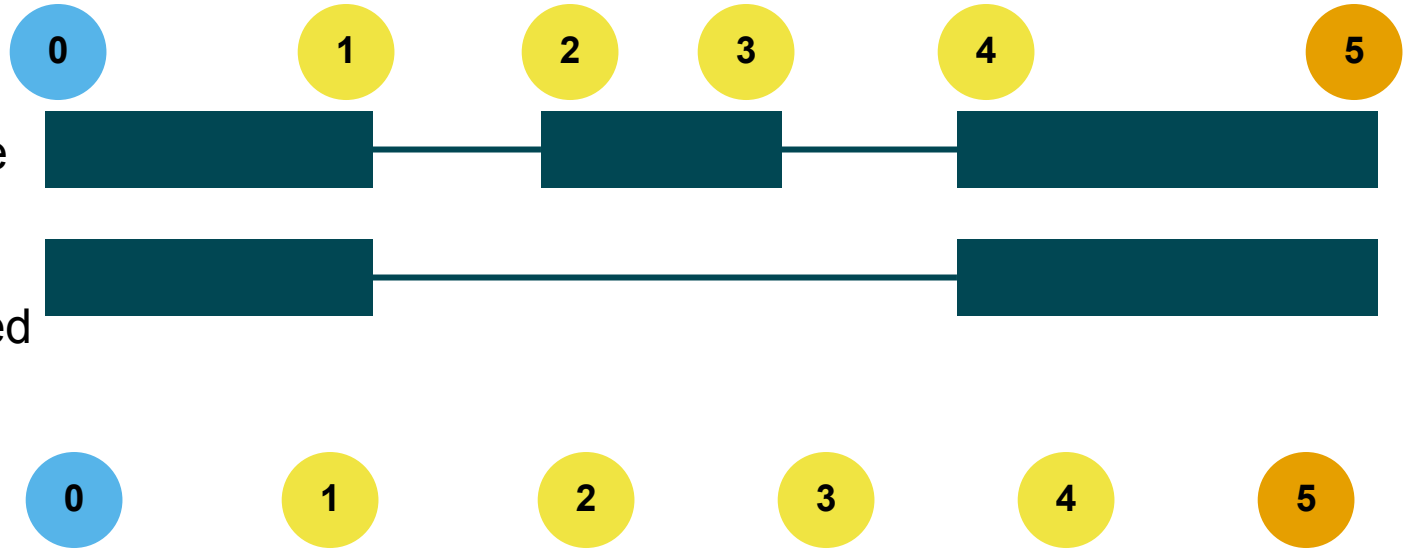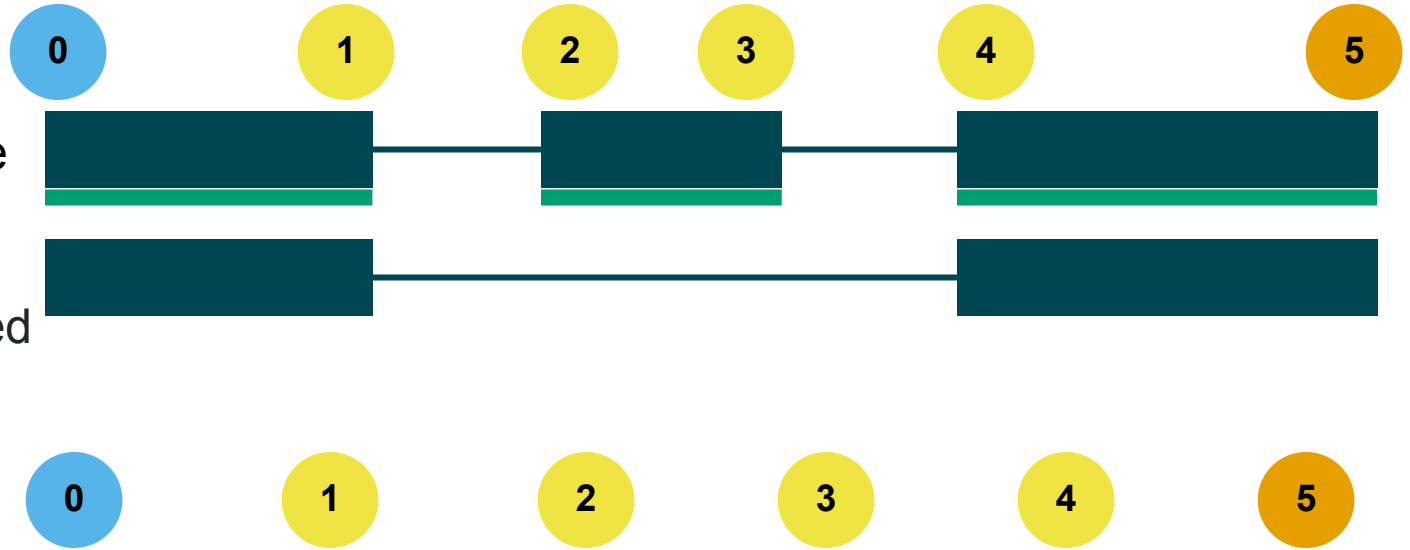
- Assign each unique splice site to a node
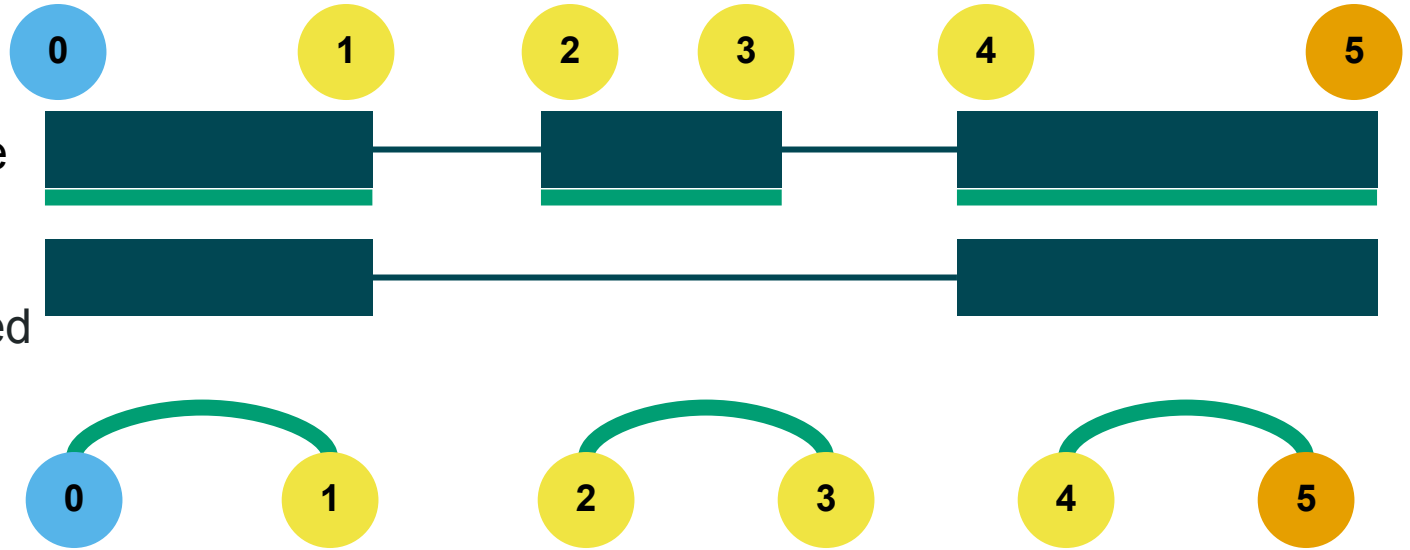

Swan representation

Swan represents transcript models as a series of splice sites and intronic or exonic connections between them

- Assign each unique splice site to a node

- Nodes are colored by their role in transcript model (**TSS**, **TES**)
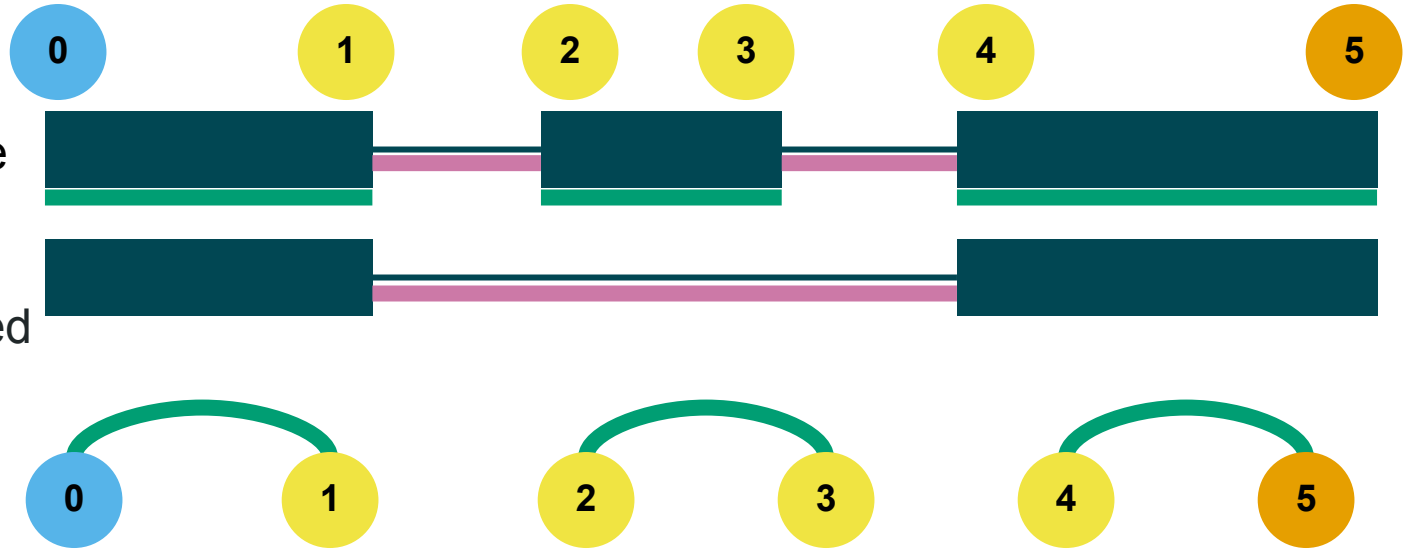
Swan represents transcript models as a series of splice sites and intronic or exonic connections between them
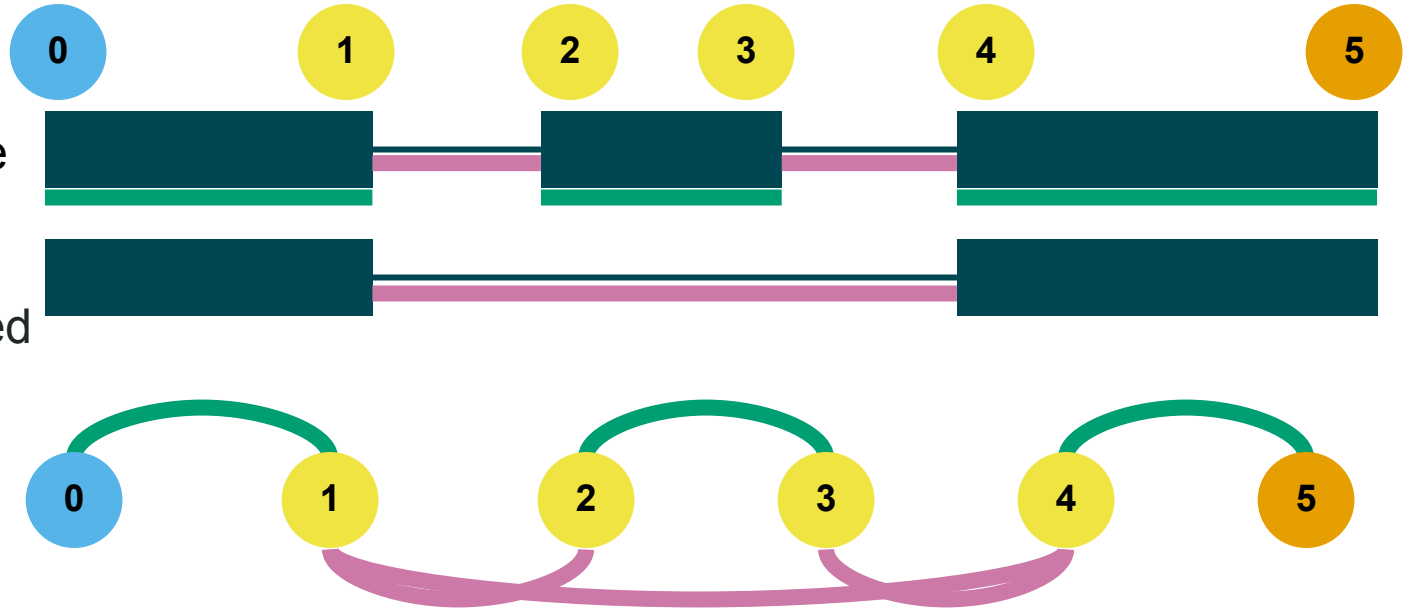

- Assign each unique splice site to a node

- Nodes are colored by their role in transcript model (**TSS**, **TES**)

- Connect nodes (splice sites) by **intronic** or **exonic** edges

Swan represents transcript models as a series of splice sites and intronic or exonic connections between them

- Assign each unique splice site to a node

- Nodes are colored by their role in transcript model (**TSS**, **TES**)

- Connect nodes (splice sites) by **intronic** or **exonic** edges

Swan represents transcript models as a series of splice sites and intronic or exonic connections between them

- Assign each unique splice site to a node

- Nodes are colored by their role in transcript model (**TSS**, **TES**)

- Connect nodes (splice sites) by **intronic** or **exonic** edges

Swan represents transcript models as a series of splice sites and intronic or exonic connections between them

- Assign each unique splice site to a node

- Nodes are colored by their role in transcript model (**TSS**, **TES**)

- Connect nodes (splice sites) by **intronic** or **exonic** edges

Swan visualizations: Gene summary plots and transcript path plots
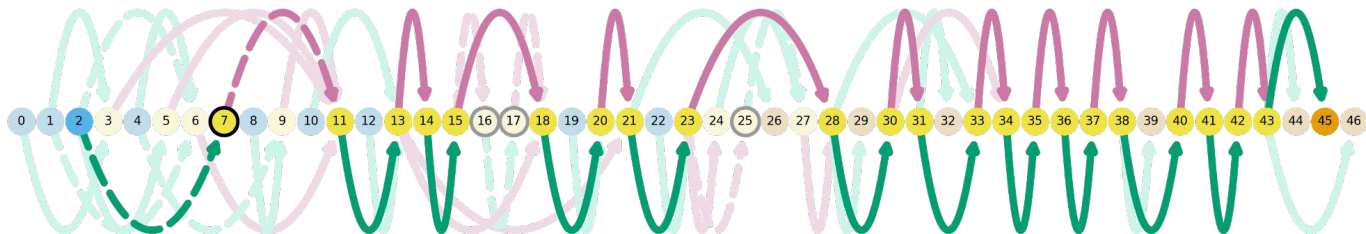
nodes: TSS TES internal

edges: exon intron

dashed edges: novel intron or exon

outlined nodes: novel splice site, TSS, or TES
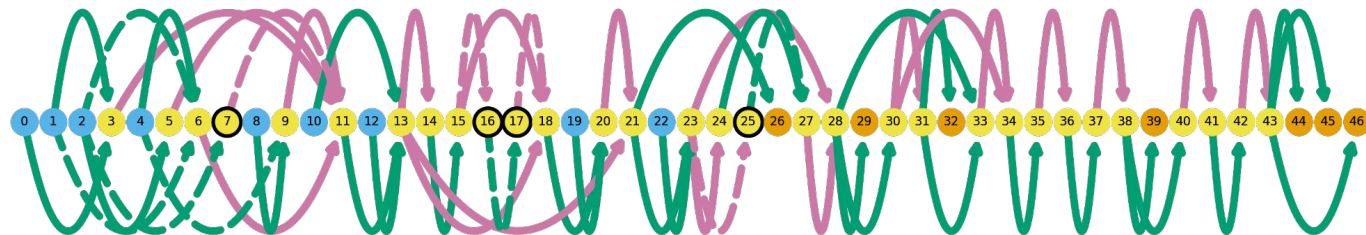

Tfdpi1 Gene Summary Plot

Tfdpi1 Transcript Path Plot

Data and data analysis overview

- Publicly available PacBio RNA-seq data

- C2C12 skeletal muscle cell line model of myogenesis
  - 0hr - undifferentiated, mononucleated myoblasts
  - 72 hr - differentiated, mutlinucleated myotubes

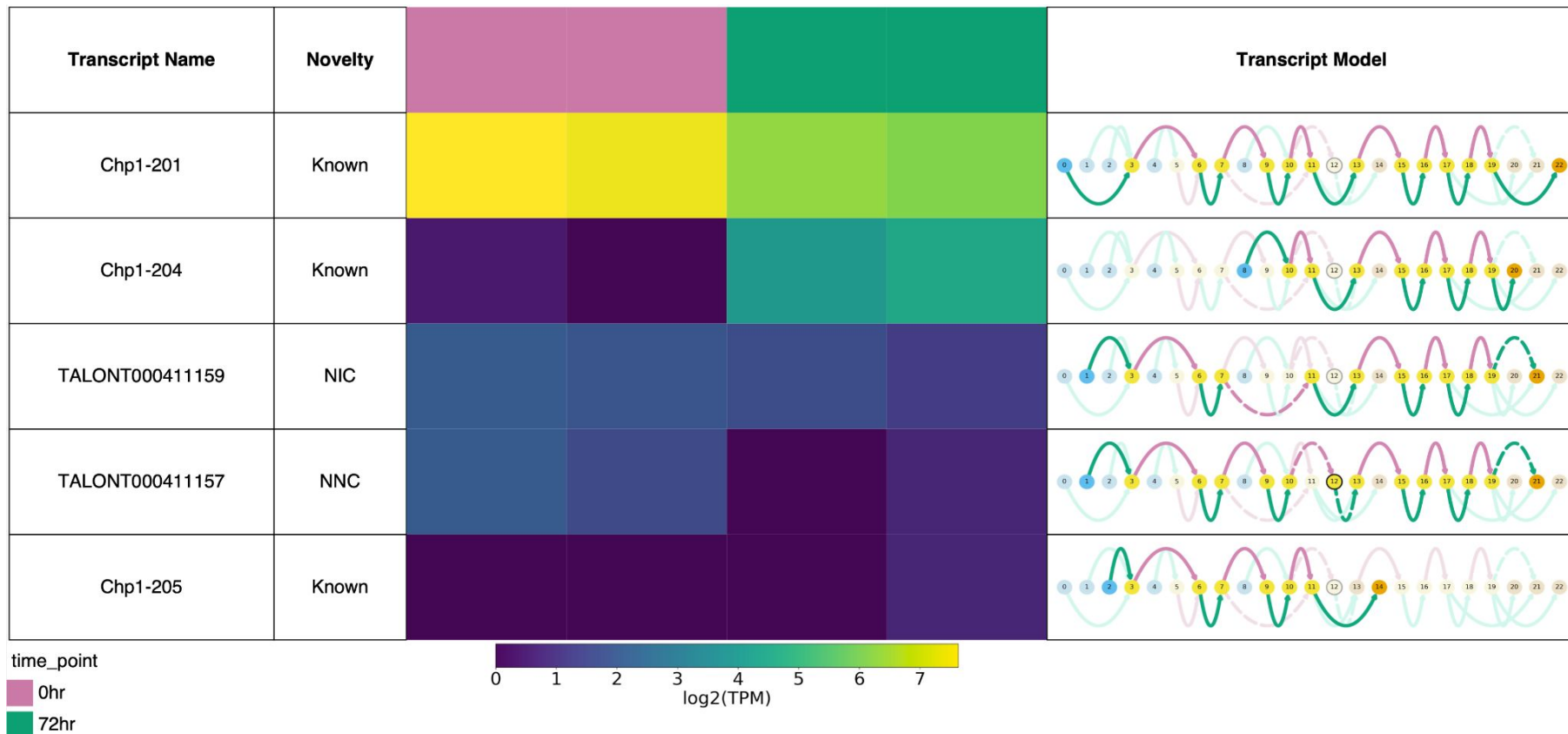- Found isoform-switching, TSS-switching, and TES-switching genes with Swan
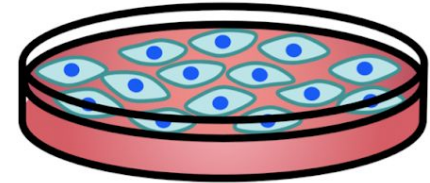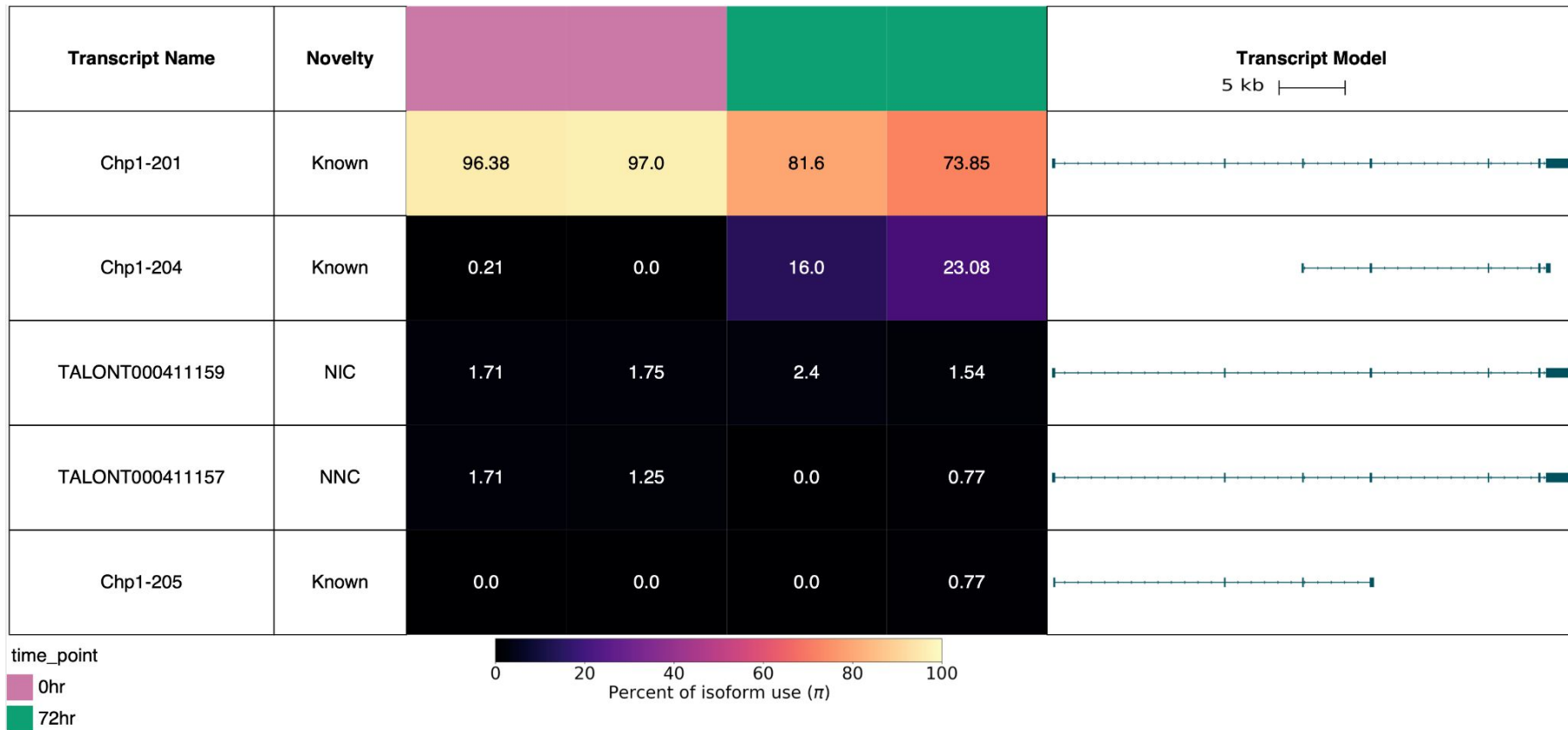
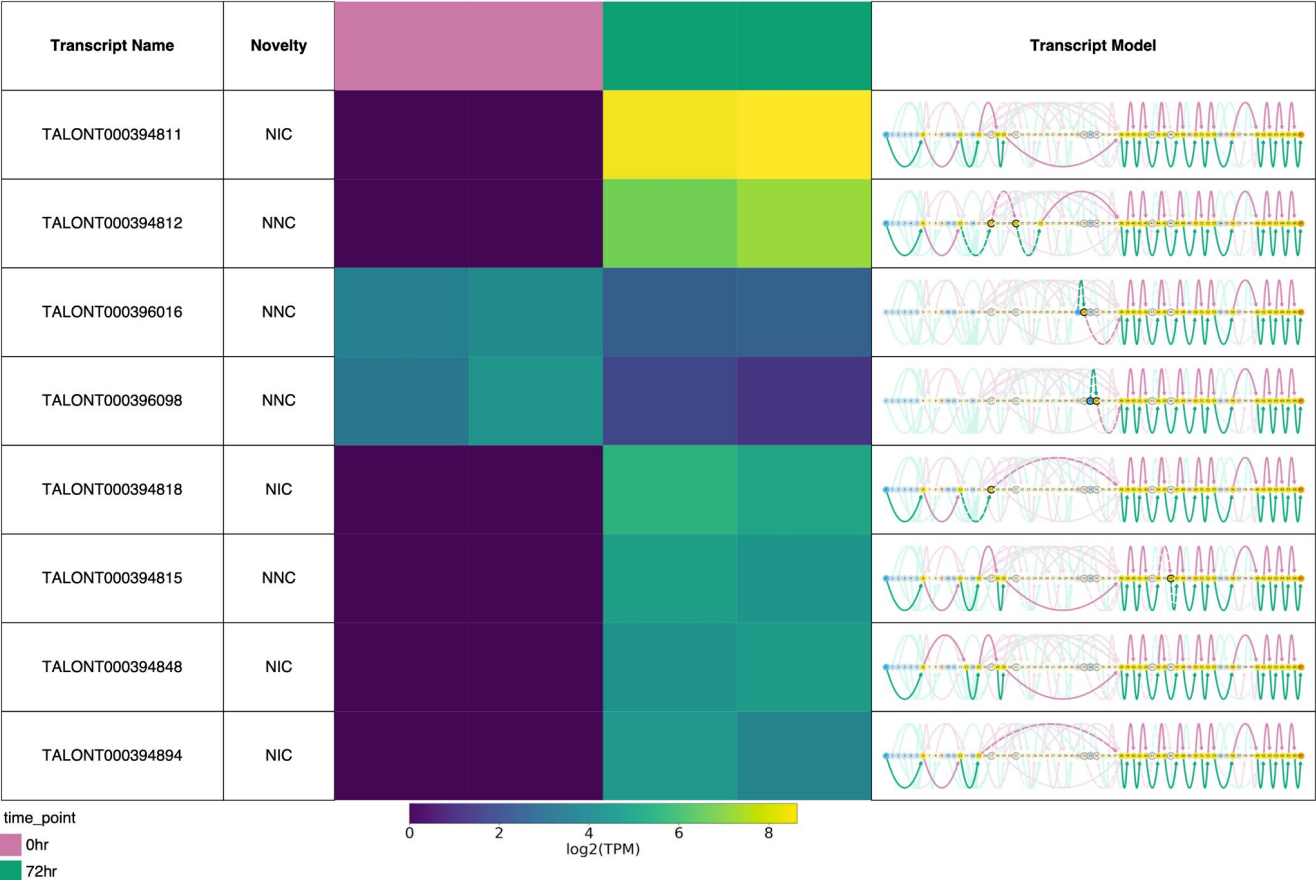C2C12 0hr myoblasts
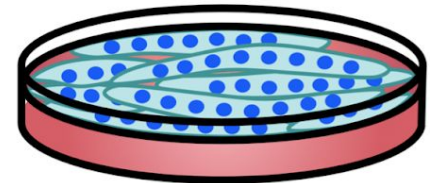
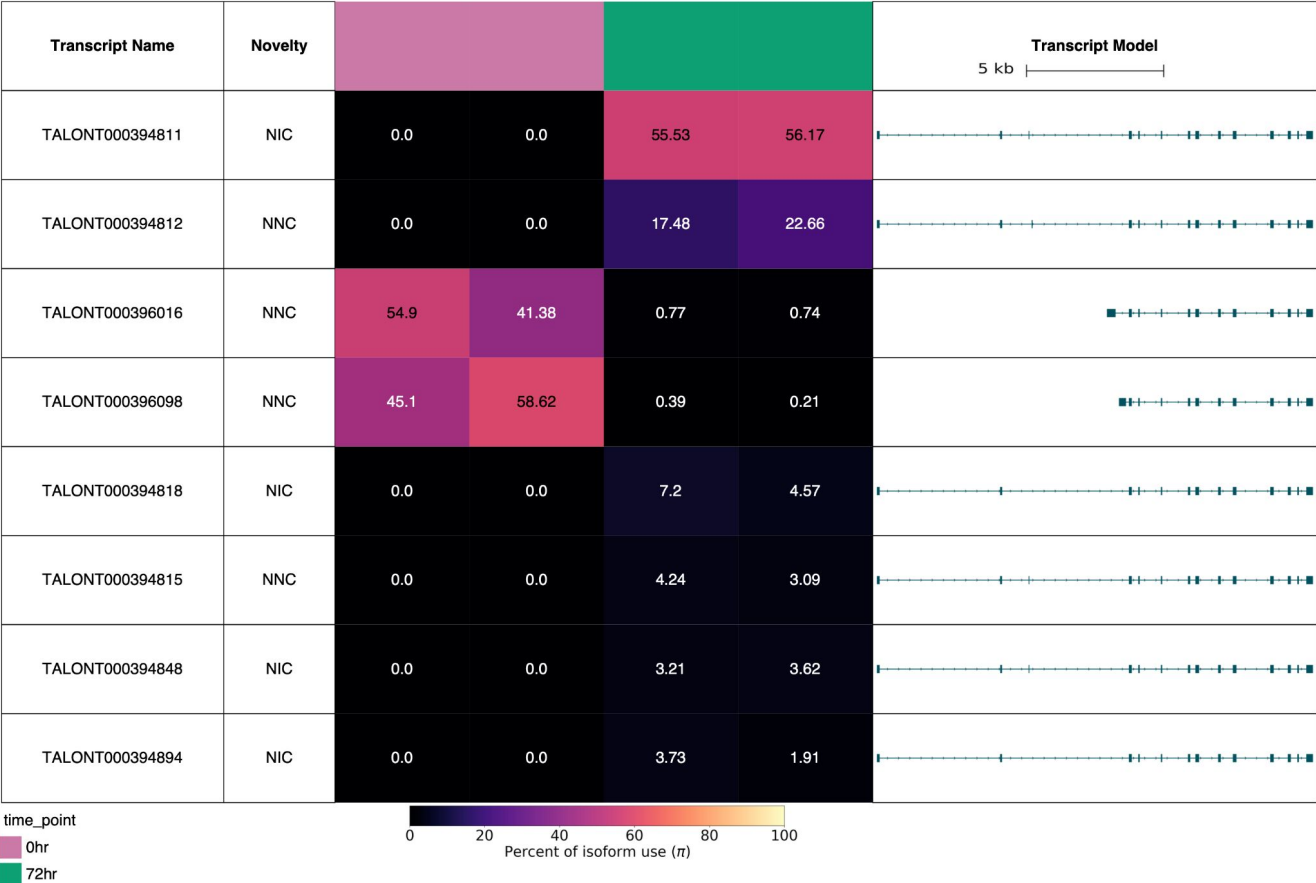C2C12 72hr myotubes

Isoform-switching gene *Chp1* report

Isoform-switching gene *Chp1* report

Transcript Name	Novelty	0hr	0hr	72hr	72hr	Transcript Model 5 kb
Chp1-201	Known	96.38	97.0	81.6	73.85	
Chp1-204	Known	0.21	0.0	16.0	23.08	
TALONT000411159	NIC	1.71	1.75	2.4	1.54	
TALONT000411157	NNC	1.71	1.25	0.0	0.77	
Chp1-205	Known	0.0	0.0	0.0	0.77	

time_point
■ 0hr
■ 72hr

Percent of isoform use (π)
0 20 40 60 80 100

TSS-switching gene *Tnnt2* report

TSS-switching gene *Tnnt2* report
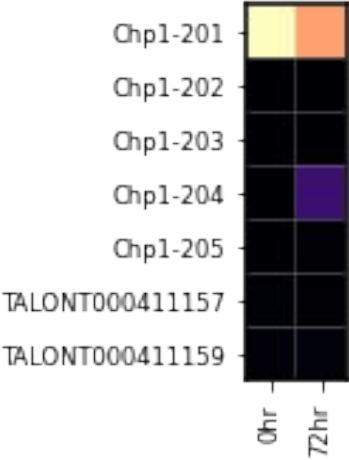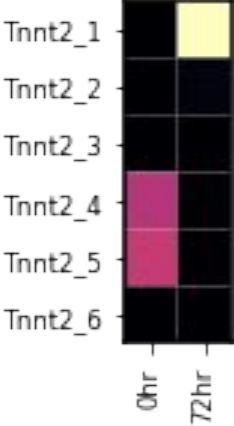
TES-switching gene *Tpm1* report
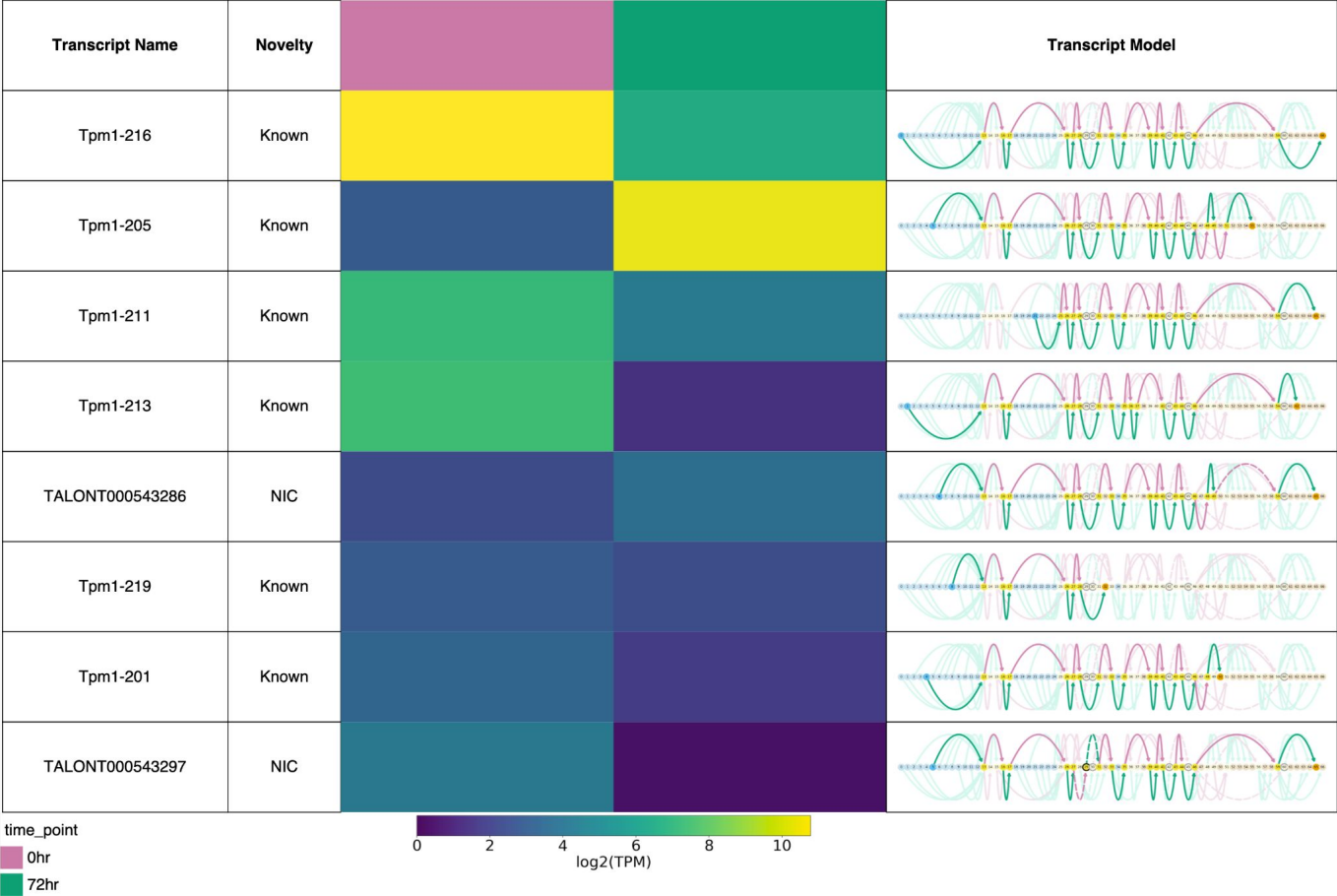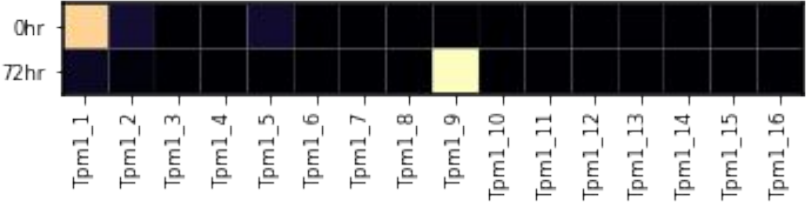
TES-switching gene *Tpm1* report
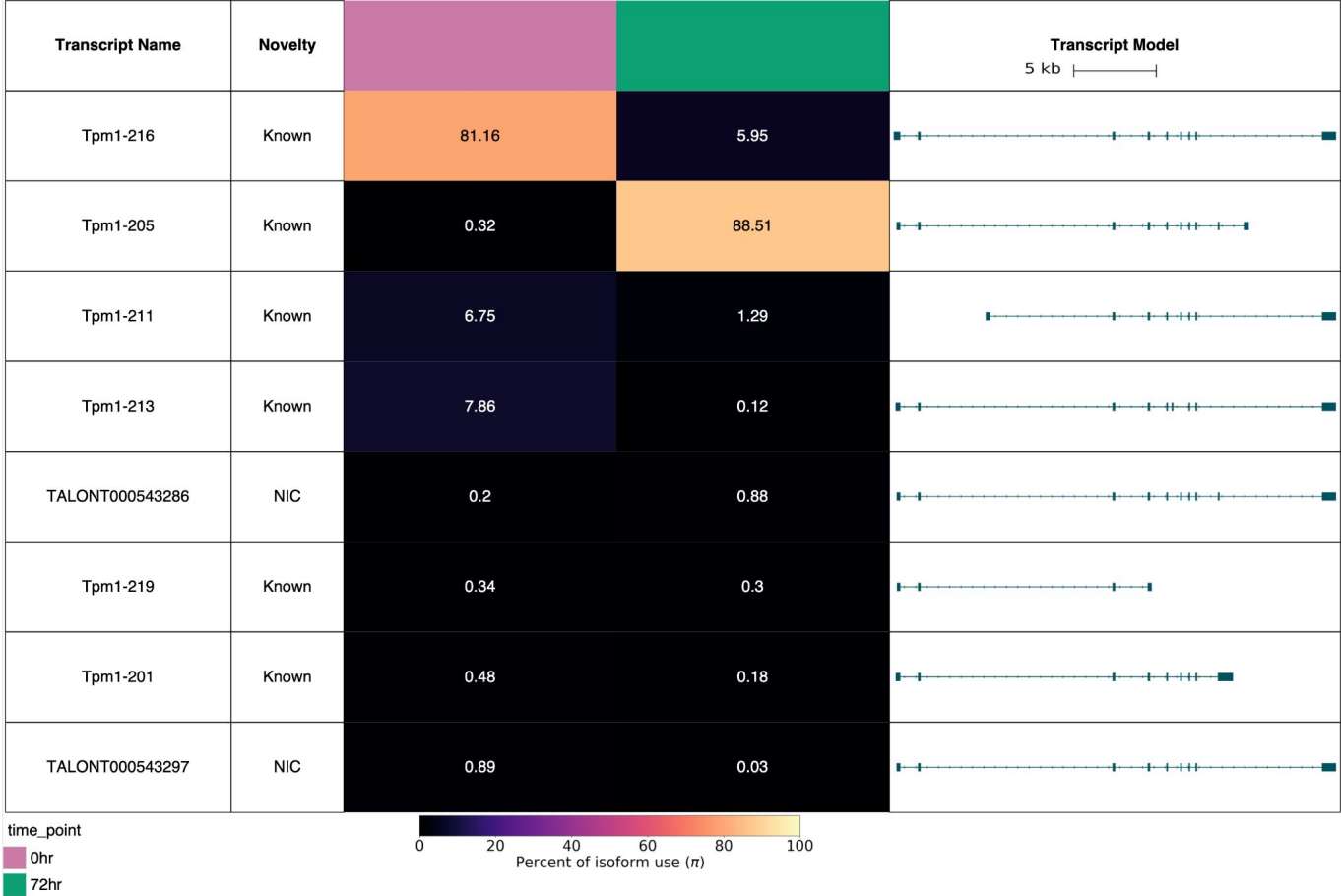
Swan's representation of TSSs, TESs, and isoforms are compatible with Scanpy plotting tools
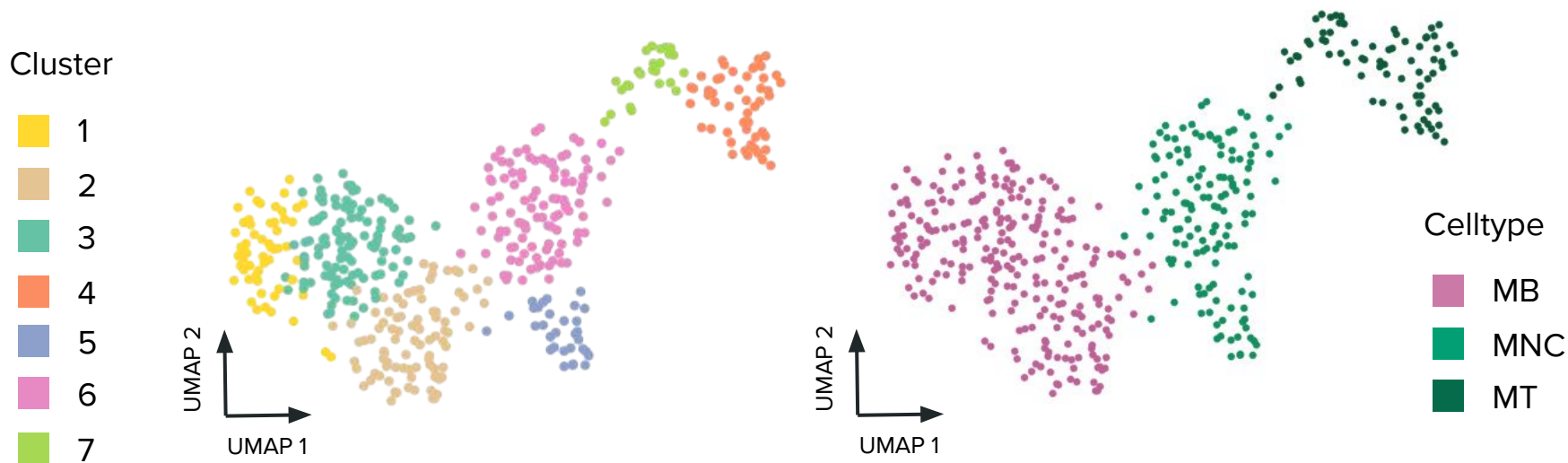

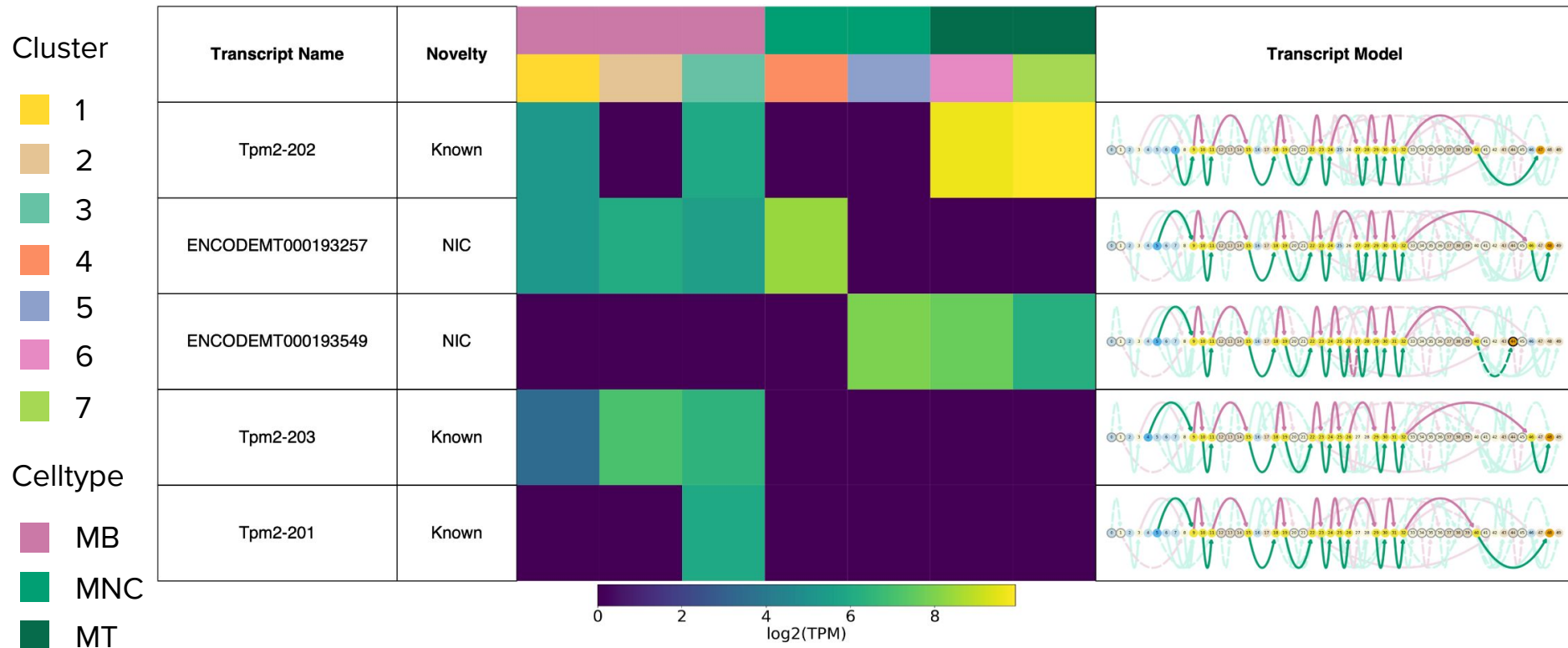Isoform expression

TSS expression

TES expression

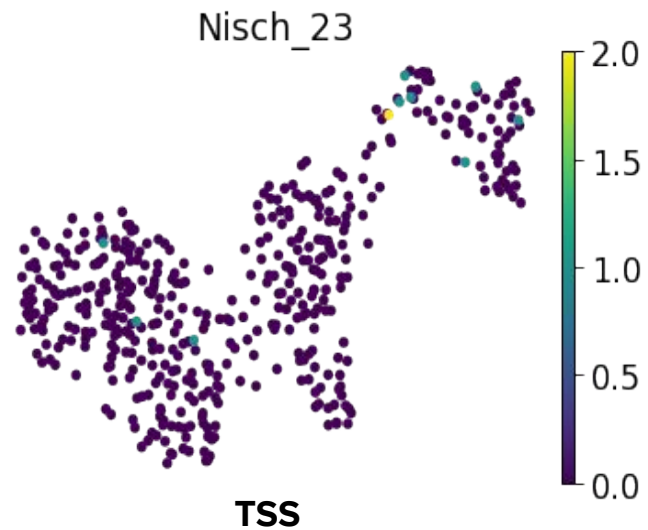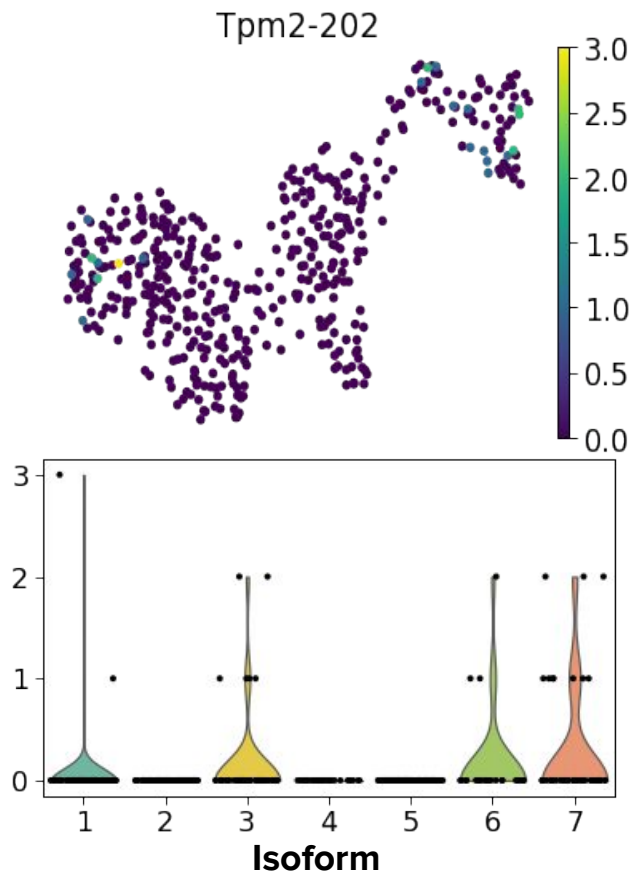# Swan can be used to analyze single-cell transcriptomes

Clusters and cell type labels of 464 differentiating C2C12 skeletal
muscle cells and nuclei

Swan can be used to analyze single-cell transcriptomes

*Tpm2* gene report in differentiating C2C12 skeletal muscle

# Swan can be used to analyze single-cell transcriptomes

Isoform and TSS expression plots in differentiating C2C12 skeletal muscle

Your turn!

- Publicly-available bulk PacBio RNA-seq data from 0hr (myoblast) and 72hr differentiated (myotube) C2C12 skeletal muscle

- Reads cleaned with TranscriptClean and annotated with TALON

- DEGs, DETs, and isoform-switching, TSS-switching, and TES-switching genes called with Swan

- Novel intron retention and exon skipping events called with Swan
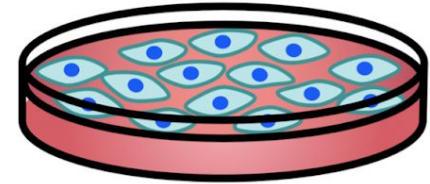
C2C12 0hr myoblasts
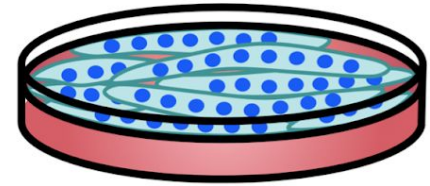
C2C12 72hr myotubes

Acknowledgements

Mortazavi Lab:

Elisabeth Rebboah

Gaby Balderrama-Gutierrez

Dana Wyman, PhD (former)

Katherine Williams

Heidi Liang

Cassandra McGill (former)

Klebea Carvalho

Christina Wilcox
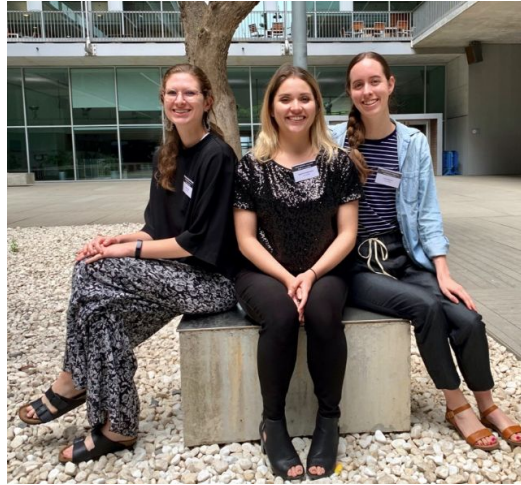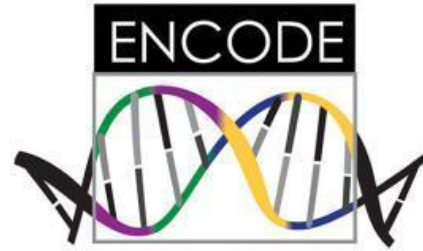
Narges Rezaie

Jasmine Sakr

Angelique Cortez

Isaryhia Rodriguez

Hasan Celik

Wold Lab (Caltech):

Diane Trout