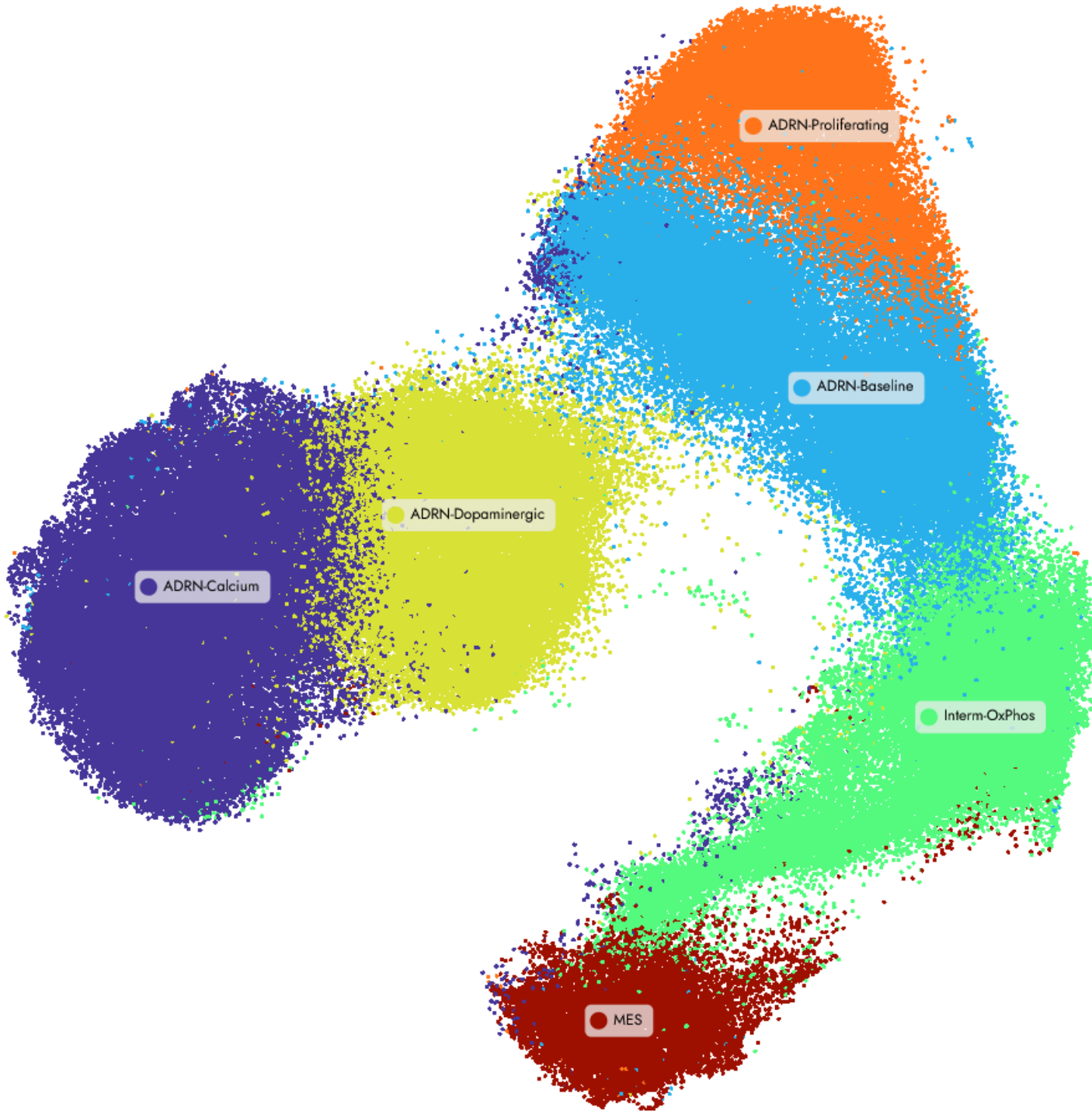
Single-cell Data at Scale

Cell-level predictions of biological activity in single cell data

Joseph Pearson, PhD
Global Product Manager, QIAGEN OmicSoft



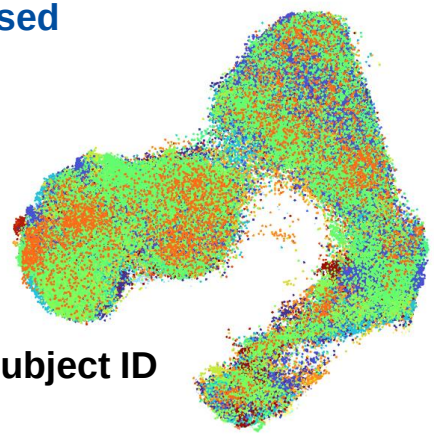
Legal disclaimer

QIAGEN products shown here are intended for research use only. Not for use in diagnostic procedures.

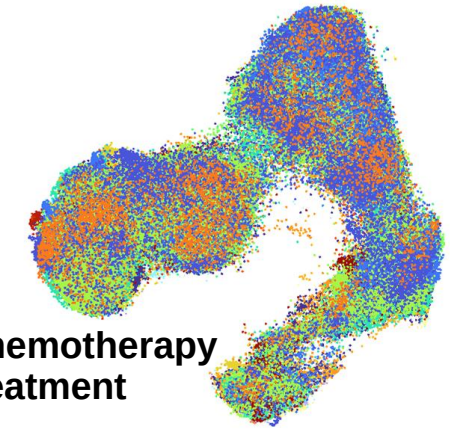
For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit instructions for use or user manual. QIAGEN instructions for use and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services (or your local distributor).

Bridging the gap from single cell metadata to biological interpretation

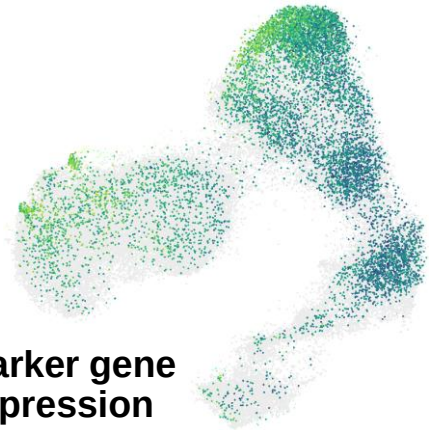
Annotation-based analysis



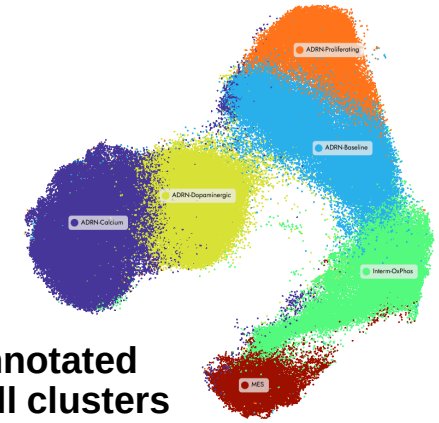
Subject ID



Chemotherapy treatment

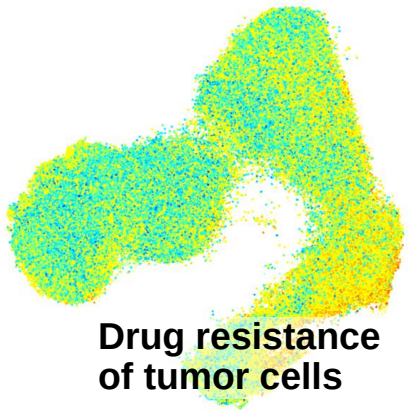


Marker gene expression

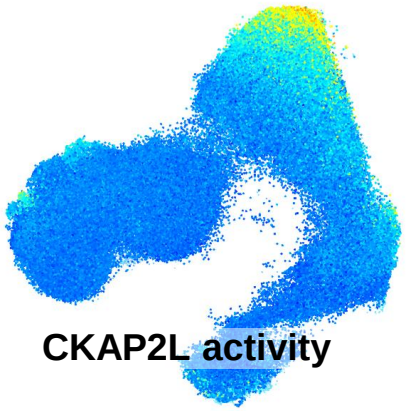


Annotated cell clusters

Map biological functions, upstream regulators and diseases at a single-cell level

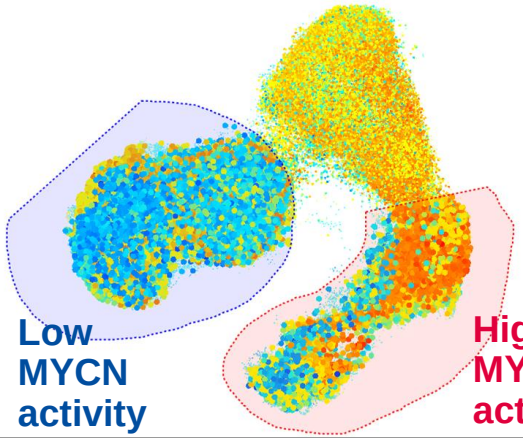


Drug resistance of tumor cells



CKAP2L activity

Perform differential expression of custom cell group selections



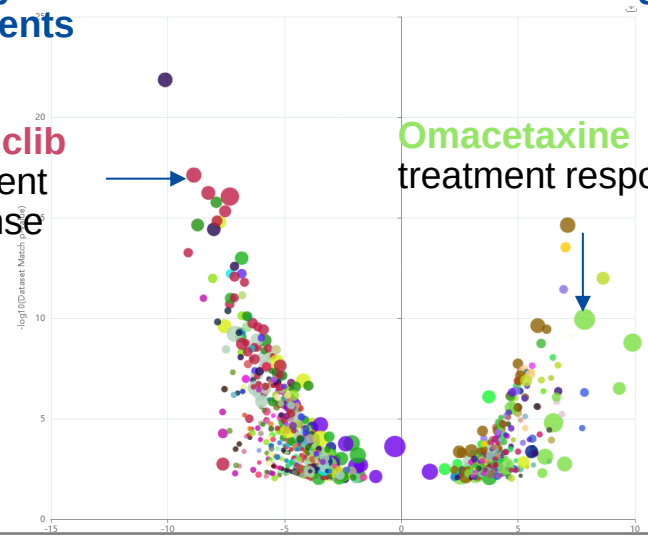
Low MYCN activity

High MYCN activity

Identify similar datasets and candidate drug treatments

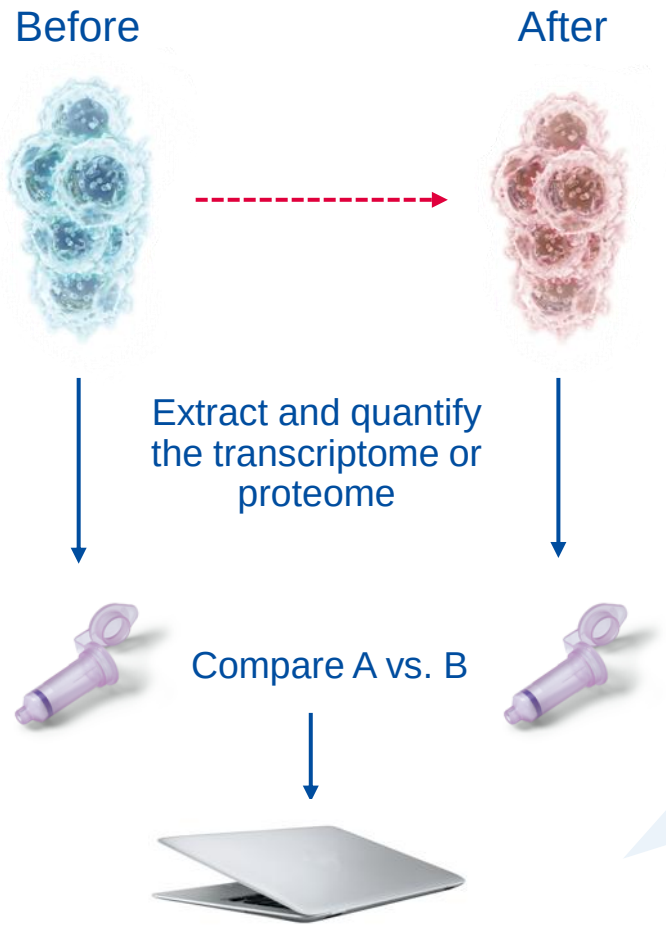
Dinaciclib treatment response

Omacetaxine treatment response



The value that QIAGEN IPA add for ‘single cell data analysis

Understand how cells differ from one another biologically to drive discovery



What is the cause and what is the effect?

What drives the expression changes?
- → Biomarkers, new targets, mechanistic understanding, drug re-purposing

What effect does that have on the biology?
- → Mechanistic hypotheses

Which well-understood pathways are activated or inhibited?
- → Biomarkers, new targets, mechanistic understanding, drug re-purposing

Do I see the biological changes I expect?
- → Confidence that the samples and experiment were of high quality

Are these results unusual or common?
- → New hypotheses, publish papers

Do the results suggest new hypotheses?
- → Publish papers, advance your research field

The data behind activity prediction: Causality drives insight

Mutation of mouse N-myc [Mycn] gene to homozygous mutant **mouse N-myc [Mycn] gene** (knockout) **in** NSC [neural stem cells] from 12.5 day-old embryonic mouse **decreases expression of mouse Hes5 mRNA in** cultured NSC [neural stem cells] from 12.5 day-old embryonic mouse.

Experiment Type: microarray analysis

22745758 Martínez-Cerdeño V, Lemen JM, Chan V, Wey A, Lin W, Dent SR, Knoepfler PS. N-Myc and GCN5 regulate significantly overlapping transcriptional programs in neural stem cells. PLoS One. 2012;7(6):e39456. Epub 2012 Jun 26.

Source: Ingenuity Expert Findings

Species

Direction of effect on gene expression

Activity of the molecule in this finding (decreased)

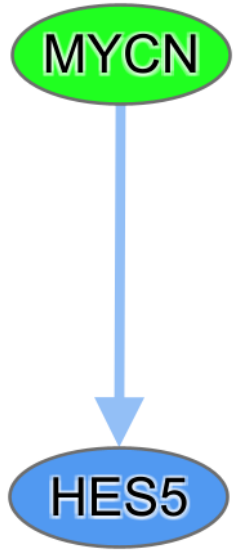
Mutation type

Downstream gene

Upstream gene

Tissue

This structure provides rich contextual detail and causality and can be used computationally.



IPA reveals biology driving differential expression between cell groups

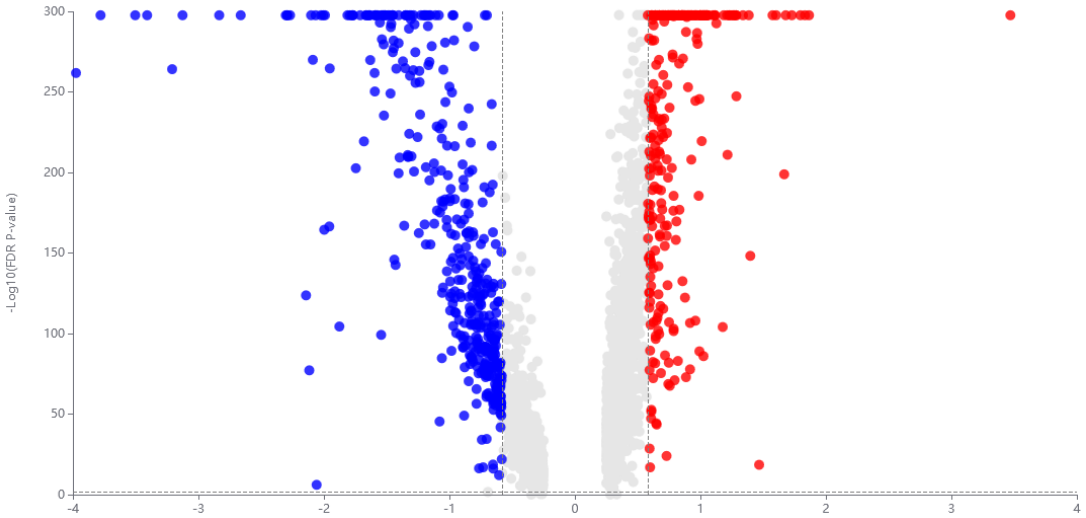


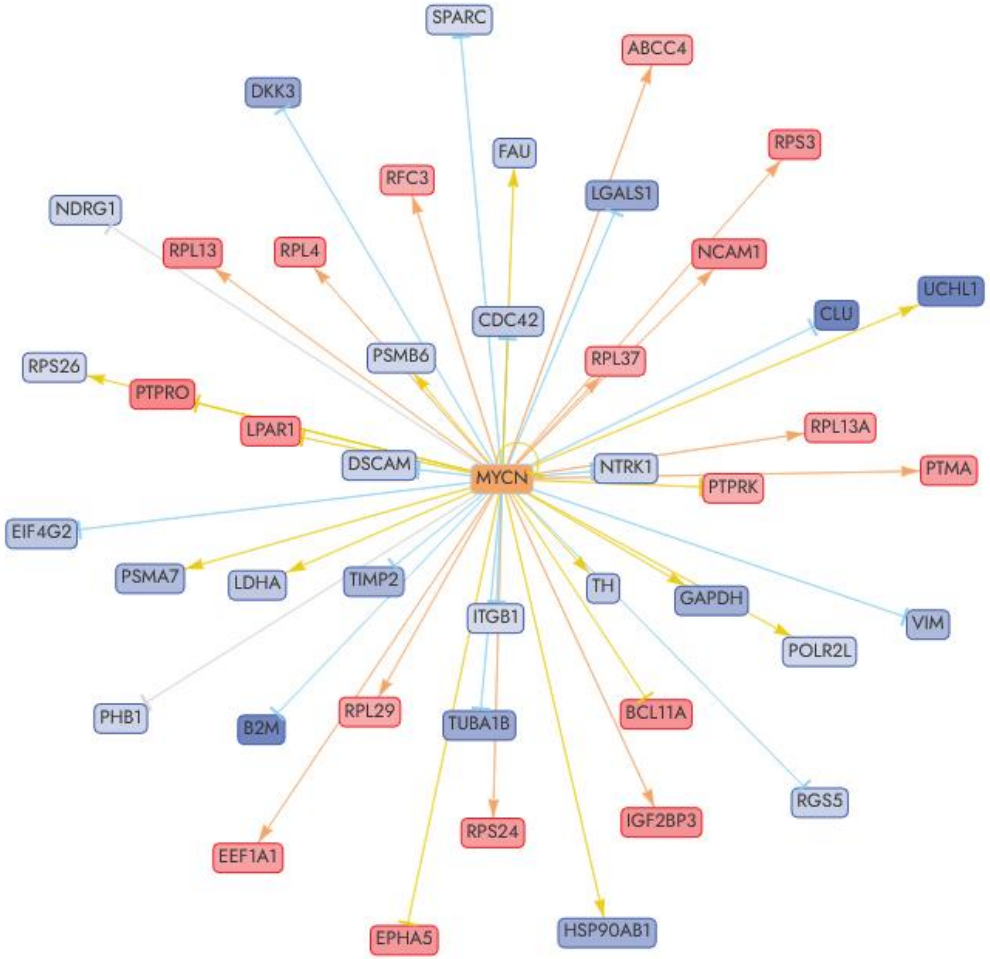
MYCN amp. neuroblastoma



MYCN normal neuroblastoma

Differential expression





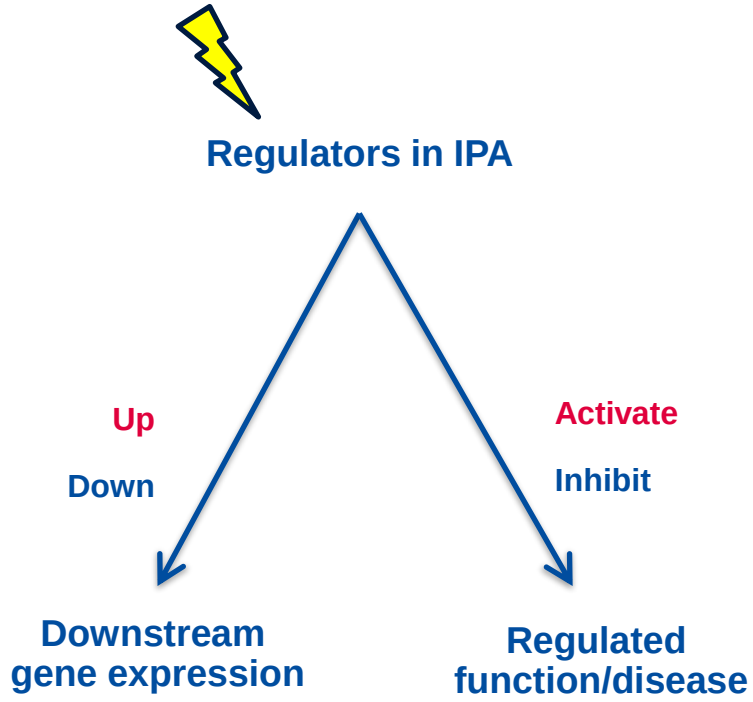
Causal regulator to gene relationships

TNF increases induction of IL-6.
(9792645)

In Hela cells, **human TNF-alpha [TNF] protein increases expression of human IL6 mRNA.**
(10903137)

In Hela cells, **human TNF-alpha [TNF] protein increases activation of** promoter fragment from **human IL6 gene.**
(10903137)

Mutant **mouse Tnf gene** (allele $Tnf^{tm1Sek}/Tnf^{tm1Sek}$) (knockout [homozygous]) in B6.Cg-Tnf mouse **decreases level of mouse Il6 protein** in blood.
(0008597)



A protein fragment (ATVIVITLVMLKKKQYTS713-730) from **APP protein decreases quantity of** primary **hippocampal neurons** from 17.0 day-old embryonic rat.
(10233838)

Fibrillar **AMYLOID beta42 [product of APP] protein(s) decreases quantity of synapse** from viable neurons.
(7991613)

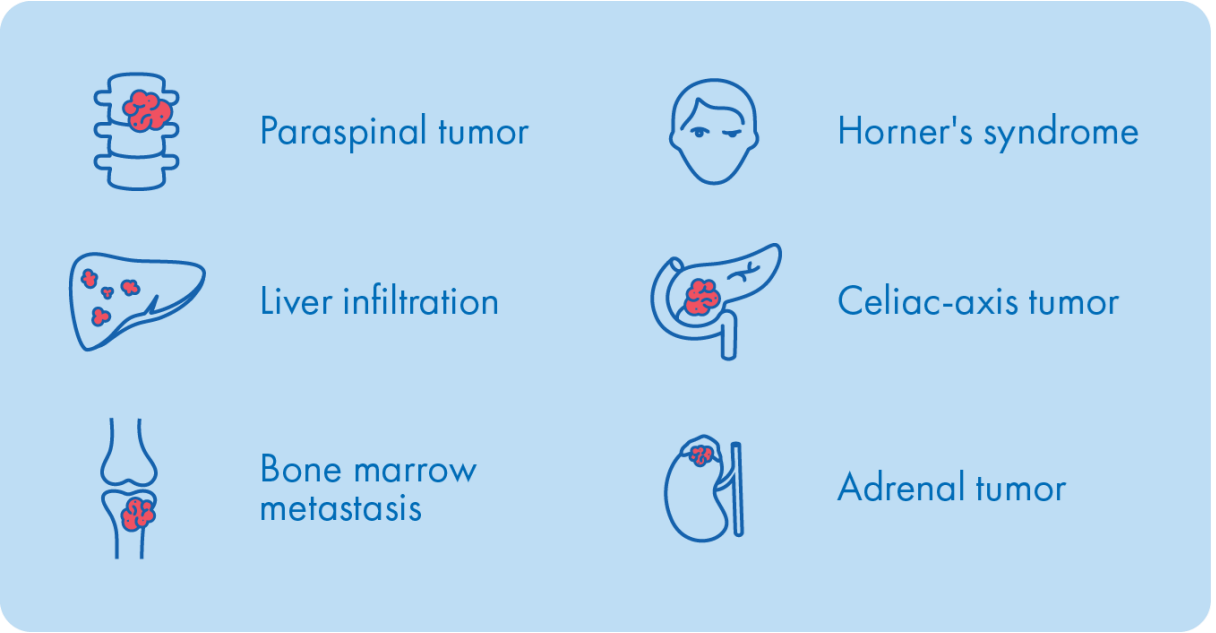
Transgenic mutant **human APP protein** (substitution K670N;M671L;V717F) in mouse **decreases quantity of pyramidal neurons** expressing **mouse Reelin [Reln] protein** in mouse entorhinal cortex.
(17360894)

- Edges have a sign +/–1
- Data collected across many contexts was used to build a general model of biological regulation

Case study: Neuroblastoma

What is neuroblastoma?

- Tumor derived from incompletely-committed neural crest, develops in infants and children
- Impacts ~700 children in the US each year with a high relapse rate, particularly for older patients
- ALK gain-of-function and chromosomal aberrations drive neuroblastoma
- Neuroblastomas with MYCN amplification associated with high risk



Why is careful single cell analysis essential for neuroblastoma?

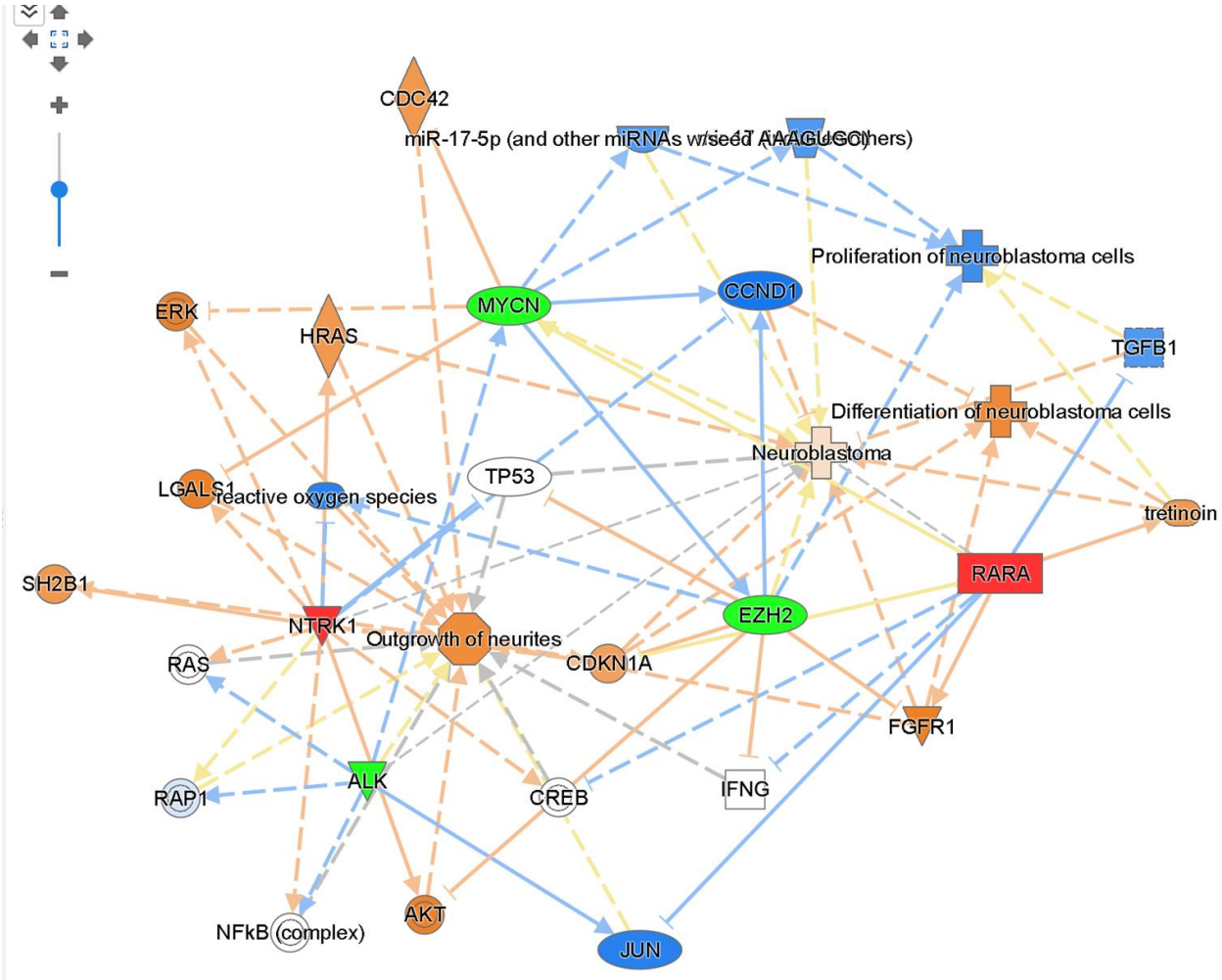
- The complexity of cell types and changes in the tumor are the key barriers to treatment and drivers of relapse
- Diverse regulatory mechanisms complicate understanding of biology
- The field needs more diverse drug candidates and cell-type-specific monitoring targets, especially for high-risk subgroups

Modified from Maris JM. Recent advances in neuroblastoma. *N Engl J Med*. 2010;362(23):2202–2211.

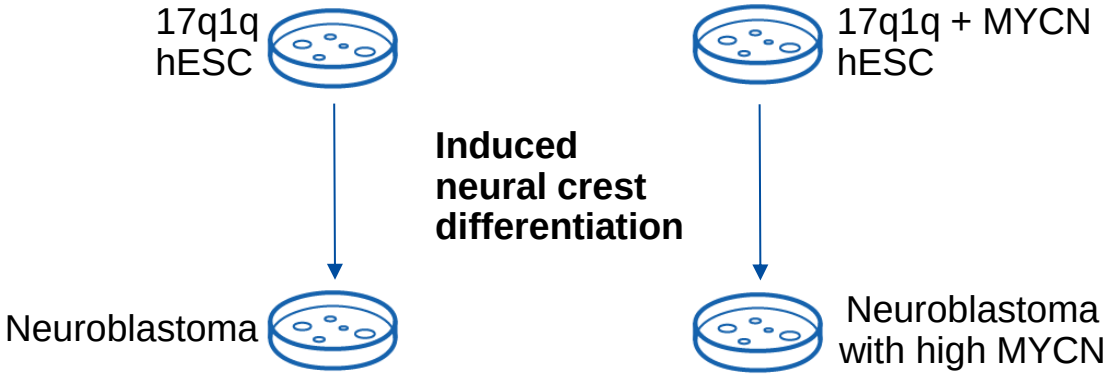
What are the key regulators of cell differentiation in neuroblastoma?

MYCN

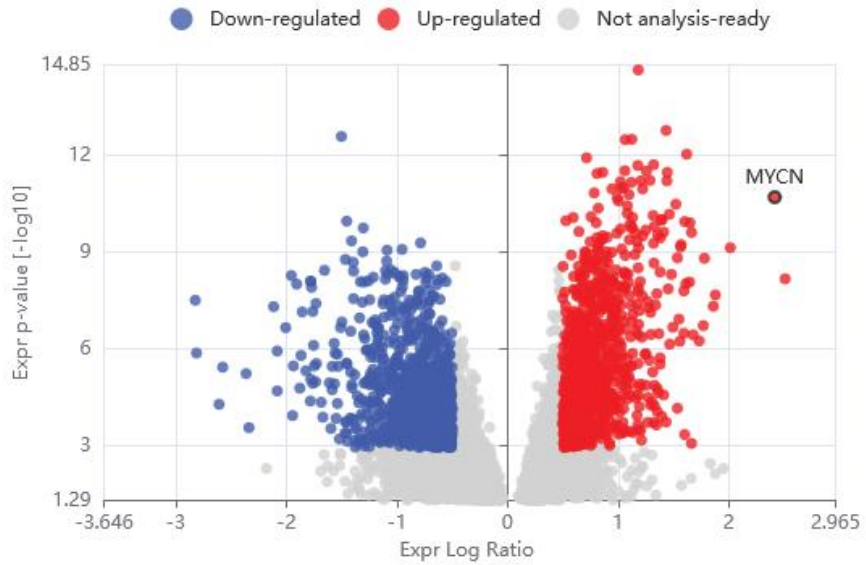
- Central oncogenic driver in high-risk neuroblastoma
- Amplification strongly correlates with poor outcome and undifferentiated phenotype
- MYCN drives proliferation and represses differentiation programs



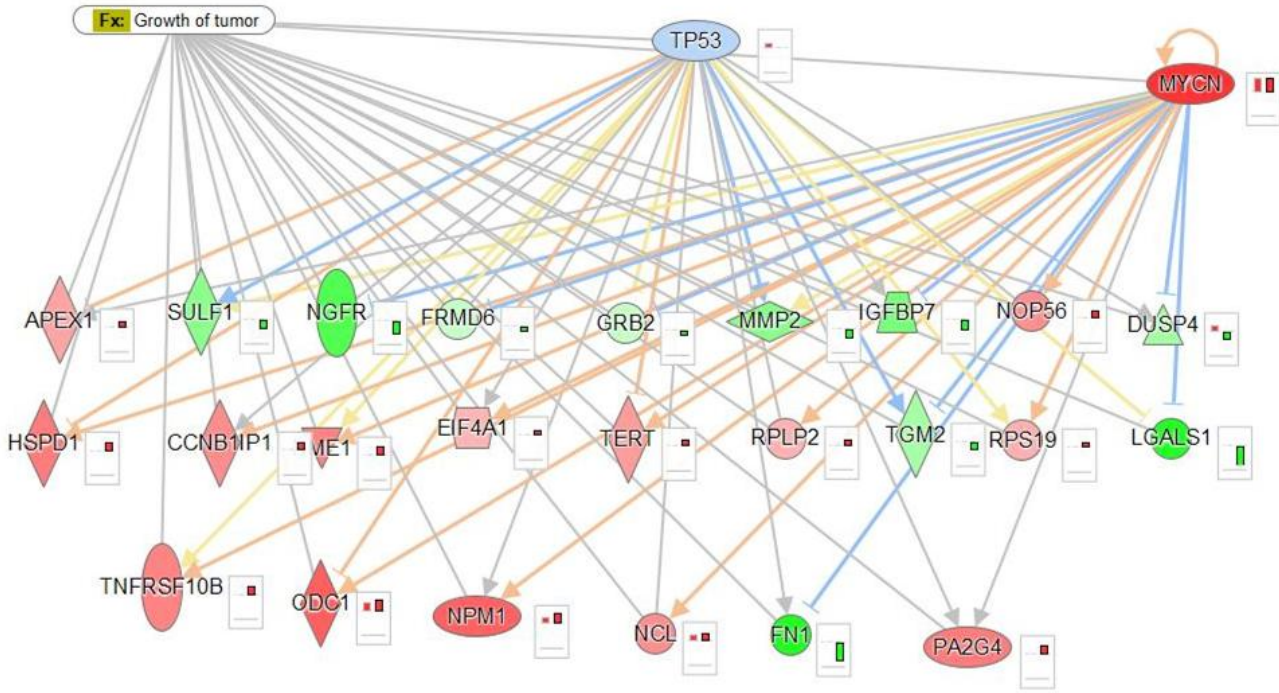
IPA reveals mechanisms activated in models of neuroblastoma



Differential expression



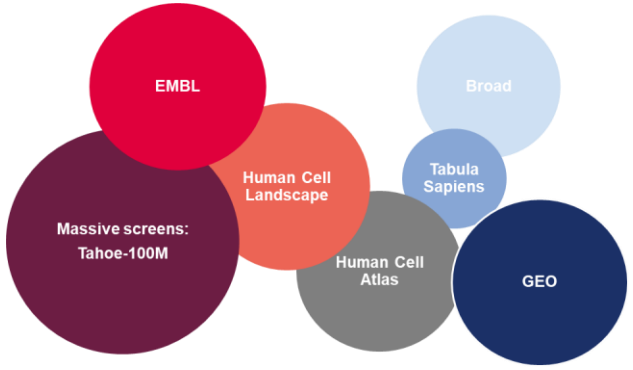
MYCN, along with other regulators such as TP53 regulate many genes involved in tumor growth and survival, but MYCN activation reduced metastatic traits.



IPA - OmicSoft Explorer contextualizes your experiments against thousands of curated datasets

Unified data accelerate the discovery of disease mechanisms in single-cell data

Scattered single-cell data



File types: .h5ad, .rds, .mtx, etc.

Standardization

- Align metadata to controlled vocabularies
- Apply detailed, consistent cell type labels
- Normalize and map expression to gene model
- Cell quality control
- Standardize file structure

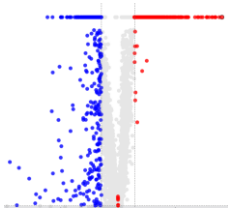
Downstream analysis

- Cell-level expression
- Precomputed comparisons
- Biological interpretation

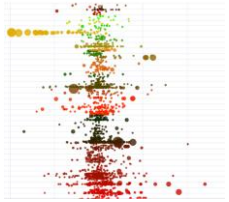
Expression specificity



Top upregulated targets



Indication expansion



Biological mechanisms



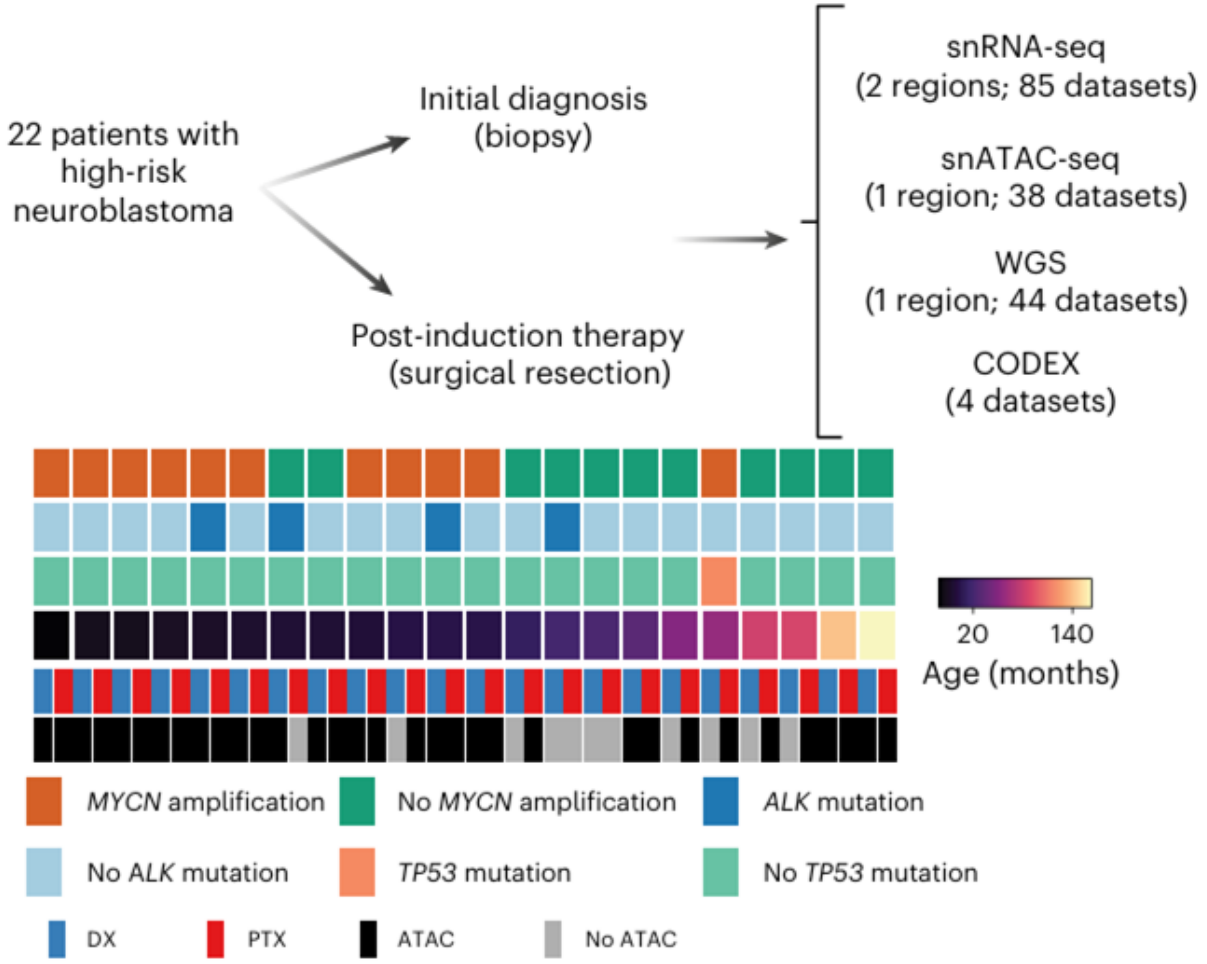
Curated neuroblastoma datasets provide context to your experiments

nature genetics

Article https://doi.org/10.1038/s41588-025-02158-6

Longitudinal single-cell multiomic atlas of high-risk neuroblastoma reveals chemotherapy-induced tumor microenvironment rewiring

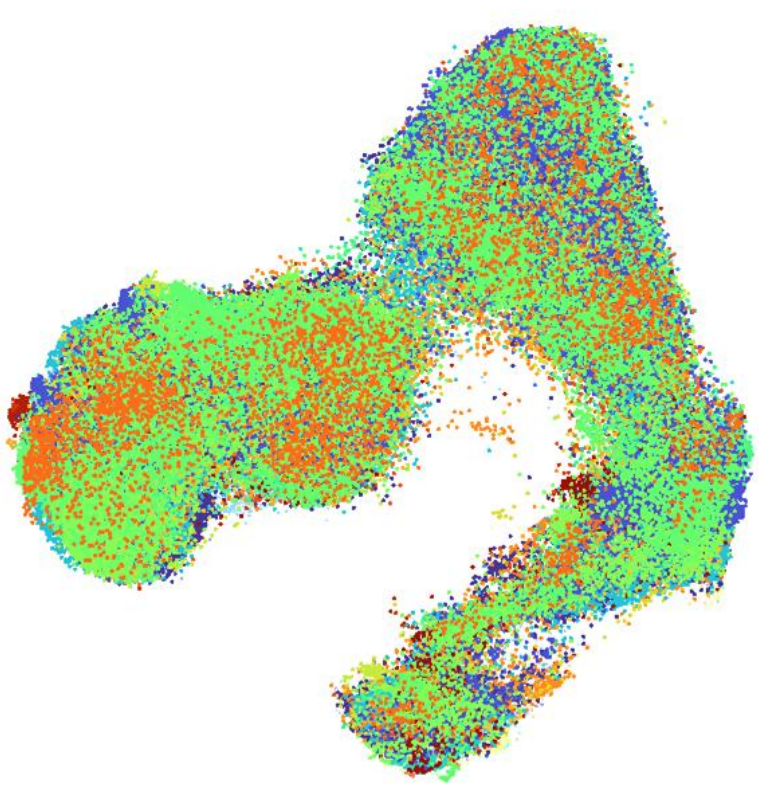
- 22 patients with neuroblastoma, pre- and post-treatment
- scRNA-seq of normal and neoplastic cells
- Profiling of amplification of common oncogenes/tumor suppressors (MYCN, TP53, ALK)



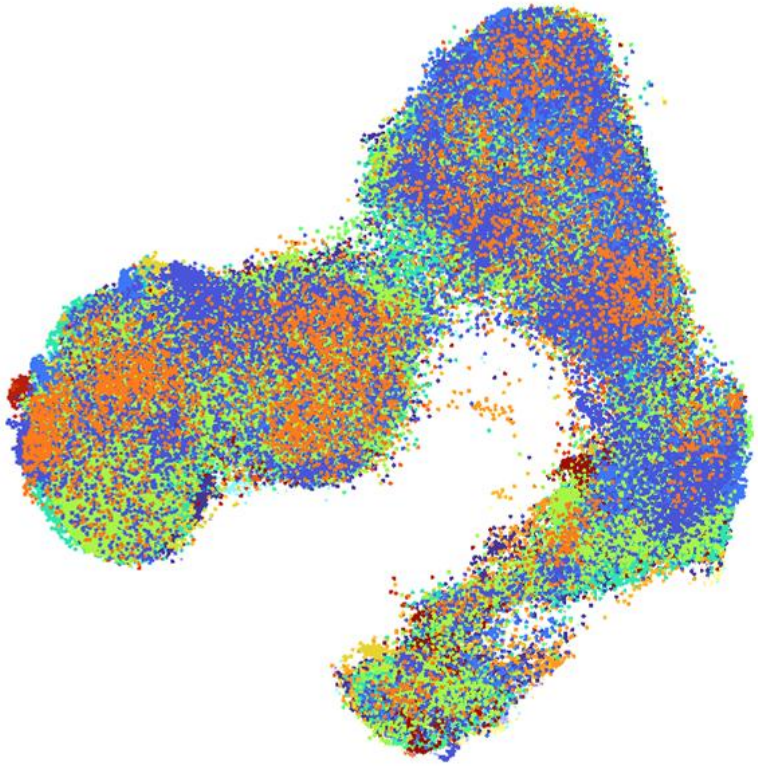
Yu W, et al. Longitudinal single-cell multiomic atlas of high-risk neuroblastoma reveals chemotherapy-induced tumor microenvironment rewiring. *Nat Genet.* 2025;57(5):1142-1154.

Neoplastic neuroblasts from 22 patients' post-chemotherapy do not reveal discrete subtypes

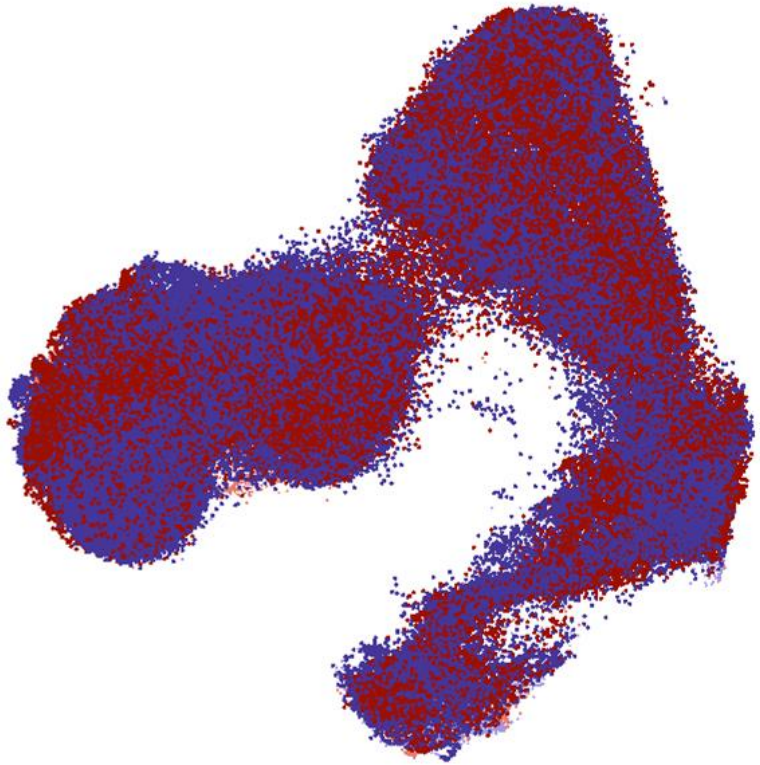
What biological insights can be gleaned from these populations?



Subject ID

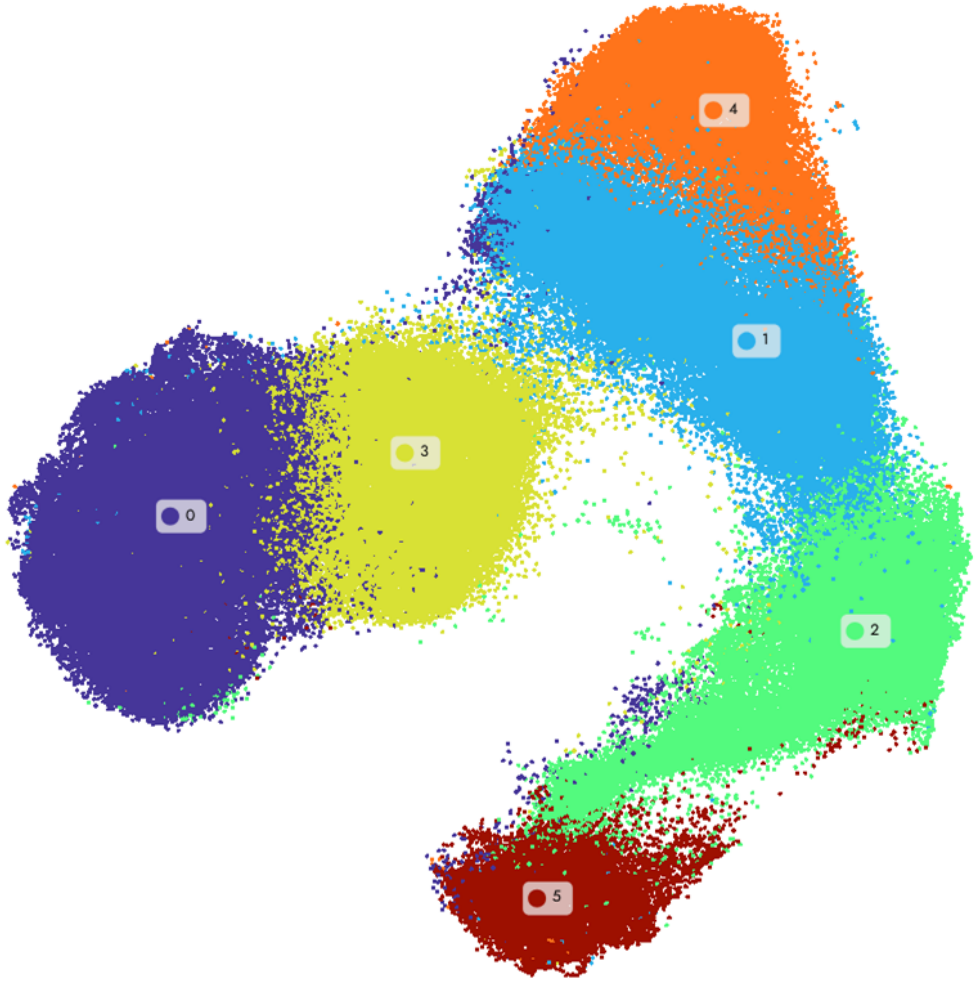


Chemotherapy treatment

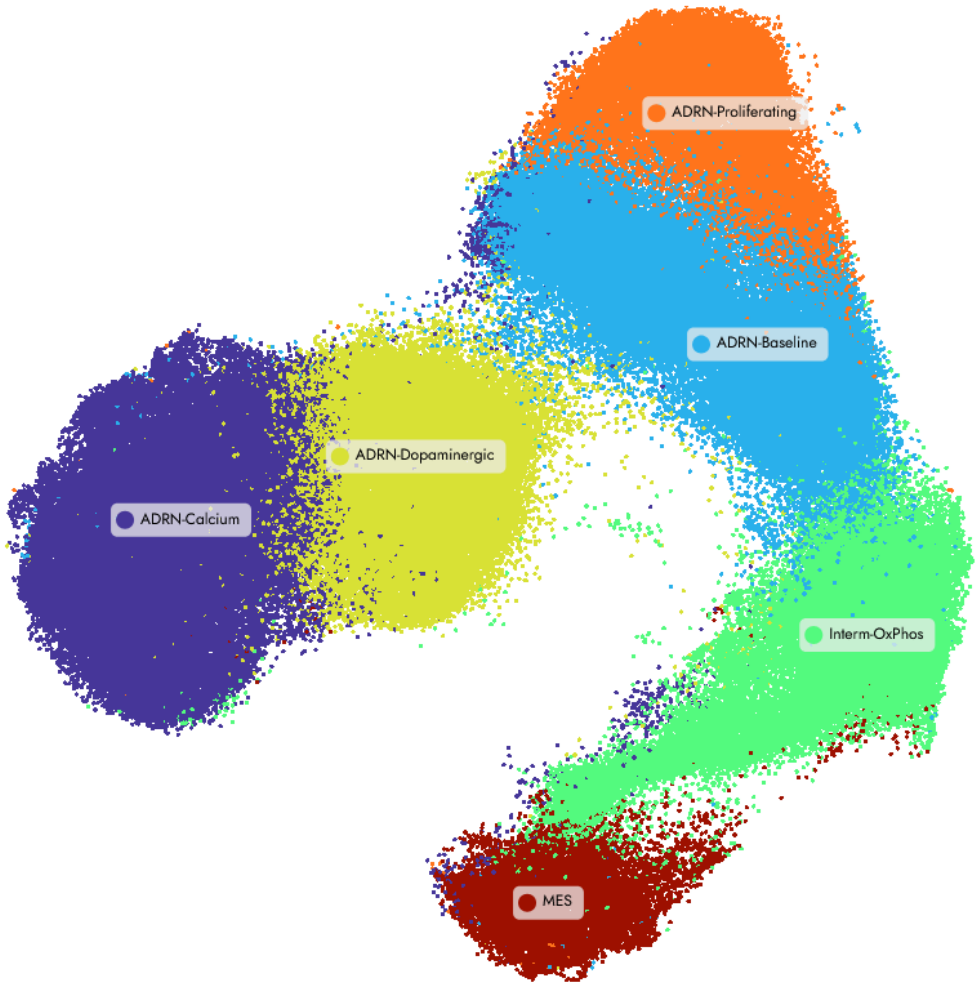


MYCN amplification status

Neuroblastomas can be subdivided for traditional analysis, but this assumes rigid cell states



Cluster identification (Louvain)



Cell state annotation (signature-based)

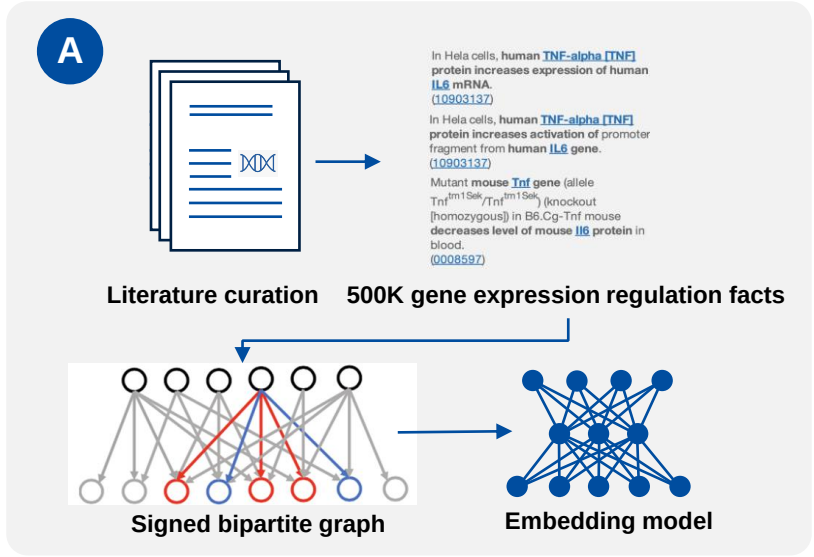
IPA cell-level activity prediction of single cell experiments, enabled by an embedding model from curated causal findings

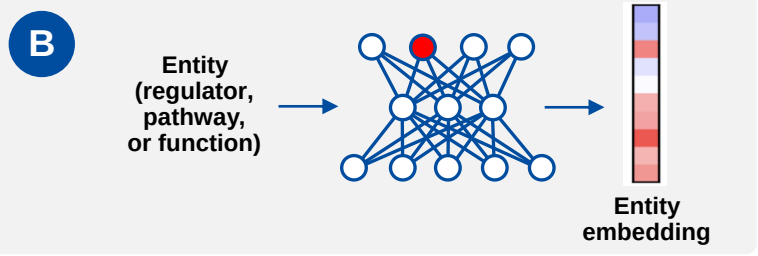
A Embedding model built from causal regulator-gene facts curated from perturbations described in the literature

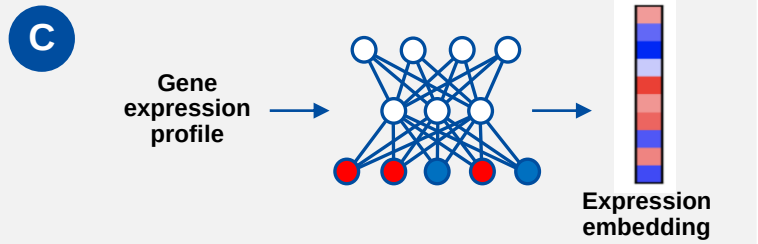
B Entity activity (regulator, pathway, or function) encoded as entity embedding

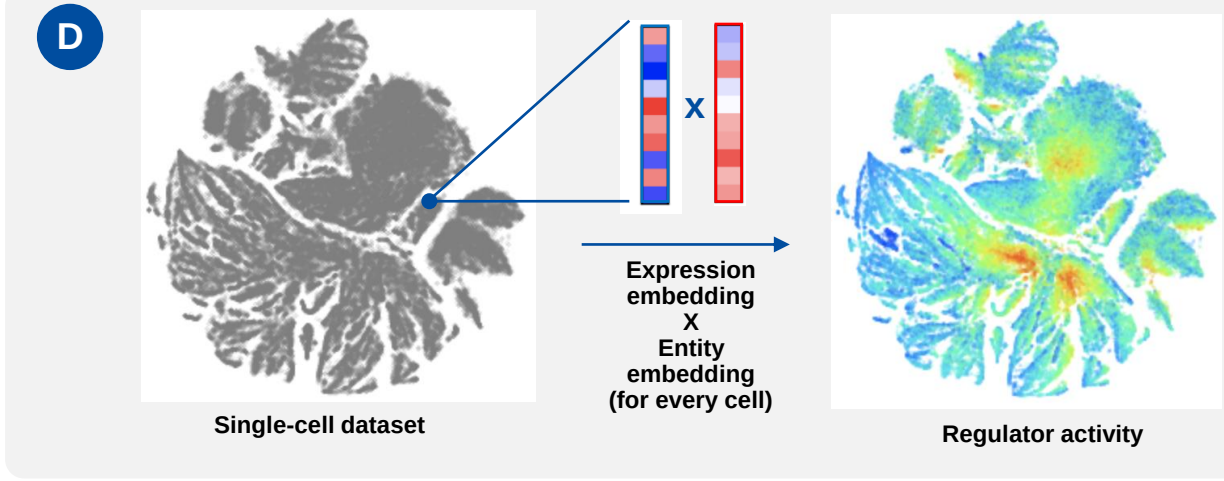
C Gene expression (normalized cell expression) encoded as expression embedding

D Similarity of expression embeddings of each cell to entity embedding (cosine similarity) reveals cell-level biological activity

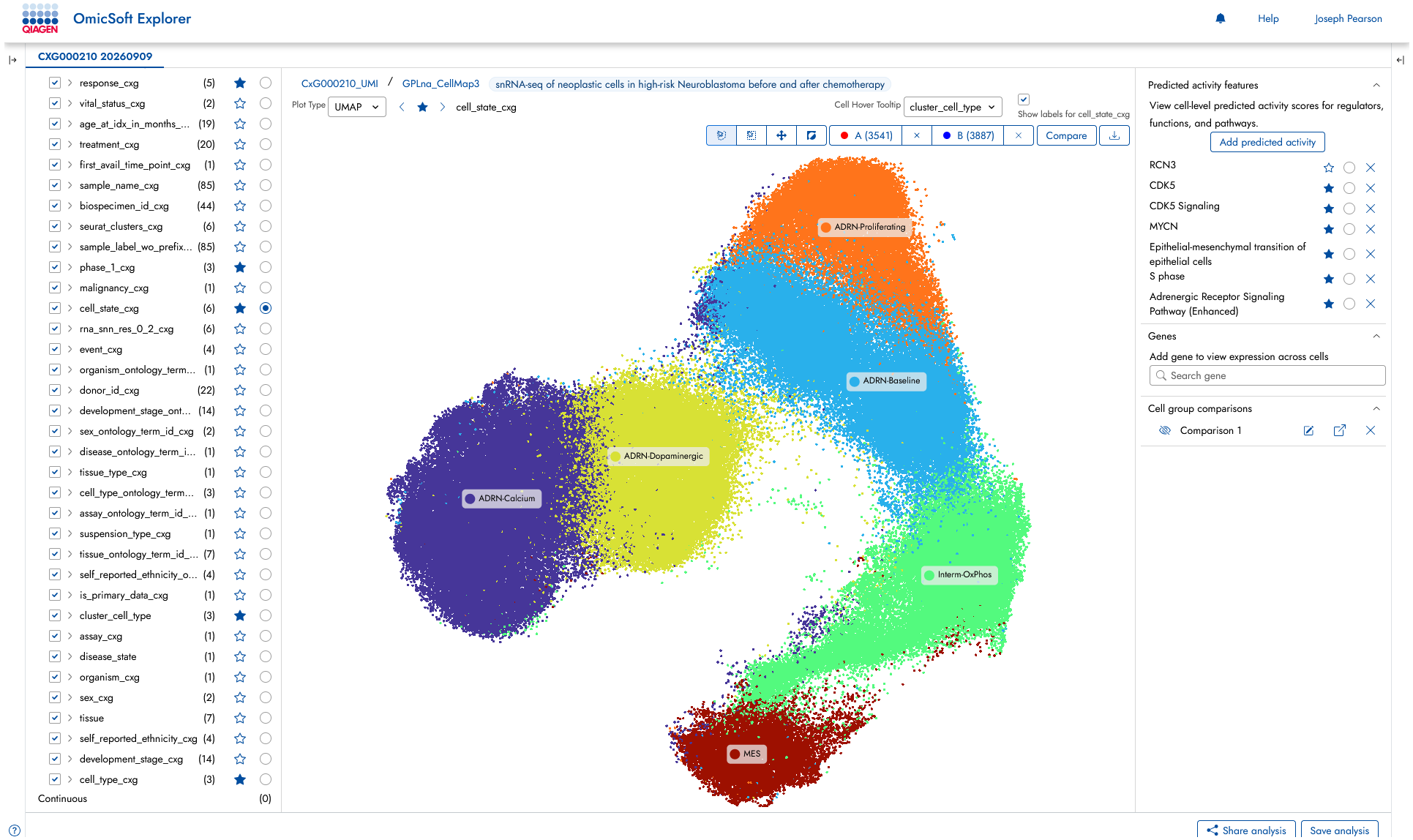




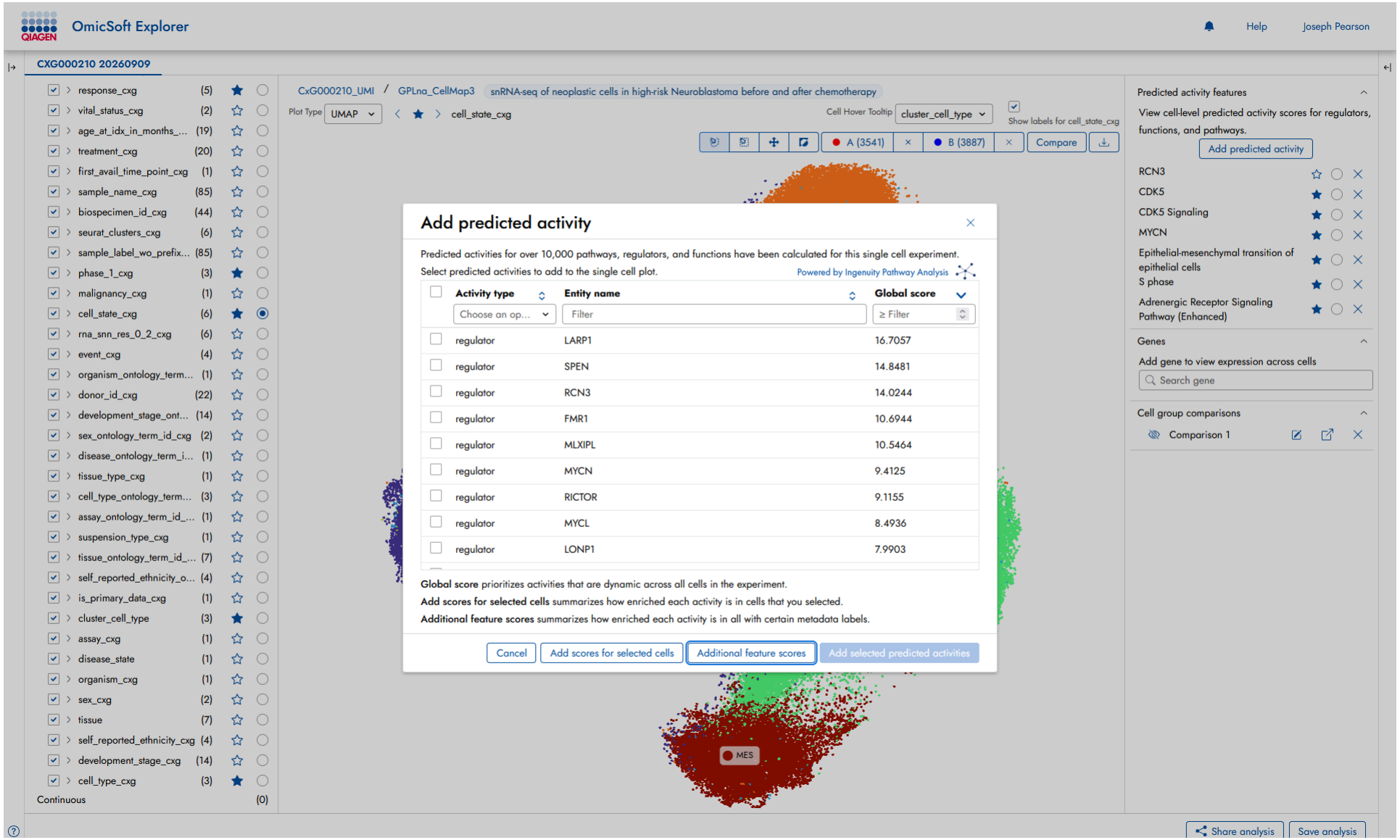




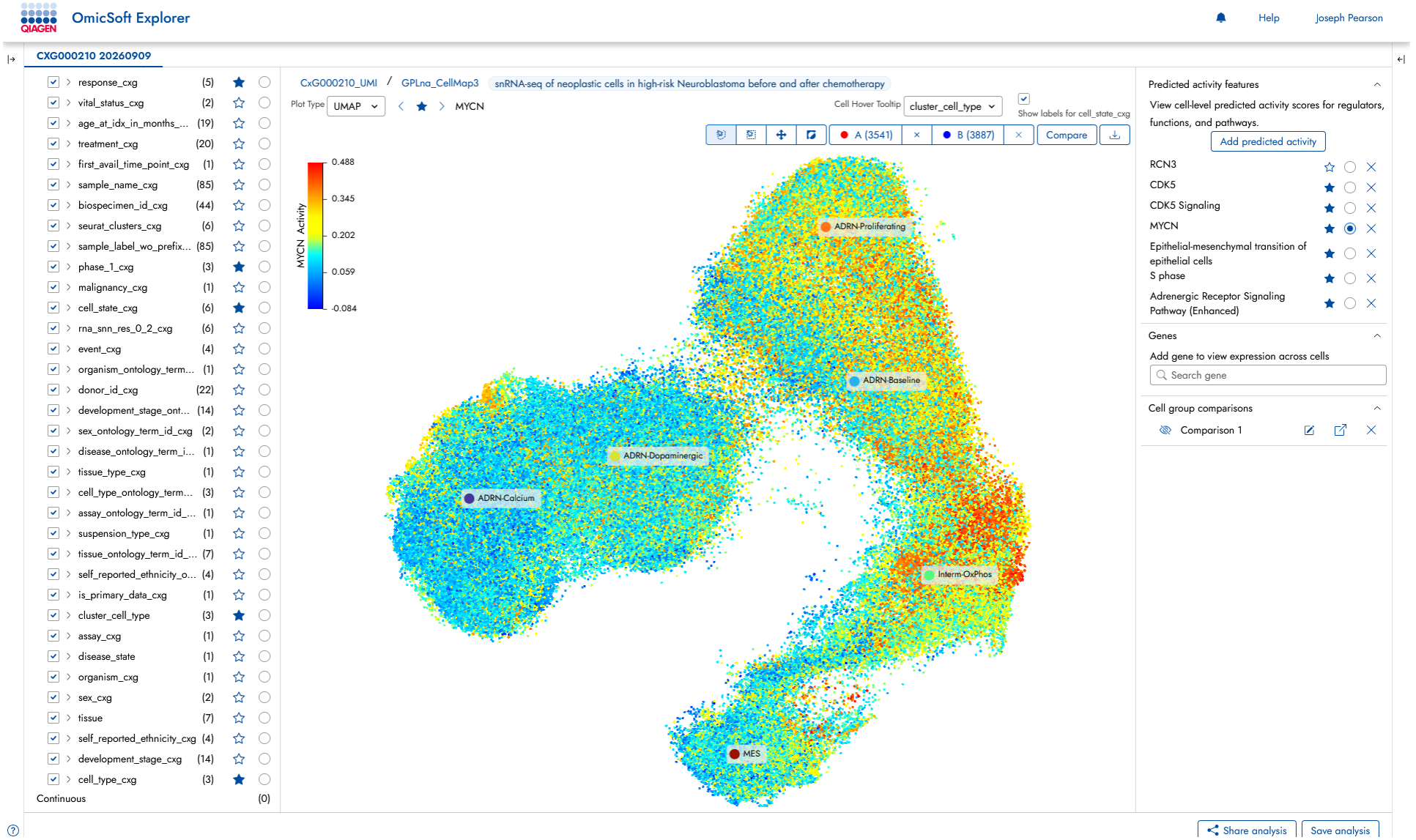
Single Cell Experiment Viewer – Upload your data, or explore curated single cell datasets



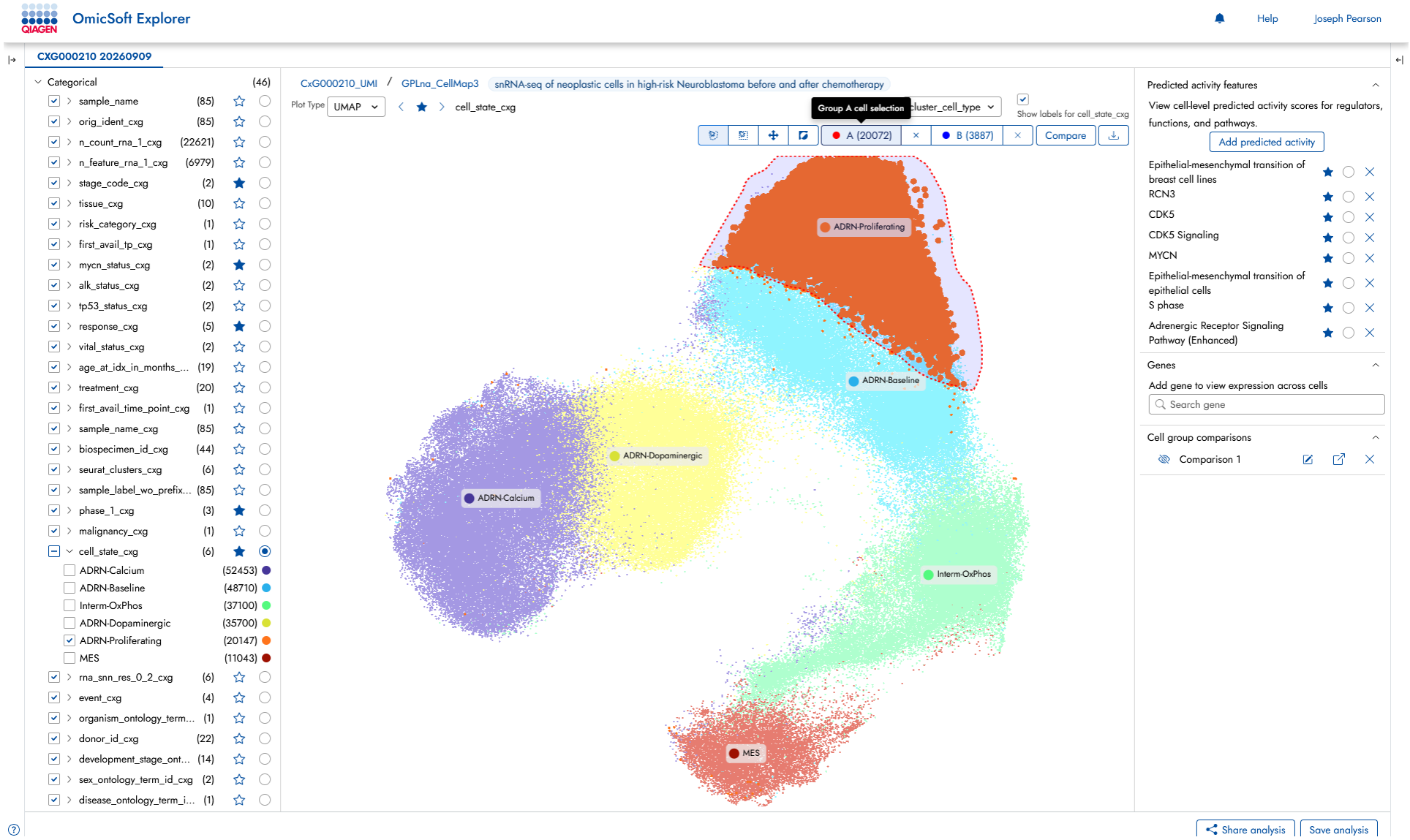
Prioritize predicted activities on global or local dynamics



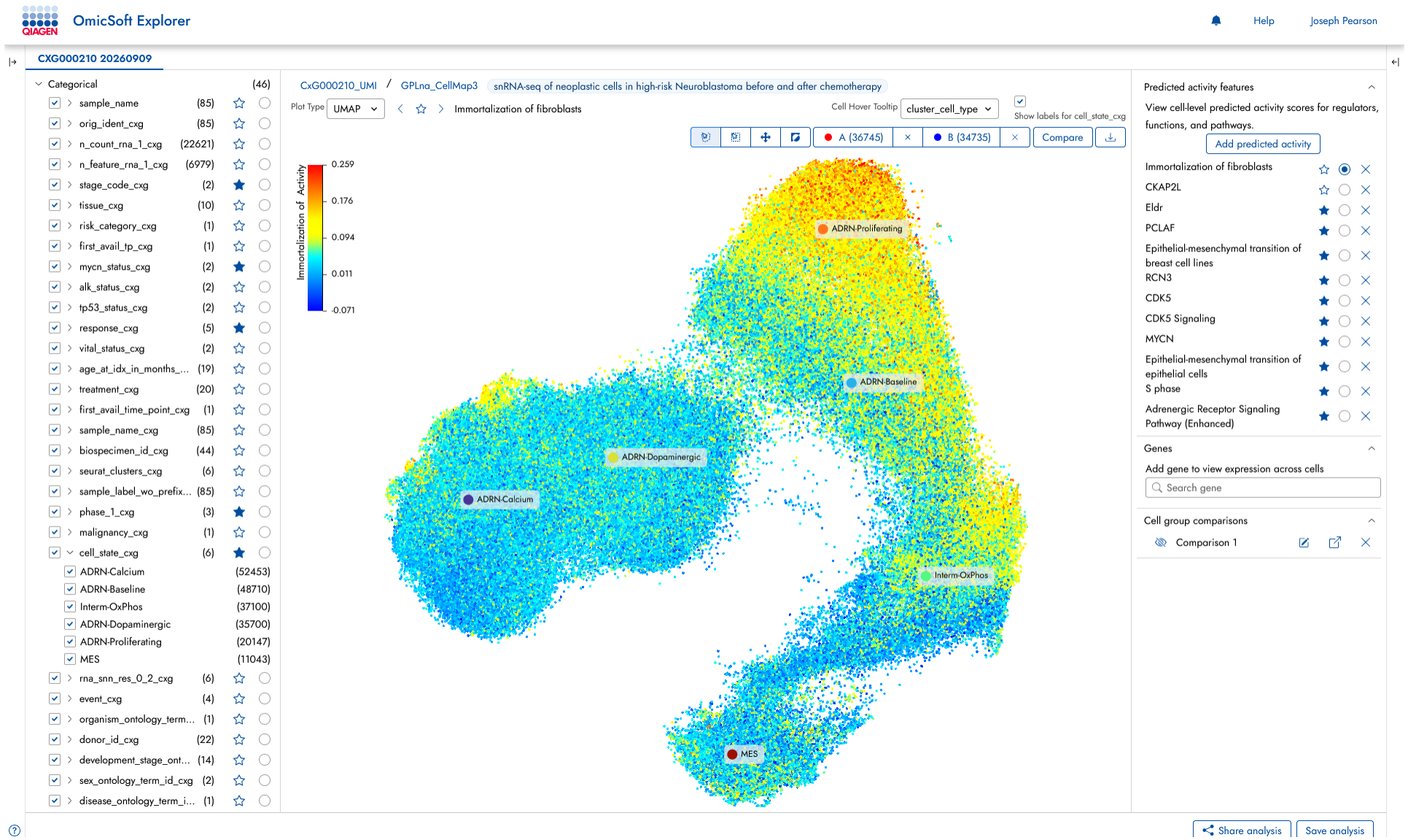
Prioritize predicted activities on global or local dynamics



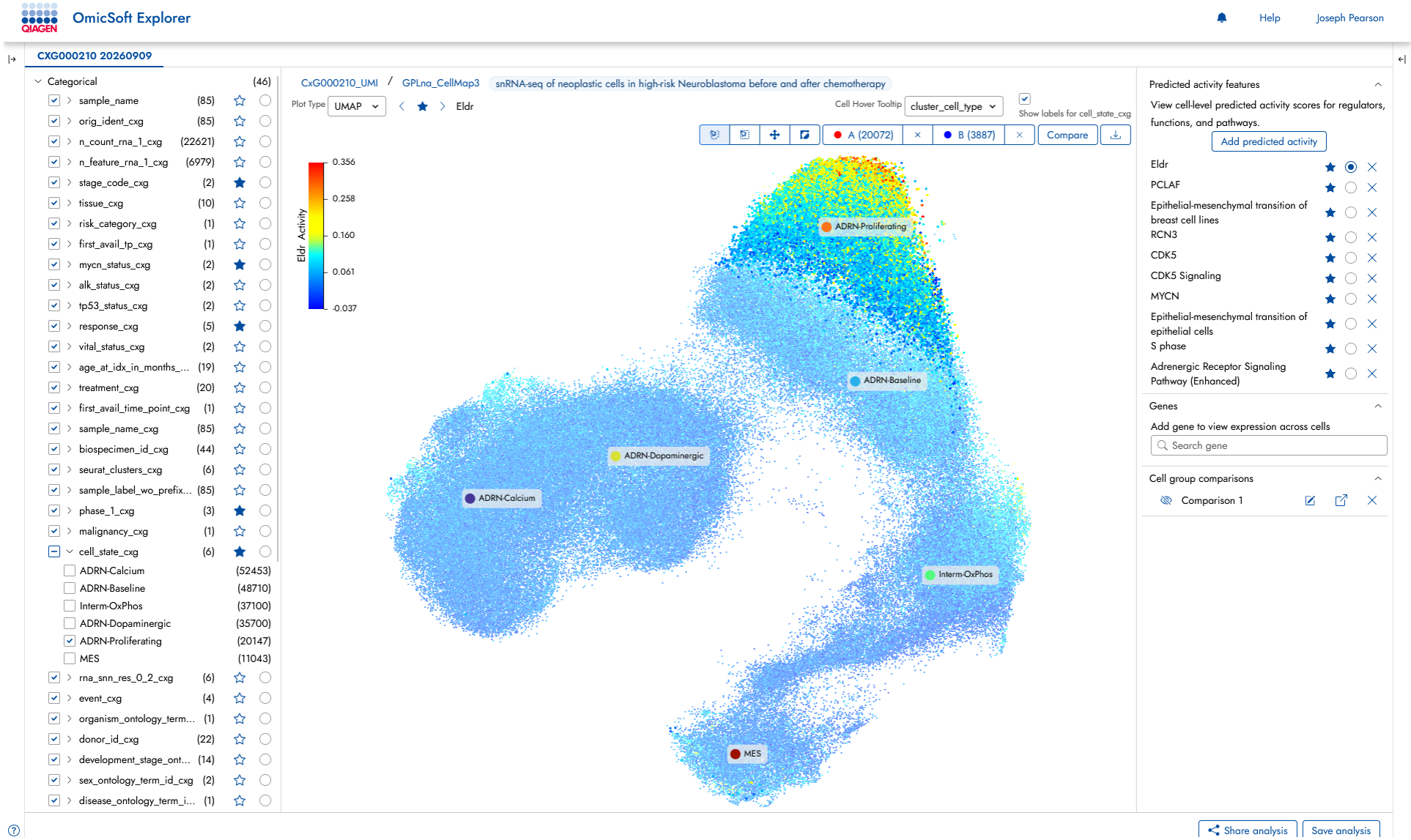
Select any group of cells to analyze predicted activity



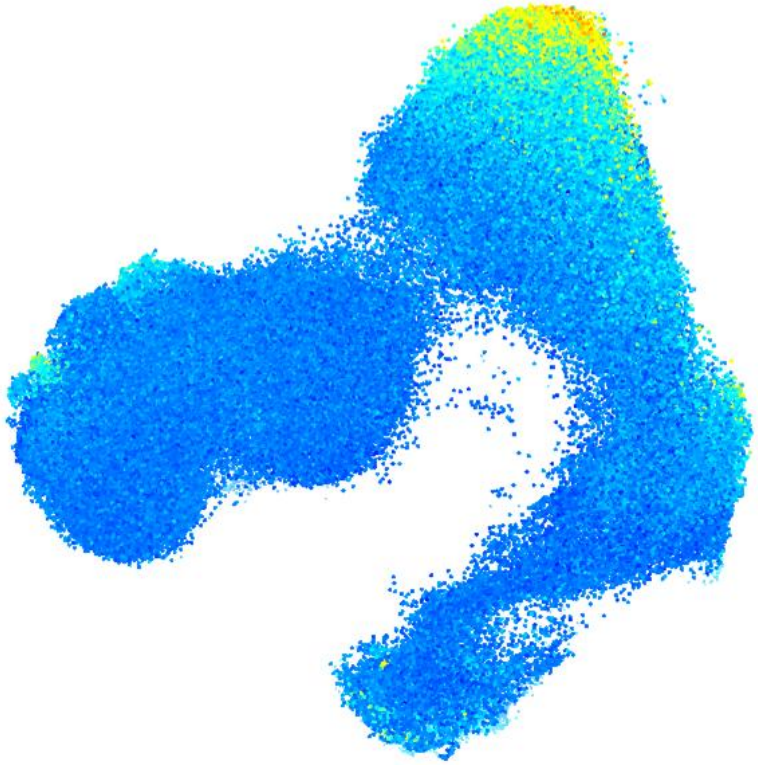
Compare selected cells against others



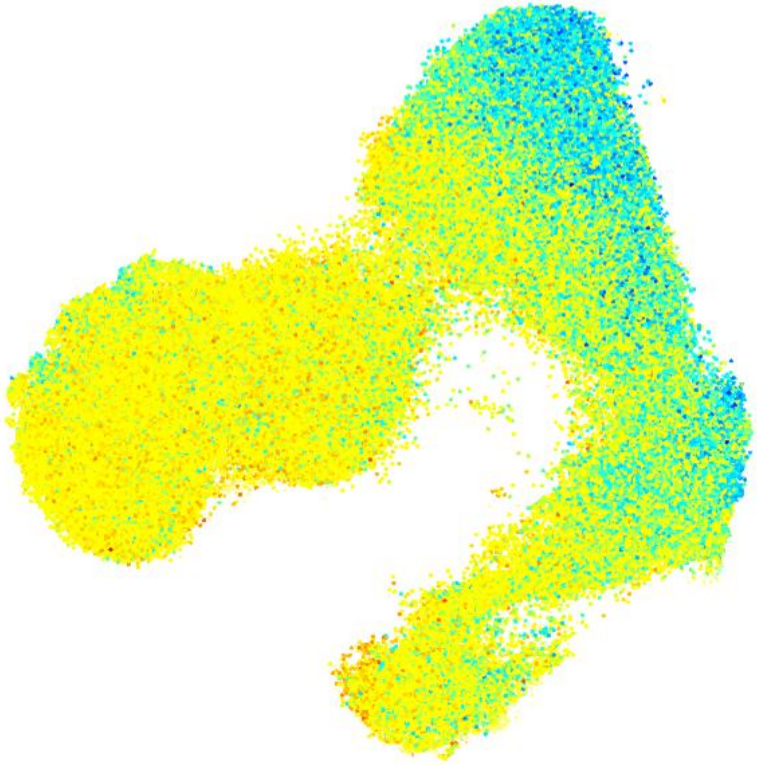
Identify dynamic regulators within any group of cells



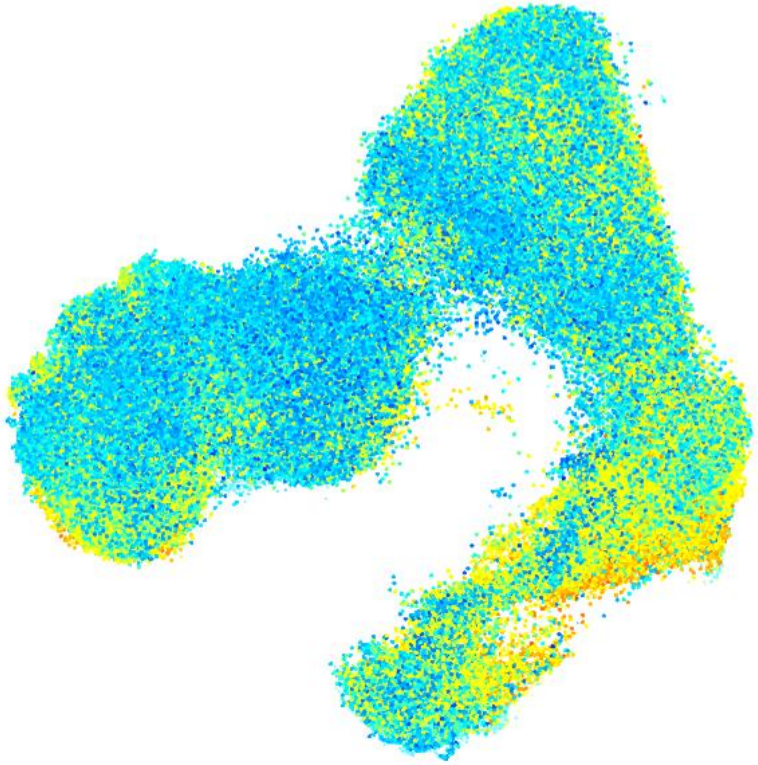
Predicted regulator activity dynamics across neoplastic neuroblasts



CKAP2L activity

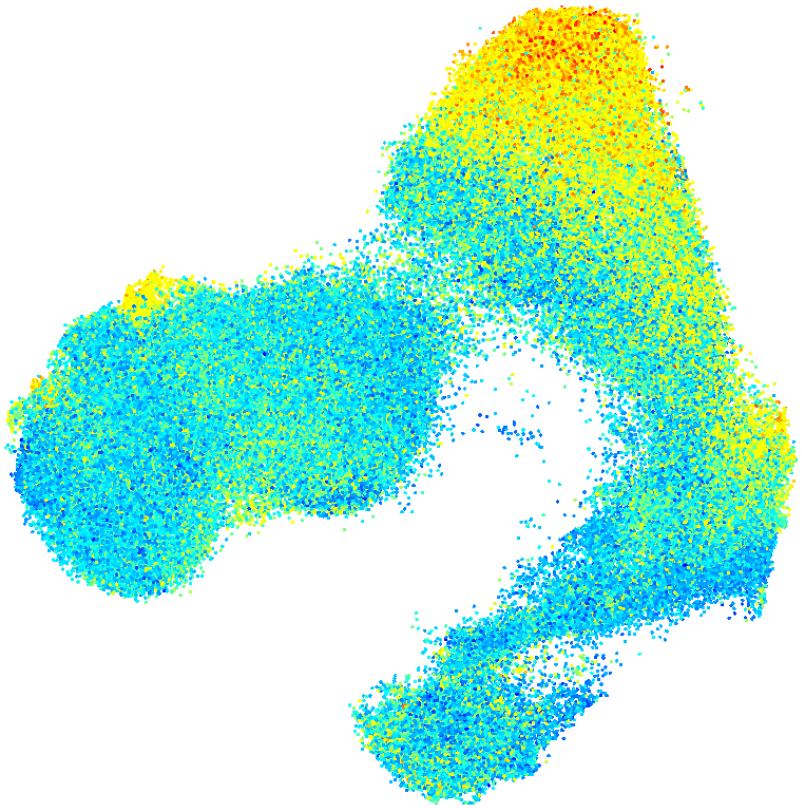


TP53 activity

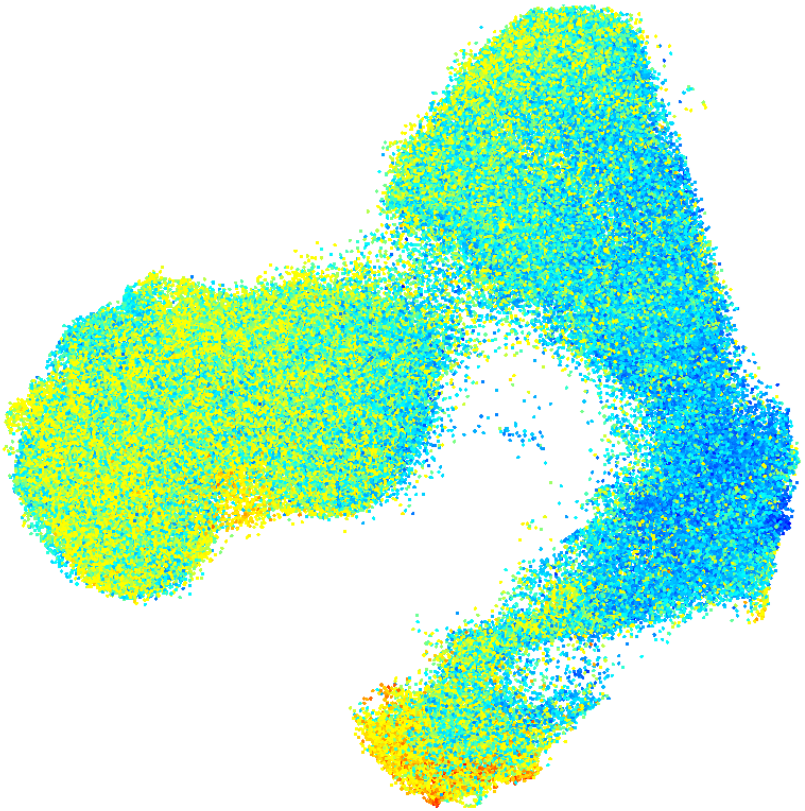


DDX5 activity

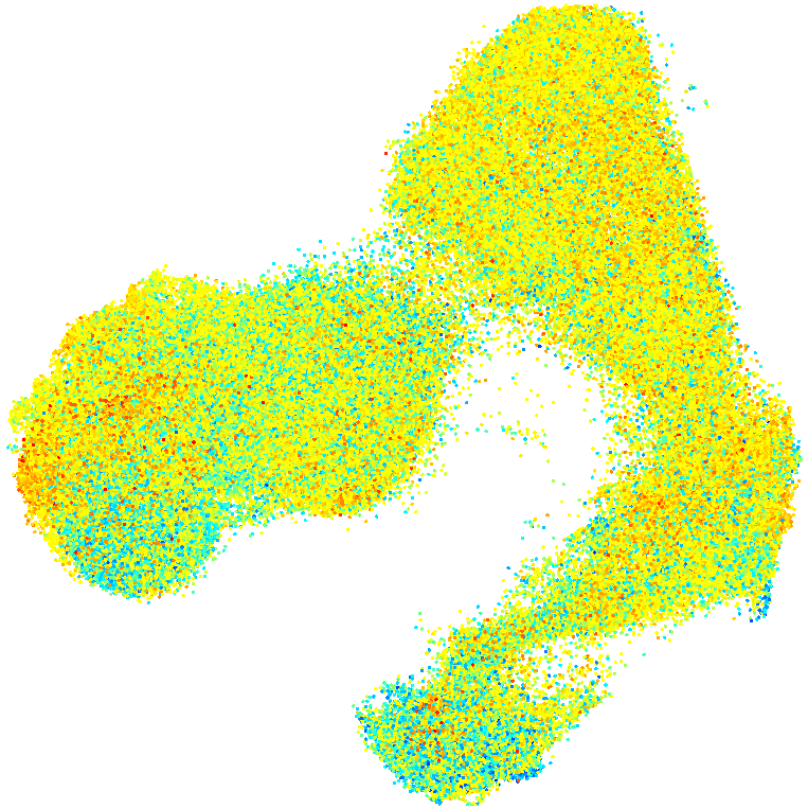
Predicted functional biological mechanisms differentiating neoplastic neuroblast cell groups



S phase

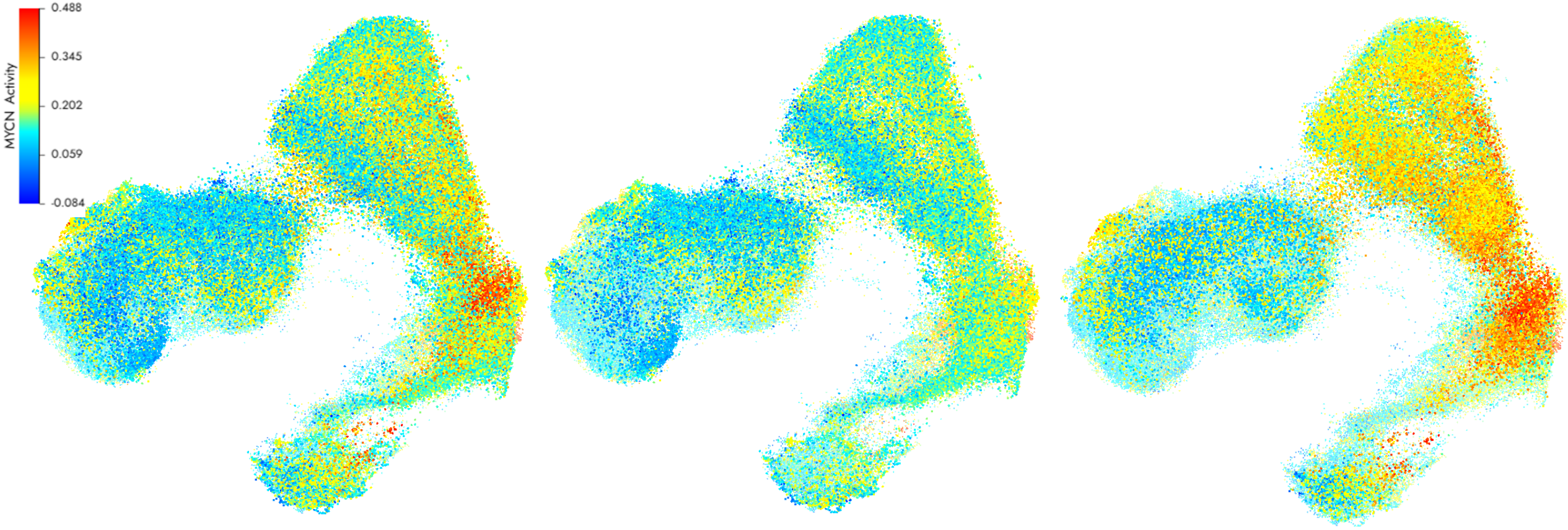


Epithelial-mesenchymal transition



Adrenergic Receptor Signaling Pathway

Cell-level activity prediction of MYCN activity with MYCN amplification, in certain neuroblastoma cell states

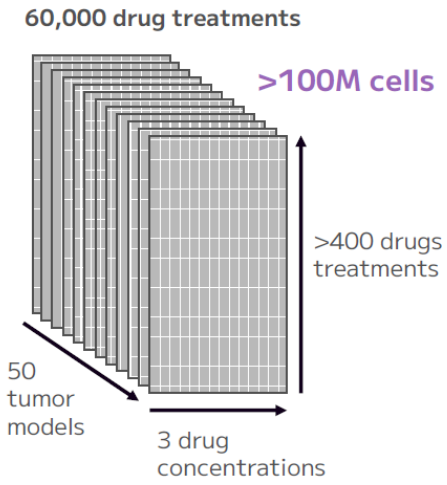


Pre-treatment, all patients

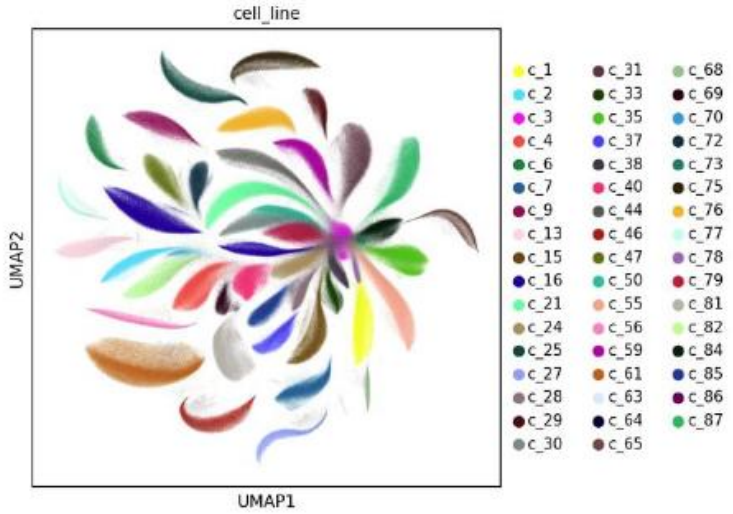
Pre-treatment, patients with no MYCN amplification

Pre-treatment, patients with MYCN amplification

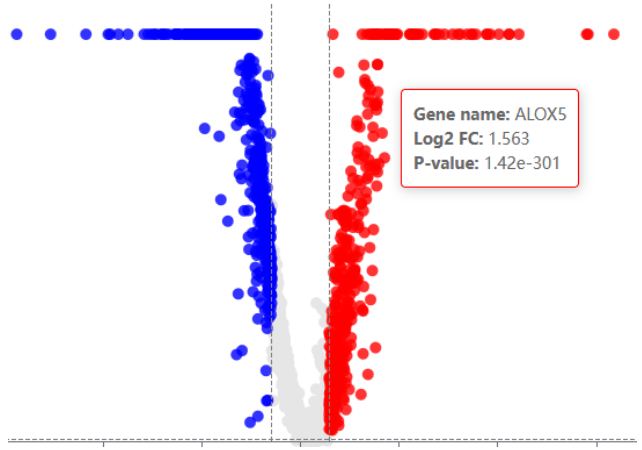
Parse Biosciences makes massive-scale single-cell studies turnkey; IPA turns the data into actionable insights



Metadata curation, unification



Statistical modeling

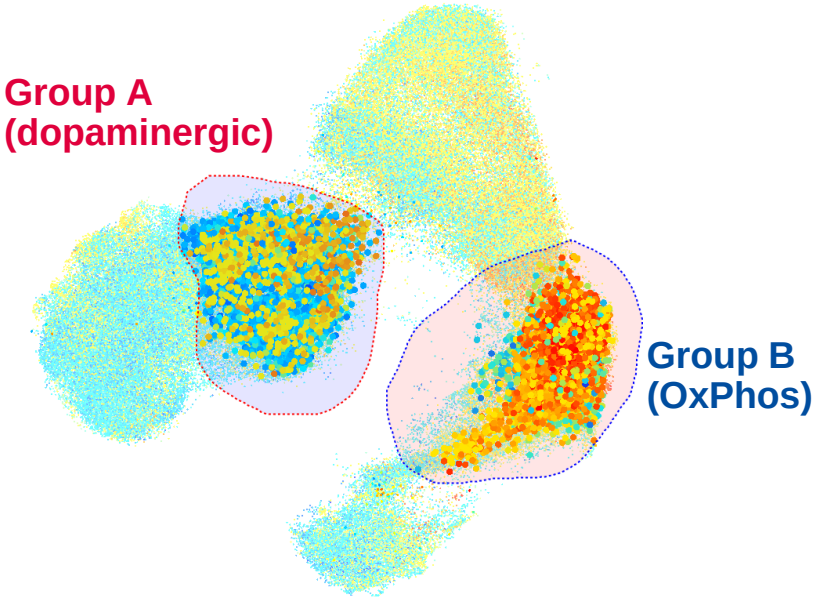


Parse Tahoe-100M experimental setup

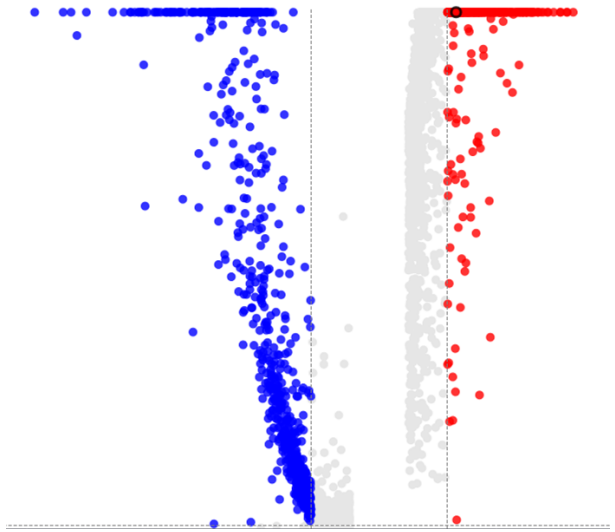
Tahoe-100M data with cell-level expression response profiles

53,000 differential expression profiles reveal transcriptional response for each compound* dose per cell line

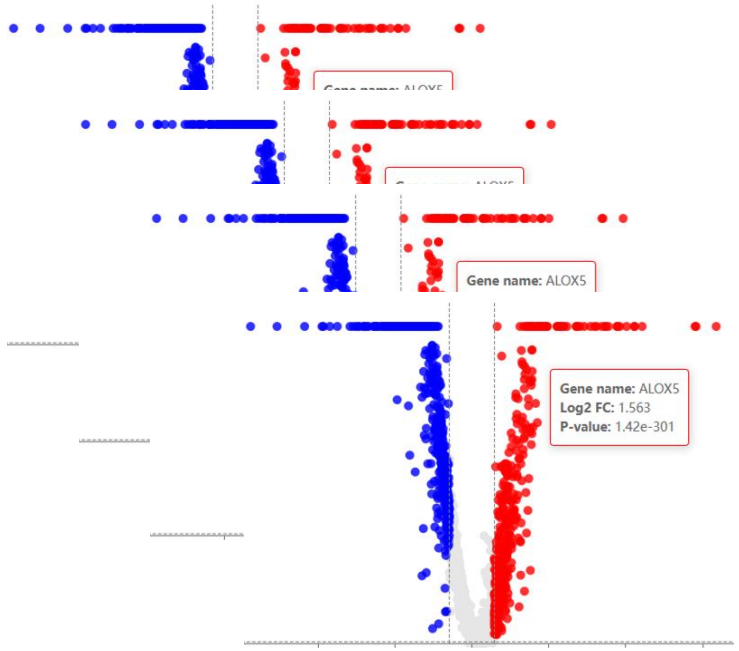
Pre-computed Tahoe-100M compound response profiles can be compared against MYCN-low vs MYCN-high neuroblastoma DEGs



Choose cell groups
(MYCN-amplified neuroblastoma, low MYCN activity
vs high MYCN activity)

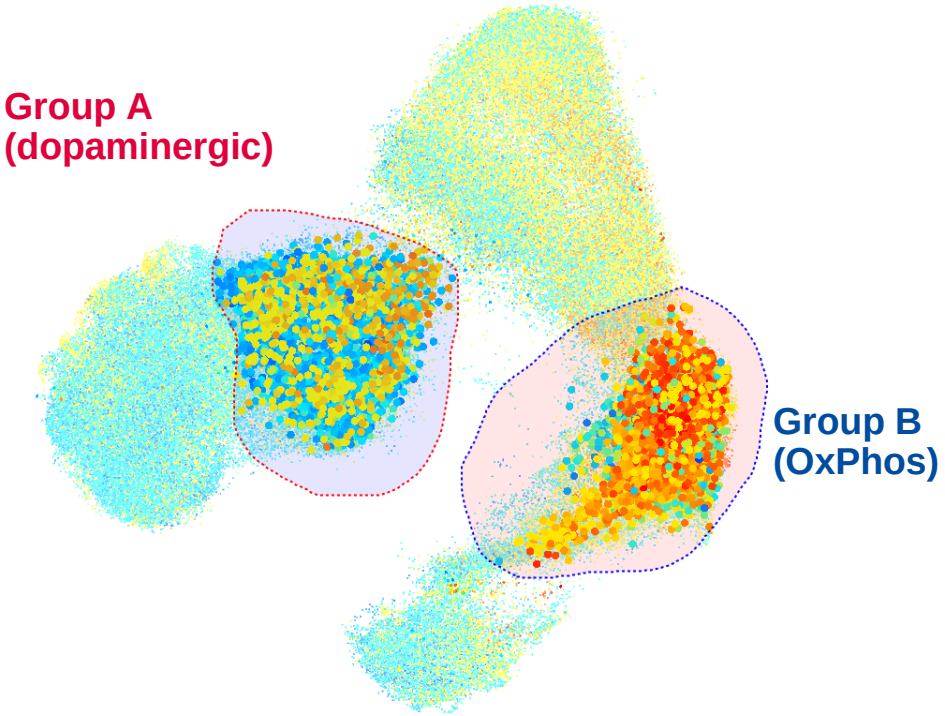


Generate differential gene expression results for genes activated in low MYCN activity vs high MYCN activity



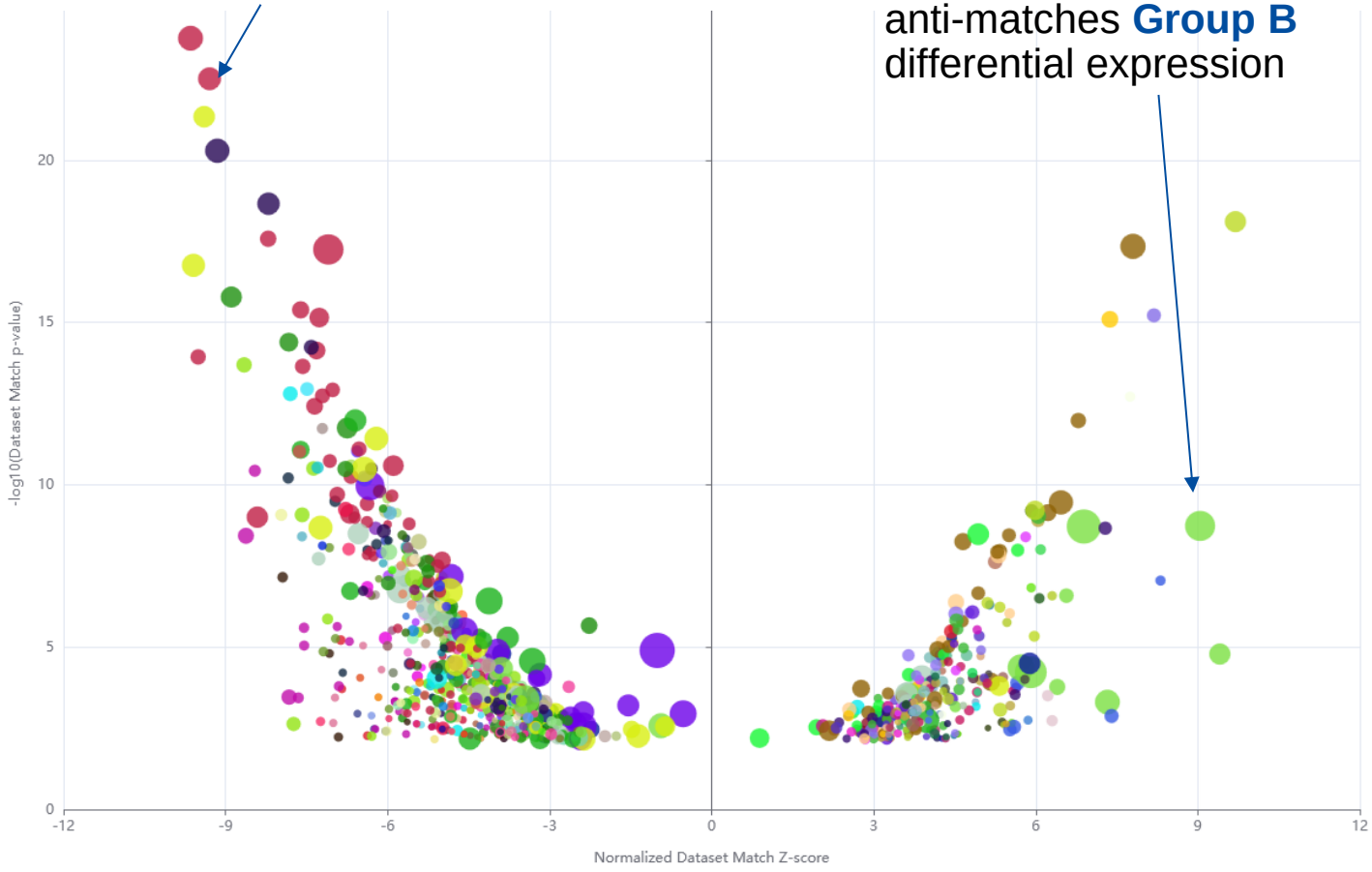
Compare against 53,000 Tahoe-100M differential expression profiles for compound x dose,
per cell line

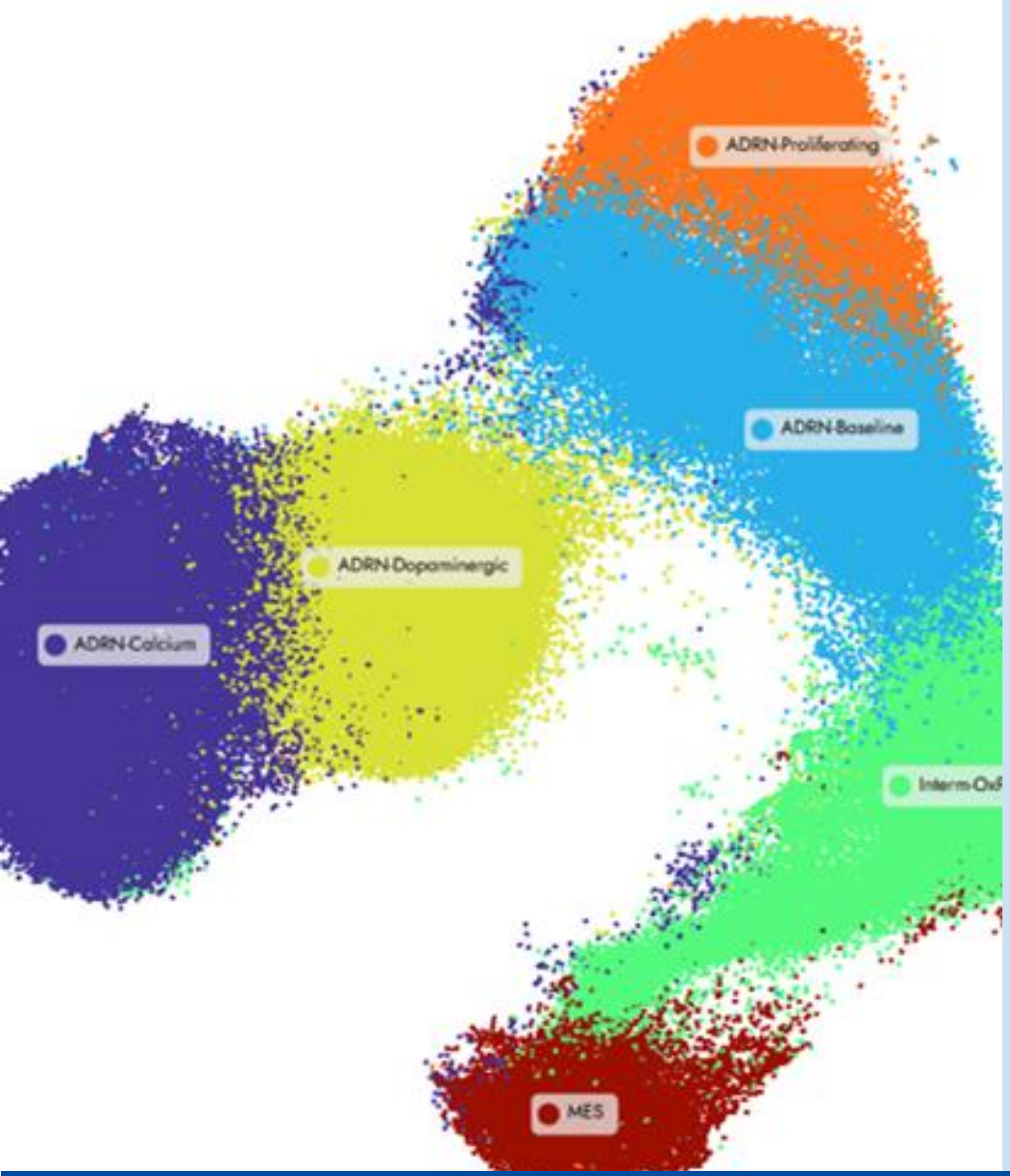
Differential expression signature of MYCN-amplified neuroblastoma groups aligns with Tahoe-100M profiles, including dinaciclib



Dinaciclib treatment response anti-matches **Group A** differential expression

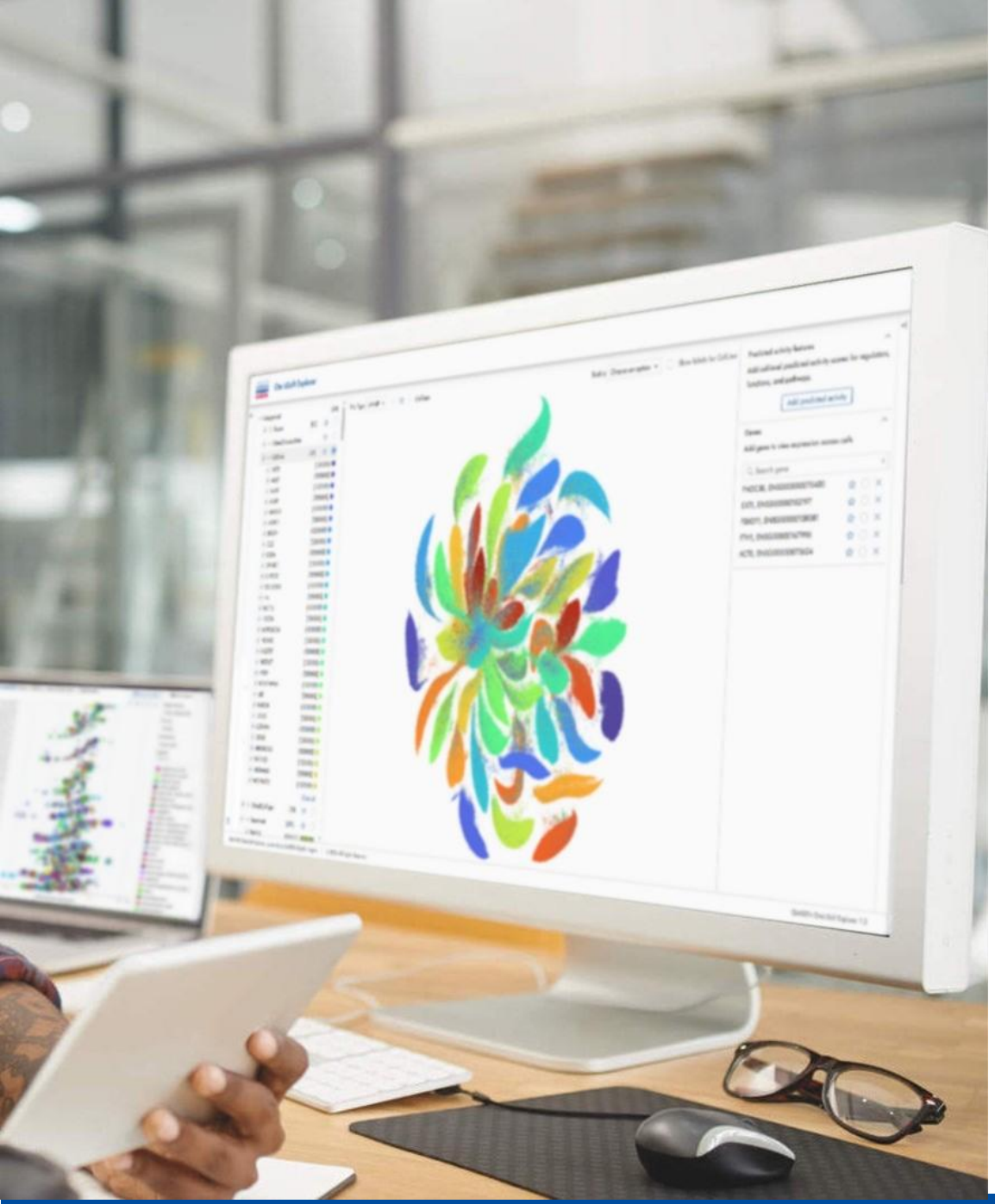
Omacetaxine treatment response anti-matches **Group B** differential expression





Key takeaway

- Marker and signature-based analysis of single cell datasets is effective but limited to well-defined or pre-annotated cell clusters
- Cell-level activity prediction reveals new biological hypotheses that can be quickly explored in pathway interpretation applications like QIAGEN Ingenuity Pathway Analysis
- Contextualization of differential expression results with curated experiments identifies matches for candidate treatments, as well as tissue, cell type, indication expansion, and more



Request a free trial of QIAGEN Ingenuity Pathway Analysis (IPA)

Scan the QR code to sign up