



Curated Single-Cell RNA-Seq for Machine Learning and Artificial Intelligence Applications in Pre-Clinical Biomarker Discovery and Validation in the Immune System

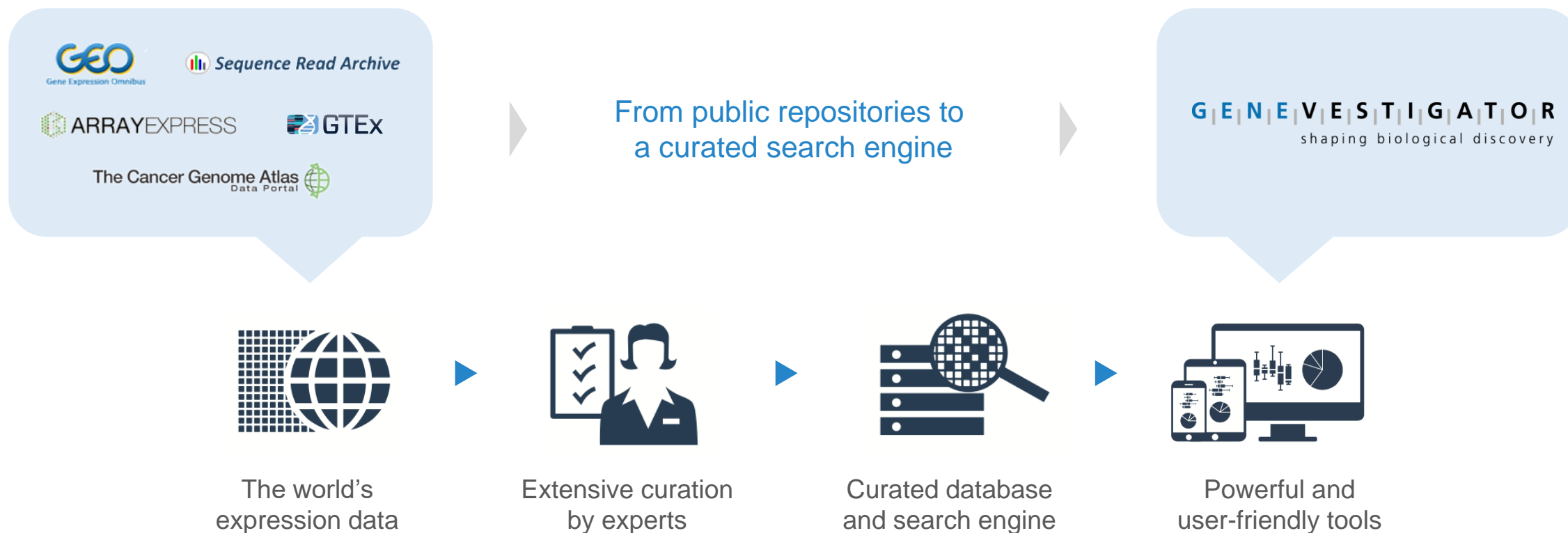
Jana Sponarova, PhD

- CSO and Head of Biocuration at NEBION, an Immunai company

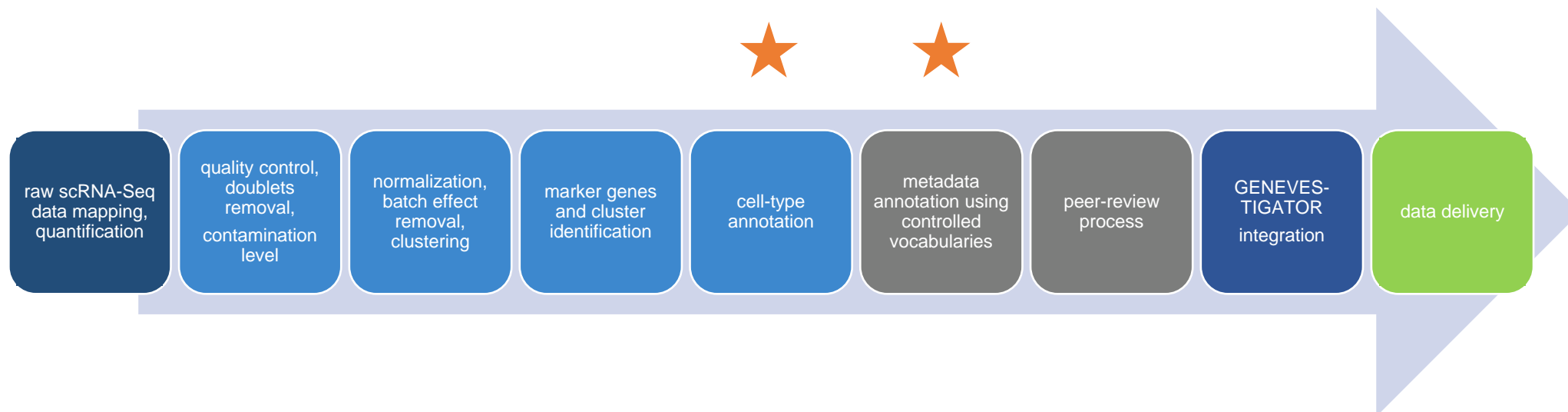
de.NBI 3rd Annual Industrial Forum Meeting 2022

24 November 2022

NEBION Core Workflow



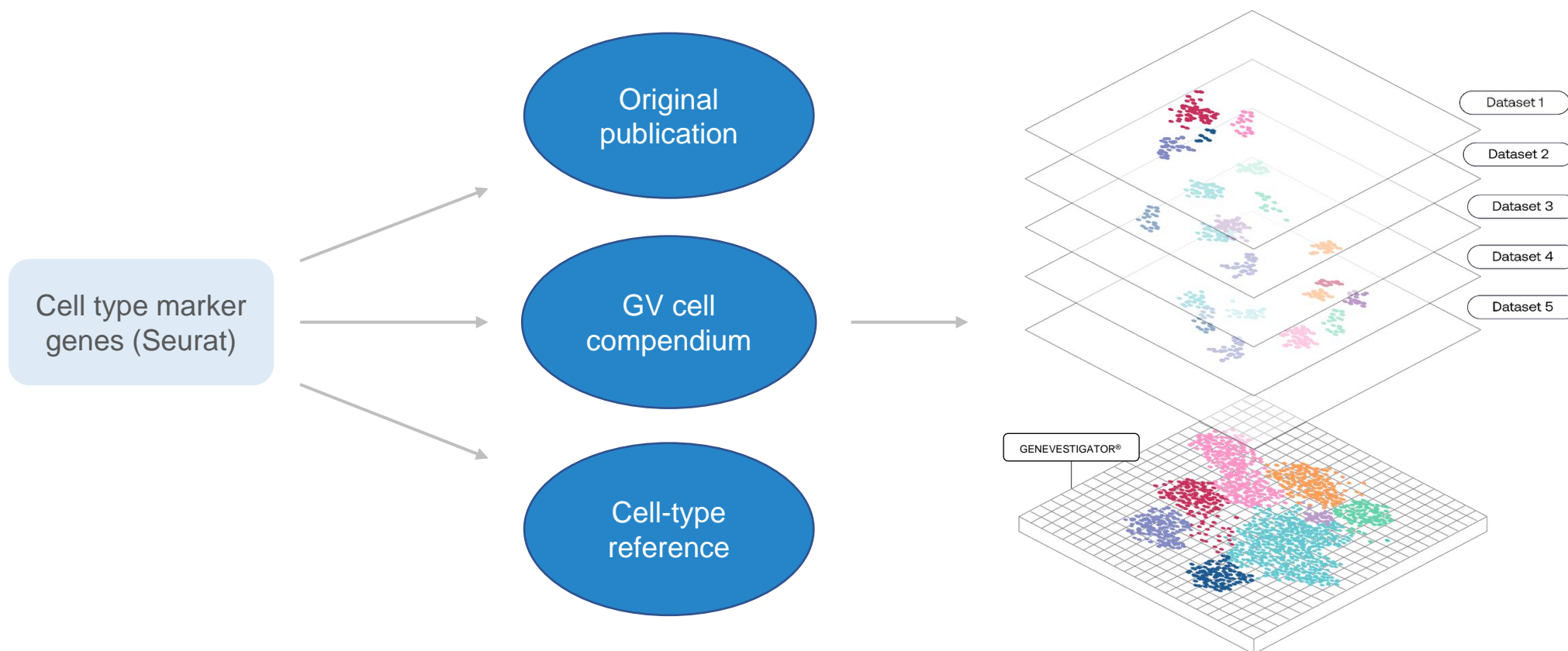
Single-Cell RNA-Seq Data – Curation Workflow



Meta-data enrichment based on publications & suppl. materials

Stimulus	Subject ID	age at assessment (HS)	sex	sample status	disease condition	disease duration	intestinal antiinflammatory agent	Ulcerative colitis (Montreal classification)
normal colonic mucosa cell sample	237	31 years (postnatal)	male	healthy	---	---	---	---
normal colonic mucosa cell sample	206	51 years (postnatal)	male	healthy	---	---	---	---
normal colonic mucosa cell sample	221	23 years (postnatal)	female	healthy	---	---	---	---
normal colonic mucosa cell sample	204	29 years (postnatal)	female	healthy	---	---	---	---
normal colonic mucosa cell sample	219	30 years (postnatal)	female	healthy	---	---	---	---
ulcerative colitis study 79	198	30 years (postnatal)	male	diseased	ulcerative colitis	0.1 years	aminosalicylic acid and similar...	E3 (extensive UC, pancolitis)
ulcerative colitis study 79	238	77 years (postnatal)	male	diseased	ulcerative colitis	8.3 years	---	E2 (left-sided UC)
ulcerative colitis study 79	202	60 years (postnatal)	female	diseased	ulcerative colitis	13.6 years	aminosalicylic acid and similar...	E2 (left-sided UC)
ulcerative colitis study 79	200	67 years (postnatal)	female	diseased	ulcerative colitis	42.2 years	aminosalicylic acid and similar...	E3 (extensive UC, pancolitis)

Cell-Type/ State Identification, Verification and Harmonization

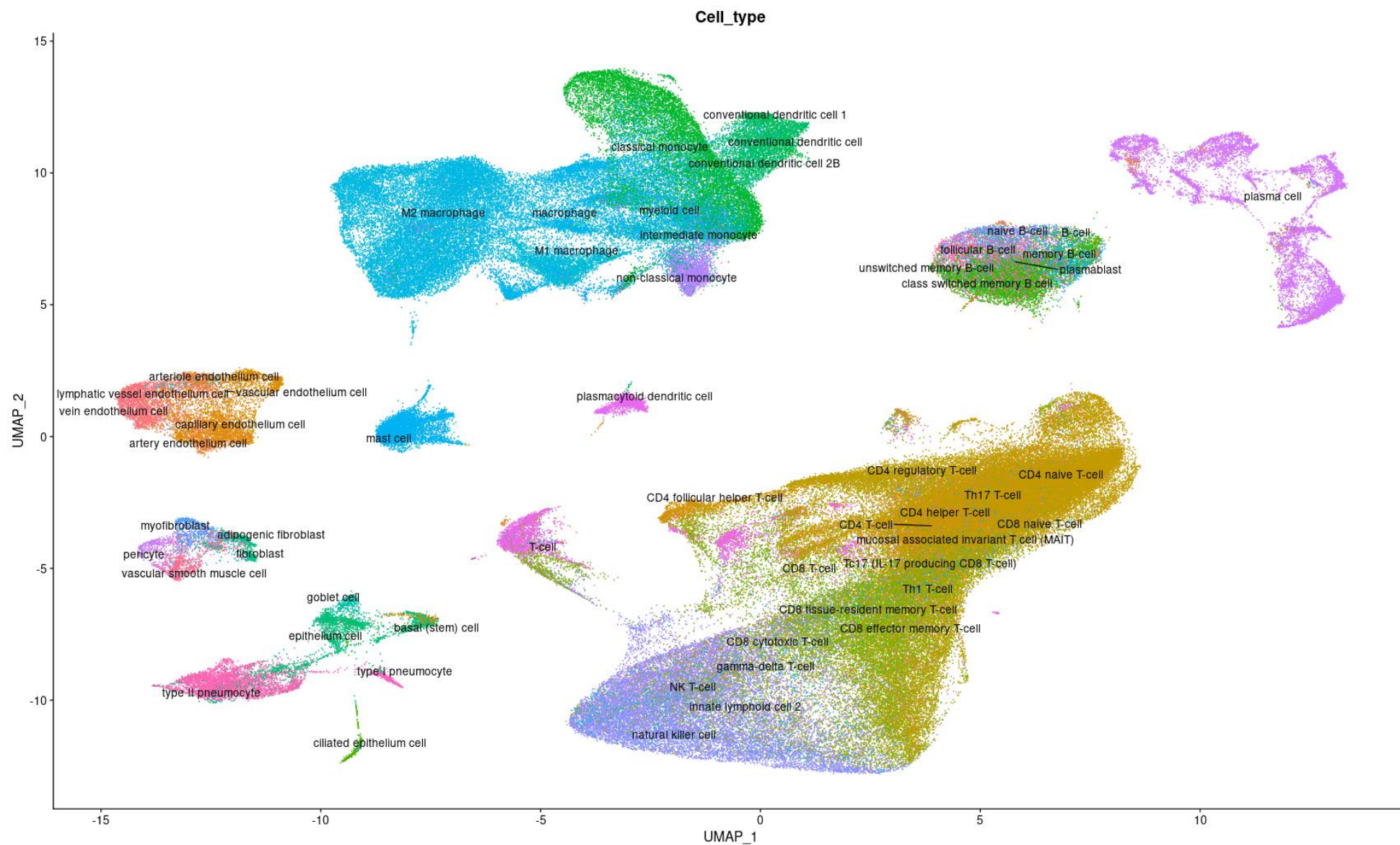


Identification
&
verification of cell types

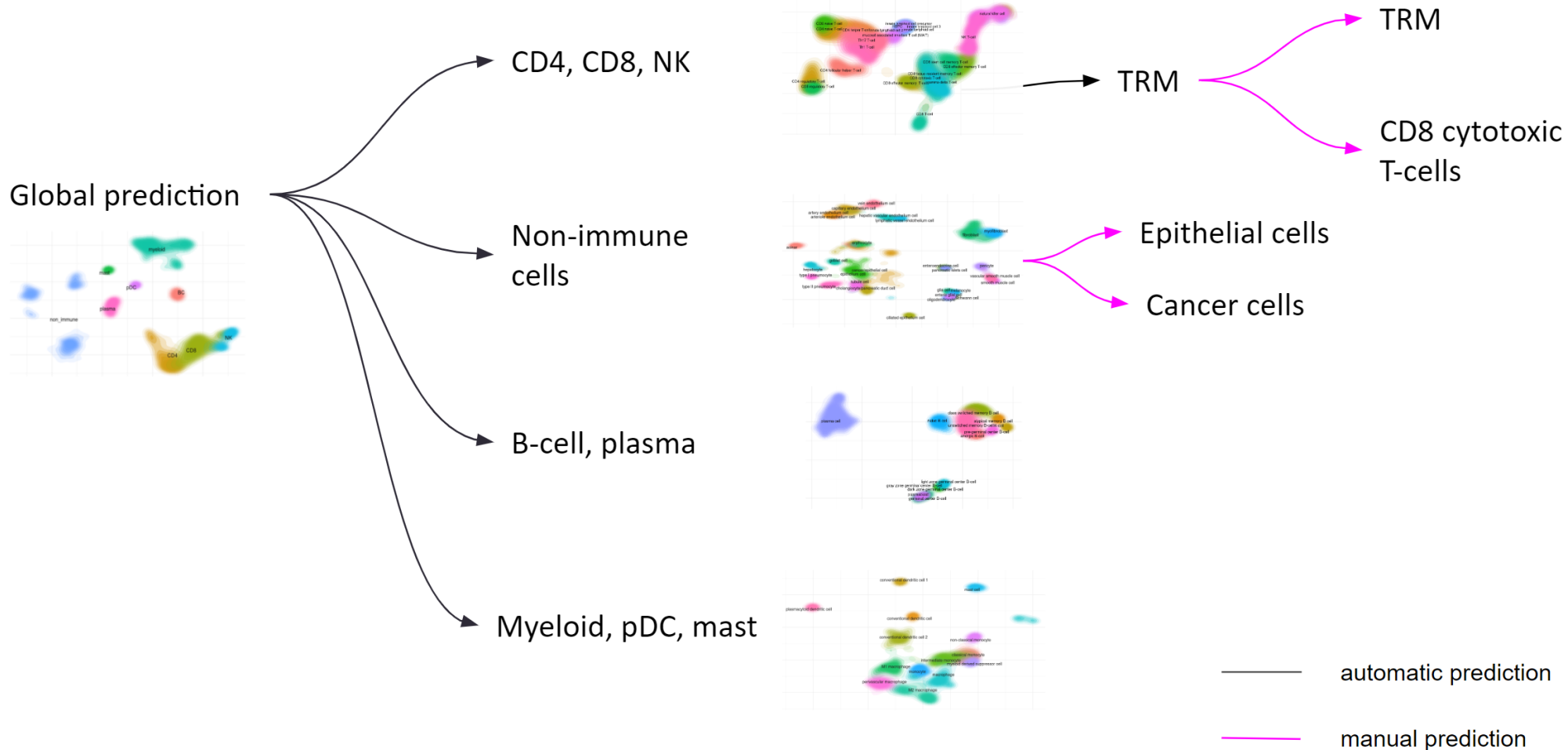
Cell-type harmonization
&
identification of rare cell types

Core expertise!

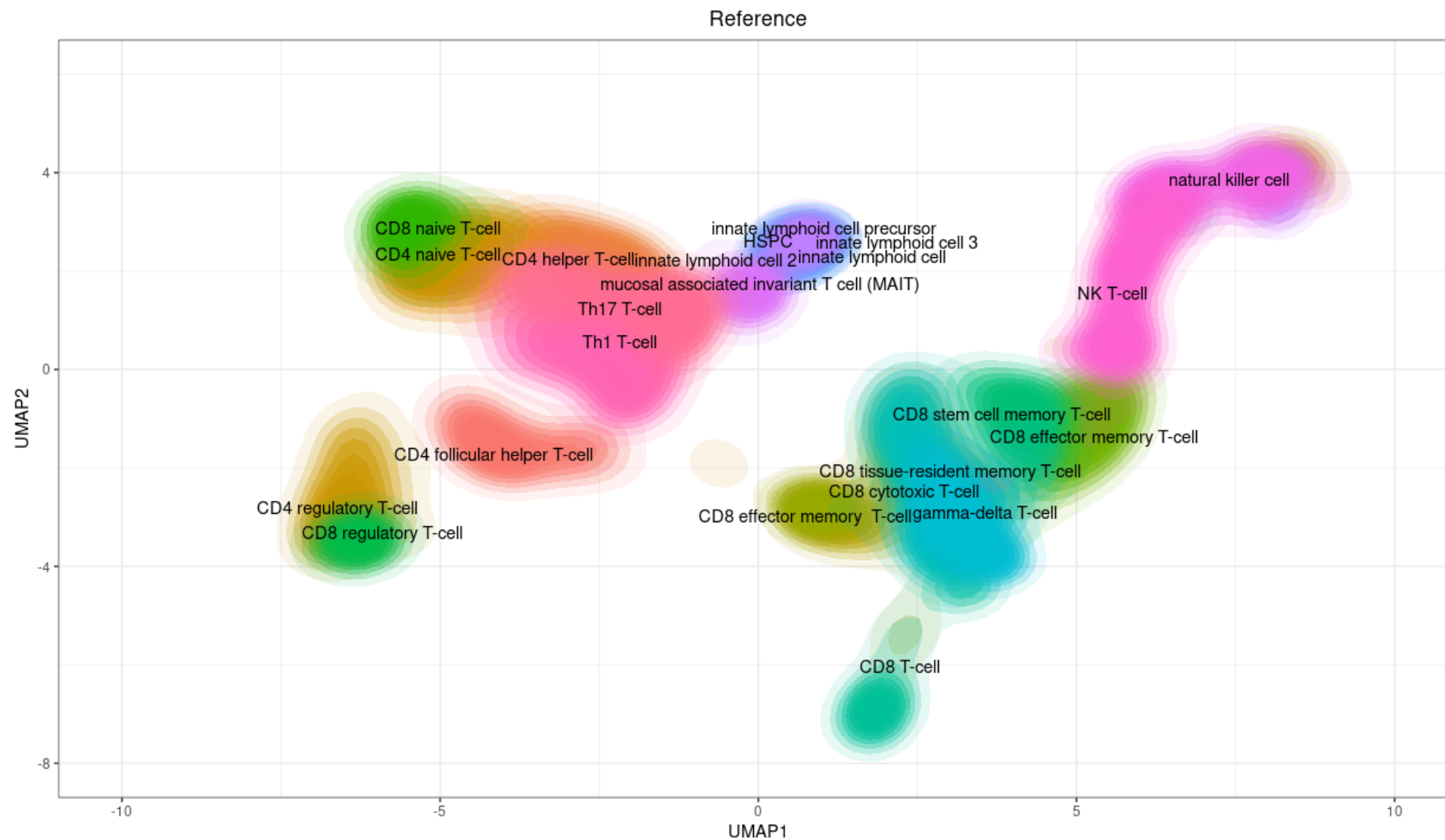
Cell-Type/ State Identification



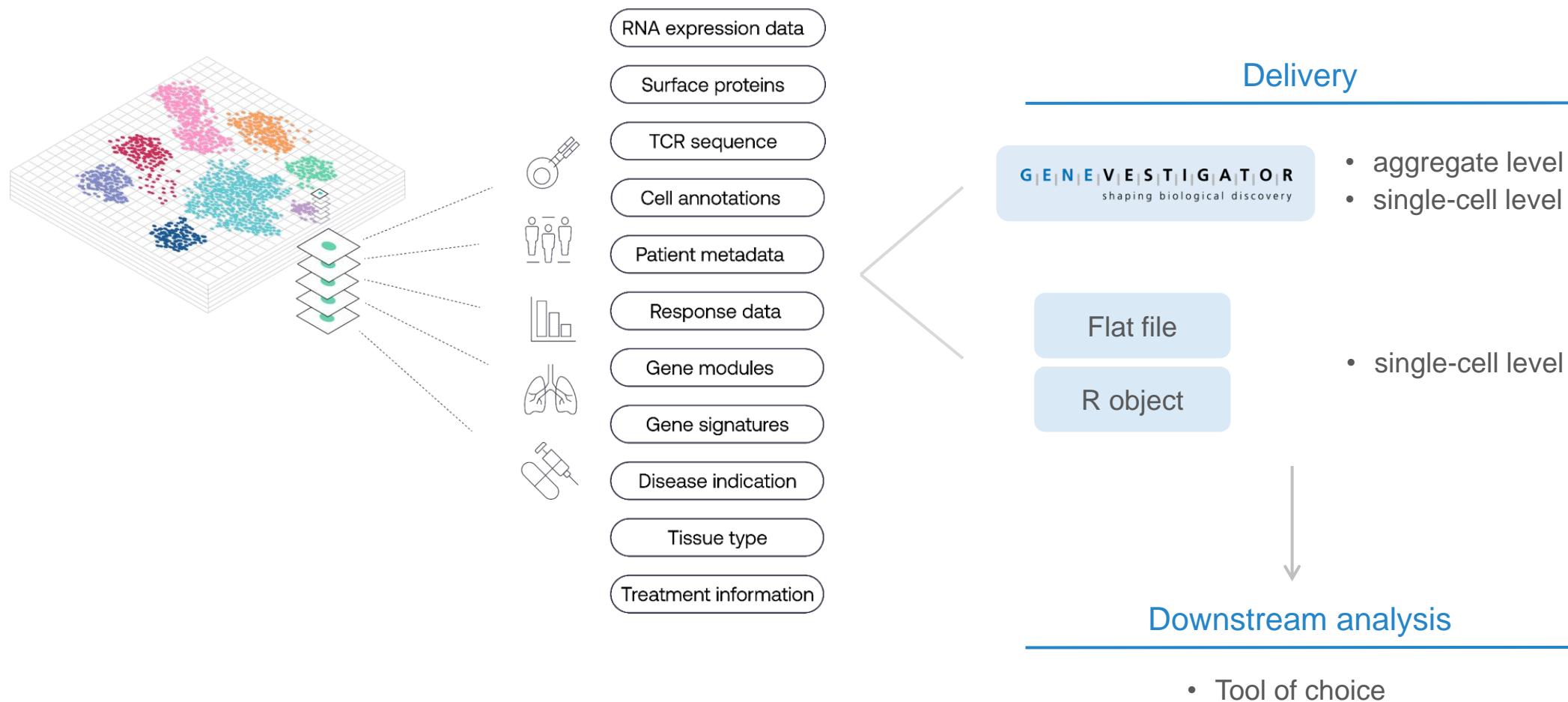
Process of Cell-Type Prediction using Tissue Reference



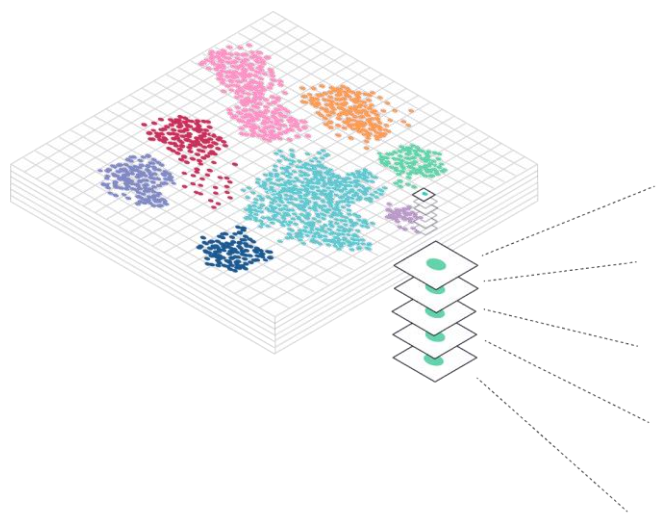
TNK Cell Reference



Curated Studies as Harmonized Compendia



GENEVESTIGATOR® Applications



RNA expression data

Surface proteins

TCR sequence

Cell annotations

Patient metadata

Response data

Gene modules

Gene signatures

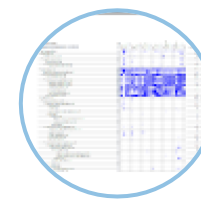
Disease indication

Tissue type

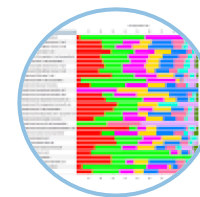
Treatment information



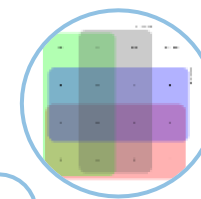
Examine cell signatures



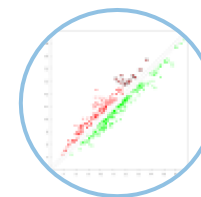
Analyze changes in cell populations



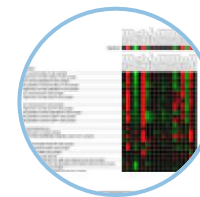
Characterize and prioritize targets & biomarkers



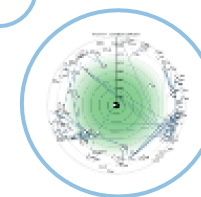
Perform differential expression



Analyze disease signatures



Identify novel targets & biomarkers



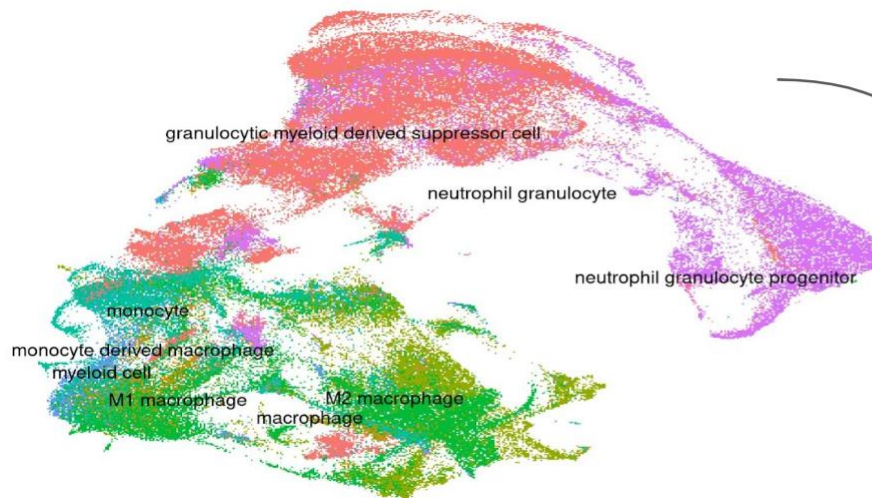
Use Case: Rare Cell Type Identification

► Aim:

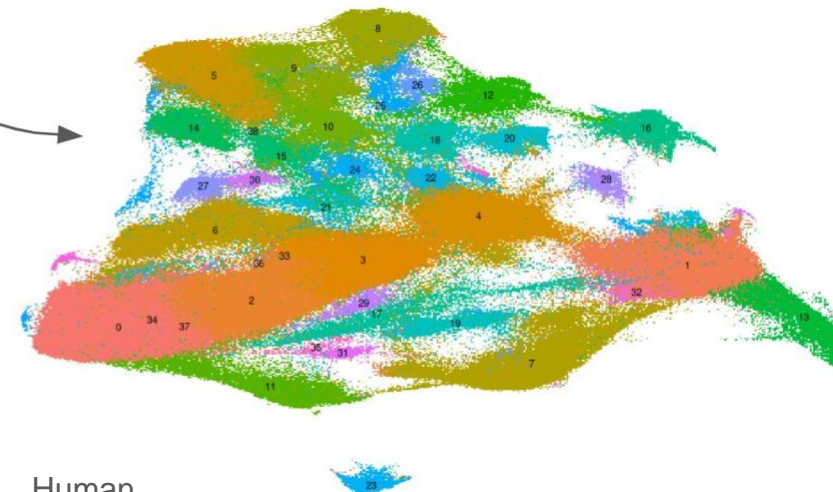
- Identification of monocytic and granulocytic MDSCs
- Derivation of MDSCs signatures
- Integration of GV myeloid cell compendium human and mouse
- Building 2 atlases:
 - Myeloid cell atlas (mouse)
 - Myeloid cell atlas (human)

► Application:

- Identification of different MDSCs in de-novo sequenced mouse/human datasets
- Correlation of this population with disease progression



Mouse
120 752 cells, 18 datasets



Human
587 685 cells, 57 datasets

Summary

- ▶ We have built a manually **curated and globally normalized** scRNA-Seq compendium, mainly consisting of immune cells obtained from studies focused on immuno-oncology, autoimmune diseases, and other therapeutic areas.
- ▶ The **processing** of a study includes raw data mapping, standardized and strict quality control, data normalization, integration, and cell-type/ -state identification.
- ▶ The **cell type annotation is synchronized** across all studies in the compendium.
- ▶ The pipeline outputs are enriched with **sample-level information** (e.g., patient-level data), and data are integrated into a user-friendly analysis software – GENEVESTIGATOR®, a high-performance visualization tool for gene expression data.
- ▶ This **deeply harmonized** compendium represents an important asset for downstream ML and AI applications in pre-clinical **biomarker discovery and validation**.

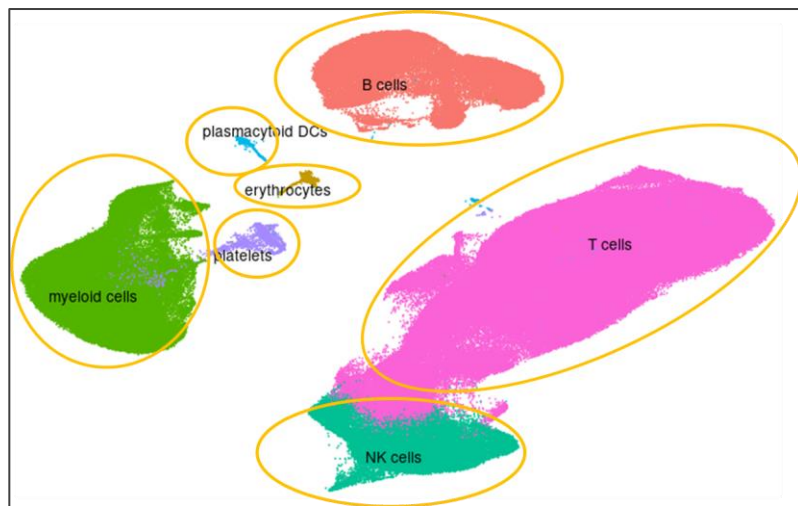


Thank you for your attention!

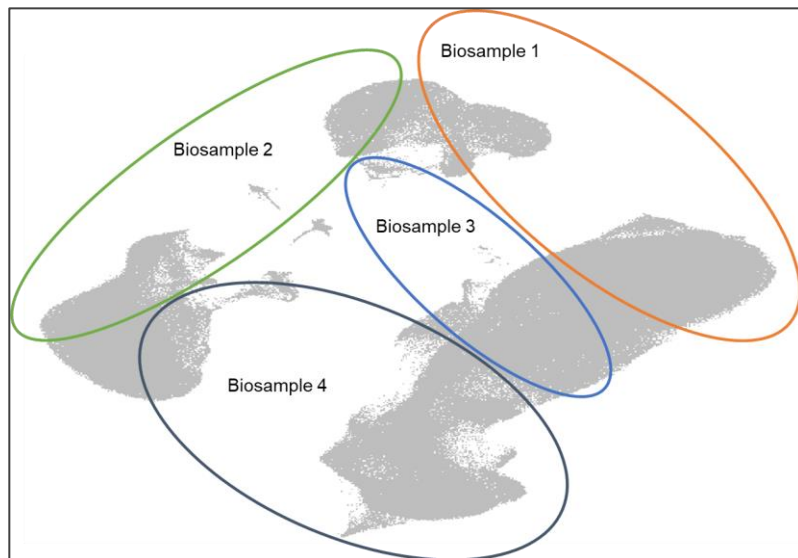


Aggregation From Cell to Cell-Type/ State/ Treatment Level

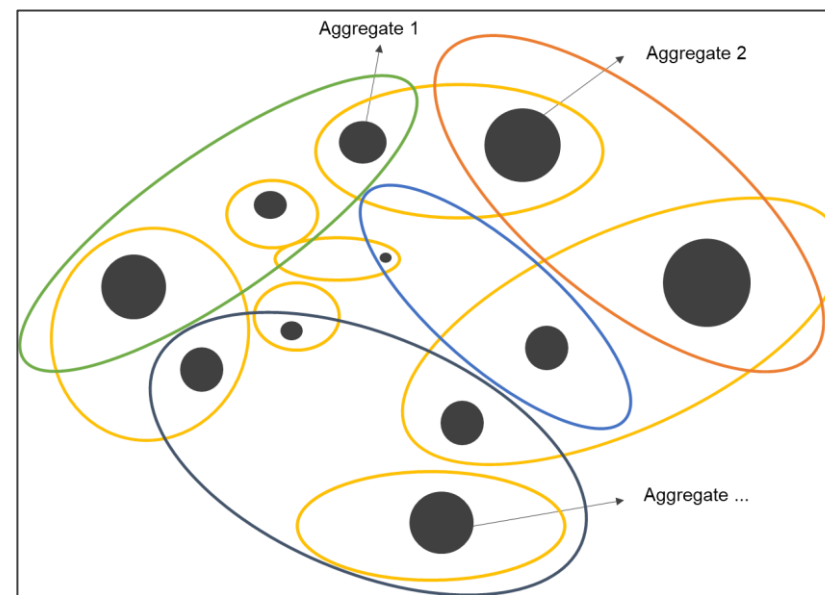
Cell-Type/ -State
Annotation



Biosample Annotation
(subject, disease, treatment)



GENEVESTIGATOR
shaping biological discovery



→ Storage of signal data in the aggregate format enables fast calculations