# Learning Single-Cell Perturbation Responses Using Neural Optimal Transport

392232 Advanced Artificial Intelligence in Biomedicine

Fatih Altundas

# Overview

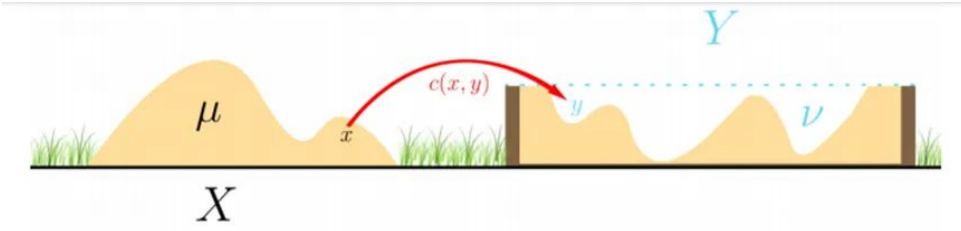
1. Introduction
2. Optimal Transport Theory
3. Model
4. Applications and Results

# Single-Cell Perturbations

- chemical or genetic effects can influence the phenotypes of cells and altering their functions
- heterogeneity of cells makes predictions of cellular responses difficult
- data from treated and control cells:
  - Iterative Indirect Immunofluorescence Imaging (4i)
    - measure the abundance and localization of proteins
    - antibodies tagged with a fluorescent dye
  - single-cell RNA sequencing
- problem: observations are unpaired

# Optimal Transport Theory

**Problem**: how to transport one distribution to another by minimizing the cost?




**Definition (transport map)**:

A measurable map $T : X \to Y$ is said to transport a probability measure $\mu \in \mathcal{P}(X)$ to a probability measure $\nu \in \mathcal{P}(Y)$ if:

$$\nu(B) = \mu(T^{-1}(B)) \quad \text{for all measurable sets } B \subseteq Y$$

where $T^{-1}(B) = \{x \in X : T(x) \in B\}$ is the preimage of $B$ under $T$.

**Notation**: $T_{\#\mu=\nu}$

**Monge's Optimal Transport Problem**:

$$\arg\min_{T:T_\sharp\mu=\nu} \mathbb{E}_{X\sim\mu} \| X - T(X)\|_2^2$$

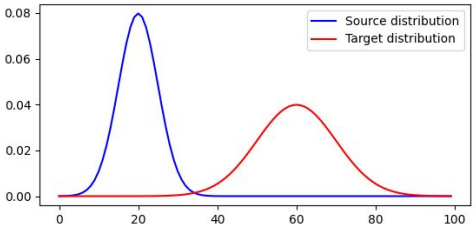
# Optimal Transport Theory

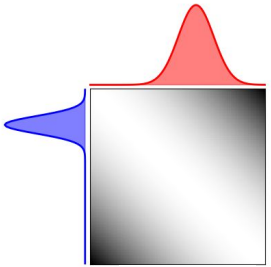
Monge's Problem is not always solvable

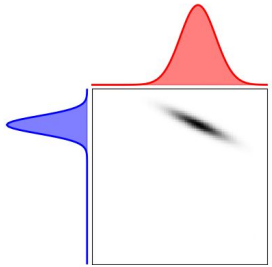
**Solution**: Kantorovich-Problem, which is a relaxation of Monge's Problem

$$W(\mu,\nu) = \min_{\gamma\in\Gamma(\mu,\nu)} \mathbb{E}_{(X,Y)\sim\gamma} \| X-Y\|_2^2,$$

$$\Gamma \text{ is the space of all probability measures on } X\times Y, \text{ with}$$
$$\gamma(A\times Y) = \mu(A) \text{ and } \gamma(X\times B) = \nu(B) \text{ for compact } A\subset X, B\subset Y$$





cost matrix


optimal transport plan

# Optimal Transport Theory

- dual of Kantorovich is defined as:

$$W(\mu,\nu) = \max_{(g,f)\in\Phi_c} \mathbb{E}_\mu[g(x)] + \mathbb{E}_\nu[f(y)]$$

$$\Phi_c := \left\{(g,f) \in L^1(\mu)\times L^1(\nu) : g(x)+f(y) \le \tfrac{1}{2}\,\|x-y\|_2^2, \forall (x,y) d\mu\otimes d\nu \text{ a.e.}\right\}$$

- the dual is constrained and concave and can further be simplified to:

$$W(\mu,\nu) = \underbrace{\frac{1}{2}\mathbb{E}\left[\|x\|_2^2 + \|y\|_2^2\right]}_{\mathcal{C}_{\mu,\nu}} - \min_{f\in\tilde{\Phi}} \mathbb{E}_\mu[f^*(x)] + \mathbb{E}_\nu\left[f(y)\right]$$

$$\text{where } \tilde{\Phi} \text{ is the set of all convex functions in } L^1(d\mu)\times L^1(d\nu)$$

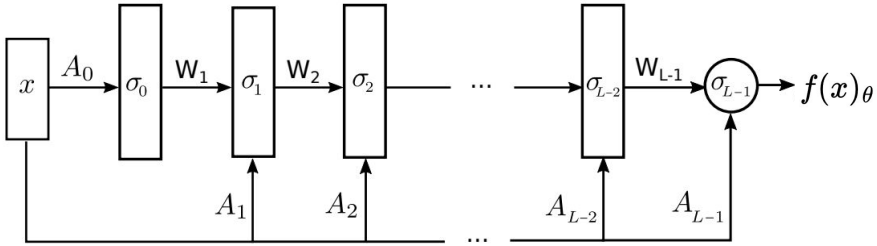
$$f^*(x) = \max_y \langle y,x\rangle - f(y)$$

- further approximated (min-max formulation):

$$W(\mu,\nu) = \max_\theta \min_\phi \mathcal{C}_{\mu,\nu} - \mathbb{E}_\mu\left[\langle x, \nabla g(x)_\phi\rangle - f(\nabla g(x)_\phi)_\theta\right] - \mathbb{E}_\nu\left[f(y)_\theta\right]$$

# Input Convex Neural Networks (ICNN)

- based on feed-forward networks
- function $x \to f(x)_\theta \in \mathbb{R}$ with $\theta = (W_l, A_l, b_l)$ is convex if:
  - activation functions are convex and non-decreasing
  - $W_l$ is non-negative

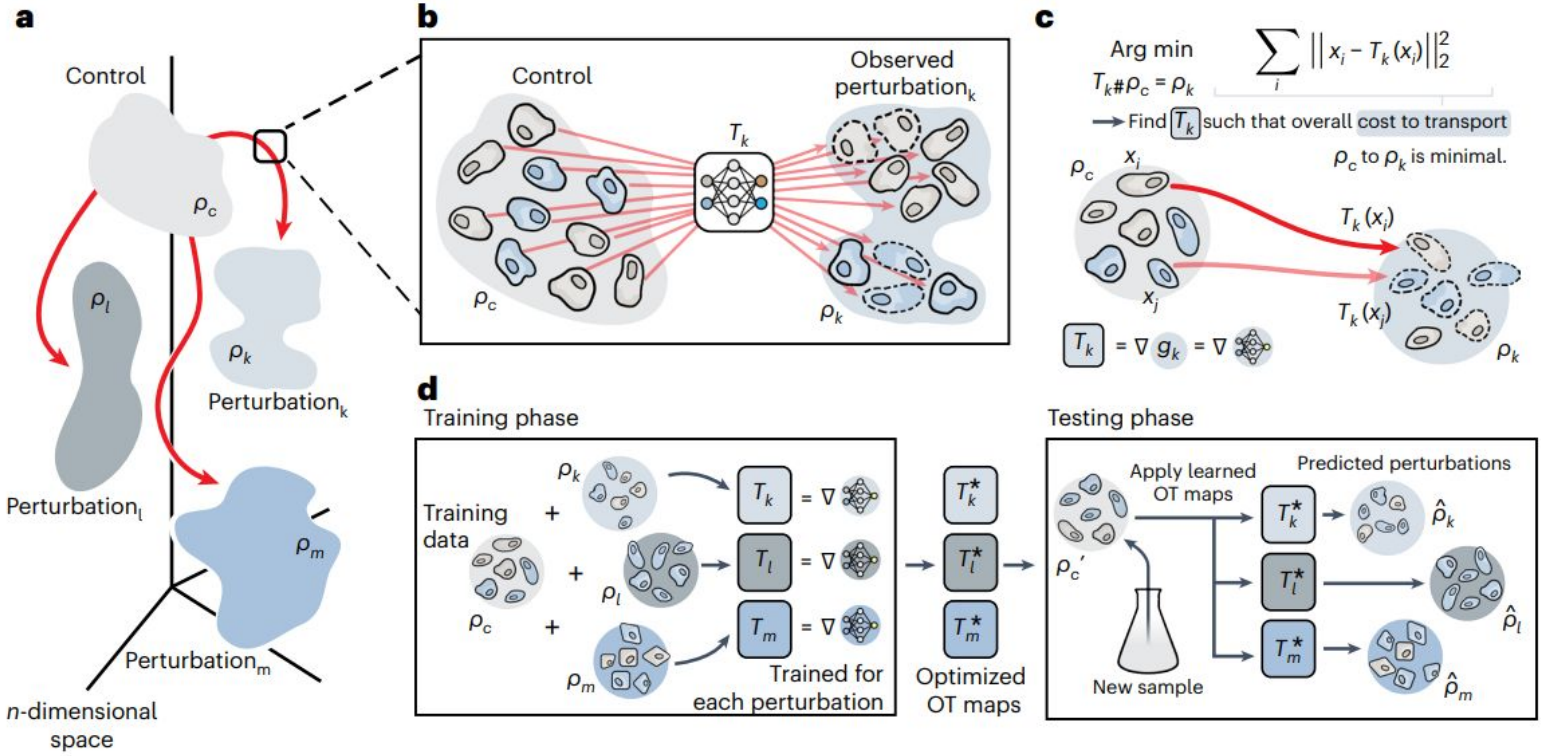



$$W(\mu,\nu) = \max_{\theta} \min_{\phi} c_{\mu,\nu} - \mathbb{E}_\mu\left[\langle x, \nabla g(x)_\phi\rangle - f(\nabla g(x)_\phi)_\theta\right] - \mathbb{E}_\nu\left[f(y)_\theta\right]$$

- optimized using two ICNNs (for $\theta$ and $\phi$)
- the optimal transport map is:

$$T = \nabla g$$

# CellOT


a
Control
$\rho_c$
$\rho_l$
$\rho_k$
Perturbation$_k$
Perturbation$_l$
$\rho_m$
Perturbation$_m$
$n$-dimensional space
b
Control
Observed perturbation$_k$
$T_k$
$\rho_c$
$\rho_k$
c
Arg min
$\sum_i ||x_i - T_k(x_i)||_2^2$
$T_k \# \rho_c = \rho_k$
Find $T_k$ such that overall cost to transport $\rho_c$ to $\rho_k$ is minimal.
$x_i$
$x_j$
$T_k(x_i)$
$T_k(x_j)$
$T_k = \nabla g_k = \nabla$
d
Training phase
Training data
$T_k = \nabla$
$T_l = \nabla$
$T_m = \nabla$
Trained for each perturbation
$T_k^*$
$T_l^*$
$T_m^*$
Optimized OT maps
Testing phase
Apply learned OT maps
Predicted perturbations
$\rho_c'$
New sample
$\hat{\rho}_k$
$\hat{\rho}_l$
$\hat{\rho}_m$

# Evaluation Methods and Metrics

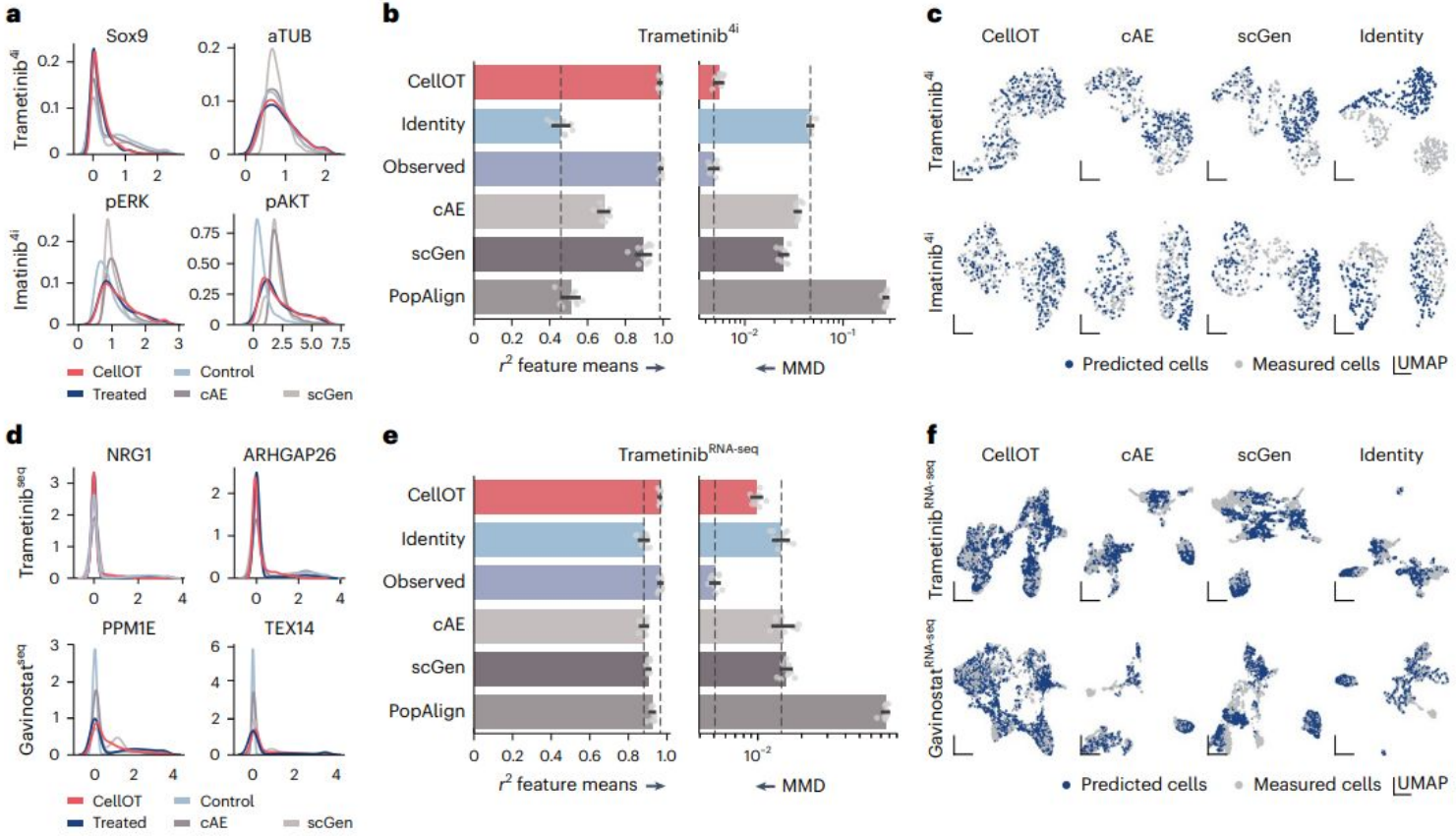
- $\ell_2$ **feature means**: distance between means of the observed and predicted distributions
- $r_2$ **feature means**: correlation of the means of the observed and predicted distributions
- **kernel maximum mean discrepancy (MMD)**: measures distance of two distributions

$$\mathrm{MMD}(p, q; \phi) = \mathbb{E}_{x,x'}[\phi(x, x')] + \mathbb{E}_{y,y'}[\phi(y, y')] - 2\mathbb{E}_{x,y}[\phi(x, y)]$$

- **uniform manifold approximation and projection (UMAP)**: dimension reduction technique used for visualisation

4i data

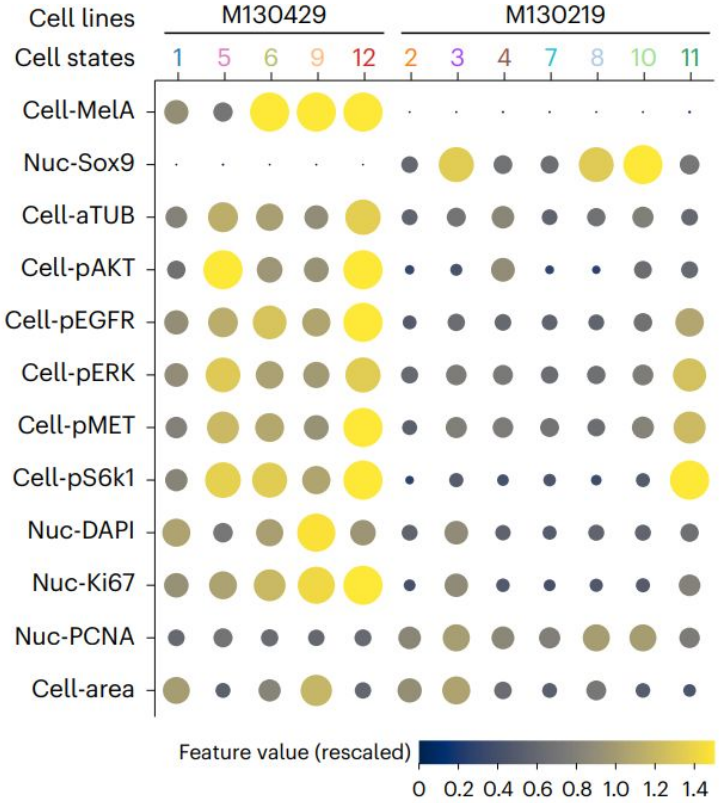
RNA seq data



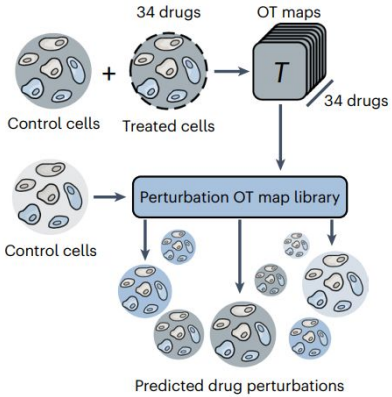

CellOT applied to predict the responses of cell populations to cancer treatments using a proteomic dataset consisting of two melanoma cell lines (M130219 and M130429)

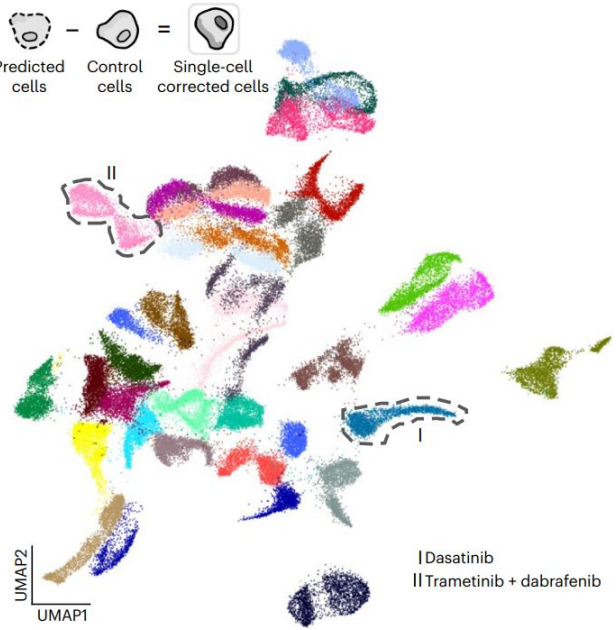
# Subpopulation-Specific Drug Effects

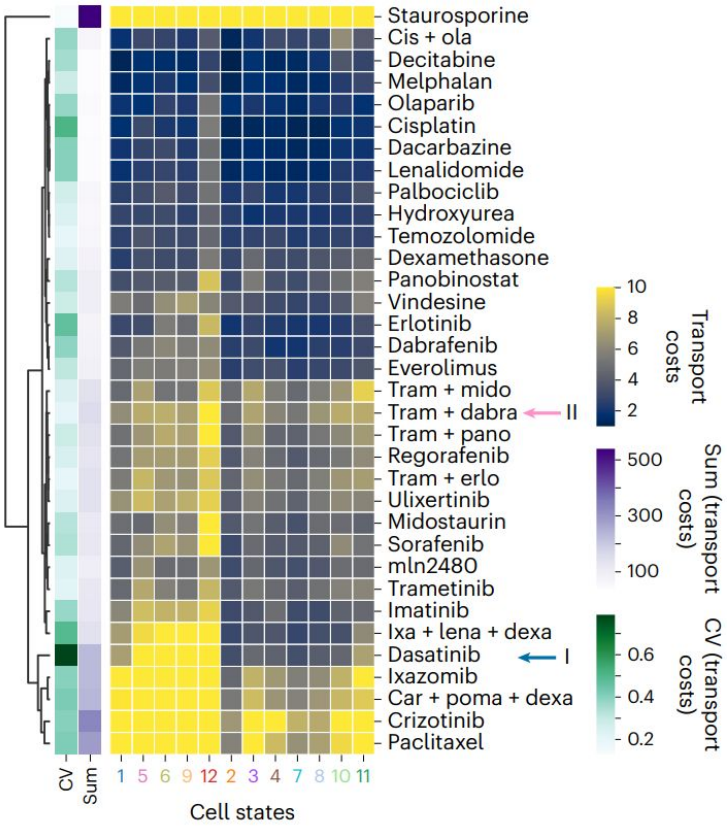
- clustering unperturbed cells into 12 cell-states
- cells 1,5,6,9,12 melanocytic cell line
  - pigment-producing cells
- cells 2,3,4,7,8,10,11 mesenchymal cell line
  - cells that develop into connective tissue, blood vessels, and lymphatic tissue

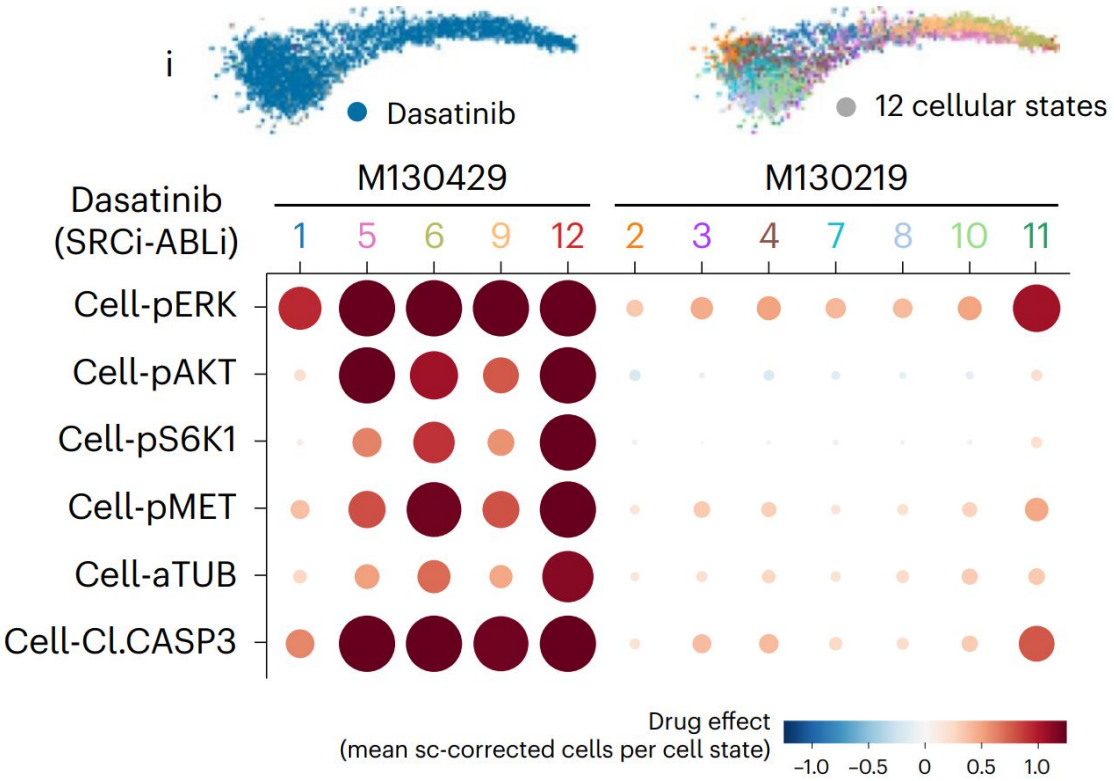


# Subpopulation-Specific Drug Effects



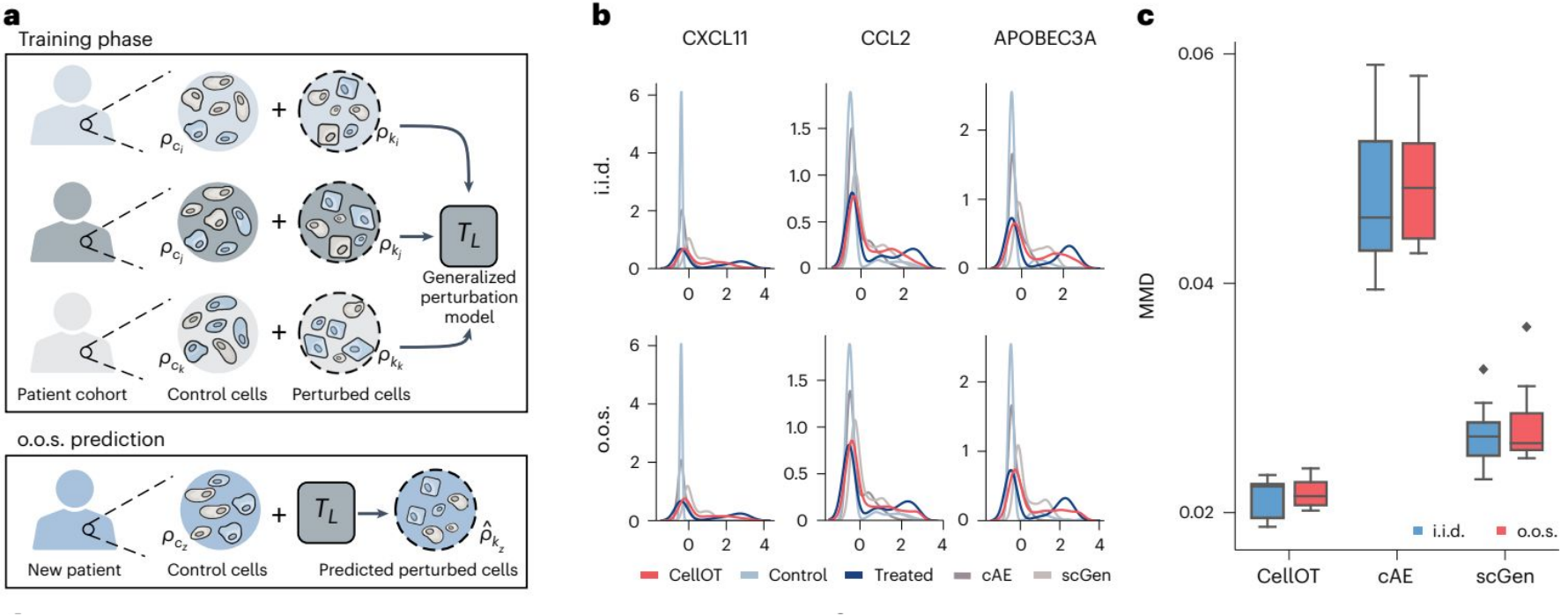

Staurosporine
Cis + ola
Decitabine
Melphalan
Olaparib
Cisplatin
Dacarbazine
Lenalidomide
Palbociclib
Hydroxyurea
Temozolomide
Dexamethasone
Panobinostat
Vindesine
Erlotinib
Dabrafenib
Everolimus
Tram + mido
Tram + dabra
Tram + pano
Regorafenib
Tram + erlo
Ulixertinib
Midostaurin
Sorafenib
mln2480
Trametinib
Imatinib
Ixa + lena + dexa
Dasatinib
Ixazomib
Car + poma + dexa
Crizotinib
Paclitaxel
II
I
CV
Sum
1 5 6 9 12 2 3 4 7 8 10 11
Cell states
Transport costs
10
8
6
4
2
Sum (transport costs)
500
300
100
CV (transport costs)
0.6
0.4
0.2



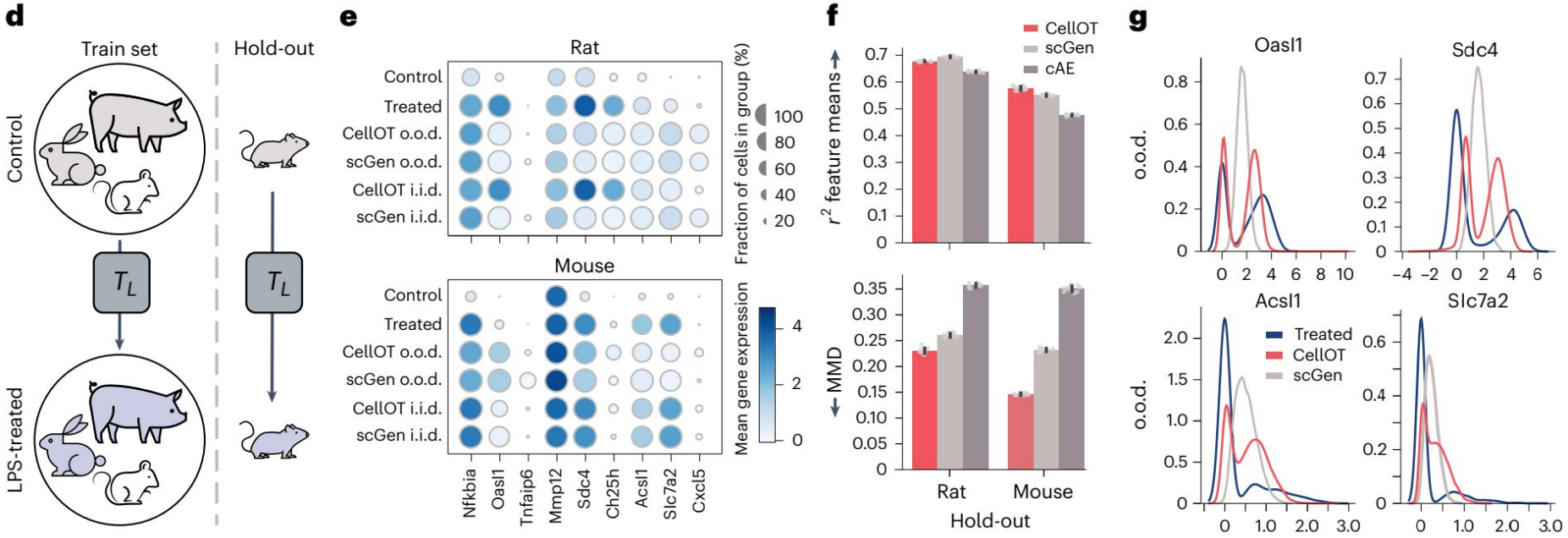

cl.CASP3 is an apoptosis marker(form of programmed cell death)

# CellOT - On Unseen Patients

- out of sample (o.o.s) setting compared against independent-identically distributed (i.i.d.) setting
- peripheral blood mononuclear cell droplet scRNA-seq dataset
- response of eight patients with lupus to interferon (IFN)-β,
  - a potent cytokine that induces genome-scale changes in immune cell transcriptional profiles
- three considered genes that are connected with autoimmune diseases
  - CXCL11, CCL2 and APOBEC3A

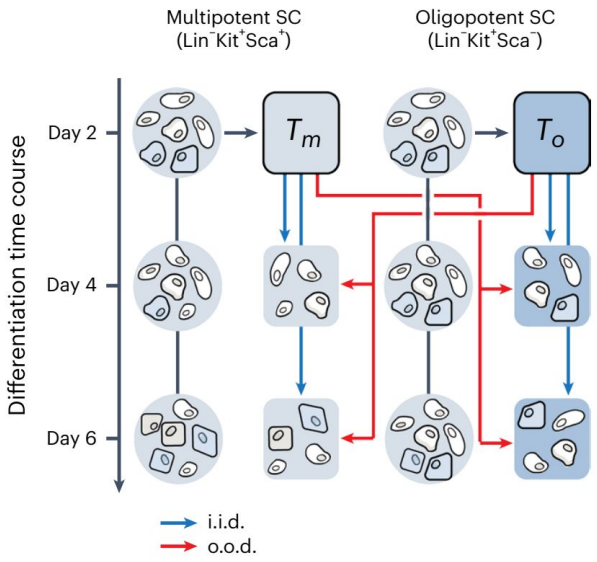

a
Training phase
Generalized perturbation model
Patient cohort
Control cells
Perturbed cells
o.o.s. prediction
New patient
Control cells
Predicted perturbed cells
b
CXCL11
CCL2
APOBEC3A
i.i.d.
o.o.s.
CellOT
Control
Treated
cAE
scGen
c
MMD
0.06
0.04
0.02
i.i.d.
o.o.s.
CellOT
cAE
scGen

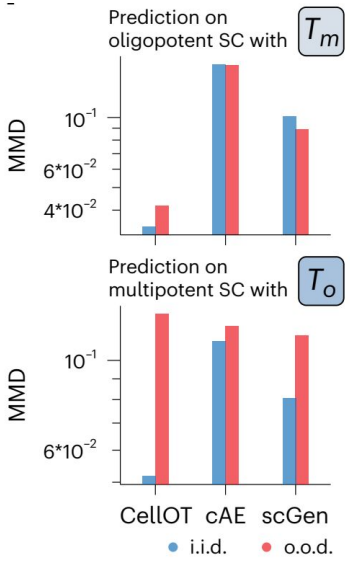
# CellOT - Out of Distribution



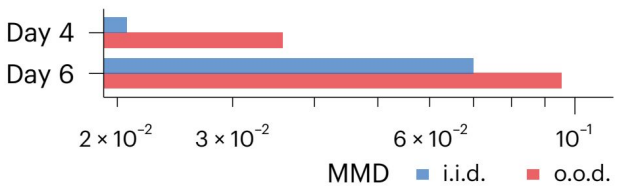
# CellOT - Differentiation Process of Cells

- hematopoietic stem and progenitor cells
  - oligopotent and multipotent progenitor cell subpopulations
- tracking cells on day 2,4 and 6 of the differentiation process
- o.o.d. setting: maps were trained only on one setting
- i.i.d. setting: trained on both populations
- tested on combination of the populations




➔ able to generalize its predictions to lower potency cell




| No. cells | Day% |
|---|---|
| 14,780 | 56.2 |
| 11,516 | 43.8 |

# Thank you for your Attention!

## Any Questions?