# PertAdapt: Unlocking Single-Cell Foundation Models for Genetic Perturbation Prediction via Condition-Sensitive Adaptation
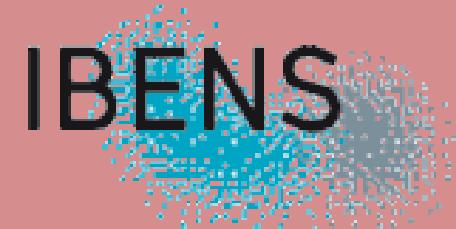
Ding Bai, Le Song, Eric Xing

Published in **bioRxiv** THE PREPRINT SERVER FOR BIOLOGY (November 21st, 2025)

03.03.2026 Clemence Reda

cnrs

IBENS
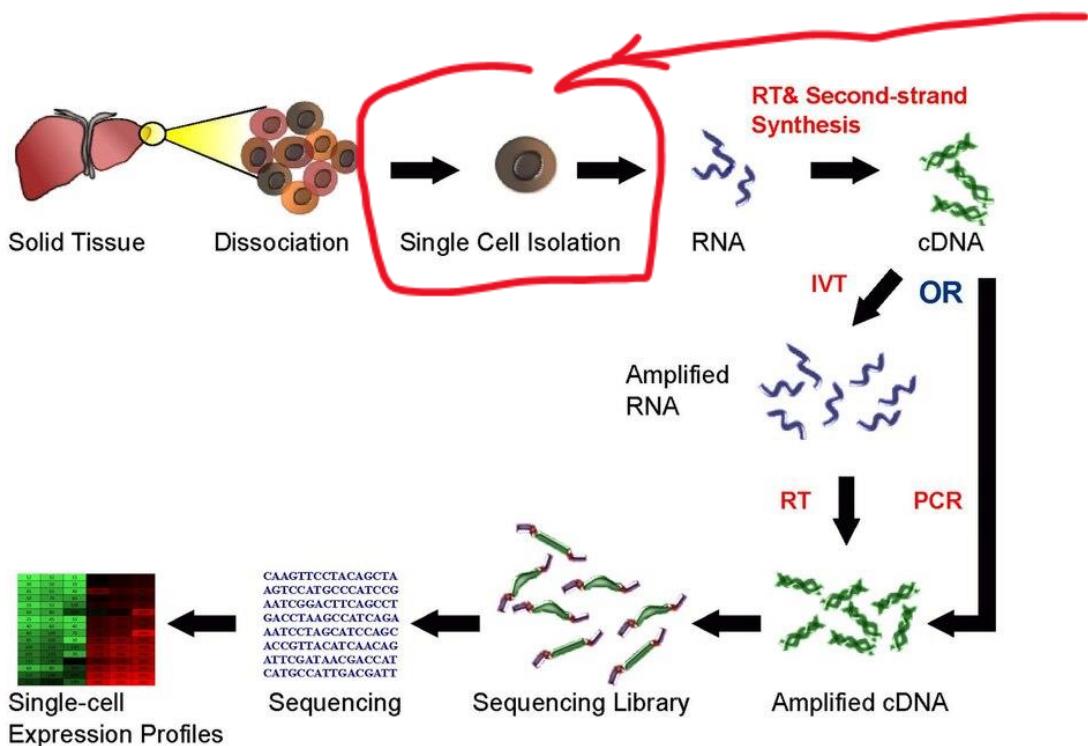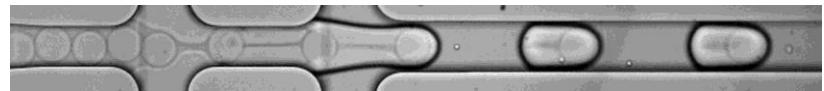
Background

1. Single-cell RNA sequencing (very short intro)
2. Gene Ontology annotations
3. Graph (Convolutional) Neural Networks (very short intro)
4. Problem setting
5. SOTA model GEARS (actually not so SOTA...)

Single-cell RNA sequencing snapshots of cell-specific expression in a tissue

Specific to single-cell seq
(e.g., droplet)

GEM-X from 10x Genomics [1]
500-20k cells/tissue
~2-4 days end-to-end
8-10k per-cell gene counts
~$3k-$8k per sample **[2]**

RT& Second-strand Synthesis

Solid Tissue   Dissociation   Single Cell Isolation   RNA   cDNA

IVT   OR

Amplified RNA

RT   PCR

CAAGTTCCTACAGCTA
AGTCCATGCCCATCCG
AATCGGACTTCAGCCT
GACCTAAGCCATCAGA
AATCCTAGCATCCAGC
ACCGTTACATCAACAG
ATTCGATAACGACCAT
CATGCCATTGACGATT

Single-cell Expression Profiles   Sequencing   Sequencing Library   Amplified cDNA

Source from Wikipedia ("Single-cell sequencing")

[1] 10xgenomics.com/blog/the-next-generation-of-single-cell-rna-seq-an-introduction-to-gem-x-technology (May, 2024)
[2] satijalab.org/costpercell/

3

cnrs IBENS

Gene Ontology (GO) annotations [3] Contribution of individual gene (products) to biological functions

One gene can be attached to multiple GO annotations

GO annotation
- Gene product
- GO term
- Reference paper
- GO Evidence code

Biological Process: e.g., DNA repair

Molecular Function: e.g., catalysis
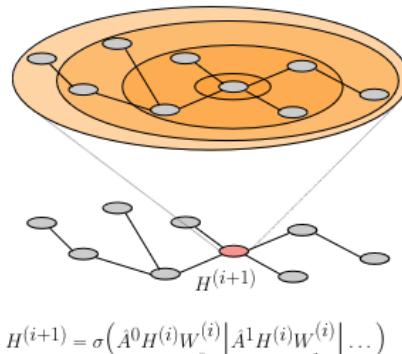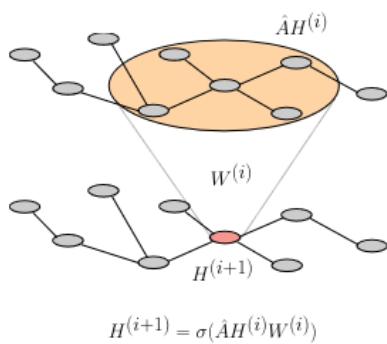
Cellular Component: e.g., cytoskeleton

Existence of a **hierarchy** for GO terms

What people do: count the number of shared GO terms in a pair of genes

[3] The Gene Ontology Consortium. (2025) *Nucleic Acids Research*;gkaf1292.

03.03.2026   Clemence Reda   cnrs IBENS

Graph (Convolutional) Neural Networks [4] Get node representations infused with graph information

$$H^{(i+1)} = \sigma(\hat{A}H^{(i)}W^{(i)})$$

$$H^{(i+1)} = \sigma\left(\hat{A}^0 H^{(i)}W^{(i)}_0 \mid \hat{A}^1 H^{(i)}W^{(i)}_1 \mid \ldots\right)$$

K=1 versus K>1: adapted from **[5]**
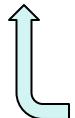
H$^{(i)}$: embedding matrix in $\mathbb{R}^{N \times d}$
at neural network layer i

in practice — simplified GCNN
$$H^{(i+1)} = \hat{A}^k H^{(i)} W^{(i)}$$
↖ normalized adj. matrix

Spatial domain $\quad H^{(i)}_v = f(\mathrm{Agg}(\{H^{(i-1)}_u \mid (u,v) \text{ in } G\})W^{(i)}, H^{(i-1)}_v)$ $\quad$ Choice of the aggregation function? **[6]**

Spectral domain $\quad H^{(i)} = UH^{(i-1)}W^{(i)}$

U can be numerically instable... **[7]**

Spectral filter: eigenvectors of Laplacian of graph
G = Degree(G) - Adjacency(G)
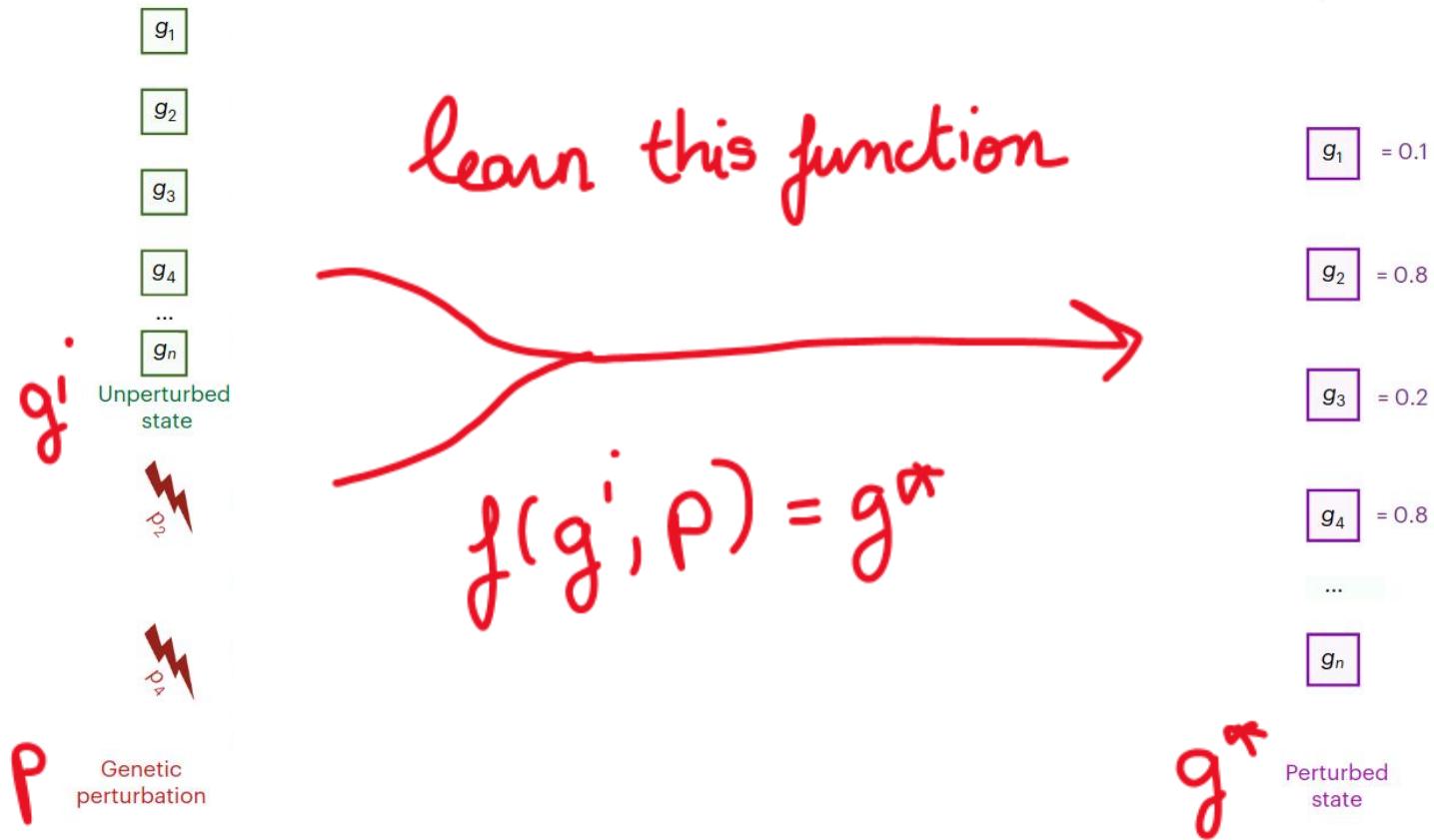
[4] Kipf and Welling. (2016) *International Conference on Learning Representations*
[5] Abu-El-Haija et al. (2019) *International Conference on Machine Learning*
[6] Xu et al. (2019) *International Conference on Learning Representations*
[7] Park et al. (2019) *IEEE/CVF Conference on Computer Vision and Pattern Recognition*
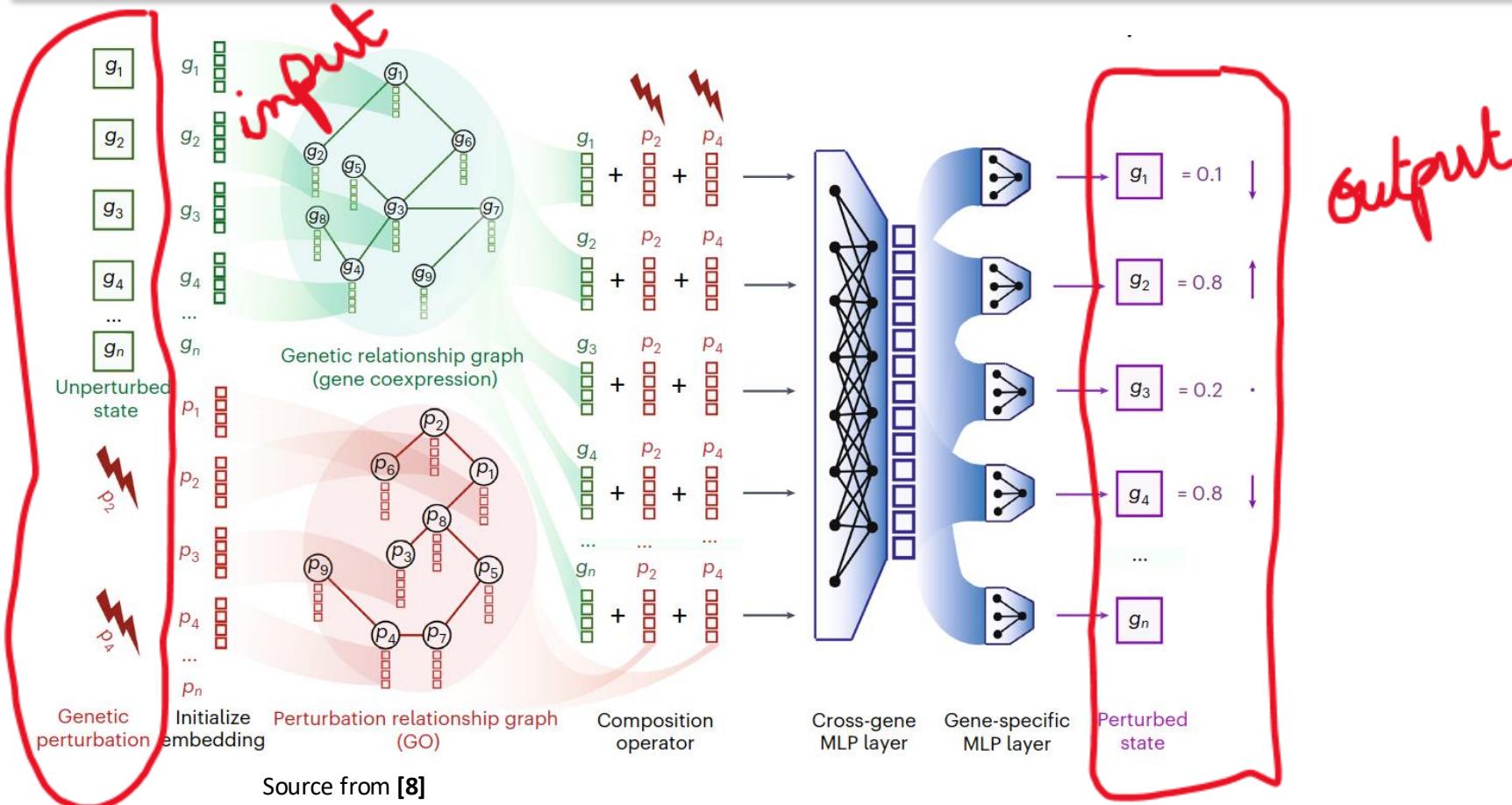
cnrs IBENS

# Problem setting Prediction of post-perturbation (on >1 gene) single-cell expression profile

learn this function

$$f(g^i, p) = g^\alpha$$

g_1	= 0.1
g_2	= 0.8
g_3	= 0.2
g_4	= 0.8
...	
g_n	

g^i

g_1
g_2
g_3
g_4
...
g_n
Unperturbed state

p_2

p_4

p Genetic perturbation

g^α Perturbed state

cnrs IBENS

SOTA model GEARS [8] predicting post-perturbation single-cell expression profiles

Source from [8]

[8] Roohani, Huang & Leskovec. (2023) *Nature Biotechnology* (1):9.

03.03.2026   Clemence Reda

SOTA model GEARS [8] predicting post-perturbation single-cell expression profiles

Correlations on training data

$\delta = 0,4$ Cut off

spatial GNNs

Source from [8]

Connect Top-N genes based on shared GO terms

$\rightarrow N = 20$

[8] Roohani, Huang & Leskovec. (2023) *Nature Biotechnology* (1):9.

Source from [8]

[8] Roohani, Huang & Leskovec. (2023) *Nature Biotechnology* (1):9.

Autofocus direction-aware loss
ĝ predicted, g* ground truth,
g^i unperturbed input
for one perturbation
of size #cells x #transcripts

$$L_{autofocus} \propto \sum_t \sum_c (g_{ct} - \hat{g}_{ct})^{2+\alpha}$$

$$\cdot L_{direction} \propto \sum_t \sum_c [\, sign(g_{ct} - g^i_{ct}) - sign(\hat{g}_{ct} - g^i_{ct})\,]^2$$

Final loss
$$L_{autofocus} + \beta L_{direction}$$

$\{\beta = 1$

Source from [8]

[8] Roohani, Huang & Leskovec. (2023) *Nature Biotechnology* (1):9.

# SOTA model GEARS [8] Is it really state-of-the-art?
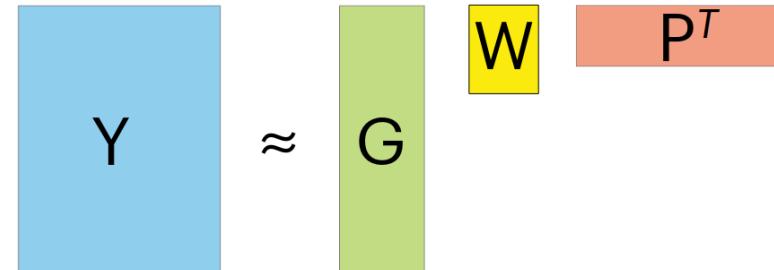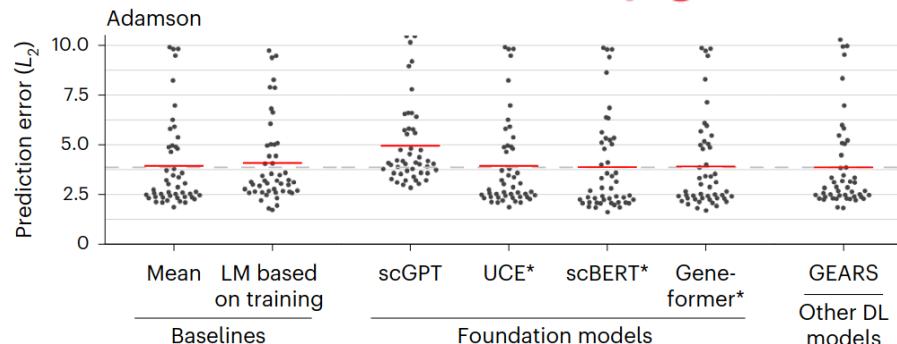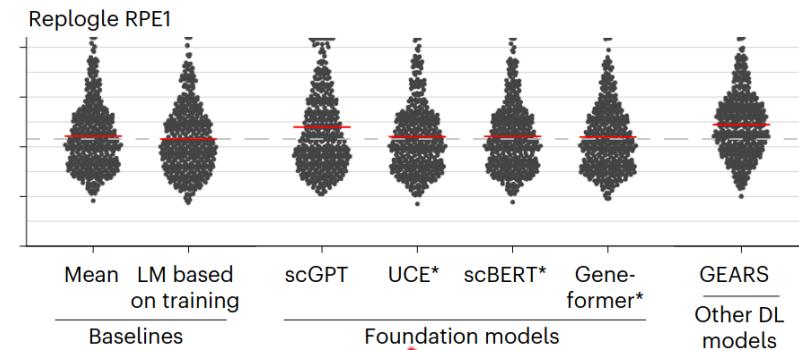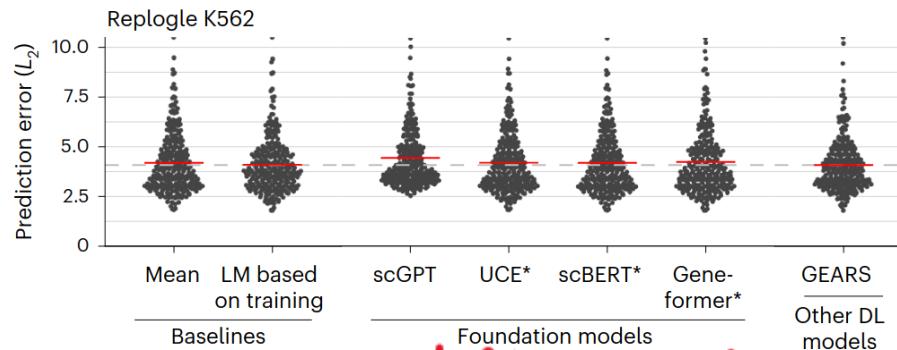
Source from **[9]**

[8] Roohani, Huang & Leskovec. (2023) *Nature Biotechnology* (1):9.
[9] Ahlmann-Eltze, Huber & Anders. (2025). *Nature Methods*, *22*(8), 1657-1661.

Does not consistently outperform linear regression... [9]

+ linear decoders

idem

Source from [9]

idem

$$Y \approx G \cdot W \cdot P^T$$

[8] Roohani, Huang & Leskovec. (2023) *Nature Biotechnology* (1):9.
[9] Ahlmann-Eltze, Huber & Anders. (2025). *Nature Methods*, *22*(8), 1657-1661.

Content of the paper

"[We] propose PertAdapt, an FM-based method [...] that integrates a plug-in perturbation adapter with an adaptive loss"

Specific issues:
- Enabling downstream task-specific tuning instead of plug-in MLPs
- Reweighting gene importance according to differential expression
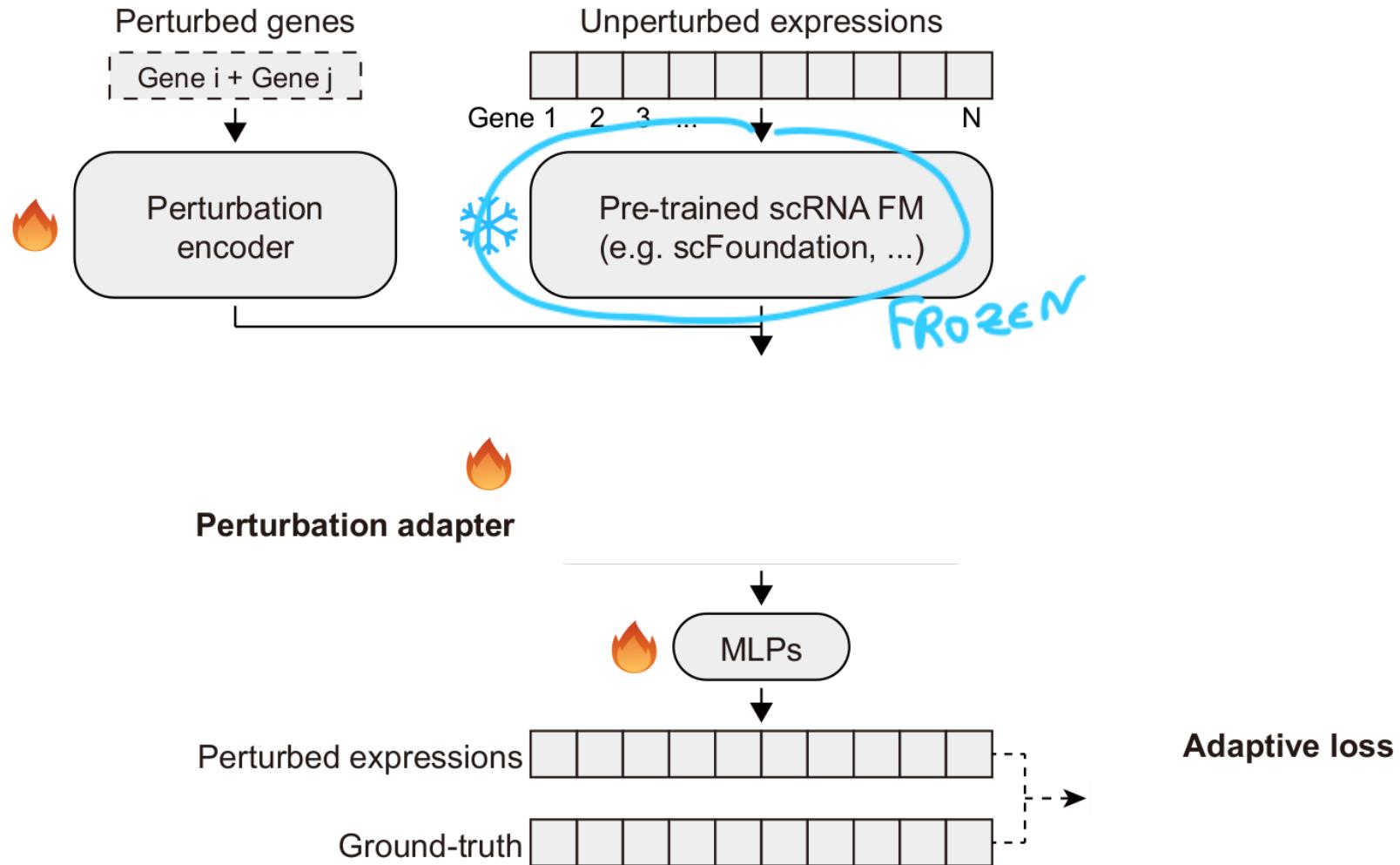
Prior approaches:
- SOTA GEARS, (repurposed) FMs: Geneformer, UCE, scFoundation,...

Content of the paper
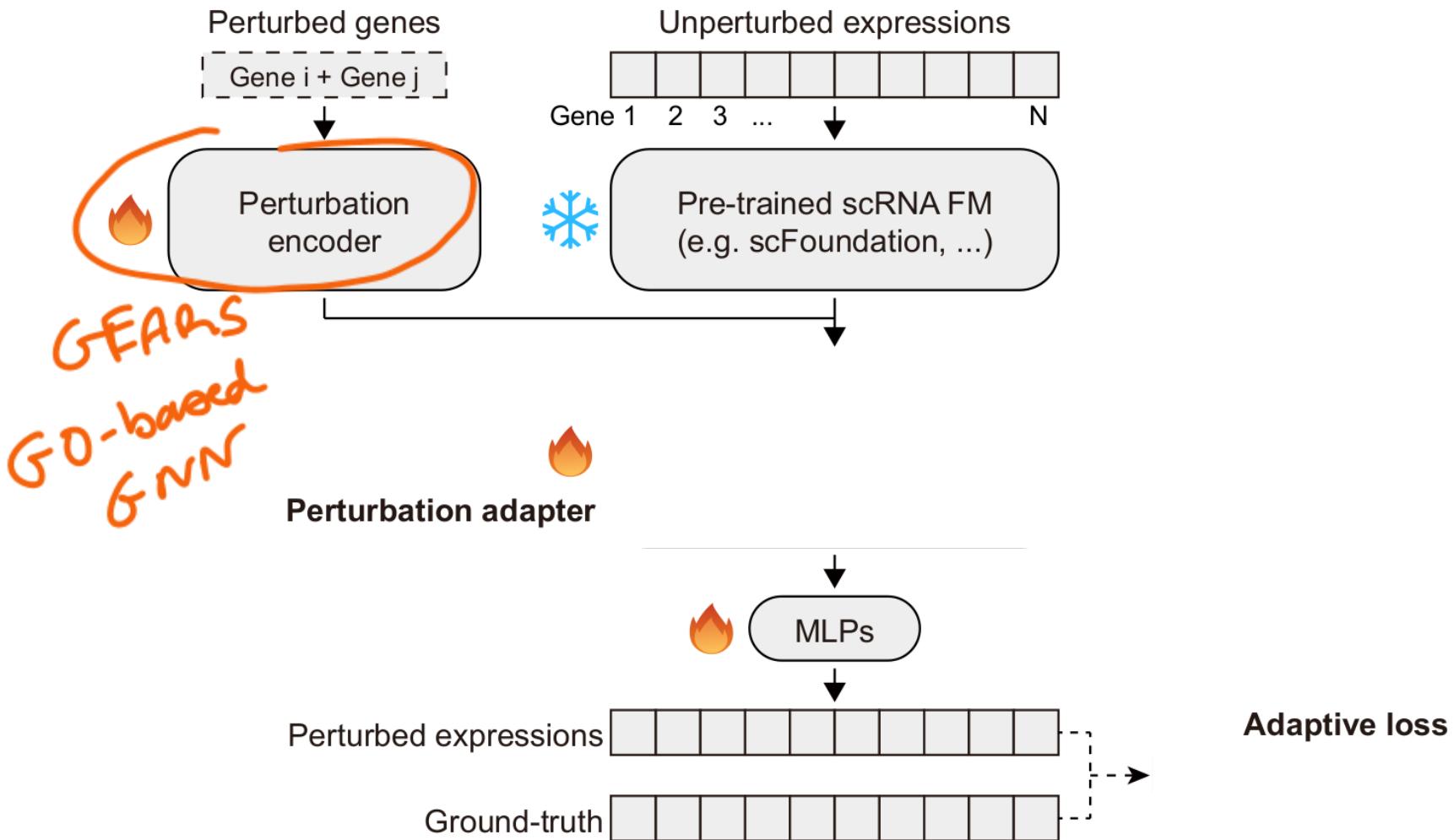
"[We] propose PertAdapt, an FM-based method […] that integrates a plug-in perturbation adapter with an adaptive loss"

1. Architecture of PertAdapt
2. Plug-in perturbation adapter
3. Adaptive loss
4. Experiments: Prediction of post-perturbation expr.
5. Experiments: Ablation study
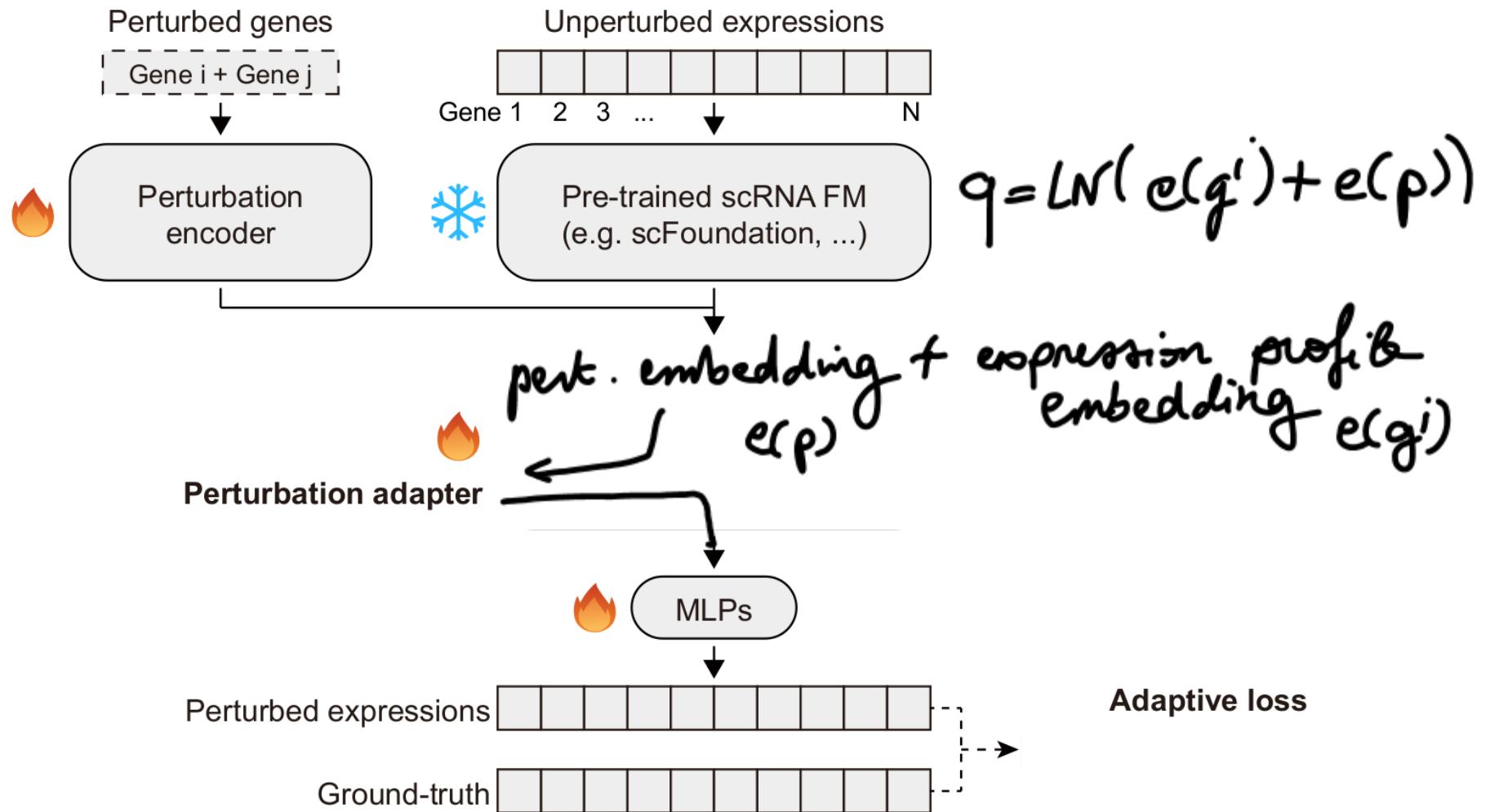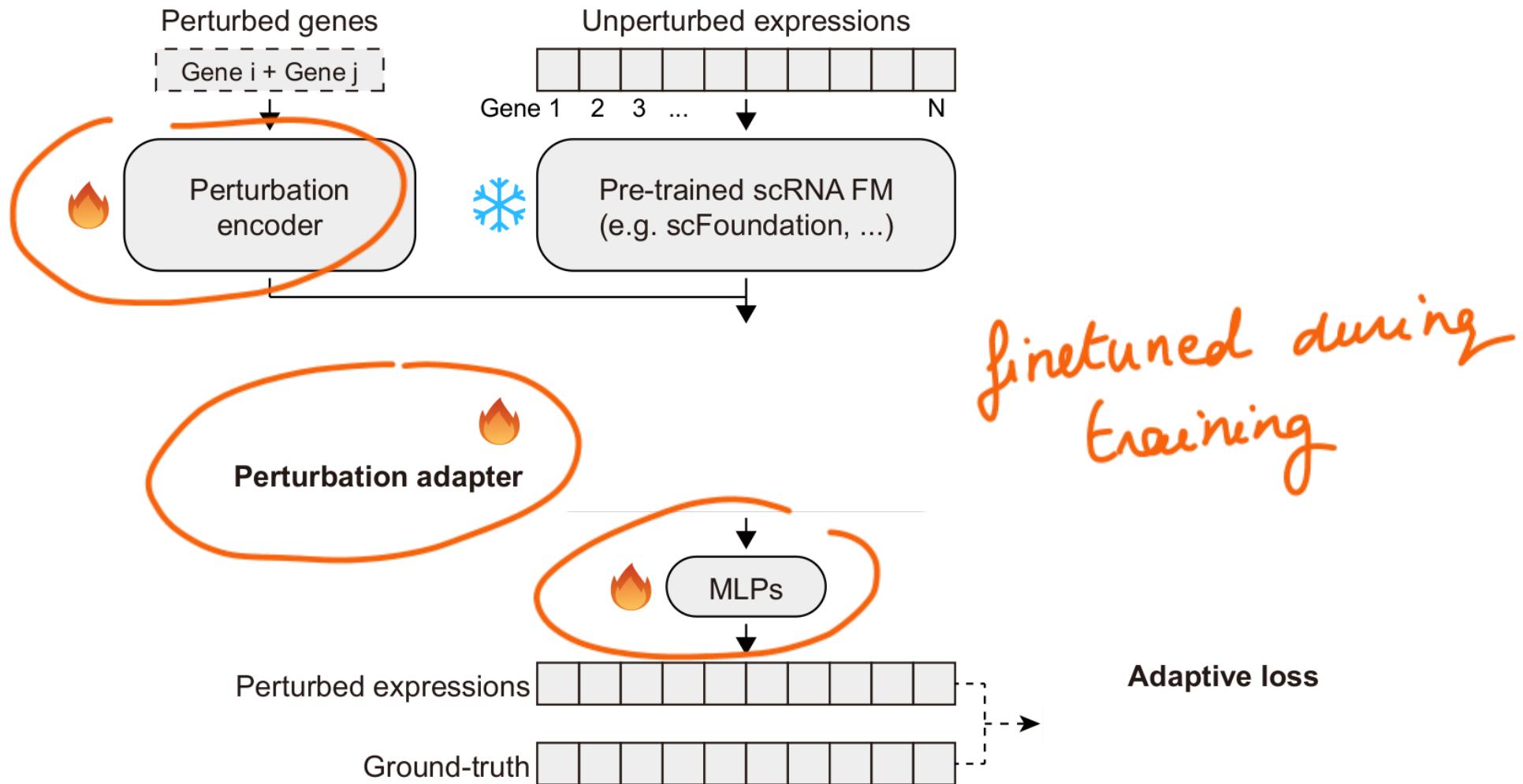
# Architecture of PertAdapt foundation model-agnostic structure
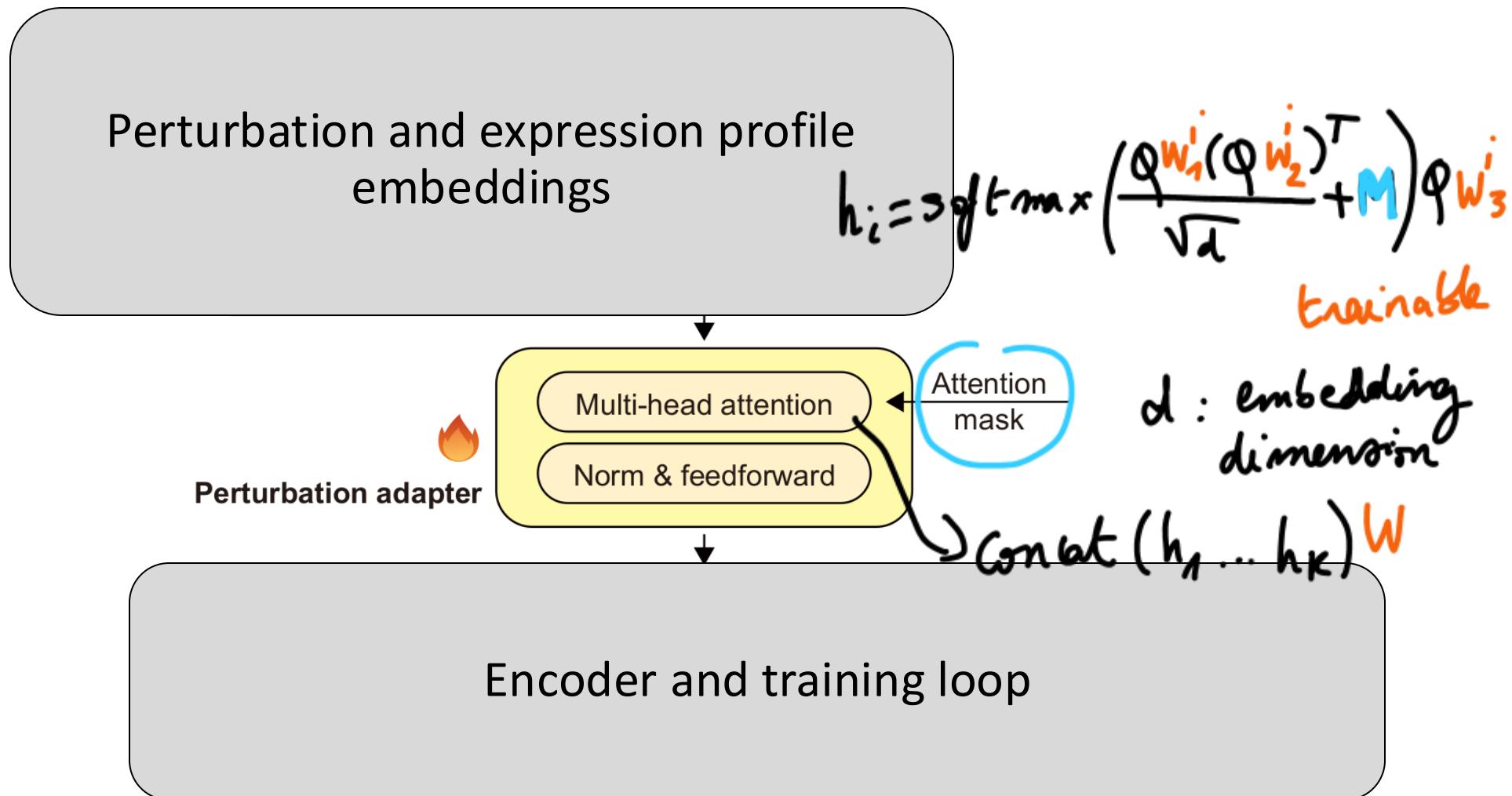
Architecture of PertAdapt foundation model-agnostic structure
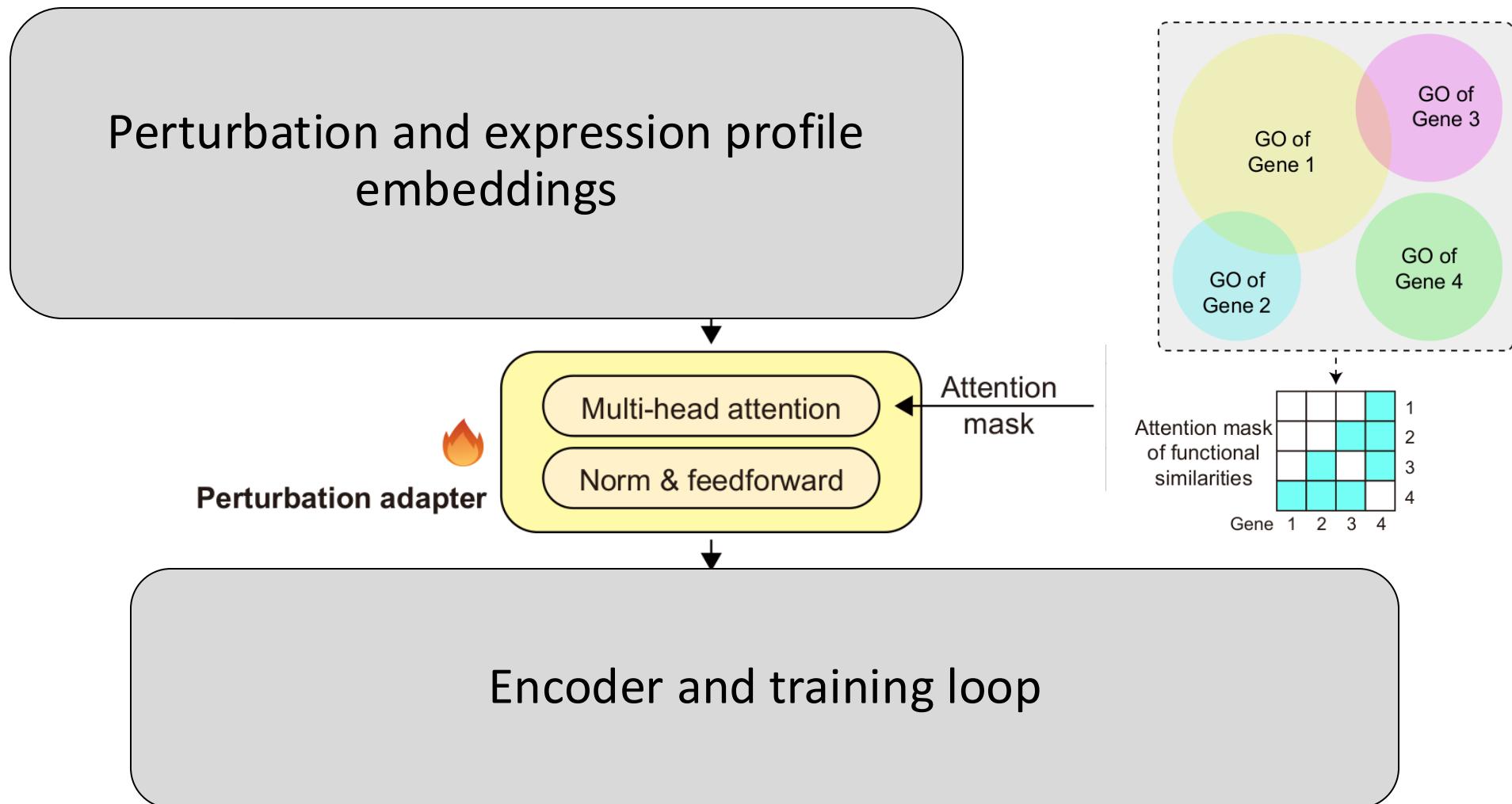
Plug-in perturbation adapter downstream-task tuning and biological prior integration
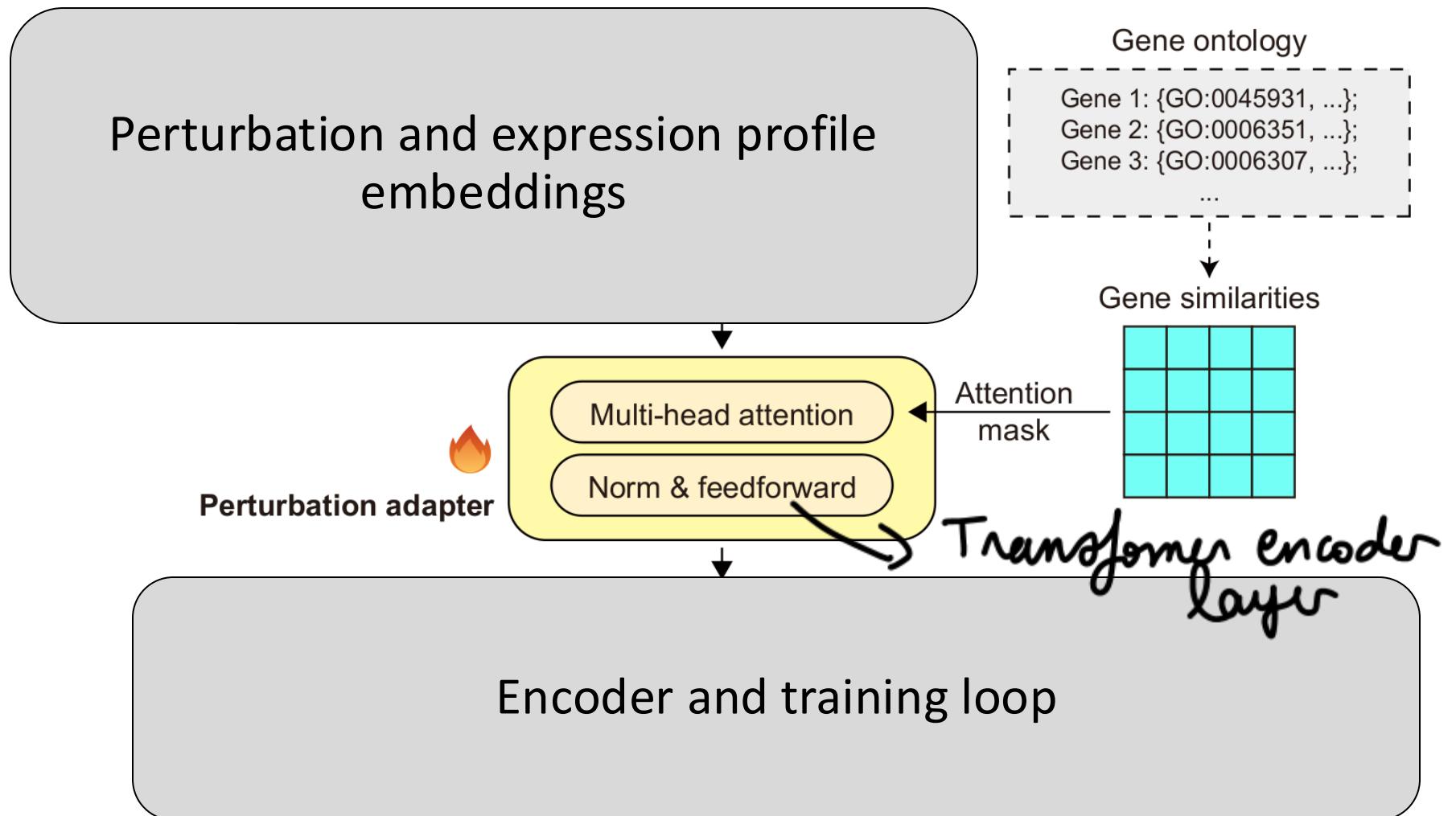
Plug-in perturbation adapter downstream-task tuning and biological prior integration

Perturbation and expression profile embeddings

$$h_i = softmax\left(\frac{\varphi W_1^i (\varphi W_2^i)^T}{\sqrt{d}} + M\right)\varphi W_3^i$$

trainable

d : embedding dimension

Multi-head attention

Attention mask

Perturbation adapter

Norm & feedforward

$Concat(h_1 \dots h_K) W$

Encoder and training loop
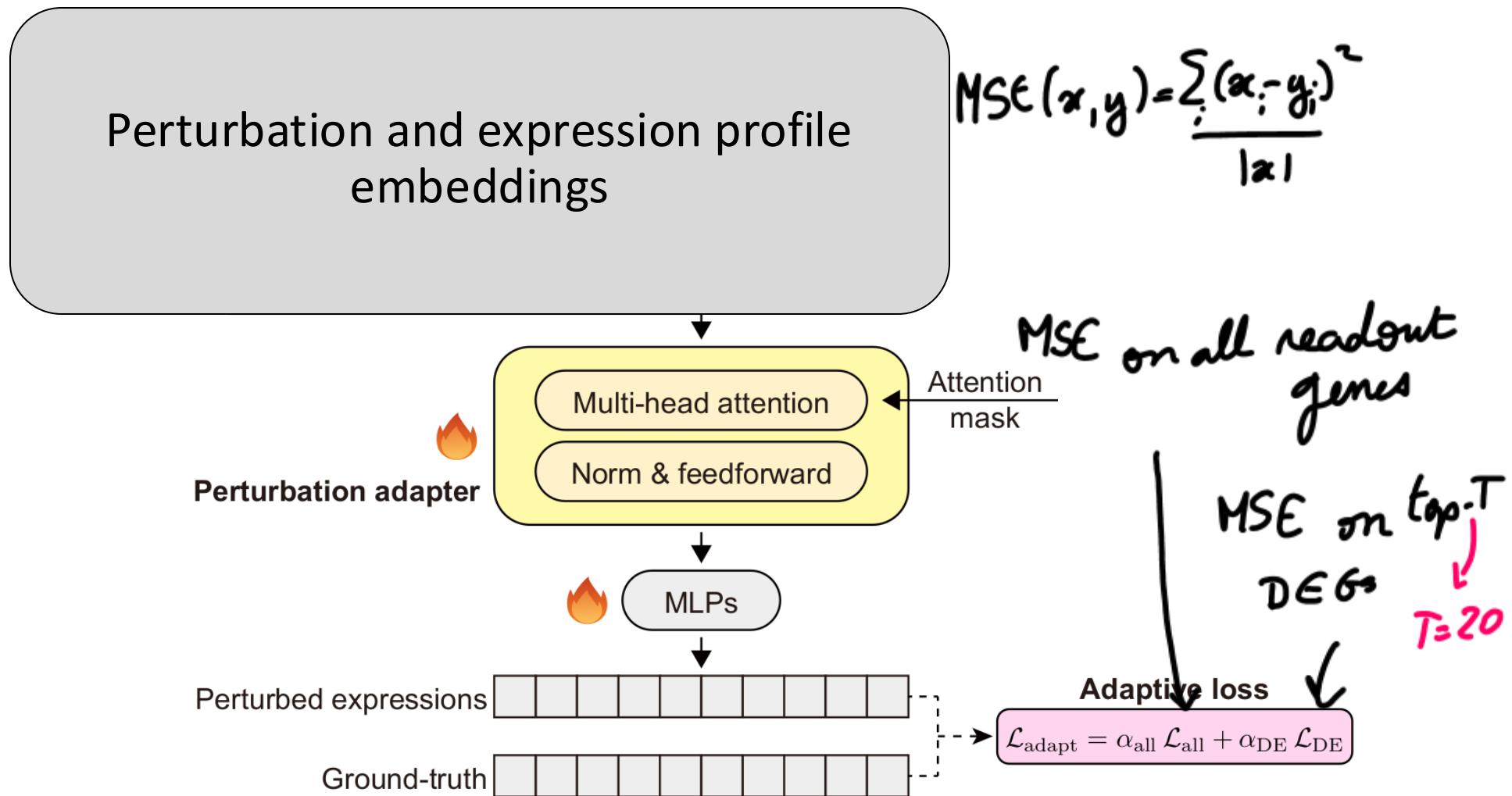
Plug-in perturbation adapter downstream-task tuning and biological prior integration
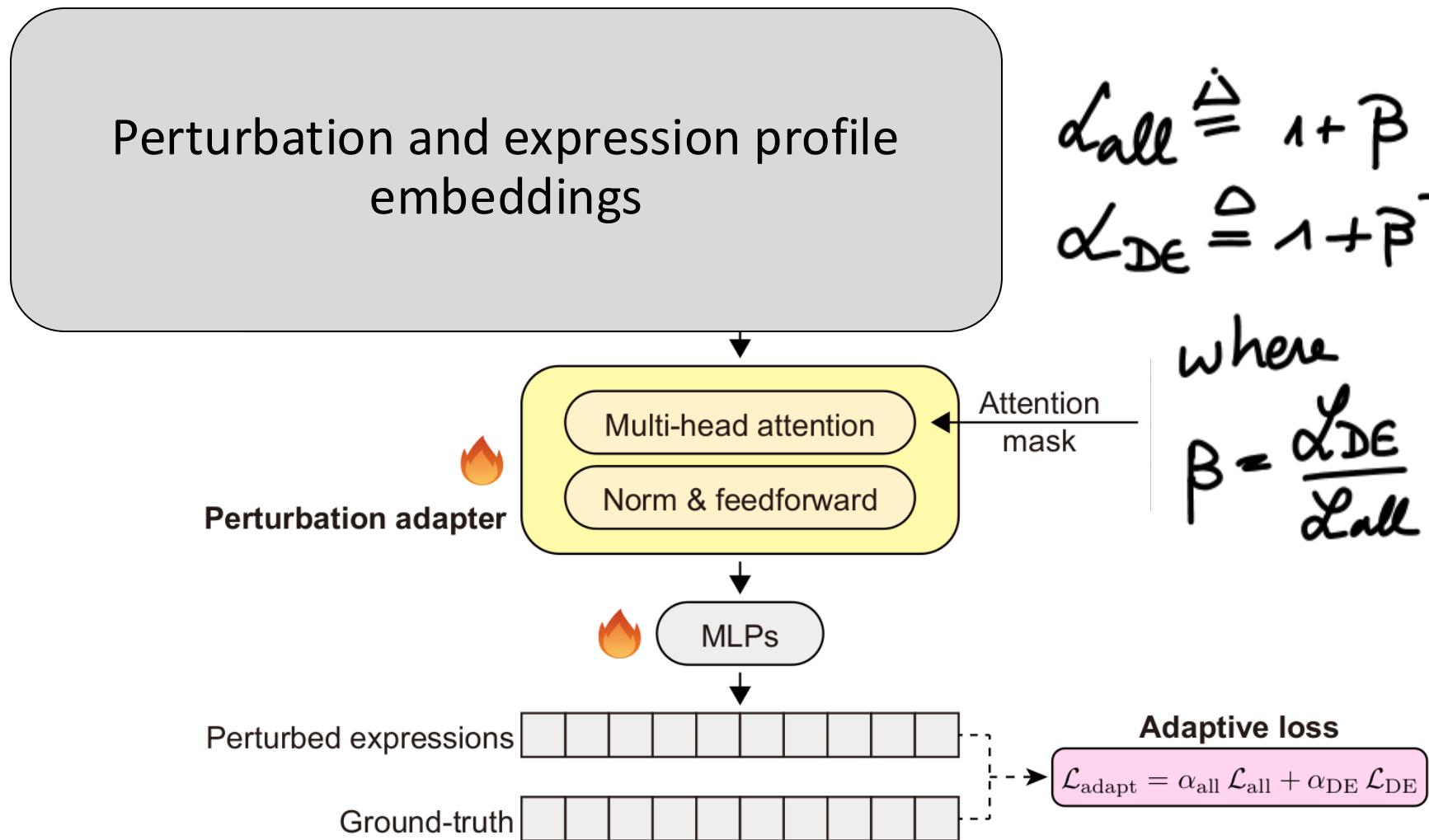
Plug-in perturbation adapter downstream-task tuning and biological prior integration
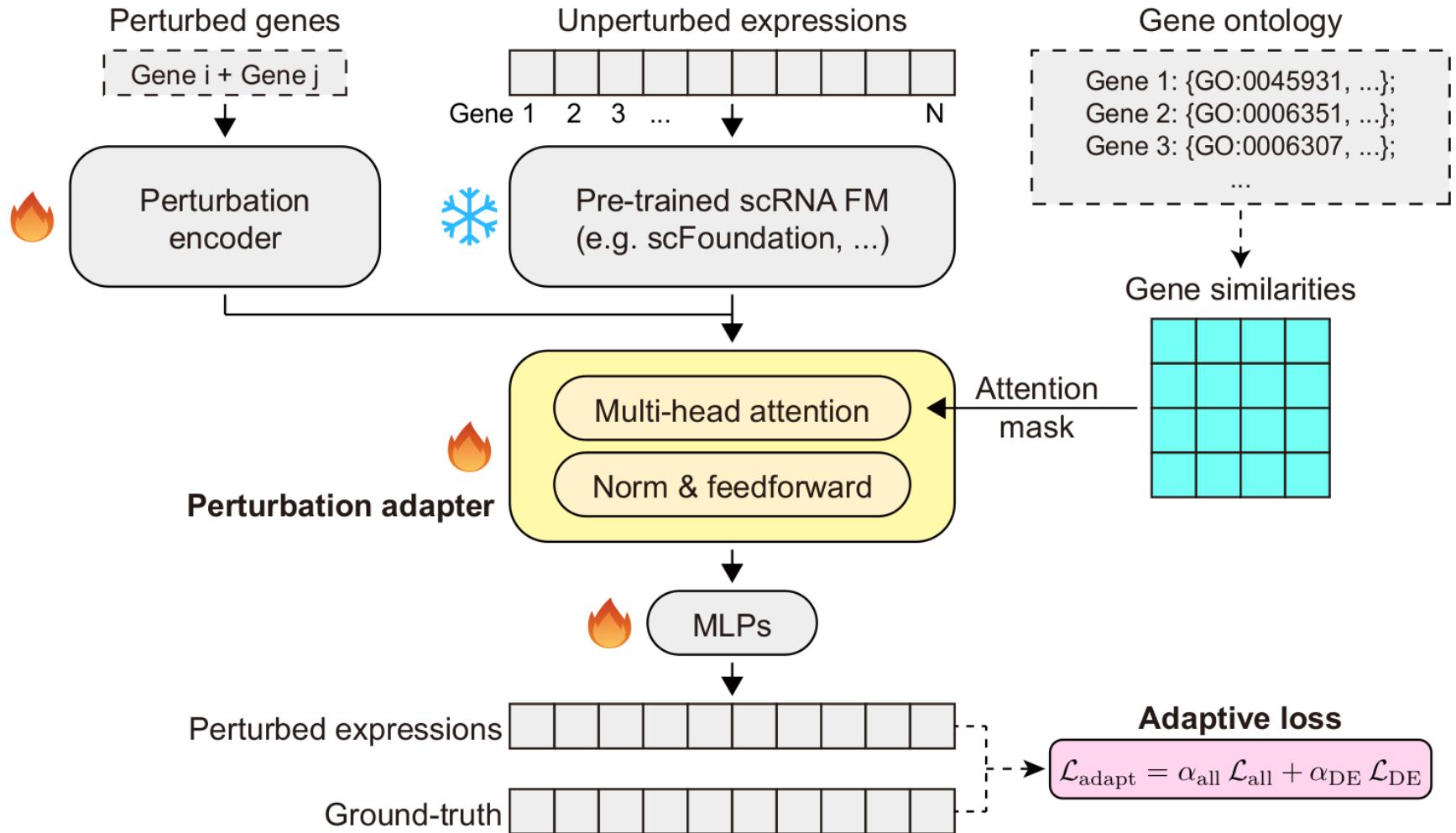
Adaptive loss put more weight on differentially expressed genes to lessen noise

$$MSE(x,y) = \frac{\sum_i (x_i - y_i)^2}{|x|}$$

Perturbation and expression profile embeddings

Multi-head attention

Norm & feedforward

Perturbation adapter

Attention mask

MLPs

Perturbed expressions

Ground-truth

MSE on all readout genes

MSE on top-T DEGs
T=20

Adaptive loss

$$\mathcal{L}_{adapt} = \alpha_{all}\,\mathcal{L}_{all} + \alpha_{DE}\,\mathcal{L}_{DE}$$

Plug-in perturbation adapter downstream-task tuning and biological prior integration

Model	Replogle RPE1		Replogle K562		Nadig HepG2	
	MSE↓	Pearson↑	MSE↓	Pearson↑	MSE↓	Pearson↑
No-Change	0.201	0.706	0.193	0.908	0.121	0.681
Linear Model	0.113	0.704	0.165	0.924	0.101	0.651
GEARS	0.159	0.719	0.124	0.929	0.101	0.652
scGPT	0.158	0.721	0.135	0.919	0.110	0.676
GeneCompass	0.170	0.727	0.121	0.921	0.104	0.680
scFoundation	0.166	0.731	0.142	0.927	0.109	0.684
AIDO.Cell	0.134	0.724	0.130	0.928	0.102	0.680
scFoundation+Ours	**0.091**	**0.737**	**0.083**	0.935	**0.062**	**0.700**
AIDO.Cell+Ours	0.094	0.725	0.088	**0.936**	0.067	0.685
	Nadig Jurkat		Adamson		Dixit	
	MSE↓	Pearson↑	MSE↓	Pearson↑	MSE↓	Pearson↑
No-Change	0.111	0.801	0.905	0.787	0.380	0.831
Linear Model	0.101	0.815	0.675	0.862	0.229	0.912
GEARS	0.095	0.817	1.028	0.795	0.271	0.896
scGPT	0.112	0.806	1.116	0.679	0.295	0.841
GeneCompass	0.099	0.808	0.986	0.808	0.257	0.890
scFoundation	0.131	0.812	0.792	0.806	0.243	0.913
AIDO.Cell	0.103	0.820	0.804	0.840	0.213	0.924
scFoundation+Ours	**0.061**	**0.826**	**0.324**	**0.904**	0.021	0.992
AIDO.Cell+Ours	0.069	0.824	0.408	0.886	**0.013**	**0.995**

Experiments: Prediction of post-perturbation expression showing only single-gene perturbations

Model	Replogle RPE1		Replogle K562		Nadig HepG2	
	MSE↓	Pearson↑	MSE↓	Pearson↑	MSE↓	Pearson↑
No-Change	0.201	0.706	0.193	0.908	0.121	0.681
Linear Model	0.113	0.704	0.165	0.924	0.101	0.651
GEARS	0.159	0.719	0.124	0.929	0.101	0.652
scGPT	0.158	0.721	0.135	0.919	0.110	0.676
GeneCompass	0.170	0.727	0.121	0.921	0.104	0.680
scFoundation	0.166	0.731	0.142	0.927	0.109	0.684
AIDO.Cell	0.134	0.724	0.130	0.928	0.102	0.680
scFoundation+Ours	**0.091**	**0.737**	**0.083**	0.935	**0.062**	**0.700**
AIDO.Cell+Ours	0.094	0.725	0.088	**0.936**	0.067	0.685
	Nadig Jurkat		Adamson		Dixit	
	MSE↓	Pearson↑	MSE↓	Pearson↑	MSE↓	Pearson↑
No-Change	0.111	0.801	0.905	0.787	0.380	0.831
Linear Model	0.101	0.815	0.675	0.862	0.229	0.912
GEARS	0.095	0.817	1.028	0.795	0.271	0.896
scGPT	0.112	0.806	1.116	0.679	0.295	0.841
GeneCompass	0.099	0.808	0.986	0.808	0.257	0.890
scFoundation	0.131	0.812	0.792	0.806	0.243	0.913
AIDO.Cell	0.103	0.820	0.804	0.840	0.213	0.924
scFoundation+Ours	**0.061**	**0.826**	**0.324**	**0.904**	0.021	0.992
AIDO.Cell+Ours	0.069	0.824	0.408	0.886	**0.013**	**0.995**

adapter   attention   loss

Backbone	FM	PA	AM	$\mathcal{L}_{adapt}$	Norman (Overall) MSE↓	Pearson↑	Replogle RPE1 MSE↓	Pearson↑
scFoundation					0.375	0.584	0.328	0.490
	✓				0.210	0.776	0.166	0.731
	✓	✓			0.208	0.779	0.143	0.730
	✓	✓	✓		0.183	0.782	0.124	0.733
	✓			✓	0.178	0.782	0.109	0.732
	✓	✓	✓	✓	**0.156**	**0.789**	**0.091**	**0.737**
AIDO.Cell					0.351	0.618	0.297	0.515
	✓				0.192	0.781	0.134	0.724
	✓	✓			0.197	0.769	0.125	0.720
	✓	✓	✓		0.179	0.782	0.104	0.723
	✓			✓	0.170	0.780	0.103	**0.726**
	✓	✓	✓	✓	**0.159**	**0.786**	**0.094**	0.725

Perspectives

1. Comments on the paper
2. Why is it interesting for BioComp?

My comments on the paper

Strengths:
- Convincing results
- I liked the plug-in perturbation adapter and GO-based attention mask
- Open-source: https://github.com/BaiDing1234/PertAdapt

Weaknesses:
- Lack of biological assessment
- Lack of runtime assessment
- "Biological prior vs exploration" dilemma

Your comments?

Why is it interesting for BioComp?

Multiple applications (e.g., drug discovery)
Many of you working on this topic

- Lots of papers on the topic
- Often not so much biology-focused
- Interpretability, practicality (time+resources),
    probably other challenges