Understanding Transcriptional Regulatory Redundancy by Learnable Global Subset Perturbations

Presented at the ACML 2024

Junhao Liu[1,#] Siwei Xu[1,#] Dylan Riffle[3] Ziheng Duan[1]
Martin Renqiang Min[2] Jing Zhang[1,*]

[1]University of California, Irvine [2]NEC Laboratories America [3]Cornell University
[#]Equal Contributions [*]Corresponding Author

December 12, 2024

Table of Contents

Table of Contents

Cis-Regulatory Elements and Transcriptional Regulatory

Cis-Regulatory Elements (CREs) are regions of DNA that regulates the expression of genes

- CREs is crucial for numerous biological functions, with its disruption potentially leading to various diseases
- CREs often exhibit redundancy, allowing them to compensate for each other in response to external disturbances
- Single-cell sequencing technologies offer researchers an opportunity to understand this regulatory redundancy

Notations

ATAC-seq $\mathbf{x} \in \{0,1\}^{d_a}$ $(d_a > 120k)$, where each dimension of this vector indicates the peak state in chromosomes

RNA-seq $\mathbf{y} \in \mathbb{R}^{d_r}$, gene expression values regulated by the ATAC-seq

Transcriptional Regulatory A complex gene regulatory process $\mathbf{y} = \mathcal{F}(\mathbf{x})$, where $\mathcal{F}$ is a black-box function.

Cis-Regulatory Elements and Transcriptional Regulatory

Cis-Regulatory Elements (CREs) are regions of DNA that regulates the expression of genes

- CREs is crucial for numerous biological functions, with its disruption potentially leading to various diseases
- CREs often exhibit redundancy, allowing them to compensate for each other in response to external disturbances
- Single-cell sequencing technologies offer researchers an opportunity to understand this regulatory redundancy

Notations

ATAC-seq $\mathbf{x} \in \{0,1\}^{d_a}$ $(d_a > 120k)$, where each dimension of this vector indicates the peak state in chromosomes

RNA-seq $\mathbf{y} \in \mathbb{R}^{d_r}$, gene expression values regulated by the ATAC-seq

Transcriptional Regulatory A complex gene regulatory process $\mathbf{y} = \mathcal{F}(\mathbf{x})$, where $\mathcal{F}$ is a black-box function.

Cis-Regulatory Elements and Transcriptional Regulatory

Cis-Regulatory Elements (CREs) are regions of DNA that regulates the expression of genes

- CREs is crucial for numerous biological functions, with its disruption potentially leading to various diseases
- CREs often exhibit redundancy, allowing them to compensate for each other in response to external disturbances
- Single-cell sequencing technologies offer researchers an opportunity to understand this regulatory redundancy

Notations

ATAC-seq $\mathbf{x} \in \{0,1\}^{d_a}$ $(d_a > 120k)$, where each dimension of this vector indicates the peak state in chromosomes

RNA-seq $\mathbf{y} \in \mathbb{R}^{d_r}$, gene expression values regulated by the ATAC-seq

Transcriptional Regulatory A complex gene regulatory process $\mathbf{y} = \mathcal{F}(\mathbf{x})$, where $\mathcal{F}$ is a black-box function.

Cis-Regulatory Elements and Transcriptional Regulatory

Cis-Regulatory Elements (CREs) are regions of DNA that regulates the expression of genes

- CREs is crucial for numerous biological functions, with its disruption potentially leading to various diseases
- CREs often exhibit redundancy, allowing them to compensate for each other in response to external disturbances
- Single-cell sequencing technologies offer researchers an opportunity to understand this regulatory redundancy

Notations

ATAC-seq $\mathbf{x} \in \{0,1\}^{d_a}$ $(d_a > 120k)$, where each dimension of this vector indicates the peak state in chromosomes

RNA-seq $\mathbf{y} \in \mathbb{R}^{d_r}$, gene expression values regulated by the ATAC-seq

Transcriptional Regulatory A complex gene regulatory process $\mathbf{y} = \mathcal{F}(\mathbf{x})$, where $\mathcal{F}$ is a black-box function.

Cis-Regulatory Elements and Transcriptional Regulatory

Cis-Regulatory Elements (CREs) are regions of DNA that regulates the expression of genes

- CREs is crucial for numerous biological functions, with its disruption potentially leading to various diseases
- CREs often exhibit redundancy, allowing them to compensate for each other in response to external disturbances
- Single-cell sequencing technologies offer researchers an opportunity to understand this regulatory redundancy

Notations

ATAC-seq $\mathbf{x} \in \{0,1\}^{d_a}$ $(d_a > 120k)$, where each dimension of this vector indicates the peak state in chromosomes

RNA-seq $\mathbf{y} \in \mathbb{R}^{d_r}$, gene expression values regulated by the ATAC-seq

Transcriptional Regulatory A complex gene regulatory process $\mathbf{y} = \mathcal{F}(\mathbf{x})$, where $\mathcal{F}$ is a black-box function.

Transcriptional Regulatory Redundancy Problem

Regulatory Redundancy Problem Given a target gene in **y**, what combinations of entries in **x** are regulating the expression of the target gene.

Examples

Let's say we have 4 cells, and a global removal is applied.

$$\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 1 & 1 \end{bmatrix} \xrightarrow{\mathcal{F}} \mathbf{Y} = \begin{bmatrix} 0.7 \\ 0.1 \\ 2.2 \\ 0.9 \end{bmatrix} \xrightarrow{\text{After}} \tilde{\mathbf{X}} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \xrightarrow{\mathcal{F}} \tilde{\mathbf{Y}} = \begin{bmatrix} 0.7 \\ 0.1 \\ 0.6 \downarrow \\ 0.9 \end{bmatrix}$$

Challenges

- The complex gene regulatory process function $\mathcal{F}$ is an indifferentiable black box
- The combinatorial effect of multiple CREs is important but has been ignored by most previous methods
- **x** is sparse, discrete, and high-dimension. The combinatorial solution space of $\text{subset}(\mathbf{x})$ is vast, which means brute-force search is intractable
- **In this work, we propose GRIDS to solve the above challenges**

Transcriptional Regulatory Redundancy Problem

Regulatory Redundancy Problem Given a target gene in **y**, what combinations of entries in **x** are regulating the expression of the target gene.

Examples

Let's say we have 4 cells, and a global removal is applied.

$$\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 1 & 1 \end{bmatrix} \overset{\mathcal{F}}{\Longrightarrow} \mathbf{Y} = \begin{bmatrix} 0.7 \\ 0.1 \\ 2.2 \\ 0.9 \end{bmatrix} \xrightarrow{\text{After}} \tilde{\mathbf{X}} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \overset{\mathcal{F}}{\Longrightarrow} \tilde{\mathbf{Y}} = \begin{bmatrix} 0.7 \\ 0.1 \\ 0.6 \downarrow \\ 0.9 \end{bmatrix}$$

Challenges

- The complex gene regulatory process function $\mathcal{F}$ is an indifferentiable black box
- The combinatorial effect of multiple CREs is important but has been ignored by most previous methods
- **x** is sparse, discrete, and high-dimension. The combinatorial solution space of $\text{subset}(\mathbf{x})$ is vast, which means brute-force search is intractable
- **In this work, we propose GRIDS to solve the above challenges**

Transcriptional Regulatory Redundancy Problem

Regulatory Redundancy Problem Given a target gene in **y**, what combinations of entries in **x** are regulating the expression of the target gene.

Examples

Let's say we have 4 cells, and a global removal is applied.

$$\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 1 & 1 \end{bmatrix} \overset{\mathcal{F}}{\Longrightarrow} \mathbf{Y} = \begin{bmatrix} 0.7 \\ 0.1 \\ 2.2 \\ 0.9 \end{bmatrix} \overset{\text{After}}{\Longrightarrow} \tilde{\mathbf{X}} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \overset{\mathcal{F}}{\Longrightarrow} \tilde{\mathbf{Y}} = \begin{bmatrix} 0.7 \\ 0.1 \\ 0.6 \downarrow \\ 0.9 \end{bmatrix}$$

Challenges

- The complex gene regulatory process function $\mathcal{F}$ is an indifferentiable black box
- The combinatorial effect of multiple CREs is important but has been ignored by most previous methods
- **x** is sparse, discrete, and high-dimension. The combinatorial solution space of $\text{subset}(\mathbf{x})$ is vast, which means brute-force search is intractable
- **In this work, we propose GRIDS to solve the above challenges**

Transcriptional Regulatory Redundancy Problem

Regulatory Redundancy Problem Given a target gene in **y**, what combinations of entries in **x** are regulating the expression of the target gene.

Examples

Let's say we have 4 cells, and a global removal is applied.

$$\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 1 & 1 \end{bmatrix} \overset{\mathcal{F}}{\Longrightarrow} \mathbf{Y} = \begin{bmatrix} 0.7 \\ 0.1 \\ 2.2 \\ 0.9 \end{bmatrix} \overset{\text{After}}{\Longrightarrow} \tilde{\mathbf{X}} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \overset{\mathcal{F}}{\Longrightarrow} \tilde{\mathbf{Y}} = \begin{bmatrix} 0.7 \\ 0.1 \\ 0.6 \downarrow \\ 0.9 \end{bmatrix}$$

Challenges

- The complex gene regulatory process function $\mathcal{F}$ is an indifferentiable black box
- The combinatorial effect of multiple CREs is important but has been ignored by most previous methods
- **x** is sparse, discrete, and high-dimension. The combinatorial solution space of $\text{subset}(\mathbf{x})$ is vast, which means brute-force search is intractable
- In this work, we propose GRIDS to solve the above challenges

Transcriptional Regulatory Redundancy Problem

Regulatory Redundancy Problem Given a target gene in **y**, what combinations of entries in **x** are regulating the expression of the target gene.

Examples

Let's say we have 4 cells, and a global removal is applied.

$$\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 1 & 1 \end{bmatrix} \overset{\mathcal{F}}{\Longrightarrow} \mathbf{Y} = \begin{bmatrix} 0.7 \\ 0.1 \\ 2.2 \\ 0.9 \end{bmatrix} \overset{\text{After}}{\Longrightarrow} \tilde{\mathbf{X}} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \overset{\mathcal{F}}{\Longrightarrow} \tilde{\mathbf{Y}} = \begin{bmatrix} 0.7 \\ 0.1 \\ 0.6 \downarrow \\ 0.9 \end{bmatrix}$$

Challenges

- The complex gene regulatory process function $\mathcal{F}$ is an indifferentiable black box
- The combinatorial effect of multiple CREs is important but has been ignored by most previous methods
- **x** is sparse, discrete, and high-dimension. The combinatorial solution space of $\text{subset}(\mathbf{x})$ is vast, which means brute-force search is intractable
- **In this work, we propose GRIDS to solve the above challenges**

Table of Contents

GRIDS: Build A Differentiable Surrogate $\hat{\mathcal{F}}$ with Neural Network

We train a differentiable surrogate $\hat{\mathcal{F}}$ to mimic the black-box $\mathcal{F}$

- a neural network to act as a differentiable surrogate $\hat{\mathcal{F}}$
- $\hat{\mathcal{F}}$ can predict the transcriptional regulatory process (an emulator to the black-box regulatory function $\mathcal{F}$).

$$\mathbf{y} = \hat{\mathcal{F}}(\mathbf{x})$$

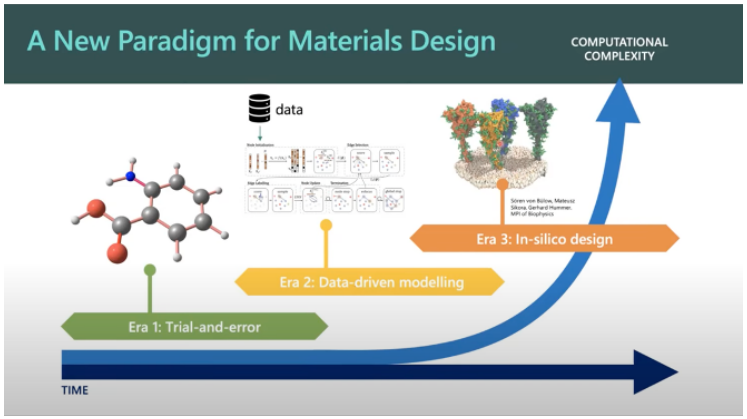


Figure: Deep Learning and the Natural Sciences: A Perfect Marriage?

Questions

- It requires an accurate surrogate model $\hat{\mathcal{F}}$ to $\mathcal{F}$
- Can we understand the transcriptional regulatory redundancy problem by interpreting the surrogate $\hat{\mathcal{F}}$?

GRIDS: Build A Differentiable Surrogate $\hat{\mathcal{F}}$ with Neural Network

We train a differentiable surrogate $\hat{\mathcal{F}}$ to mimic the black-box $\mathcal{F}$

- a neural network to act as a differentiable surrogate $\hat{\mathcal{F}}$
- $\hat{\mathcal{F}}$ can predict the transcriptional regulatory process (an emulator to the black-box regulatory function $\mathcal{F}$).

$$\mathbf{y} = \hat{\mathcal{F}}(\mathbf{x})$$

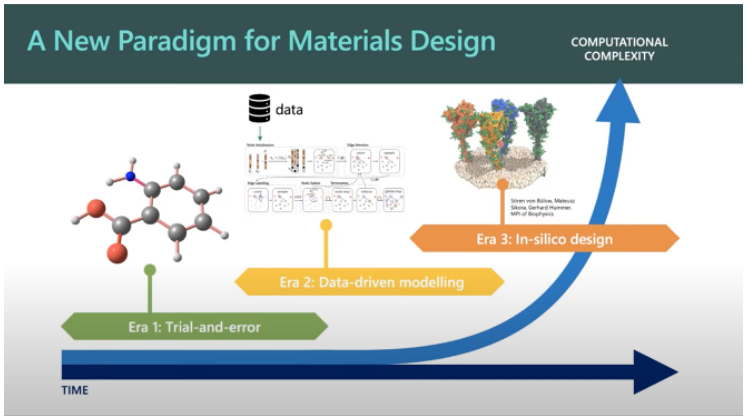


Figure: Deep Learning and the Natural Sciences: A Perfect Marriage?

Questions

- It requires an accurate surrogate model $\hat{\mathcal{F}}$ to $\mathcal{F}$
- Can we understand the transcriptional regulatory redundancy problem by interpreting the surrogate $\hat{\mathcal{F}}$?

GRIDS: Cross-modality Surrogate Mapping $\hat{\mathcal{F}}$

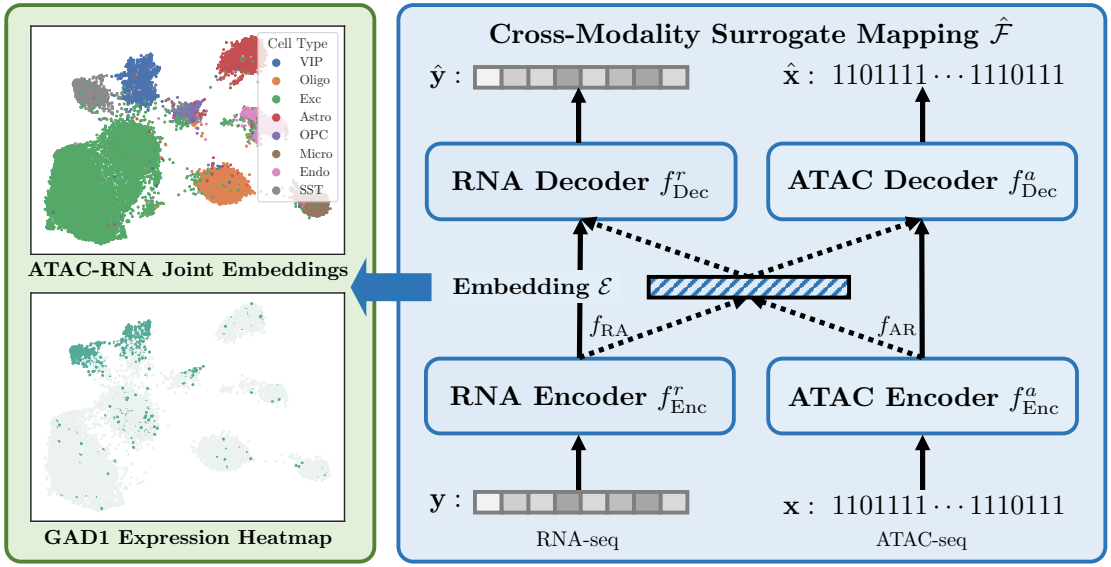
- $\hat{\mathcal{F}}$ modeled by Auto-encoders

$$\mathbf{h}_a^{(i)} = f_{\text{Enc}}^a(\mathbf{W}_{\text{Emb}}^a(\mathbf{x}^{(i)})),\ \mathbf{h}_r^{(i)} = f_{\text{Enc}}^r(\mathbf{W}_{\text{Emb}}^r(\mathbf{y}^{(i)}))$$

- Learn a joint embedding space to align RNA and ATAC sequences by adversarial training

$$\tilde{\mathbf{h}}_r^{(i)} = f_{\text{AR}}(\mathbf{h}_a^{(i)}),\ \tilde{\mathbf{h}}_a^{(i)} = f_{\text{RA}}(\mathbf{h}_r^{(i)})$$

- Enables cross-modality generations (e.g. ATAC-seq $\rightarrow$ RNA-seq)



GRIDS: Cross-modality Surrogate Mapping $\hat{\mathcal{F}}$

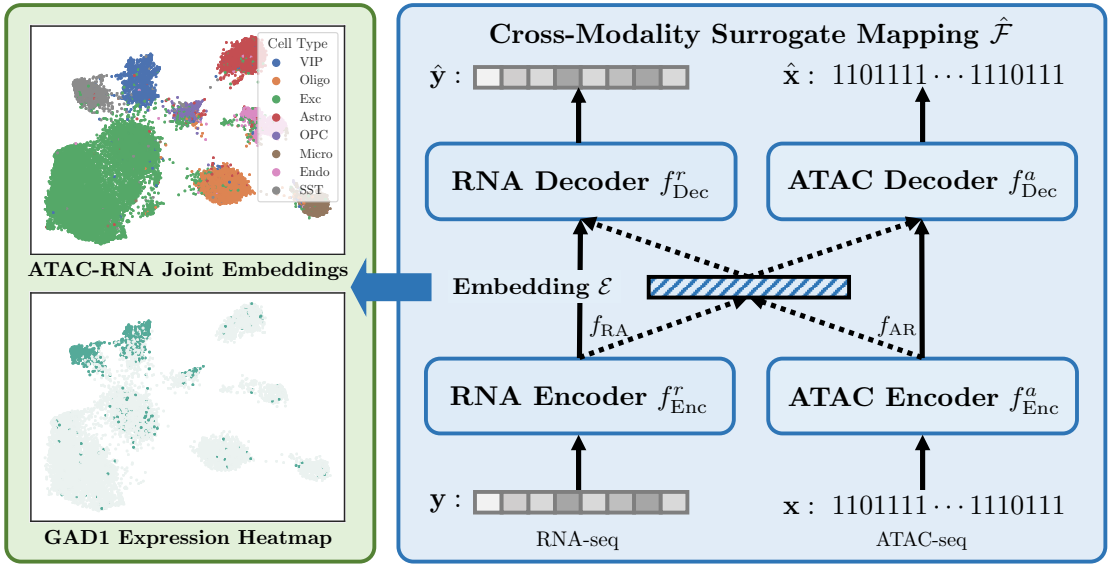
- $\hat{\mathcal{F}}$ modeled by Auto-encoders

$$\mathbf{h}_a^{(i)} = f_{\text{Enc}}^a(\mathbf{W}_{\text{Emb}}^a(\mathbf{x}^{(i)})),\ \mathbf{h}_r^{(i)} = f_{\text{Enc}}^r(\mathbf{W}_{\text{Emb}}^r(\mathbf{y}^{(i)}))$$

- Learn a joint embedding space to align RNA and ATAC sequences by adversarial training

$$\tilde{\mathbf{h}}_r^{(i)} = f_{\text{AR}}(\mathbf{h}_a^{(i)}),\ \tilde{\mathbf{h}}_a^{(i)} = f_{\text{RA}}(\mathbf{h}_r^{(i)})$$

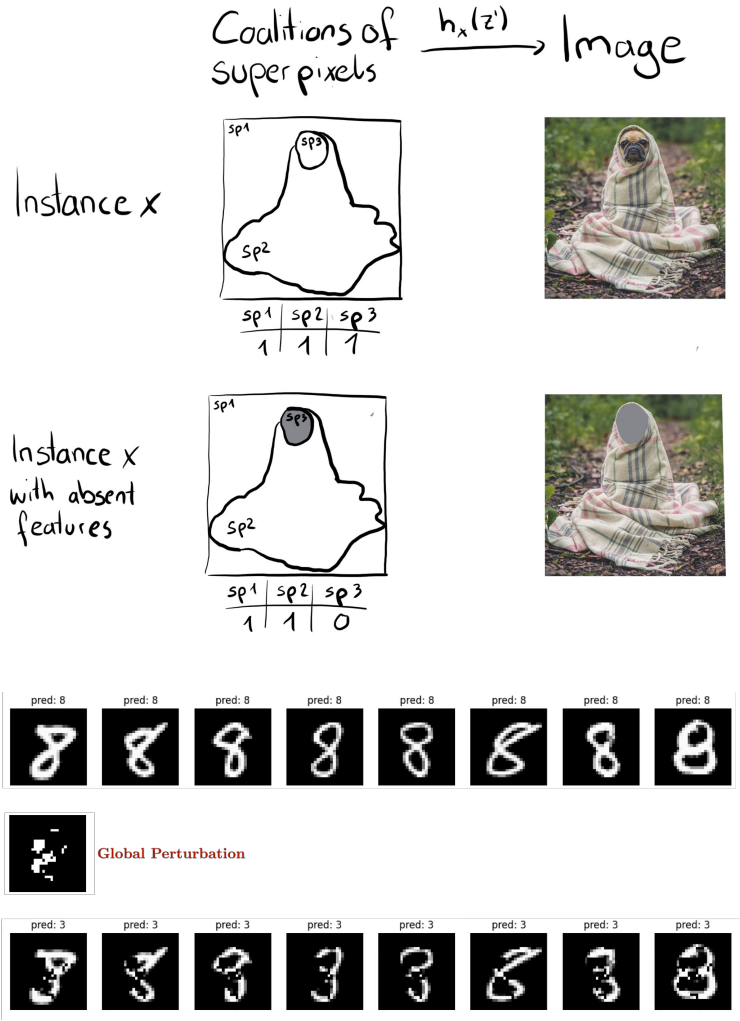
- Enables cross-modality generations (e.g. ATAC-seq $\rightarrow$ RNA-seq)



GRIDS: Global Feature Explanation for Regulatory Redundancy

Use global feature explanation to understand regulatory redundancy

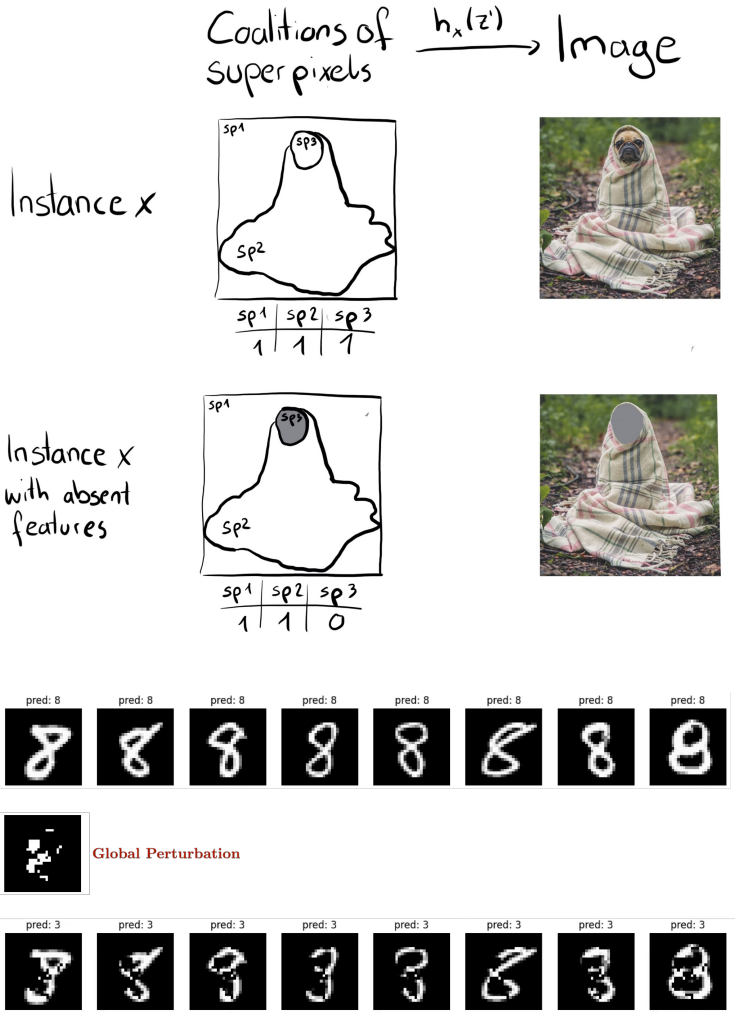
- **Feature Explanation** explains the reason of decision made by the model through adding perturbations on inputs
- Given a target gene in **y**, how to use global explanation to find a subset **r** of k indices out of the entire dimension of **x**?
- **Local Feature Explanation** seeks to explain individual predictions
  - Instance-wise explanation
- **Global Feature Explanation** seeks to characterize a model's decisions across a population of instances
  - MNIST example (flip model predictions from 8 to 3)



GRIDS: Global Feature Explanation for Regulatory Redundancy

Use global feature explanation to understand regulatory redundancy

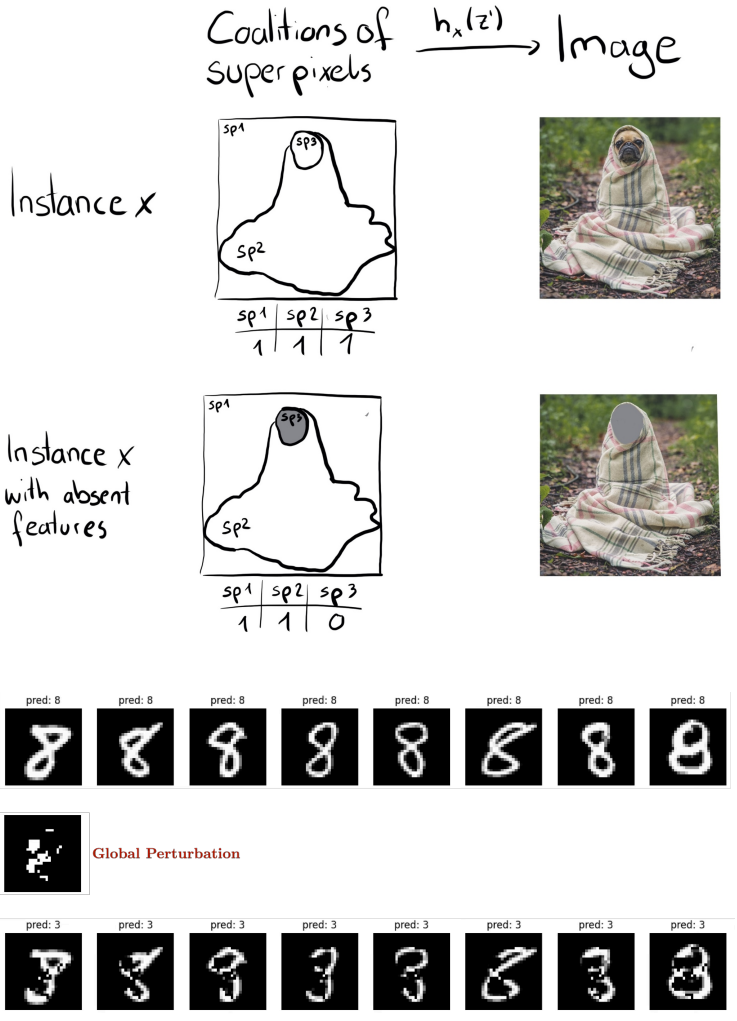
- **Feature Explanation** explains the reason of decision made by the model through adding perturbations on inputs
- Given a target gene in **y**, how to use global explanation to find a subset **r** of k indices out of the entire dimension of **x**?
- **Local Feature Explanation** seeks to explain individual predictions
  - Instance-wise explanation
- **Global Feature Explanation** seeks to characterize a model's decisions across a population of instances
  - MNIST example (flip model predictions from 8 to 3)



GRIDS: Global Feature Explanation for Regulatory Redundancy

Use global feature explanation to understand regulatory redundancy

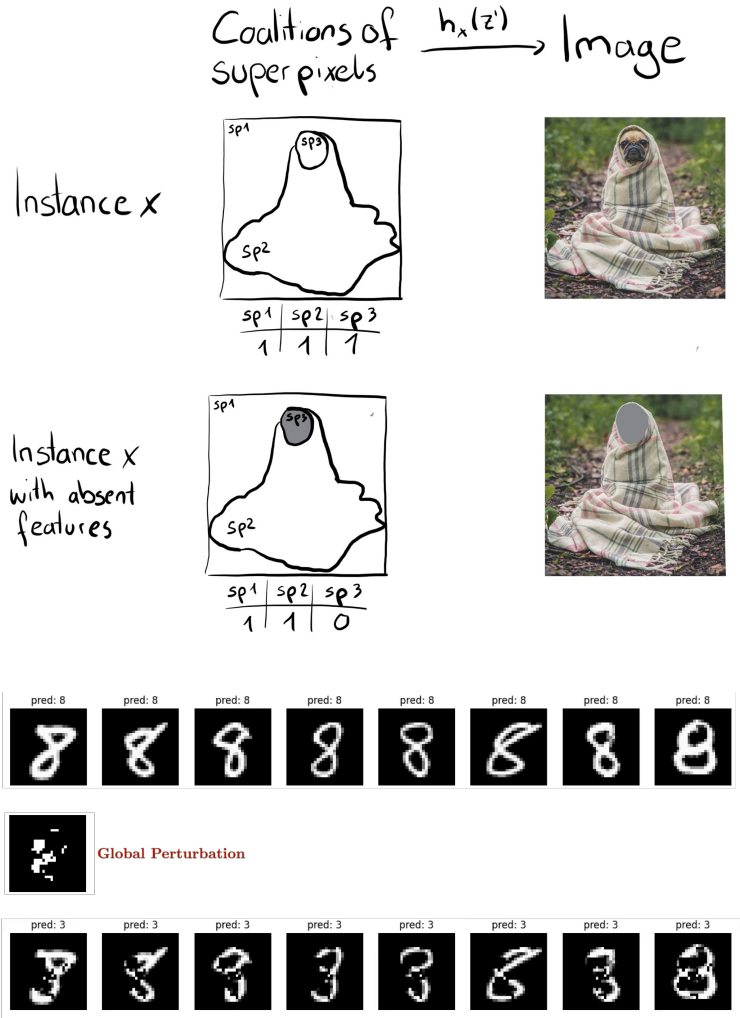
- **Feature Explanation** explains the reason of decision made by the model through adding perturbations on inputs
- Given a target gene in **y**, how to use global explanation to find a subset **r** of k indices out of the entire dimension of **x**?
- **Local Feature Explanation** seeks to explain individual predictions
  - Instance-wise explanation
- **Global Feature Explanation** seeks to characterize a model's decisions across a population of instances
  - MNIST example (flip model predictions from 8 to 3)



GRIDS: Global Feature Explanation for Regulatory Redundancy

Use global feature explanation to understand regulatory redundancy

- **Feature Explanation** explains the reason of decision made by the model through adding perturbations on inputs
- Given a target gene in **y**, how to use global explanation to find a subset **r** of k indices out of the entire dimension of **x**?
- **Local Feature Explanation** seeks to explain individual predictions
  - Instance-wise explanation
- **Global Feature Explanation** seeks to characterize a model's decisions across a population of instances
  - MNIST example (flip model predictions from 8 to 3)



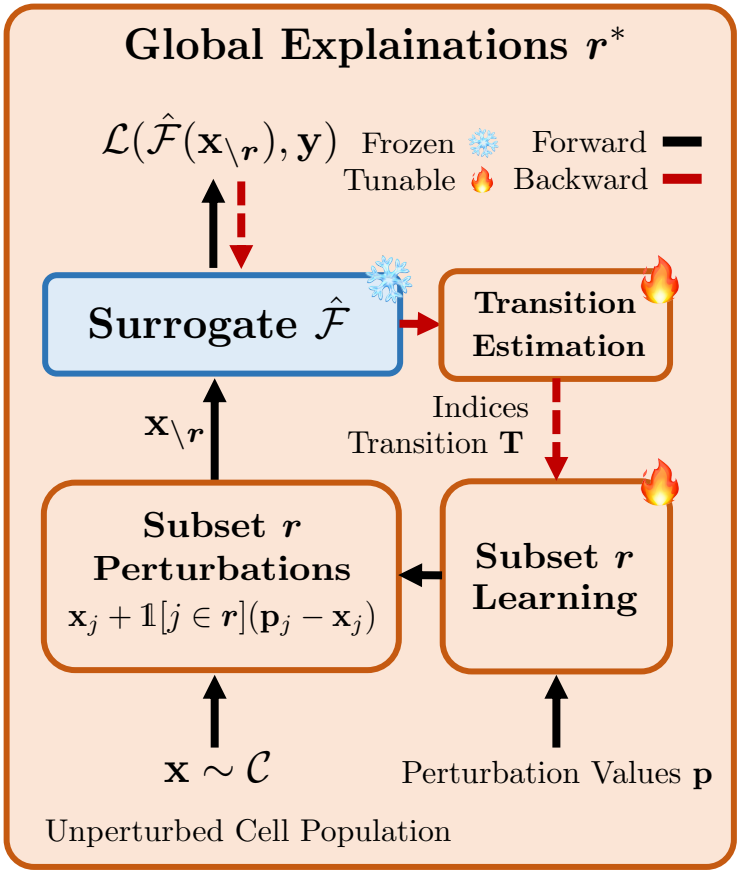
GRIDS: Finding Global Subset Perturbations by End-to-End Learning

Directly applying pervious global feature explanation method is challenging

- ATAC-seq $\mathbf{x}$ is sparse, discrete, and high-dimension
- Finding a combinatorial subset is too vast to be computationally tractable

Our solution: a global explanation method capable of performing end-to-end training on discrete data

- Optimize over dataset
- Utilize auto-differentiaion
- A unified perturbation operator
  - To make replacement operations be directly optimized as a continuous variable
  - Can be easily extended to different perturbation forms



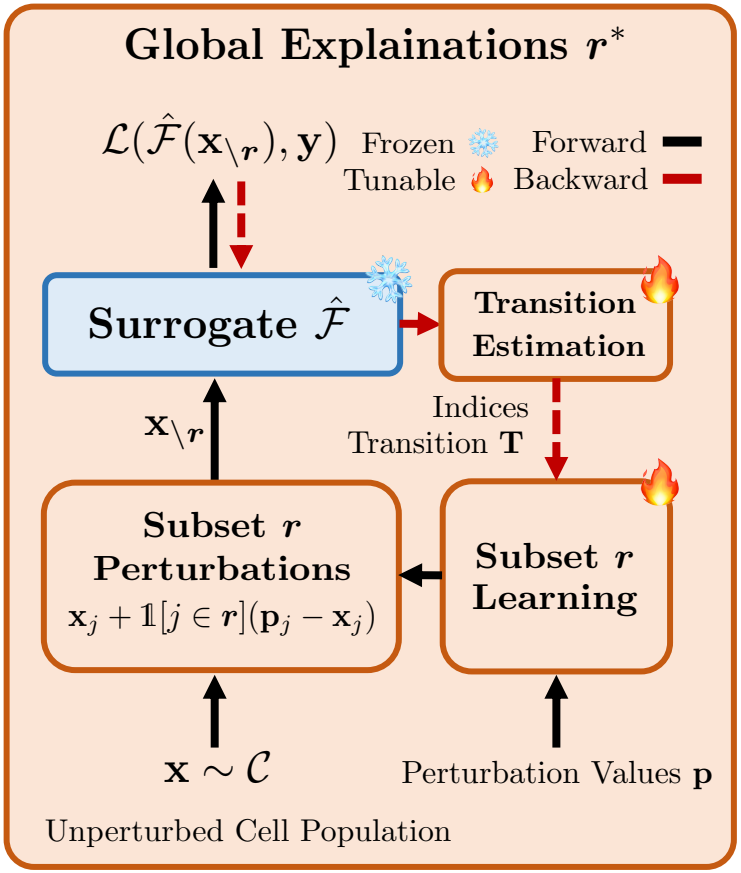
GRIDS: Finding Global Subset Perturbations by End-to-End Learning

Directly applying pervious global feature explanation method is challenging

- ATAC-seq $\mathbf{x}$ is sparse, discrete, and high-dimension
- Finding a combinatorial subset is too vast to be computationally tractable

Our solution: a global explanation method capable of performing end-to-end training on discrete data

- Optimize over dataset
- Utilize auto-differentiaion
- A unified perturbation operator
  - To make replacement operations be directly optimized as a continuous variable
  - Can be easily extended to different perturbation forms



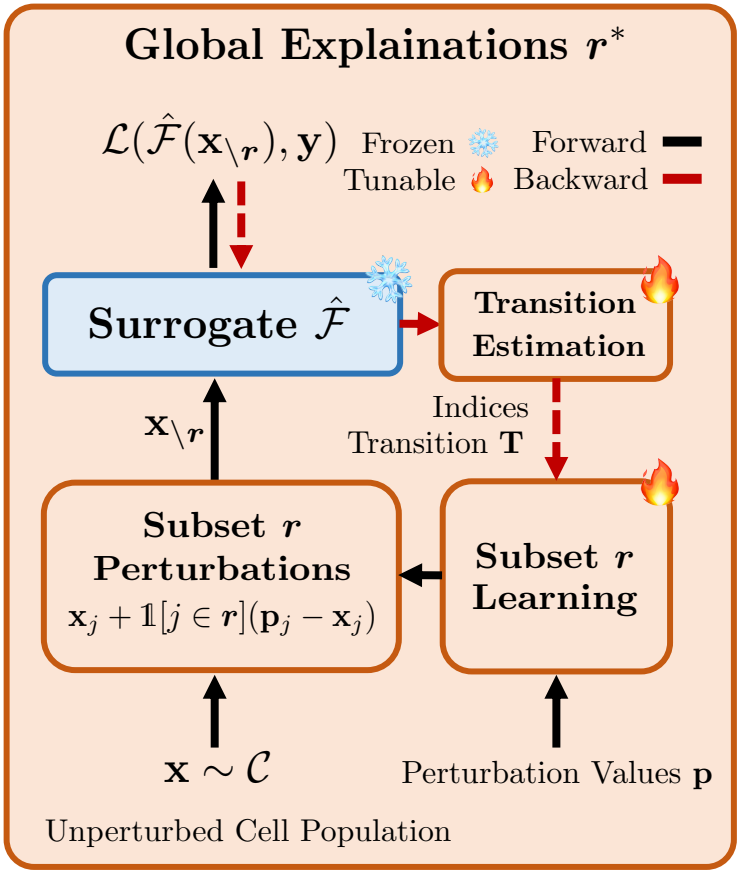
GRIDS: Finding Global Subset Perturbations by End-to-End Learning

Directly applying pervious global feature explanation method is challenging

- ATAC-seq $\mathbf{x}$ is sparse, discrete, and high-dimension
- Finding a combinatorial subset is too vast to be computationally tractable

Our solution: a global explanation method capable of performing end-to-end training on discrete data

- Optimize over dataset
- Utilize auto-differentiaion
- A unified perturbation operator
  - To make replacement operations be directly optimized as a continuous variable
  - Can be easily extended to different perturbation forms



GRIDS: Understanding Transcription via Global Feature Explanation

Global Explanation Measurement

How much a model's performance degrades over observed samples when features are perturbed

$$\boldsymbol{r}^* = \underset{\boldsymbol{r}}{\operatorname{argmin}}\, \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}[\mathcal{L}(\hat{\mathcal{F}}(\mathbf{x}_{\setminus \boldsymbol{r}}), \mathbf{y})]$$

$\boldsymbol{r}$ a subset of L peak indices $\boldsymbol{r} = \{r_1, \ldots, r_L\}$ to be perturbed
- across a population of cells $\mathcal{C}$
- solution space: $\binom{d_a}{L}$

$\mathbf{x}_{\setminus \boldsymbol{r}}$ denotes the perturbed features indicated by $\boldsymbol{r}$ (i.e, $\mathbf{x}_{\setminus \boldsymbol{r}, r_j} = \mathbf{p}_{r_j}$) (special case, removing features if $\mathbf{p} = \mathbf{0}$)

$\mathcal{L}$ a loss measurement for expected performance degradation

Problem

- The objective involves discrete operation, which is reformulated as $\mathbf{x}_{\setminus \boldsymbol{r}, j} = \mathbf{x}_j + \mathbf{1}[j \in \boldsymbol{r}](\mathbf{p}_j - \mathbf{x}_j)$ to enable auto-differentiation
- The challenge of combinatorial subsets persists

GRIDS: Understanding Transcription via Global Feature Explanation

Global Explanation Measurement

How much a model's performance degrades over observed samples when features are perturbed

$$\boldsymbol{r}^* = \underset{\boldsymbol{r}}{\operatorname{argmin}}\, \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}[\mathcal{L}(\hat{\mathcal{F}}(\mathbf{x}_{\setminus \boldsymbol{r}}), \mathbf{y})]$$

- $\boldsymbol{r}$ a subset of L peak indices $\boldsymbol{r} = \{r_1, \ldots, r_L\}$ to be perturbed
  - across a population of cells $\mathcal{C}$
  - solution space: $\binom{d_a}{L}$
- $\mathbf{x}_{\setminus \boldsymbol{r}}$ denotes the perturbed features indicated by $\boldsymbol{r}$ (i.e, $\mathbf{x}_{\setminus \boldsymbol{r}, r_j} = \mathbf{p}_{r_j}$) (special case, removing features if $\mathbf{p} = \mathbf{0}$)
- $\mathcal{L}$ a loss measurement for expected performance degradation

Problem

- The objective involves discrete operation, which is reformulated as $\mathbf{x}_{\setminus \boldsymbol{r}, j} = \mathbf{x}_j + \mathbf{1}[j \in \boldsymbol{r}](\mathbf{p}_j - \mathbf{x}_j)$ to enable auto-differentiation
- The challenge of combinatorial subsets persists

GRIDS: Understanding Transcription via Global Feature Explanation

Global Explanation Measurement

How much a model's performance degrades over observed samples when features are perturbed

$$\boldsymbol{r}^* = \underset{\boldsymbol{r}}{\operatorname{argmin}}\, \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}[\mathcal{L}(\hat{\mathcal{F}}(\mathbf{x}_{\setminus \boldsymbol{r}}), \mathbf{y})]$$

- $\boldsymbol{r}$ a subset of L peak indices $\boldsymbol{r} = \{r_1, \ldots, r_L\}$ to be perturbed
  - across a population of cells $\mathcal{C}$
  - solution space: $\binom{d_a}{L}$
- $\mathbf{x}_{\setminus \boldsymbol{r}}$ denotes the perturbed features indicated by $\boldsymbol{r}$ (i.e, $\mathbf{x}_{\setminus \boldsymbol{r}, r_j} = \mathbf{p}_{r_j}$) (special case, removing features if $\mathbf{p} = \mathbf{0}$)
- $\mathcal{L}$ a loss measurement for expected performance degradation

Problem

- The objective involves discrete operation, which is reformulated as $\mathbf{x}_{\setminus \boldsymbol{r}, j} = \mathbf{x}_j + \mathbf{1}[j \in \boldsymbol{r}](\mathbf{p}_j - \mathbf{x}_j)$ to enable auto-differentiation
- The challenge of combinatorial subsets persists

GRIDS: Understanding Transcription via Global Feature Explanation

Global Explanation Measurement

How much a model's performance degrades over observed samples when features are perturbed

$$\boldsymbol{r}^* = \underset{\boldsymbol{r}}{\operatorname{argmin}}\, \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}[\mathcal{L}(\hat{\mathcal{F}}(\mathbf{x}_{\setminus \boldsymbol{r}}), \mathbf{y})]$$

- $\boldsymbol{r}$ a subset of L peak indices $\boldsymbol{r} = \{r_1, \ldots, r_L\}$ to be perturbed
  - across a population of cells $\mathcal{C}$
  - solution space: $\binom{d_a}{L}$
- $\mathbf{x}_{\setminus \boldsymbol{r}}$ denotes the perturbed features indicated by $\boldsymbol{r}$ (i.e, $\mathbf{x}_{\setminus \boldsymbol{r}, r_j} = \mathbf{p}_{r_j}$) (special case, removing features if $\mathbf{p} = \mathbf{0}$)
- $\mathcal{L}$ a loss measurement for expected performance degradation

Problem

- The objective involves discrete operation, which is reformulated as $\mathbf{x}_{\setminus \boldsymbol{r}, j} = \mathbf{x}_j + \mathbf{1}[j \in \boldsymbol{r}](\mathbf{p}_j - \mathbf{x}_j)$ to enable auto-differentiation
- The challenge of combinatorial subsets persists

GRIDS: Learning Subset Perturbations by Estimating State Transition

Subset Transition Matrix for Gradient Estimation

- gradient w.r.t the input embedding

$$\mathbf{G} = \partial \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}[\mathcal{L}(\hat{\mathcal{F}}(\mathbf{x}_{\setminus \boldsymbol{r}}), \mathbf{y})] / \partial \mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{x}_{\setminus \boldsymbol{r}})$$

- construct the transition matrix $\mathbf{T} \in \mathbb{R}^{L \times d_a}$ via first-order approximation
- $\mathbf{T}_{i,j}$ in the matrix represents the advantage value of transitioning from replacing the previous index r_i with the new index j

$$\mathbf{d}_j = \mathbf{G}_j \cdot (\mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{p})_j - \mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{x})_j)$$
$$\mathbf{T}_{i,j} = \mathbf{1}[j \notin \boldsymbol{r}]\mathbf{d}_j - \mathbf{1}[j \neq r_i]\mathbf{d}_{r_i}$$

Perturbation Subset Update Mechanism

- Coordinate descent method for updating the subset with one element at a time
- Need to implement a custom PyTorch optimizer

GRIDS: Learning Subset Perturbations by Estimating State Transition

Subset Transition Matrix for Gradient Estimation

- gradient w.r.t the input embedding

$$\mathbf{G} = \partial \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}[\mathcal{L}(\hat{\mathcal{F}}(\mathbf{x}_{\setminus \boldsymbol{r}}), \mathbf{y})] / \partial \mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{x}_{\setminus \boldsymbol{r}})$$

- construct the transition matrix $\mathbf{T} \in \mathbb{R}^{L \times d_a}$ via first-order approximation
- $\mathbf{T}_{i,j}$ in the matrix represents the advantage value of transitioning from replacing the previous index r_i with the new index j

$$\mathbf{d}_j = \mathbf{G}_j \cdot (\mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{p})_j - \mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{x})_j)$$
$$\mathbf{T}_{i,j} = \mathbf{1}[j \notin \boldsymbol{r}]\mathbf{d}_j - \mathbf{1}[j \neq r_i]\mathbf{d}_{r_i}$$

Perturbation Subset Update Mechanism

- Coordinate descent method for updating the subset with one element at a time
- Need to implement a custom PyTorch optimizer

GRIDS: Learning Subset Perturbations by Estimating State Transition

Subset Transition Matrix for Gradient Estimation

- gradient w.r.t the input embedding

$$\mathbf{G} = \partial \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}[\mathcal{L}(\hat{\mathcal{F}}(\mathbf{x}_{\setminus \boldsymbol{r}}), \mathbf{y})] / \partial \mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{x}_{\setminus \boldsymbol{r}})$$

- construct the transition matrix $\mathbf{T} \in \mathbb{R}^{L \times d_a}$ via first-order approximation
- $\mathbf{T}_{i,j}$ in the matrix represents the advantage value of transitioning from replacing the previous index r_i with the new index j

$$\mathbf{d}_j = \mathbf{G}_j \cdot (\mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{p})_j - \mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{x})_j)$$
$$\mathbf{T}_{i,j} = \mathbf{1}[j \notin \boldsymbol{r}]\mathbf{d}_j - \mathbf{1}[j \neq r_i]\mathbf{d}_{r_i}$$

Perturbation Subset Update Mechanism

- Coordinate descent method for updating the subset with one element at a time
- Need to implement a custom PyTorch optimizer

GRIDS: Learning Subset Perturbations by Estimating State Transition

Subset Transition Matrix for Gradient Estimation

- gradient w.r.t the input embedding

$$\mathbf{G} = \partial \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}[\mathcal{L}(\hat{\mathcal{F}}(\mathbf{x}_{\setminus \boldsymbol{r}}), \mathbf{y})] / \partial \mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{x}_{\setminus \boldsymbol{r}})$$

- construct the transition matrix $\mathbf{T} \in \mathbb{R}^{L \times d_a}$ via first-order approximation
- $\mathbf{T}_{i,j}$ in the matrix represents the advantage value of transitioning from replacing the previous index r_i with the new index j

$$\mathbf{d}_j = \mathbf{G}_j \cdot (\mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{p})_j - \mathbf{W}^{a}_{\mathrm{Emb}}(\mathbf{x})_j)$$
$$\mathbf{T}_{i,j} = \mathbf{1}[j \notin \boldsymbol{r}]\mathbf{d}_j - \mathbf{1}[j \neq r_i]\mathbf{d}_{r_i}$$

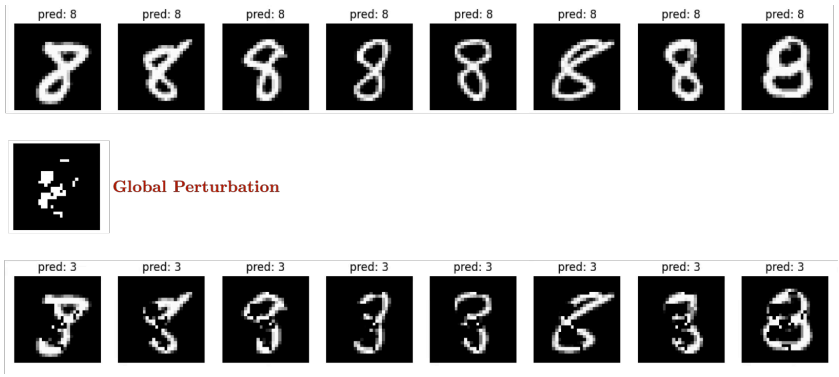
Perturbation Subset Update Mechanism

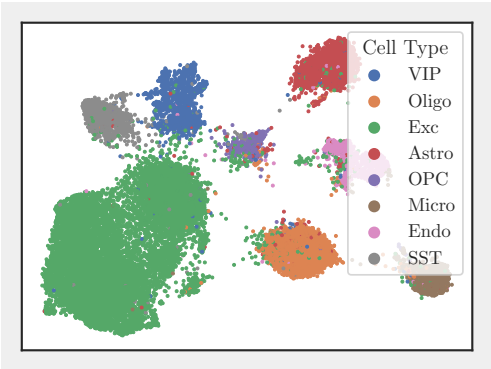
- Coordinate descent method for updating the subset with one element at a time
- Need to implement a custom PyTorch optimizer

Table of Contents

Datasets & Experimental Setup

- **MNIST Dataset** To find combinatorial effects found by GRIDS, we conducted experiments on the binary digit classification using MNIST as a start, to better explain our idea.
  - Flip model predictions from 8 to 3
- **Real Single-cell Dataset** We curated a set of deeply-sequenced single-cell multi-modal data from postmortem human pre-frontal cortex (PFC).
  - $10,266$ cells with 8 different cell types

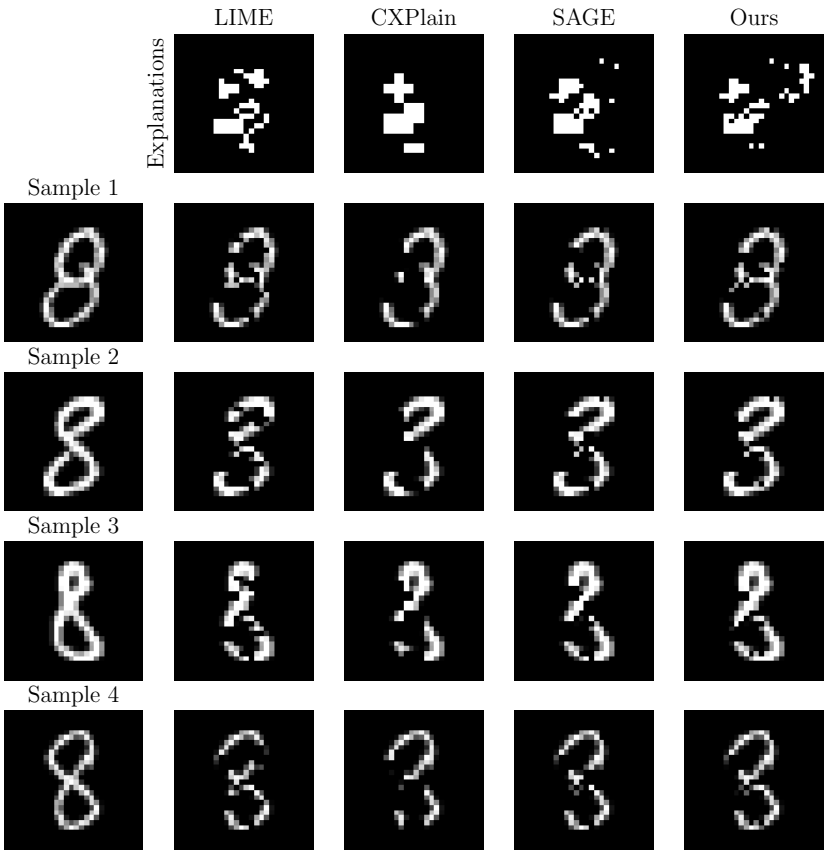




Demo Experiments on MNIST with GRIDS

Experiment Setup and Results

- A pretrained binary classifier (8 vs. 3) that achieves 97.9% accuracy over the test set
- Flip prediction from 8 to 3 by masking $L = 64$ the most important pixels (i.e., $\mathbf{p} = \mathbf{0}$)
- GRIDS achieves a combinatorial pattern similar to that of SAGE



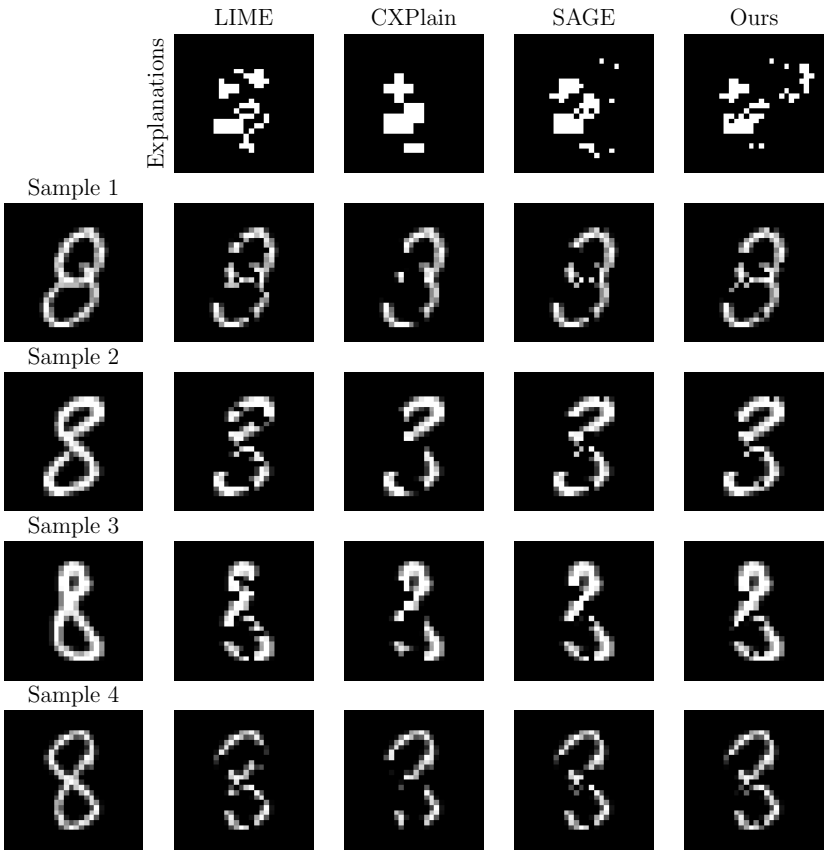
Apply perturbations to ATAC-seq $\mathbf{x}$

The pixel here is analogous to the binary entry in the ATAC-seq $\mathbf{x}$, while the binary classifier corresponds to the surrogate $\hat{\mathcal{F}}$

Demo Experiments on MNIST with GRIDS

Experiment Setup and Results

- A pretrained binary classifier (8 vs. 3) that achieves 97.9% accuracy over the test set
- Flip prediction from 8 to 3 by masking $L = 64$ the most important pixels (i.e., $\mathbf{p} = \mathbf{0}$)
- GRIDS achieves a combinatorial pattern similar to that of SAGE



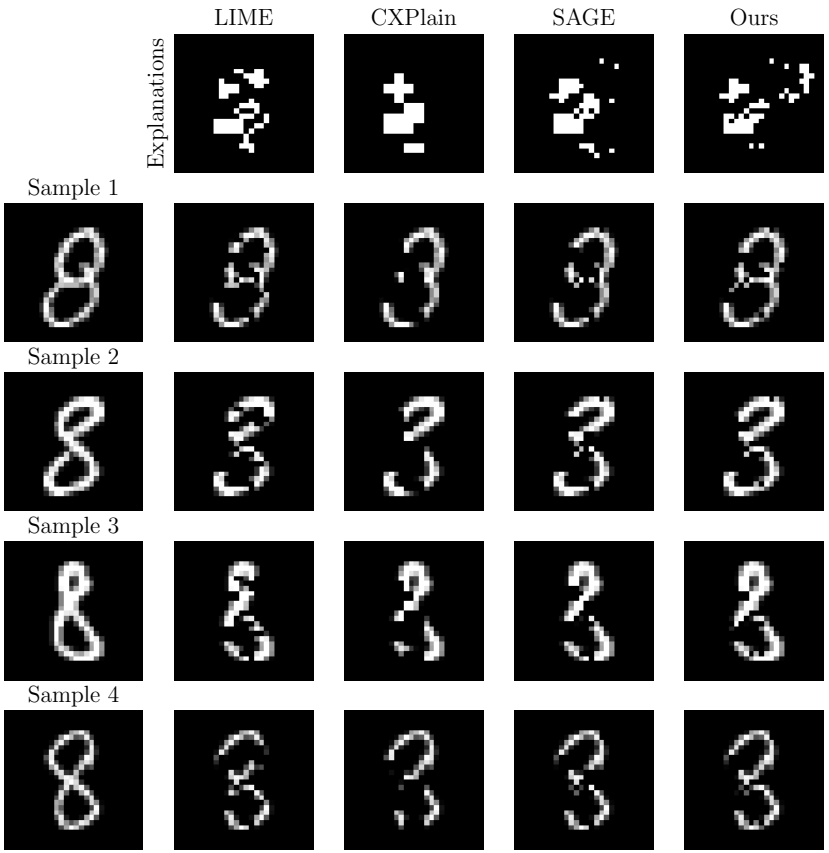
Apply perturbations to ATAC-seq $\mathbf{x}$

The pixel here is analogous to the binary entry in the ATAC-seq $\mathbf{x}$, while the binary classifier corresponds to the surrogate $\hat{\mathcal{F}}$

Demo Experiments on MNIST with GRIDS

Experiment Setup and Results

- A pretrained binary classifier (8 vs. 3) that achieves 97.9% accuracy over the test set
- Flip prediction from 8 to 3 by masking $L = 64$ the most important pixels (i.e., $\mathbf{p} = \mathbf{0}$)
- GRIDS achieves a combinatorial pattern similar to that of SAGE

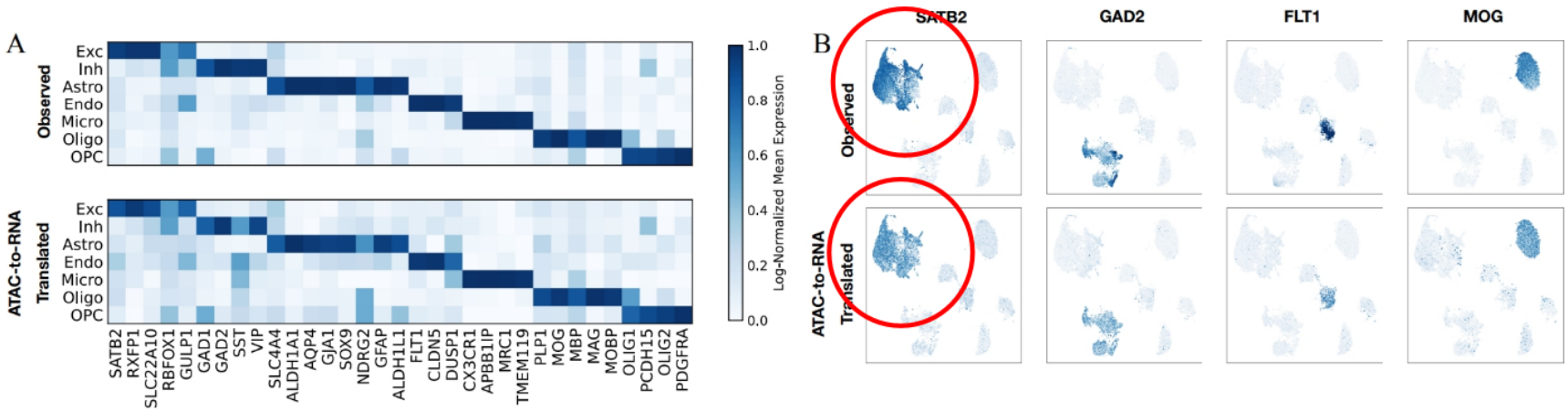


Apply perturbations to ATAC-seq $\mathbf{x}$

The pixel here is analogous to the binary entry in the ATAC-seq $\mathbf{x}$, while the binary classifier corresponds to the surrogate $\hat{\mathcal{F}}$

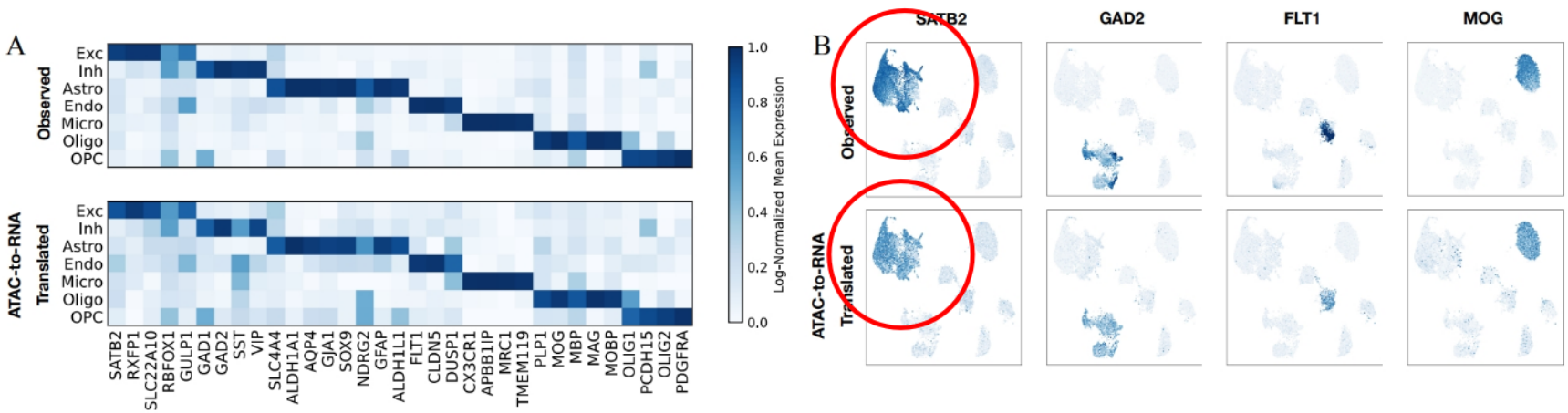
The Surrogate Model $\hat{\mathcal{F}}$ Accurately Models the ATAC-to-RNA Relationship

- Comparing the **mean expressions** between cell types and between the observed and translated cohort
- The averaged expression change (Avg. Δ) and the ratio of expression change against the original value (Rel. Δ)
- The predicted marker gene expression with actual values are pretty same
- UMAP of four key marker genes also similar



The Surrogate Model $\hat{\mathcal{F}}$ Accurately Models the ATAC-to-RNA Relationship

- Comparing the **mean expressions** between cell types and between the observed and translated cohort
- The averaged expression change (Avg. Δ) and the ratio of expression change against the original value (Rel. Δ)
- The predicted marker gene expression with actual values are pretty same
- UMAP of four key marker genes also similar



GRIDS consistently outperforms all baselines across each cell type on finding global explanations

- Learning global perturbations for specific marker gene of each cell type
- Among all the model, GRIDS achieves worst avg, which means it can find important CREs

Cell Type	Random		Saliency		SmoothGrad		FIMAP		GRIDS	
	Avg. Δ	Rel. Δ(%)	Avg. Δ	Rel. Δ(%)	Avg. Δ	Rel. Δ(%)	Avg. Δ	Rel. Δ(%)	Avg. Δ	Rel. Δ(%)
Astro	−0.085	−0.015	−2.163	−0.601	−2.155	−0.621	−13.502	−4.254	**−16.696**	**−5.837**
Endo	−1.073	−0.138	−4.974	−0.372	−9.726	−0.995	−38.997	−9.303	**−57.477**	**−11.816**
Micro	−0.012	−0.026	−23.757	−1.545	−32.944	−2.083	−73.752	−6.248	**−90.607**	**−7.671**
OPC	+0.823	−0.087	−54.645	−2.338	−48.438	−2.067	−77.167	−6.260	**−96.661**	**−8.256**
Oligo	−0.058	+0.026	−0.558	−0.173	−0.939	−0.220	−10.917	−4.252	**−16.760**	**−6.896**
SST	+0.159	+0.080	−5.201	−2.006	−5.201	−2.006	−16.453	−5.660	**−17.677**	**−6.365**
VIP	+0.012	+0.001	−0.654	−1.189	−0.634	−1.160	−2.732	−3.797	**−6.804**	**−7.195**
Avg.	+0.016	−0.021	−12.988	−1.209	−13.519	−1.290	−30.268	−5.367	**−39.103**	**−7.300**
Astro	−1.793	−0.533	−15.511	−4.853	−18.505	−6.217	−82.565	−24.766	**−100.556**	**−34.633**
Endo	+2.554	+0.468	−46.160	−6.217	−52.383	−7.893	−252.338	−41.790	**−259.920**	**−44.601**
Micro	−9.091	−0.490	−131.512	−9.122	−145.561	−10.116	−451.210	−39.695	**−470.430**	**−44.114**
OPC	−1.848	−0.165	−193.739	−10.260	−186.235	−9.891	**−415.231**	−35.687	−392.326	**−36.380**
Oligo	−1.134	−0.211	−19.809	−6.382	−21.136	−7.630	−69.460	−28.175	**−93.518**	**−38.982**
SST	−1.681	−0.615	−33.589	−11.675	−32.275	−11.115	−86.191	−29.198	**−93.772**	**−33.708**
VIP	+0.071	+0.002	−4.014	−4.876	−3.872	−4.782	−13.054	−16.757	**−19.703**	**−27.221**
Avg.	−1.843	−0.237	−68.620	−7.618	−70.292	−8.212	−202.368	−30.787	**−209.583**	**−36.893**

GRIDS consistently outperforms all baselines across each cell type on finding global explanations

- Learning global perturbations for specific marker gene of each cell type
- Among all the model, GRIDS achieves worst avg, which means it can find important CREs

Cell Type	Random		Saliency		SmoothGrad		FIMAP		GRIDS	
	Avg. Δ	Rel. Δ(%)	Avg. Δ	Rel. Δ(%)	Avg. Δ	Rel. Δ(%)	Avg. Δ	Rel. Δ(%)	Avg. Δ	Rel. Δ(%)
Astro	−0.085	−0.015	−2.163	−0.601	−2.155	−0.621	−13.502	−4.254	**−16.696**	**−5.837**
Endo	−1.073	−0.138	−4.974	−0.372	−9.726	−0.995	−38.997	−9.303	**−57.477**	**−11.816**
Micro	−0.012	−0.026	−23.757	−1.545	−32.944	−2.083	−73.752	−6.248	**−90.607**	**−7.671**
OPC	+0.823	−0.087	−54.645	−2.338	−48.438	−2.067	−77.167	−6.260	**−96.661**	**−8.256**
Oligo	−0.058	+0.026	−0.558	−0.173	−0.939	−0.220	−10.917	−4.252	**−16.760**	**−6.896**
SST	+0.159	+0.080	−5.201	−2.006	−5.201	−2.006	−16.453	−5.660	**−17.677**	**−6.365**
VIP	+0.012	+0.001	−0.654	−1.189	−0.634	−1.160	−2.732	−3.797	**−6.804**	**−7.195**
Avg.	+0.016	−0.021	−12.988	−1.209	−13.519	−1.290	−30.268	−5.367	**−39.103**	**−7.300**
Astro	−1.793	−0.533	−15.511	−4.853	−18.505	−6.217	−82.565	−24.766	**−100.556**	**−34.633**
Endo	+2.554	+0.468	−46.160	−6.217	−52.383	−7.893	−252.338	−41.790	**−259.920**	**−44.601**
Micro	−9.091	−0.490	−131.512	−9.122	−145.561	−10.116	−451.210	−39.695	**−470.430**	**−44.114**
OPC	−1.848	−0.165	−193.739	−10.260	−186.235	−9.891	**−415.231**	−35.687	−392.326	**−36.380**
Oligo	−1.134	−0.211	−19.809	−6.382	−21.136	−7.630	−69.460	−28.175	**−93.518**	**−38.982**
SST	−1.681	−0.615	−33.589	−11.675	−32.275	−11.115	−86.191	−29.198	**−93.772**	**−33.708**
VIP	+0.071	+0.002	−4.014	−4.876	−3.872	−4.782	−13.054	−16.757	**−19.703**	**−27.221**
Avg.	−1.843	−0.237	−68.620	−7.618	−70.292	−8.212	−202.368	−30.787	**−209.583**	**−36.893**

GRIDS consistently outperforms all baselines across highly-expressed gene sets from two representative cell types

- Top 100 Highly Expressed Genes Expression Changes in VIP and Microglia by masking L important CREs
- GRIDS can also gives global explanation even with a wide range of genes

Type	L	Method	Avg. Δ	Rel. Δ (%)
VIP-100	10	Random	−0.448	−0.009
		Saliency	−18.822	−0.915
		SmoothGrad	−18.424	−0.927
		FIMAP	−56.469	−3.087
		GRIDS	**−64.016**	**−3.827**
Microglia-100	10	Random	−0.333	−0.008
		Saliency	−42.372	−1.941
		SmoothGrad	−44.125	−2.073
		FIMAP	−115.092	−5.863
		GRIDS	**−141.339**	**−7.466**

GRIDS consistently outperforms all baselines across highly-expressed gene sets from two representative cell types

- Top 100 Highly Expressed Genes Expression Changes in VIP and Microglia by masking L important CREs
- GRIDS can also gives global explanation even with a wide range of genes

Type	L	Method	Avg. Δ	Rel. Δ (%)
VIP-100	10	Random	−0.448	−0.009
		Saliency	−18.822	−0.915
		SmoothGrad	−18.424	−0.927
		FIMAP	−56.469	−3.087
		GRIDS	**−64.016**	**−3.827**
Microglia-100	10	Random	−0.333	−0.008
		Saliency	−42.372	−1.941
		SmoothGrad	−44.125	−2.073
		FIMAP	−115.092	−5.863
		GRIDS	**−141.339**	**−7.466**

GRIDS can identify CREs with biology insights

- Calculate CRE-to-gene distance with Soft Hit Ratio and Hit Ratio
- Soft Hit Ratio (SHR) measures how many of the reported L CREs are located in this neighborhood
- Hit Ratio (HR) calculates an exact match
- GRIDS's global explanations gives a larger percent of directly interacting CREs verified in experiments

Method	$L = 10$		$L = 128$	
	HR ↑	SHR ↑	HR ↑	SHR ↑
Saliency	0.00	0.00	6.25	18.75
SmoothGrad	0.00	0.00	0.00	18.75
FIMAP	12.50	12.50	18.75	56.25
GRIDS	**18.75**	**25.00**	**31.25**	**68.75**

Table of Contents

Conclusion

- GRIDS uses a combination of surrogate modeling and learnable perturbations to analyze the regulatory redundancy problem.
- The findings indicate that GRIDS provides more semantically meaningful feature importance values, enabling effective analysis of regulatory redundancy across extensive genome regions.
- To our knowledge, this study is the first to integrate global feature explanations with regulatory redundancy analysis in the context of single-cell multi-modal data.
- GRIDS has the potential to significantly impact biological research.

Conclusion

- GRIDS uses a combination of surrogate modeling and learnable perturbations to analyze the regulatory redundancy problem.
- The findings indicate that GRIDS provides more semantically meaningful feature importance values, enabling effective analysis of regulatory redundancy across extensive genome regions.
- To our knowledge, this study is the first to integrate global feature explanations with regulatory redundancy analysis in the context of single-cell multi-modal data.
- GRIDS has the potential to significantly impact biological research.

Conclusion

- GRIDS uses a combination of surrogate modeling and learnable perturbations to analyze the regulatory redundancy problem.
- The findings indicate that GRIDS provides more semantically meaningful feature importance values, enabling effective analysis of regulatory redundancy across extensive genome regions.
- To our knowledge, this study is the first to integrate global feature explanations with regulatory redundancy analysis in the context of single-cell multi-modal data.
- GRIDS has the potential to significantly impact biological research.

Conclusion

- GRIDS uses a combination of surrogate modeling and learnable perturbations to analyze the regulatory redundancy problem.
- The findings indicate that GRIDS provides more semantically meaningful feature importance values, enabling effective analysis of regulatory redundancy across extensive genome regions.
- To our knowledge, this study is the first to integrate global feature explanations with regulatory redundancy analysis in the context of single-cell multi-modal data.
- GRIDS has the potential to significantly impact biological research.

Acknowledgments

Funding

National Institutes of Health
[R01HG012572, R01NS128523]

Thanks

Prof. Jing Zhang
Dr. Martin Renqiang Min
Mr. Yaqi Hu

Links

Paper GitHub Positions

Zhang Lab at UC Irvine

Thanks for your listening!
Q & A