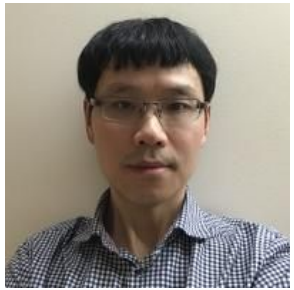
Spatial Transcriptomics

Project Proposal

November 8, 2022

Project Sponsor



Dr. Jia Rui Ding
PhD

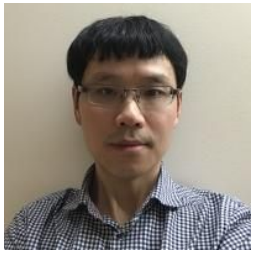
ENPH 479 - team 2255

Sayem Zaman

Daniel Chen

David Wang

Project Sponsor



Dr. Jia Rui Ding

- Assistant Professor in UBC Computer Science Department
- Postdoc Associate, Broad Institute of MIT & Harvard (2017 - 2021)

Research Interest

- Bioinformatics
- Computational Biology
- Machine Learning
- Probabilistic Deep Learning
- Single-Cell Genomics

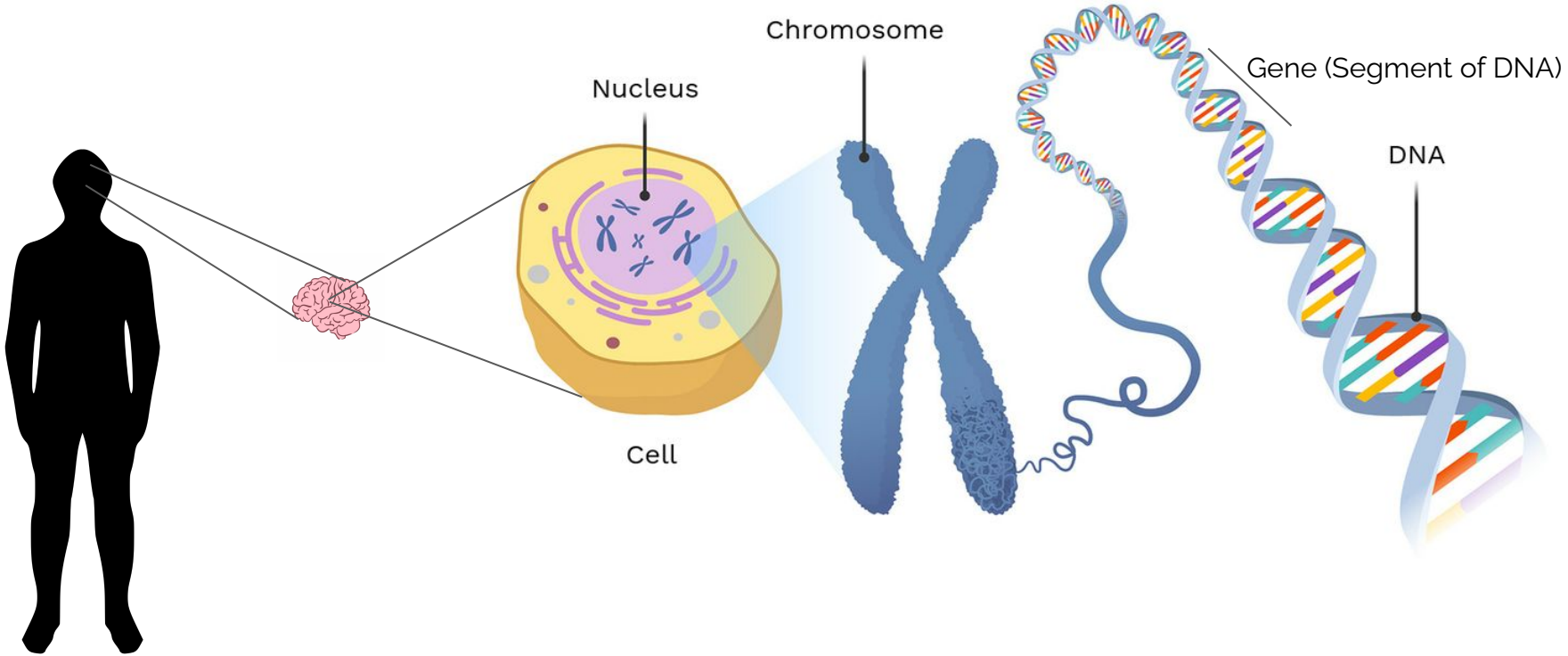
Project Proposal Outline

- Introduction
  - Impacts
  - RNA Sequencing & Spatial Transcriptomics
- Project Detail
  - Project Goal
  - Project Workflow
- Technical Implementation Details
  - VAE & VGAE
  - Map Function between Latent Spaces
  - Testing and Validation
- System Level Diagram
- Project Timeline
- Deliverables

Roughly **9.6 million** people die from cancer every year

1 in every 6 deaths is due to cancer

Cells and DNA



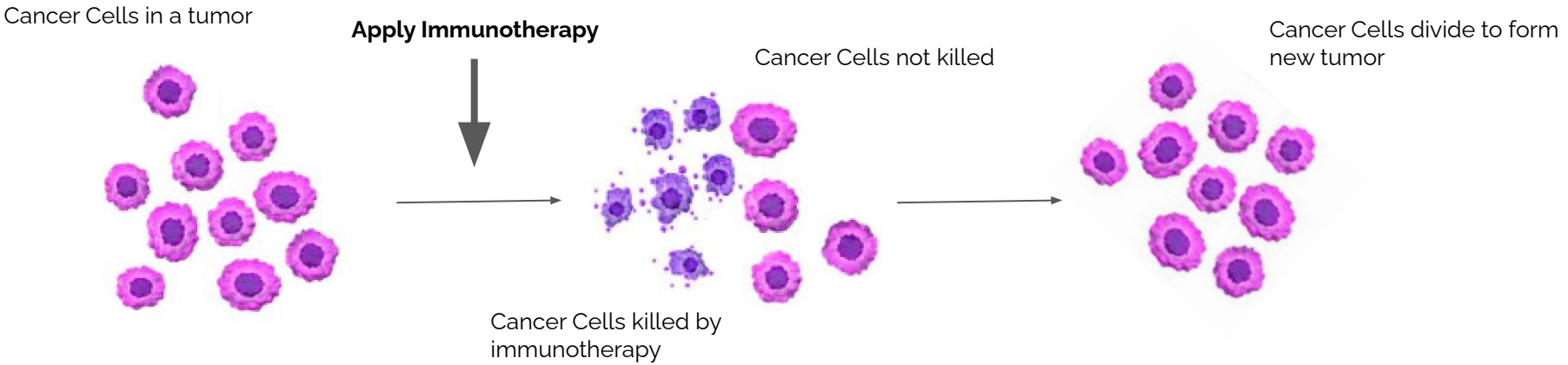
Immunotherapy for treating cancer

Our immune system has the ability to find and destroy cancer cells.

But cancer cells can sometimes **hide from the immune system** and avoid being destroyed. Cancer cells may also stop the immune system from working properly.

Immunotherapy helps to **strengthen or restore the immune system's natural ability to fight cancer** (with very little damage to patient's body)

Cancer resistance to Immunotherapy

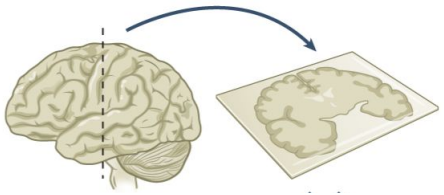


A promising approach...

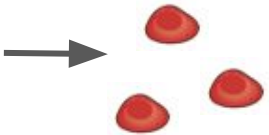
If we know the **gene expression information** (which genes) and **spatial information** (where they are) of cancer cells not affected by immunotherapy, we can learn to make better treatments

How to get **gene expression** information from cells

Single Cell RNA Sequencing (scRNA-seq)



Slice of tissue



Isolate individual cells

Extract RNA

	Cell 1	Cell 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 20,000	22	0	…

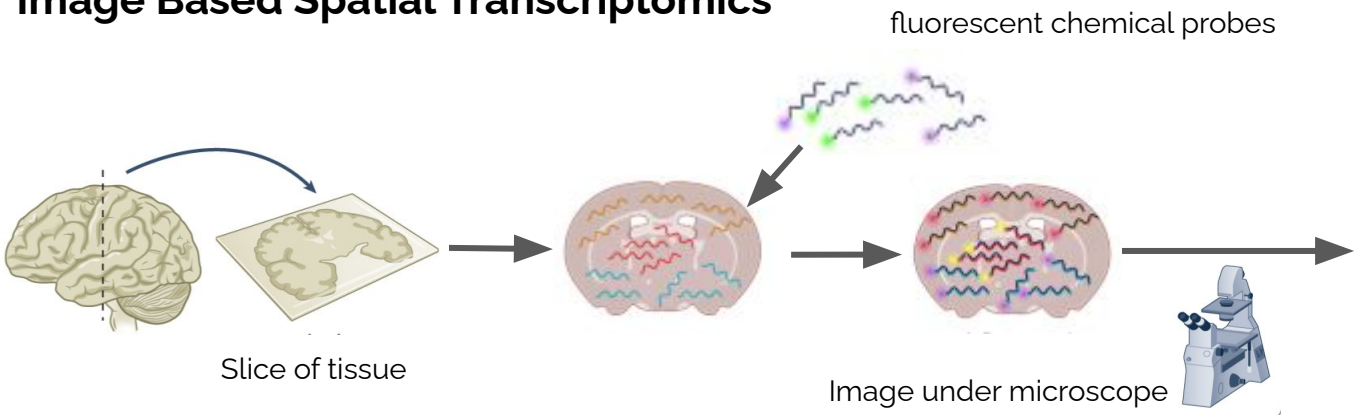
Cell-gene matrix

Extracts all genetic data in the cell (i.e. 20,000 genes)

Problem: we don't know what part of the original tissue the genes came from

How to get **spatial** information from cells

Image Based Spatial Transcriptomics



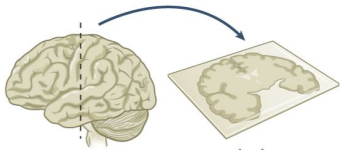
	Spot 1	Spot 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 500	22	0	…

X	137.9	74.8	…
Y	24.7	64.0	…

Identifies spatial location of genes, but due to chemical limits only a smaller number of genes can be profiled

Problem: we can only recover a small number of genes and their locations (~500)

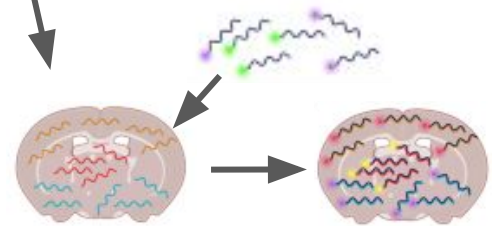
Goal: Get both **spatial** and **genetic** information from cells



scRNA sequencing

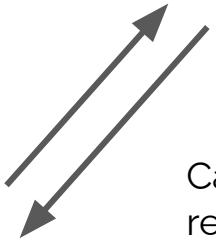
	Cell 1	Cell 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 20,000	22	0	…

Spatial Transcriptomics



	Spot 1	Spot 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 500	22	0	…

X	137.9	74.8	…
Y	24.7	64.0	…



Can the relationship be learned?

Project Goal

Our project aims to develop a **probabilistic machine learning model** to map the image-based spatial transcriptomics data to scRNA sequencing data.

- This will allow us to **infer the spatial information of cells based on the genetic information** of the RNA molecules within each cell.

Data Collection → **Generate ML Models** → **Relationship Between Models**

- **Image-Based Spatial Transcriptomics** and **Single-Cell RNA Sequencing**

- High dimensional data (tens of thousands of features)

How can we analyze the data using machine learning?

scRNA seq data

	Cell 1	Cell 2	...
Gene 1	18	0	...
Gene 2	1010	506	...
...	...	...	...
Gene 20,000	22	0	...

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & n \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array}$$

Sparse Matrix (Lots of zeros in data matrix)

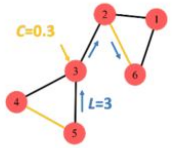
Spatial transcriptomics data

	Spot 1	Spot 2	...
Gene 1	18	0	...
Gene 2	1010	506	...
...	...	...	...
Gene 500	22	0	...

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & n \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array}$$

X	137.9	74.8	...
Y	24.7	64.0	...

0	1	0	0	0	1
1	0	1	0	0	1
0	1	0	1	1	0
0	0	1	0	1	0
0	0	1	1	0	0
1	1	0	0	0	0



Adjacency Matrix / Graph (Shows the spatial relationship to other cells)

High dimensional dataset

Another high dimensional dataset

Data is very complex and very different, hard to find direct mapping

Can we first transform the data into a more similar space, and still preserve the important features?

scRNA seq data

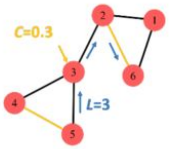
	Cell 1	Cell 2	...
Gene 1	18	0	...
Gene 2	1010	506	...
...	...	...	...
Gene 20,000	22	0	...

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & n \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array}$$

Very Hard

Spatial transcriptomics data

0	1	0	0	0	1
1	0	1	0	0	1
0	1	0	1	1	0
0	0	1	0	1	0
0	0	1	1	0	0
1	1	0	0	0	0



	Spot 1	Spot 2	...
Gene 1	18	0	...
Gene 2	1010	506	...
...	...	...	...
Gene 500	22	0	...

X	137.9	74.8	...
Y	24.7	64.0	...

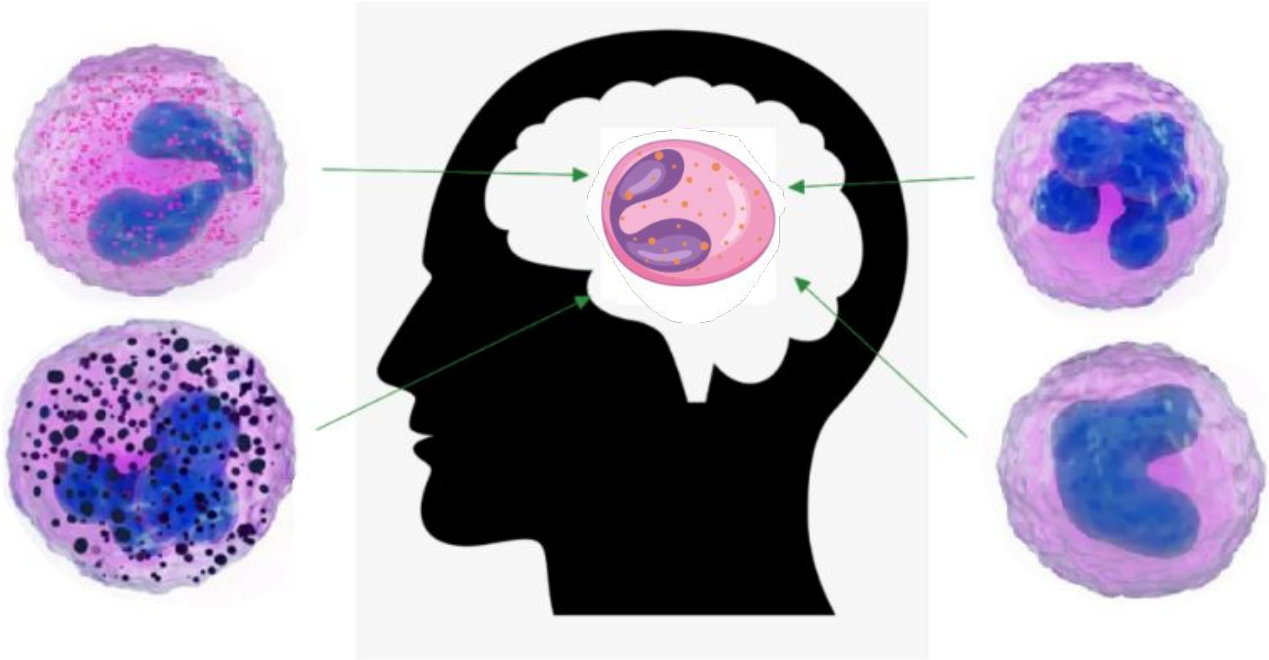
"Latent Space"

Transformed Representation

Easy

Transformed Representation

Latent space representation



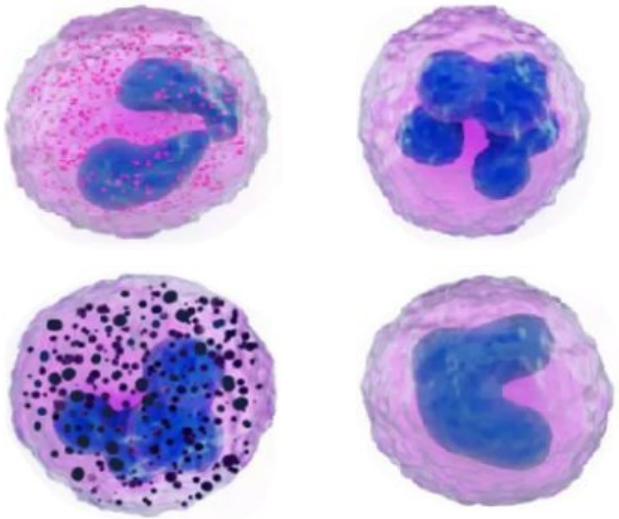
Shared features:

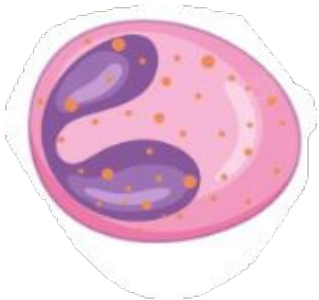
1. Light or pinkish color
2. Relatively round shape
3. Well defined nucleus

Understanding Latent spaces

A Latent Space is an abstract space that **encodes** a meaningful **compressed** internal representation of externally observed **high dimensional** events

High Multidimensional Space → **Lower Dimensional Representation**

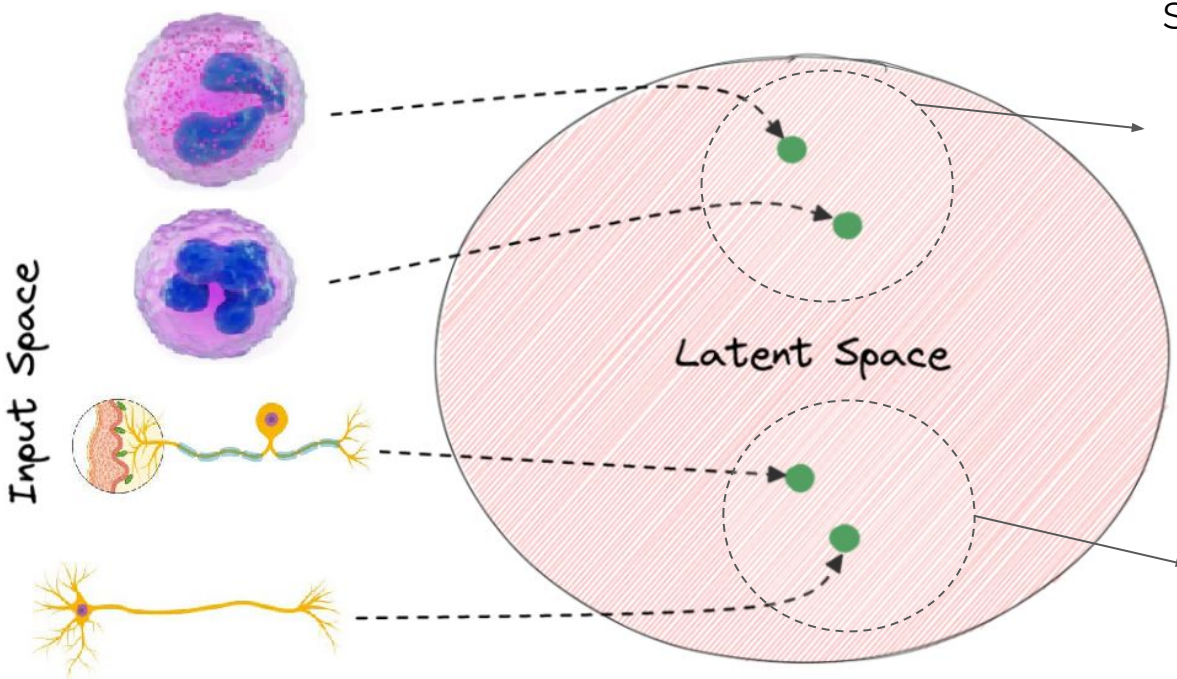




Lots of specific features which are useless for representation

Key features which encapsulate fundamental observations

Understanding Latent spaces



Shared features:

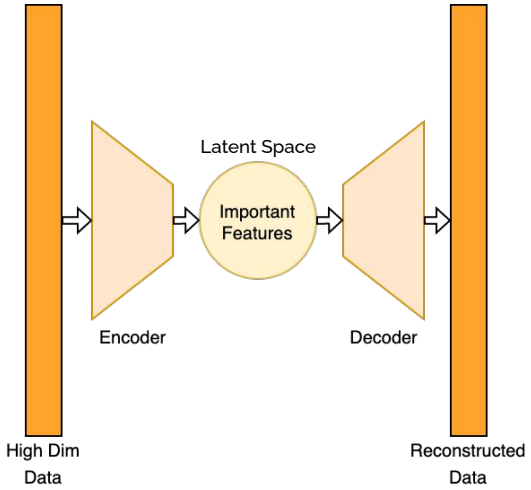
1. Light or pinkish color
2. Relatively round shape
3. Well defined nucleus

Shared features:

1. Long body
2. Many branches

How do we obtain a good latent space for the dataset?

Variational Autoencoders



Probabilistic machine learning method that **encodes** and **decodes** a latent space

Encoder

- learns the "**important features**" that are worth preserving

Decoder

- learns how to "**interpret**" coded information to reconstruct the original information

Variational Autoencoders (VAE)

2 parts: encoder and decoder (neural networks)

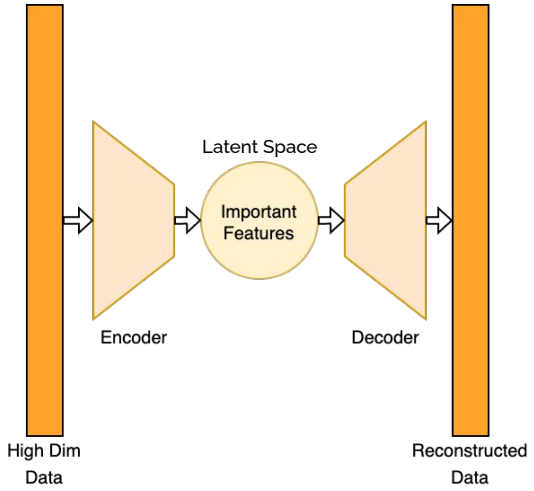
Encoder Network

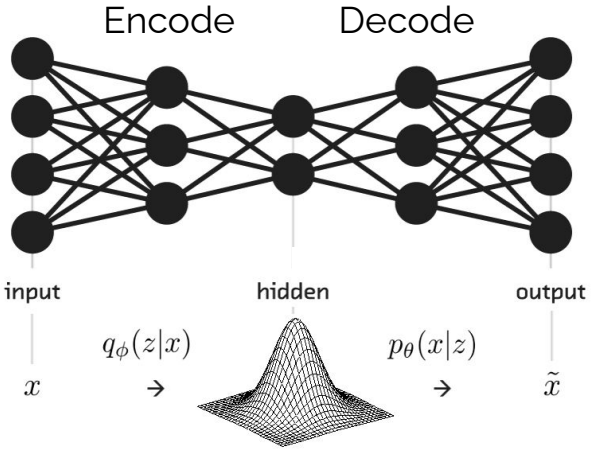
- Transforms high-dim data into low-dim data

Decoder Network

- Reconstruct high-dim data from low dim data

Minimize information loss of reconstructed data

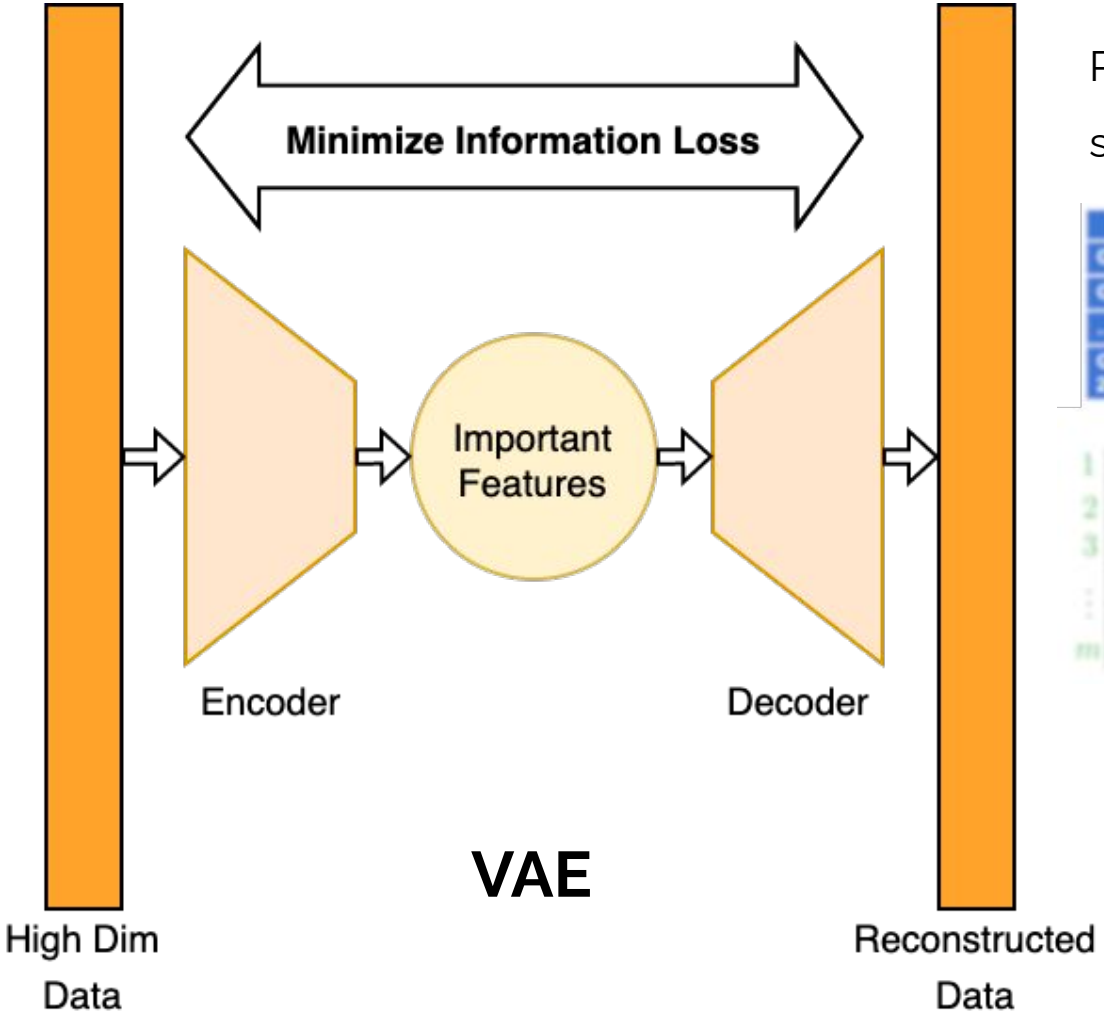




scRNA seq data

	Cell 1	Cell 2	...
Gene 1	18	0	...
Gene 2	1010	506	...
...	...	...	...
Gene 20,000	22	0	...

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & n \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array}$$



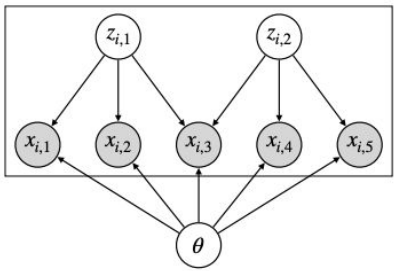
Reconstructed scRNA seq data

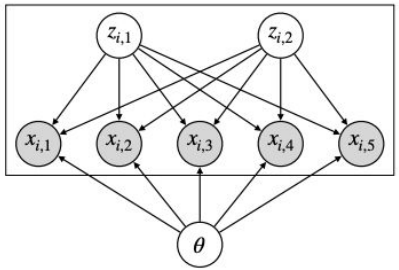
	Cell 1	Cell 2	...
Gene 1	18	0	...
Gene 2	1010	506	...
...	...	...	...
Gene 20,000	22	0	...

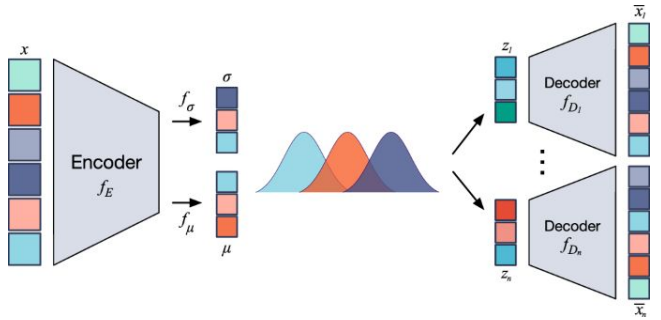
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & n \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array}$$

Benchmarking VAE Methods

- Many VAE algorithms in the field
- We will analyze them with the same dataset and pick the one with **best performance (lowest loss)**
- **Iterative Process**: continuously tune model structure and parameters

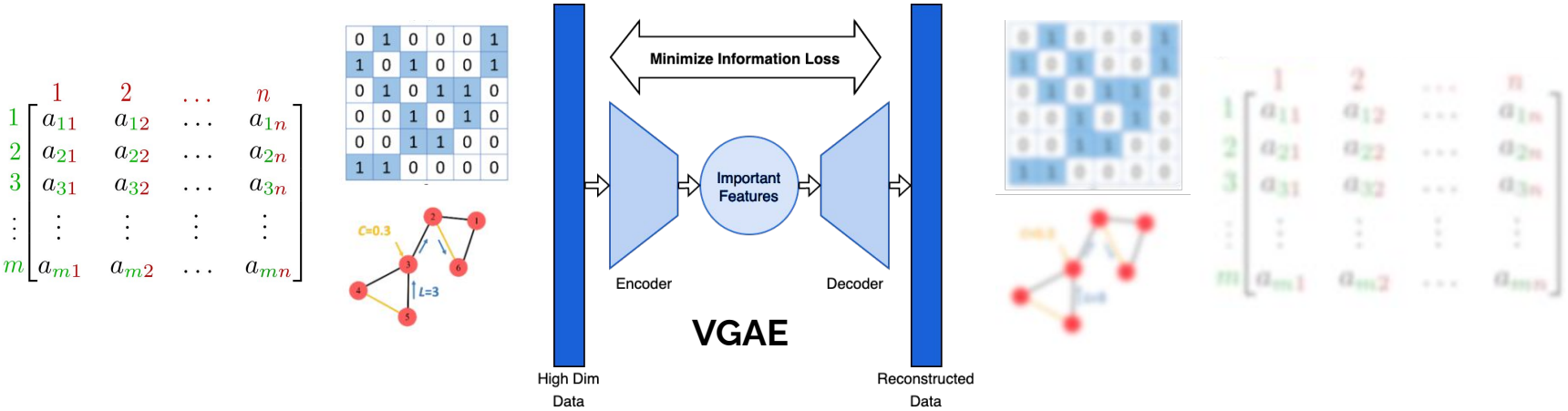


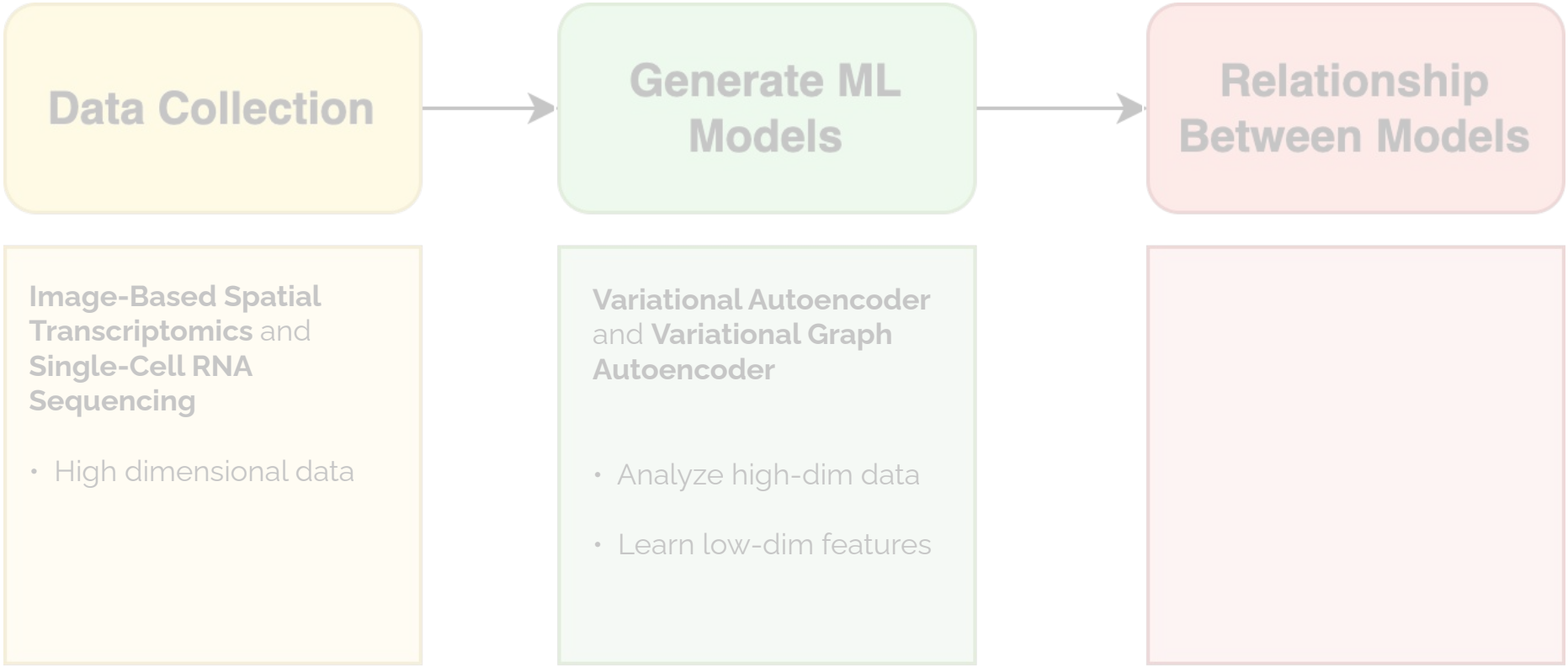




Variational Graph Autoencoder (VGAE)

- Similar to Variational Autoencoder (VAE)
- Additionally allows **spatial information (x, y)** as input
- Encodes **sequencing data** and **adjacency matrix (spatial graph)** together to a latent space

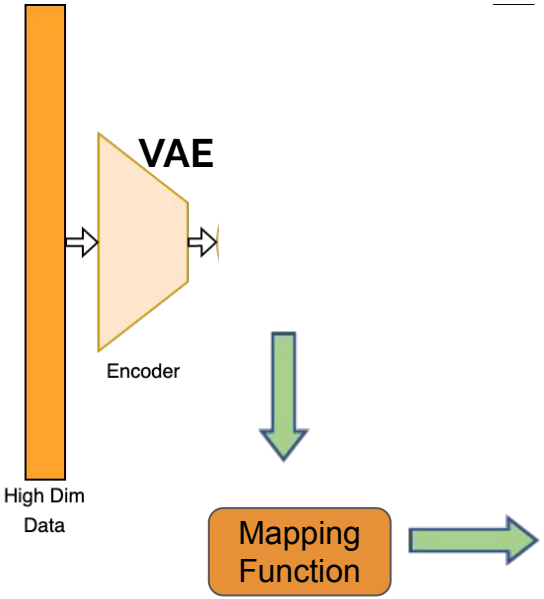



Data Collection
Generate ML Models
Relationship Between Models
Image-Based Spatial Transcriptomics and Single-Cell RNA Sequencing
• High dimensional data
Variational Autoencoder and Variational Graph Autoencoder
• Analyze high-dim data
• Learn low-dim features

Connecting between the Latent Spaces

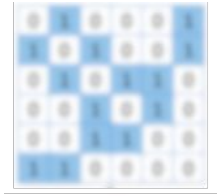
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & n \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array}$$

gene information



gene information and

spatial information



How do we create

this mapping?

Theories for Bridging Two Latent Spaces

- Many theories developed for non-bioinformatics fields
- Potential Methods:
  - Domain Transfer
  - Latent Mapping
  - Latent Translation
- **Evaluate and apply those implementations on our latent spaces**

Data Collection

Image-Based Spatial Transcriptomics and **Single-Cell RNA Sequencing**

- High dimensional data

Generate ML Models

Variational Autoencoder and **Variational Graph Autoencoder**

- Analyze high-dim data
- Learn low-dim features

Relationship Between Models

Mapping Function between two Latent Spaces

- Understand what the features represent
- Minimize reconstruction loss

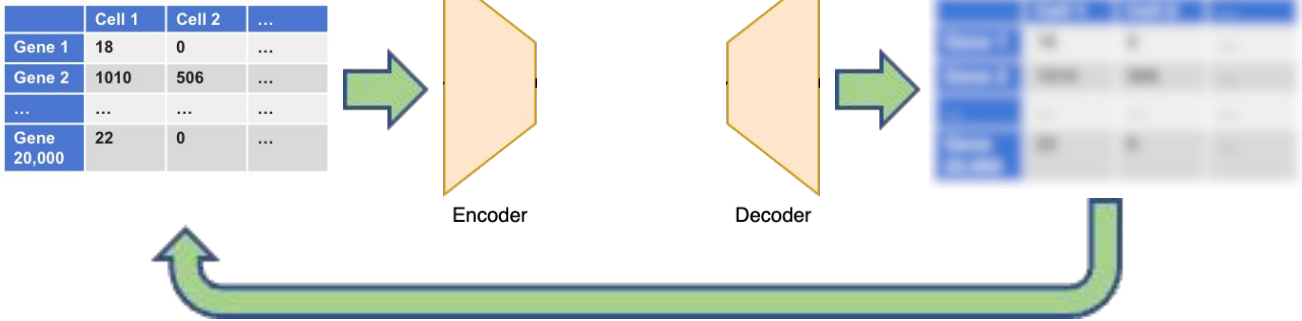
Validation

Important to verify our model on **new, unseen data**.

- **VAE Model**
  - Accuracy of encoding and decoding of scRNA-seq data (reduce information loss)

VAE

NEW scRNA-seq data



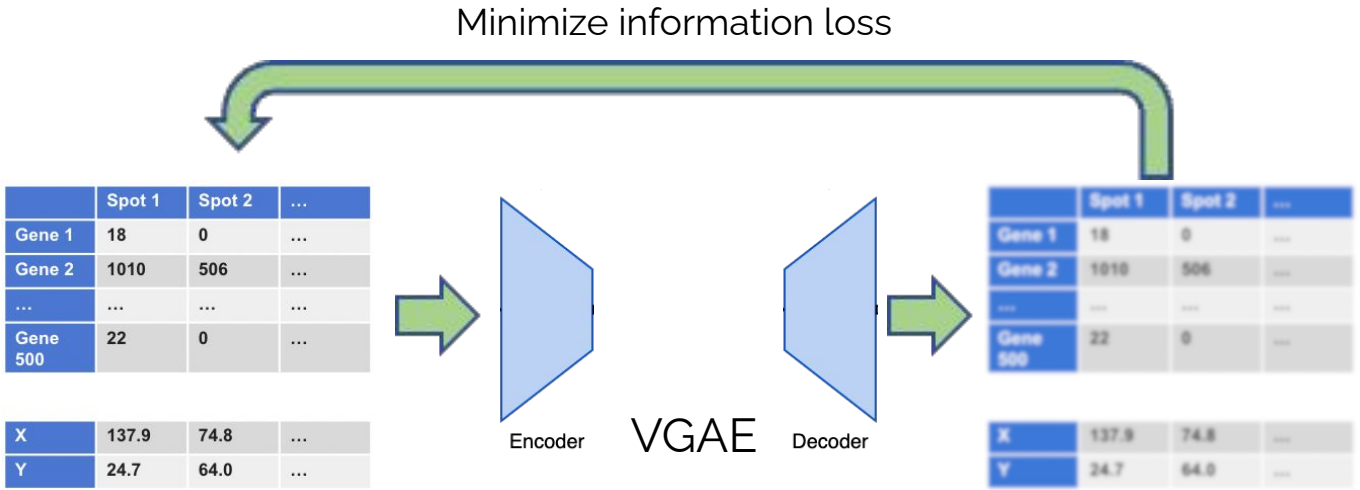
Minimize information loss

Validation

Important to verify our model on **new, unseen data**.

- VAE Model
  - Accuracy of encoding and decoding of scRNA-seq data (reduce information loss)
- **VGAE Model**
  - Accuracy of encoding and decoding of spatial transcriptomics data (reduce information loss)

NEW spatial transcriptomics data



Validation

Important to verify our model on **new, unseen data**.

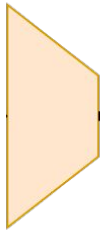
- VAE Model
  - Accuracy of encoding and decoding of scRNA-seq data (reduce information loss)
- GVAE Model
  - Accuracy of encoding and decoding of spatial transcriptomics data (reduce information loss)
- **Joint Model**
  - Accuracy of latent space mapping between scRNA-seq and spatial transcriptomics data (learning associations)

NEW spatial transcriptomics data

	Spot 1	Spot 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 500	22	0	…

X	137.9	74.8	…
Y	24.7	64.0	…

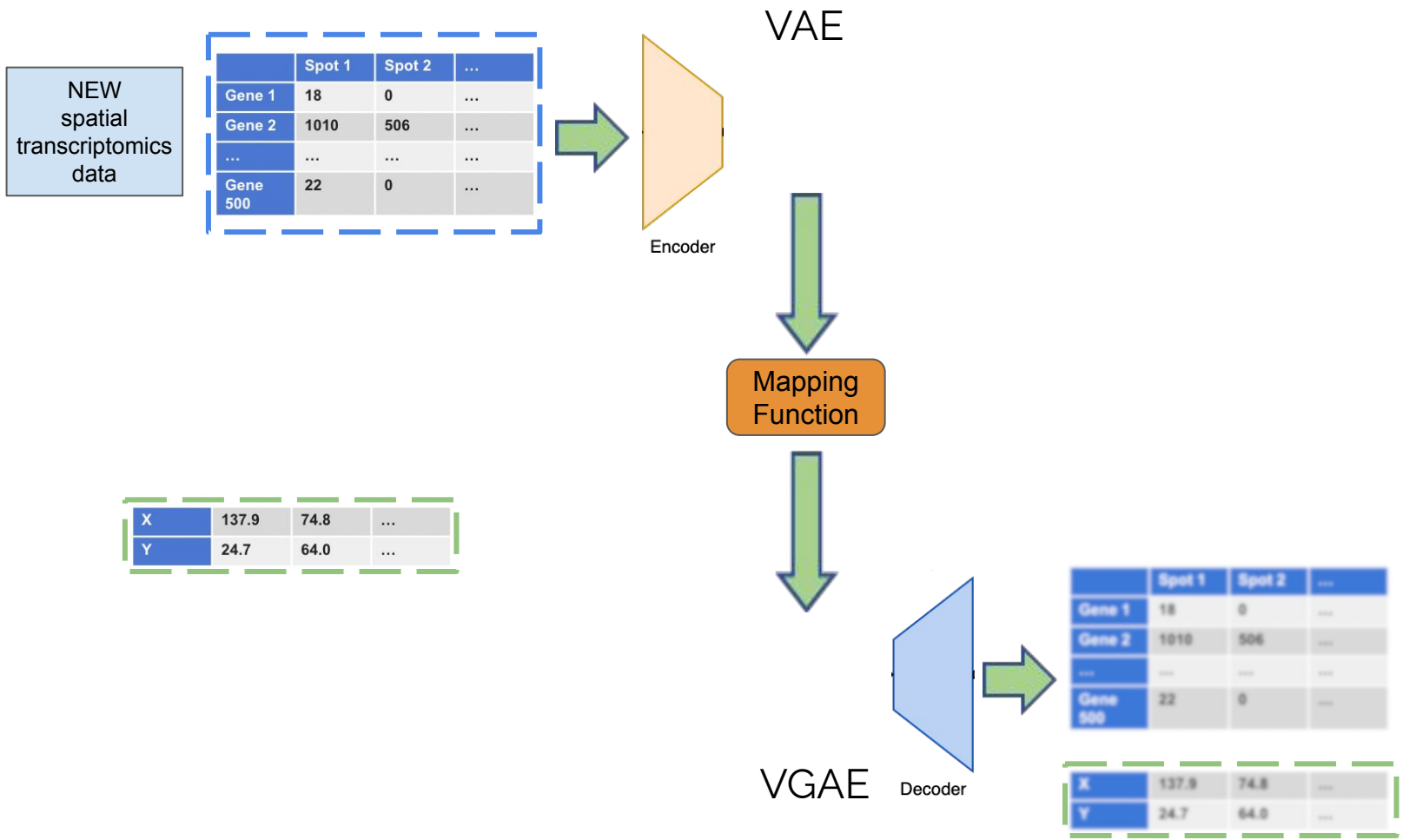
VAE

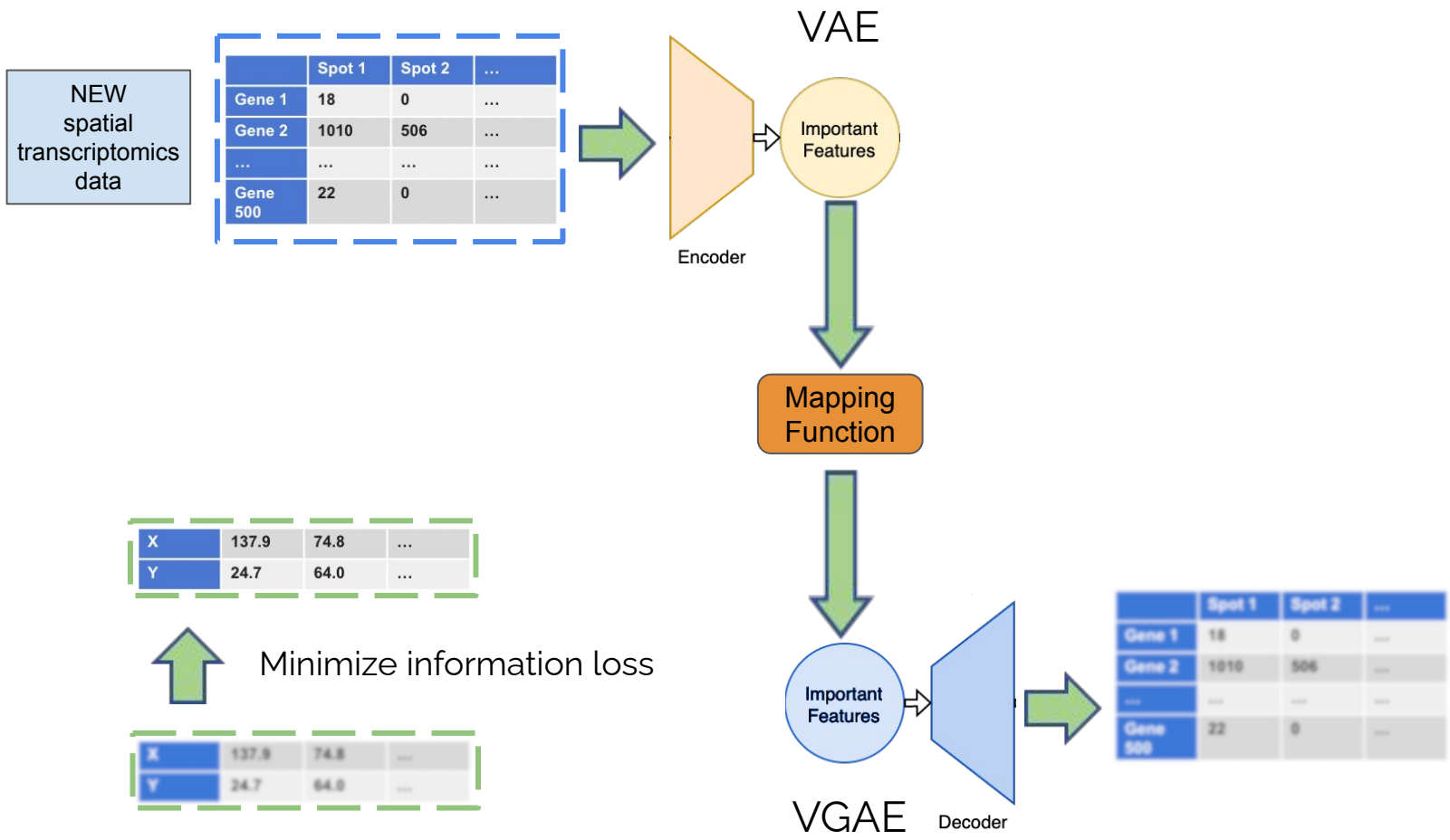


Encoder

VGAE

Decoder

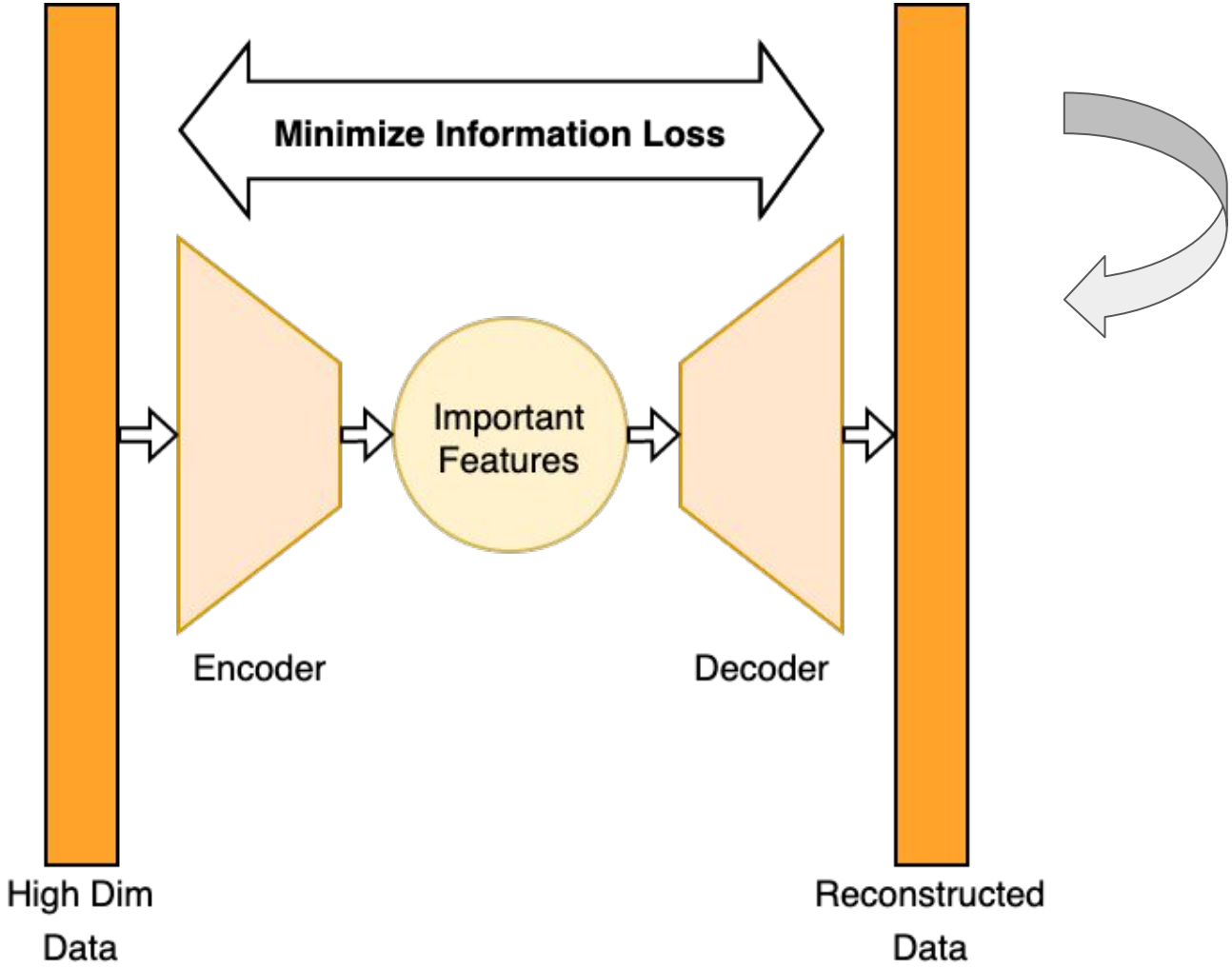

NEW spatial transcriptomics data
Spot 1
Spot 2
...
Gene 1
18
0
Gene 2
1010
506
Gene 500
22
VAE
Encoder
Mapping Function
X
137.9
74.8
Y
24.7
64.0
VGAE
Decoder

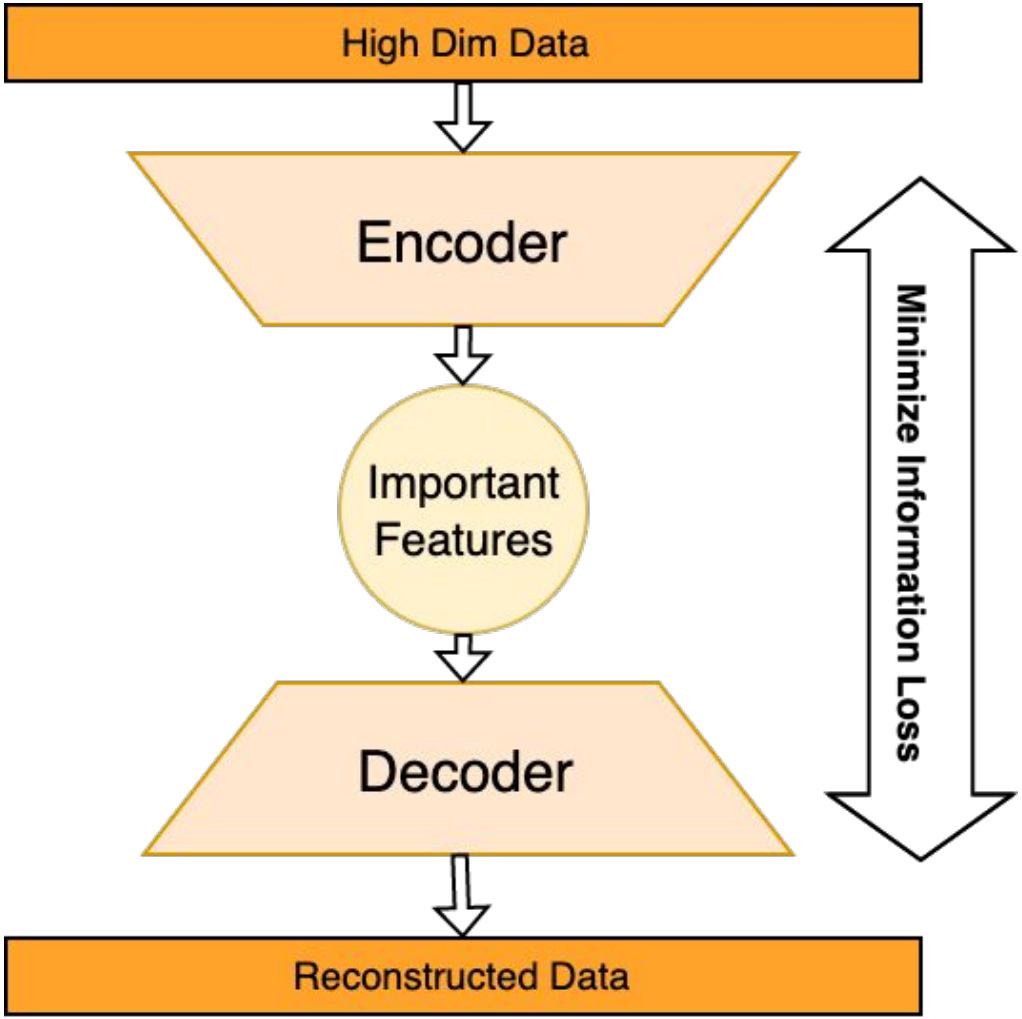

NEW spatial transcriptomics data
Spot 1
Spot 2
...
Gene 1
18
0
Gene 2
1010
506
Gene 500
22
VAE
Encoder
Important Features
Mapping Function
X
137.9
74.8
Y
24.7
64.0
Minimize information loss
Important Features
VGAE
Decoder

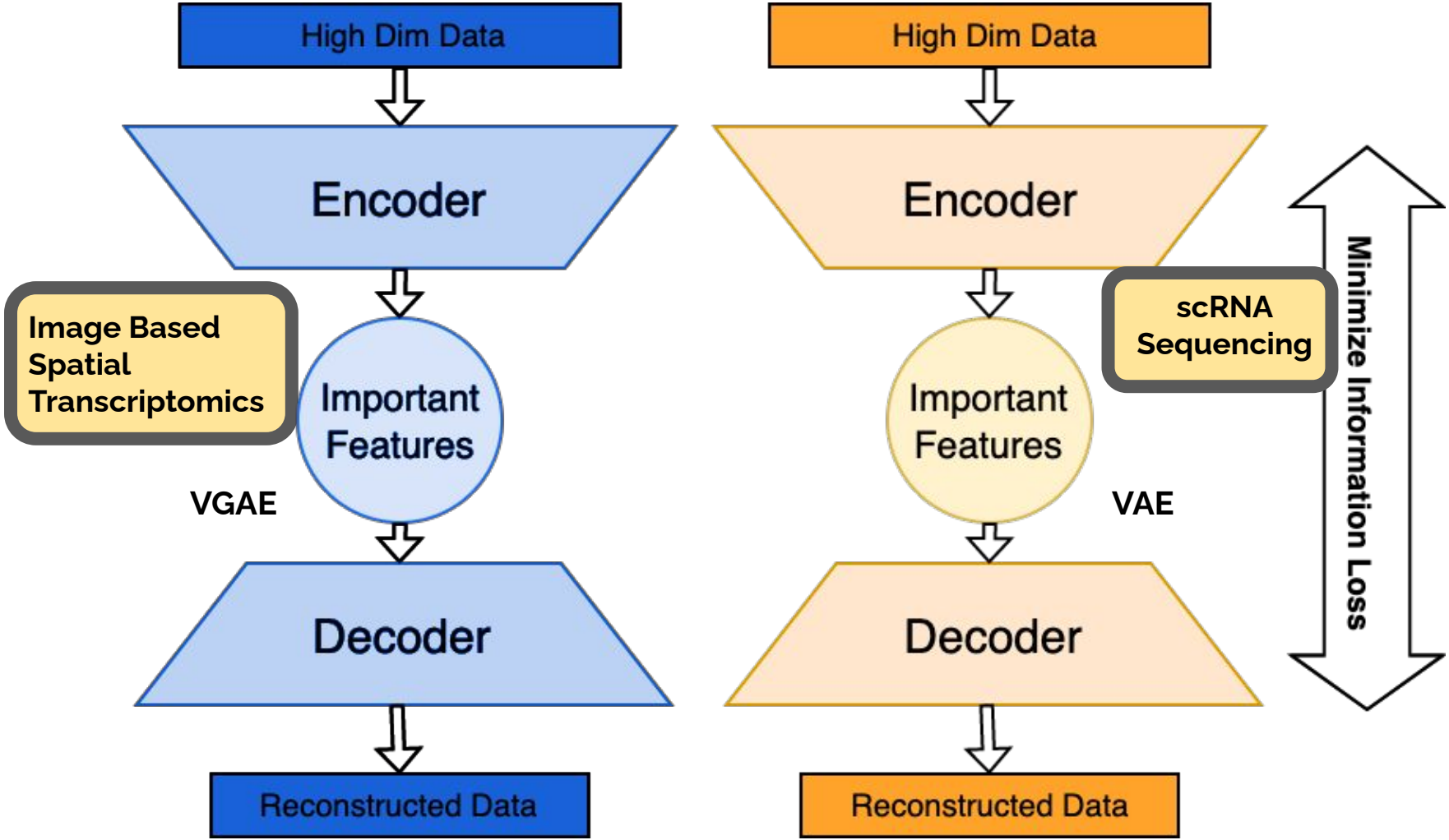
Validation

Important to verify our model on **new, unseen data**.

- VAE Model
  - Accuracy of encoding and decoding of scRNA-seq data (reduce information loss)
- GVAE Model
  - Accuracy of encoding and decoding of spatial transcriptomics data (reduce information loss)
- Joint Model
  - Accuracy of latent space mapping between scRNA-seq and spatial transcriptomics data (learning associations)


Minimize Information Loss
Important Features
Encoder
Decoder
High Dim Data
Reconstructed Data


High Dim Data
Encoder
Important Features
Decoder
Reconstructed Data
Minimize Information Loss


High Dim Data
Encoder
Image Based Spatial Transcriptomics
Important Features
VGAE
Decoder
Reconstructed Data
High Dim Data
Encoder
scRNA Sequencing
Important Features
VAE
Decoder
Reconstructed Data
Minimize Information Loss

	Spot 1	Spot 2	...
Gene 1	18	0	...
Gene 2	1010	506	...
...	...	...	...
Gene 500	22	0	...

X	137.9	74.8	...
Y	24.7	64.0	...

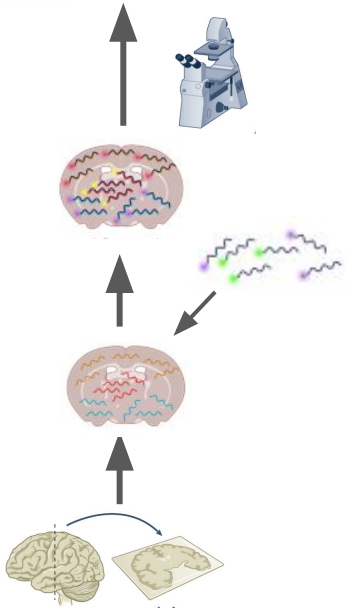
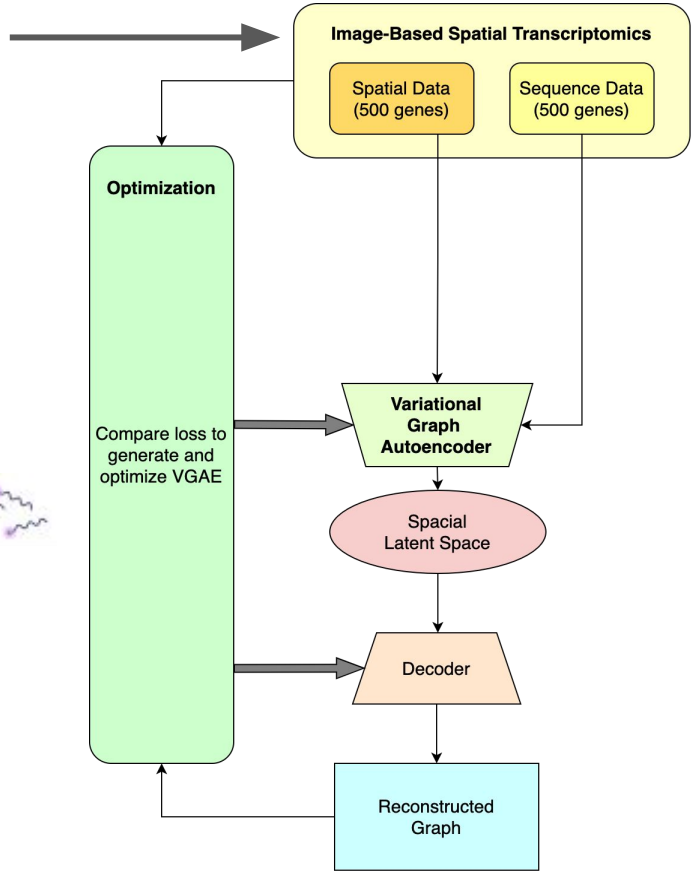
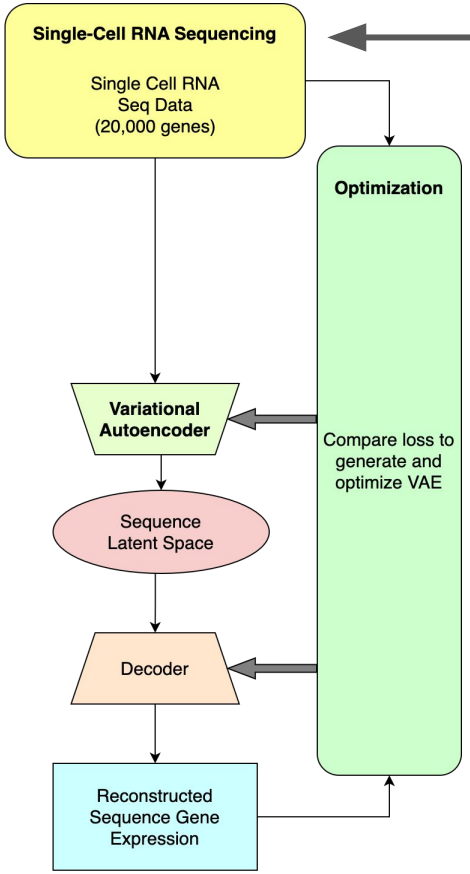


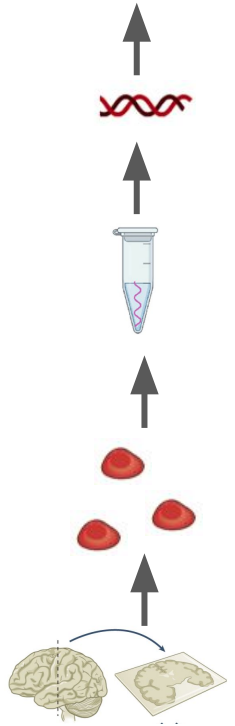
Image Based

Spatial Transcriptomics





	Cell 1	Cell 2	...
Gene 1	18	0	...
Gene 2	1010	506	...
...	...	...	...
Gene 20,000	22	0	...



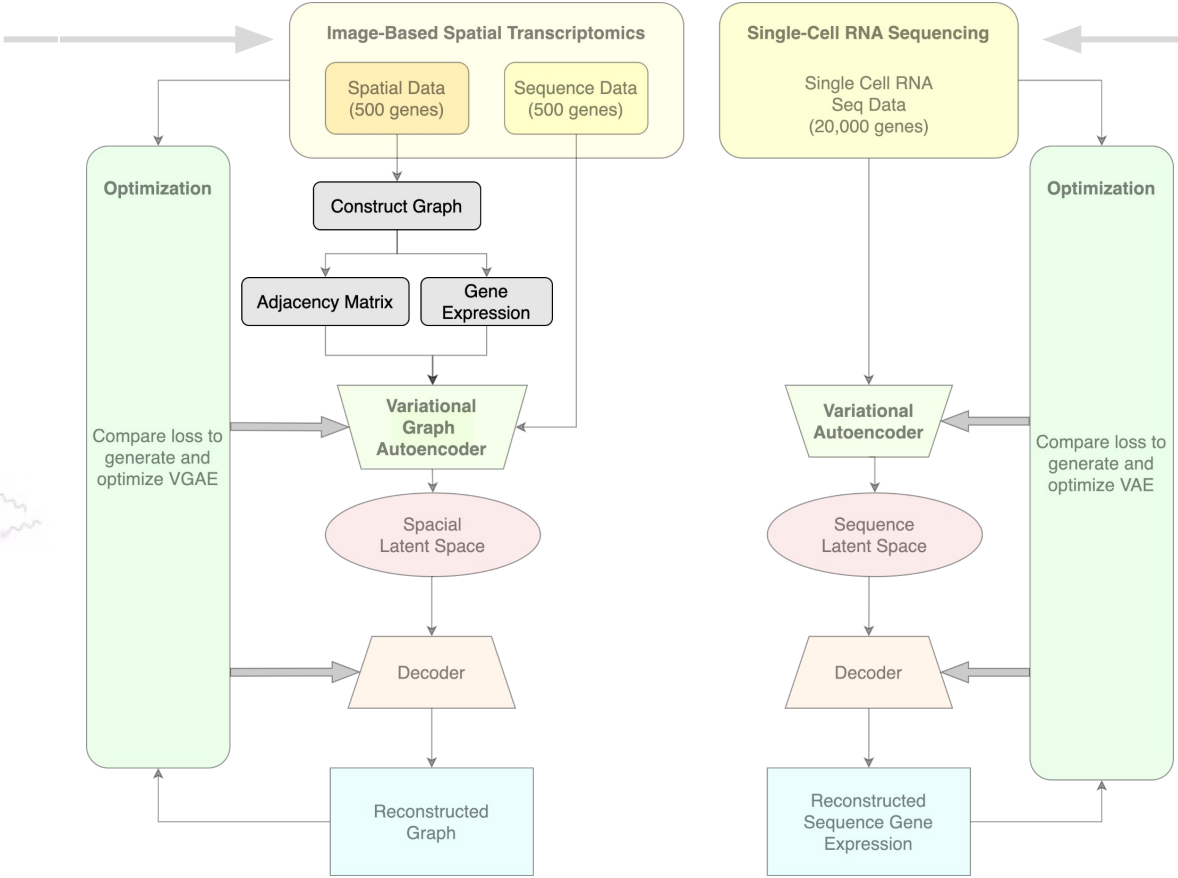
Single Cell

RNA Sequencing

	Spot 1	Spot 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 500	22	0	…

X	137.9	74.8	…
Y	24.7	64.0	…

Image Based
Spatial Transcriptomics



	Cell 1	Cell 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 20,000	22	0	…

Single Cell
RNA Sequencing

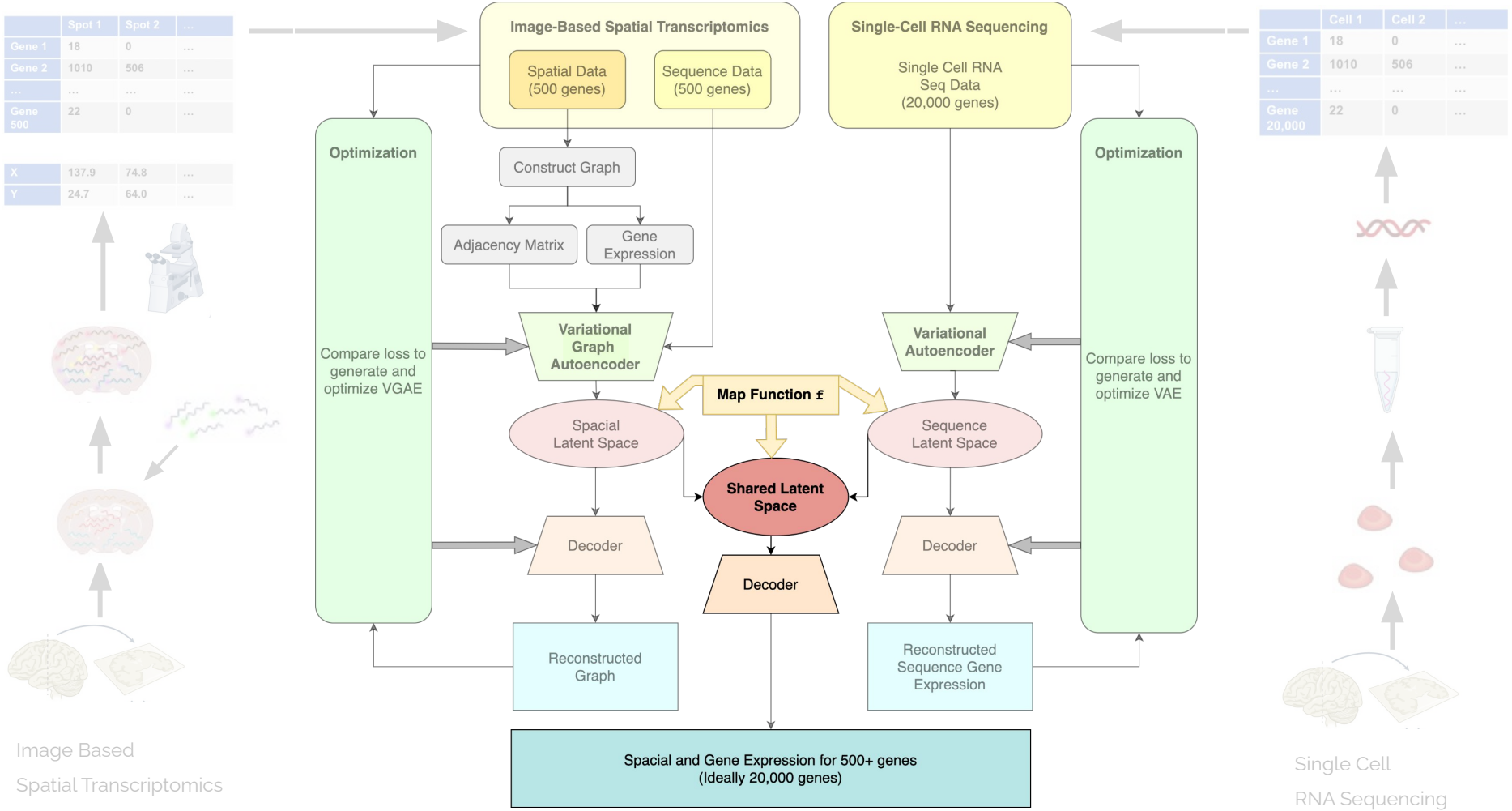


Image-Based Spatial Transcriptomics
Spatial Data (500 genes)
Sequence Data (500 genes)
Single-Cell RNA Sequencing
Single Cell RNA Seq Data (20,000 genes)
Optimization
Construct Graph
Adjacency Matrix
Gene Expression
Variational Graph Autoencoder
Compare loss to generate and optimize VGAE
Map Function £
Spacial Latent Space
Shared Latent Space
Sequence Latent Space
Variational Autoencoder
Compare loss to generate and optimize VAE
Optimization
Decoder
Decoder
Decoder
Reconstructed Graph
Reconstructed Sequence Gene Expression
Spacial and Gene Expression for 500+ genes
(Ideally 20,000 genes)
Image Based
Spatial Transcriptomics
Single Cell
RNA Sequencing

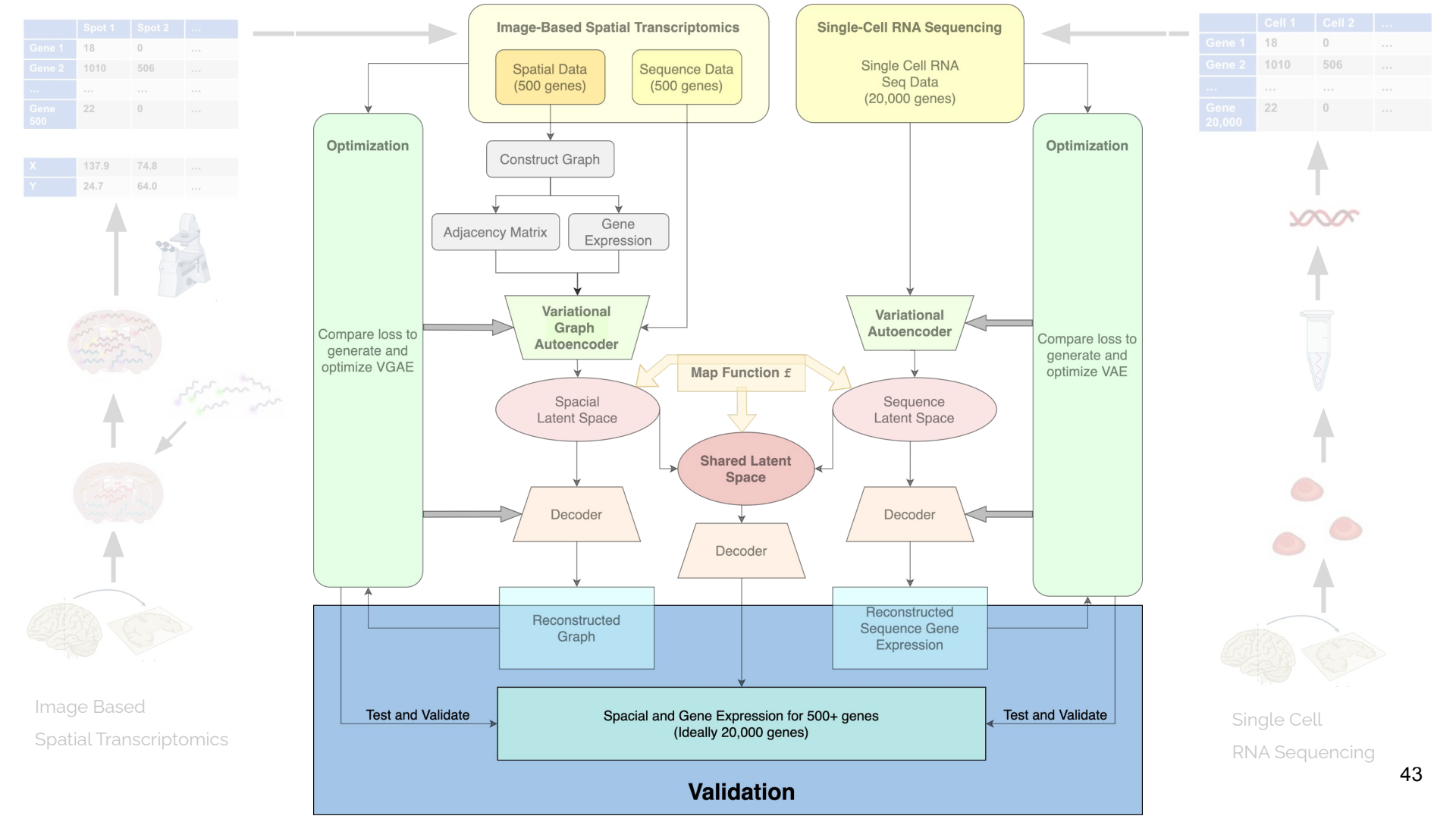


Image-Based Spatial Transcriptomics
Spatial Data (500 genes)
Sequence Data (500 genes)
Single-Cell RNA Sequencing
Single Cell RNA Seq Data (20,000 genes)
Optimization
Construct Graph
Adjacency Matrix
Gene Expression
Variational Graph Autoencoder
Variational Autoencoder
Compare loss to generate and optimize VGAE
Compare loss to generate and optimize VAE
Map Function £
Spacial Latent Space
Sequence Latent Space
Shared Latent Space
Decoder
Decoder
Decoder
Reconstructed Graph
Reconstructed Sequence Gene Expression
Test and Validate
Spacial and Gene Expression for 500+ genes (Ideally 20,000 genes)
Test and Validate
Validation
Image Based Spatial Transcriptomics
Single Cell RNA Sequencing

	Spot 1	Spot 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 500	22	0	…

X	137.9	74.8	…
Y	24.7	64.0	…

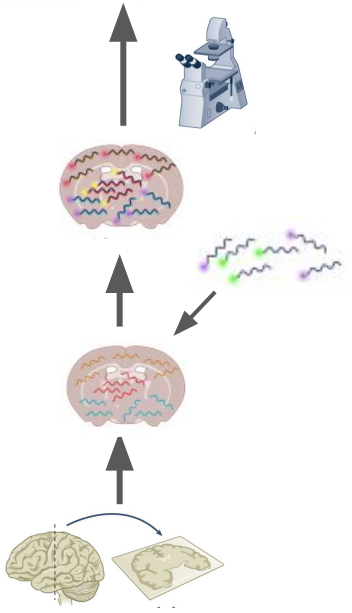


Image Based

Spatial Transcriptomics

Image-Based Spatial Transcriptomics

Spatial Data (500 genes)

Sequence Data (500 genes)

Optimization

Construct Graph

Adjacency Matrix

Gene Expression

Variational Graph Autoencoder

Compare loss to generate and optimize VGAE

Map Function £

Spacial Latent Space

Shared Latent Space

Decoder

Decoder

Reconstructed Graph

Test and Validate

Spacial and Gene Expression for 500+ genes (Ideally 20,000 genes)

Validation

Single-Cell RNA Sequencing

Single Cell RNA Seq Data (20,000 genes)

Optimization

Variational Autoencoder

Compare loss to generate and optimize VAE

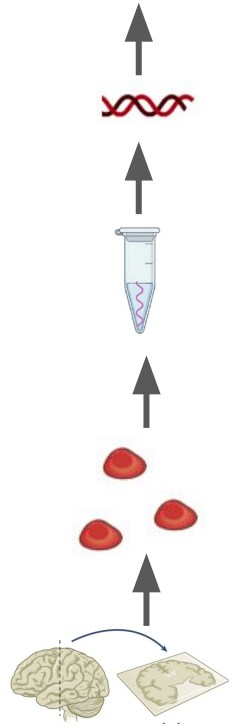
Sequence Latent Space

Decoder

Reconstructed Sequence Gene Expression

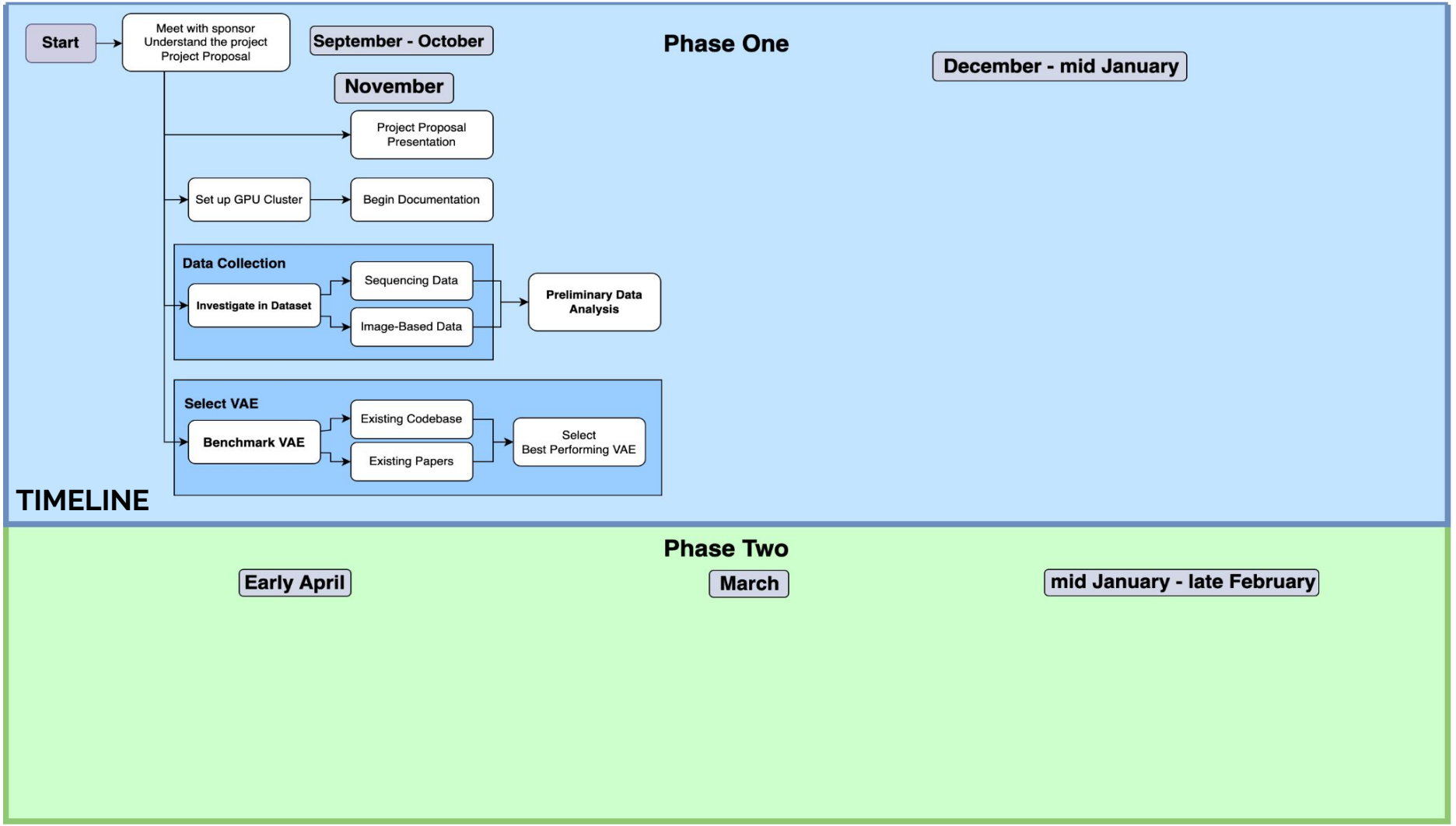
Test and Validate

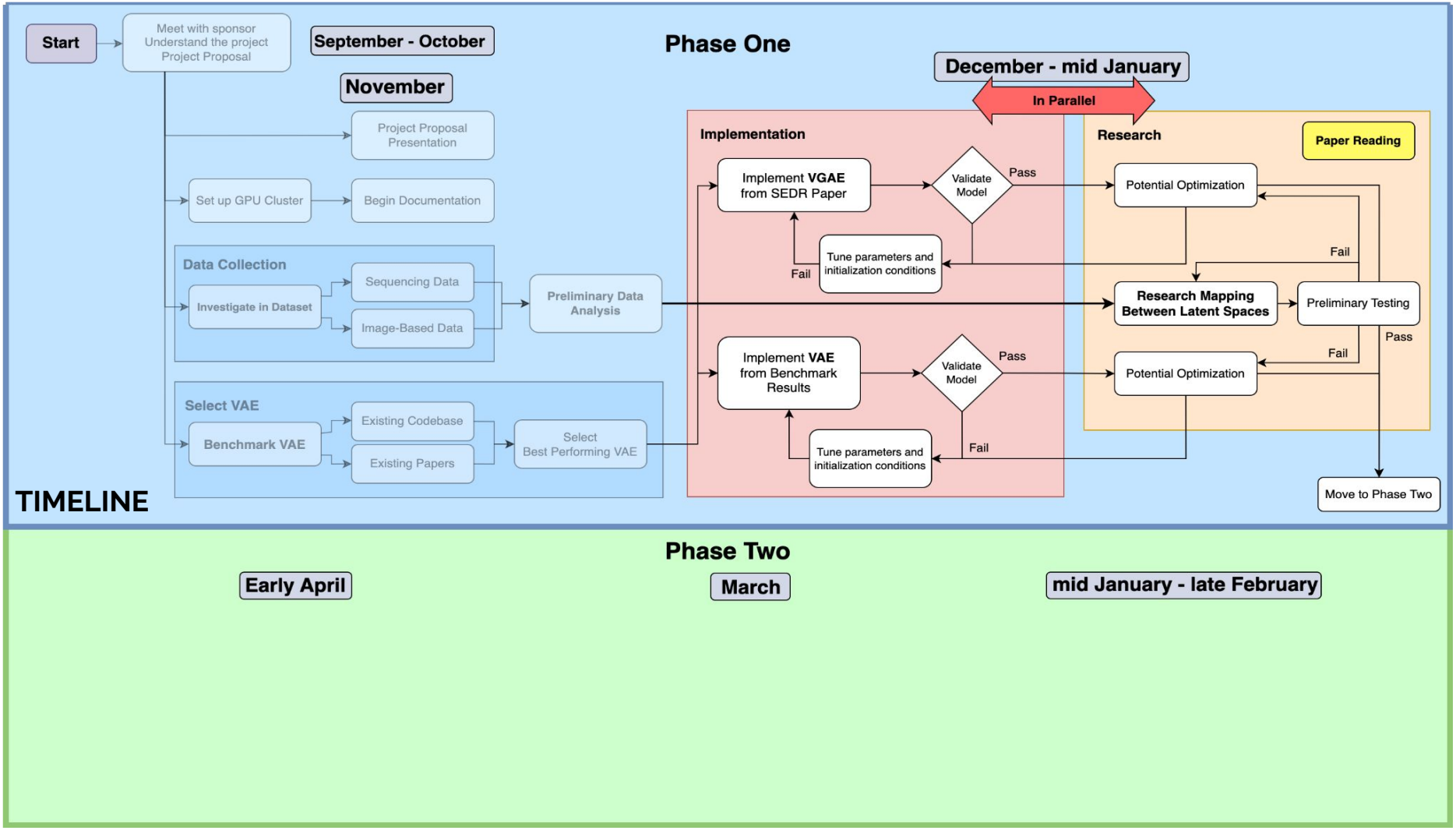
	Cell 1	Cell 2	…
Gene 1	18	0	…
Gene 2	1010	506	…
…	…	…	…
Gene 20,000	22	0	…

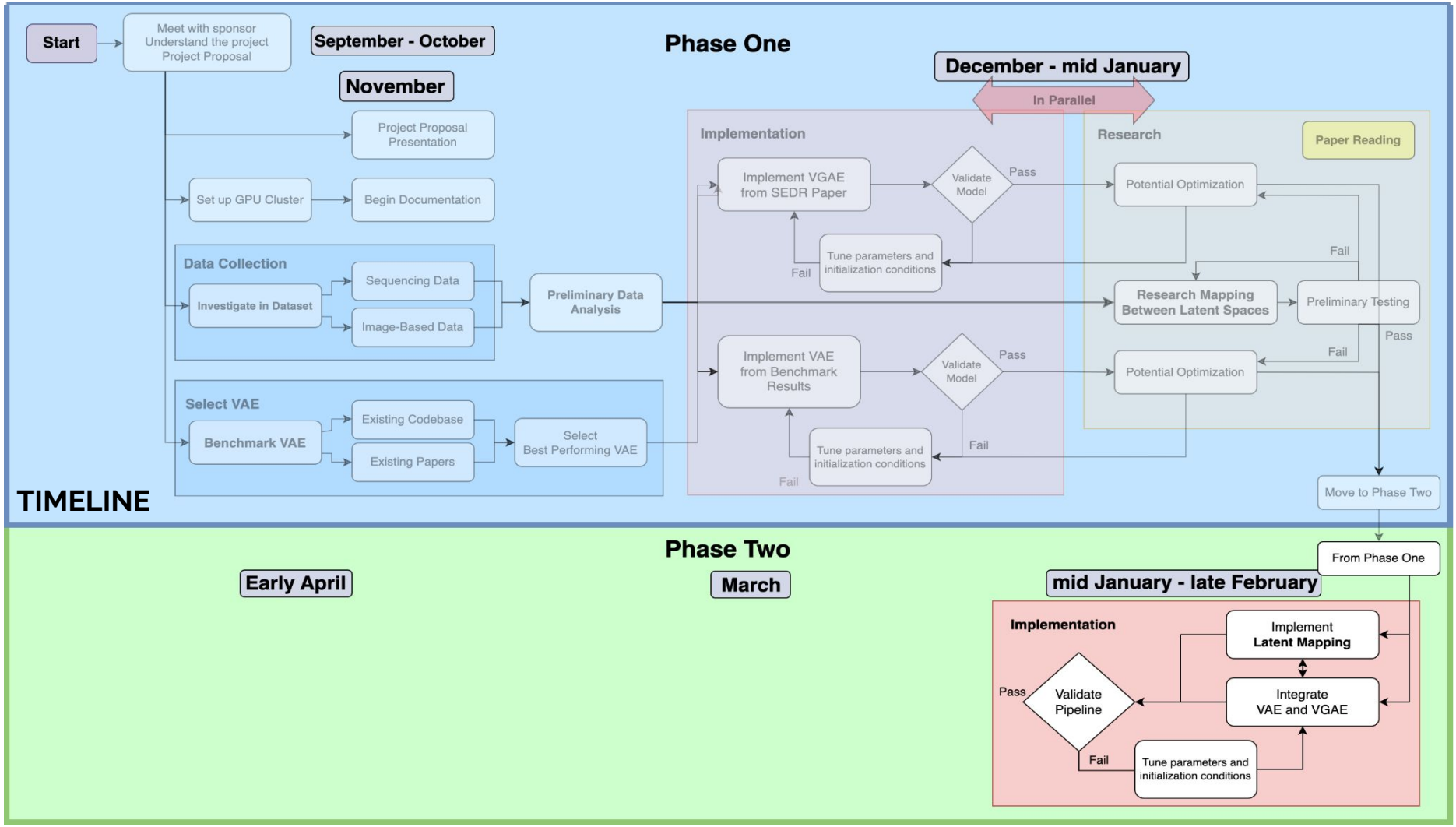


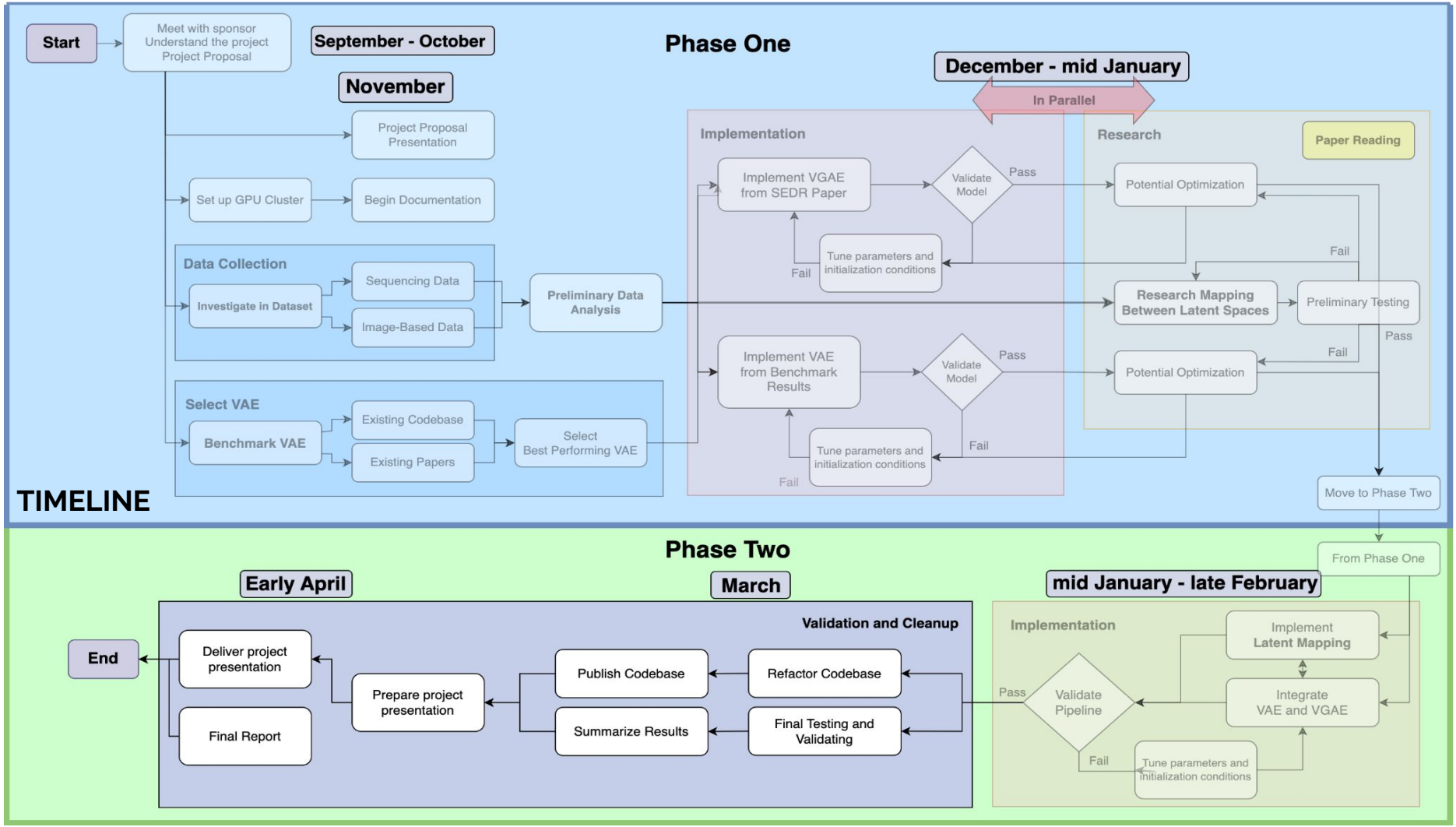
Single Cell

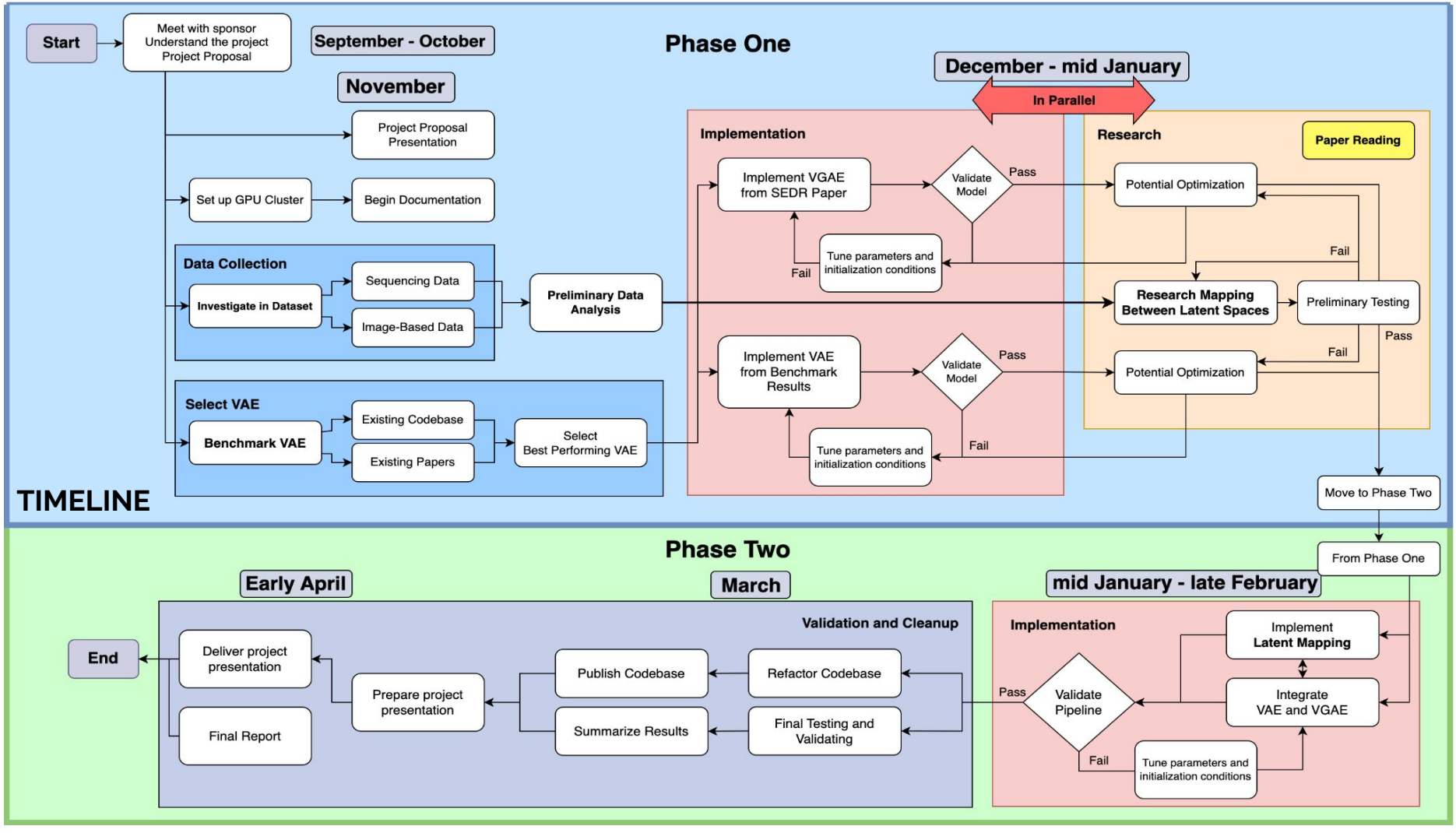
RNA Sequencing


Start
Meet with sponsor
Understand the project
Project Proposal
September - October
Phase One
December - mid January
November
Project Proposal
Presentation
Set up GPU Cluster
Begin Documentation
Data Collection
Investigate in Dataset
Sequencing Data
Image-Based Data
Preliminary Data
Analysis
Select VAE
Benchmark VAE
Existing Codebase
Existing Papers
Select
Best Performing VAE
TIMELINE
Phase Two
Early April
March
mid January - late February


Start
Meet with sponsor
Understand the project
Project Proposal
September - October
Phase One
November
Project Proposal Presentation
Set up GPU Cluster
Begin Documentation
Data Collection
Investigate in Dataset
Sequencing Data
Image-Based Data
Preliminary Data Analysis
Select VAE
Benchmark VAE
Existing Codebase
Existing Papers
Select Best Performing VAE
December - mid January
In Parallel
Implementation
Implement VGAE from SEDR Paper
Validate Model
Pass
Fail
Tune parameters and initialization conditions
Implement VAE from Benchmark Results
Validate Model
Pass
Fail
Tune parameters and initialization conditions
Research
Paper Reading
Potential Optimization
Fail
Research Mapping Between Latent Spaces
Preliminary Testing
Pass
Fail
Potential Optimization
Move to Phase Two
TIMELINE
Phase Two
Early April
March
mid January - late February


Start
Meet with sponsor
Understand the project
Project Proposal
September - October
Phase One
November
Project Proposal Presentation
Set up GPU Cluster
Begin Documentation
Data Collection
Investigate in Dataset
Sequencing Data
Image-Based Data
Preliminary Data Analysis
Select VAE
Benchmark VAE
Existing Codebase
Existing Papers
Select Best Performing VAE
December - mid January
In Parallel
Implementation
Implement VGAE from SEDR Paper
Validate Model
Pass
Fail
Tune parameters and initialization conditions
Implement VAE from Benchmark Results
Validate Model
Pass
Fail
Tune parameters and initialization conditions
Research
Paper Reading
Potential Optimization
Fail
Research Mapping Between Latent Spaces
Preliminary Testing
Pass
Fail
Potential Optimization
Move to Phase Two
TIMELINE
Phase Two
Early April
March
From Phase One
mid January - late February
Implementation
Implement Latent Mapping
Integrate VAE and VGAE
Validate Pipeline
Pass
Fail
Tune parameters and initialization conditions

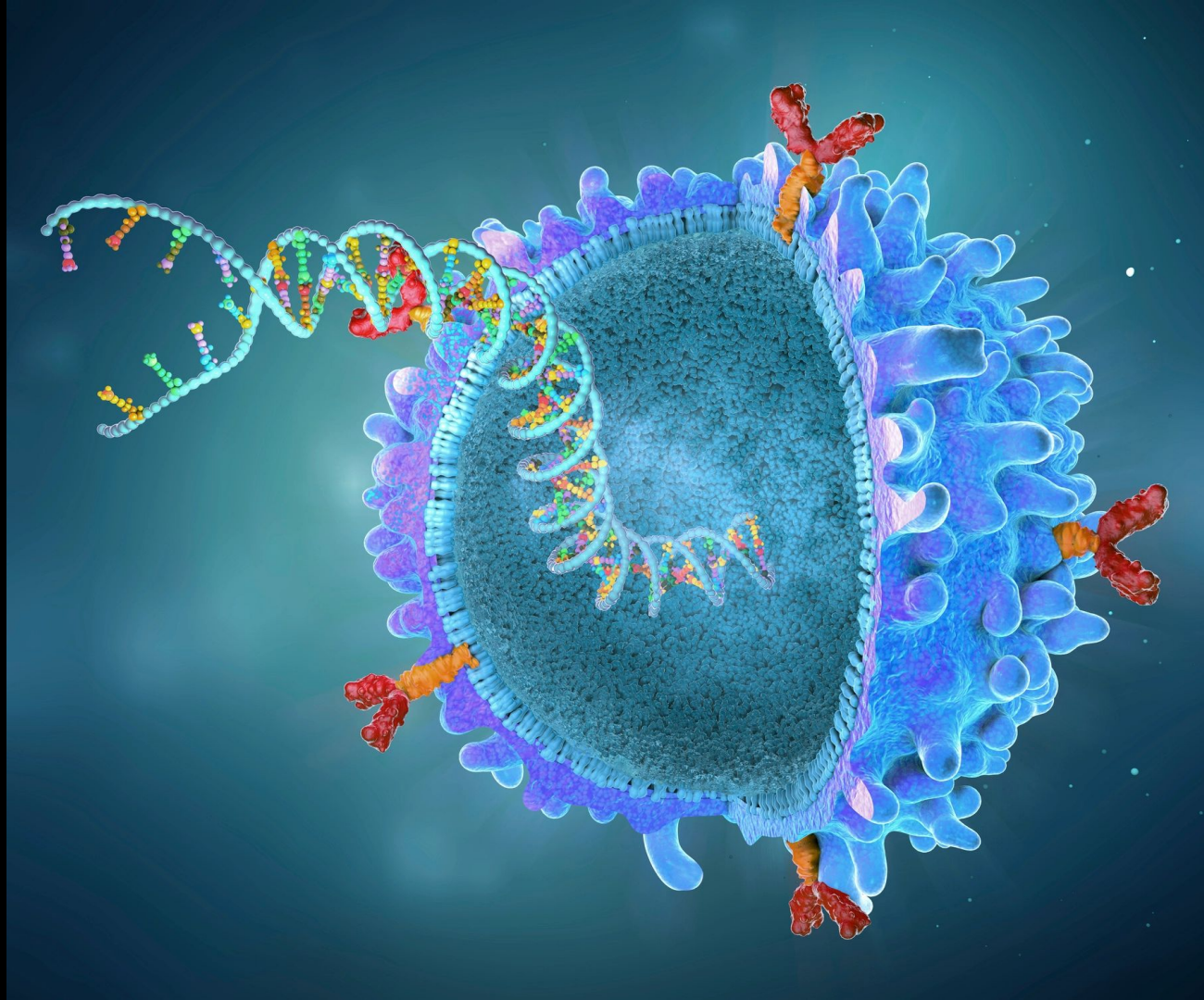

TIMELINE
Phase One
Start
Meet with sponsor
Understand the project
Project Proposal
September - October
November
Project Proposal Presentation
Set up GPU Cluster
Begin Documentation
Data Collection
Investigate in Dataset
Sequencing Data
Image-Based Data
Preliminary Data Analysis
Select VAE
Benchmark VAE
Existing Codebase
Existing Papers
Select Best Performing VAE
December - mid January
In Parallel
Implementation
Implement VGAE from SEDR Paper
Validate Model
Pass
Fail
Tune parameters and initialization conditions
Implement VAE from Benchmark Results
Validate Model
Pass
Fail
Tune parameters and initialization conditions
Research
Paper Reading
Potential Optimization
Fail
Research Mapping Between Latent Spaces
Preliminary Testing
Pass
Fail
Potential Optimization
Move to Phase Two
Phase Two
From Phase One
Early April
March
Validation and Cleanup
End
Deliver project presentation
Final Report
Prepare project presentation
Publish Codebase
Summarize Results
Refactor Codebase
Final Testing and Validating
mid January - late February
Implementation
Implement Latent Mapping
Integrate VAE and VGAE
Validate Pipeline
Pass
Fail
Tune parameters and initialization conditions


TIMELINE
Start
Meet with sponsor
Understand the project
Project Proposal
September - October
Phase One
November
Project Proposal Presentation
Set up GPU Cluster
Begin Documentation
Data Collection
Investigate in Dataset
Sequencing Data
Image-Based Data
Preliminary Data Analysis
Select VAE
Benchmark VAE
Existing Codebase
Existing Papers
Select Best Performing VAE
December - mid January
In Parallel
Implementation
Implement VGAE from SEDR Paper
Validate Model
Pass
Fail
Tune parameters and initialization conditions
Implement VAE from Benchmark Results
Validate Model
Pass
Fail
Tune parameters and initialization conditions
Research
Paper Reading
Potential Optimization
Fail
Research Mapping Between Latent Spaces
Preliminary Testing
Pass
Fail
Potential Optimization
Move to Phase Two
Phase Two
From Phase One
Early April
March
mid January - late February
Validation and Cleanup
End
Deliver project presentation
Final Report
Prepare project presentation
Publish Codebase
Summarize Results
Refactor Codebase
Final Testing and Validating
Implementation
Pass
Validate Pipeline
Fail
Tune parameters and initialization conditions
Implement Latent Mapping
Integrate VAE and VGAE

Deliverables

1. **Fully functional VAE and VGAE pipeline with a map function** between two latent spaces that passes our validation methods, including
    - Successfully reconstructing missing genes for image-based spatial transcriptomics
    - Achieve acceptable accuracy for predicting on validation dataset
2. **Publishing codebase** up to open source standard

Understanding cellular composition provide powerful insights into treatment of disease and cancer

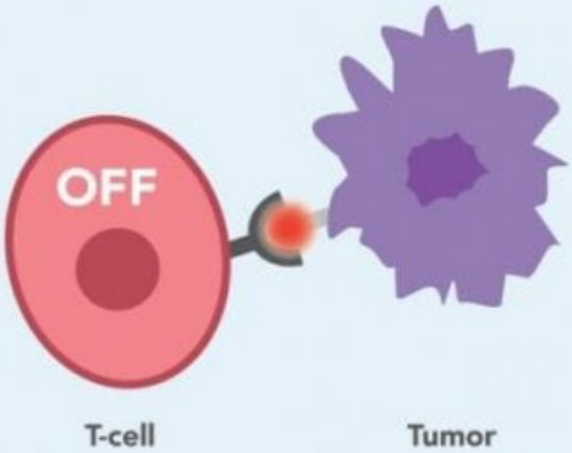


Thank you!

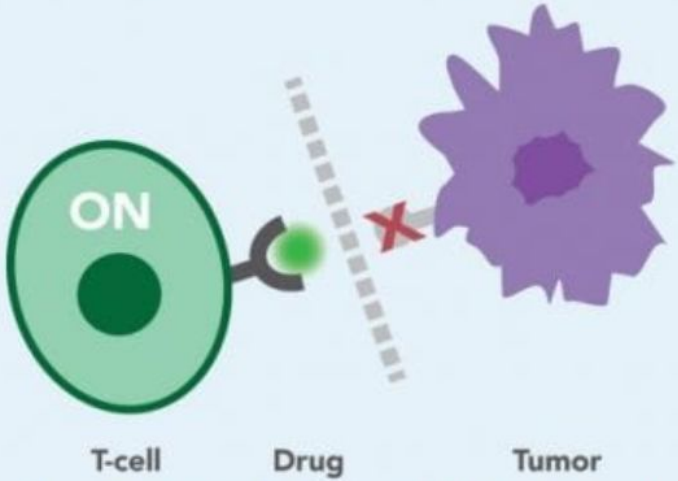
Appendix

How Does Immunotherapy Work?

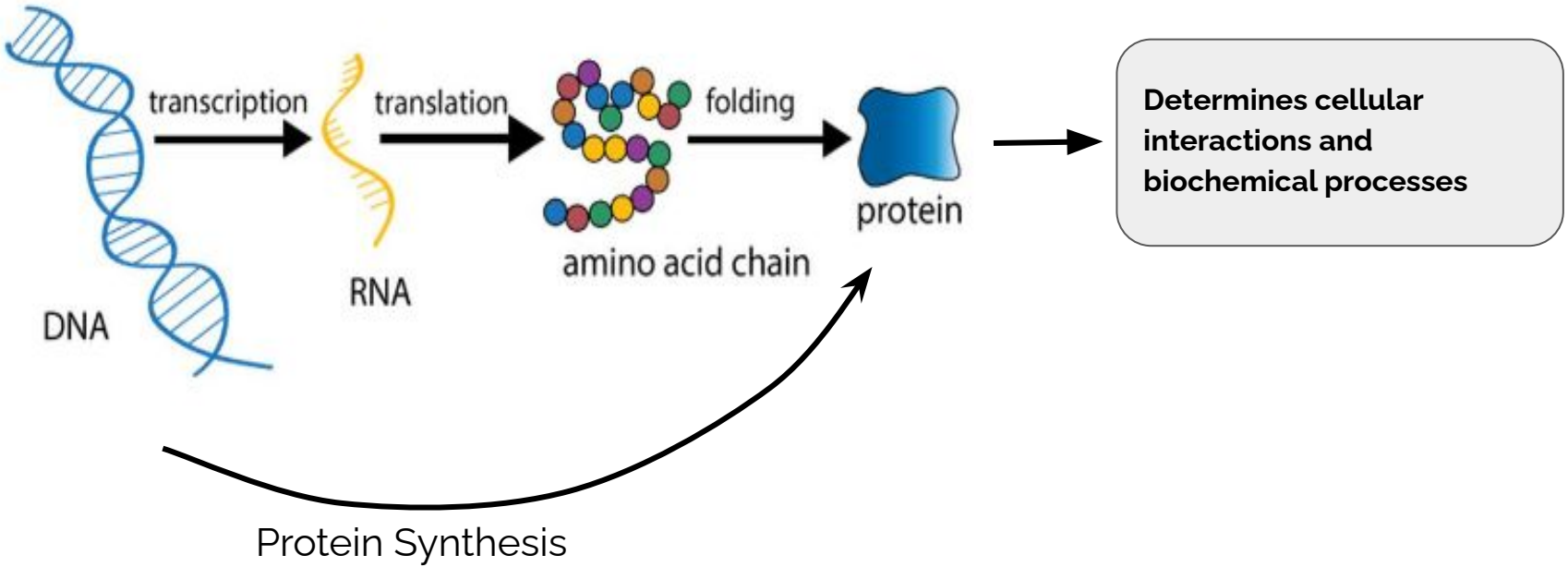
Tumor cells bind to T-cells to deactivate them



Immunotherapy drugs can block tumor cells from deactivating T-cells



DNA and proteins

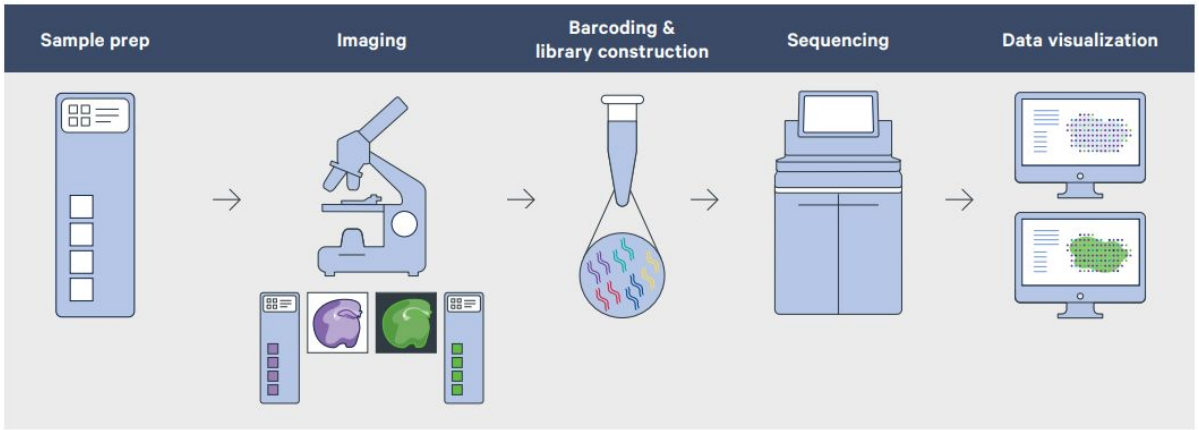


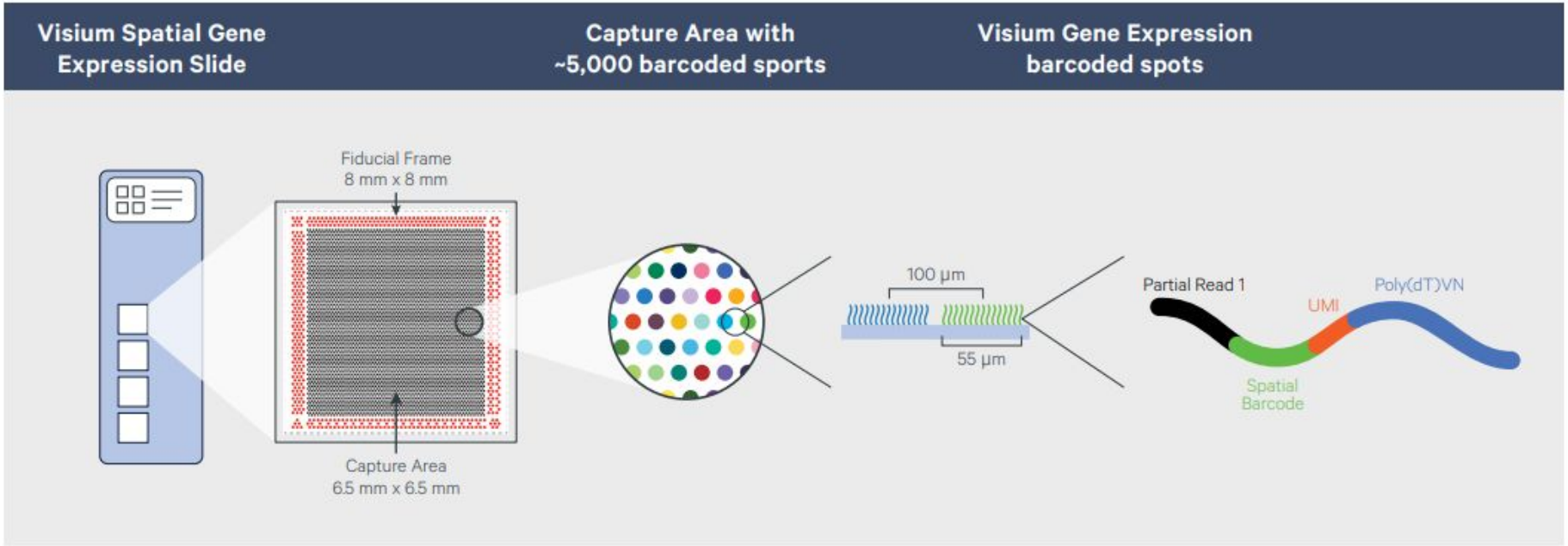
Determines cellular interactions and biochemical processes

Producing spatial transcriptomics data

Another way we can study the RNA molecules inside the cells is by spatial transcriptomics.

In addition to producing genetic data, just like sequential RNA data generation method, spatial transcriptomics also provides insight into the spatial information of cells (x and y coordinates of cells).




Visium Spatial Gene Expression Slide
Capture Area with ~5,000 barcoded sports
Visium Gene Expression barcoded spots
Fiducial Frame
8 mm x 8 mm
Capture Area
6.5 mm x 6.5 mm
100 µm
55 µm
Partial Read 1
UMI
Poly(dT)VN
Spatial Barcode

Variational Graph Auto-Encoders

Thomas N. Kipf
University of Amsterdam
T.N.Kipf@uva.nl

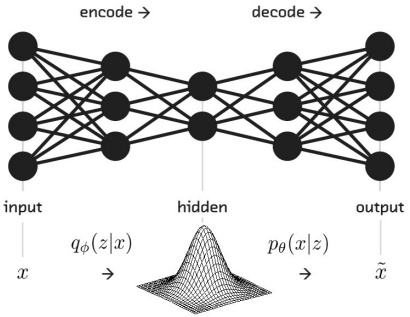
Max Welling
University of Amsterdam
Canadian Institute for Advanced Research (CIFAR)
M.Welling@uva.nl

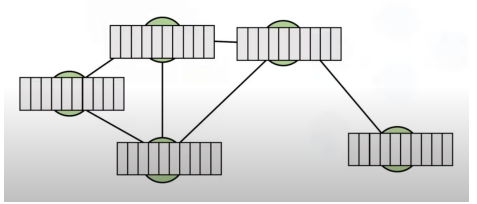
SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS

Thomas N. Kipf
University of Amsterdam
T.N.Kipf@uva.nl

Max Welling
University of Amsterdam
Canadian Institute for Advanced Research (CIFAR)
M.Welling@uva.nl

- Variational autoencoders
  - Encodes high dimensional data into lower-dimensional latent space.
  - Reconstruct original data by decoding latent space back into full-dimensional data
- Graph CNNs
  - Nodes contain vector data
  - Graph to contain relationship of each node to its neighbors
- Graph variational autoencoder can encode the data of an adjacency matrix of cellular data



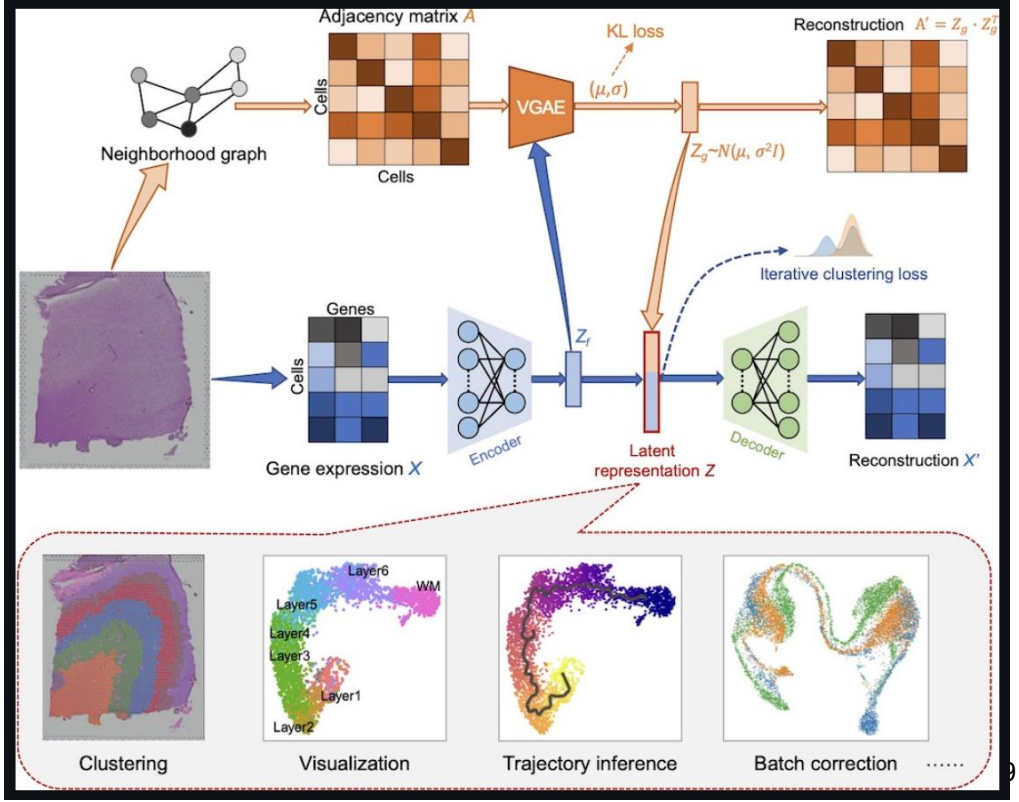


SEDR: paper + code

First step of our project

- Variational graph autoencoder to reconstruct adjacency matrix for spatial transcriptomics data
- Open source github repo with results on various spatial transcriptomics datasets

SEDR (spatial embedded deep representation) learns a low-dimensional latent representation of gene expression embedded with spatial information for spatial transcriptomics analysis. SEDR method consists of two main components, a deep autoencoder network for learning a gene representation, and a variational graph autoencoder network for embedding the spatial information. SEDR has been applied on the 10x Genomics Visium spatial transcriptomics dataset as well as Stereo-seq dataset, and demonstrated its ability to achieve better representation for various follow-up analysis tasks including clustering, visualization, trajectory inference and batch effect correction.



Ann-data

- Both sequence and iBST data are stored in this format.
- Helps us understand what is stored with the data and how to parse it