

## ti\_evaluation

### Statement

Dynamic processes can be modelled and studied with TI methods

### Experiment

- \* Characterise and organise TI methodology @characterisation
- \* Thoroughly set up experiment design to ensure correct scientific reasoning @experiment\_design
- \* Train TI methods on synthetic data @synthetic\_testing
- \* Evaluate them on biological data @biological\_validation
- \* Provide reader with a set of instructions on which TI method to use on which kinds of datasets @guidelines

### Expected results

Some TI methods will perform well on some datasets. Performance will depend on parameters.

### Motivation

An extensive TI evaluation will:

- \* Stimulate the adoption of TI methods
- \* Stimulate the development of more robust and accurate TI methods
- \* Stimulate the evolution of the TI to tackle more complex tasks

## characterise\_platforms

### Statement

We collected properties of possible datasets and determined which TI task could be used on the dataset

### Experiment

TODO

### Expected results

### Motivation

## characterise\_methods

### Statement

We characterised TI methods in terms of its inputs, outputs, methodology and type of TI task it tries to solve

### Experiment

TODO

### Expected results

### Motivation

## characterisation

### Statement

We characterised and organised TI methodology in terms of which biological question is being asked, how a TI method tries to solve this question, and what characteristics it has

### Experiment

We need to characterise

- \* the tasks / biological questions @characterise\_tasks
- \* the platforms @characterise\_platforms
- \* the methods @characterise\_methods

### Expected results

### Motivation

There is a lot of confusion about what TI is, what problems can be solved with it, and what sort of solution you might expect from a TI method. The characterisation will provide a structure to link everything together and clarify ambiguities

## experiment\_design

### Statement

Our evaluation of TI methods is scientifically sound

### Experiment

Construct and evaluate the:

- \* synthetic gold standard
- \* biological gold standard
- \* benchmark pipeline

### Expected results

### Motivation

Good and scientifically sound TI evaluation methodology is required in order for the field to progress further

## characterise\_tasks

### Statement

We characterised the different TI tasks

### Experiment

TODO

### Expected results

### Motivation

## construct\_synthetic\_gold

### Statement

We generated realistic synthetic expression data of cells as part of a dynamic process

### Experiment

Determine:

- \* the gene regulation model @regulation\_model
- \* the transcription @transcription\_model
- \* the dynamic process model @dynamic\_process\_model
- \* the sequencing model @platforms\_model

### Expected results

### Motivation

## regulation\_model

### Statement

We use GRNs as a model for gene regulation. These GRNs are derived from real regulatory networks

### Experiment

TODO

### Expected results

### Motivation

## transcription\_model

### Statement

We can generate expression given a GRN

### Experiment

TODO

### Expected results

### Motivation

## dynamic\_process\_model

### Statement

We can translate a dynamic process into a GRN such that generated expression therefrom resembles observations of cells following this dynamic process

### Experiment

TODO

### Expected results

### Motivation

## platforms\_model

### Statement

We determined a model of the technical noise generated by observing the state of a cell through RNA sequencing

### Experiment

Compare real data across platforms with the simulated data across platforms

- Drop-out rates
- Distributions
- Compare with "ideal" platform

### Expected results

### Motivation

Platforms could induce a bias in the performance of certain methods, so it is important that our platform data is comparable to real platform data  
This will allow us to investigate the advantages and disadvantages of different platforms

## generate\_synthetic\_data

### Statement

We generated synthetic data in accordance to our models:

- \* @regulation\_model
- \* @transcription\_model
- \* @dynamic\_process\_model
- \* @platforms\_model

### Experiment

We simulate using SDEs (with enough noise) and Gillespie SSA (with noise and single-molecule) which takes into account

- \* Technical bias introduced by platforms @simulating\_platform\_noise
- \* Biological
  - Modular
  - Dim red and example of single-cell paths

### Expected results

N/A

### Motivation

We need to generate realistic (or at least useful) expression data from GRNs in order to be able to prioritise TI methods for real datasets

## timings\_synthetic\_generator

### Statement

Our SSA implementation is /blazing fast/.

Goal: simulate 100'000 cells with 20'000 genes in <24 hours.

10-fold sampling of single cells can be allowed.

### Experiment

Perform timings experiments on our implementation of fastgssa versus that of GillespieSSA.

### Expected results

fastgssa is very faster, but still outputs very similar data for the same given network

### Motivation

If we want to generate realistic data, we need to be able to simulate as many cells and as many genes as would hopefully occur in near-future experiments

## construct\_biological\_gold

### Statement

We collected high-quality datasets containing cells along a dynamic process, of which we know the state of each of the cells

### Experiment

- \* Collect data
- \* Perform some kind of tests to ensure there is enough signal-to-noise?

### Expected results

### Motivation

## benchmarking\_pipeline

### Statement

We use a benchmarking pipeline to train the parameters of the methods and ensure that the datasets the methods are evaluated on are unrelated to avoid overfitting

### Experiment

For each of the TI tasks, we train the parameters of the methods on synthetic datasets and evaluate them on biological datasets

### Expected results

### Motivation

## parameter\_crossvalidation

### Statement

For each of the tasks, we executed the different methods and trained their parameters using the metrics specified by @metrics

### Experiment

Train parameters for each of the TI tasks either by generating a grid of parameters or by using a smarter parameter optimisation algorithm

### Expected results

### Motivation

## metrics

### Statement

We have several good metrics to evaluate different aspects of predicting a good trajectory:

- \* the main structure of the predicted trajectory should be good
- \* the ordering of the cells should be good
- \* metric with distance from origin

### Experiment

Unit test metrics with toy examples

### Expected results

Each metric makes sense theoretically and fulfill the requirements of the toy example, although they can focus different aspects (or levels) of the TI model

### Motivation

Good metrics are the foundation of a good evaluation

## benchmarking

### Statement

We benchmarked TI methods using the real and simulated data and constructed a set of guidelines for the user to follow

### Experiment

We need to:

- \* generate synthetic data @generate\_synthetic\_data
- \* perform parameter cross-validation @parameter\_crossvalidation
- \* find methods (& parameters) which work well for certain datasets/task characteristics @characteristics\_vs\_performance
- \* construct a set of guidelines for the reader with which to decide which method to use @construct\_guidelines

### Expected results

### Motivation

## method\_fairness

### Statement

We treat each TI method fairly

### Experiment

- \* No TI method is left out
- \* Used implementation from original authors if available
- \* Made own implementation if necessary; contacted original authors to verify implementation
- \* Described all sorts of parameters, which will be optimised later

### Expected results

### Motivation

## characteristics\_vs\_performance

### Statement

We could link performance to dataset and method metadata

### Experiment

Use dataset and method metadata to determine which methods and parameters work well for certain TI tasks

### Expected results

We hope to see links between the performance and the metadata

### Motivation

We need to do this in order to start constructing guidelines

## construct\_guidelines

### Statement

We constructed a set of usage guidelines

### Experiment

Attempt to construct TI usage guidelines depending on the TI task.

### Expected results

This will likely not be feasible, but would be extremely useful.

### Motivation

Guide the biologists in their selection of methods

Guide the bio-informatician in the choice of future methods

## simulate\_inferred\_networks

### Statement

We can generate expression data on inferred networks of developing cells

### Experiment

0. Collect expression of cells in a dynamic process
1. Perform NI
2. Generate expression of GRN
3. Apply TI on real expression data
4. Apply TI on synthetic expression data
5. TI results on real expression data should be comparable to synthetic expression data

### Expected results

Data looks similar, as long as we can trust the NI

### Motivation

It is a validation of the whole synthetic data generation workflow

## biological\_validation

### Statement

Results from real and simulated data are similar, validating the simulated data, especially on more complex tasks for which no decent real datasets are available. The selected datasets cover the range of TI tasks as well as possible.

### Experiment

Apply guidelines on real data and show that the selected methods work well in comparison to when the guidelines are not followed. A TI method that obtains a good relative performance on synthetic data should also obtain a good performance on the biological data

### Expected results

### Motivation

## tradeoff\_cost\_performance

### Statement

We compared current platforms, estimated the cost to improve certain parameters, and predicted which parameters to improve to obtain the best performance increase versus cost increase

### Experiment

We should attempt to quantify the cost associated with improving each of the sources of technical noise, and make a trade-off between TI performance and total cost

### Expected results

### Motivation

scRNA-seq platforms will improve. We can find out what the most cost-effective way of choosing/improving a platform is in order to study dynamic processes.

## future

### Statement

We also look to the future of single-cell technologies and TI

### Experiment

We simulate data which will be available soon  
We also generate data which can't be modelled yet by current methods

### Expected results

N/A

### Motivation

We have only seen the tip of the iceberg of which can be modeled using current TI methods, and we thus stimulate the development of future methods and the generation of data from more complex biological settings

## mrna\_vs\_protein

### Statement

Observing protein levels might improve TI

### Experiment

Add flowcyto and proseq platforms. Evaluate

### Expected results

As cells differentiate, mRNA levels increase first, protein levels increase later.  
Observing protein expression levels might thus improve TI.

### Motivation

Integrative TI with index sorting

### multiple\_processes

**Statement**

Most existing methods will have problems detecting multiple processes. These might be synchronised or not.

**Experiment**

Generate networks and expression data containing multiple dynamic processes

**Expected results**

Watch TI methods fail horribly

**Motivation**

Promoting the development of better methods

### intercellular\_communication

**Statement**

We generated data in which multiple cells interact with each other, influencing each other's expression. And develop methods to synchronize multiple cells

**Experiment**

Add networks between cells  
Slightly different metric to comparison

**Expected results****Motivation**

scRNA-seq will in the future allow the inference of networks between cells

### modular\_ti

**Statement**

Each TI approach can be disassembled into smaller components, new TI approaches can be constructed from a combination of components

**Experiment**

TODO

**Expected results****Motivation**

### manuscript

**Statement**

We communicated to the 100 most interested people

**Experiment**

Write a clear and concise manuscript

**Expected results****Motivation**

### large\_communication

**Statement**

We communicated to the 1000 most interested people

**Experiment**

We make our evaluation code available as a package. We post the results on the github page. When the evaluation pipeline is modified (i.e. new methods or datasets have been added), the results are also updated

**Expected results****Motivation**

### mass\_communication

**Statement**

We communicated to 10000+ people

**Experiment**

Go to conferences and/or contact the press?

**Expected results**

Fame and glory

**Motivation**

### communication

**Statement**

We communicated our results to a broad audience in an interesting manner

**Experiment**

TODO

**Expected results****Motivation**