Sergieo

Data generation

Using the Sergio Single cell simulator, we will produce a “clean” data set (biological variance included in the simulator but no technical noise)

The GRN used to produce the “Clean” data set will be read to produce an equivalent regulatory connection graph.

The “clean data” will be used for the generation of further datasets with increasing level of the four categories of technical noise that are capacities of the Sergio simulator.

Question(s)

Data \_confirms reality:

What sample size and # of replicates from the clean simulated data are required to confidently\* confirm the veracity of the regulatory network generated from the GRN?

Next, what are the sample sizes and # of replicates are required as we move further away from the “Clean” data to the more technically noisy data sets to make the same confirmation to an acceptable confidence.\*\*

Lastly, at what distance in technical noise from “Clean” are we no longer able to confidently confirm the veracity of the correct Regulatory network.

Data\_informs\_our\_understanding:

For these experiments will modify the Regulatory network away from that correctly produced from the GRN, and using the same outline as above, we will find the sample size and replicate numbers required at each technical noise distance form “clean” where the data can invalidate the “incorrect” regulatory network, and the point that it can confidently suggest the correction to return to the regulatory network which correctly reflects the GRN.