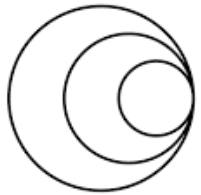


wellcome
connecting
science

Getting started with logical modelling

David Shorthouse

Ben Hall

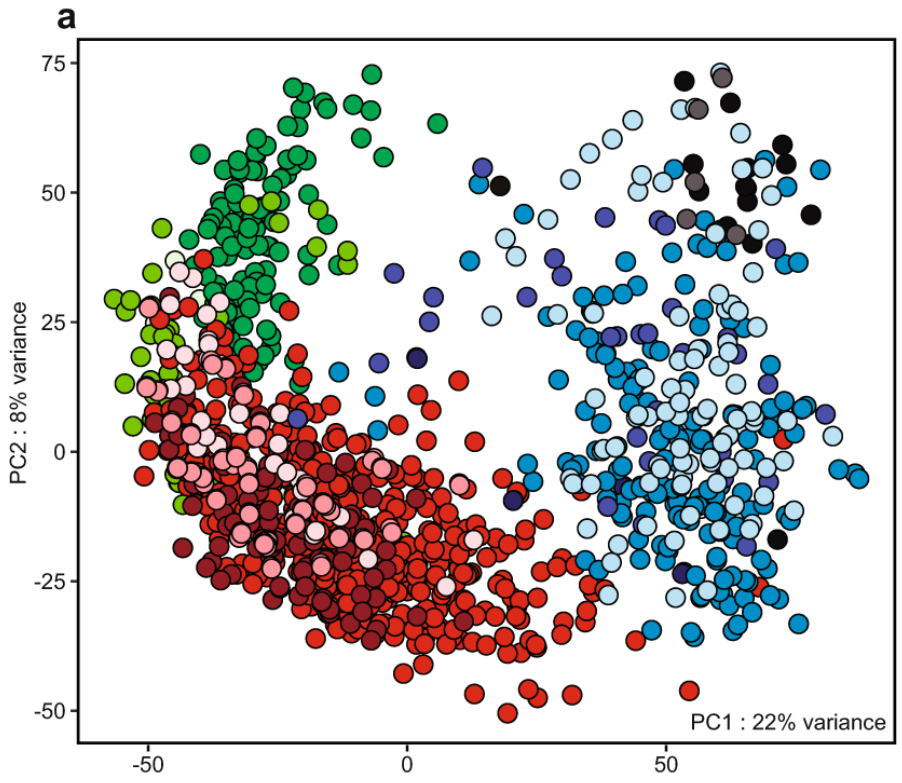
This exercise aims to get you exposed to all the steps of generating and analysing a logical model based on biological data (in this case transcriptomics)

The idea is for you to see the entire “pathway” behind generating a dynamic model from data, so that you can more easily place each tool we will cover over the week in context

~15/20 minute intro, then a practical exercise

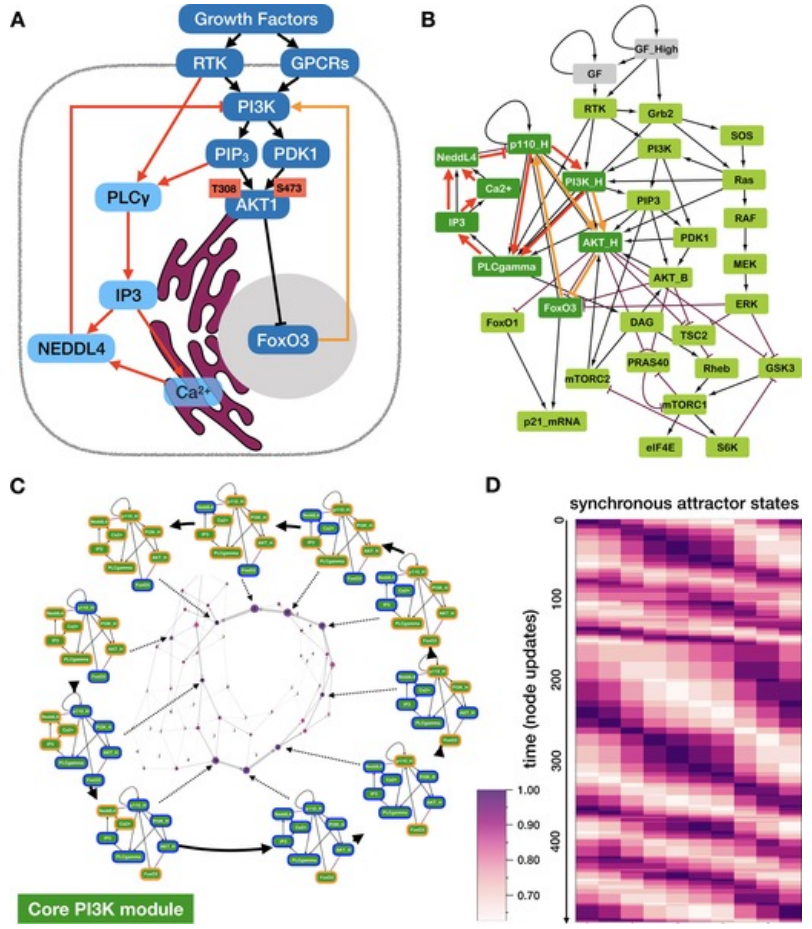
Please ask questions!

Expectation



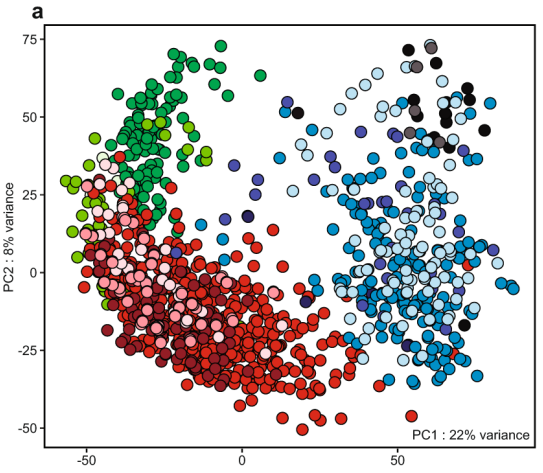
Complex –omics data

AI wizardry

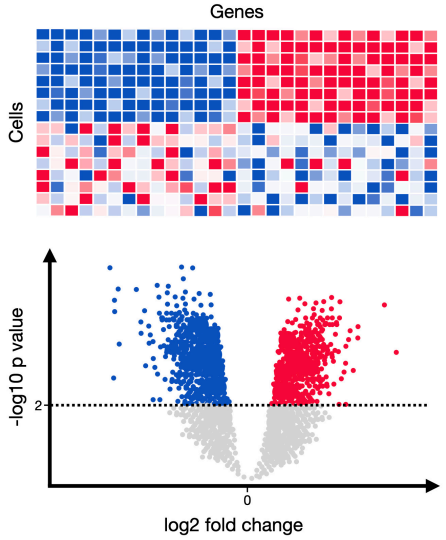


Biological mechanisms
Explainability

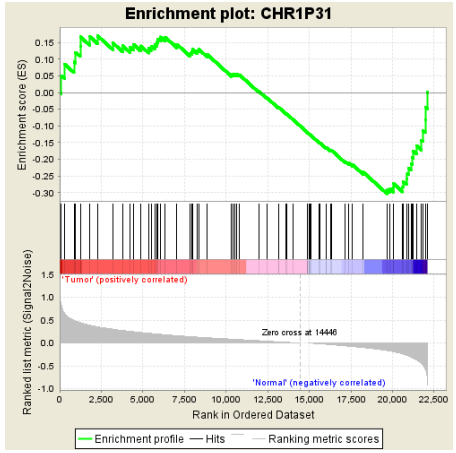
Reality



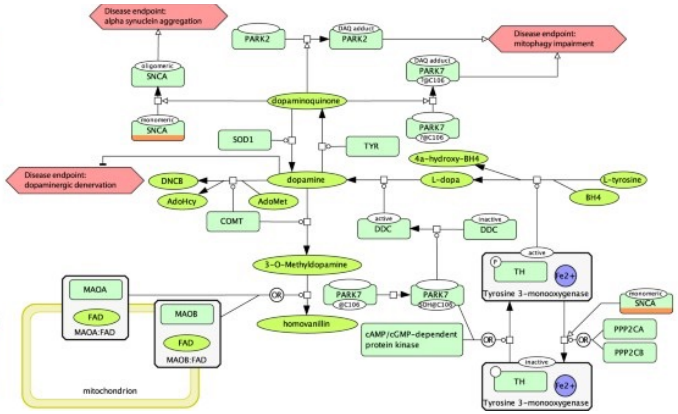
Complex –omics data



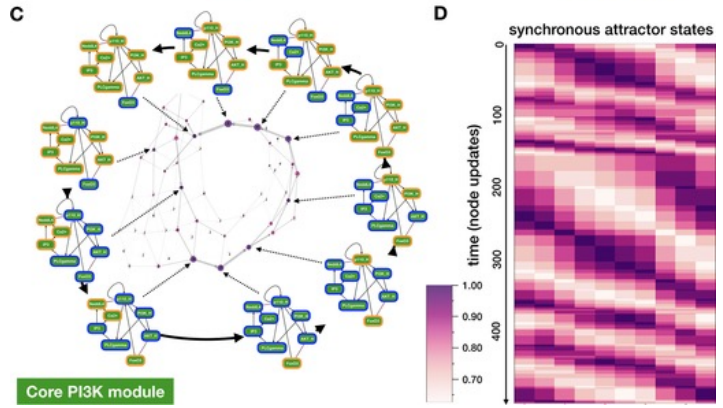
Differentially expressed/identified genes/proteins/metabolites



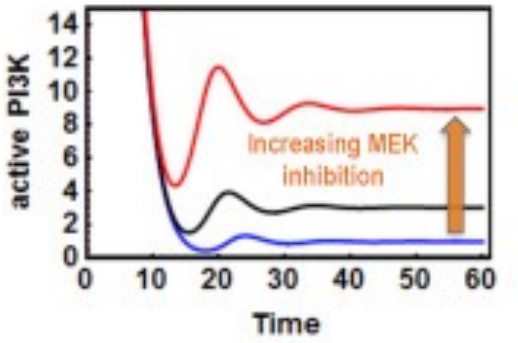
Pathways + enriched sets of genes



Static disease map

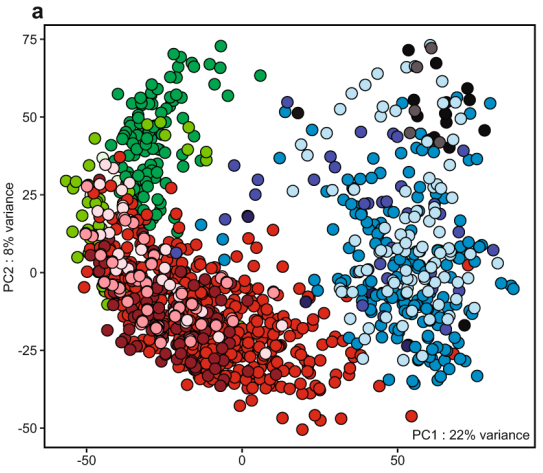


Dynamic model



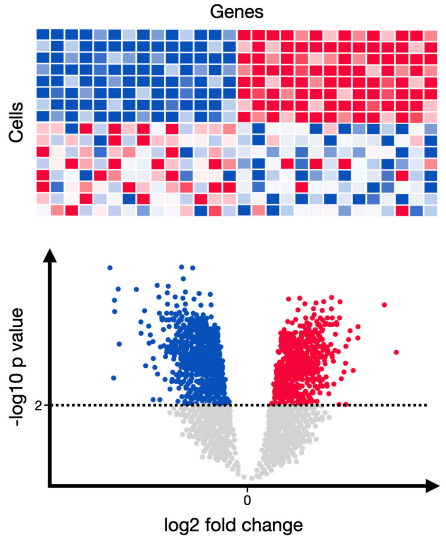
Biological mechanisms
Explainability

Reality



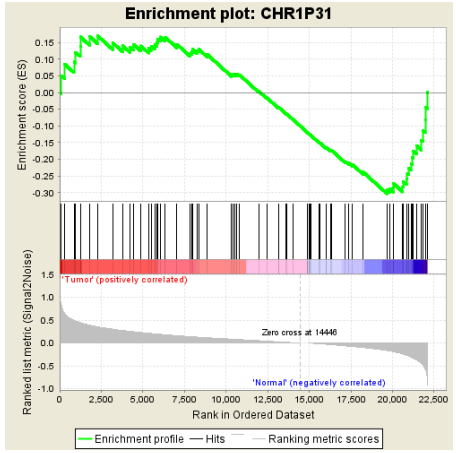
Complex –omics data

Manual



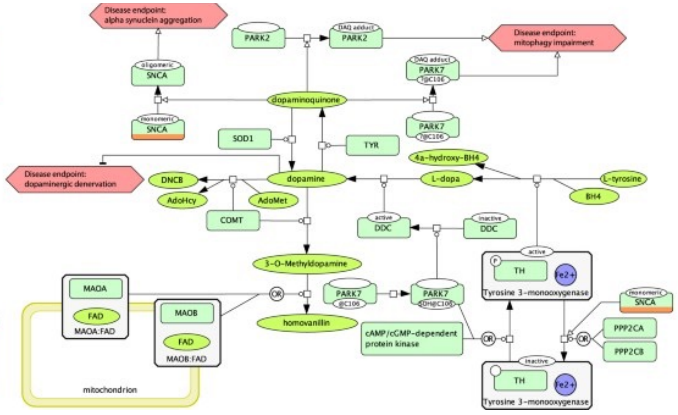
Differentially expressed/identified genes/proteins/metabolites

Manual



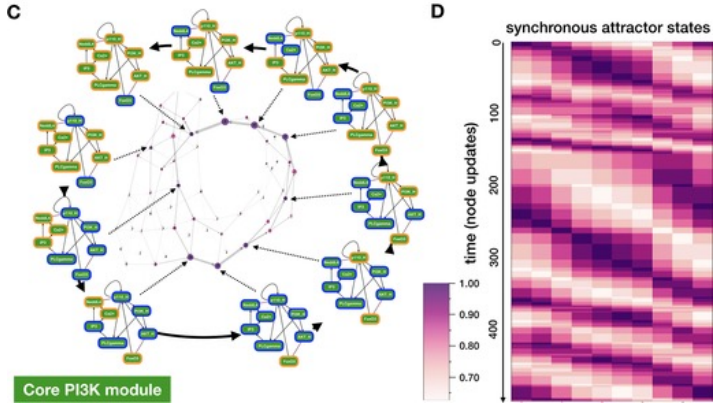
Pathways + enriched sets of genes

Manual



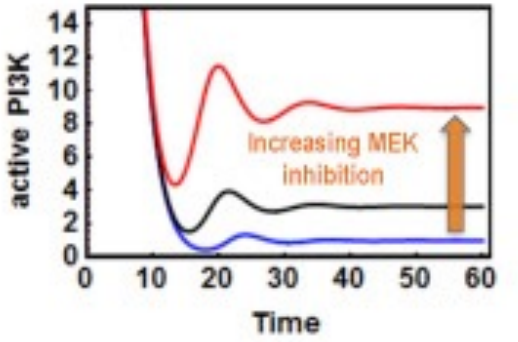
Static disease map

Manual



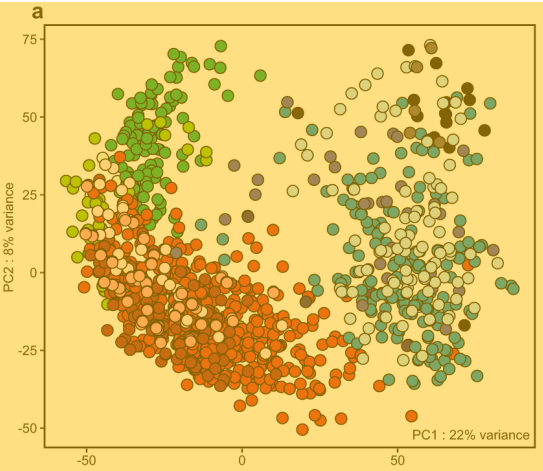
Dynamic model

Manual



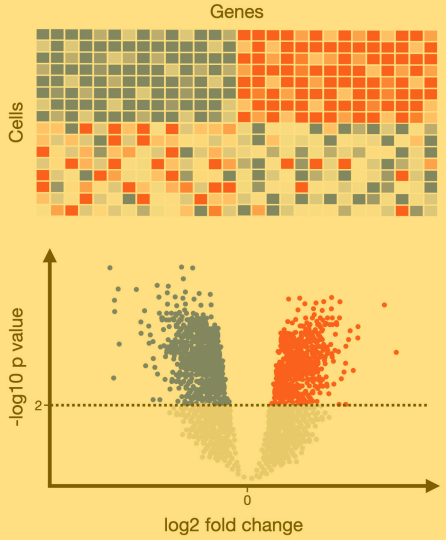
Biological mechanisms
Explainability

Reality



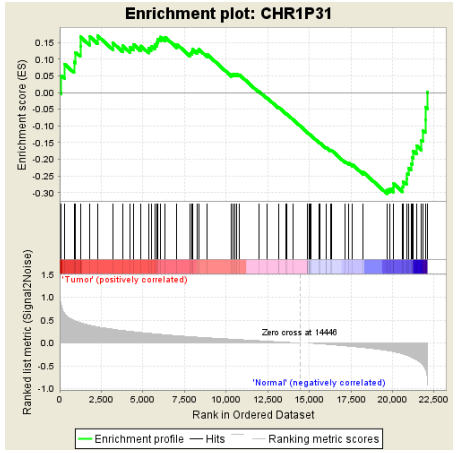
Complex –omics data

Manual



Differentially expressed/identified genes/proteins/metabolites

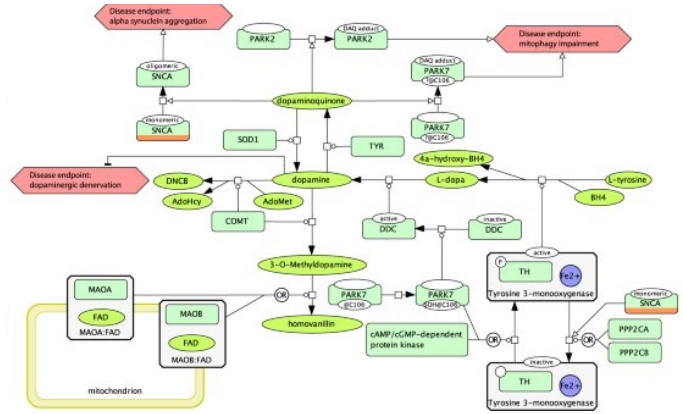
Manual



Pathways + enriched sets of genes

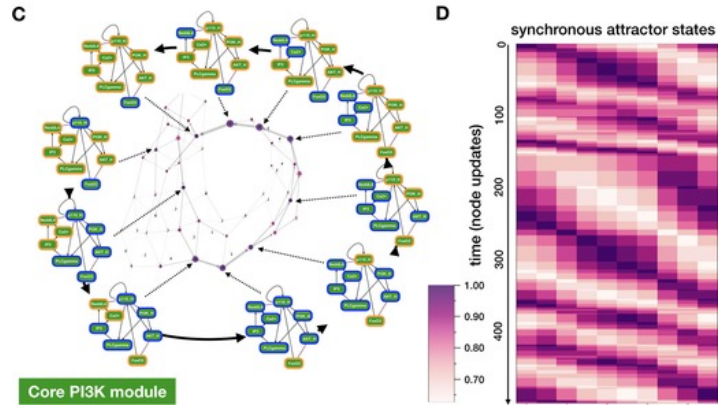
Not covered extensively this week

Manual



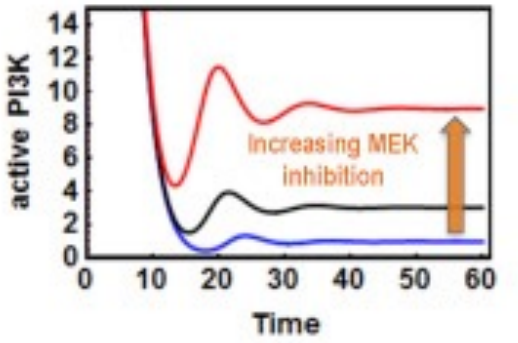
Static disease map

Manual



Dynamic model

Manual



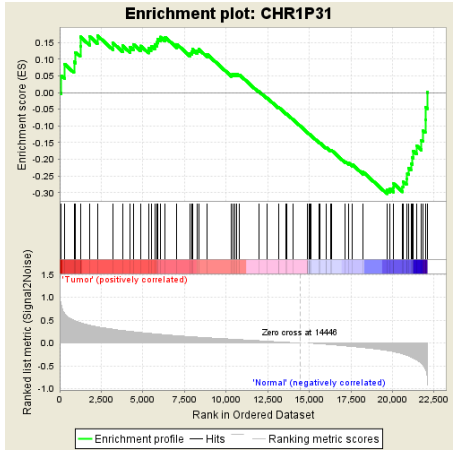
Biological mechanisms
Explainability

Some tools:

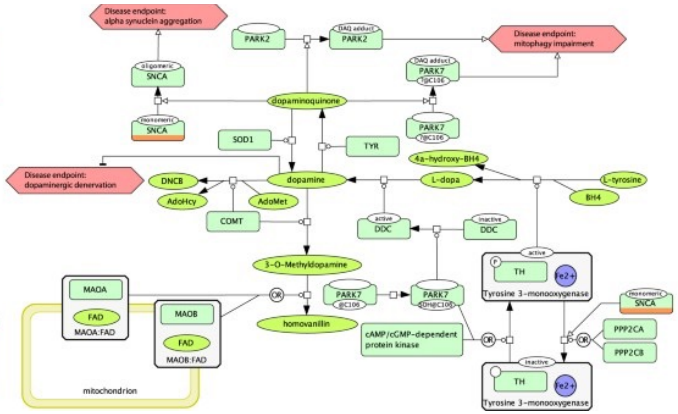
Celldesigner
Cytoscape

CaSQ

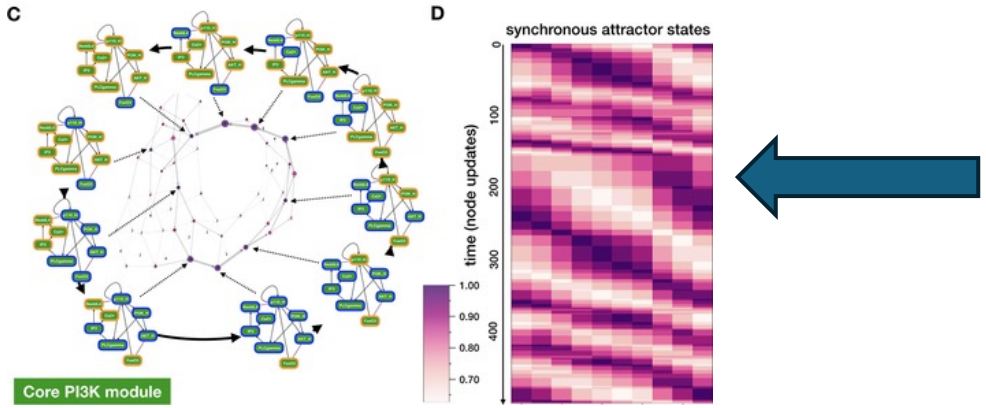
MaBoSS
CellCollective
BioModelAnalyzer
CoLoMoTo



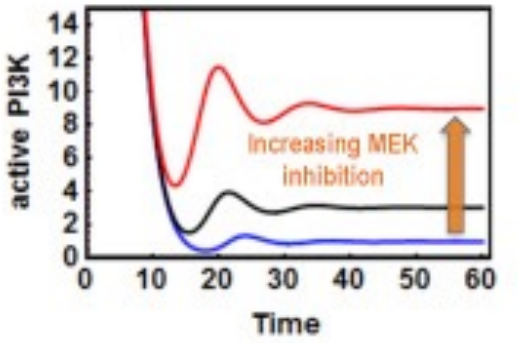
Pathways + enriched sets of genes



Static disease map

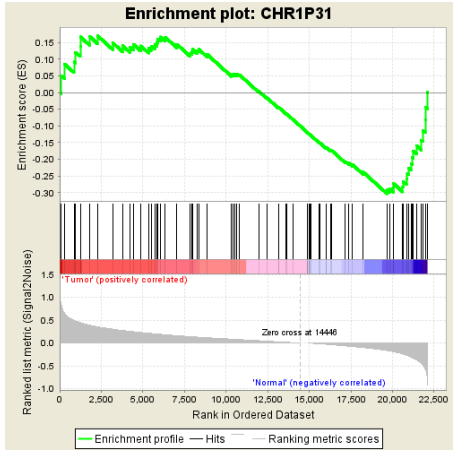


Dynamic model



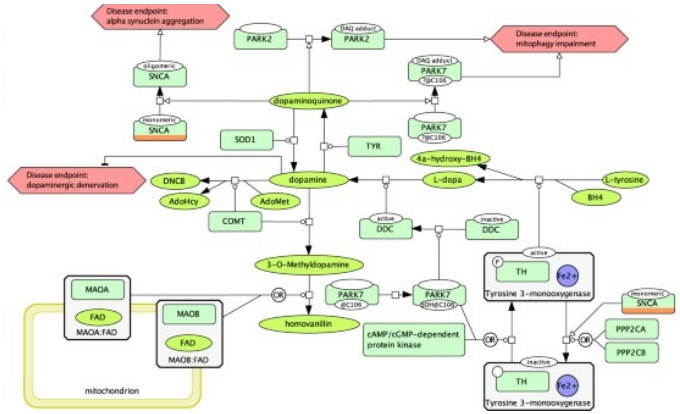
Biological mechanisms
Explainability

Some tools:



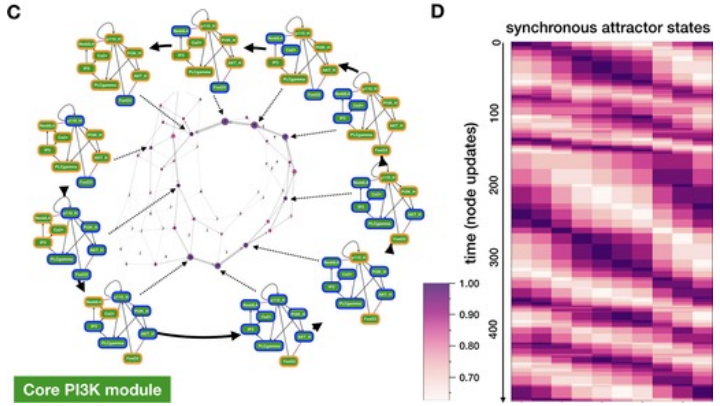
Pathways + enriched sets of genes

Celldesigner
Cytoscape



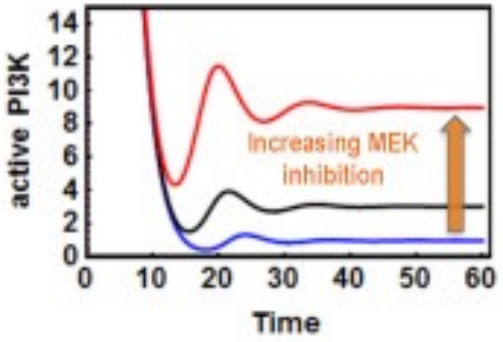
Static disease map

CaSQ



Dynamic model

MaBoSS
CellCollective
BioModelAnalyzer
CoLoMoTo



Biological mechanisms
Explainability

General idea (very similar to bioinformatics):

- Some stages of this process can be automated to some extent

- Manual curation and intervention is **almost always required** along this process

- A dynamic model can be analysed in different ways with different tools, to give insights into different facets of the network

Disclaimer

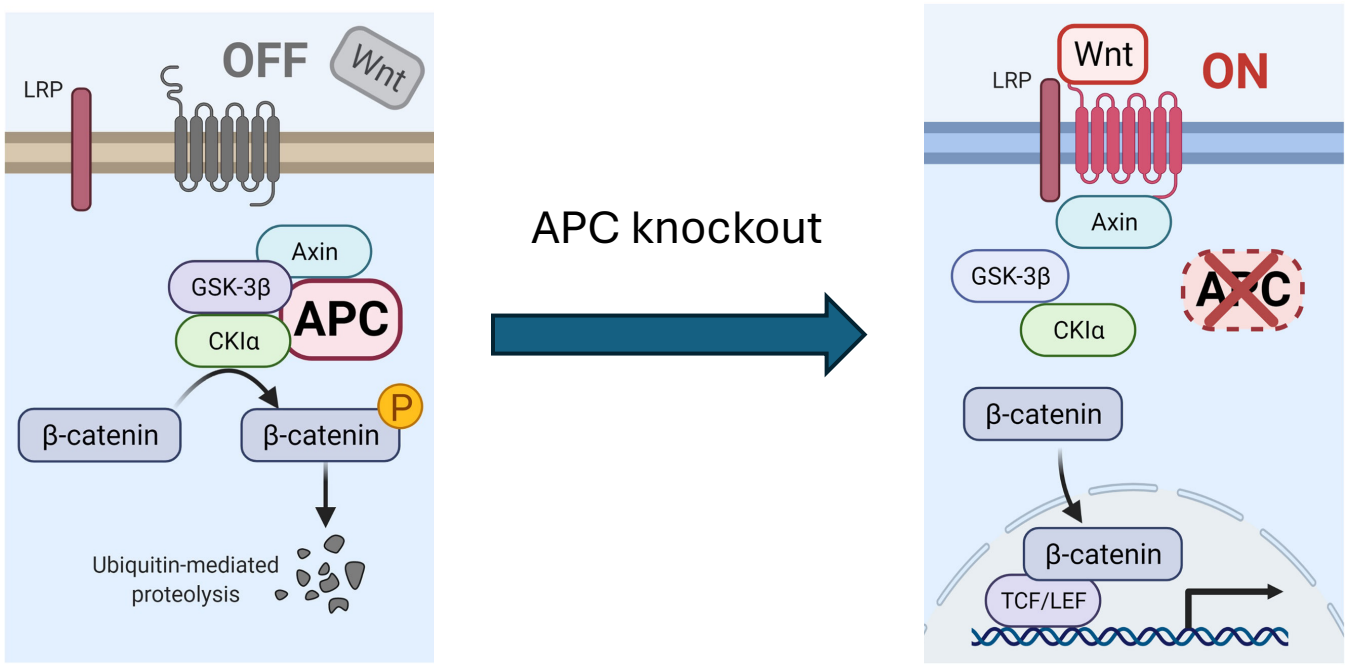
Our workflow today is not necessarily “best practice”, and our example is arbitrary and simple

Example system:

APC – Adenomatous Polyposis Coli

Loss of function mutations are present in %80 of colorectal cancers

Well-understood as a mediator of the WNT pathway



Example system:

In King et al. (2016), a colorectal cancer cell line that is mutant for APC is **restored to WT**

RNA-seq analysis is performed

https://doi.org/10.1016/j.gdata.2016.02.001

Genomics Data 7 (2016) 293–296

Contents lists available at ScienceDirect

Genomics Data

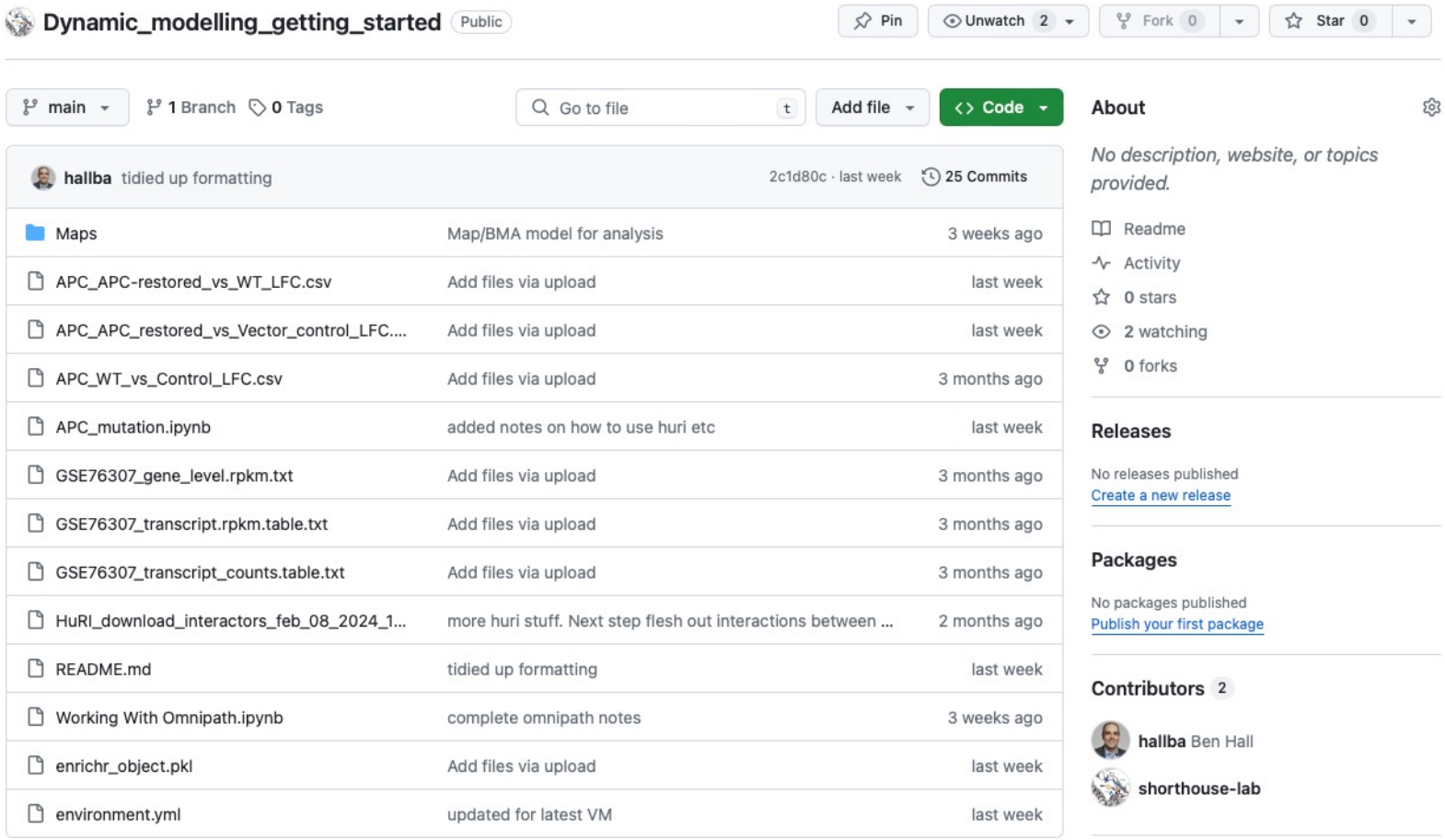
journal homepage: www.elsevier.com/locate/gdata

Differential RNA-seq analysis comparing APC-defective and APC-restored SW480 colorectal cancer cells

Lauren E. King [a,b,c,1], Christopher G. Love [b,c,d,1], Oliver M. Sieber [b,c,d,e,f], Maree C. Faux [a,b,c], Antony W. Burgess [a,b,c,f,*]

Example system:

https://github.com/shorthouse-lab/Dynamic_modelling_getting_started



Github contains all the data needed to run everything, and jupyter-notebooks with code

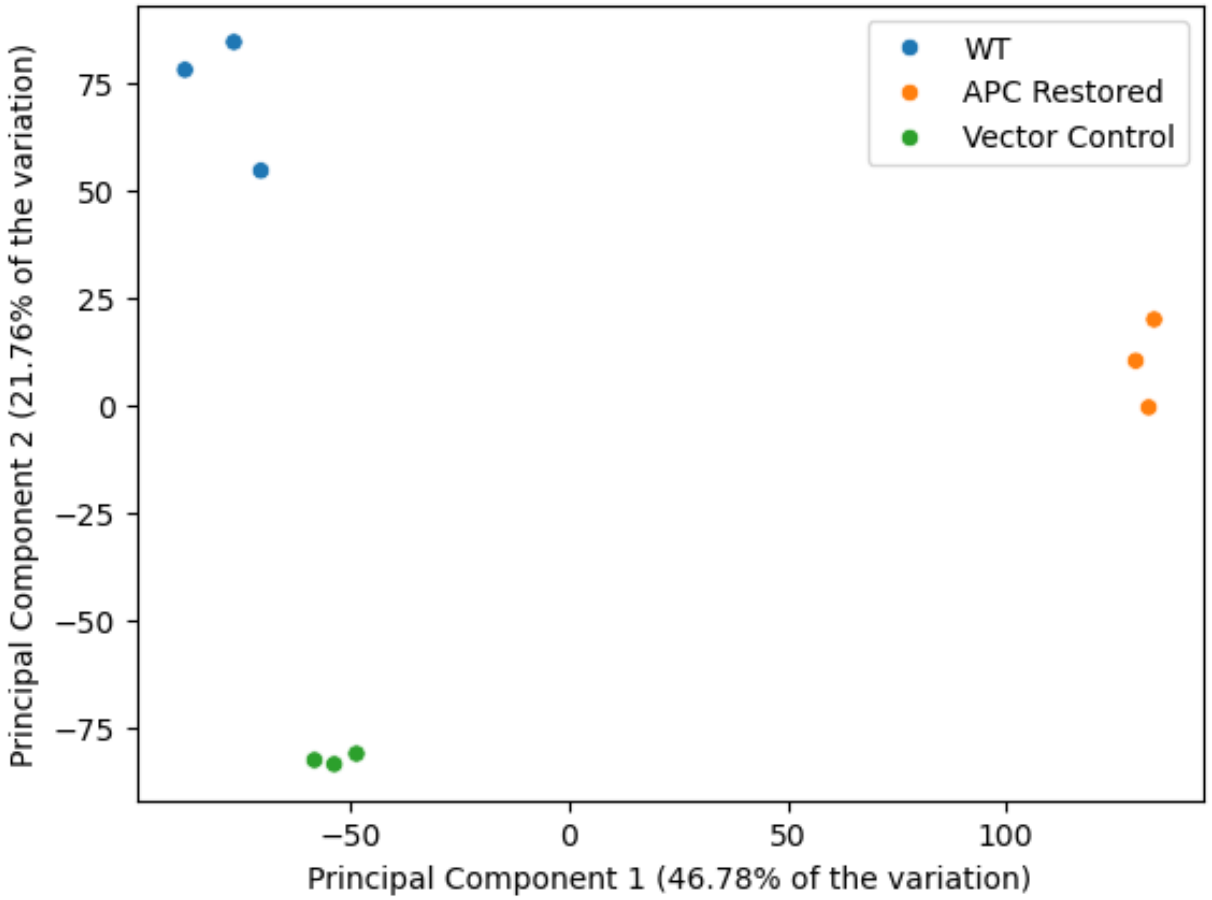
Example system:

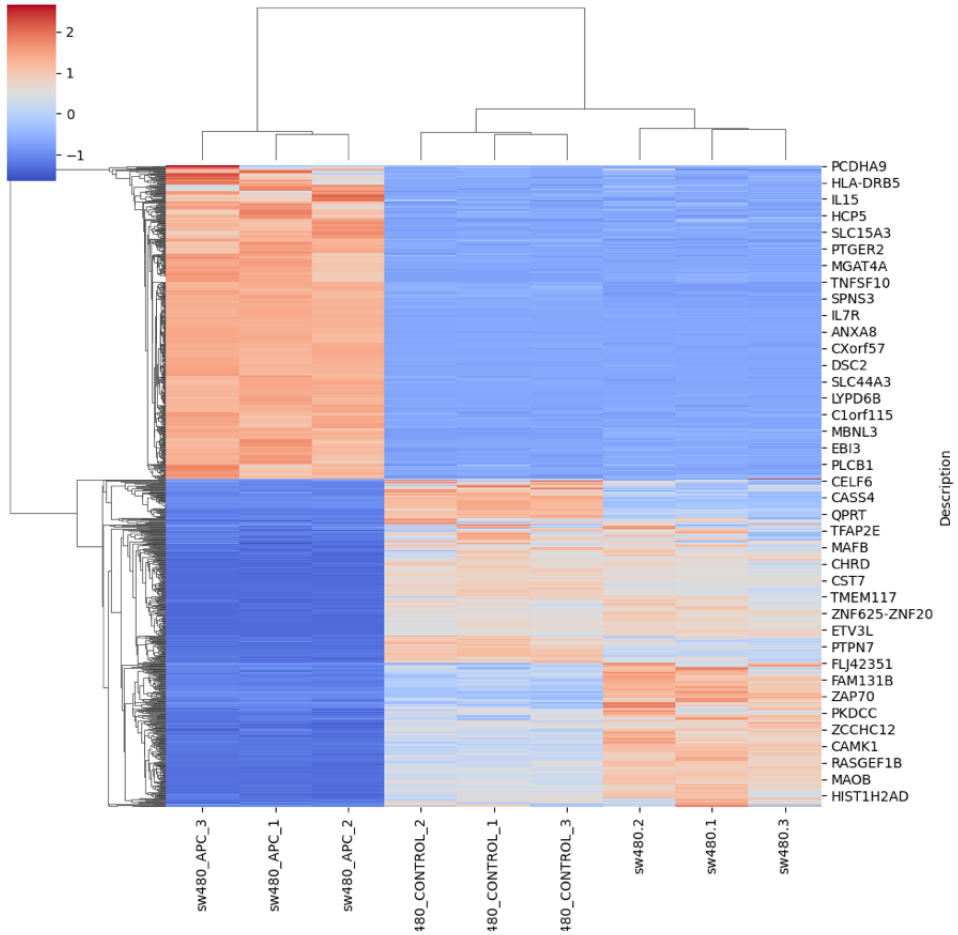
Some dependencies are required before running everything (from the readme):

```
conda env create -f environment.yml
conda activate quickstart
curl https://sh.rustup.rs -sSf | sh -s -- -y
export PATH="$PATH:$HOME/.cargo/bin"
pip install gseapy casq
```

General workflow:

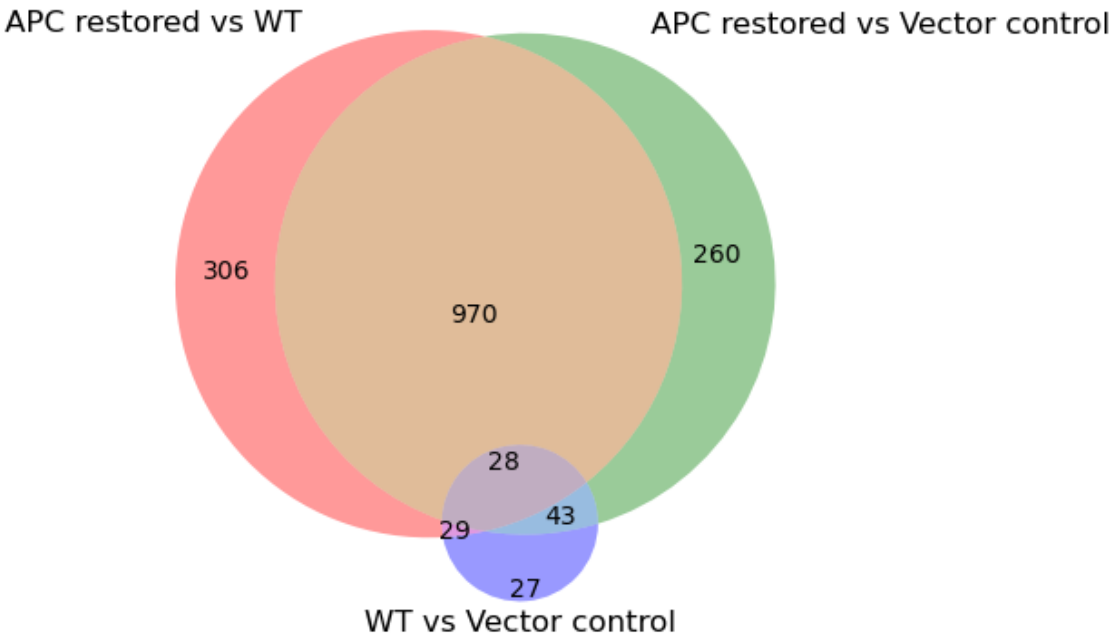
Some basic RNAseq analysis to check the experiment has worked





General workflow:

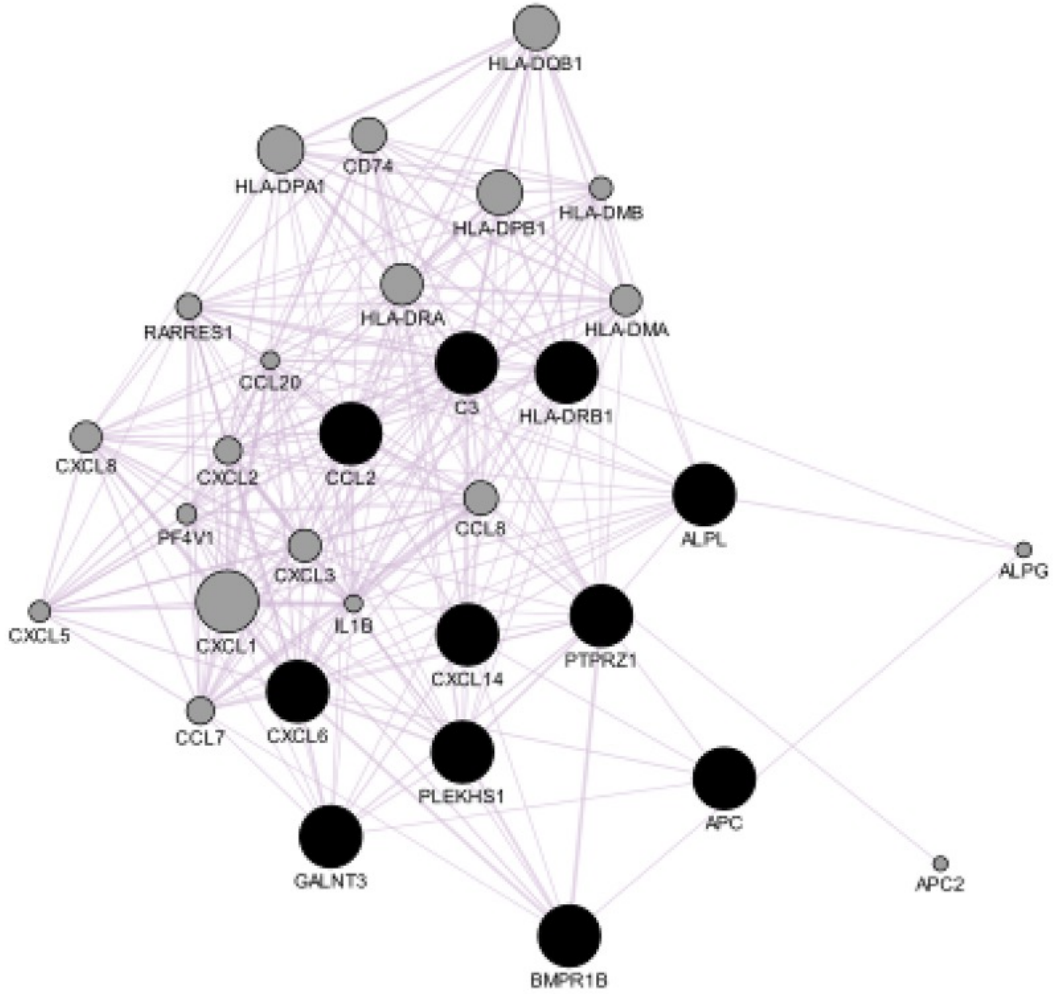
Identify sets of differentially expressed genes and pathways



	Gene_set	Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score
0	MSigDB_Hallmark_2020	TNF-alpha Signaling via NF-kB	2/200	0.005157	0.027504	0	0	22.212121	117.000045
1	MSigDB_Hallmark_2020	Interferon Gamma Response	2/200	0.005157	0.027504	0	0	22.212121	117.000045
2	MSigDB_Hallmark_2020	Inflammatory Response	2/200	0.005157	0.027504	0	0	22.212121	117.000045
16	KEGG_2021_Human	Cytokine-cytokine receptor interaction	4/295	0.000014	0.000356	0	0	38.680412	432.021505
17	KEGG_2021_Human	Rheumatoid arthritis	3/93	0.000016	0.000356	0	0	82.912500	917.527195
18	KEGG_2021_Human	Viral protein interaction with cytokine and cy...	3/100	0.000019	0.000356	0	0	76.902062	834.256871
19	KEGG_2021_Human	Chemokine signaling pathway	3/192	0.000136	0.001867	0	0	39.285714	349.817130

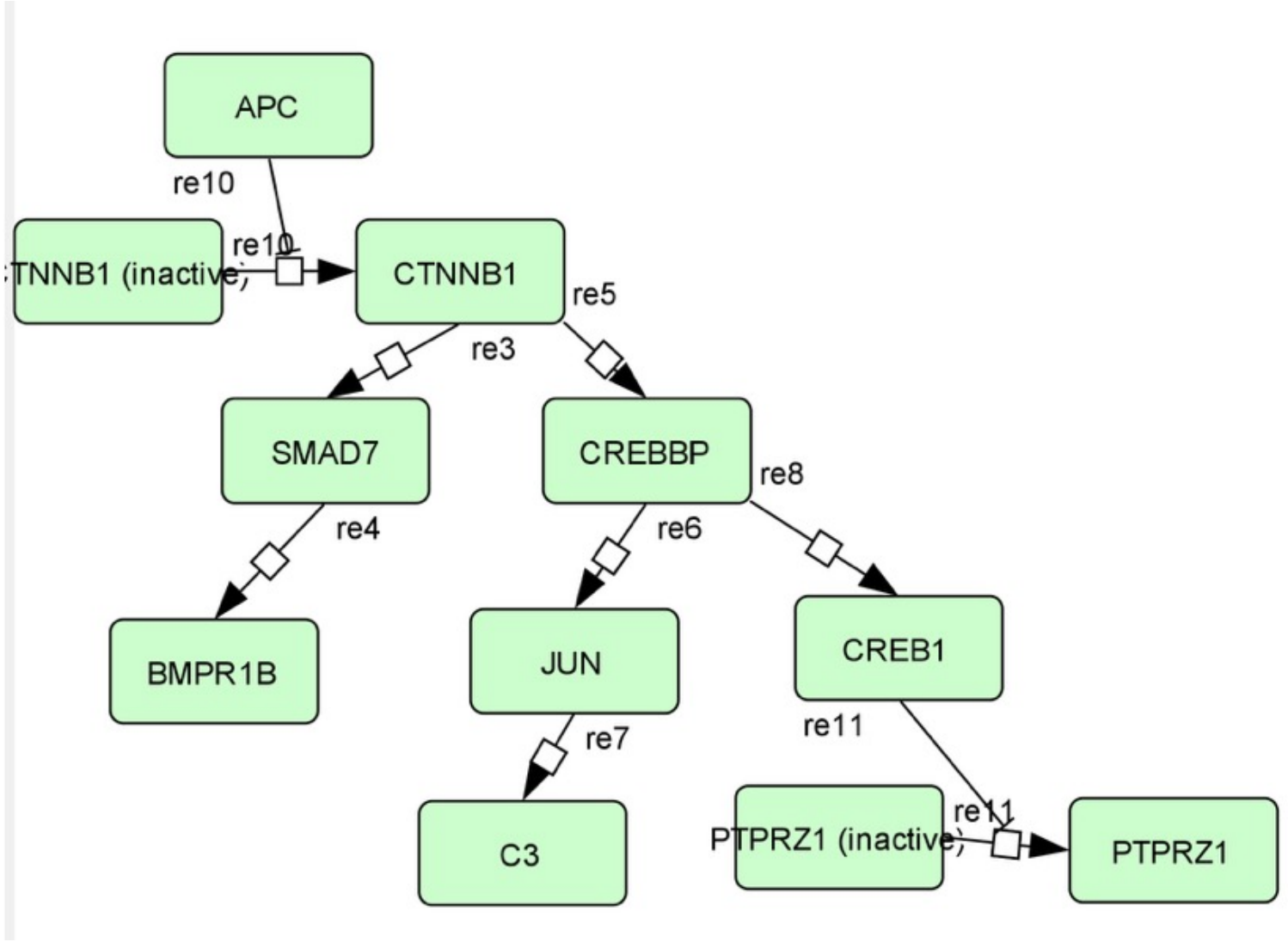
General workflow:

Generate static interaction networks



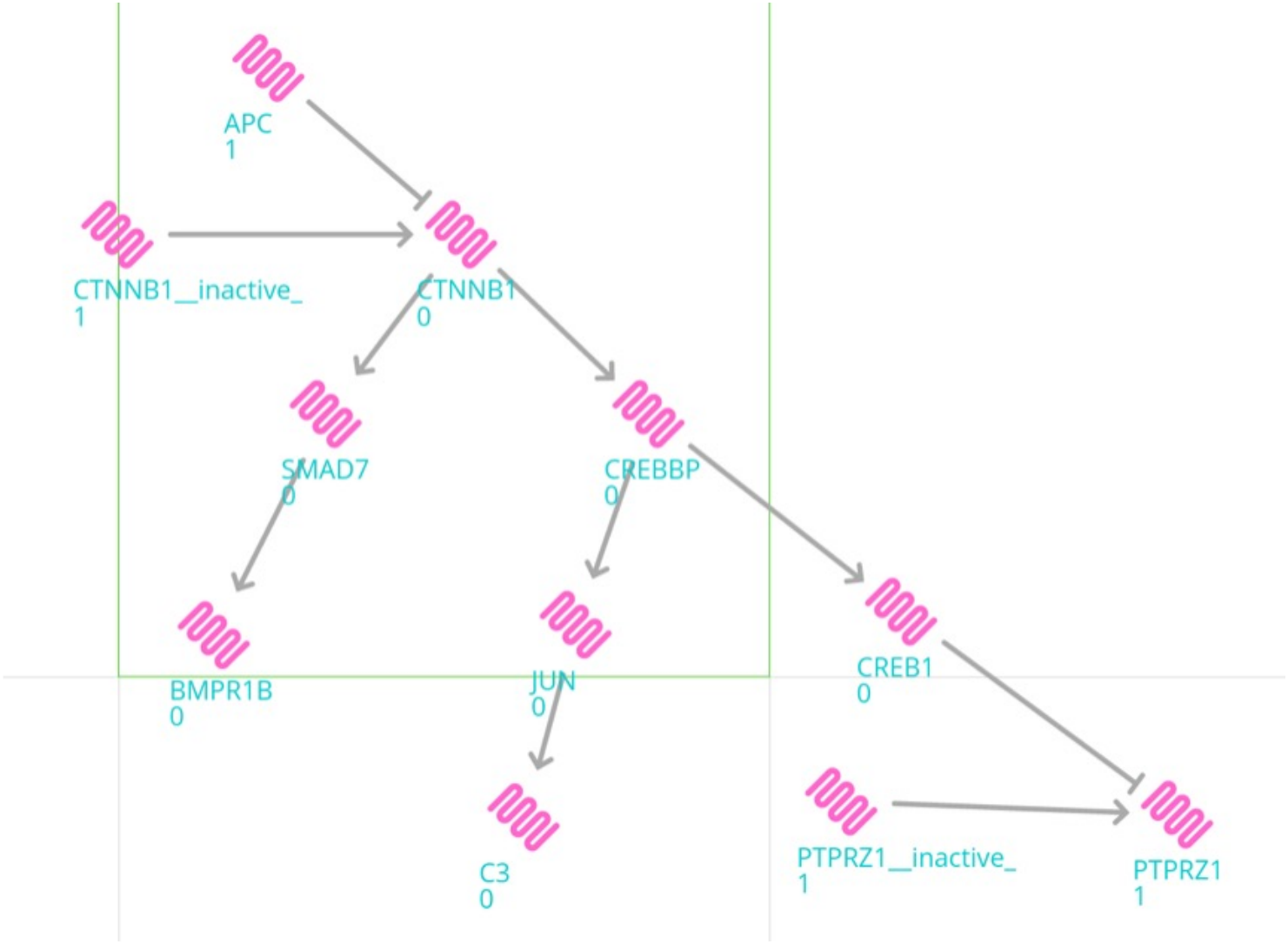
General workflow:

Build a basic map



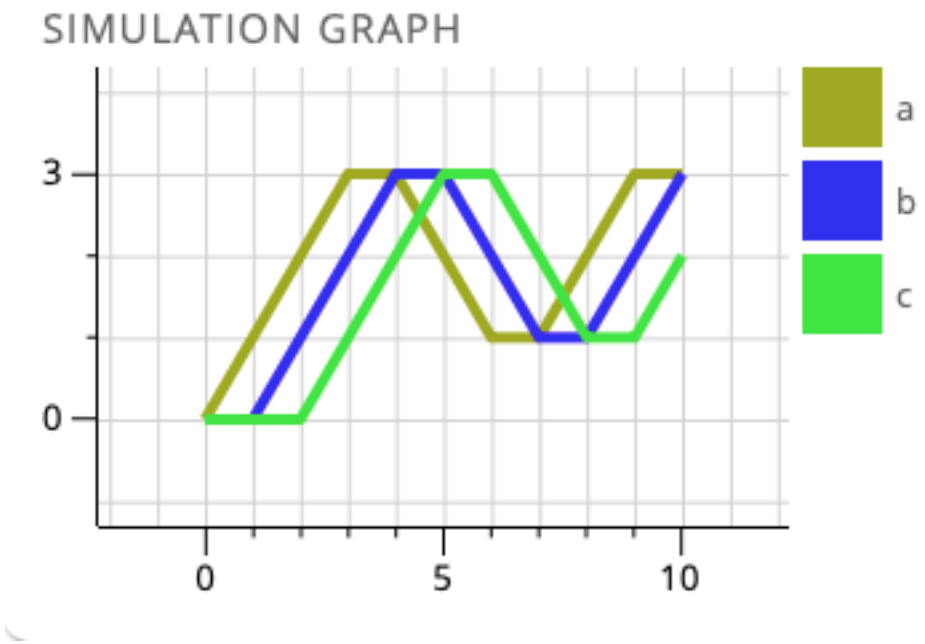
General workflow:

Add some dynamics



General workflow:

Do some basic analysis



Files are all in the course github:

CompSysBio24 / Modules /

Add file ···

shorthouse-lab Update readme.md 8dc7f52 · now History

Name	Last commit message	Last commit date
..		
MaBoSS	Initial commit of MaBoSS material	2 weeks ago
Monday_DynamicModellingQuickStart	Update readme.md	now
Monday_Lectures_HandsOn_files_AN	Add files via upload	4 days ago
Projects	Add files via upload	4 days ago
Suggested_reading	Add files via upload	3 days ago
Sunday_presentation	Add files via upload	4 days ago
Tuesday_Lectures_HandsOn_files_AN	Add files via upload	4 days ago
README.md	Create README.md	3 months ago
module_base.md	Create module_base.md	3 months ago

Download and run the jupyter notebooks:

CompSysBio24 / Modules / Monday_DynamicModellingQuickStart / APC_mutation.ipynb

shorthouse-lab Add files via upload 3820a16 · 3 minutes ago History

Preview Code Blame 5409 lines (5409 loc) · 1.78 MB Raw

Quickstart guide to mapping and modelling disease

This practical is written to give you a rapid overview of how to get started building and analysing models. The learning aim here is not to teach everything, but to present one workflow that illustrates how you can move from an experimental dataset, to gene network maps, to models. Everything in this guide is expanded on in more depth later, including the practical and theoretical aspects, but you should be able to adapt this as a starting point for your own data.

How mutation of APC alters cancer signalling

We have chosen an example to get started:

"Differential RNA-seq analysis comparing APC-defective and APC-restored SW480 colorectal cancer cells"
https://doi.org/10.1016/j.gdata.2016.02.001

Adenomatous polyposis coli (APC) tumour suppressor gene is mutated in 80% of colorectal cancers (CRC). Its mutation is a well studied step in the Vogelstein model of tumour development. In this dataset APC has been restored in a CRC cell line, and gene expression changes measured with RNASeq. This dataset allows us to explore the role of APC in the control of the cancer phenotype.

In this tutorial we:

- Process the raw data
- Perform differential expression analysis to identify genes activated and inhibited by the reintroduction of APC
- Use GSEA to identify pathways changed by the presence or absence of APC (to be confirmed)
- Develop maps of signalling
- Convert these maps into models

In [1]:
```
import pandas as pd
```

Once finished with the notebooks you could:

- Add extra genes from the pathway analysis

- Think about ways of connecting other genes of interest

- What questions could you ask of this network? How would you improve it to answer those questions?

Ask us questions!