Readme

Link to paper: https://arxiv.org/abs/2401.11858

The python code for generating the figures is divided into two main sections:

(1) For generating pathways
(2) For computing Lyapunov exponents

For (1) the main sections with self descriptive labels in the colab are:

(a) Load data into dataframe
(b) Preprocess data, which includes:
    (i) Split data into “Reference vs pathological data”
    (ii) Define the HUGO symbols of Nuclear Receptors
    (iii) Search for annotate the cyclins
(c) Use cyclins to sort and then compute matrix of increments, which includes:
    (i) Sort by age and compute diffusivity matrix
(d) Solve the compressed sensing problem (Algorithm 1)
(e) Compute perturbations (Algorithm 2)
(f) The main loop of the algorithm to generate pathways (Algorithm 3)
(g) Display pathways

For (2) the main sections with self descriptive labels in the colab are:

(a) Basket trial candidates (which is the start of code for (2))
(b) Generate Heatmap (which generates the heatmap in Fig 1) (Algorithm 4)
(c) Generate equilibrium values by brute force (which generates the equilibrium from Eqn 5)

All of the code is self-descriptively named. In order to run the pipeline, RNA-seq data containing various cyclins (or another suitable approximate clock is required). Note that the code does not split samples into homogeneous subcohorts. We also do not try to clean or filter data in any substantial way (for e.g. adjusting for drop outs etc). The latter is not a limitation as we work with nuclear receptors which are fairly prominently expressed, but some filtering may be required when dealing with other gene sets.