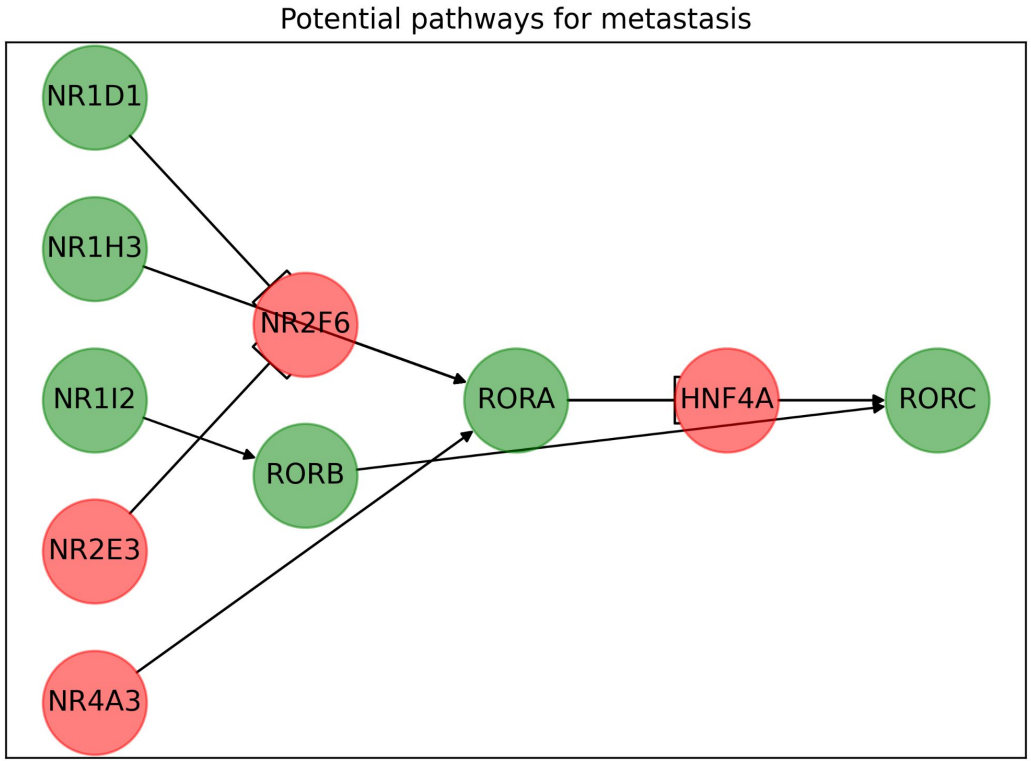
Changes in nuclear receptor networks in cancer

Pratik Worah

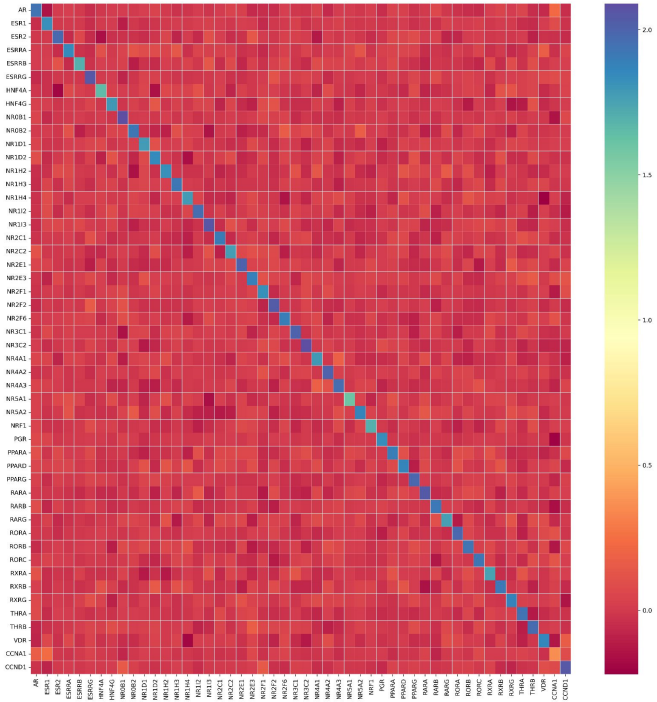
Problem

1. Gene network: each gene is directly influenced by a handful of different genes and in turn directly influences other genes, this interaction map is the “gene network”.
    a. Note: one can replace “directly influences” by a weaker notion of influence, say influences within a small number of steps.
2. Goal: To compute critical pathological changes, i.e., perturbations, in ***nuclear receptor networks*** of metastasis vs. non-metastasis cases in breast cancer.
3. Input: RNA-seq data from brca_tcga_2012.
4. Output:
    a. Heatmaps for:
        i. The gene network for a selected subset of ~50 genes amongst cancer patients in those datasets, and
        ii. The perturbations to the network for non-metastasis (M0) vs. metastasis (M1) cases.
    b. Predicted genomic pathways that may be responsible for metastasis (recovered using the perturbation map above)
5. ***Summary: perturbation map can give much more information about underlying mechanisms than differential gene expression.***

Prominently perturbed pathways

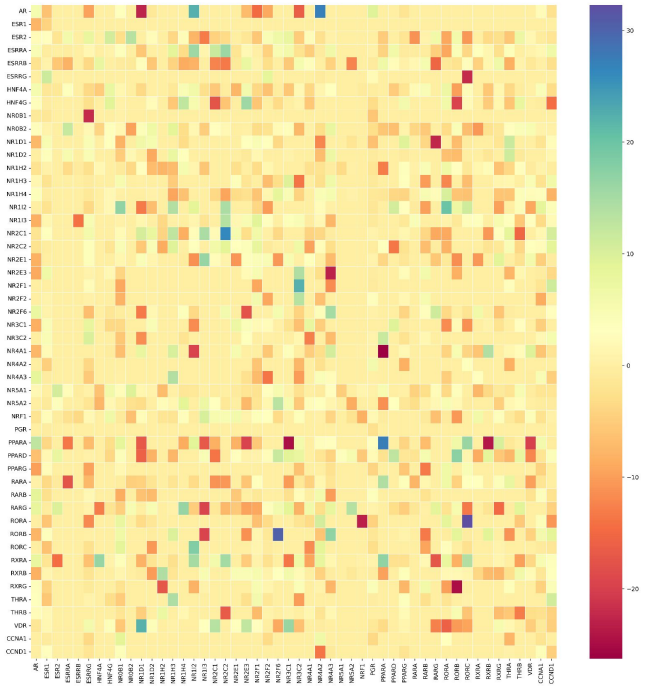


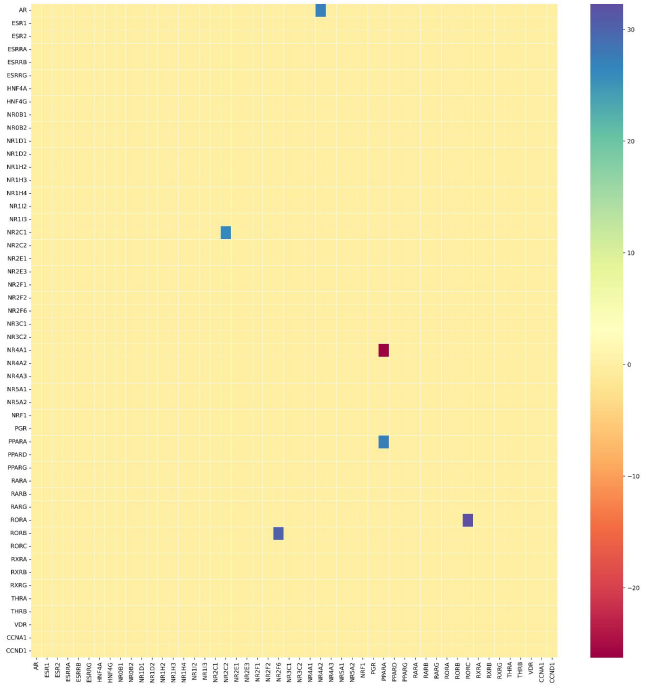
Nuclear receptor network in M0 vs M1 in brca_tcga_2012



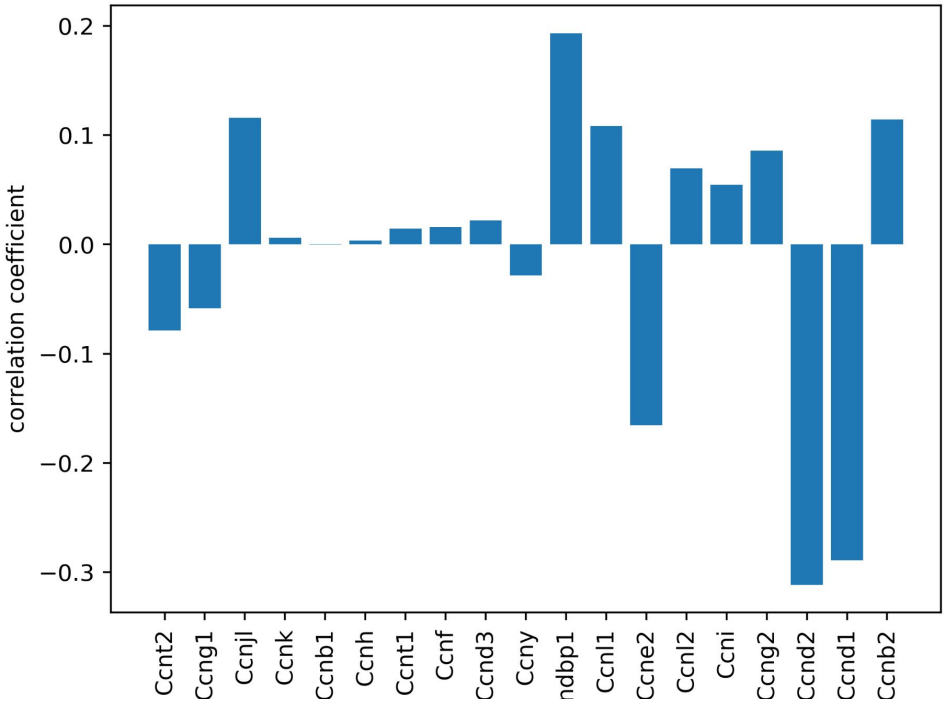
- As expected many entries are ~0.0, as most genes will not be closely influenced by many other genes
- Each matrix/heatmap entry denotes the influence of a row gene on a column gene in the underlying dynamical system.
- Cyclin expression (CCNA1 and CCND1) was used to “timestamp” cell age/time to compute the interaction matrix.
- Note: The number of M1 cases was far smaller than M0 cases.

Perturbation to the network in metastasis





Cyclins vs scVelo latent time



Questions,limitations,…

- One can use the perturbation map to find genes or epigenetic information that differentiates M0 vs M1 breast cancer. What, if any, clinical purpose can we use it for?
  - If yes, what's the next step?
  - Note that epigenetic information is not explicitly baked into the algorithm, but can be inferred from perturbation matrix entries
- The selected ~50 genes are nuclear receptors. We can use other subsets of genes. What's a good set of genes to focus on?
  - Limited computational resources in google colab prevents use of all ~20K genes.
- Is there a dataset available with more M1 cases?
  - The brca_tcga_2012 dataset has only 14 such cases.