Cytokine Surrogates from Gene Expression Data

Coinfection Lab Presentation: 07/18/24

Benjamin Whipple

Overview

- We seek to better understand differences between mock-PR8 and RV-PR8 groups.
- Cytokines such as IFN-alpha, IFN-gamma are likely important to the difference.
- We don’t have direct cytokine measurements.
- Old data collected in Dr. Miura’s lab provided gene expression data.
- It may be possible to construct surrogate of cytokines from gene expression data.
- **We construct surrogates of IFN-alpha, IFN-beta using Gaussian process models.**

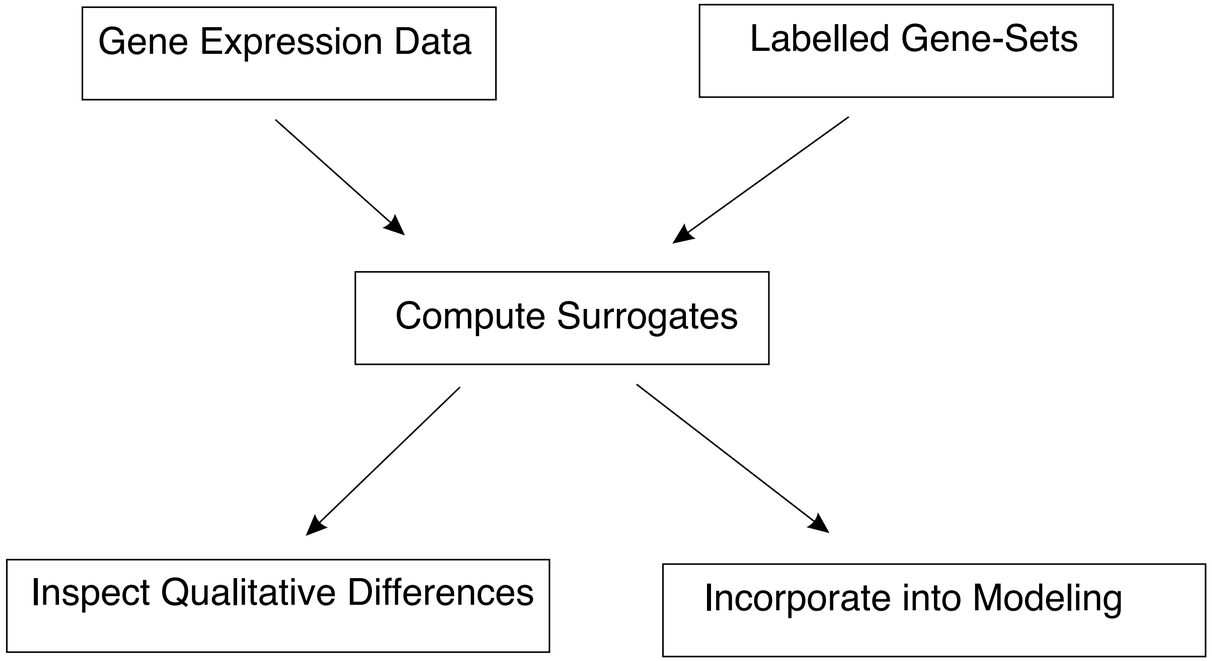
Gene Expression Data

- Collected in previous experiment (see Van Leuven 2021 paper)
- 3 replicates taken from days 0, 2, 4, 6.
- We only consider mock-PR8 and RV-PR8 groups (for now).
- Roughly 24,000 genes per replicate.
- **How do we identify which genes are related to which cytokines?**

Hallmark Gene Sets

- Hallmark gene sets of genes associated to phenomena exist (https://www.gsea-msigdb.org/gsea/msigdb/mouse/genesets.jsp?collection=MH)
- These include IFN-alpha, IFN-gamma, and several others which are interesting (inflammation, IL-2, IL-6, TGF-beta, TNF-alpha).
- We focus on IFN-alpha and IFN-gamma for brevity.
- These sets are described as being genes identified to be unregulated in response to IFN (alpha/gamma) proteins.
- **We can use these curated collections as a defensible choice of genes to base a surrogate off.**

Building a Surrogate Model



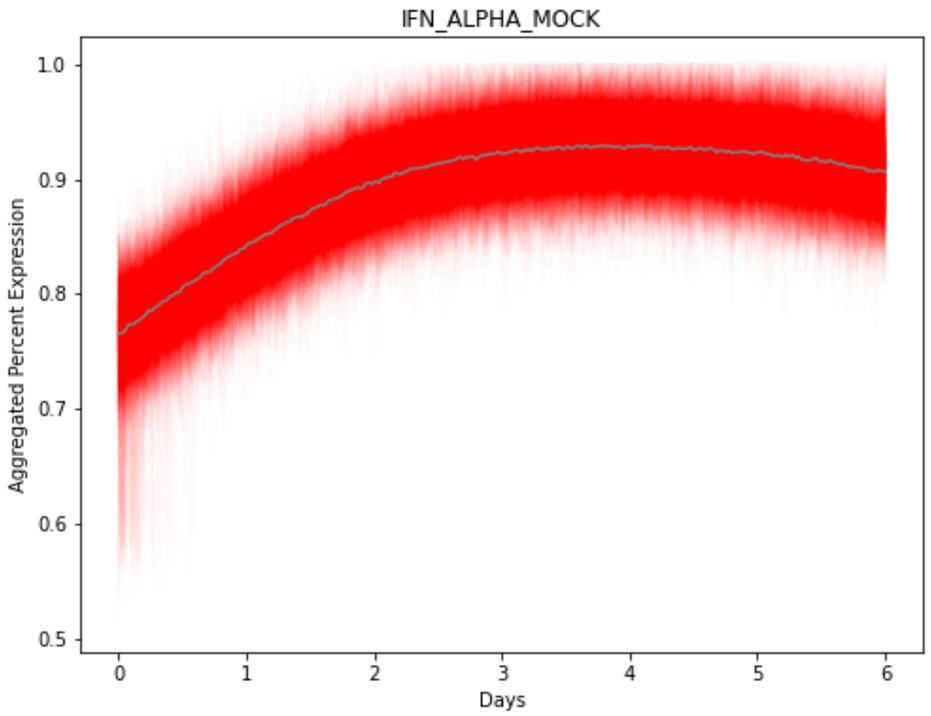
- The aim will be to establish qualitative differences.
- Gene magnitude should not matter.
- We assume that expression of a gene is monotone regarding phenomena.
- The outcome should be usable for future modeling.

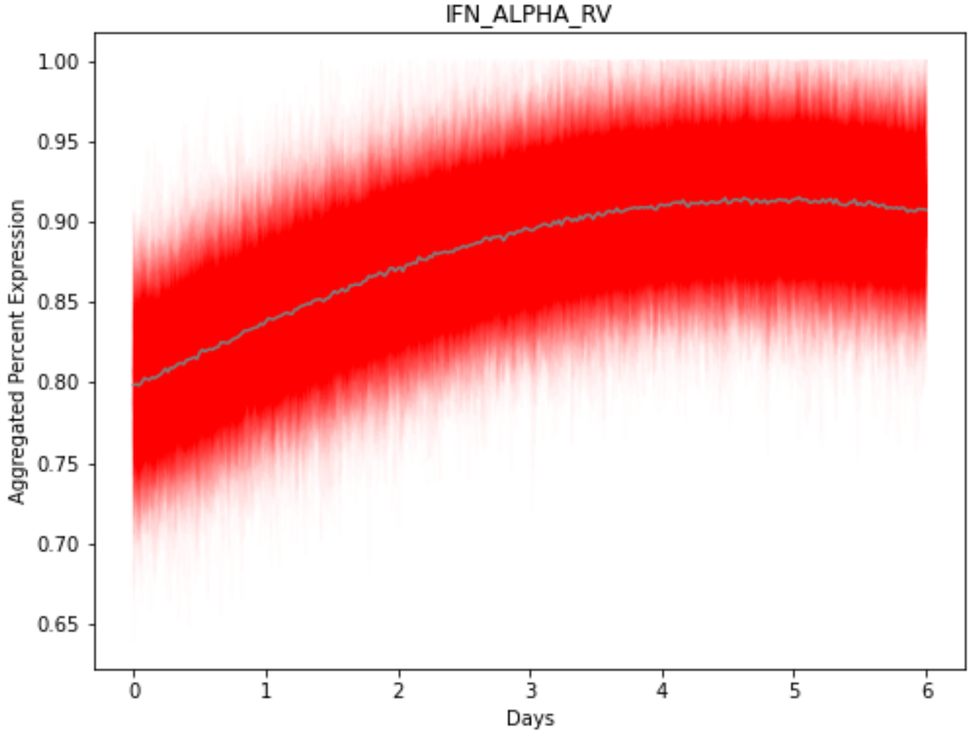
Procedure

Surrogate Model Procedure (inputs: gene set and data)

1. Compute a gaussian process surrogate for each gene
2. Generate N samples from each surrogate
3. Normalize each individual sample path
4. Aggregate across genes using geometric mean for each sample
5. Return N aggregated samples

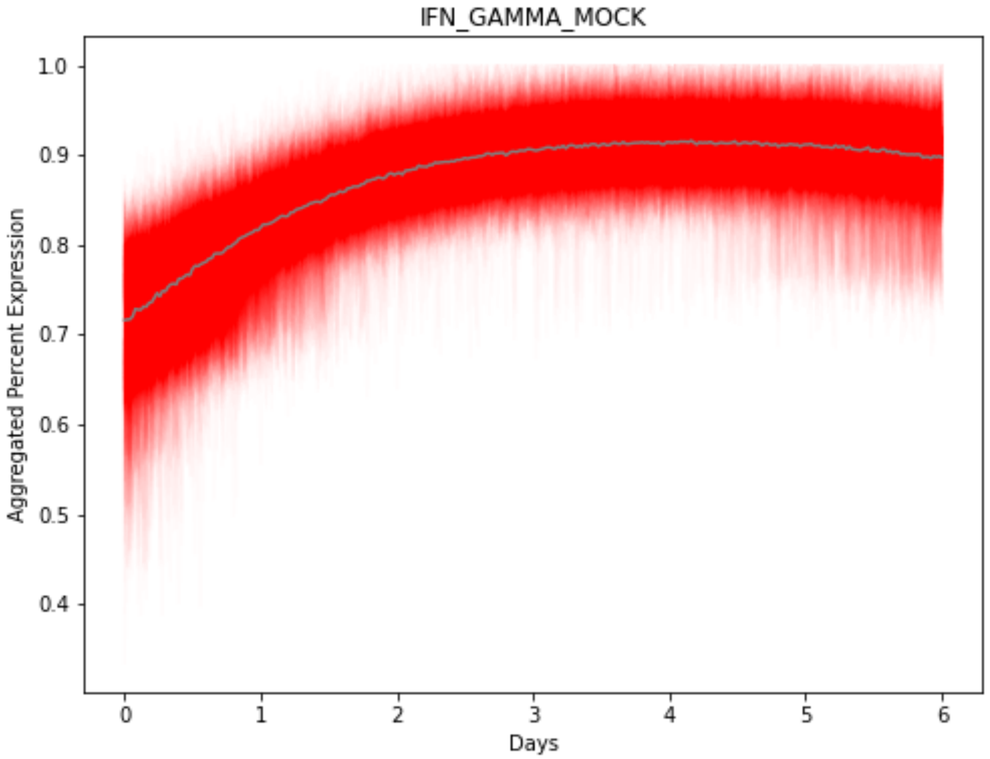
Results: IFN-alpha

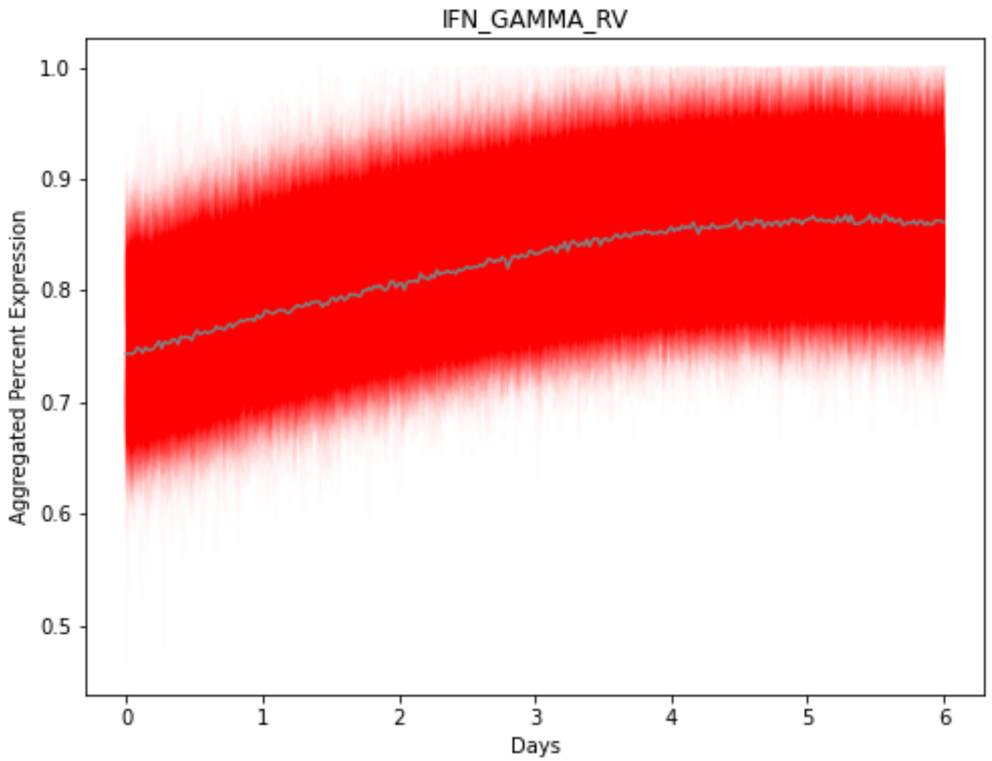




- We can see that IFN alpha begins slightly higher (0.8 vs 0.75) for the treatment group, and peaks slightly later (day 4-5 vs day 3-4).

Results: IFN-gamma





- A similar story holds for IFN-gamma.

However, or “The Trouble With Genes”

- Genes should not be assumed to have a simple relationship with a phenomena.
  - We have collections of up-regulated genes, but do not consider downrelated or suppressing genes.
  - Magnitude of gene expression need not map to magnitude of phenomenon.
- As such, all our work moving forward needs to be viewed as highly approximate.