TBRU Computational Update:

Shortest-Path Network Analyses

BRAINSTORMING WHAT DO WE WANT OUT OF THIS?

Overarching goal: Find mechanisms of human innate immune response involved in rejecting TB infection. Specifically genes/pathways/systems in macrophage that are most responsible for resistor phenotype.

Is there a clinical dataset or a subset of clinical data that would be worth running Shortest Paths net analysis on (probably with TB-PPI)?

Sticking with the CRISPR KD/KO targets…

BRAINSTORMING WHAT DO WE WANT OUT OF THIS?

Using TNFa/iNOS and IFNb KDs, we're going to be looking for

- TB secreted factors that interfere with host immune response
- TB secreted factors that are detected by host and facilitate immune response
- Critical genes, pathways, or systems for macrophage innate immune response

What kind of results should we expect?

How can we get the most descriptive information out of our analysis that's well grounded in the experimental data we already have?

BRAINSTORMING APPROACHES?

Most grounded in data approaches:

- Find KD and PPI hits that are in both data sets, find direct interactors, and find indirect interactors connected by only one intermediate protein
- Create shortest path subnetworks and find high traffic genes, central genes, and enriched pathways (GO, KEGG, MSigDB, IPA… Corum?)
- Track the TB genes responsible for most paths total and most traffic on genes of interest
- Break up shortest path networks by TB-secreted factor or by Crispr phenotype and enrich/analyze separately (or find shortest paths with everything in the same basket but track origin).

BRAINSTORMING APPROACHES?

Most Abstract/Generative network approaches:

- Modularize the network to come out of shortest path analysis by clustering on a DSD matrix or by finding network communities.
- Make nets using multiple methods: k-shortest and ‘for-each’, between whole gene sets and gene sets divided by phenotype or TB-interactor, based on STRING network and others (BioPlex, GIANT, ConsensusPathDB…) and then compare different networks’ similarity by taking the difference of ‘S-matrices’.
- Take another look at network propagation—try tuning different heat dispersion algorithms to find overlap using quantitative heat outputs instead of thresholded output.

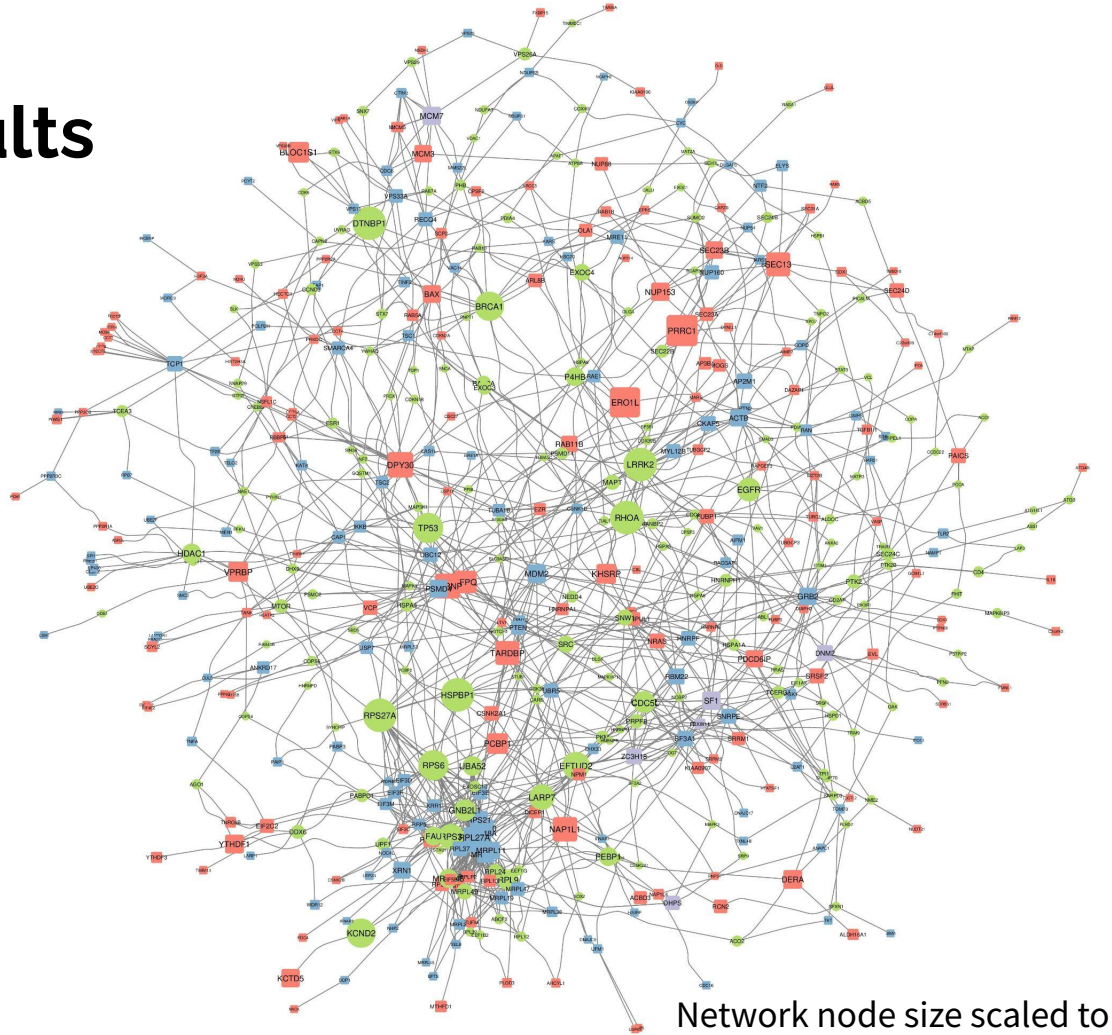
What I've been working on.

Fixed normalization techniques and pruned the network

Created/analyzed subnetworks of shortest paths from PPI -> allKD, PPI -> eachKD, and all KD -> PPI

Simple analysis of direct/indirect interactors

PPI -> all KD results



Network node size scaled to N Paths

PPI -> all KD results

New Proteins from Path Analysis involved in many paths (roughly ordered, from 25-130 paths):

- LRRK2
- RHOA
- HSBP1
- DTNBP1
- TP53
- BRCA1
- RPS27A
- DTNBP1
- RPS6/RPS3
- KCND2
- EFTUD2
- LARP7
- HDAC1

PPI -> all KD results

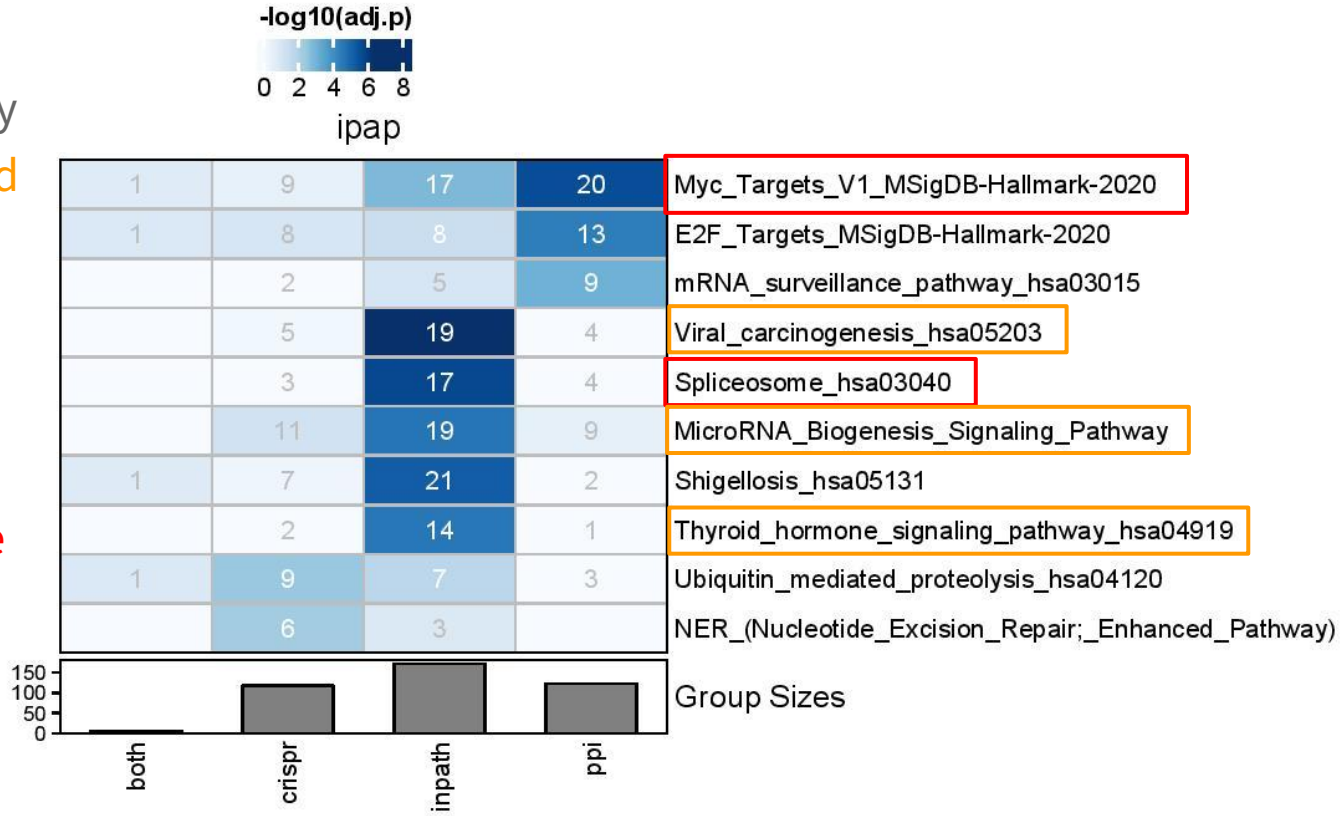
New Proteins from Path Analysis with High Normalized Betweenness (> 20) :

- LARP7
- LRRK2
- PRPF8
- MAPK3
- Honorable mentions (>10): HSPBP1, DTNBP1, STX6, PHB, RPL24, EXOSC10, RPL9, EIF1AX, PTK2B, KCND2

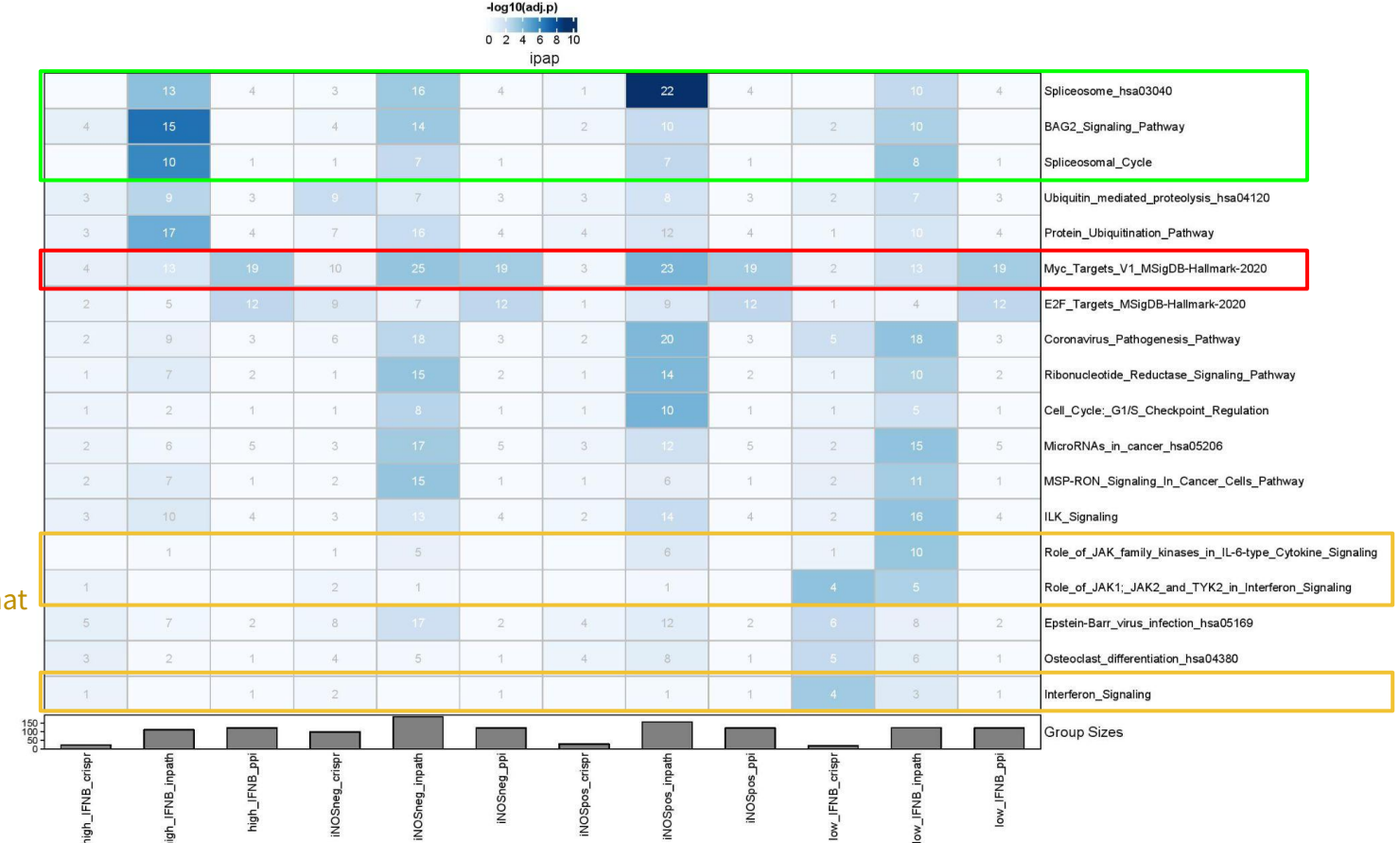
PPI -> all KD results

MAPK1, MAPK3, PIK3R1, P53, and HRAS shared by viral carc, microRNA, and Thyroid gene sets.

NCBP2, HNRNPU, SNRPD3, TRA2B, and SRSF1 shared by Myc targets and spliceosome gene sets.



PPI -> each KD results



Enriched in KD that caused low IFNb and pathways to those KD, not in anything else

All KDs -> all PPI results