Subnetworks of Disrupted Genes in Cancer

Lucas Lochovsky
Gerstein lab
Group Meeting
2012-04-10

Overview

- Produce a supernetwork that captures our most current knowledge of molecular interaction networks
  - Protein-protein interaction (PPI)
  - Regulatory networks (TFs and their binding sites/regulated genes)
  - miRNA networks (miRNA and their target transcripts)
- Map cancer genome data onto nodes of this network
  - Mutations (currently SNPs, small multi-base substitutions, and small indels)
  - Cancer/normal expression ratio
- Find subnetworks of disrupted nodes

Outline

- Network Data
- Cancer Data
- Subnetwork discovery process
- Results
- Future Directions

Outline

- Network Data
- Cancer Data
- Subnetwork discovery process
- Results
- Future Directions

Network Data

- **PPI Interactions:** HPRD Release 9 (latest release)
  - ~37,000 interactions between some ~9000 proteins
- **TFBS:** ~400,000 binding sites of 74 human TFs (same ones from GENCODE v7) based on Peakseq coordinates
  - Follow disrupted regulatory relationships
  - Assume binding site regulates gene within 1 kb of TSS
- **miRNA:** ~900 human miRNA and their targets (~20,000) from microRNA.org database

Network Data

- Genes shared by PPI and TF networks:
  - 7452
- Genes shared by PPI and miRNA networks:
  - 9458
- Genes shared by TF and miRNA networks:
  - 12,993
- Given these nontrivial intersections, this demonstrates that there is considerable “cross-talk” possible between these networks if merged

Outline

- Network Data
- Cancer Data
- Subnetwork discovery process
- Results
- Future Directions

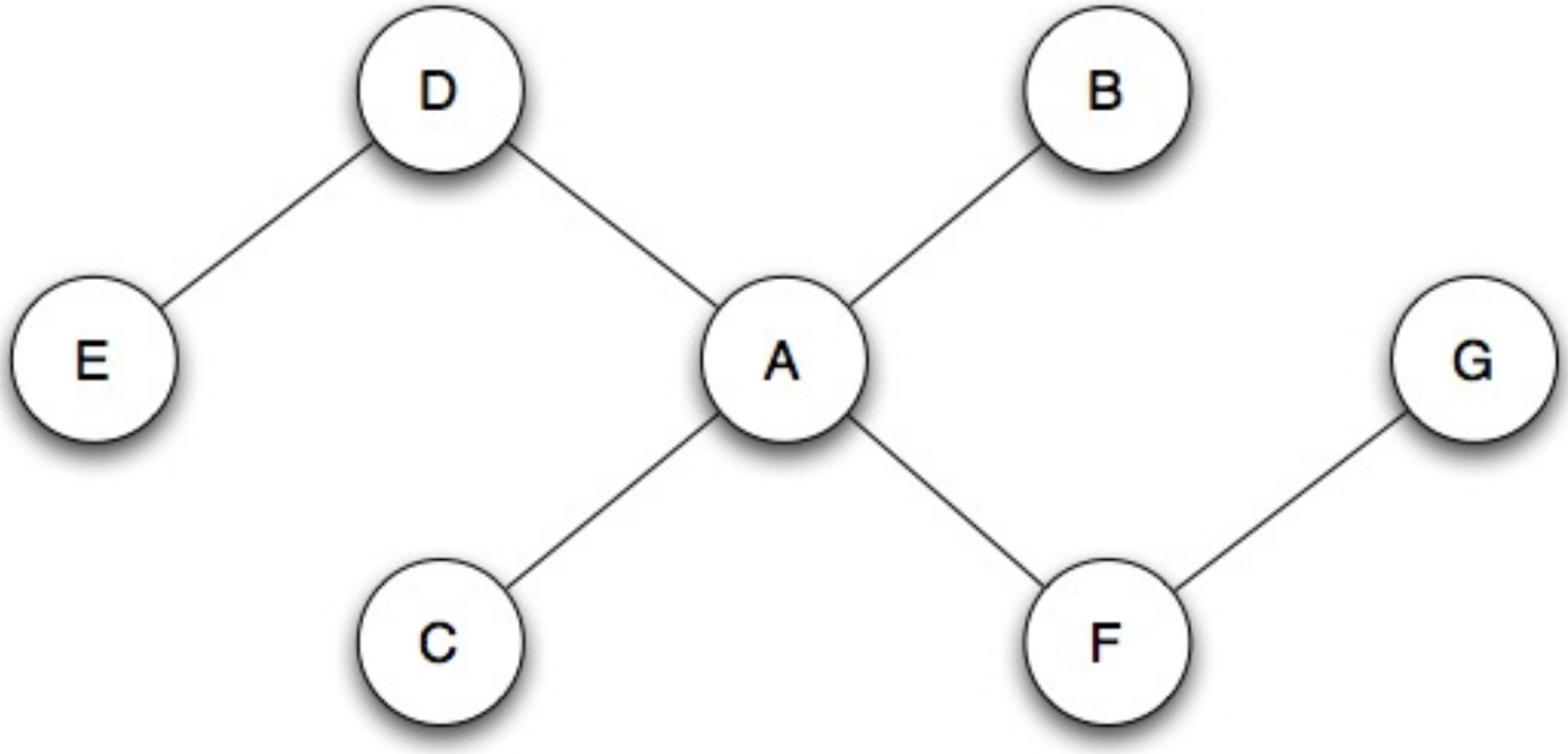
Cancer Data

- Set of 7 prostate cancer genomes from the Rubin lab
  - All have mutation data
  - Two have expression data
- Nine colorectal cancer genomes from Peter Good (published in *Nature Genetics*)
  - Whole genome mutation data
- Whole genome mutation data for various cancers from TCGA
  - 44 colon adenocarcinoma genomes (BCM; SOLiD_DNASeq)
  - 126 glioblastoma multiforme genomes (BI (Broad Institute); ABI)
  - 59 acute myeloid leukemia genomes (WUSM (Washington University School of Medicine), IlluminaGA_DNASeq)
  - 142 ovarian serous cystadenocarcinoma genomes (BI; IlluminaGA_DNASeq)
  - 35 rectum adenocarcinoma genomes (BCM; SOLiD_DNASeq)

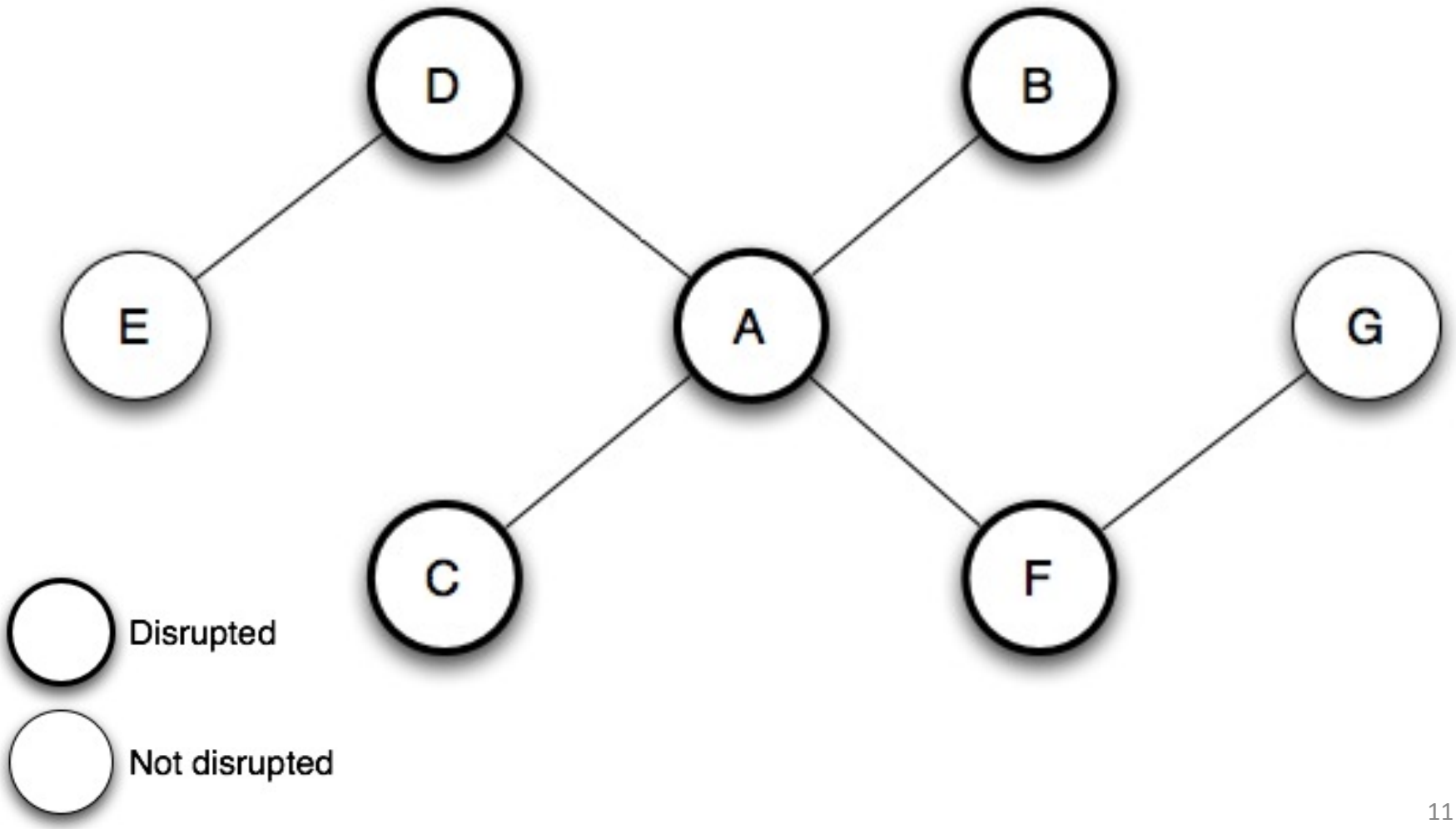
Outline

- Network Data
- Cancer Data
- Subnetwork discovery process
- Results
- Future Directions

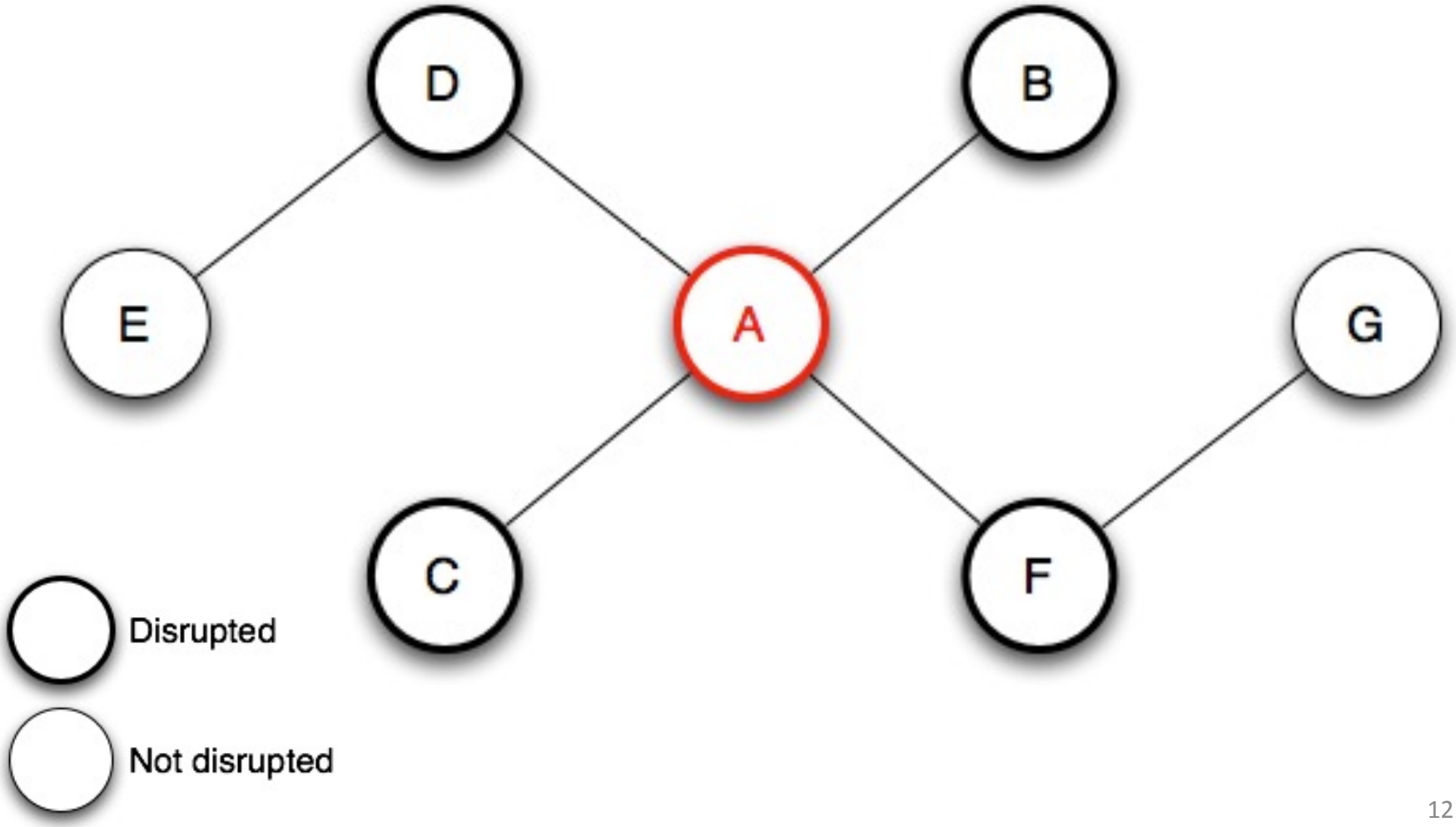
Subnetwork Discovery Process



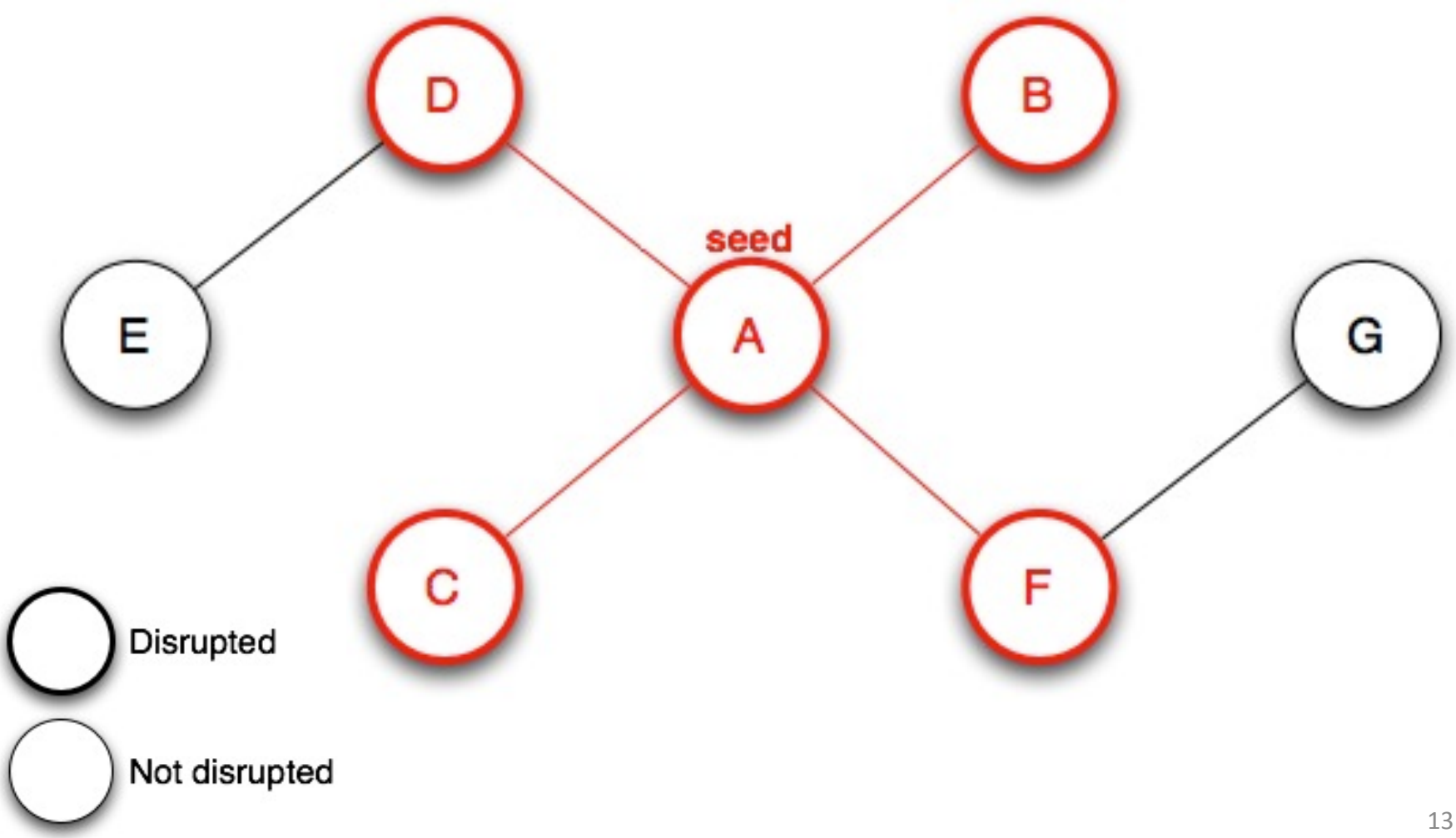
Subnetwork Discovery Process



Subnetwork Discovery Process



Subnetwork Discovery Process



Outline

- Network Data
- Cancer Data
- Subnetwork discovery process
- Results
- Future Directions

Results

- Each cancer had one dominant subnetwork
  - Each contains some fraction of the seed (OMIM) genes
  - Likely corresponds to the main disease module
  - Stricter parameters can be used to filter size of the network → pick out most important sub-subnetworks

Cancer	**Subnetwork Size**
Prostate	6868
Colorectal	10410
AML	423
COAD	733
GBM	0
OV	3192
READ	4566

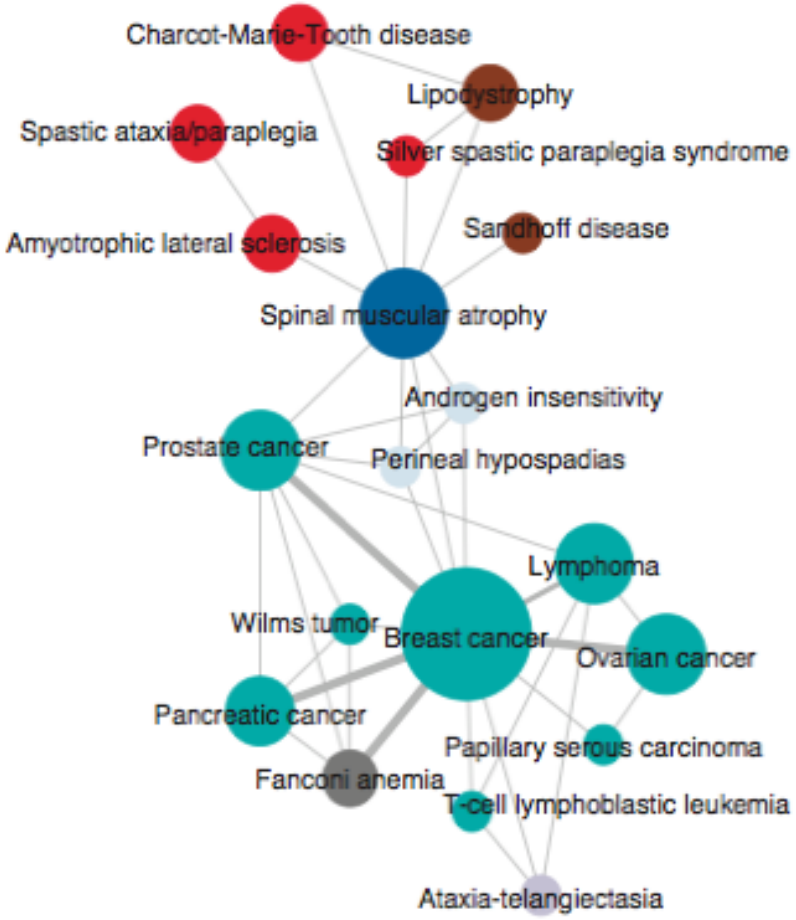
Results

Pairwise intersection matrix

	Prostate	Colorectal	AML	COAD	OV	READ
Prostate		4680	153	268	1081	1541
Colorectal			214	392	1492	2245
AML				27	78	109
COAD					183	287
OV						862
READ						

- Each subnetwork occupies a unique space in the entire network
- However, a small fraction of each disease subnetwork is shared with at least one other subnetwork → potential for development of one therapy that affects multiple diseases?

Human Diseasome Network

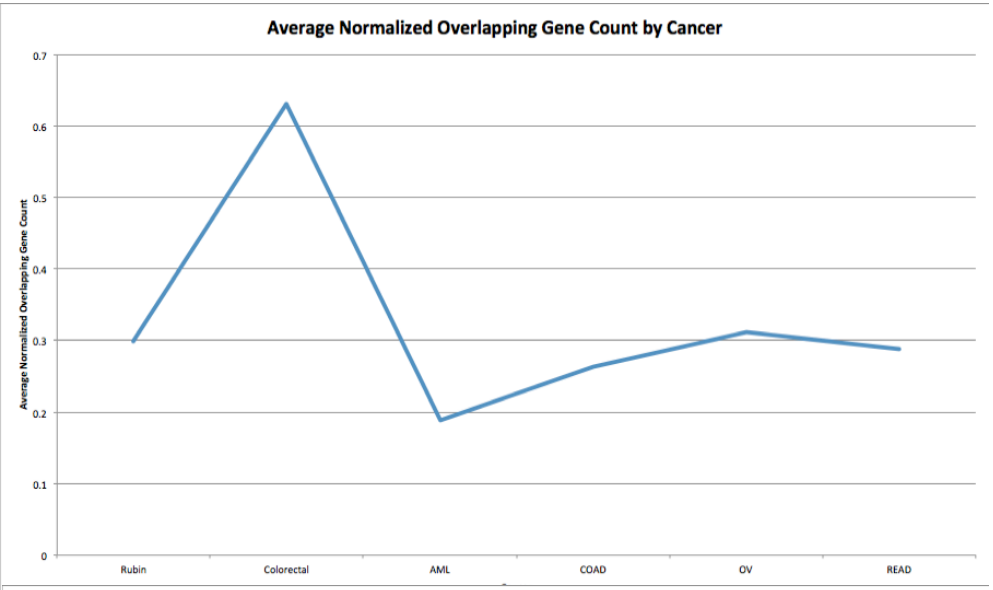


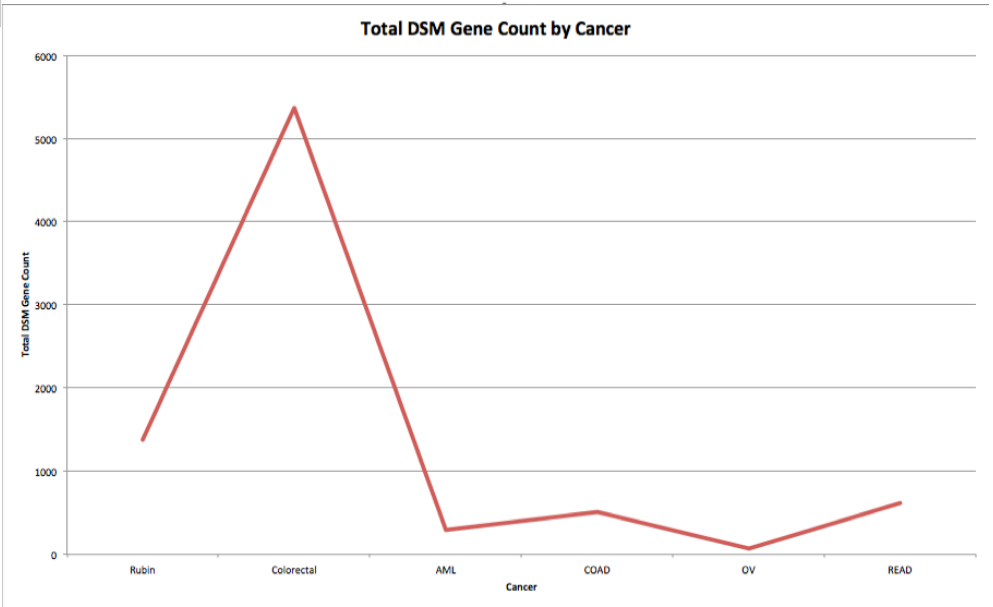
Goh *et al*. "The human disease network." *PNAS May 22, 2007 vol. 104 no. 21 8685-8690*

Intersection Results

- Subnetwork first-neighbor miRNA typically included at least 2/3 of all known human miRNA, if not all
  - Not very informative
- Switched focus to subnetwork member genes
  - Pairwise intersection sizes were proportional to the # of subnetwork genes
  - Found very strong correlation between the average normalized overlapping gene count and subnetwork gene count

Intersection Results



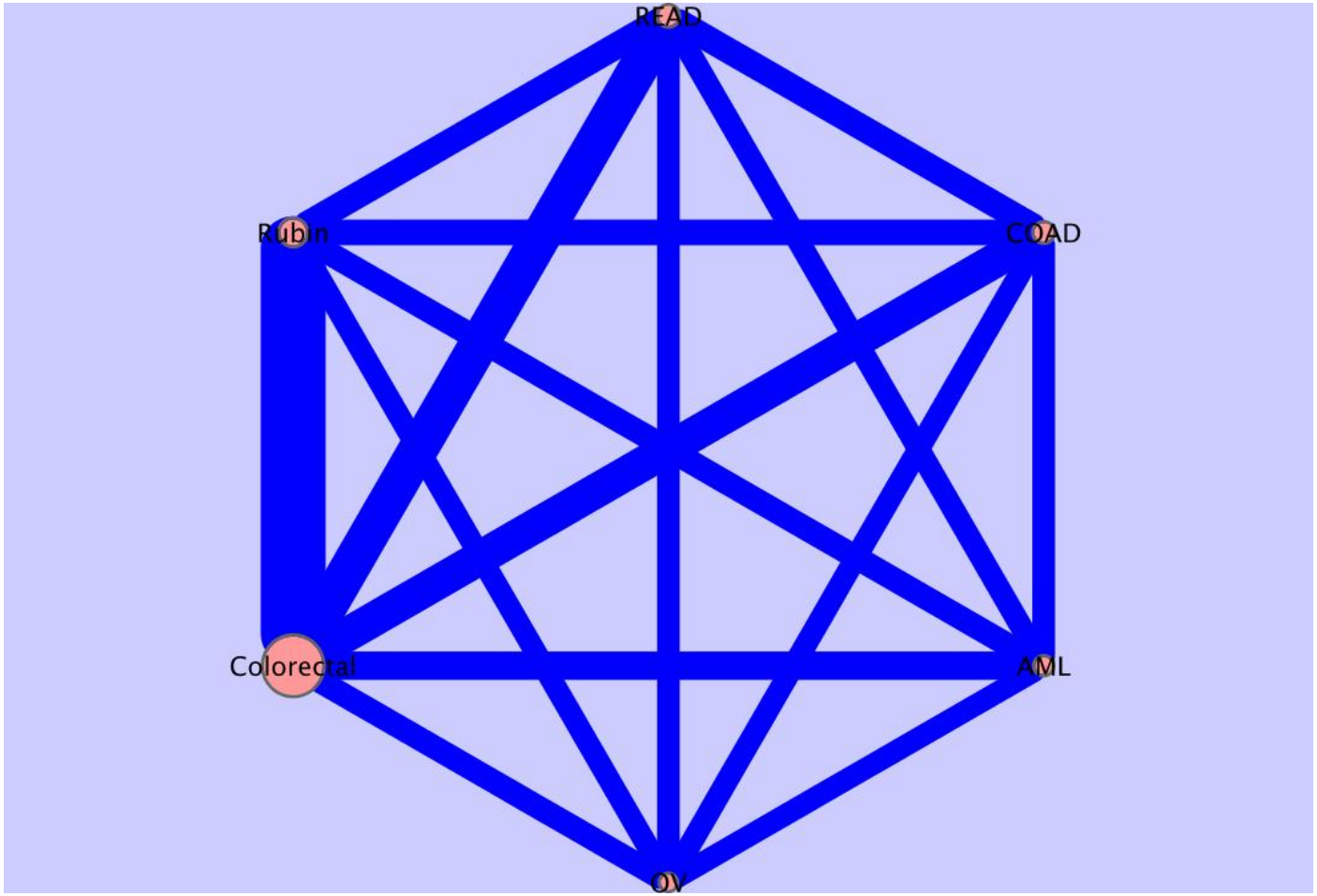


Top graph:
Average normalized overlapping gene count by cancer
(for each cancer, start with the number of genes in each pairwise intersection, divide by total gene count, and find arithmetic mean)

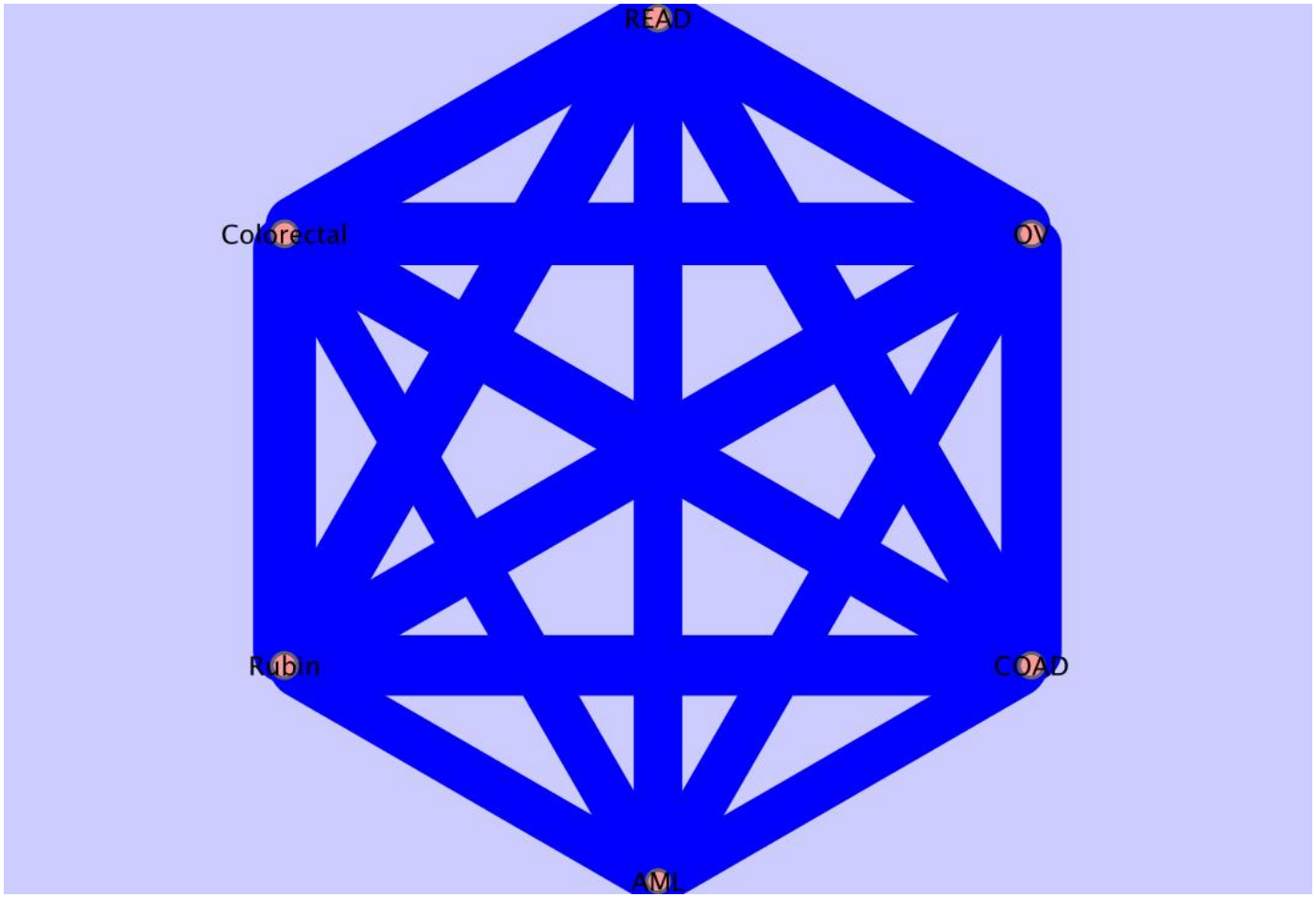
Bottom graph:
Total subnetwork gene count by cancer

Correlation: 0.952

Diseasome Network: Common Genes



Diseasome Network: Common miRNA



Summary of DSM miRNA

Shared by 3 cancer subnetworks

- hsa-miR-1267 – Colorectal, AML, prostate

Shared by 2 cancer subnetworks

- hsa-miR-1302 – Colorectal, prostate
- hsa-miR-548f/h/q – Colorectal, prostate

Singletons

- hsa-miR-1185 - colorectal
- hsa-miR-142-3p - AML
- hsa-miR-142-5p - AML
- hsa-miR-1973 - colorectal
- hsa-miR-320d - colorectal
- hsa-miR-548a-3p - prostate
- hsa-miR-891a - AML
- hsa-miR-921 - colorectal

miRNA Functional Characterization

- Looked in miRNA databases to see what was already known about these miRNA
- Found miR-Ontology
  - Data warehouse developed at the Univ of Cantania in Italy
  - Integrates all sorts of data on miRNA, including sequence features, expression data, experimental and predicted targets
  - Includes disease associations --> targets' disease annotations are applied to parent miRNA

miRNA Functional Characterization

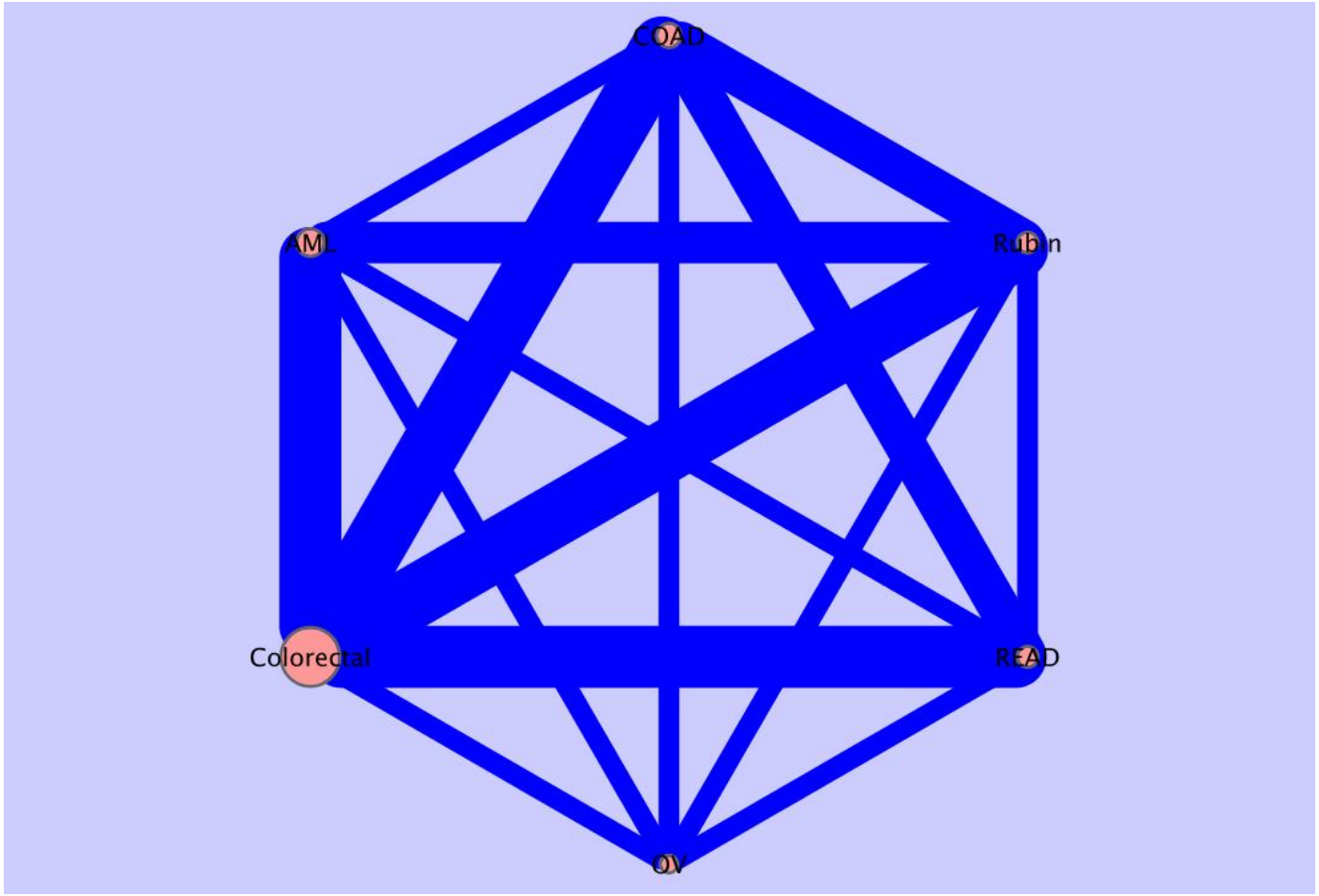
- miR-Ontology confirmed the disease associations of almost all miRNA discovered through DSMs
- With 3 exceptions:
  - hsa-miR-1973 in colorectal cancer
  - hsa-miR-548q in colorectal and prostate cancer
  - hsa-miR-891a in AML
- Hence, we can conclude that many of the miRNAs implicated in DSMs have genuine associations with their respective diseases
- And there is the potential for DSMs to implicate miRNA not previously implicated

miRNA targets: Cancer subnetworks vs. miR-Ontology

miRNA	Cancer Type	DSM #	miR-Ontology #	Intersection #
hsa-miR-1185	Colorectal	756	4	1
hsa-miR-1267	AML	64	1	0
hsa-miR-1267	Colorectal	1145	6	4
hsa-miR-1267	Prostate	338	7	0
hsa-miR-1302	Colorectal	854	13	3
hsa-miR-1302	Prostate	234	10	2
hsa-miR-142-3p	AML	50	1	0
hsa-miR-142-5p	AML	128	2	1
hsa-miR-1973	Colorectal	144	0	0
hsa-miR-320d	Colorectal	1567	20	5
hsa-miR-548-3p	Prostate	639	16	0
hsa-miR-548f	Colorectal	2322	20	9
hsa-miR-548f	Prostate	624	18	0
hsa-miR-548h	Colorectal	2302	28	14
hsa-miR-548h	Prostate	663	12	0
hsa-miR-548q	Colorectal	779	0	0
hsa-miR-548q	Prostate	223	0	0
hsa-miR-891a	AML	14	0	0
hsa-miR-921	Colorectal	859	15	2

- For each miRNA found in a cancer subnetwork, find how many targets appear in its subnetwork and how many appear in miR-Ontology (also find how many in the intersection set)
- Subnetworks have far more targets than miR-Ontology
  - Did the subnetwork algorithm find many targets not previously implicated in disease?
- Also have many targets in miR-Ontology not in subnetworks
  - Poor sensitivity?

Diseasome Network: Common TFs



DSM transcription factors

TF	# of Cancers	Cancers
Elf1	3	Prostate, colorectal, AML
Taf1	3	Colorectal, COAD, READ
Zeb1	3	Prostate, colorectal, COAD
Max	2	Colorectal, AML
Tcf12	2	Colorectal, READ
Bclaf1	1	Colorectal
Brca1	1	OV
Ets1	1	Colorectal
Nrf1	1	Colorectal
Pbx3	1	Colorectal
Smc3	1	AML
Stat1	1	Colorectal
Stat3	1	Colorectal
Tbp	1	Colorectal
Zbtb33	1	COAD
Znf143	1	Colorectal
Zzz3	1	Colorectal

- List of all transcription factors that appear in DSMs
- Ordered by # of cancers involved
- Indicates shared and unique components of cancer regulatory disruptions
- Most cancers appear to affect a unique part of the network (lots of singletons)

Outline

- Network Data
- Cancer Data
- Subnetwork discovery process
- Results
- Future Directions

Future Directions

- Functionally characterize the intersecting components of DSMs
- Map additional cancer data to the supernetwork
  - 118 prostate adenocarcinoma exomes from Rubin lab
- Investigate correspondence of DSM components to pathways and complexes
- Subsample cancer genome data to correct for representation bias

Acknowledgements

- The acknowledgement wiki:
  - http://goo.gl/E844W