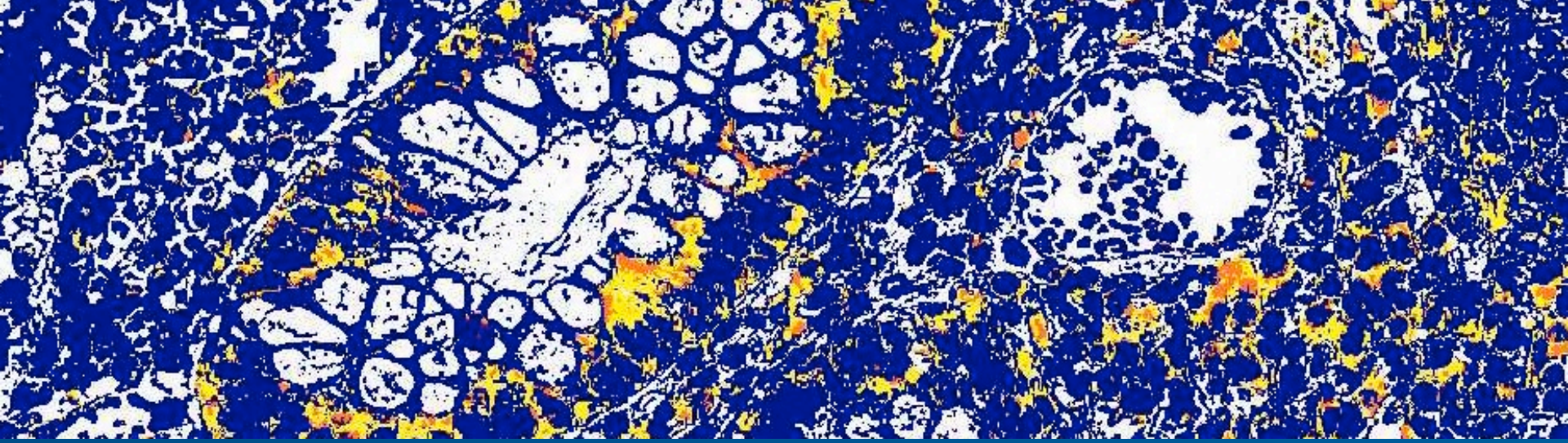


Gaining Insights into lncRNA Function in Inflammatory Bowel Disease (IBD)

Network Analysis via a Shiny Dashboard

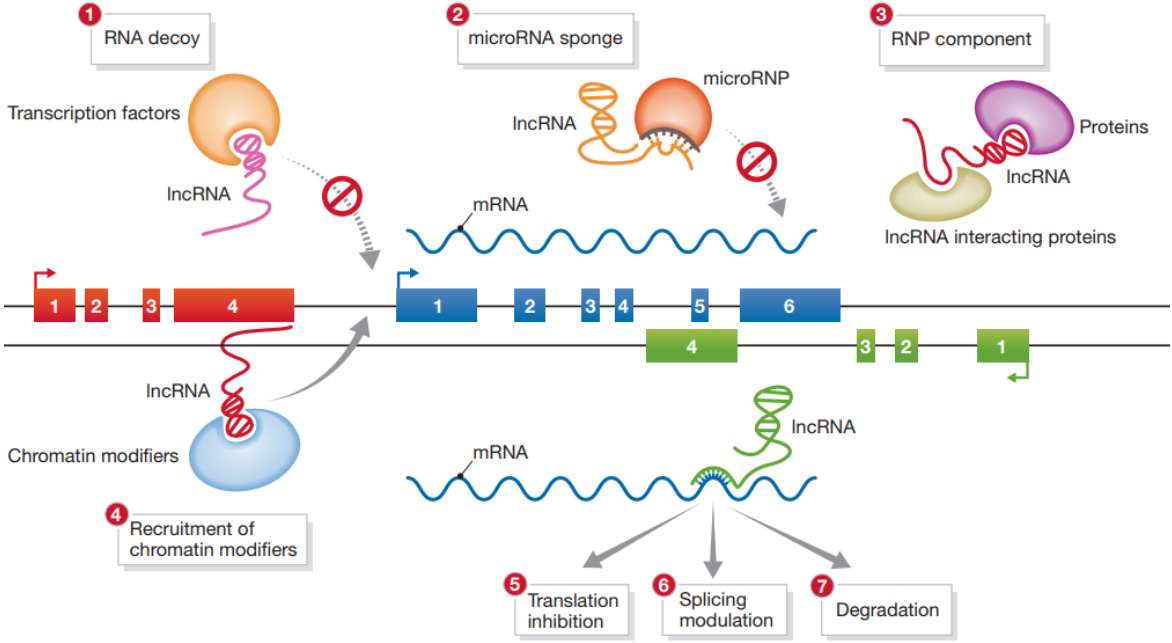
Pictured above: Ulcerative Colitis

Stefan Avey, PhD, Senior Statistician in Discovery Statistics
June 18, 2019

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

Long Non-coding RNAs (lncRNAs) have diverse structures, function, and expression patterns

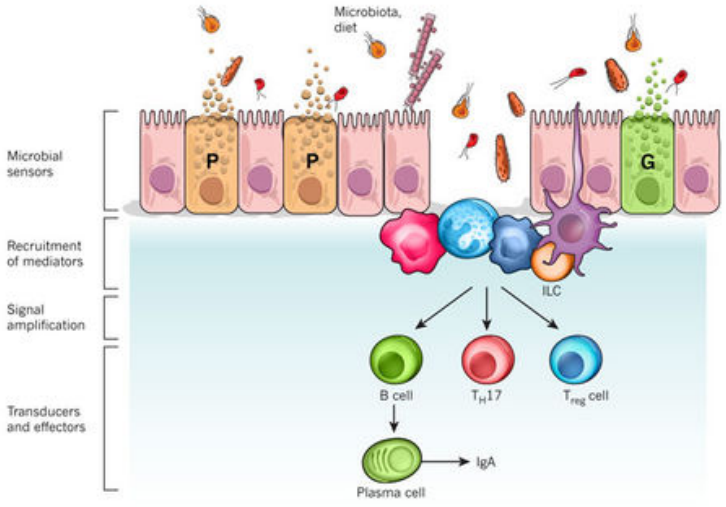
- Arbitrarily ≥200nt in length and lowly expressed
- Bind RNA, DNA, and protein
- ~20,000 protein coding genes | ~60,000 lncRNA genes
- Transcribed and processed like mRNAs for protein coding genes
- Expressed in a temporal, cell type, and tissue-specific manner



lncRNAs regulate fundamental biological processes through highly tissue and cell-specific regulation of gene expression, making them attractive for targeted therapy

Image: Regulation of mammalian cell differentiation by long non-coding RNAs
Wenqian Hu, Juan R Alvarez-Dominguez, Harvey F Lodish
DOI 10.1038/embor.2012.145| Published online 16.10.2012EMBO reports (2012) 13, 971-983

Genome-wide association studies have identified over 200 loci associated with IBD, majority are in non-coding regions



IBD-related processes

Epithelial barrier
GNA12*, HNF4A, CDH1, ERRFI1, MUC19, ITLN1*

Restitution
REL, PTGER4, NKX2-3, STAT3, ERRFI1, HNF4A, PLA2G2A/E

Solute transport
SLC9A4, SLC22A5, SLC22A4*, AQP12A/B, SLC9A3, SLC26A3

Paneth cells
ITLN1*, NOD2*, ATG16L1*, XBP1*

Innate mucosal defence
NOD2*, ITLN1*, CARD9*, REL, SLC11A1, FCGR2A*/B

Immune cell recruitment
CCL11/CCL2/CCL7/CCL8, CCR6, IL8RA/IL8RB, MST1*

Antigen presentation
ERAP2*, LNPEP, DENND1B

IL-23/T_H17
IL23R*, JAK2, TYK2*, STAT3, ICOSLG, IL21, TNFSF15*

T-cell regulation
NDFIP1, TNFSF8, TAGAP, IL2, IL2RA, TNFRSF9, PIM3, IL7R*, IL12B, IL23R*, PRDM1, ICOSLG, TNFSF8, IFNG, IL21

B-cell regulation
IL5, IKZF1, BACH2, IL7R*, IRF5

Immune tolerance
IL10, IL27*, SBNO2, CREM, IL1R1/IL1R2, NOD2*

Cellular responses

Autophagy
ATG16L1*, IRGM, NOD2*, LRRK2, CUL2, PARK7, DAP

ER stress
CPEB4, ORMDL3, SERINC3, XBP1*

Intracellular logistics
VAMP3, KIF21B, TTLL8, FGFR1OP, CEP72, TPPP

Cell migration
ARPC2, LSP1, AAMP

Apoptosis/necroptosis
FASLG, THADA*, DAP, PUS10, MST1*

Carbohydrate metabolism
GCKR*, SLC2A4RG

Oxidative stress
PRDX5, BACH2, ADO, GPX4, GPX1*, SLC22A4, LRRK2, NOD2*, CARD9*, HSPA6, DLD, PARK7, UTS2*, PEX13

UC
CD
UC/CD
cis-eQTL
*Coding mutation

Majority of disease-associated SNPs (>80%) are in noncoding regions, causative SNPs are elusive

Do risk alleles in lncRNAs contribute to dysregulated inflammatory responses in IBD?

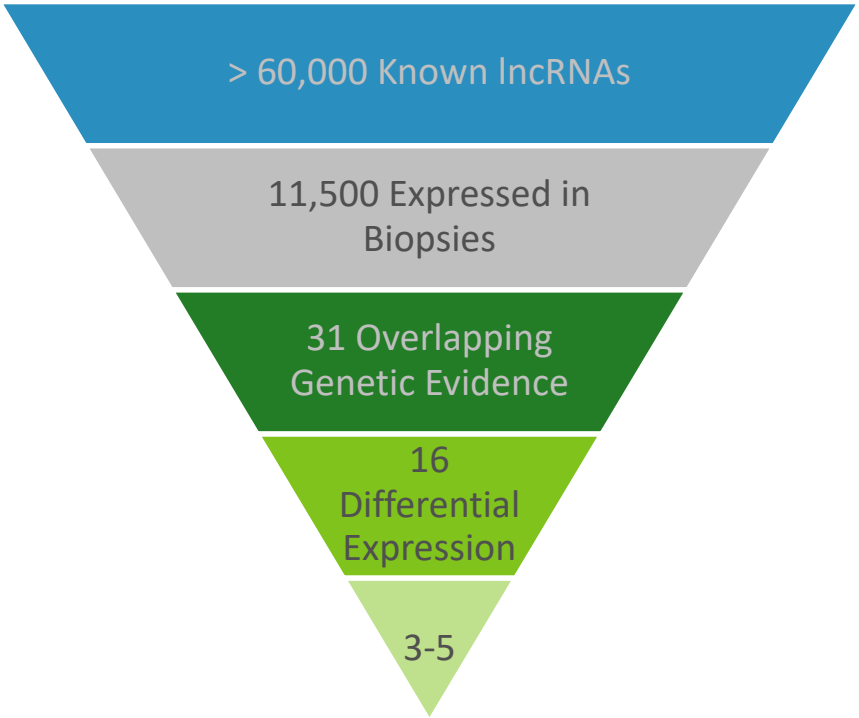
SNP = Single Nucleotide Polymorphism
Khor et al, *Nature* 2011; Liu et al., *Nat. Genetics* 2015; Jostins et al., *Nature* 2012

How to begin understanding the role of many uncharacterized lncRNAs in IBD?

General Discovery Approach:

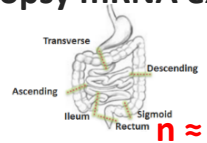
1. Prioritize lncRNAs that may be involved in IBD
2. Correlate lncRNA expression with coding gene modules
3. Annotate coding gene modules with biological databases (gene set enrichment)
4. Functional validation of lncRNA expression (e.g., cell type specific) or function

lncRNAs Prioritized in IBD by Leveraging Genetic and Transcriptional Data



Mount Sinai School of Medicine collaboration

Biopsy mRNA expression

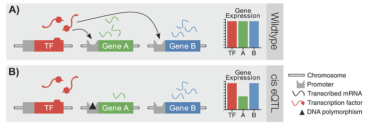


n ≈ 2,500 samples

- Control
- UC
- CD

n ≈ 1,200 subjects

Genetics



Gene Modules Generated from Co-Expression Network Analysis

- Multiscale Embedded Gene Co-expression Network Analysis (MEGENA)

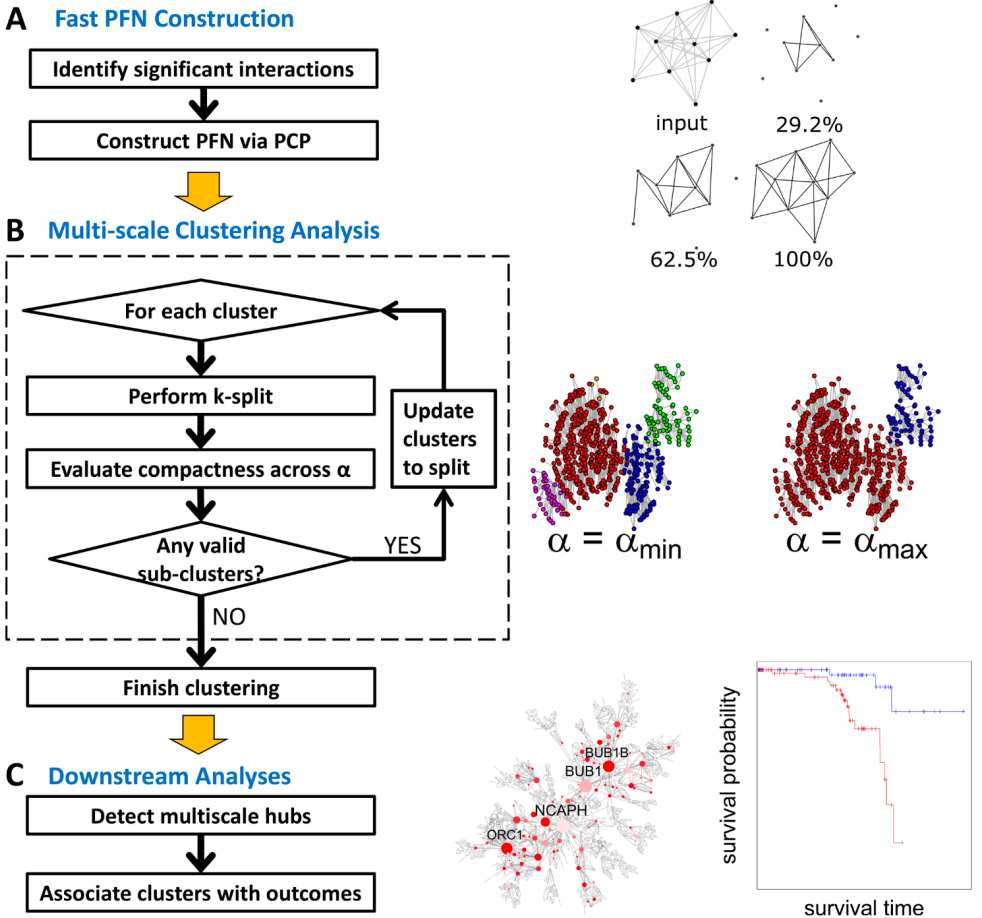
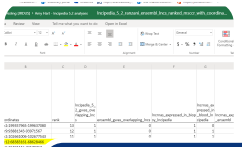


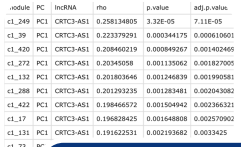
Fig 1. Flow chart of MEGENA. A) Fast planar filtered network construction. Significant interactions are first identified and then embedded on topological

Discovery Workflow

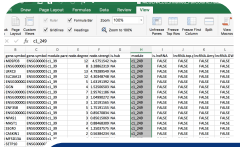
Previous Workflow was Time Consuming and Error-Prone



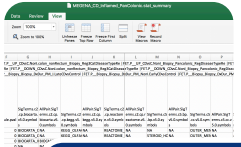
Query lncRNA in ranked list



Obtain top correlated MEGENA modules

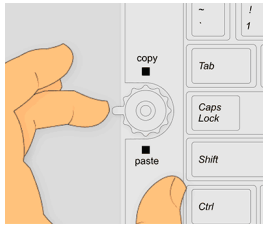


Extract module gene lists from module annotation files



Extract pre-computed gene set enrichments from summary files

Query Internal Gene Set Enrichment Tool (Trial Data, Cell Barcodes)



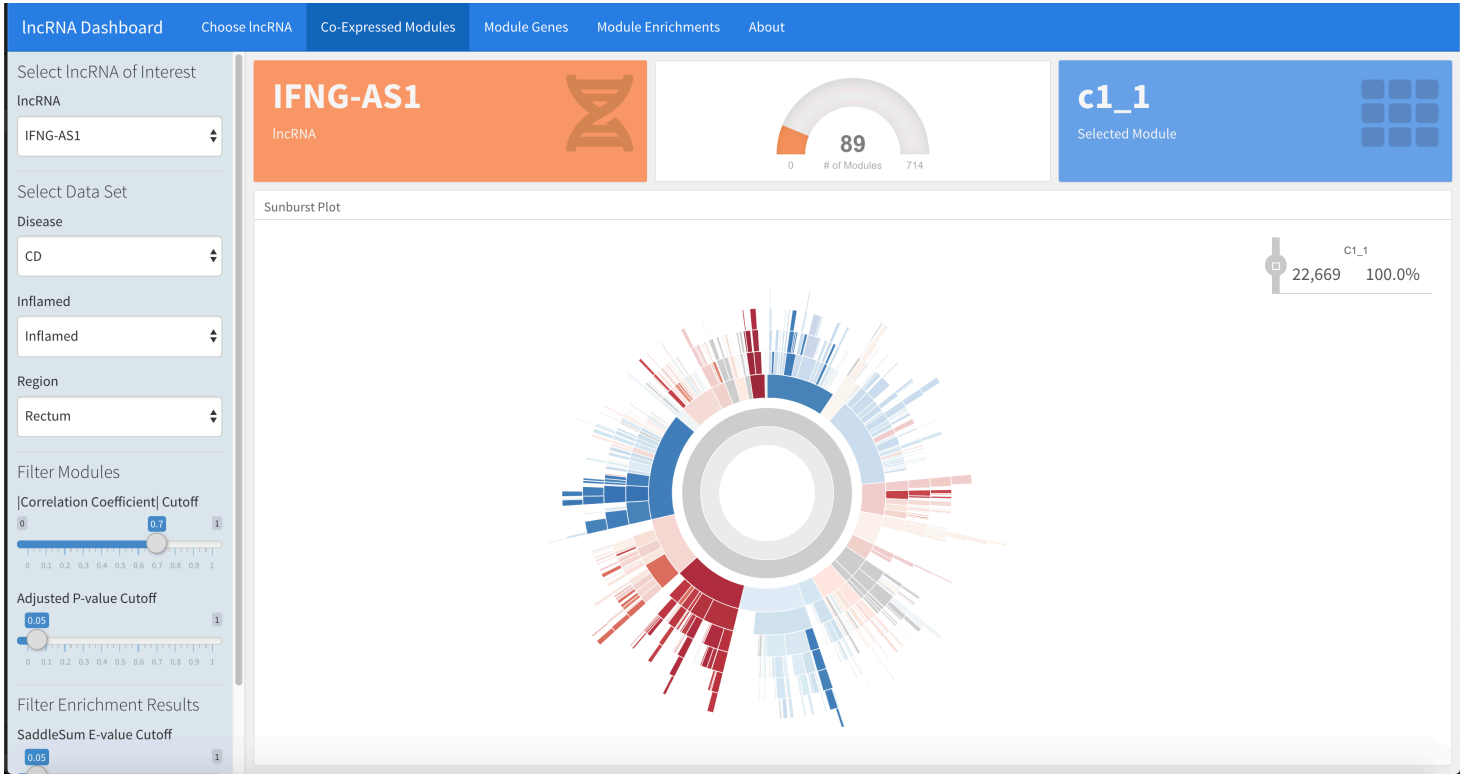
Summary file (one for each lncRNA, for each disease, for each biopsy region, for each tissue type)

Flexdashboard Shiny App

lncRNA Dashboard

Visually Inspect Expression Data to Choose Dataset of Interest for a given lncRNA

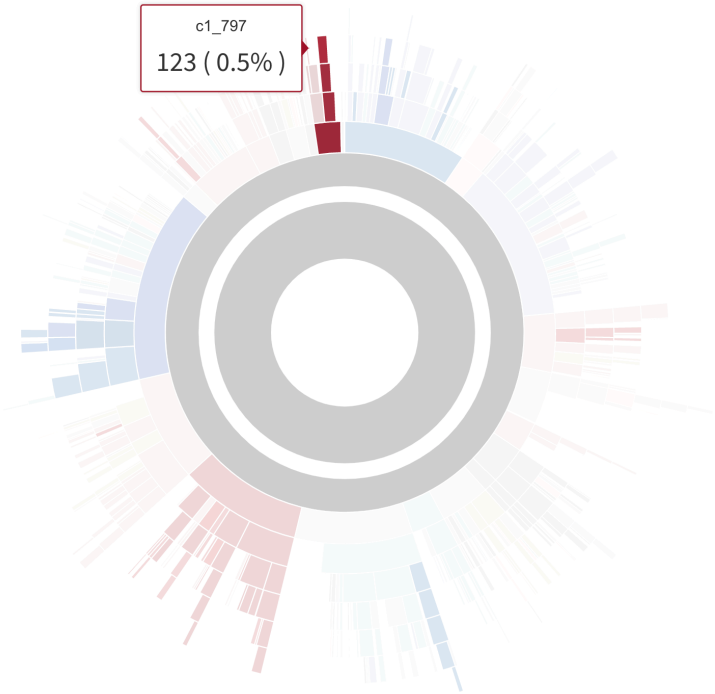
Sunburst Plots are Visual Representation of Co-Expression with Coding Gene Modules



Sunburst Plots are Visual Representation of Co-Expression with Coding Gene Modules

- Represents modules that are co-expressed with your gene of interest
- Red is positively correlated, blue is negatively correlated
- The darker the color the higher the correlation coefficient
- Transparent modules do not meet user-defined cutoffs on correlation coefficient and FDR
- Modules form hierarchical clusters so modules further out are nested within larger modules closer to the center of the circle

Modules are Interactively Selected



ROOT
22,669 100.0%

C1_1
22,669 100.0%

C1_7
465 2.1%

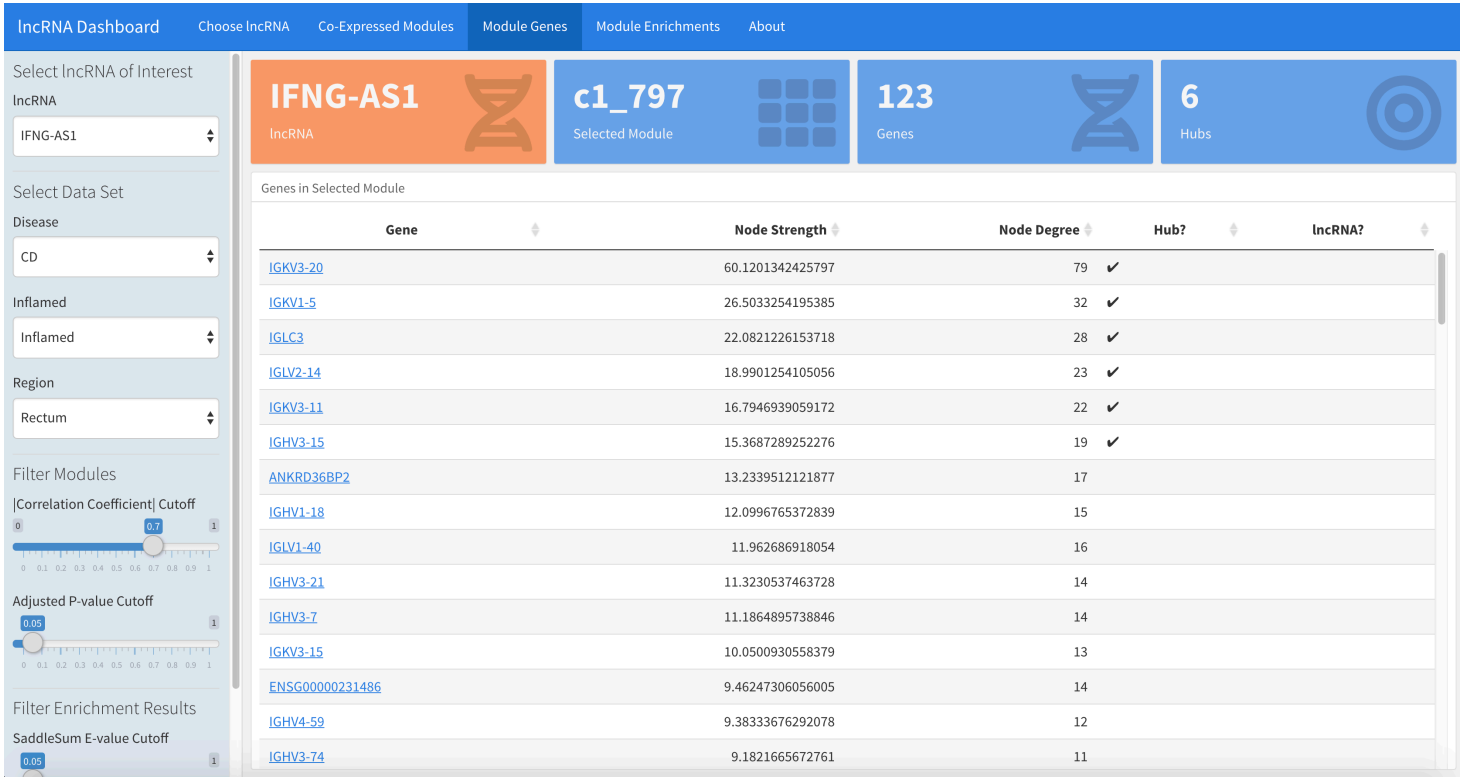
C1_51
192 0.8%

C1_306
146 0.6%

C1_797
123 0.5%

sunburstR: https://github.com/timelyportfolio/sunburstR

Module Genes are Listed for Selected Module



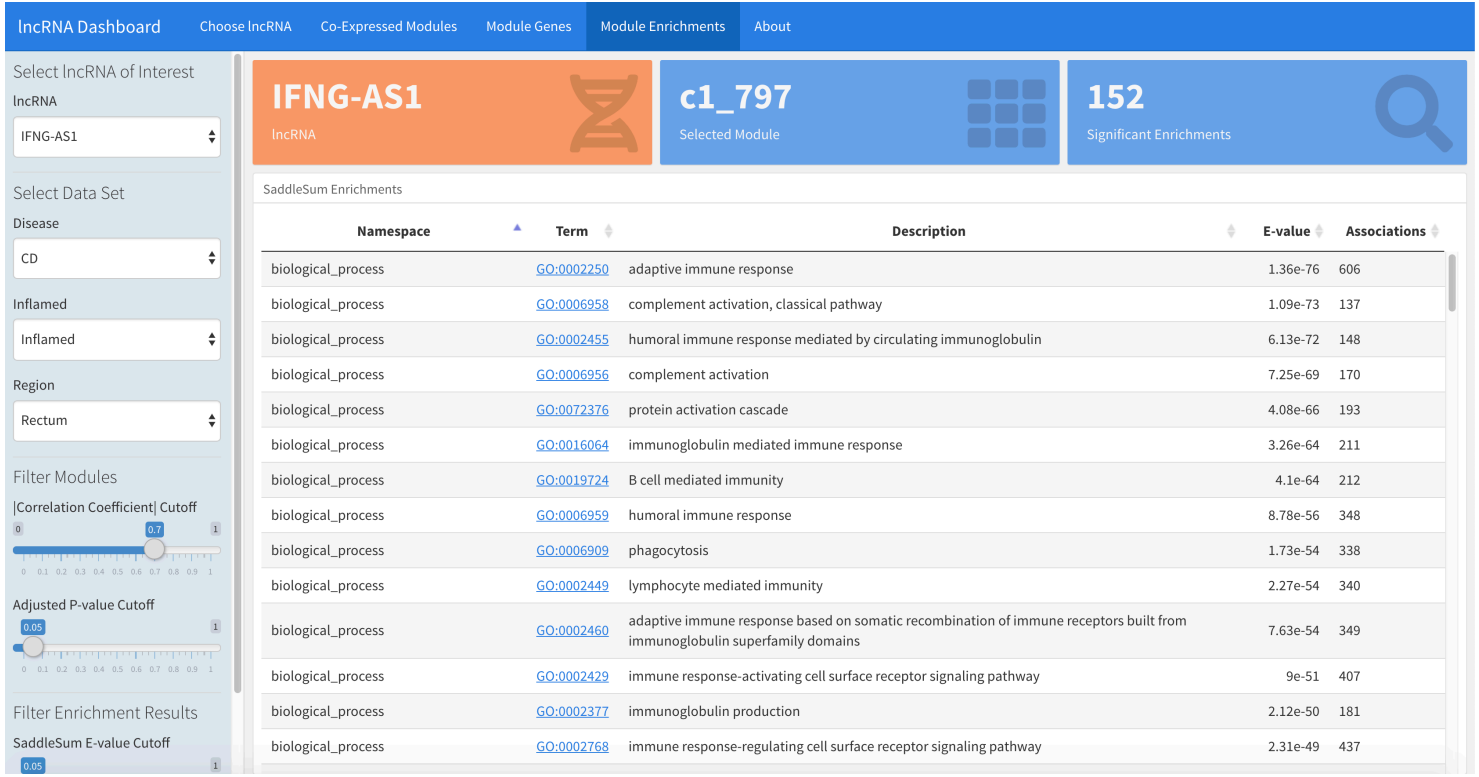
Gene	Node Strength	Node Degree	Hub?	lncRNA?
IGKV3-20	60.1201342425797	79	✔	
IGKV1-5	26.5033254195385	32	✔	
IGLC3	22.0821226153718	28	✔	
IGLV2-14	18.9901254105056	23	✔	
IGKV3-11	16.7946939059172	22	✔	
IGHV3-15	15.3687289252276	19	✔	
ANKRD36BP2	13.2339512121877	17		
IGHV1-18	12.0996765372839	15		
IGLV1-40	11.962686918054	16		
IGHV3-21	11.3230537463728	14		
IGHV3-7	11.1864895738846	14		
IGKV3-15	10.0500930558379	13		
ENSG00000231486	9.46247306056005	14		
IGHV4-59	9.38333676292078	12		
IGHV3-74	9.1821665672761	11		

Gene Module Enrichment Calculated by SaddleSum (Internal + External Gene Sets)

- Genes in each cluster have weights indicating strength of association with 1st eigengene.
- Let *m* = # of genes mapping to a term
- SaddleSum calculates a p-value by using the saddlepoint approximation to the empirical distribution function derived from all weights.
  - "How likely is it to randomly randomly pick *m* genes whose sum of weights exceed observed weights?"

Stojmirović A, Yu YK. Robust and accurate data enrichment statistics via distribution function of sum of weights. Bioinformatics 2010; 26: 2752–275. https://doi.org/10.1093%2Fbioinformatics%2Fbtq511

IFNG-AS1 module is implicated in adaptive immunity and other immune processes



Namespace	Term	Description	E-value	Associations
biological_process	GO:0002250	adaptive immune response	1.36e-76	606
biological_process	GO:0006958	complement activation, classical pathway	1.09e-73	137
biological_process	GO:0002455	humoral immune response mediated by circulating immunoglobulin	6.13e-72	148
biological_process	GO:0006956	complement activation	7.25e-69	170
biological_process	GO:0072376	protein activation cascade	4.08e-66	193
biological_process	GO:0016064	immunoglobulin mediated immune response	3.26e-64	211
biological_process	GO:0019724	B cell mediated immunity	4.1e-64	212
biological_process	GO:0006959	humoral immune response	8.78e-56	348
biological_process	GO:0006909	phagocytosis	1.73e-54	338
biological_process	GO:0002449	lymphocyte mediated immunity	2.27e-54	340
biological_process	GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	7.63e-54	349
biological_process	GO:0002429	immune response-activating cell surface receptor signaling pathway	9e-51	407
biological_process	GO:0002377	immunoglobulin production	2.12e-50	181
biological_process	GO:0002768	immune response-regulating cell surface receptor signaling pathway	2.31e-49	437

IFNG-AS1 is a lncRNA previously associated with IBD and shown to regulate IFNγ expression

The NeST Long ncRNA Controls Microbial Susceptibility and Epigenetic Activation of the Interferon-γ Locus

J. Antonio Gomez,[1] Orly L. Wapinski,[2] Yul W. Yang,[2] Jean-François Bureau,[3] Smita Gopinath,[1] Denise M. Monack,[1] Howard Y. Chang,[2] Michel Brahic,[1] and Karla Kirkegaard[1,*]

A long noncoding RNA signature for ulcerative colitis identifies IFNG-AS1 as an enhancer of inflammation

David Padua,[1] Swapna Mahurkar-Joshi,[1] Ivy Ka Man Law,[1] Christos Polytarchou,[2,3] John P. Vu,[4] Joseph R. Pisegna,[4] David Shih,[5] Dimitrios Iliopoulos,[1,2] and Charalabos Pothoulakis[1]

Regulation of the Th1 Genomic Locus from *Ifng* through *Tmevpg1* by T-bet

Sarah P. Collier,* Melodie A. Henderson,† John T. Tossberg,† and Thomas M. Aune*,†

IFNG-AS1 is also known as NeST or TMEVPG1

Conclusions

- Creating a Shiny App allowed automation and improvement of the existing workflow, saving time for scientists.
- Most of the app is general purpose and can be re-used for other co-expression analyses beyond the lncRNA project.
- Utility of approach has been confirmed with well characterized lncRNAs as well as novel lncRNAs through experimental validation.

Acknowledgements

Janssen R&D

- **Amy Hart**
- **Eric Neiman**
- Aleksandar Stojmirovic
- Joshua R. Friedman
- Carrie Brodmerkel
- Scott Plevy
- **Jacqueline Perrigoue**
- **Aisling O'Hara**

Icahn School of Medicine at Mount Sinai

- **Won-min Song**
- **Carmen Argmann**
- Bojan Losic
- Roman Kosoy
- Antonio Di Narzo
- Mayte Suárez-Fariñas
- Ruiqi Huang
- Lauren Peters
- Ke Hao
- Bin Zhang
- Eric Schadt
- Andrew Kasarskis
- Marla Dubinsky
- Bruce Sands

Thank you!

Stefan Avey

savey@its.jnj.com

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson


janssen
PHARMACEUTICAL COMPANIES
OF Johnson&Johnson