

Network and modularity analysis

Lou Shaoke

Department of Molecular Biophysics and Biochemistry

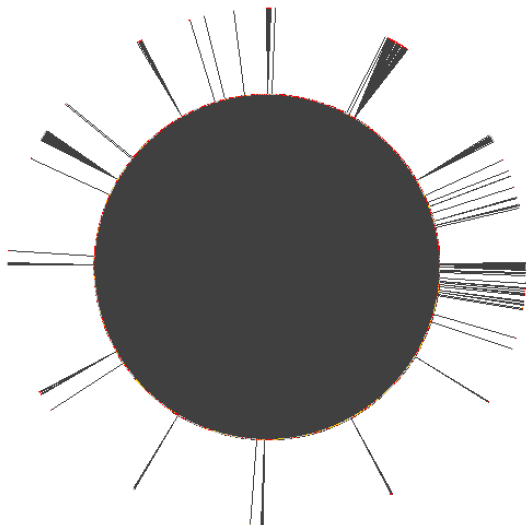
loushaoke@gmail.com

January 5, 2017

Network

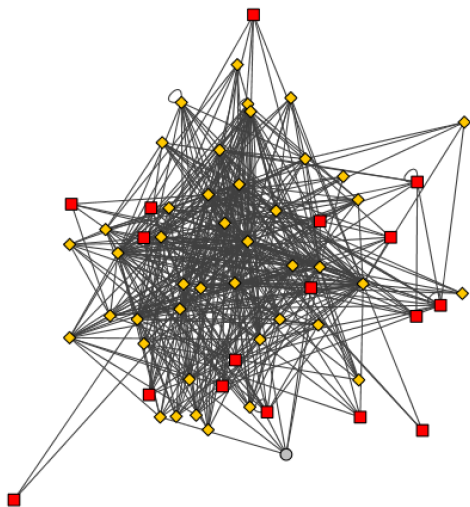
- ▶ enhancer distal regulatory network: enhancer gene linkage with ChIP-Seq peaks
- ▶ TF regulatory network: ChIP-Seq
- ▶ TIP: ChIP-Seq
- ▶ PPI and gene expression

Dense network

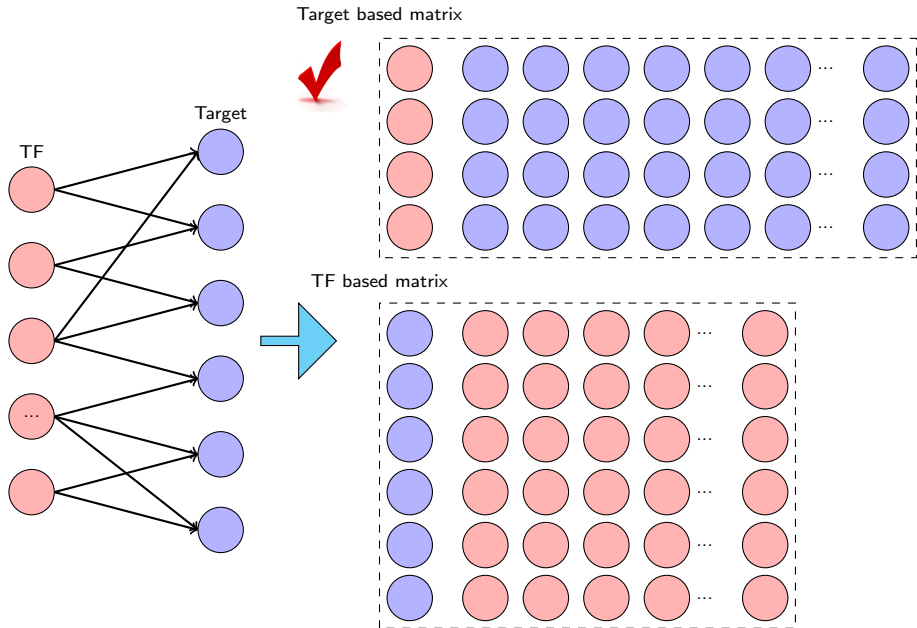


It is difficult to identify modules from such a dense network (above is circular layout for TF regulatory network).

TF-TF only network

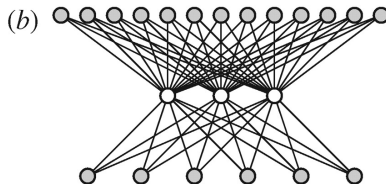
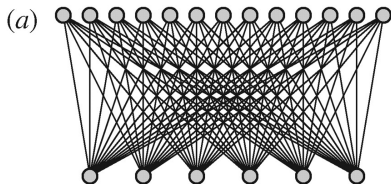


Even only consider TF-TF regulations, the structure is complex and hard to compare between cell lines (TF-TF interactions in TF regulatory network)



Regulator network can be transformed into two different views. Today I will only focus on using target to infer the hidden classes for TFs and also estimate the rewiring/changes between Gm12878 and K562 cell lines.

Mix Membership Model



target gene layer (top): $J = 1, \dots, j$

Hidden membership (class) layer: $H = 1, \dots, k, k=4$

TF layer (bottom): $T = 1, \dots, n$

Membership layer link the target gene and TF. each TF has a proportion λ for hidden membership, which is drawn from Dirichlet distribution;

For each hidden membership, H is drawn from multinomial distribution for all the target genes with parameter θ .

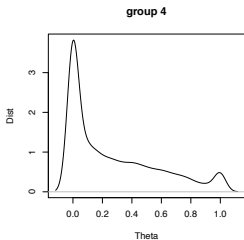
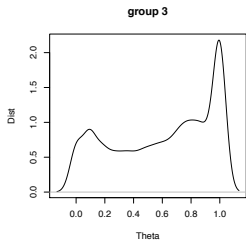
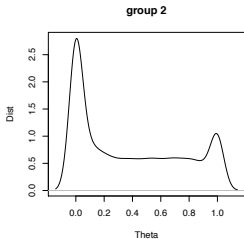
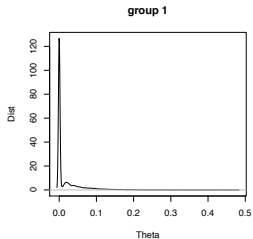
Each target gene is categorical variable 0,1 for each TF, where 1 means is target, 0 means not.

Clustering

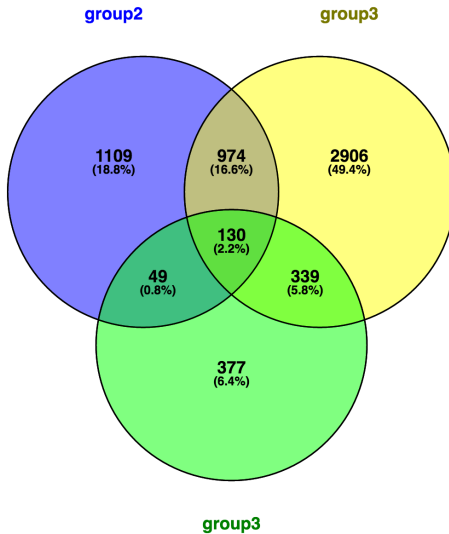


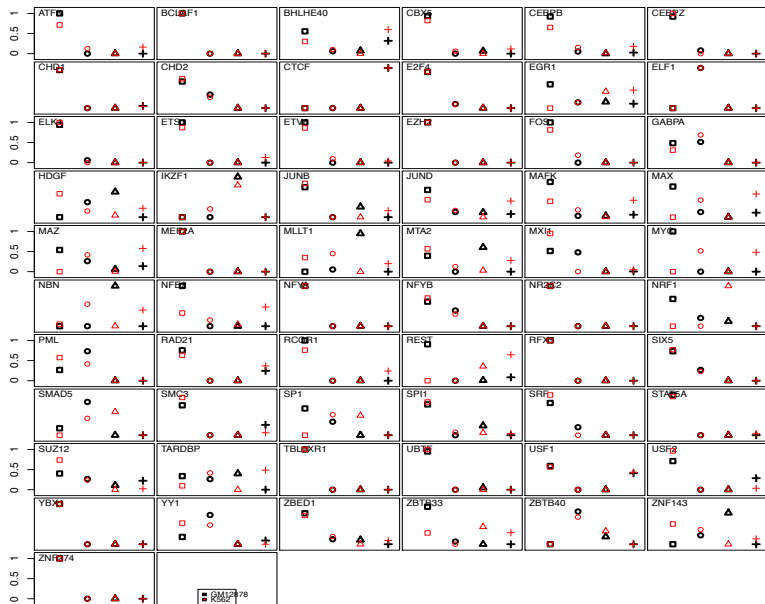
Simple clustering cannot explore the complex structure and compare two different samples.

Theta distribution for gene be a target of TF in four hidden group/memberships, and membership change if assign the group for each TF using max probabability in all the four classes.(group=class=membership)



Overlap of gene between groups

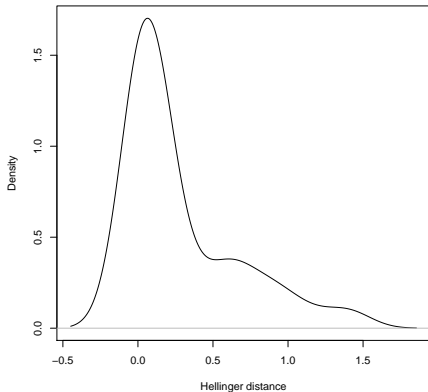




visualization of membership changes between Gm12878 and K562

	K.g1	K.g2	K.g3	K.g4
G.g1	36	2	2	7
G.g2	2	3	1	0
G.g3	3	2	1	1
G.g4	0	0	0	1

Table: TF groups/membership changes

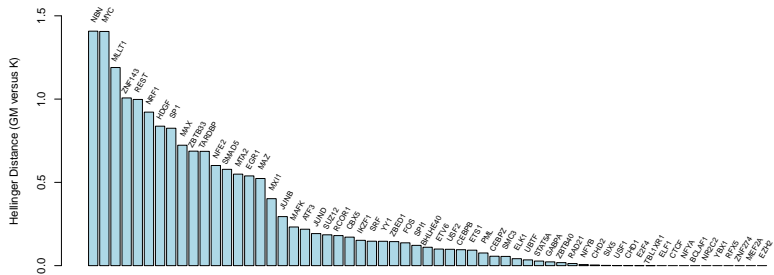


Hellinger Distance $d = \frac{1}{\sqrt{2}} \sqrt{\sum_i (\sqrt{p_{Gm,i}} - \sqrt{p_{K,i}})^2}$, is used to quantify the membership change between G_m and K cell lines

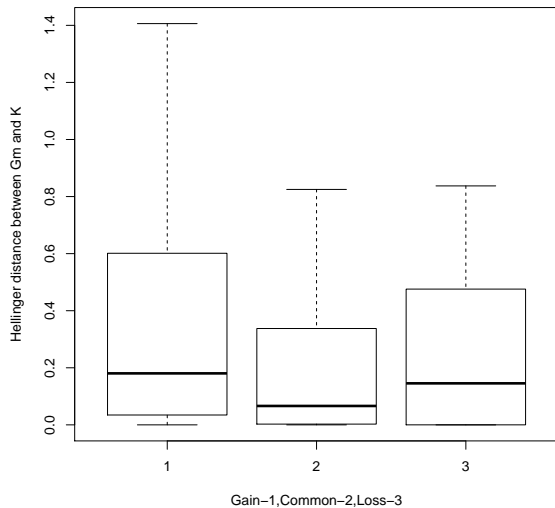
Top changed TF:

NBN,MYC,MLLT1,
 ZNF143,REST,NRF1,HDGF,
 SP1,MAX,ZBTB33,TARDBP,
 NFE2,SMAD5,MTA2,EGR1,
 MAZ

Rank of Hellinger distance

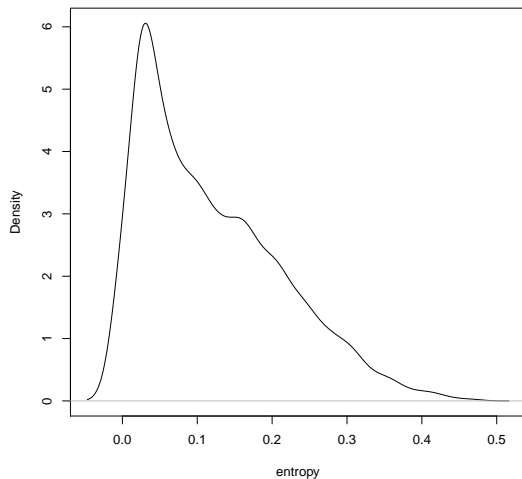


Compare with rewiring classes defined by DL



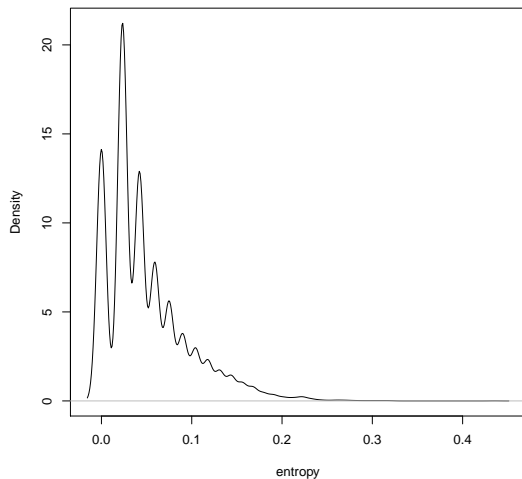
ATF3 □ ○ ▲ +	BCLAF1 □ ○ ▲ +	BHLHE40 □ ○ ▲ +	CBX5 □ ○ ▲ +	CEBPB □ ○
CHD1 □ ○ ▲ +	CHD2 □ ○ ▲ +	CTCF □ ○ ▲ +	E2F4 □ ○ ▲ +	EGR1 □ ○
ELK1 □ ○ ▲ +	EP300 □ ○ ▲ +	ETS1 □ ○ ▲ +	ETV6 □ ○ ▲ +	EZH2 □ ○
GABPA □ ○ ▲ +	HMGF □ ○ ▲ +	IKZF1 □ ○ ▲ +	JUNB □ ○ ▲ +	JUND □ ○
MAX □ ○ ▲ +	MAZ □ ○ ▲ +	MEF2A □ ○ ▲ +	MILL1 □ ○ ▲ +	MTA2 □ ○
MYC □ ○ ▲ +	NFE2 □ ○ ▲ +	NFYA □ ○ ▲ +	NFYB □ ○ ▲ +	NR2C2 □ ○
PML □ ○ ▲ +	POLR2A □ ○ ▲ +	POLR2AphosphoS2 □ ○ ▲ +	POLR2AphosphoS5 □ ○ ▲ +	POLR3G □ ○
RCOR1 □ ○ ▲ +	RES1 □ ○ ▲ +	RF-X5 □ ○ ▲ +	SIN3A □ ○ ▲ +	SIX5 □ ○
SMC3 □ ○ ▲ +	SP1 □ ○ ▲ +	SP11 □ ○ ▲ +	SRF □ ○ ▲ +	STAT5A □ ○
TAF1 □ ○ ▲ +	TARDBP □ ○ ▲ +	TBL1XR1 □ ○ ▲ +	TBP □ ○ ▲ +	UBTF □ ○
USF2 □ ○ ▲ +	YY1 □ ○ ▲ +	ZBED1 □ ○ ▲ +	ZBTB33 □ ○ ▲ +	ZBTB40 □ ○
ZNF274 □ ○ ▲ +	<div style="display: flex; align-items: center;"> <div style="width: 15px; height: 15px; background-color: black; margin-right: 5px;"></div> GM12878 <div style="width: 15px; height: 15px; background-color: red; margin-right: 5px; margin-left: 10px;"></div> K562 </div>			

The same analysis fails in TIP and Ehn network



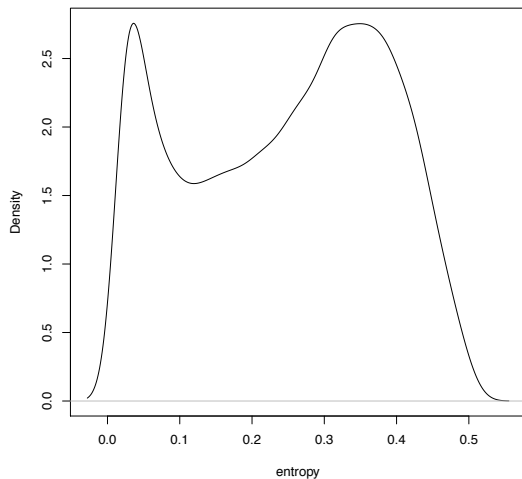
Enhancer distal network

The same analysis fails in TIP and Ehn network



TIP network

The same analysis fails in TIP and Ehn network



Regulatory network