# Network and modularity analysis

Lou Shaoke

Department of Molecular Biophysics and Biochemistry

*loushaoke@gmail.com*

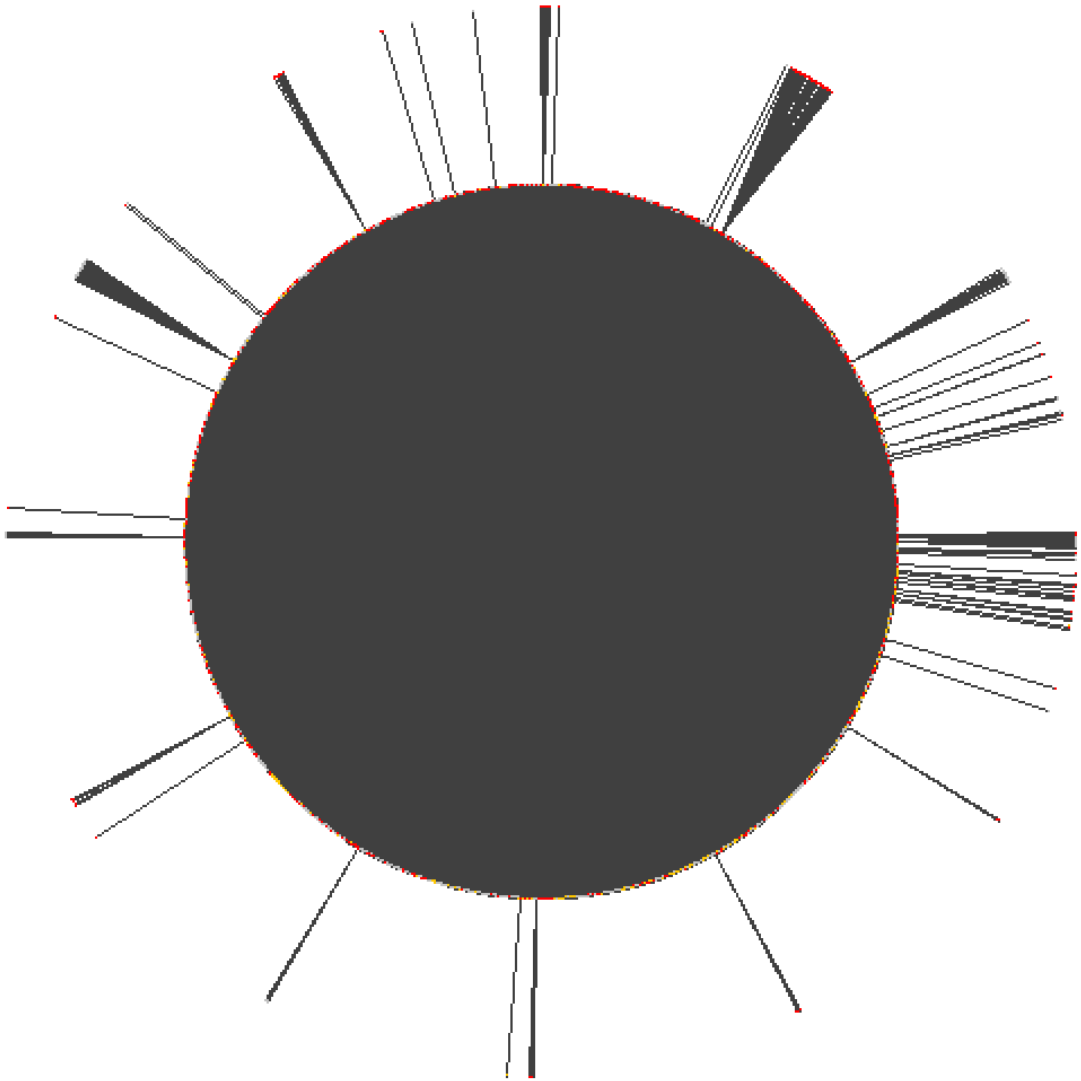
December 21, 2016

Yale

# Network

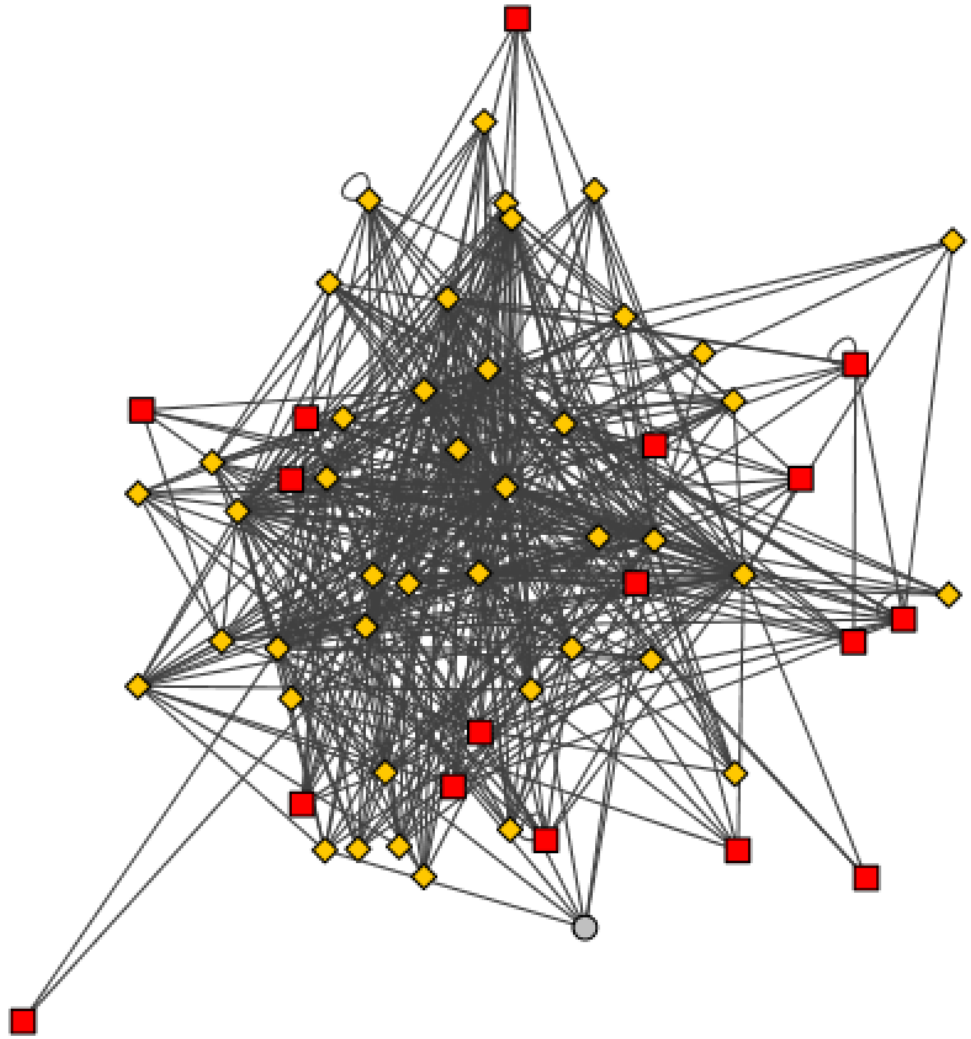
- enhancer distal regulatory network: enhancer gene linkage with ChIP-Seq peaks
- TF regulatory network: ChIP-Seq
- TIP: ChIP-Seq
- PPI and gene expression

# Dense network

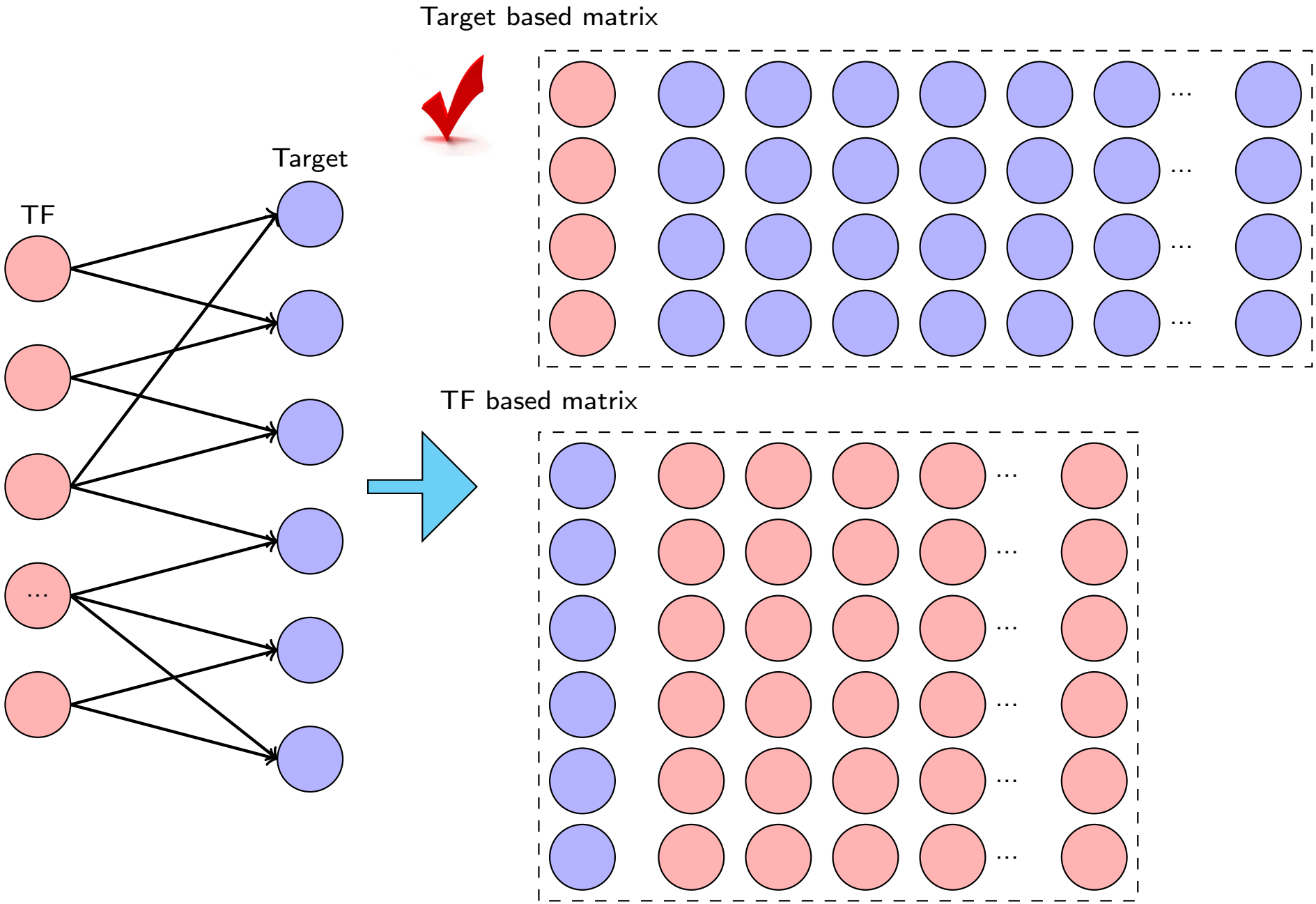


It is difficult to identify modules from such a dense network (above is circular layout for TF regulatory network).

# TF-TF only network

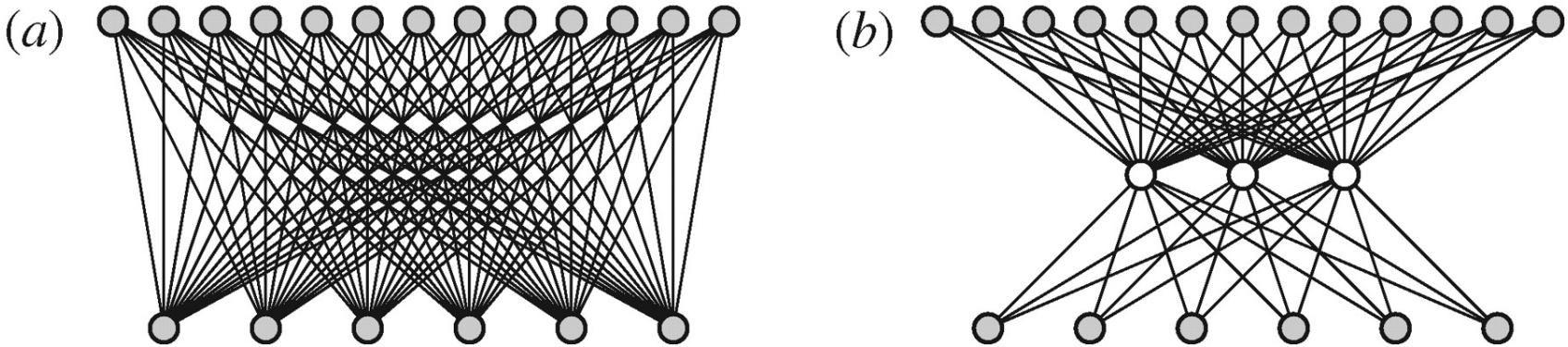


Even only consider TF-TF regulations, the structure is complex and hard to compare between cell lines (TF-TF interactions in TF regulatory network)




Regulator network can be tranformed 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

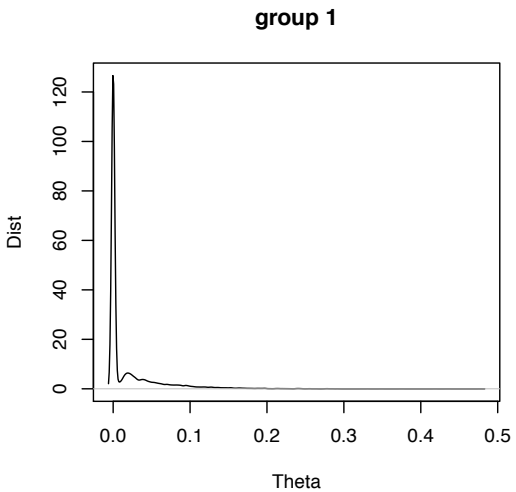


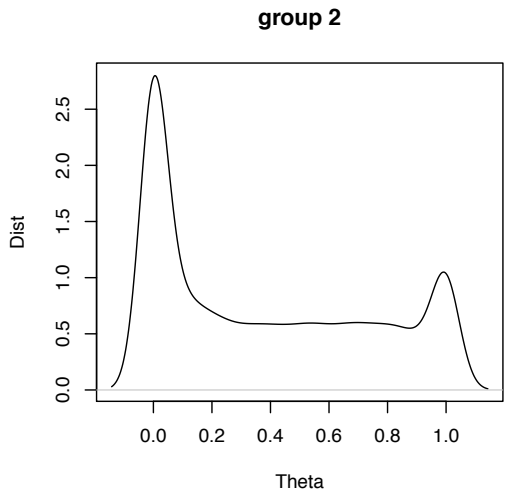

target gene layer (top): $J = 1, ..., j$
Hiden membership (class) layer: $H = 1, .., k$, k=4
TF layer (bottom): $T = 1, ..., n$
Membership layer link the target gene and TF. each TF has a proportion $\lambda$ for hidden membership, which is drawn from Dirichlet distribution;
For each hidden membership, H is drawn from multinomial distribution for all the target genes with parameter $\theta$.
Each target gene is categorical variable 0,1 for each TF, where 1 means is target, 0 means not.

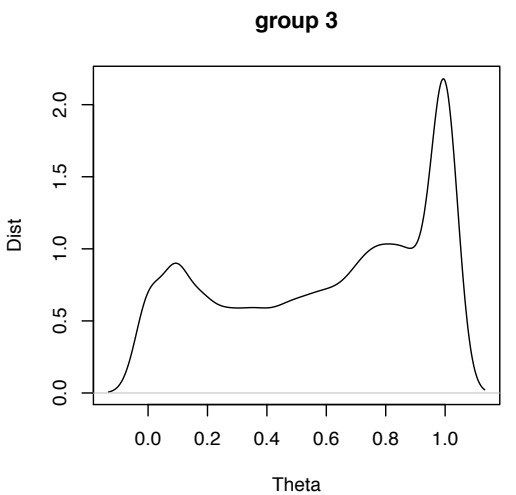
# Clustering

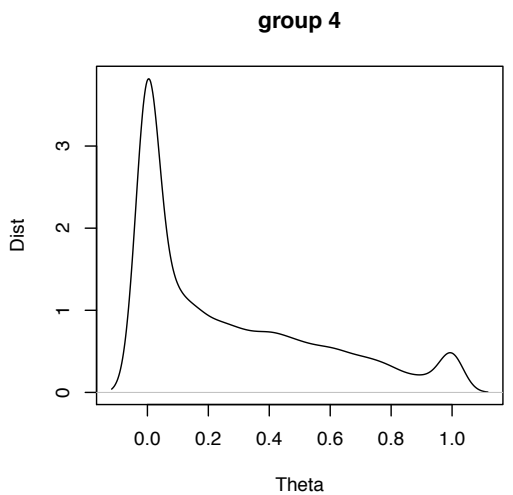

Simple clustering cannot explore the complex structure and compare two different samples.

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 probabablity in all the four classes.(group=class=membership)



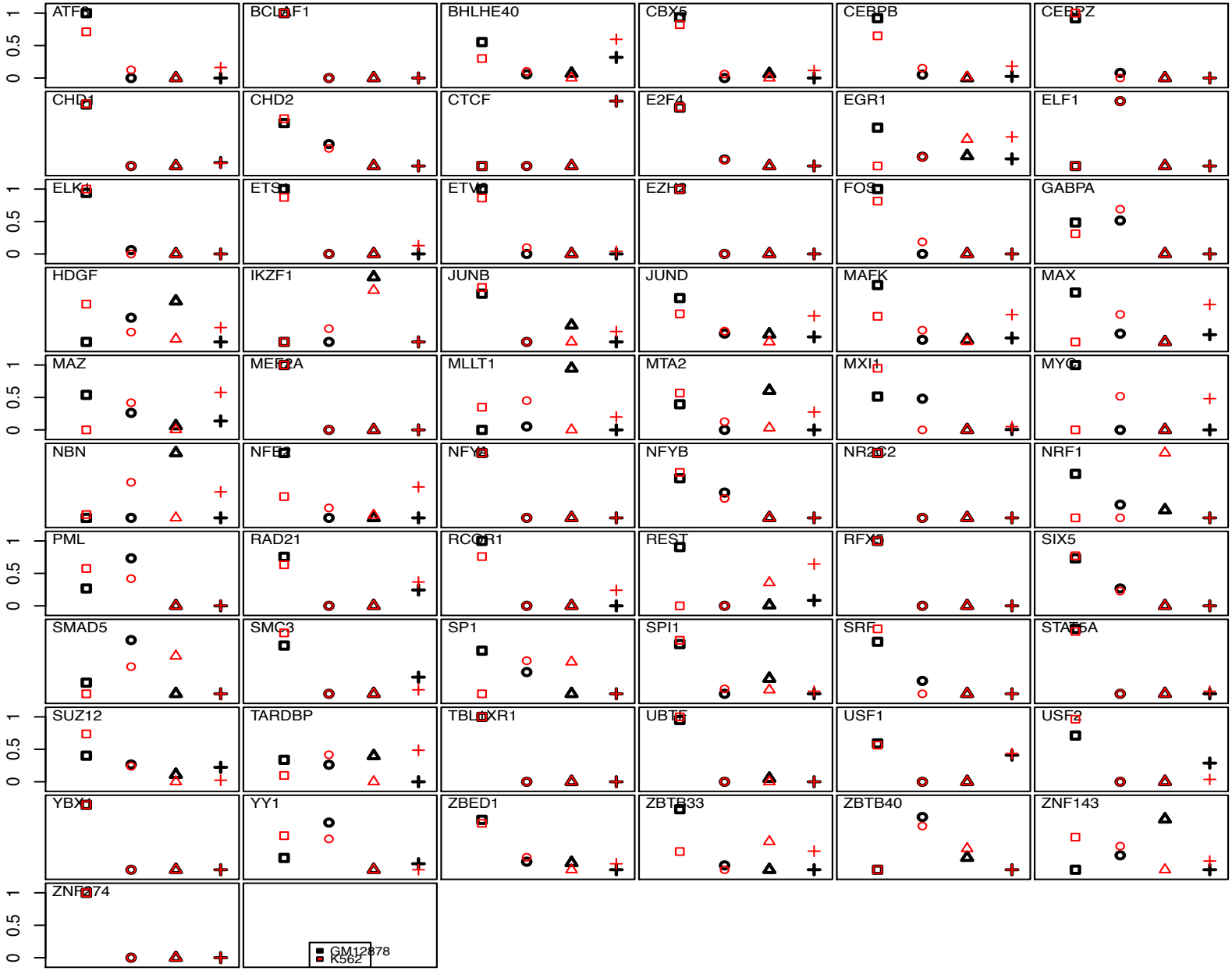



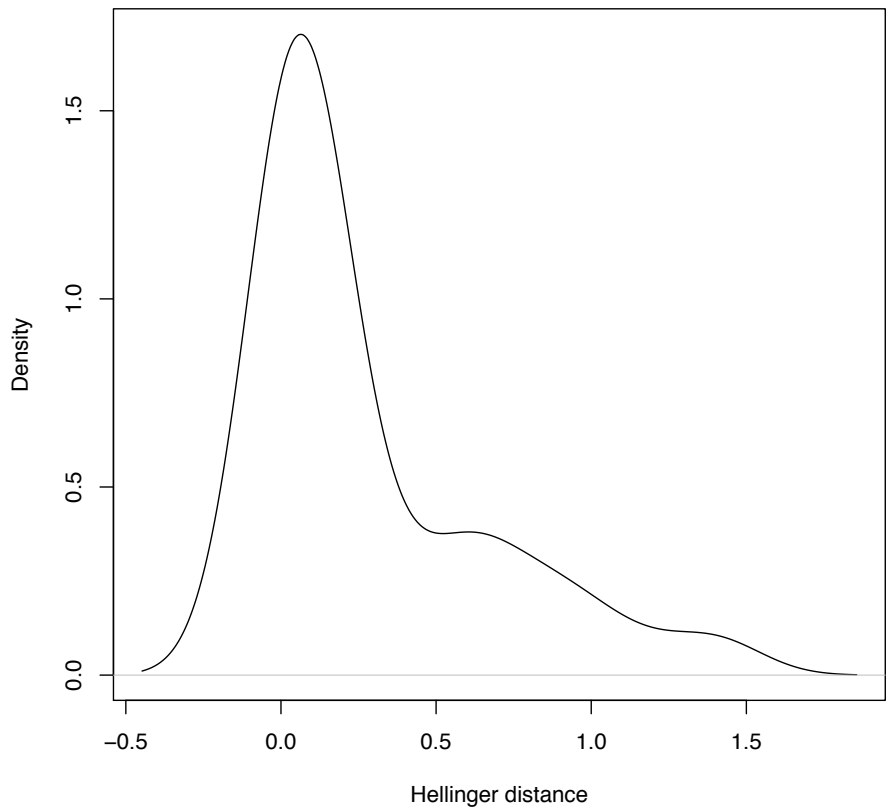

| | K.g1 | K.g2 | K.g3 | K.g4 |
|---|---|---|---|---|
| G.g1 | 36 | 2 | 2 | 7 |
| G.g2 | 2 | 3 | 1 | 0 |
| G.g3 | 3 | 2 | 1 | 1 |
| G.g4 | 0 | 0 | 0 | 1 |

Table: TF groups/membership changes




visulization of membership changes between Gm12878 and K562

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 Gm and K cell lines




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