



Genome-wide Analysis of the Distances



between Human Transcription Factor Binding Sites

12 / 10 / 05

Hyunmin Kim and Dr. Lawrence Hunter

Center for Computational Pharmacology
University of Colorado Health Sciences Center





Workflow



❖ Step 1: Data Acquisition

- ❖ 19,925 distinct promoter regions (-10k~10k bps) from the UCSC hg16
- ❖ 488,191 TFBSs (transcription factor binding sites) from the human-mouse-rat conserved TFBS track of GALA database (Giardine et al. 2003)

❖ Step 2: Preprocessing

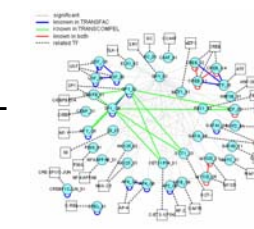
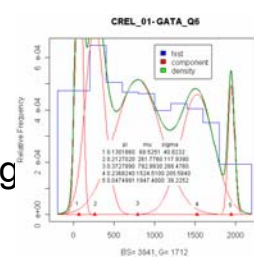
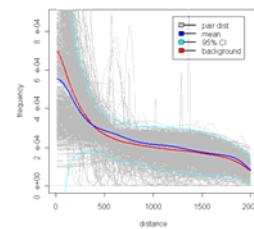
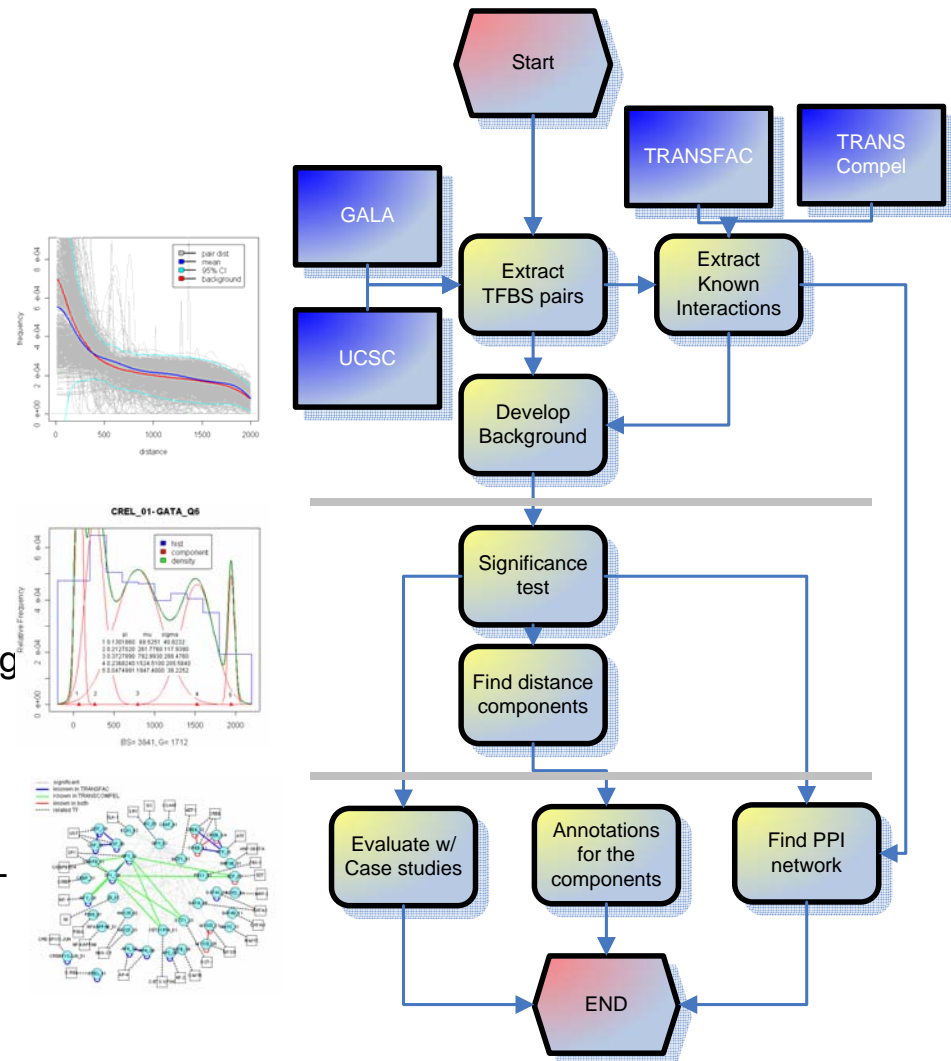
- ❖ 11,983,355 PWM (position weight matrix) pairs from the TFBSs

❖ Step 3: Data Analysis

- ❖ Komogorov-Siminov test
- ❖ w/ False Discovery Rate multiple testing correction

❖ Step 3: Evaluation

- ❖ w/ known TF-TF interactions of TRANSFAC v6.1 and TRANSCOMPEL v2.4
- ❖ w/ GO terminology





TF-TF Interactions

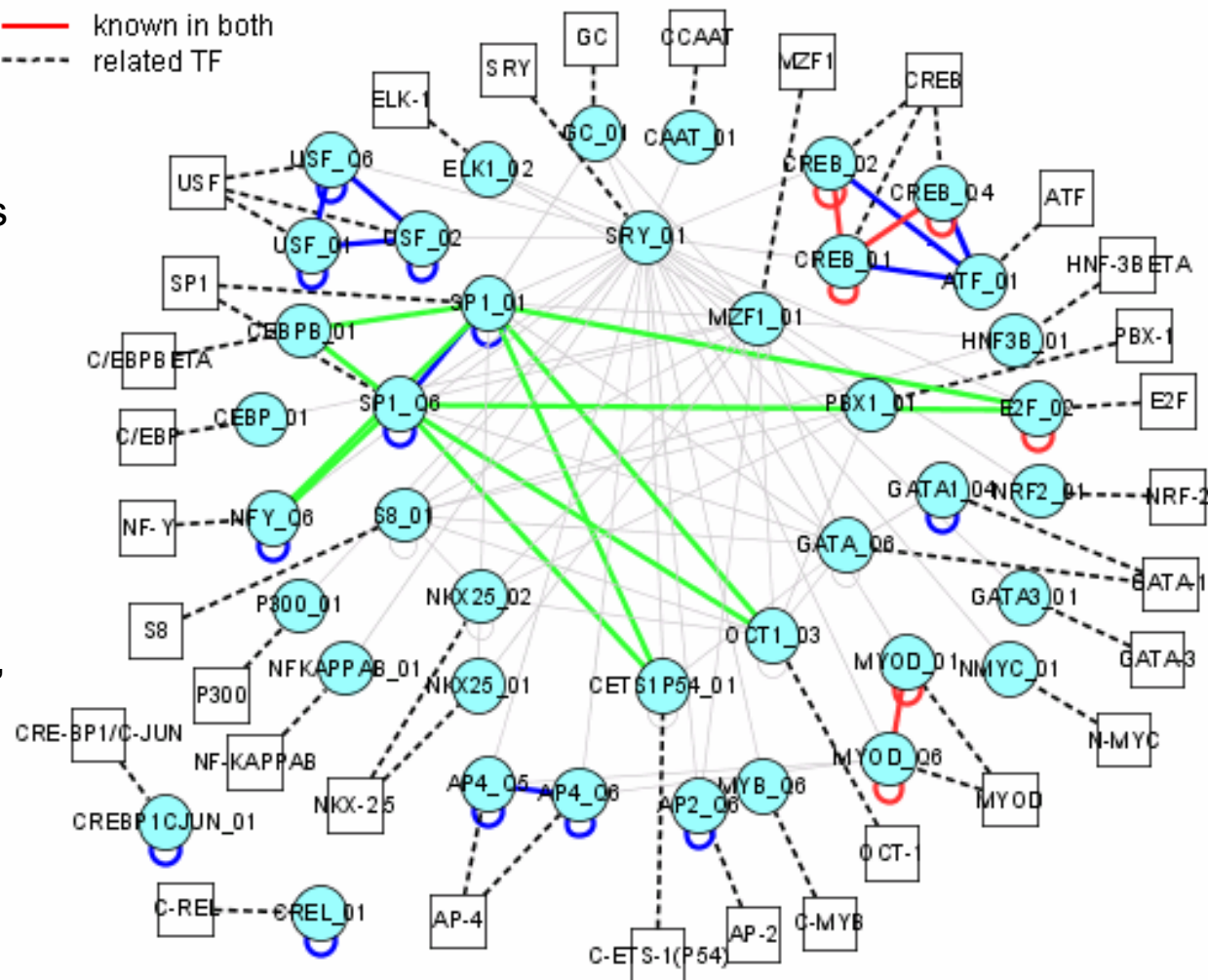


❖ Among 437 known PWM pairs that could have been found, 101 (23.1%) were among the 988 significant pairs at the threshold ($9.79e-005$)

❖ 39/100 are known PWM pairs (17/31 known TF pairs)

c.f. 10/100 known TF pairs (Rateitschak et al, 2004)

- significant
- known in TRANSFAC
- Known in TRANSCOMPEL
- known in both
- - - related TF





Functional analysis



- ❖ 186/988 significant pairs showed significant enrichment for a GO Biological Process term (p-value < 0.05, GOTermFinder)
- ❖ Each pair distribution is divided into 3,888 components using Finite Gaussian Mixture Model (Fraley and Raftery 2002)
- ❖ 790/3,888 components were associated with genes that had significantly enriched GO biological process annotations (p-value < 0.05)
- ❖ Examples
 - ❖ Direct association for SRF-SRF (two genes at a distance of 1647 ~ 1744 bps)
 - ❖ “regulation of cell growth” (GO:0001558)
 - ❖ “regulation of development” (GO:0050793)
 - ❖ Indirect association for MYOD-MYOD (eleven genes, 1938 ~ 2000 bps)
 - ❖ “cholesterol bio-synthesis” (GO: 0006695)
 - ❖ “MyoD gene family is affected by ARP-1 participating in cholesterol synthesis” (Muscat et. al, 1995)
 - ❖ A new association for ATF1-CREB (ten genes, 10 ~ 19 bps)
 - ❖ “glucosamine metabolism” (GO: 0006041) (Singh 2003)





Conclusions



- ❖ 10% of the significant pairs found by this method were previously published
- ❖ 90% potential interactions require experimental validation
- ❖ Hypothesized interactions were validated indirectly via demonstration of enrichment of Gene Ontology biological process terms
- ❖ We obtained consistent results of muscle-specific TFBS pairs and known TFBS interactions in the LCR(locus control region) and promoter regions of beta-globin genes
 - ❖ Muscle-specific genes
 - ❖ Matched known: Myf-Myf (p-value 4.98E-08), Srf-Srf (4.22E-05), and Mef2-Myf (7.79E-05)
 - ❖ Missed: Mef2-Mef2(0.132712)
 - ❖ Beta-globin cluster
 - ❖ Matched known: GATA1-NF-E2 (1.19E-14), GATA1-Sp1 (1.93e-12), and GATA1-Tal1(9.08e-7)

