

# TFBS of Pseudogenes

Annotation Meeting

Baikang Pei

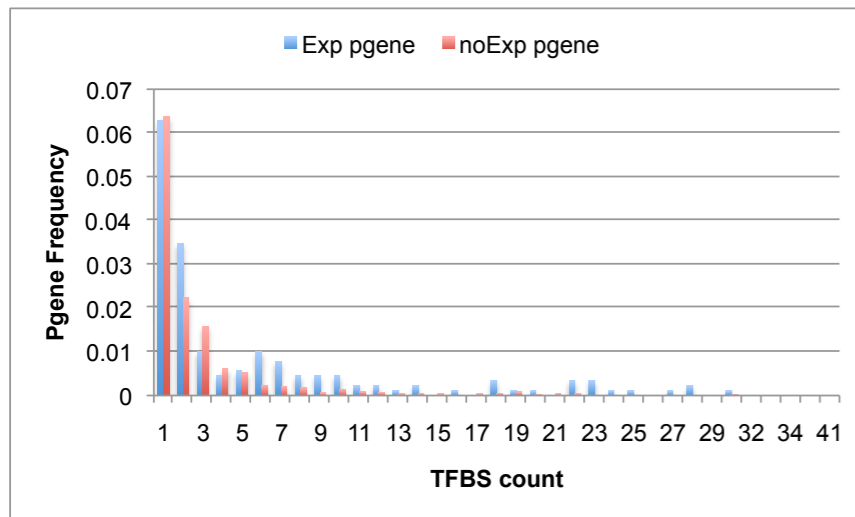
0/10/2011

## Upstream TFBS of Pseudogenes

- HAVANA pseudogene based on GENCODE v7
- Filter out pgenes whose TSS is within 4kb of protein coding TSS (transcribed: 892, non-transcribed: 9631)
- Collect TF binding peaks 2kb upstream of Pseudogenes
- TFBS datasets from the ENCODE analysis (January 2011) freeze
- TF peaks called by PeakSeq, using “optimal” setting
- Cell-line specific

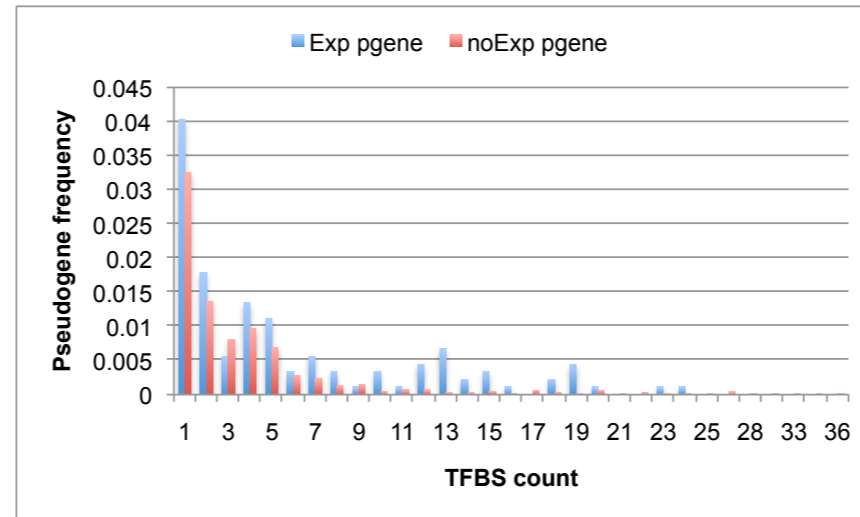
# Upstream TFBS of Pseudogenes

K562



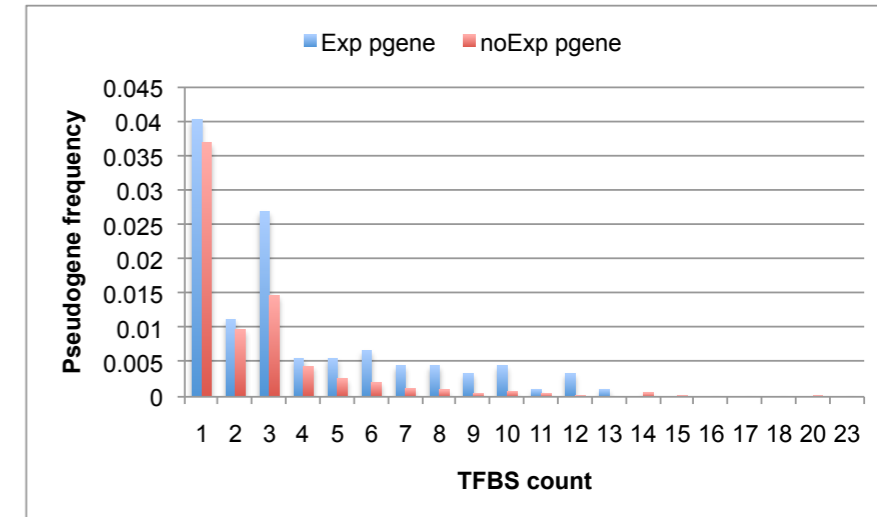
TFBS enriched in expressed pgenes, p-value = 0.0038

Gm12878



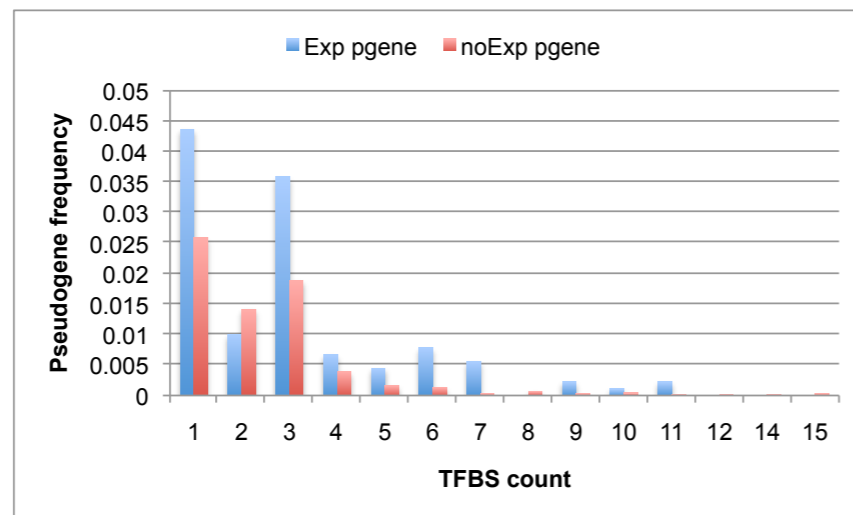
P-value = 0.006

HeLa3



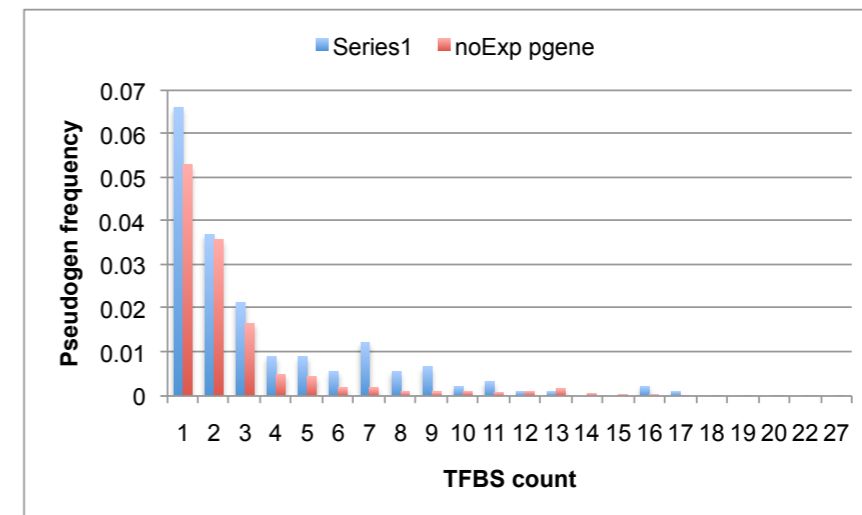
P-value = 0.014

H1-hesc



P-value = 0.004

Hepg2



P-value = 0.0016

This indicates the transcribed pgenes have more TFs bound to their upstream than non-transcribed pgenes

# Enriched TFBS in Upstream of Transcribed Pseudogenes

- TFBS enriched in transcribed pgenes upstream comparing to non-transcribed pgenes upstream in different cells
- Fisher exact test, with FDR = 0.05

Cell	TFBS
Gm12878	Pol2, Elf1sc631, Pol24h8, Zeb1sc25388, Oct2, Pou2f2, Sin3ak20, Pax5c20, Nrf1, Egr1, Znf143166181ap, Gabp, Usf2, Sp1, Tcf12, Usf1, Yy1, Srf, Six5, Rfx5n494, Cmyc, Taf1
K562	Yy1, E2f6h50, Pol2, Hey1, Gabp, Max, Tbp, Elf1sc631, Sin3ak20, Cmyc, Ets1, Zbtb7asc34508, E2f6, Pol24h8, E2f4, Ccnt2, Taf1, Nrf1, Hmgn3, Hdac2sc6296, Pol2b, Egr1, Cfos, Irf1, Thap1sc98174, Six5, Tf3c110, Nfya, Sp2sc643, Rad21, Usf1, Chd21250, Usf2
Helas3	Pol2, Ini1, Pol2s2, Gabp, Tbp, Taf1, Hae2f1, Nrf1, Mxi1bhlh, E2f1, Max, Elk4, Rfx5n494, Usf2, Cmyc, Smc3ab9263, Stat1
HI-hesc	Pol2, Taf1, Tbp, Pol24h8, Nrf1, Usf2, Gabp, Six5, Jun, Ctcfc5916, Yy1c20, Usf1, Taf7sq8, Rfx5n494, Sp1
Hepg2	Hey1, Sin3ak20, Pol2, Taf1, Elf1sc631, Cmyc, Tbp, Gabp, Usf1, Hdac2sc6296, Rfx5n494, Hnf4gsc6558, Nrf1, Sp1, Fosl2, Zbtb33, Hnf4ah171, Rad21, Ctcf, Jun