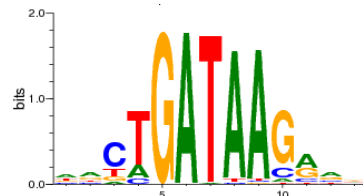
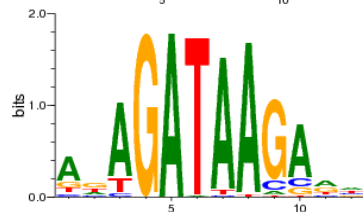


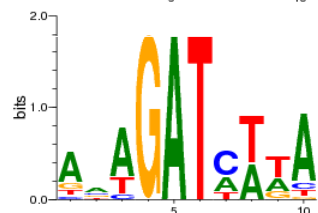
Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML
Diversity and complexity in DNA recognition by transcription factors.
 Science 324, 1720-3 (2009)



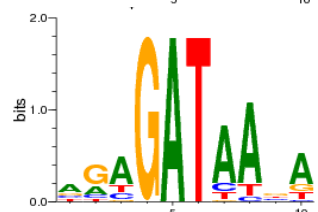
Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML
Diversity and complexity in DNA recognition by transcription factors.
 Science 324, 1720-3 (2009)



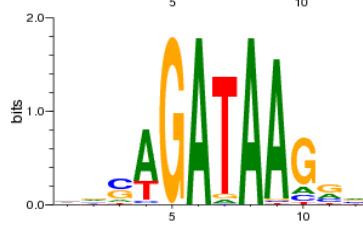
Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML
Diversity and complexity in DNA recognition by transcription factors.
 Science 324, 1720-3 (2009)



Ko LJ, Engel JD
DNA-binding specificities of the GATA transcription factor family.
 Mol Cell Biol 13, 4011-22 (1993)



Ko LJ, Engel JD
DNA-binding specificities of the GATA transcription factor family.
 Mol Cell Biol 13, 4011-22 (1993)

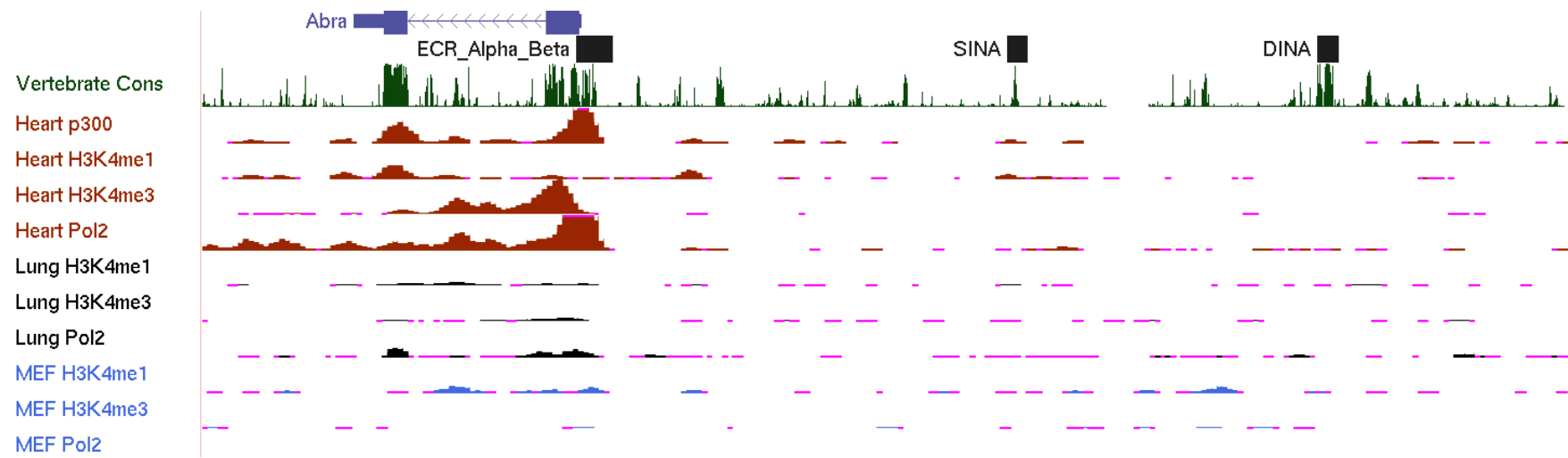


Fujiwara T, O'Geen H, Keles S, Blahnik K, Linnemann AK, Kang YA, Choi K, Farnham PJ, Bresnick EH
Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy.
 Mol Cell 36, 667-81 (2009)

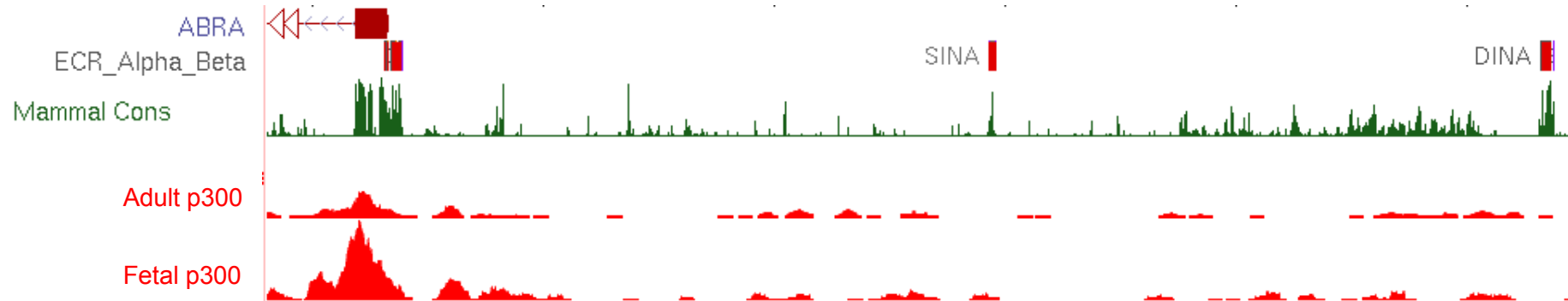
from MatBase by Genomatix

Supplemental Figure 1. Position weight matrix plots of GATA binding motifs. Experimentally generated GATA binding motif position weight matrices are represented in frequency bit score versus residue location plots. These matrix plots were taken from MatBase by genomatix. Primary source of binding site specificity data is cited in right panel adjacent to each plot.

A)

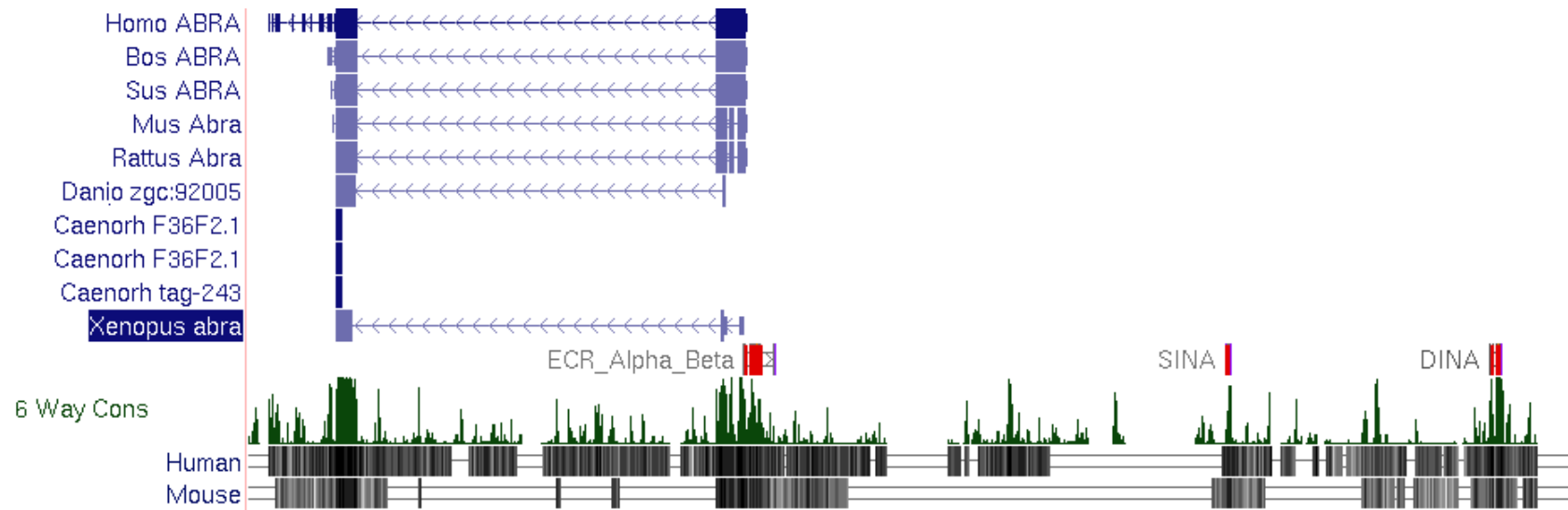


B)

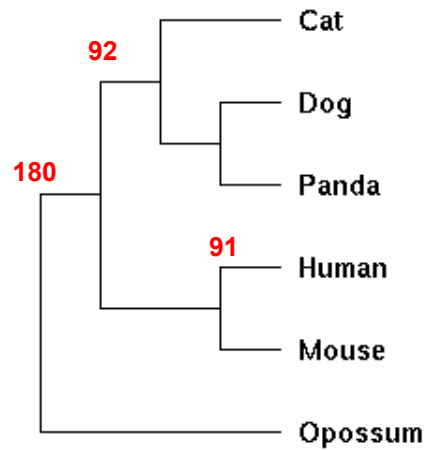


Supplemental Figure 2. Occupancy of chromatin modifications and transcription co-factors in the mouse and human *ms1* locus. Chromatin and transcription co-factor features in mouse 8 week adult heart, lung, mouse embryonic fibroblast (MEF) (A) and human fetal and adult human heart (B) are displayed for the mouse and human MS1 locus. The top track shows the annotated gene structure of mouse (A) and human (B) MS1 from Ensemble (A) and GENCODE (B) gene annotations. The second track shows the location of the mouse (A) and orthologous human (B) ECRs α , β , SINA and DINA (black horizontal boxes). The third track highlights sequence conservation using multispecies mammalian conservation plots from the UCSC genome browser. The densities of two histone modifications (H3K4me1 and H3K4me3), p300 and Pol2 are plotted as enrichment peaks isolated from respective tissues. ChIP-Seq data sets from mouse was obtained via the UCSC browser and encompassed within the freely available ENCODE data sets. Human fetal and adult heart p300/CBP ChIP-Seq can be freely downloaded from NCBI GEO database (GSE32587).

A)



B)



Supplemental Figure 3. Comparative genomic analysis of the cat *ms1/STARS* gene. Schematic of cat *ms1/STARS* locus with orthologous trans-mapped genes from other species in the top track. The second track shows mouse ECRs trans-mapped (via BLAT on UCSC browser) to the cat genome. The third track shows 6-way nucleotide sequence homology (MULAN alignment engine) plot of the ~30 kbp 5'-flanking region in cat, dog, panda, human, mouse and opossum. The bottom track specifically highlights orthologous regions of high sequence conservation in mouse and human (represented by vertical dashed lines, with strength of colour denoting percentage sequence identity) (A). Species used for 6-way nucleotide sequence and the evolutionary relationships between them are illustrated as a phylogenetic tree. Numbers indicate million years since sharing a last common ancestor for species linked at each branch point.

TABLE S1. *In silico* and experimental analysis supporting GATA4 binding site authenticity

GATA site	Sequence	Evidence
1	tCTATt	#Krejci et al .2004; GATA4 binding motif via EMSA and luciferase reporter assays in H9c2 cells
2	cATAGt	“ “
3	aCATCt	Matbase source matrix (Genomatrix Matinspector) – Supplementary Figure 1
4	tGATGa	“ “
5	aTATCc	“ “
6	cAATGt	“ “
7	tAACTt	“ “
8	cGATCa	“ “
9	cCTAGt	“ “
10	aGATAc	“ “
11	aGATGt	“ “
12	tTTAGc	“ “
13	tTATCt	“ “

#Krejci A, et al. 2004. Multiple promoters drive tissue-specific expression of the human M muscarinic acetylcholine receptor gene. *J. Neurochem.* **91**: 88-98.

TABLE S2A. Primers used in rat *msl*-luciferase reporter studies

Oligonucleotide name	Sense/Antisense	Sequence (5'-3')
MS1 (-1585)	Sense	TATTCAATGCTTAGTCCTGC
MS1 (-1382)	Sense	CAACTGATTGATGAGCTTGTC
MS1 (-1148)	Sense	GAGATATATGGGAGGTTGTAG
MS1 (-365)	Sense	TTACAGAGGTTTAAGTGAGAGC
MS1 (-300)	Sense	CGGAGCTCATGATTCAATCTAGTACTTTCC
MS1 (-127)	Sense	CGGAGCTCAGAACACCGTCAGAGCCATAGC
MS1 (+60)	Antisense	CCAAGCTTCAGGCTACCTGTTTCTTCTC
MS1 (-16702)	Sense	CGCTCGAGGCCCCACGAGTTAAAGCTCA
MS1 (-16432)	Antisense	TGGGCTGTGTGGGTTAGGAGA
MS1 (-9498)	Sense	TTGCTATGGGTGGGGAAGTGG
MS1 (-8608)	Antisense	GGCTCGAGGGGAAGGTTGGTCAGGAAAAGG

TABLE S2B. Primers used in ChIP studies

Oligonucleotide name	Sense/Antisense	Sequence (5'-3')
Cat B	Sense	CGGAGCTCAGAACACCGTCAGAGCCATAGC
Cat B	Antisense	CCAAGCTTCAGGCTACCTGTTTCTTCTC
Cat A	Sense	CTCACCAAAGAATAAAGAAGACTAGC
Cat A	Antisense	TAGAAGAGACAGATAACAGCTTCCT
Cat C	Sense	GAGGGGTAAATATAATCACGATACA
Cat C	Antisense	AGGAACTACAAGAAATTTAAATGC
Cat D	Sense	GTGTTGGATGAAGCGGAGAG
Cat D	Antisense	AACATGGTTTGCATAGCTGCA
RMS1 ECR α	Sense	CGGAGCTCAGAACACCGTCAGAGCCATAGC
RMS1 ECR α	Antisense	CCAAGCTTCAGGCTACCTGTTTCTTCTC
RMS1 ECR DINA	Sense	CTCACCAAAGAATAAAGAAGACTAGC
RMS1 ECR DINA	Antisense	TAGAAGAGACAGATAACAGCTTCCT
RMS1 Exon 2	Sense	GAGGGGTAAATATAATCACGATACA
RMS1 Exon 2	Antisense	AGGAACTACAAGAAATTTAAATGC
RNcx1 -50 GATA	Sense	GTGTTGGATGAAGCGGAGAG
RNcx1 -50 GATA	Antisense	AACATGGTTTGCATAGCTGCA
RMS1 ECR α /B	Sense	CTTGCCAAGATGTGGTGCTA
RMS1 ECR α /B	Antisense	AGTGTCTTTGCCACCAGAG
RMS1 ECR DINA/A	Sense	GAGAGGCTGCCTGAGAGCTA
RMS1 ECR DINA/A	Antisense	TGCCAGGTTTAGCCATCTCT
RNon-specific C	Sense	CCACCTCAGTGCCCTAACAT
RNon-specific C	Antisense	TGCGATTGTCTCAGAGGATG
mDINA/A	Sense	GGAAGCTGTGTGGGAAGGTA
mDINA/A	Antisense	AGCCTCTCCAGTCTCCCTGT
mAlpha/B	Sense	CCCTGACGGTGTCTGTTTT
mAlpha/B	Antisense	ACTCTGGCACTTGGAGGAGA
mSmarcd3	Sense	CCGCATTCTGCACTGATA
mSmarcd3	Antisense	ACGCGGAGTGCACAGGAG
mNegative/F	Sense	ATTTTGTGCTGCATAACCTCCT
mNegative/F	Antisense	TAGCAACATCCTAAGCTGGACA
mNon specific C	Sense	TAGTTGAAACGCGGCTACTG
mNon specific C	Antisense	GAAACCAGCAGAGTGGAAGC

ECR, Evolutionary conserved region; DINA, Distal Intergenic Activator; ChIP, Chromatin immunoprecipitation ; R, rat ; m, mouse