**Transcription Factor Binding Site Database (TFBS\_DB)**

**Two use cases:**

1. Search by Transcription Factor (TF)
2. Search by Gene

**Use Case 1: Search by TF –**

User enters from front page and enters a TF into a text box or chooses from a drop down of pre-populated TFs. The user clicks submit and is sent to a new page with detailed information about the TF. Multiple motifs may exist for a single TF. A list of the different motifs for the TF should be on the top of the page. Each motif should have the following information:

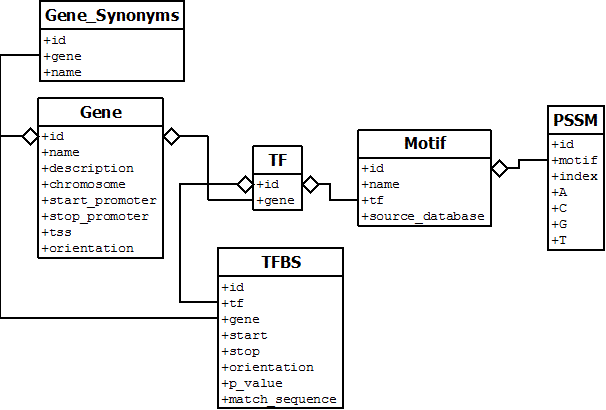
1. PSSM using Wei-Ju’s HTML5 representation
2. Histogram of positions of binding sites for this TF relative to the transcriptional start site (TSS). Boundaries of this plot will be 5Kbp upstream and 500bp downstream of the TSS.
3. Table that is a list of all targets aggregated by gene. The columns headers and data are:
   1. Gene – Gene symbol and Entrez ID with links to the page for the gene
   2. Chromosome – Chromosome the gene resides on
   3. Promoter Definition – Start and stop of promoter
   4. TSS – Transcriptional start site
   5. Number of Sites – Total number of motif instances in the promoter of the gene

**Use Case 2: Search by Gene –**

User enters from front page and enters a gene into a text box. We allow them to enter Entrez Gene IDs, Official Gene Symbols, Ensembl Gene and RefSeq Transcript IDs. The user clicks submit and is sent to a new page with detailed information about the Gene. Multiple TFs may be identified per gene. A list of the TF binding sites (TFBS) will be displayed:

1. HTML5 promoter in genome web browser with the following tracks
   1. TFBS in the positive orientation
   2. Our promoter definition track including the TSS
   3. TFBS in the negative orientation
   4. DNAse I hypersensitivity footprints
2. Legend for TFs in the promoter image
3. Table that is a list of all TFBS. The column headers and data are:
   1. TF – Gene symbol and Entrez ID with links to the page for the TF
   2. Motif – Name of the and the corresponding source database (JASPAR, TRANSFAC, Jolma,et al. 2013)
   3. Location – Chromosome, start and stop locations (hg19)
   4. Distance to TSS – Base-pairs to TSS, positive means upstream, negative means downstream.
   5. Orientation – Orientation of binding site and in parentheses relative to gene
   6. Sequence of Match – Sequence of the matching sequence in the promoter
   7. P-Value of Match – P-value of the match from FIMO

**Database Schema:**



**Additional Resources:**

1. UCSC track of TFBS for hg19
2. CSV file with all TFBS