



L.B.B.

Laboratory of Systems Biology
and Bioinformatics

Introduction to Transcription Factor Databases: Part2

Fereshteh Noroozi

Advised by: Prof. Ali Masoudi-Nejad

Winter 2024



Review of Previous Presentation

Database Name	Subject
TmAlphaFold	Transmembrane structure proteins
BioGRID	Protein genetic and chemical interaction
HIPPIE	Human integrated protein-protein interaction
dcGO	Domain-centric ontologies
MatrisomeDB	ECM COMPOSITION
DEPICTER2	Disorder regions
DescribePROT	Predicted amino acid-level descriptors



L.B.B.

Laboratory of Systems Biology
and Bioinformatics

UniBind



Research | [Open access](#) | [Published: 26 June 2021](#)

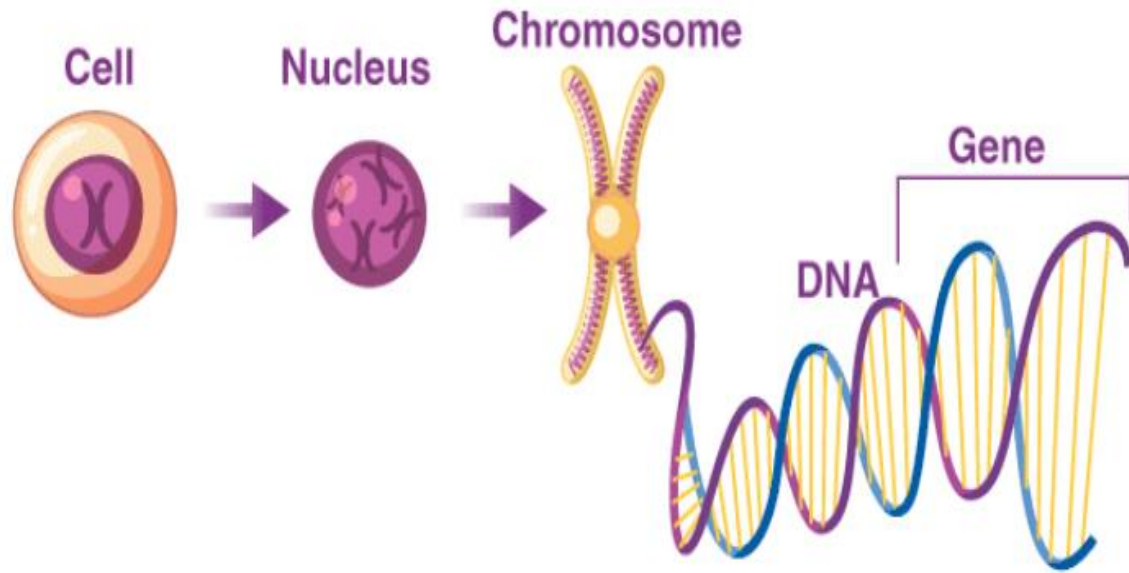
UniBind: maps of high-confidence direct TF-DNA interactions across nine species

[Rafael Riudavets Puig](#), [Paul Boddie](#), [Aziz Khan](#), [Jaime Abraham Castro-Mondragon](#) & [Anthony Mathelier](#) 

Oslo University, Norway

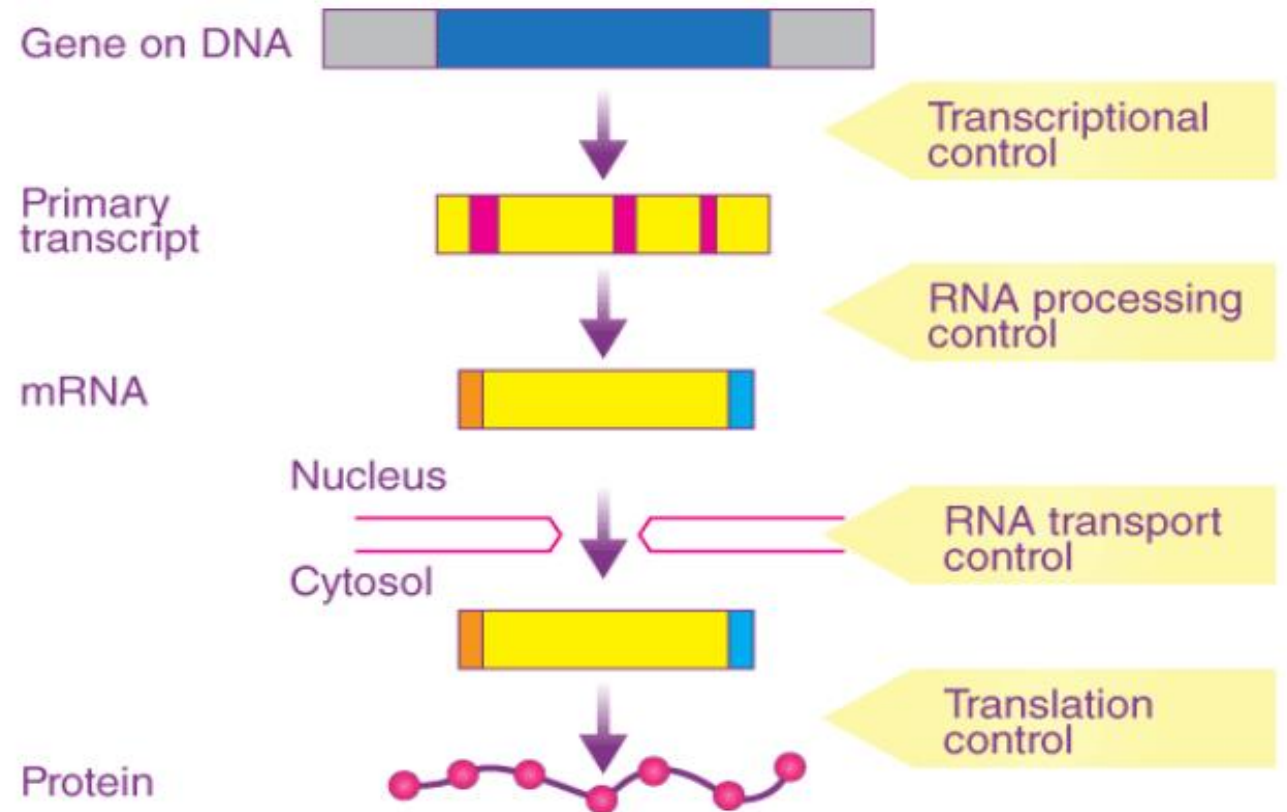
[BMC Genomics](#) **22**, Article number: 482 (2021) | [Cite this article](#)

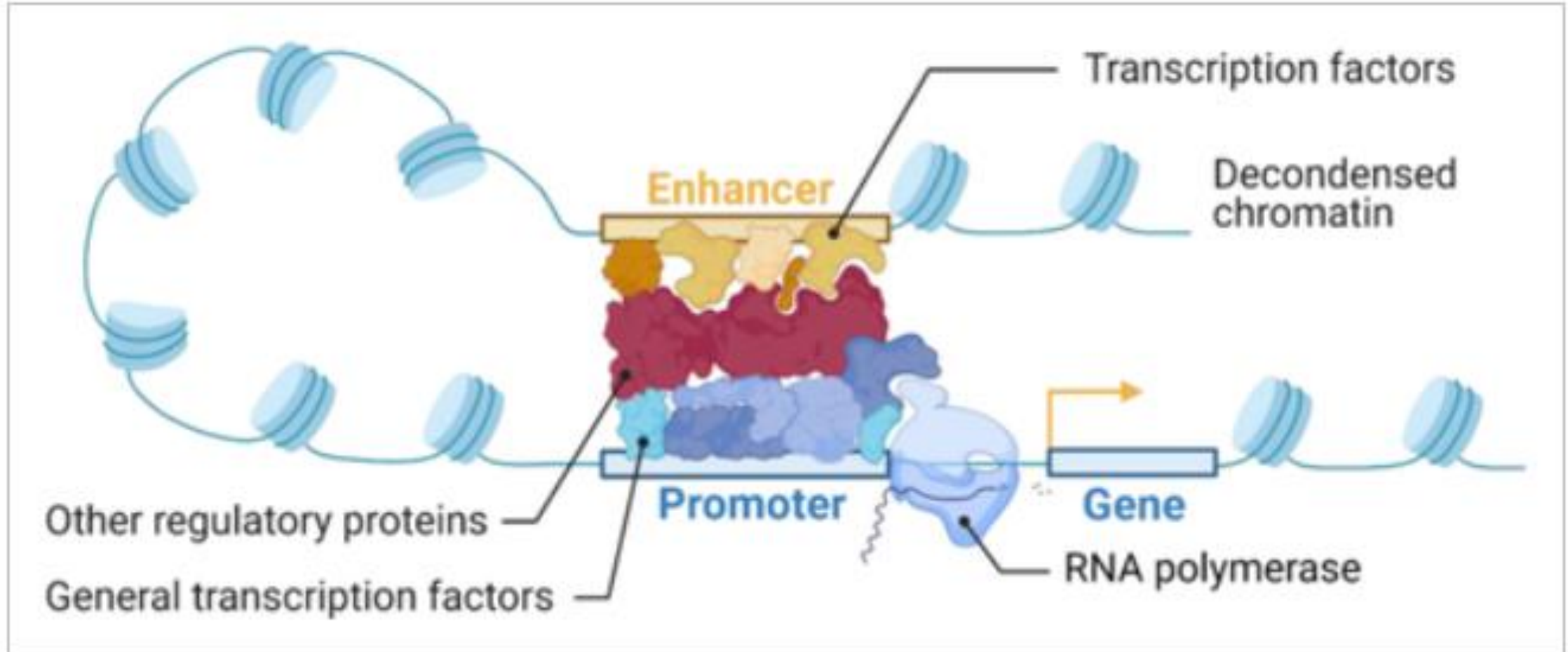
- Processed ~10,000 public ChIP-seq datasets from nine species
- Predicted ~56 million TFBs with experimental and computational support
- Identified TF-DNA interactions for 644 TFs in >1000 cell lines and tissues



What is gene?

Regulation of gene expression







L.B.B.

Laboratory of Systems Biology
and Bioinformatics

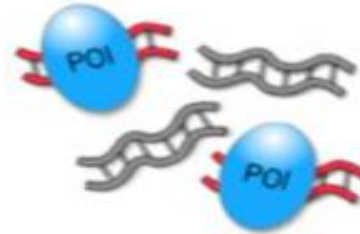
UniBind



DNA + bound protein



Fragment DNA



Immunoprecipitate



Release DNA



Map sequence
tags to genome
& identify
peaks



Sequence

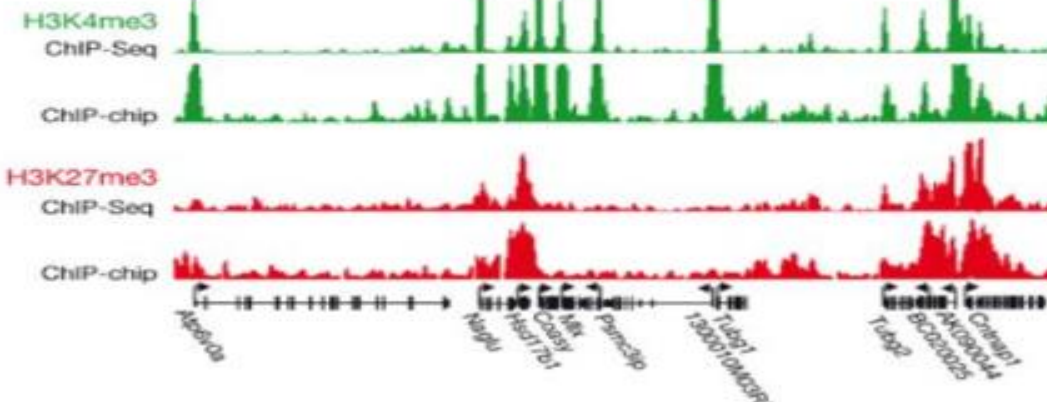
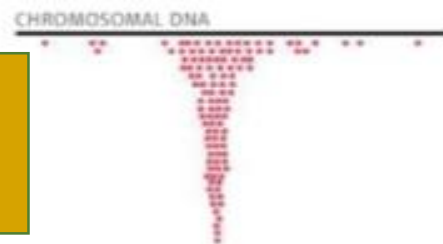


Prepare
sequencing
library

Adding sequence
adaptor

By heating the
samples

Regions of the genome
bound by the protein



Adapted from slide set by: Stuart M. Brown, Ph.D.,
Center for Health Informatics & Bioinformatics, NYU School of Medicine



9654

ChIP-seq datasets



1316

Cell lines & Tissues



841

Transcription Factors



9

Species



Arabidopsis thaliana



Caenorhabditis elegans



Danio rerio



Drosophila melanogaster



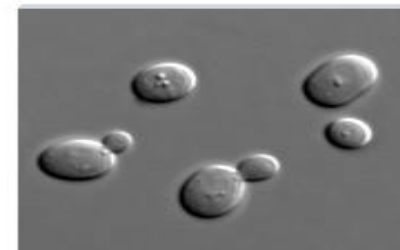
Homo sapiens



Mus musculus



Rattus norvegicus



Saccharomyces cerevisiae



Schizosaccharomyces pombe



L.B.B.

Laboratory of Systems Biology
and Bioinformatics

UniBind



Species:

Arabidopsis thaliana

TF name:

HY5

leucine zipper (bZIP)
family

Data source:

All sources

All sources, GEO, GTRD

Collection:

Robust


All collections, Permissive, Robust

☒ Has log(p-value) below defined centrality
threshold

**Centrality
threshold:**

0.5

log(p-value); maximum 0

Use the search field below to restrict the cell lines shown. Clearing the
field  will show all relevant cell lines again.

Search cell lines...

Search 

Cell line/tissue:

☒ **total tissue**

☐ flowers, 5-week-old plants

☐ inflorescence

☐ inflorescence meristem and flower from stages 1 to 11

☐ inflorescence meristems and floral buds (5-7 weeks)

☐ seedling

☐ seedlings 2 weeks old

☐ transition apices

☐ whole 3 days seedlings

☒ whole plants



2 datasets found **Robust**

Download ▾

Display 10 ▾ profiles

Filter:

TF Name ▴▾	Cell-line ▴▾	Condition/Source ▴▾	Collections ▴▾	Species ▴▾	Identifiers ▴▾	JASPAR ID ▴▾
HY5	whole plants	Treatment: blue light treated; Ecotype: Ws; Genotype: HY5 mutant;	Permissive Robust	Arabidopsis thaliana	GSE117797 ...	MA0551.1
HY5	whole plants	Treatment: red light treated; Ecotype: Ws; Genotype: HY5 mutant;	Permissive Robust	Arabidopsis thaliana	GSE117797 ...	MA0551.1

Copy

CSV

Showing 2 records in page 1 from 1 pages

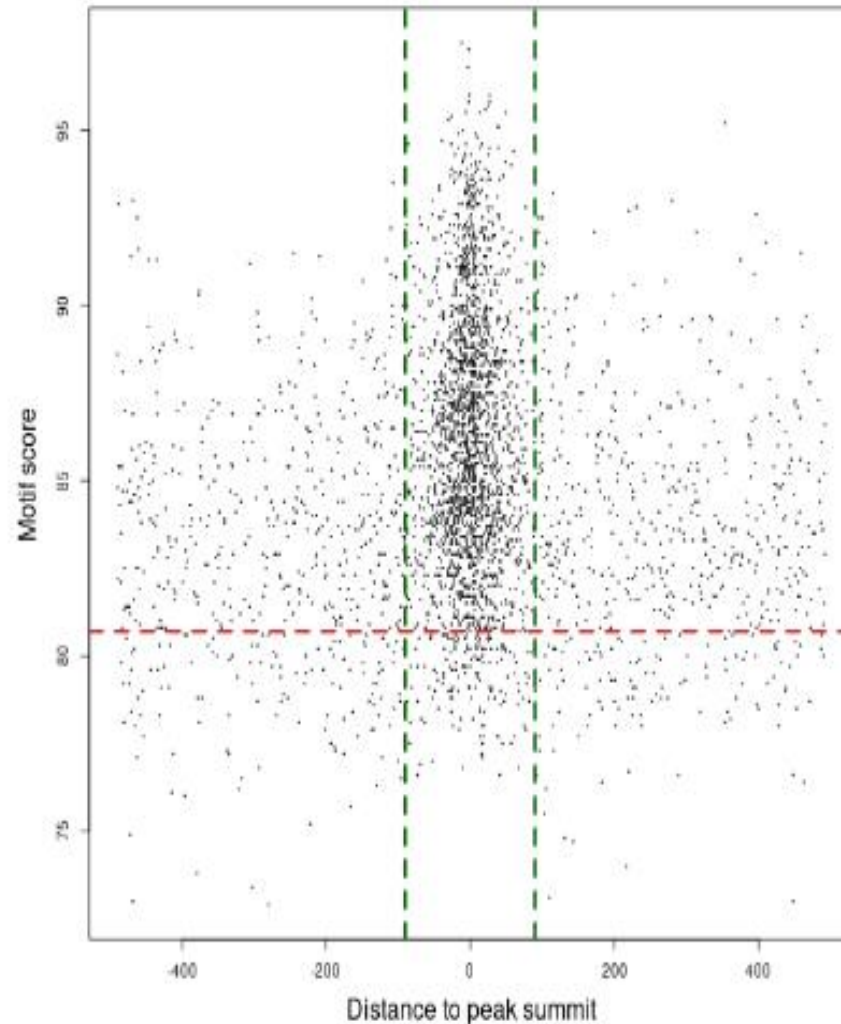
< 1 >

Blue/Red light-treated Arabidopsis thaliana (Ws ecotype) carrying a mutation in the HY5 gene.

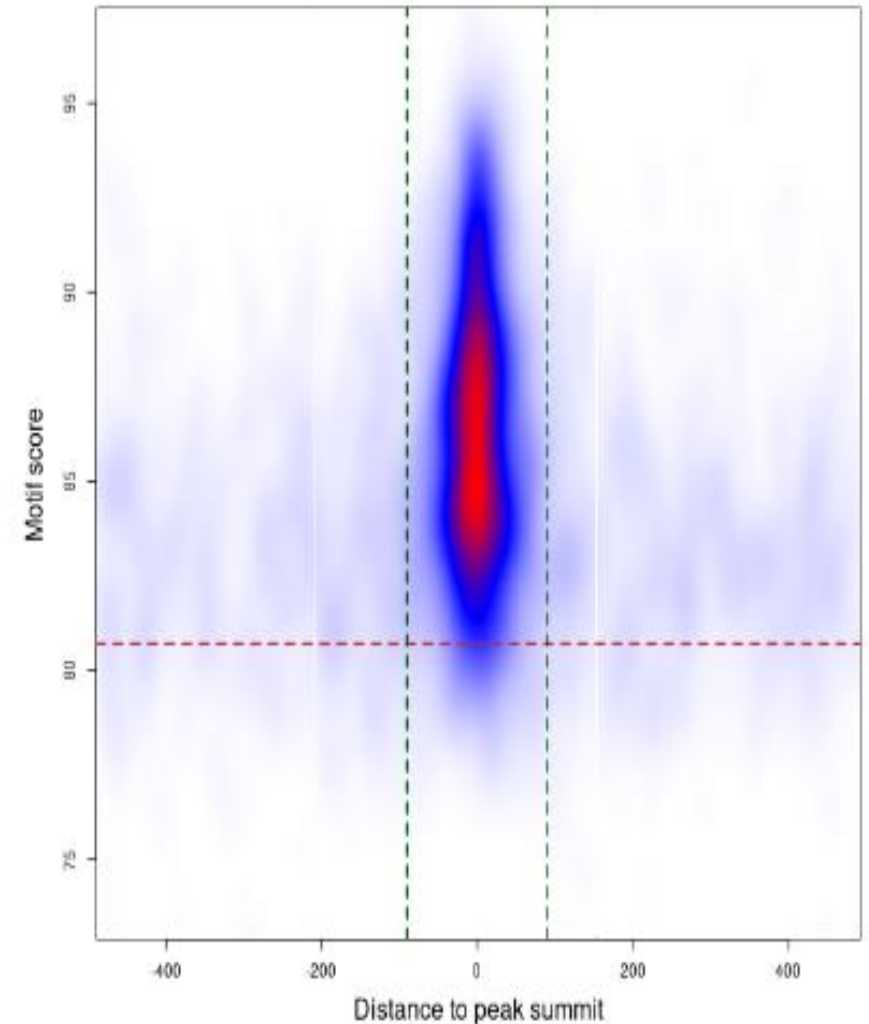


TF Name: HY5

Top hit for each peak



Density heat map for the top hits



Quantitative measure of the similarity between a predicted TFBS sequence and a known transcription factor binding motif or consensus sequence.





Advantages	Disadvantages
Comprehensive dataset covering 9 species	Data limitations: Certain TFs, cell types, or species may be underrepresented
High-quality predictions based on robust QC	Quality control thresholds may affect dataset inclusion
Predictive modeling enhances accuracy	Algorithmic assumptions may influence predictions
User-friendly interface for easy exploration	Interpretation challenges require experimental validation
Public accessibility promotes collaboration	Maintenance and updates are essential for relevance



Database

7190 ChIP-seq samples of 659 TFs and high-confidence binding sites of 699 TFs

hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets

Qiong Zhang¹ #, Wei L
An-Yuan Guo¹  

China, 2020

- Browse or search general targets of a query TF across datasets.
- Browse TF-target regulations for a query TF in a specific dataset or tissue.
- Search potential TFs for a given target gene or non-coding RNA.
- Investigate co-association between TFs in cell lines.
- Explore potential co-regulations for given target genes or TFs.
- Predict candidate TF binding sites on given DNA sequences.
- Visualize ChIP-seq peaks for different TFs and conditions in a genome browser.



Database of Human Transcription Factor Targets

hTFtarget

TF

Target

Peak

Co-regulation

Co-association

Prediction

Document

Contact

Download

Query gene

ASCL2

Submit

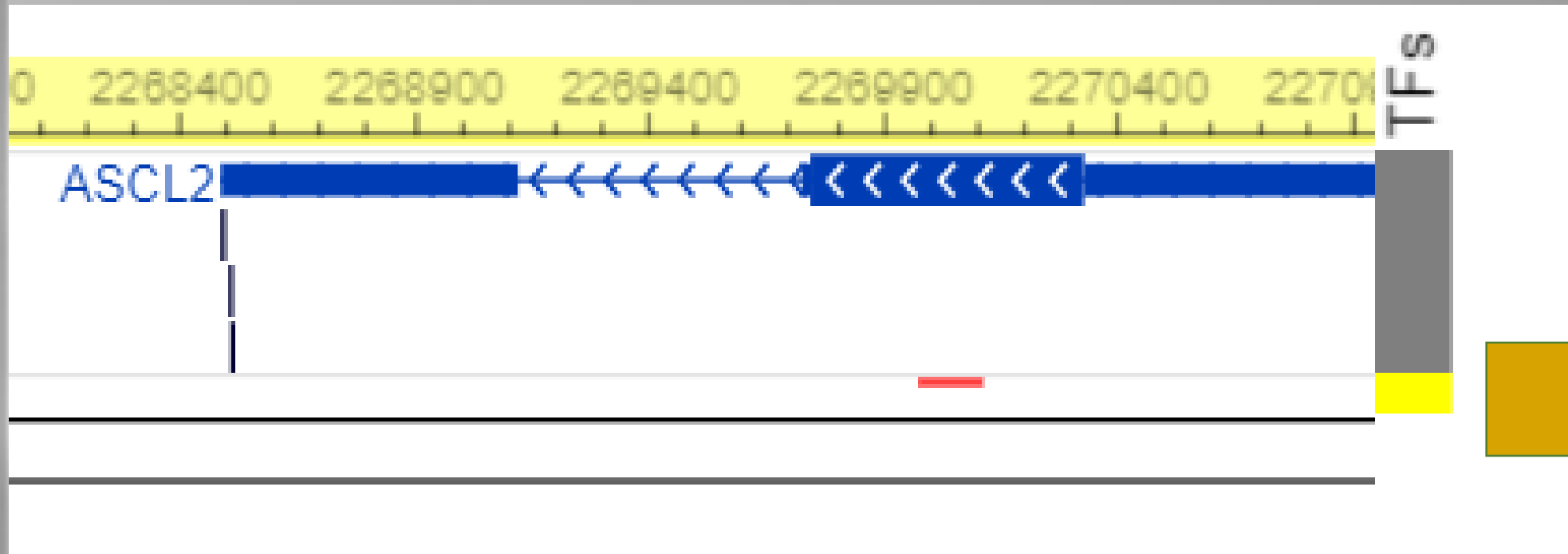
Query gene(s): **ASCL2** regulated by Transcription Factors:

#	Ensembl ID	Official Name	Chromosome	Start	End	Biotype
1	ENSG00000183734	ASCL2	chr11	2268495	2270952	protein_coding




TF	Tissue	No. of peaks (total/average)	No. of peaks in gene body (total/average)	No. of peaks around TSS (total/average)	The peak close to TSS	The peak with strongest signal
CTCF	Aortic adventitial	1 / 1	1 / 1	0 / 0	Chr11,2269980,2270117,5.18,835,gb,dataset-676	Chr11,2269980,2270117,5.18,835,gb,dataset-676

CCCTC-
binding
factor



CTCF



hTFtarget

hTFtarget TF Target **Peak** Co-regulation Co-association

Cell line/condition:

22RV1



TF:

AR



 Visualize

Reset

User could select interested cell line(s)/condition(s) and TFs in the boxes (and datasets), and then view ChIP-Seq peaks by the "Visualize" function.



 View peaks ?

Dataset

TF

Cell line

Tissue

Disease

Dataset-36

AR

22RV1

Prostate

Carcinoma

Dataset-37

AR

22RV1

Prostate

Carcinoma

Dataset-78

AR

22RV1

Prostate

Carcinoma

Androgen Receptor: mediating the biological effects of androgen hormones, such as testosterone



L.B.B.

Laboratory of Systems Biology
and Bioinformatics

hTFtarget



hTFtarget

Datab

hTFtarget

TF

Target

Peak

Co-regulation

Co-association

Pro

Please input TFs in gene symbol format, se

MYC, GATA1

MYC is a proto-oncogene that encodes the c-MYC protein, a transcription factor involved in the regulation of cell proliferation, growth, differentiation, and apoptosis.

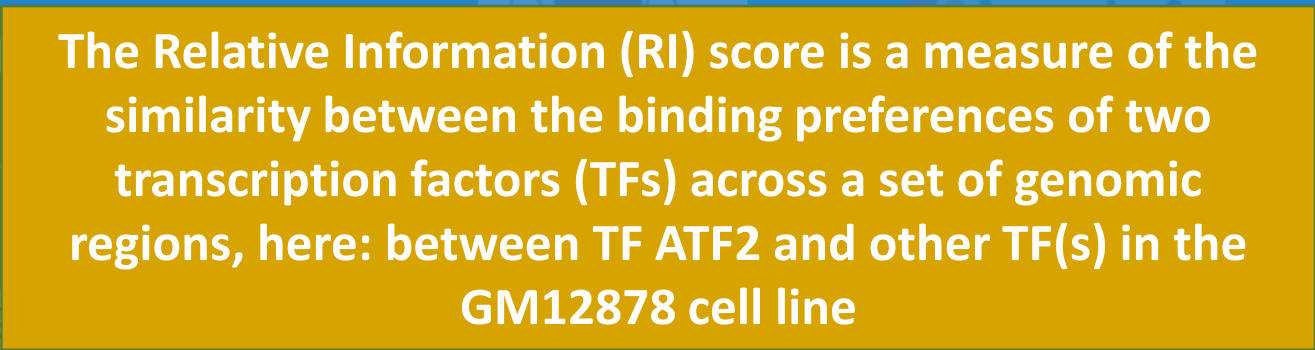
GATA1 is a transcription factor belonging to the GATA family of zinc-finger proteins, which are characterized by their ability to bind to DNA sequences containing the GATA motif




Gene id	Gene name	Chromosome	Start	End
ENSG00000110514	MADD	Chr11	47269161	47330031

Submit

Reset

Download



Sample Information: dataset-2692 RELA	
ID	2692
TF	RELA
Project	GSE31477 
Cell Line	GM12878
Cell Type	lymphoblastoid cell
Tissue	blood
Samples	GSM935478 
Control	GSM935564 
Targets?	details



Database of Human Transcription Factor Targets

[hTFtarget](#) [TF](#) [Target](#) [Peak](#) [Co-regulation](#) [Co-association](#) **[Prediction](#)** [Document](#) [Contact](#) [Download](#)

>SEQ1

ATCGATCGACTAGCTAGCTAGCTACGATCGATCGATCAGCTGACTGACTAGCTACGTCG

TF	Pattern name	Source	Sequence name	Start	Stop	Strand	Score	P value	Q value	Matched motif
AP1		database	SEQ1	39	49	+	11.8265	4.26e-05	0.00418	GCTGACTGACT
AP4		database	SEQ1	35	44	-	11.9633	5.36e-05	0.00536	GTCAGCTGAT
FOS		database	SEQ1	38	49	+	11.0678	6.96e-05	0.00668	AGCTGACTGACT
AP1		database	SEQ1	39	49	+	11.3061	7.11e-05	0.00696	GCTGACTGACT
AP1		database	SEQ1	39	49	+	10.7195	7.95e-05	0.00779	GCTGACTGACT



Advantages	Disadvantages
Comprehensive Coverage	Does not link to articles
Large Dataset	Co-association information is incomplete
Diverse Conditions	



L.B.B.


Laboratory of Systems Biology
and Bioinformatics

Tfcancer



JOURNAL ARTICLE

Tfcancer: a manually curated database of transcription factors associated with human cancers FREE

Qingqing Huang, Zhengtang Tan, Yanjing Li, Wenzhu Wang, Mei Lang, Changying Li, Zhiyun Guo  Author Notes

Bioinformatics, Volume 37

<https://doi.org/10.1093/bi>

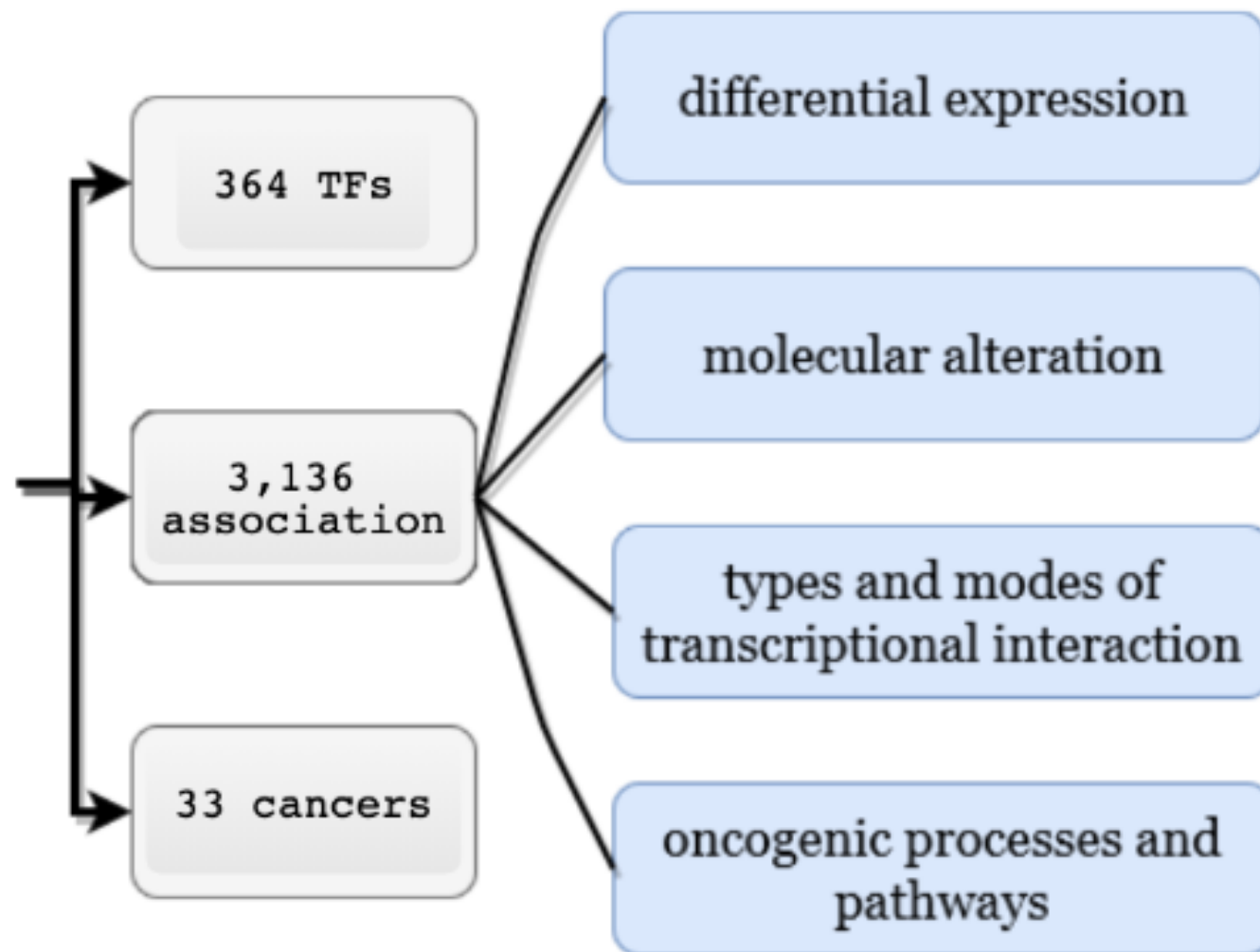
Published: 26 May 2021

China

- Contains 3136 experimentally supported associations between 364 TFs and 33 TCGA cancers.
- Curated through the manual examination of over 1800 literature sources.
- Offers a user-friendly interface for browsing and searching.
- Allows flexible data downloading.
- Facilitates user data submission, promoting collaboration and data sharing among researchers.



TFcancer





TF

EGR1

Early Growth Response 1

Select a Cancer Type

Cholangiocarcinoma

Reset

Search



Mastermind Like Transcriptional
Coactivator 1

<input type="checkbox"/>	Cancer	TF	Characteristics	Gene	Regulation type	Processes & Pathways	PMID	Details
<input type="checkbox"/>	KIRC	EGR1	targeted by	MAML1	positive	tumor progression	23029358	details....



Select a Cancer Type

Breast invasive carcinoma

TF FOXA1 Forkhead box protein A1

Select TF Characteristics

Select a tf characteristic

Gene Samd3

Hallmark apoptosis

Select Regulation Modes

Select a regulation type

positive
phosphorylation
negative
promoter binding
nuclear translocation

targeted by
regulate
high expression
DNA binding site methylation
assosiate with
differential binding

Reset **Search**

<input type="checkbox"/>	Cancer	TF	Characteristics	Gene	Regulation type	Processes & Pathways	Pmid	Details
<input type="checkbox"/>	BRCA	FOXA1	regulate	Smad3	negative; nuclear translocation	apoptosis	30206966	details....
<input type="checkbox"/>	BRCA	FOXA1	regulate	Smad3	negative; nuclear translocation	apoptosis	30206966	details....



A

AHR ALX4 AR ARNT ASCL1 ATF2 ATF3 ATF4 ATF5 ATF6 ARNTL ARNTL2 AIRE ATF1 ARID3B

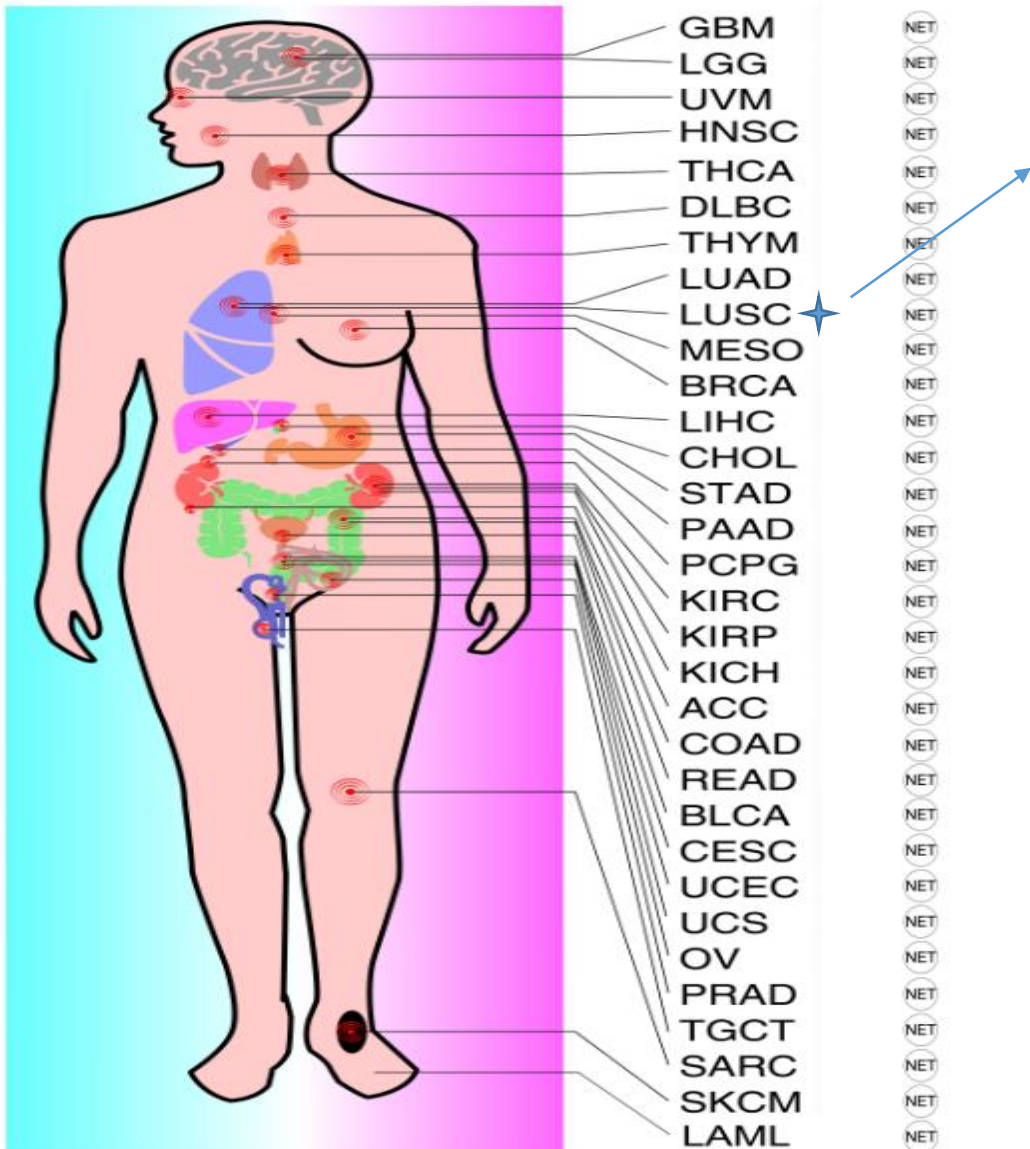
<input type="checkbox"/>	Cancer	TF	Characteristics	Gene	Regulation type	Processes & Pathways	PMID	Details
<input type="checkbox"/>	BRCA	AR	polymorphism; CAG repeat length	N/A	N/A	N/A	10817350	details....
<input type="checkbox"/>	BRCA	AR	N/A	N/A	N/A	proliferation	12203367	details....
<input type="checkbox"/>	BRCA	AR	N/A	N/A	N/A	hypoxia signaling; ER signaling pathway	17140257	details....
<input type="checkbox"/>	BRCA	AR	regulate	E-cadherin	positive	bone metastasis	25447306	details....
<input type="checkbox"/>	BRCA	AR	high expression; targeted by	HER2	positive	N/A	12912973	details....
<input type="checkbox"/>	BRCA	AR	regulate	MMP9; TIMP1; TIMP3	positive	tumor growth; invasion	16636675	details....



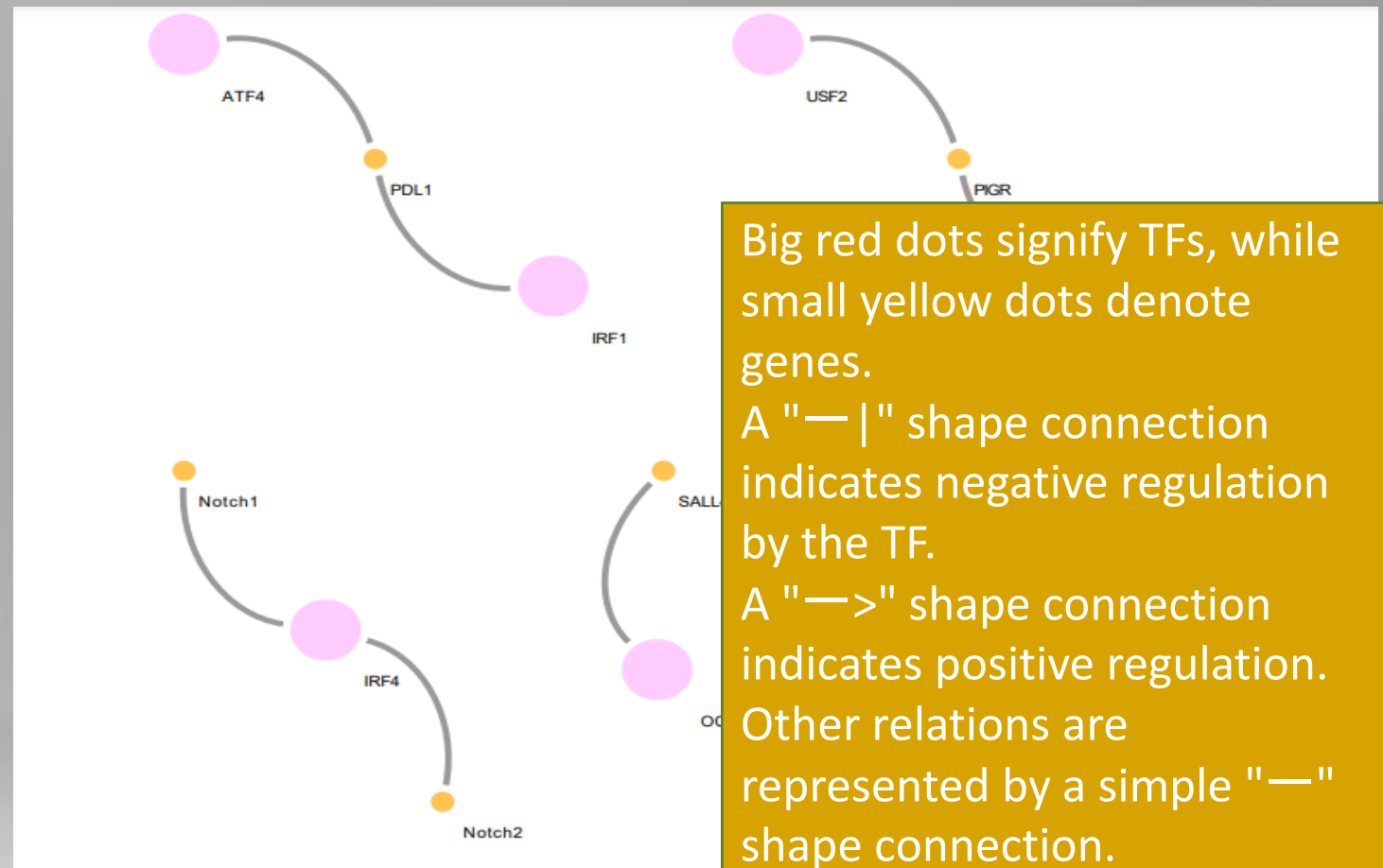
L.B.B.

Laboratory of Systems Biology
and Bioinformatics

Tfcancer-By Cancer



<input type="checkbox"/>	Cancer	TF	Characteristics	Gene	Regulation type	Processes & Pathways	PMID	Detail
<input type="checkbox"/>	LUSC	NR1H2	low expression	N/A	N/A	N/A	27335465	details
<input type="checkbox"/>	LUSC	NR1H3	low expression	N/A	N/A	N/A	27335465	details





Welcome to submit new entries for Tfcancer.

PubMed ID	Please enter a PMID (e.g. '546782')
Cancer	Please enter a cancer type (e.g. 'Breast invasive carcinoma')
TF	Please enter a TF gene symbol (e.g. 'BCL6' or 'ENSG00000113916.18')
Characteristics	Please enter a TF characteristics (e.g. 'high expression')
Gene	Please enter a gene name (e.g. 'GADD45A')
Regulation type	Please enter a regulation mode (e.g. 'positive')
Processes and Pathways	Please enter the cancer hallmark (e.g. 'tumorigenesis')
Original text	Original descriptions of the association users input above in the article.
Your email	

Submit



Advantages	Disadvantages
Comprehensive Coverage of TFs in human cancers	Incomplete coverage of all TF-cancer associations
Manual curation ensures accuracy and reliability	Potential for bias in manual curation process
Diverse Conditions	Limited functional annotations or mechanistic insights



JOURNAL ARTICLE

Factorbook: an updated catalog of transcription factor motifs and candidate regulatory motif sites

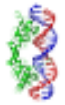


Henry E Pratt, Gregory R Andrews, Nishigandha Phalke, Jack D H
Arjan van der Velde, Jill E Moore, Zhiping Weng ✉ Author No

Nucleic Acids Research, Volume 50, Issue D1, 7 January 2022, Pa

Worcester, MA, USA

- Factorbook expansion: Includes more cell types and TFs.
- Improved motif catalog: Adds motifs from thousands of experiments.
- New machine learning tools: Simplify motif model use.
- Variant annotation and disease analysis: Helps assess heritability and annotate variants.

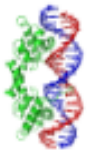


FACTORBOOK MAIN TF SEARCH



Search Human TFs

2688 experiments · 881 transcription factors · 170 cell types



A highly conserved zinc finger protein that functions as a transcriptional regulator and insulator protein in the genome, CTCF binds to specific DNA sequences known as CCCTC-binding motifs .

CTCF

Sequence-specific TF

chr16:67,562,467-67,639,177

[ENCODE](#)[Ensembl](#)[GO](#)[GeneCards](#)[HGNC](#)[RefSeq](#)[UCSC Genome Browser](#)[UniProt](#)[Wikipedia](#)



L.B.B.

Laboratory of Systems Biology
and Bioinformatics

Factorbook



Function

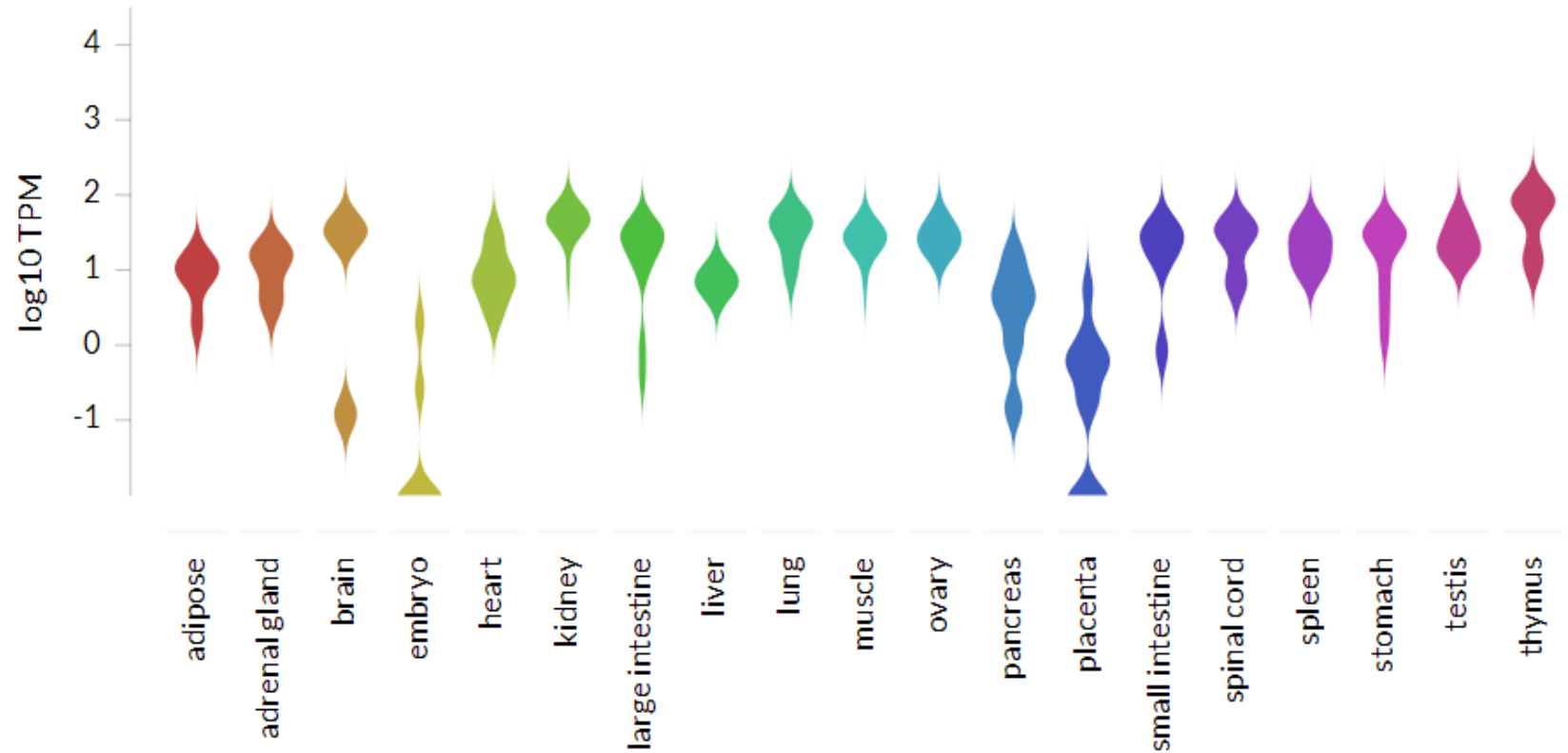
Expression (RNA-seq)

Measure the abundance of RNA transcripts in a sample, providing insights into gene expression levels.

Search CTCF peaks by region

Visualizes the expression pattern of transcription factor across different tissues or conditions, with the X-axis representing tissues or conditions and the Y-axis representing the expression level on a logarithmic scale.

The log10 transformed transcripts per million (TPM) values





Function

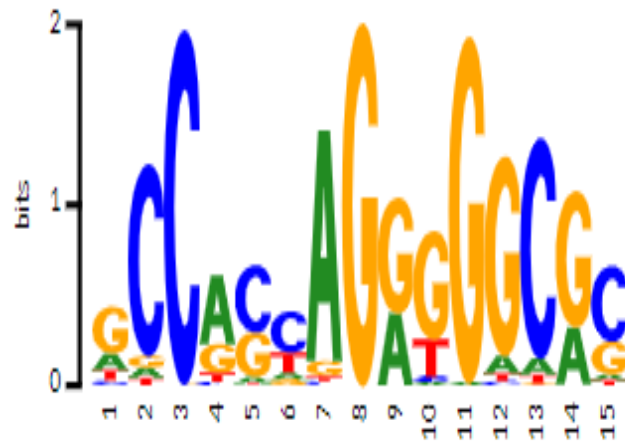
Expression (RNA-seq)

Motif Enrichment (MEME, ChIP-seq)

Motif Enrichment (SELEX)

Epigenetic Profile

Search CTCF peaks by region



Best external database match:
CTCF_MOUSE.H11MO.0.A (HOCOMOCO)

E-value provides a quantitative measure of the likelihood

How many times the motif or sequence pattern was identified within the set of genomic regions being analyzed

GRCh38, also known as hg38, is the latest version of the human reference genome assembly

? E-value < 1e-300

? Occurrences 50,581 / 87,157 peaks

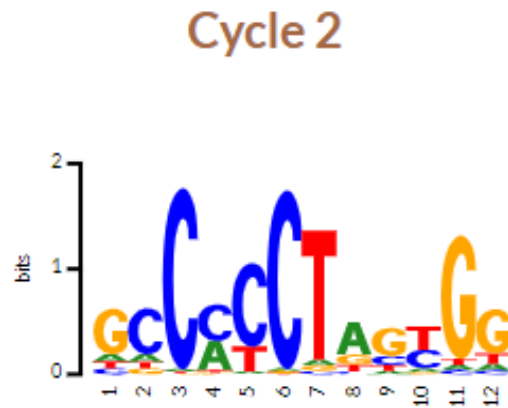
Download Peak

? Sites
(hg38 genome) 



Function Expression (RNA-seq) Motif Enrichment (MEME, ChIP-seq) **Motif Enrichment (SELEX)** Epigenetic Profile Search CTCF peaks by region

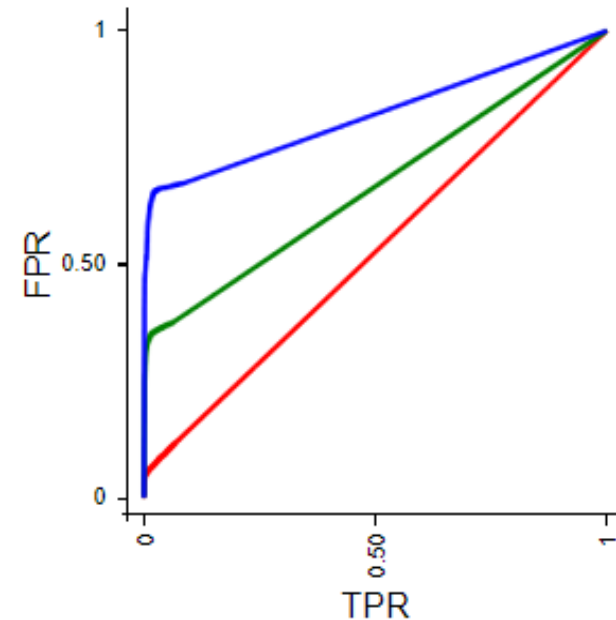
It is a laboratory technique used to identify specific nucleic acid sequences (DNA or RNA) that bind to a particular protein of interest.



↓ export motif (MEME)

↓ Export Logo

↔ Reverse Complement

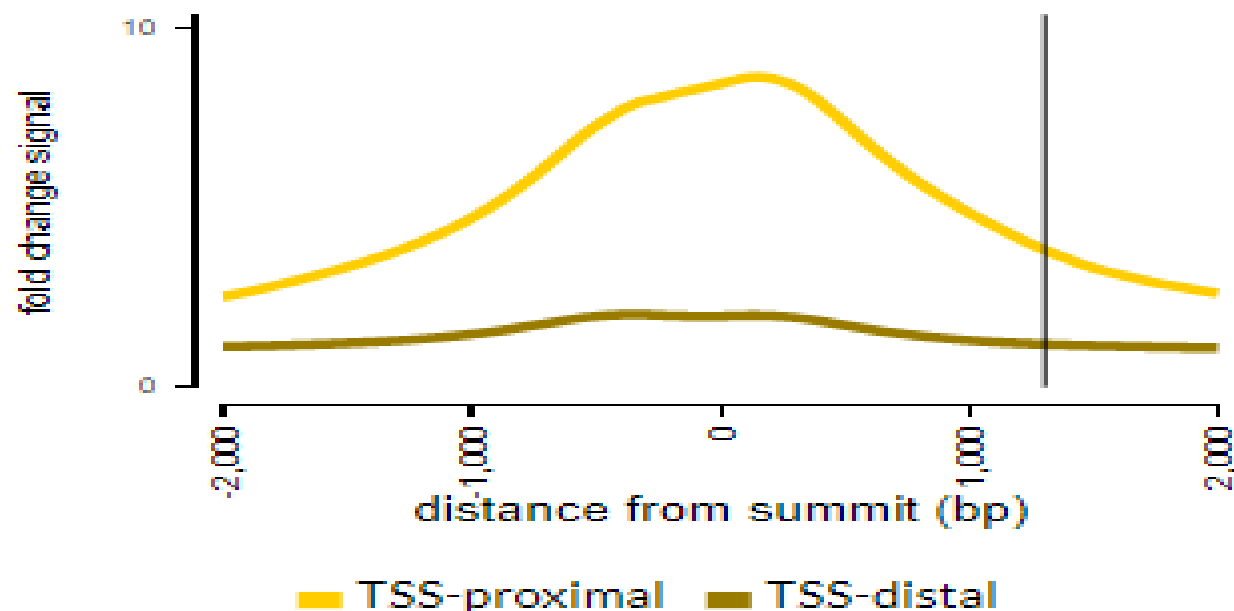




Function Expression (RNA-seq) Motif Enrichment (MEME, ChIP-seq) Motif Enrichment (SELEX) **Epigenetic Profile** Search CTCF peaks by region

H3K27ac

Histone modification involving the acetylation of lysine 27 on histone H3.



Fold change is a measure of the relative increase or decrease in signal intensity between the experimental condition (e.g., treatment or cell type) and the control condition.



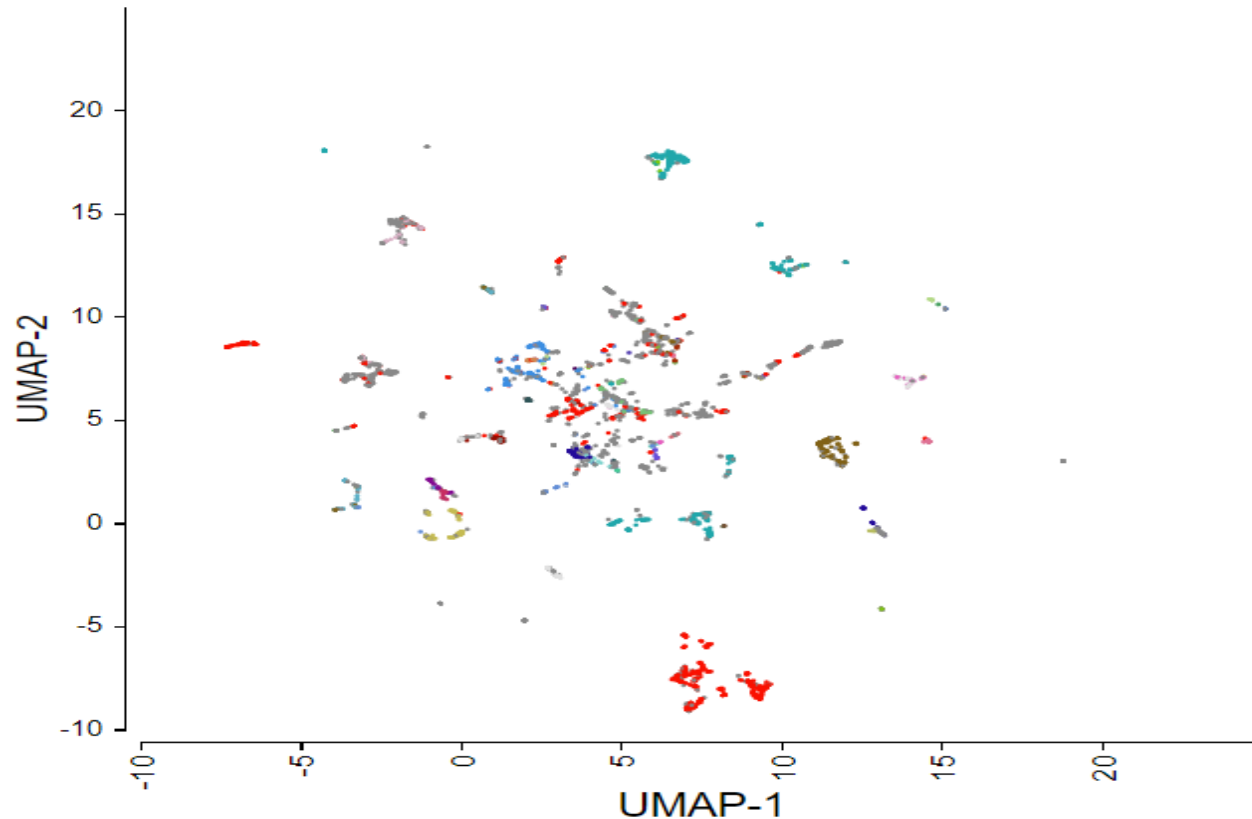
MOTIF AND MOTIF SITE CATALOG

SEARCH THE MOTIF CATALOG

Enter a consensus sequence or regex:

ccascagrgggcgd

Collection of information regarding the presence and characteristics of specific DNA sequence motifs within the genome or a set of genomic regions.



UMAP (Uniform Manifold Approximation and Projection).

UMAP-1 represents the first dimension (or component) resulting from the UMAP projection. UMAP-2 represents the second dimension resulting from the UMAP projection.



Advantages	Disadvantages
Comprehensive coverage of TF binding motifs	Lack of clear analysis
Integration of in vitro and in vivo profiling data	Lack of guidance in visualization
Novel tools for integrative analysis	



L.B.B.


Laboratory of Systems Biology
and Bioinformatics

PlantPAN



JOURNAL ARTICLE

PlantPAN 4.0: updated database for identifying conserved non-coding sequences and exploring dynamic transcriptional regulation in plant promoters

Chi-Nga Chow, Chien-Wen Yang, Nai-Yun Wu, Hun
Yu-Hsuan Chiu, Tzong-Yi Lee, Wen-Chi Chang 

Nucleic Acids Research, Volume 52, Issue D1, 5 Jan

<https://doi.org/10.1093/nar/gkad945>

Published: 28 October 2023 **Article history** ▼

- Integrative resource for building regulatory networks in 115 plant species.
- Expanded gene annotation and promoter sequences.
- Identification of conserved non-coding sequences.
- Updated repository with 3428 TF binding site matrices.
- Improved statistical analysis for ChIP-seq data.
- Easy-to-read experimental condition clusters.
- Peak visualization for all regulatory factors.

National Cheng Kung University ,Taiwan



L.B.B.

Laboratory of Systems Biology
and Bioinformatics

PlantPAN



Choose species

Please choose a model plant of interest:

Arabidopsis thaliana

Submit

or Quick Select:



Arabidopsis thaliana



Solanum lycopersicum



Gossypium

Input Gene ID or keyword

Please enter Gene ID, Locus or keyword to search:

☒ Gene ID/Locus

☐ Keyword

AT1G15820.1

Search

Reset

Go To PCBase

Selected

- Species: *Arabidopsis thaliana*
- Function: Gene Search - Gene ID
- Query: AT1G15820.1

Results: 1 record

Gene ID	Description	Symbol
AT1G15820.1	light harvesting complex photosystem II subunit 6	LHCB6



Please choose a function of interest:

Explore a TF

Explore a TFBS by keyword or ID

Explore a TFBS by sequence



Explore a TF of interest:

- ☒ TF Name
- ☐ TF Locus
- ☐ UniProt ID

AGL15

Search

Selected

- **Function:** Explore a TF
- **TF Name:** AGL15

Results: 1 record

TF ID	TF family	Species
AT5G13790	MADS box; MIKC	<i>Arabidopsis thaliana</i>

These proteins are characterized by the presence of a highly conserved DNA-binding domain(60 amino acids) known as the MADS-box domain.



Choose species

Please choose a model plant of interest:

Arabidopsis thaliana

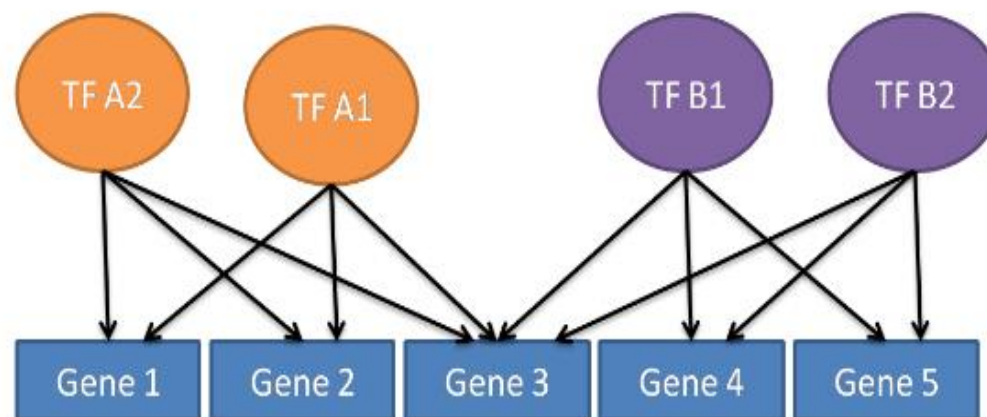
or Quick Select:



Arabidopsis thaliana

Selected

- Species: *Aegilops tauschii*



Step 1

Please input a group of transcript IDs and separate with ";

>AET6Gv20916900.1;AET7Gv20385300.2;AET4Gv20629900.1;AET7Gv20535700.3;AET4Gv20030000.2



Step 2 Set parameters

- How many co-occurrence TFBSs you want to analyze?

☒ 1 ☐ 2

- Input the threshold for analysis

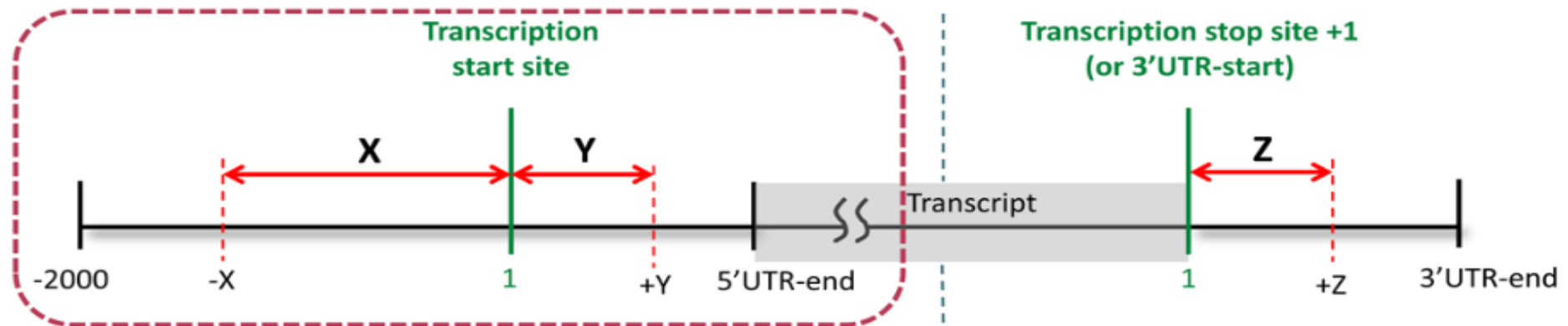
Support >= % (the frequency of promoters containing the TF/TFBS)

Individual TFBSs (1) or pairs of TFBSs (2) within genomic regions

Step 3 Please customize upstream and downstream coordinates of promoter

- Choose transcription start site/5'UTR-End or transcription stop site/3'UTR-End

☒ transcription start site/5'UTR-End ☐ transcription stop site/3'UTR-End



X: (MAX=2000) / Y: ☒ (MAX=500) or ☐ 5'UTR-End



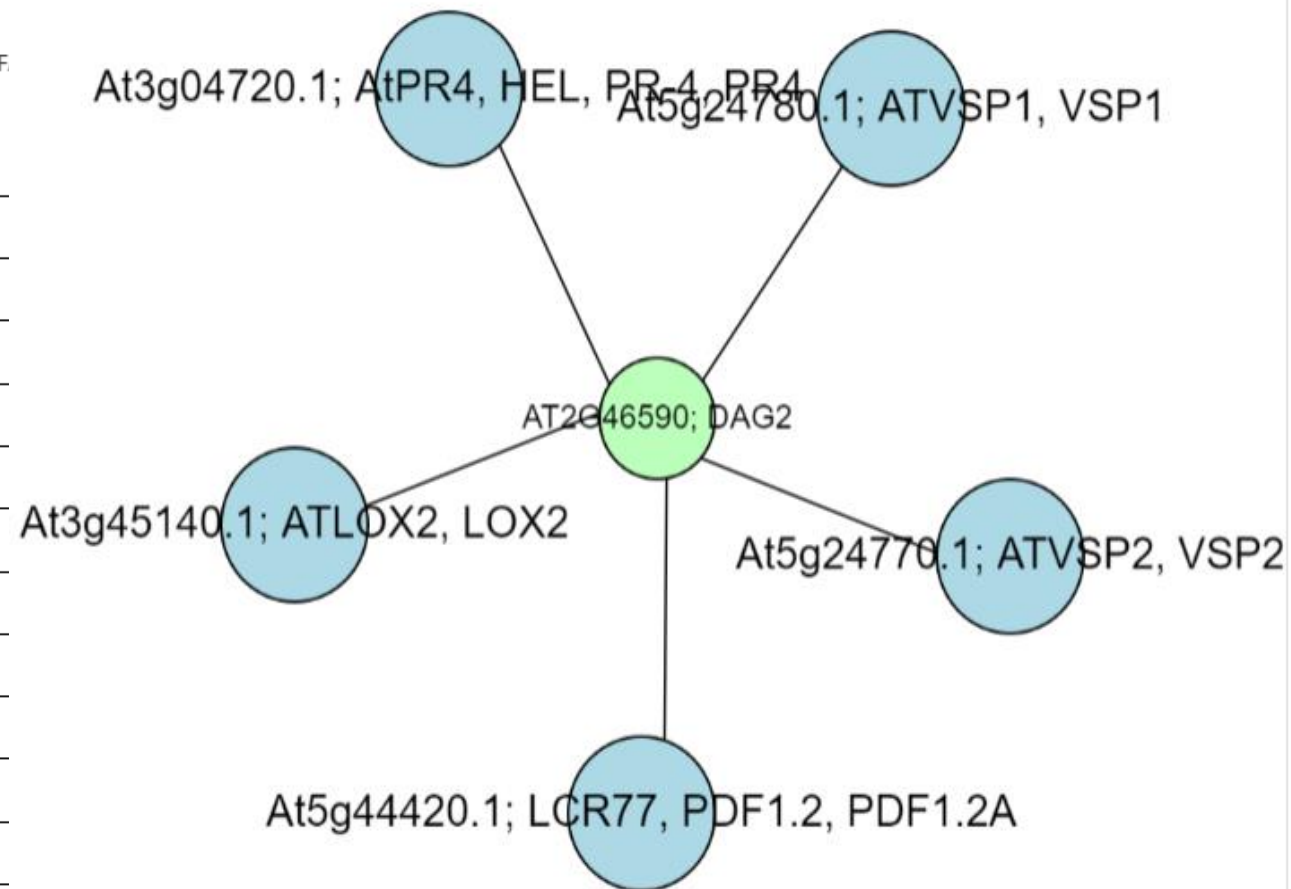
Gene Group

ALL

TRANSCRIPTION FACTOR BINDING SITE WITH TRANSCRIPTION F

Transcription Factor Binding site with Transcription Factor

TFBS ID	TF Family of TFBS
TFmatrixID_0221	CG-1; CAMTA
TFmatrixID_0220	CG-1; CAMTA
TFmatrixID_0199	bZIP
TF_motif_seq_0237	GATA; tify
TF_motif_seq_0239	Dof
TFmatrixID_0174	bHLH
TF_motif_seq_0241	ZF-HD
TF_motif_seq_0243	GATA; tify
TF_motif_seq_0244	SBP
TF_motif_seq_0246	Homeodomain; TALE
TF_motif_seq_0251	TCP
TF_motif_seq_0252	Myb/SANT; MYB; ARR-B





Promoter Analysis

🔗🔗🔗 For optimal use of this database, we recommend to use Google Chrome or Firefox browser and above at 1680 x 1050 resolution.

- If there are two or more promoters for scanning, please use [Multiple promoter analysis](#).
- If you want to use matrices from ChIP-seq dataset for scanning, please use [Promoter analysis in PCBase](#).

Step 1 Please input sequence name and the promoter sequence in FASTA format.

```
>promoter_sequence
GTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACG
GGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAA
GACGTATTTCCACCGGTGATCGGGGTTGCTGCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGA
TCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGGT
ACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGCAATAAAGC
CCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATTTGAT
GAAGGTAAACTGCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAACTGTTGCTCATGACTATGATGTCATAGTTAT
TGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGATGTGCTGCTGATGTGCTGATTGTTCCACGCCTGCTGAG
TTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGGGTTCGAGCC
TGATGTACGTATTTTGCTTACCAAATACAGCAATAGTAATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCC
TGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTT
GAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTT
CGATCGTCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCGTGCGCCTGTTATTCAAAACATACGCTCAATACTCAACC
GGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGATGGTGGATTGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGG
TAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTGAGAAGACC
TCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACT
GACTGGTCAACAGACACCGGCGTTTCGGTCAAGAGTATCTGGTGTGATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGC
TGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGT
AACGATTATCGCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTC
```



Step 2 Please specify TF binding motifs:

☒ PlantPAN 3.0 database

Please select transcription factors from the following species:

☐ All species

☒ Choose species

Arabidopsis thaliana

Brachypodium distachyon

Chlamydomonas reinhardtii

Glycine max

Malus domestica

Oryza sativa

Populus trichocarpa

Sorghum bicolor

Volvox carteri

Zea mays

Physcomitrella patens

☐ User-customized motifs

Input custom motif sequence (max=10, separate by line feed) with IUPAC code and group with "[]" (EX: [GC]SS[GC]GC);

NAWWWAN
RTWWWTR
ACGT

Step 3 Please select the optional promoter elements:

☒ Tandem Repeat

☒ CpNpG

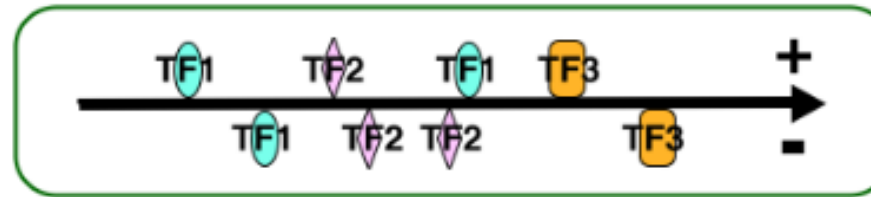
Search

Reset



— Visualization —

- **Visualize TF binding sites within the promoter region.**
Visualize TF binding sites within one promoter region.



Measures the randomness or disorder within the repeat sequence, with higher values indicating more variability.

— Tandem Repeat — TIP

Location	Period size	Copy number	Consensus size	Percent matches	Percent indels	Score	Number of the nucleotides				Entropy (0-2)	Seed
							A	T	C	G		
2082-2541	43	10.7	43	93	0	749	16	28	25	29	1.97	TCTGGGACCACGGTCCCACTCGTATCGTC GGTCTGATTATTAG
3319-3923	178	3.4	178	94	0	1059	33	15	22	28	1.95	CGAATATGACTTGATGTCATGTGTATGATT GAGTATAAGAACTTAAACCGCAACCCGAT CTTAAAAGCCTAAGTAGTGTTGCCTTGTTA GAAGACACAAAGCCAAAGACTCATATGGA

Insertions or deletions (indels)

A short, conserved sequence motif or pattern that serves as a starting point for identifying or detecting similar sequences or motifs within a larger dataset.



CpG Island — TIP

Begin site	End site	Length	G+C frequency	CpG o/e ratio	AT Skew	CG Skew	Start-p	Strand	Strand-p
1032	1580	549	0.49	1.07	-0.09	0.14	0.55	+	0.95
1904	3021	1118	0.5	1.12	-0.19	0.01	0.7	+	0.91

observed/expected
ratio

p-value associated
with.....

```

CACGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCCACTGTTCCAC      2100
GTGCAAAATCAGATGCAATAGACAGAAAATGAATTACAGGAAACAATGTCCGGTCTTTCTGATTGACCGGACTTATAAGAGAGACCCGGGTGACAAGGTG
*****
TTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTG      2200
AACATAGCAGCCAGACTATTAGTCTGACCTGGTGCCAGGGTGAGCATAGCAGCCAGACTAATAATCAGACCTGGTGCCAGGGTGAGCATAGCAGCCAG
*****
TGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTG      2300
ACTAATAATCAGACCTGGTGCCAGGGTGAGCATAGCAGCCAGACTATTAGTCTGACCTGGTGCCAGGGTGAGCATAGCAGCCAGACTAATAATCAGAC
*****
GGACCATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCA      2400
CCTGGTACCAGGGTGAGCATAGCAGCCAGACTAATAATCAGACCTGGTGCCAGGGTGAGCATAGCAGCCAGACTAATAATCAGACCTTGGTGCCAGGGT
*****
CTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTGTGCGGT      2500
GAGCATAGCAGCCAGACTAATAATCAGACCTGGTGCCAGGGTGAGCATAGCAGCCAGACTAATAATCAGACCTGGTGCTAGGGTGAGCACAACAGCCA
*****
CTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGAC      2600
GACTAATAGCCAGACCTGGTGCCAGGGTGAACATAACAGCTAGTCTGATAGTCGACTCTGATGCTAAGGTAGTTACGGACAGTTCCC GTTCATAACTG
*****

```




Pattern Search Results

- ☐ **TFmatrixID_0008** / AT-Hook (CpG Island)
- ☐ **TFmatrixID_0011** / NAC ; NAM
- ☐ **TFmatrixID_0012** / NAC ; NAM (CpG Island)
- ☐ **TFmatrixID_0015** / MYB ; ARR-B
- ☐ **TFmatrixID_0016** / MYB ; ARR-B
- ☐ **TFmatrixID_0018** / Myb/SANT ; MYB ; ARR-B
- ☐ **TFmatrixID_0019** / Myb/SANT ; MYB ; ARR-B
- ☐ **TFmatrixID_0020** / AP2 ; ERF (CpG Island, Tandem Repeat)
- ☐ **TFmatrixID_0021** / C2H2 (CpG Island)
- ☐ **TFmatrixID_0022** / Trihelix (CpG Island)
- ☐ **TFmatrixID_0024** / AP2 ; ERF (CpG Island)
- ☐ **TFmatrixID_0025** / Homeodomain ; bZIP ; HD-ZIP (Tandem Repeat)
- ☐ **TFmatrixID_0028** / bZIP (CpG Island)
- ☐ **TFmatrixID_0029** / MYB-related
- ☐ **TFmatrixID_0032** / AP2 ; ERF (CpG Island)
- ☐ **TFmatrixID_0033** / AP2 ; ERF (CpG Island)
- ☐ **TFmatrixID_0034** / AP2 ; ERF (CpG Island, Tandem Repeat)
- ☐ **TFmatrixID_0035** / AP2 ; ERF (CpG Island, Tandem Repeat)
- ☐ **TFmatrixID_0037** / AP2
- ☐ **TFmatrixID_0038** / AP2 (CpG Island, Tandem Repeat)
- ☐ **TFmatrixID_0040** / B3 ; ARF (CpG Island, Tandem Repeat)
- ☐ **TFmatrixID_0041** / B3 ; ARF (CpG Island)

Gene name (TF name)	TF Family
AHL25; AGF1	AT-Hook

The AT hook family refers to a group of DNA-binding proteins that contain a characteristic DNA-binding motif known as the AT hook. This motif consists of a short peptide sequence that specifically recognizes and binds to regions of DNA rich in adenine (A) and thymine (T), which are the two nucleotides forming the DNA double helix.



Advantages	Disadvantages
<p>Full coverage and friendly user.</p>	<p>Its only for plants.</p>



L.B.B.

Laboratory of Systems Biology
and Bioinformatics

THANKS FOR YOUR ATTENTION

Enter your sub headline here



**Thanks
For
Your
Attention**