

---

# JASPAR: an open-access database of transcription factor binding profiles

**Dr. Oriol Fornes**

Deputy Group Leader @ Wasserman Lab  
Centre for Molecular Medicine and Therapeutics  
BC Children's Hospital Research Institute  
University of British Columbia



[oriol@cmmmt.ubc.ca](mailto:oriol@cmmmt.ubc.ca)



@OForbes

---

**GREEKC Meeting – February 13, 2018**

# Outline

---

- **Overview**
- **Manual Curation**
- **Tools:**
  - **Profile Inference**
  - **Matrix Clustering**
  - **Genome Tracks**
- **New Web Interface**
- **RESTful API**
- **Perspectives**
- **MANTA2**

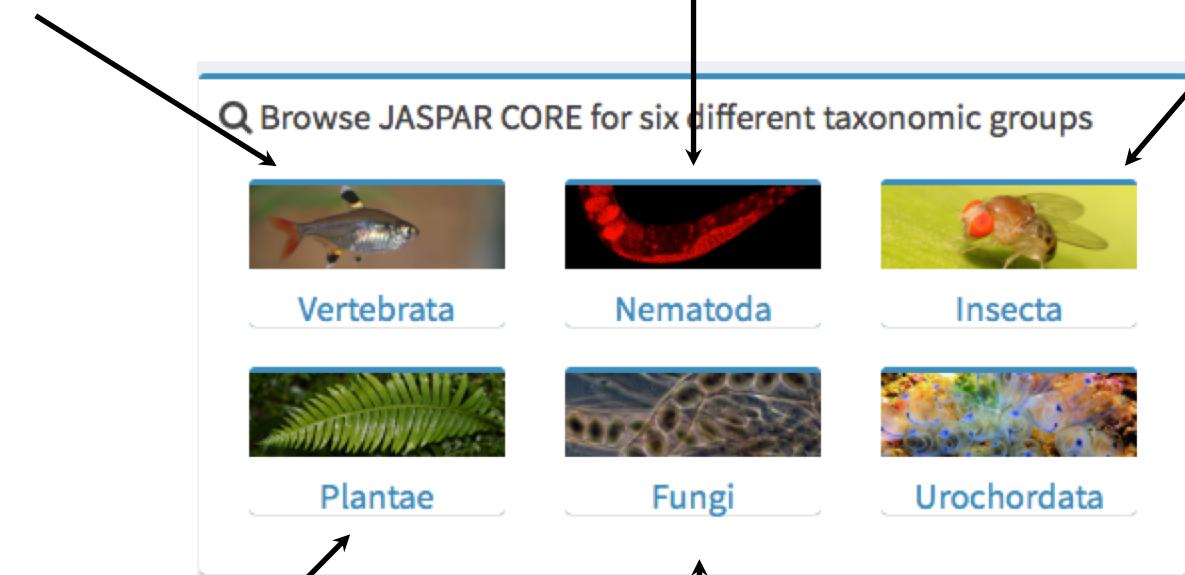


# Overview

human, mouse, rat,  
zebra fish...

*C. elegans*

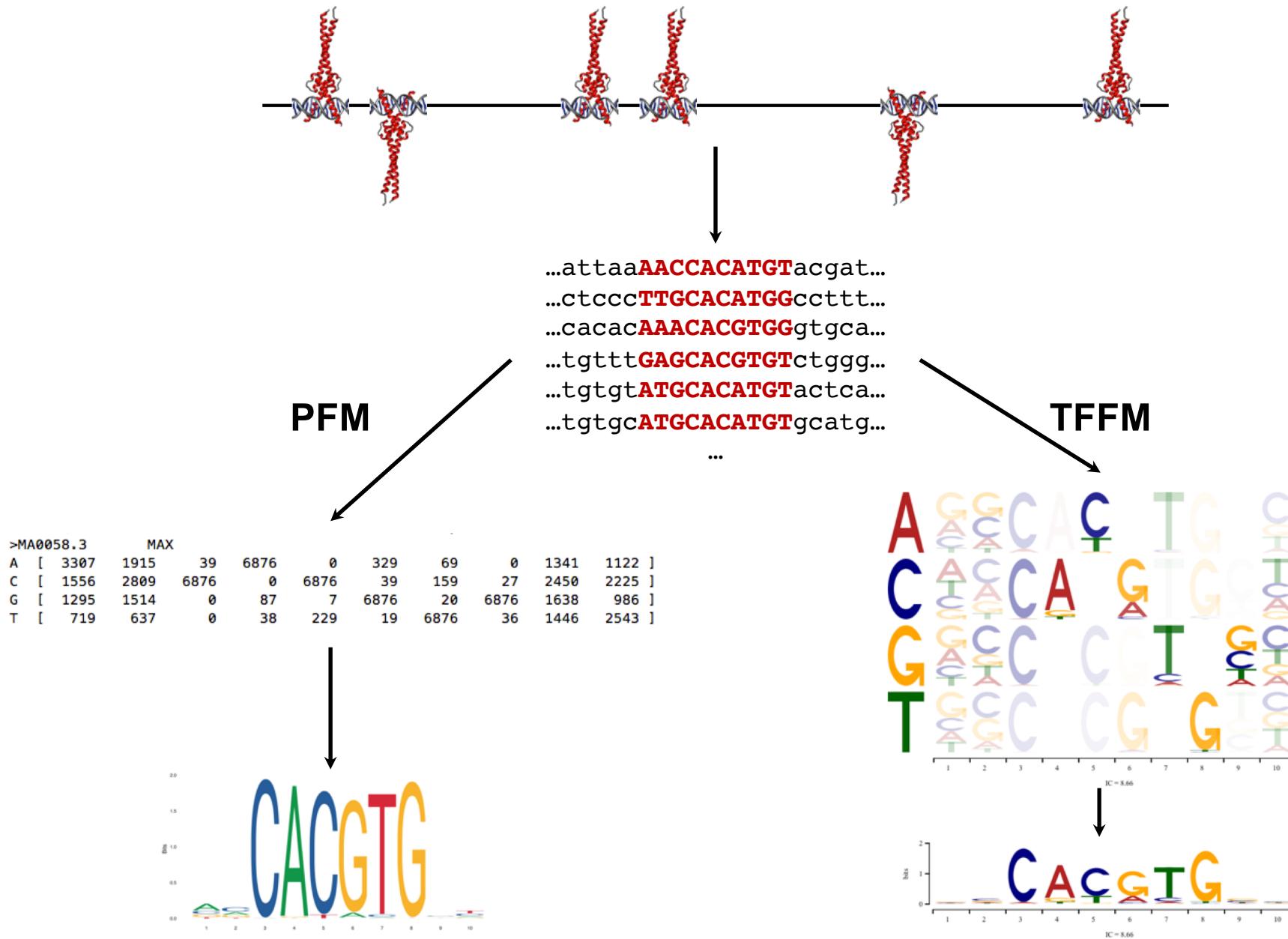
*Drosophila*



*Arabidopsis*

yeast

# Overview



## Table Of Contents

Welcome to TFFM's documentation!  
System requirements  
Contents  
Tutorial  
Download  
Reference  
Licence  
Indices and tables

## Next topic

tffm Module

## This Page

Show Source

## Quick search

 Go

Enter search terms or a module,  
class or function name.

## Welcome to TFFM's documentation!

We provide here the documentation of the TFFM-framework developed in [Python](#). The **Transcription Factor Flexible Models (TFFMs)** represent TFBSs and are based on hidden Markov models (HMM). They are flexible and are able to model both position interdependence within TFBSs and variable length motifs within a single dedicated framework.

The framework also implements methods to generate a new graphical representation of the modeled motifs that convey properties of position interdependences.

TFFMs have been assessed on ChIP-seq data sets coming from the [ENCODE project](#), revealing that the new HMM-based framework performs, in most cases, better than both PWMs and the dinucleotide weight matrix (DWM) extension in discriminating motifs within ChIP-seq sequences from background sequences. Under the assumption that ChIP-seq signal values are correlated with the affinity of the TF-DNA binding, we find that TFFM scores correlate with ChIP-seq peak signals. Moreover, using available TF-DNA affinity measurements for the Max TF, we observe that TFFMs constructed from ChIP-seq data correlate with published experimentally measured DNA-binding affinities. These results demonstrate the capacity of TFFMs to accurately model DNA-protein interactions, while providing a single unified framework suitable for the next generation of TFBS predictions. All the details have been published in [Mathelier and Wasserman, The Next Generation of Transcription Binding Site Prediction, PLOS Computational Biology](#) , Sept. 2013, 9(9):e1003214, DOI:[10.1371/journal.pcbi.1003214](https://doi.org/10.1371/journal.pcbi.1003214).

TFFMs can be saved and opened from files using the XML format already used by the [GHMM library](#).

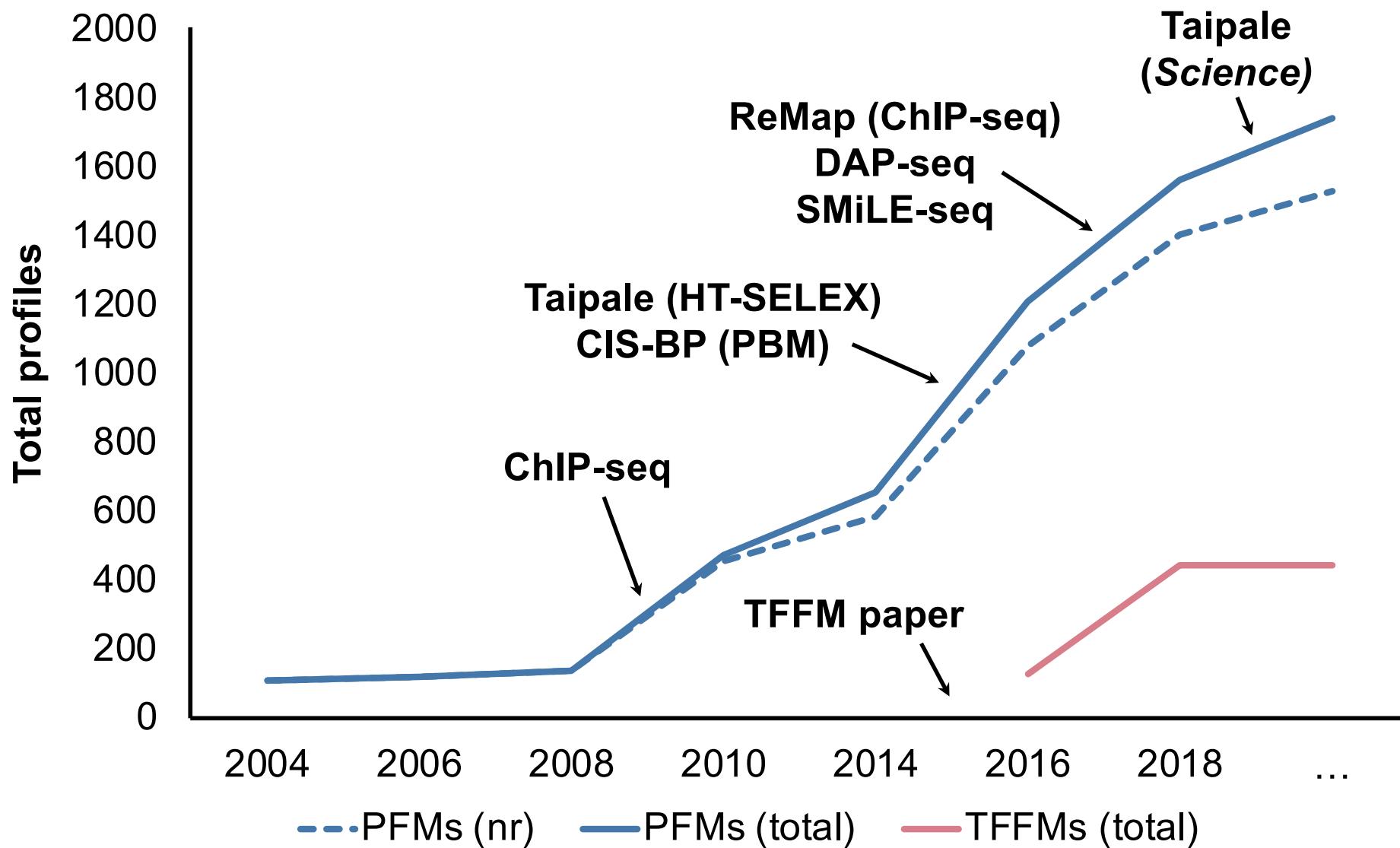
## System requirements

- The TFFM-framework 2.0 has been developed and tested under [Ubuntu Linux](#) operating system. It has also been tested on [CentOS](#).
- [Python](#) should be installed (version 2.7 has been used successfully).
- [Biopython](#) (at least version 1.61) should be installed and accessible from your Python executable. See <http://biopython.org> for instructions on how to install it.
- The [GHMM](#) library should be installed and accessible from Python. See <http://ghmm.org> for instructions on how to install it.

## Contents

- tffm Module
- hit Module
- drawing Module
- util Module

# Overview



# Overview

---

## The JASPAR database as of 2018:

Taxon	PFMs (nr)	PFMs (total)	TFFMs (total)	TF classes	TF families
Vertebrates	579	719	225	38	95
Plants	489	501	218	22	24
Insects	133	140	3	18	25
Nematodes	26	26	0	10	14
Fungi	176	177	0	19	11
Urochordata	1	1	0	1	0
<b>Total</b>	<b>1,404</b>	<b>1,564</b>	<b>446</b>	<b>64</b>	<b>115</b>

**Experimental data sources:** ChIP-chip/seq, DAP-seq, EMSA, PBM, (HT-)SELEX, SMiLE-seq...

# Outline

---

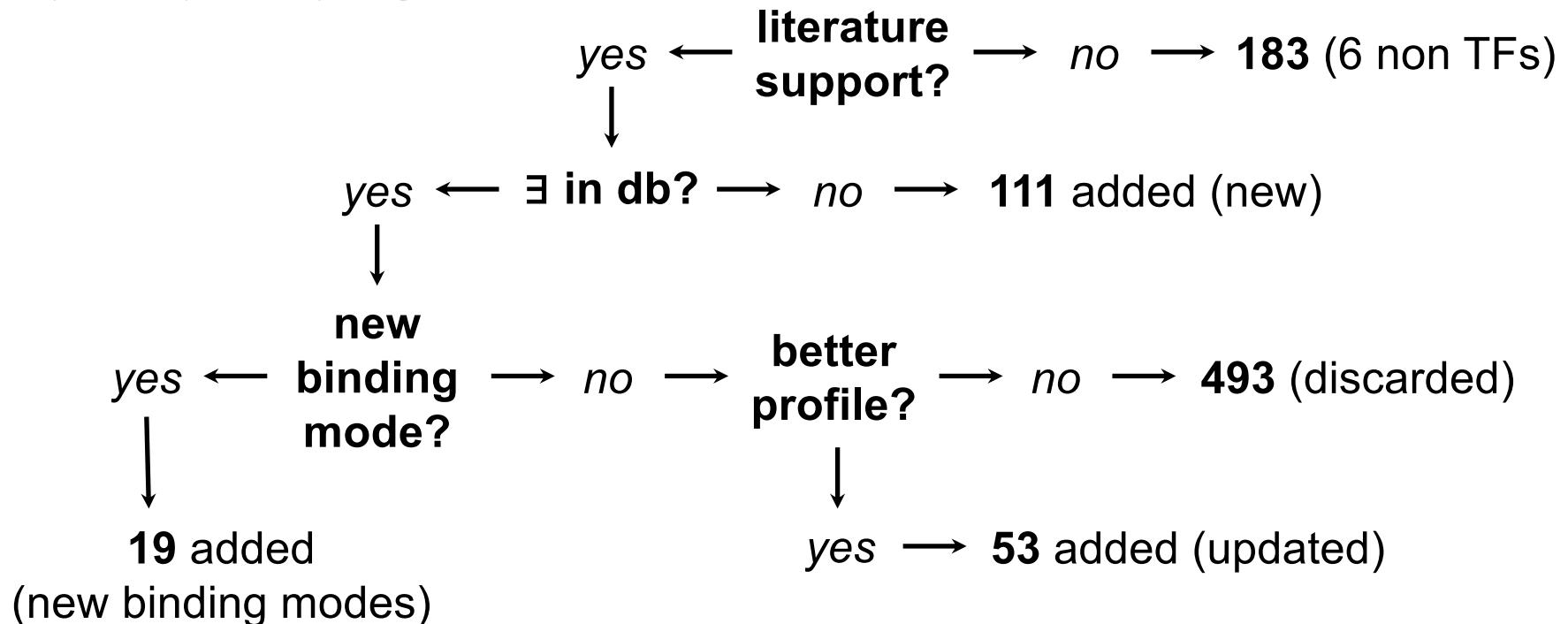
- Overview
- **Manual Curation**
- Tools:
  - Profile Inference
  - Matrix Clustering
  - Genome Tracks
- New Web Interface
- RESTful API
- Perspectives
- MANTA2

# Manual Curation

## Impact of cytosine methylation on DNA binding specificities of human transcription factors

Yimeng Yin, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, Teemu Kivioja, Kashyap Dave, Fan Zhong, Kazuhiro R. Nitta, Minna Taipale, Alexander Popov, Paul A. Ginno, Silvia Domcke, Jian Yan, Dirk Schübeler, Charles Vinson, Jussi Taipale\*

859 PFMs



# Manual Curation

---

## Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences

Michael F. Berger,<sup>1,3,8</sup> Gwenael Badis,<sup>5,8</sup> Andrew R. Gehrke,<sup>1,8</sup> Shaheynoor Talukder,<sup>5,8</sup> Anthony A. Philippakis,<sup>1,3,6</sup> Lourdes Peña-Castillo,<sup>4</sup> Trevis M. Alleyne,<sup>5</sup> Sanie Mnaimneh,<sup>4</sup> Olga B. Botvinnik,<sup>1,7</sup> Esther T. Chan,<sup>5</sup> Faiqua Khalid,<sup>4</sup> Wen Zhang,<sup>5</sup> Daniel Newburger,<sup>1</sup> Savina A. Jaeger,<sup>1</sup> Quaid D. Morris,<sup>4,5</sup> Martha L. Bulyk,<sup>1,2,3,6,\*</sup> and Timothy R. Hughes<sup>4,5,\*</sup>

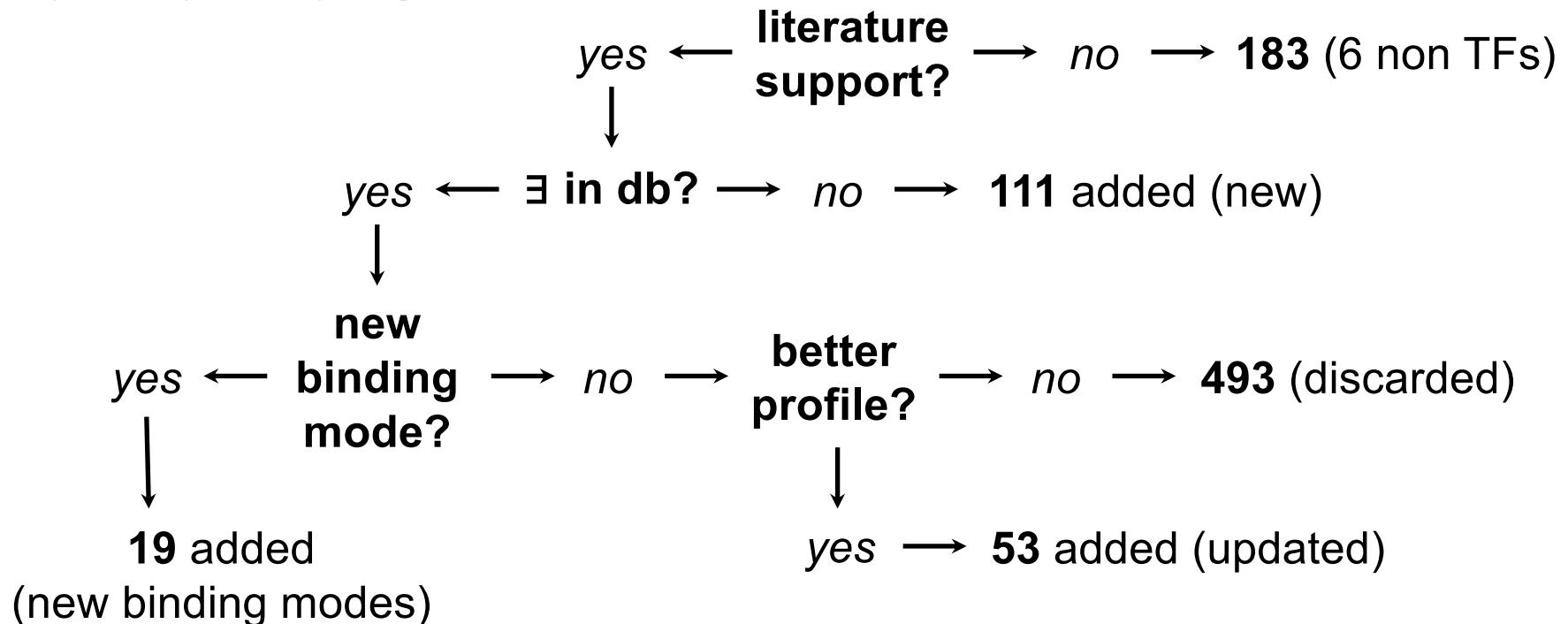
For 71 of the proteins we analyzed, there is no in vitro or in vivo binding site data, and for the majority, there is no PWM, in either mouse or the closest homolog in any species. To our knowledge, for several families, we describe a relatively uniform and apparently distinct binding profile for the first time. These encompass the **Irx family** (preferring sequences resembling **TACATGTA**), the **Obox family** (**GGGGATTAA**), the **Six family** [**G(G/A)TATCA**], **Gbx1/2 (CTAATTAG)**, and **Pknox1/2 (CCTGTCA)**. Our data also include individual proteins with apparently unique sequence preferences, including **Dux1 (CAATCAA)**, **Hdx [(C/A)AATCA]**, **Hmbox (TAACTAG)**, **Homez (ATCGTTT)**, and **Rhox11 [GCTGT(T/A)(T/A)]**. The variety in motifs we obtained motivated us to further explore the similarities and differences among homeodomains within our data set.

# Manual Curation

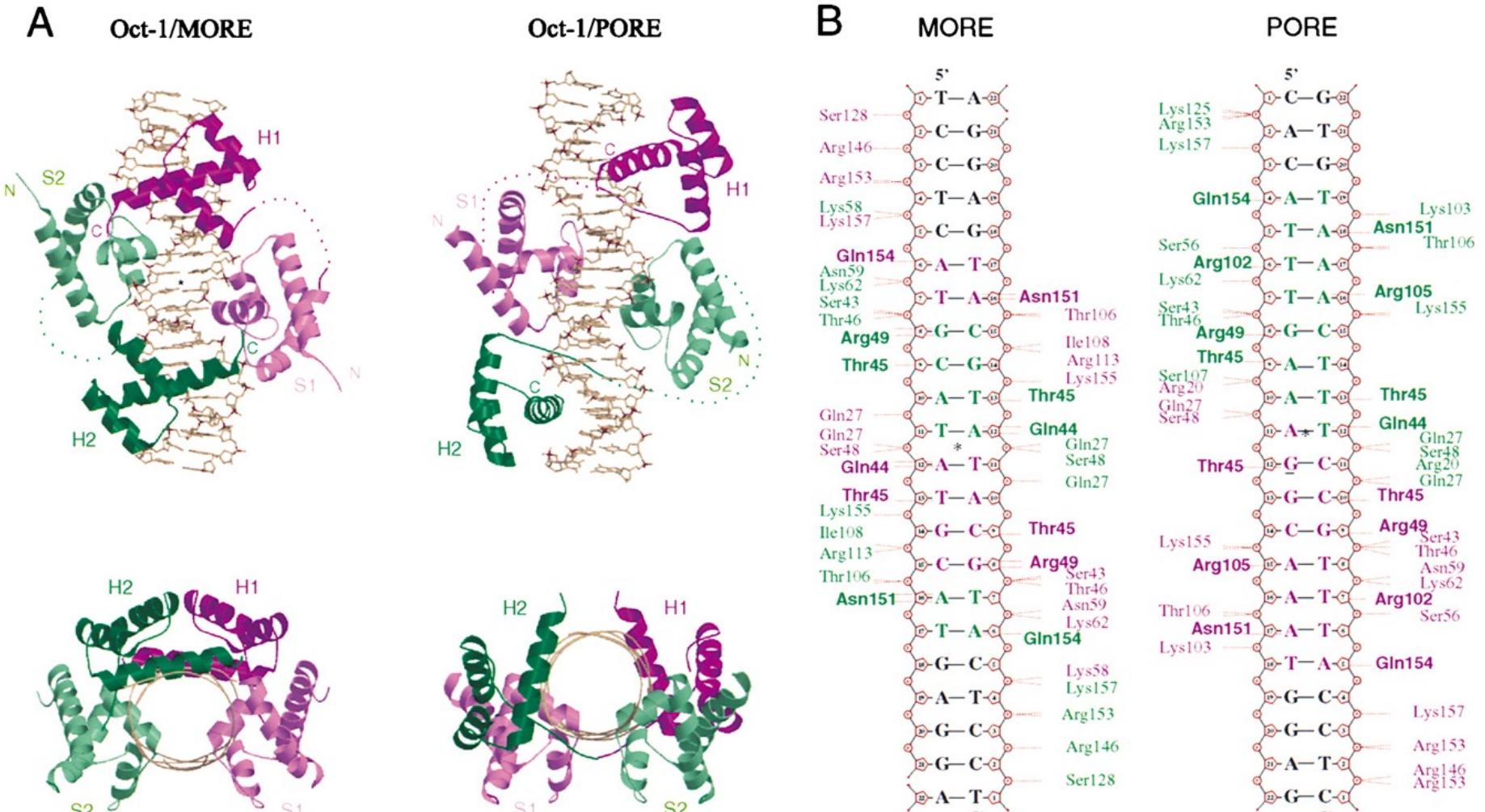
## Impact of cytosine methylation on DNA binding specificities of human transcription factors

Yimeng Yin, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, Teemu Kivioja, Kashyap Dave, Fan Zhong, Kazuhiro R. Nitta, Minna Taipale, Alexander Popov, Paul A. Ginno, Silvia Domcke, Jian Yan, Dirk Schübeler, Charles Vinson, Jussi Taipale\*

859 PFMs



# Manual Curation



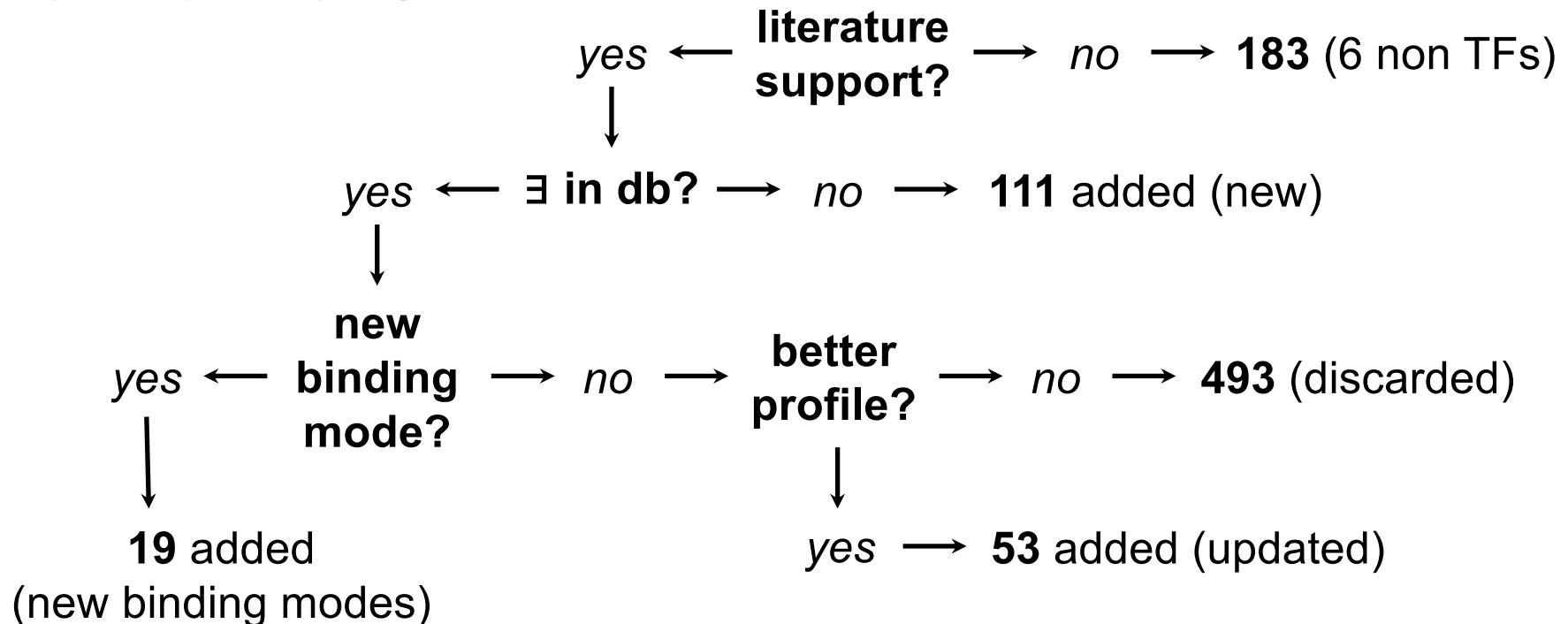
Reményi A, et al. (2001)

# Manual Curation

## Impact of cytosine methylation on DNA binding specificities of human transcription factors

Yimeng Yin, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, Teemu Kivioja, Kashyap Dave, Fan Zhong, Kazuhiro R. Nitta, Minna Taipale, Alexander Popov, Paul A. Ginno, Silvia Domcke, Jian Yan, Dirk Schübeler, Charles Vinson, Jussi Taipale\*

859 PFMs



# Manual Curation

J 18 Cart 0 JASPAR Blog

## JASPAR CORE Profile versions

Note: A logo is only shown when it first appears or changes with the exception that it always appears in the JASPAR 2016 column. This is to distinguish between the case when it has been removed from JASPAR vs. the case where it just didn't change since a previous release.

Display 10 profiles Search based on ID/Name:

ID	Name	JASPAR profile versions				
		2.1	2006	2008	2010	2014
MA0001	AGL3					
MA0002	RUNX1					
MA0003	TFAP2A					
MA0004	Arnt					
MA0005	AG					

# Outline

---

- Overview
- Manual Curation
- Tools:
  - **Profile Inference**
  - **Matrix Clustering**
  - **Genome Tracks**
- New Web Interface
- RESTful API
- Perspectives
- MANTA2

# Profile Inference

JASPAR 2018

Home

About

Search

Browse JASPAR CORE

Browse Collections

Tools

Profile Inference

Matrix Align

External tools

RESTful API

Download Data

Matrix Clusters

Genome Tracks



JASPAR Blog

## Profile Inference

Home > Profile Inference

### Profile Inference

Please input a TF protein sequence for which to look for a JASPAR TF binding profile.

Paste a protein sequence below

```
MQNSHSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRY  
YETGSIRPRAIGGSKPRVATPEVVKIAQYKRECPSPFAWEIRDRLLSEGVTNDNIPSV  
SSINRVLRLNASEKQQMGADGMYDKLRLQNRNTSGTGSWGRPGWYPGTSVPQGPTQDGCCQQQ  
EGGGENTNSISSNGEDSDEAQMRQLKRKLQRNRNTFTQEIQALEKEFERTHYPDVFAR  
ERLAALKIDLPEARIQVWFNSNRAKWRREEKLRNQRQASNTPSHIPISSSFSTSVDQPIP  
QPTTPVSSFTSGSMLGRTDTALTNTYSALPPMPSFTMANNLPMQPPVPSQTSSYSCMLPT  
SPSVNGRSYDTYTPPHMQTHMNSQPMGTSGTTSTGLISPVGVSVPVQVPGSEPDMSQYWPR  
LQ
```

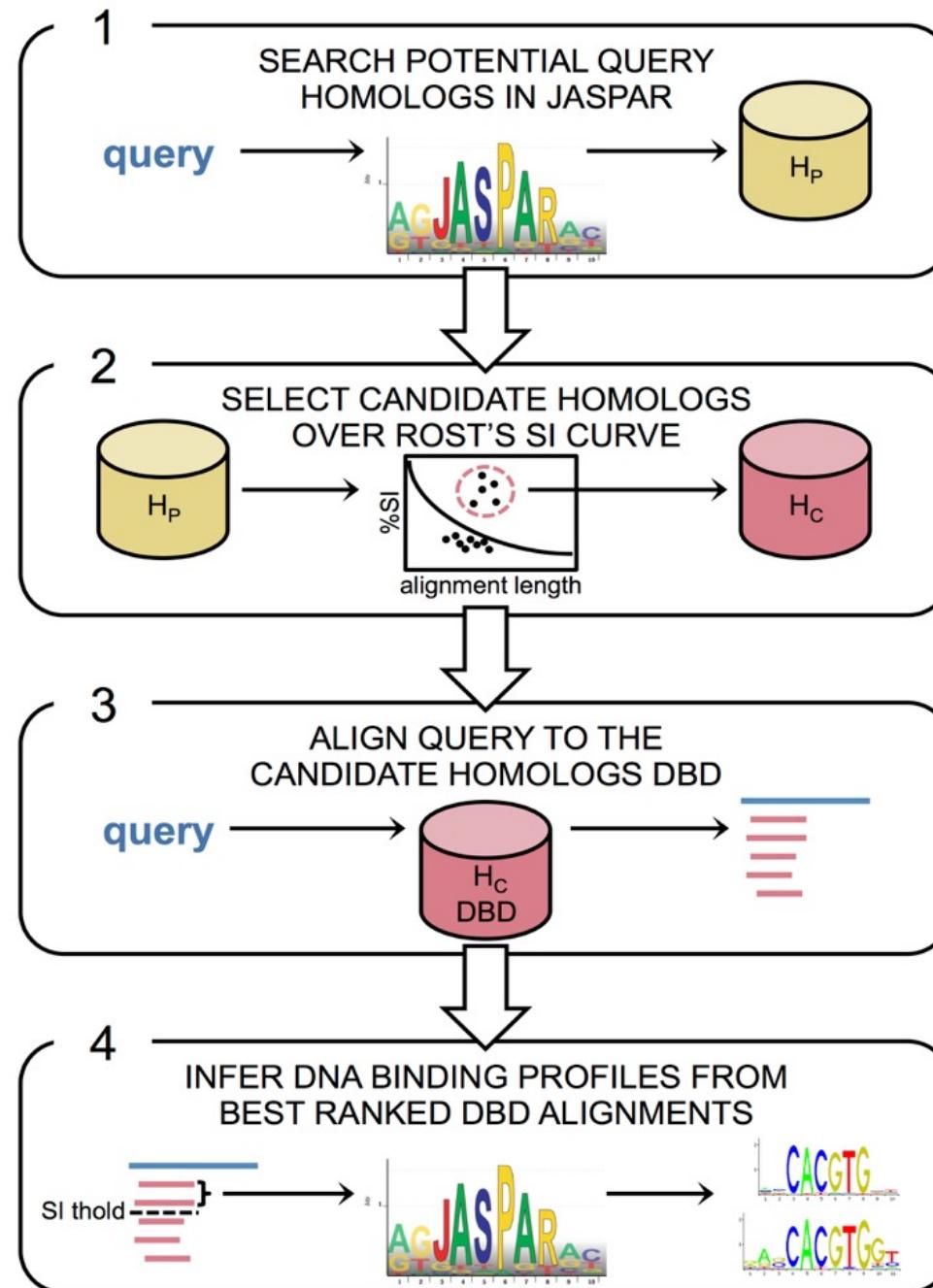
	Matrix ID	Name	DBD	E-value	Sequence logo
	MA0781.1	PAX9	0.704	3.51197e-57	
	MA0069.1	Pax6	1.0	0.0	
	MA0014.2	PAX5	0.776	2.26217e-70	
	MA0779.1	PAX1	0.704	4.72903e-55	
	MA0208.1	al	0.719298245614	3.93817e-23	

JASPAR Profile Inference

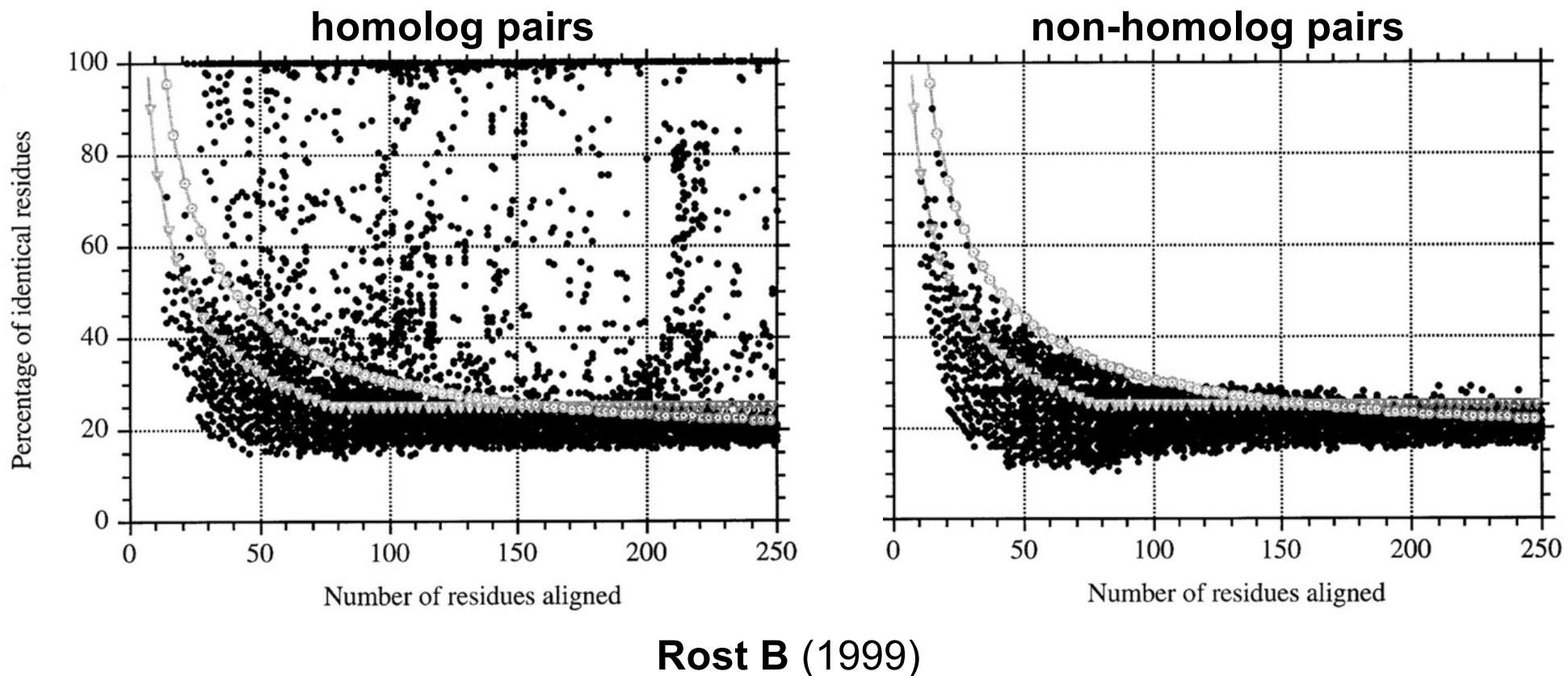


Follow @jaspar\_db 114 followers

# Profile Inference

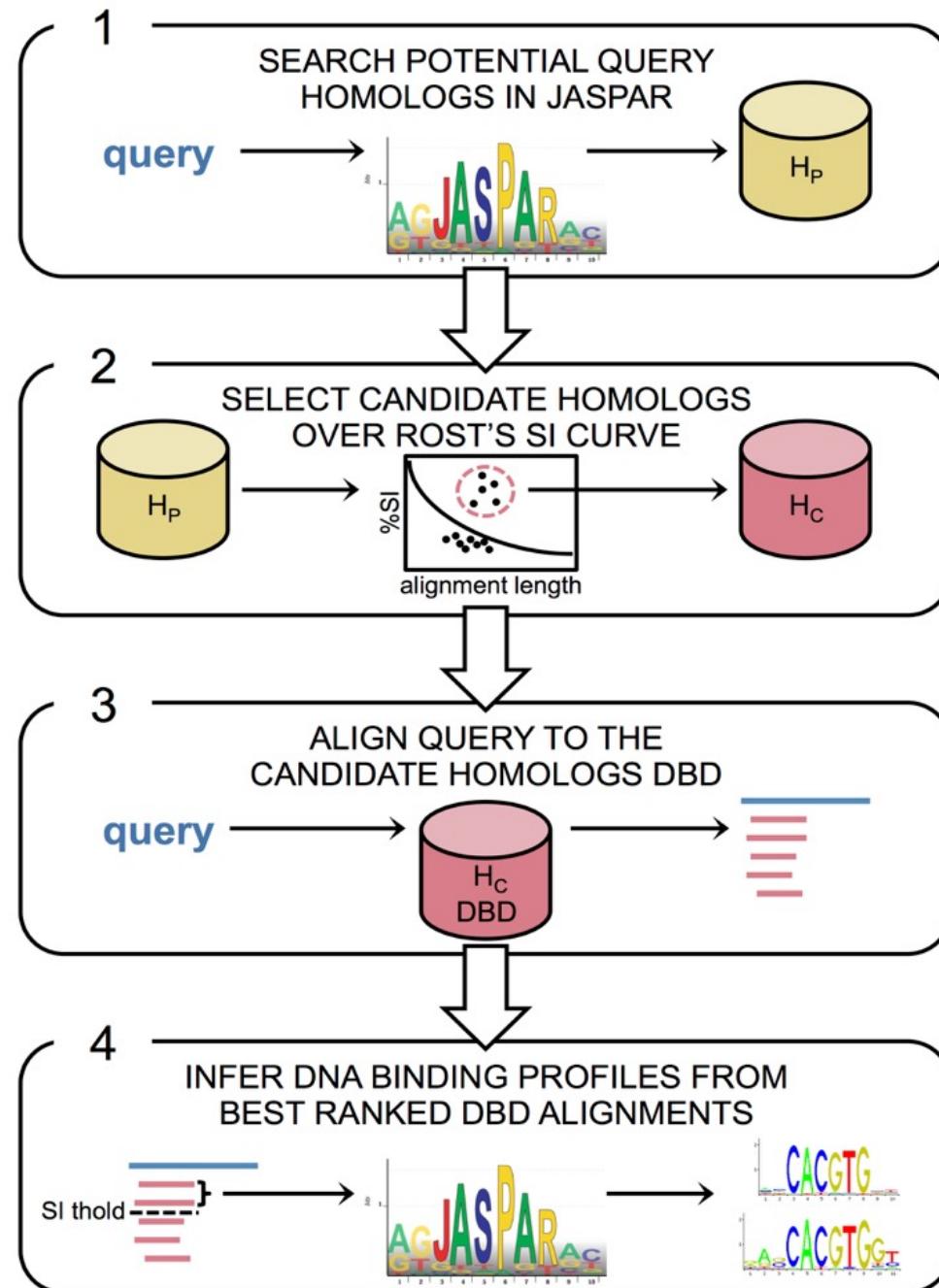


# Profile Inference

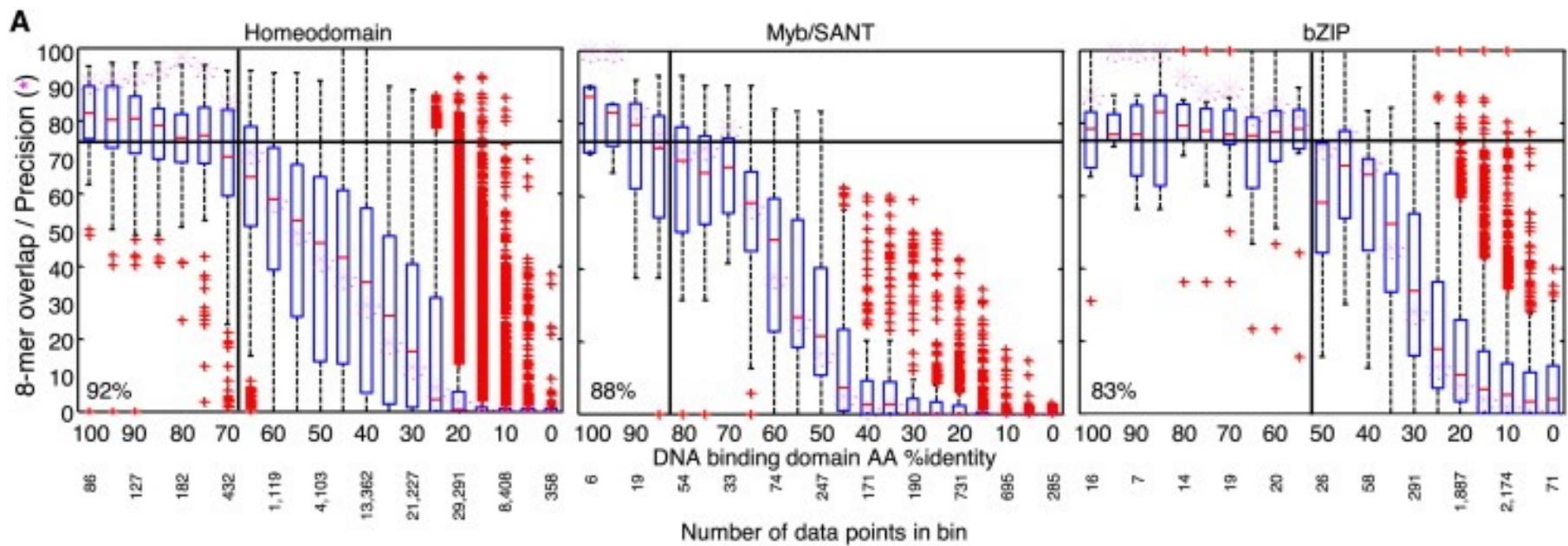


Rost B (1999)

# Profile Inference



# Profile Inference



Weirauch M, et al. (2014)

# Outline

---

- Overview
- Manual Curation
- **Tools:**
  - Profile Inference
  - **Matrix Clustering**
  - Genome Tracks
- New Web Interface
- RESTful API
- Perspectives
- MANTA2

# Matrix Clusters

JASPAR 2018

Cart 0 JASPAR Blog

Home About Search Browse JASPAR CORE Browse Collections Tools RESTful API

## JASPAR CORE matrix clustering

Browse JASPAR CORE matrix clustering for five taxonomic groups

Vertebrata Nematoda Insecta Plantae Fungi

Radial tree Cluster summary Radial tree Cluster summary Radial tree Cluster summary Radial tree Cluster summary

The main visualization is a circular phylogenetic tree with nodes representing different taxonomic groups. A specific cluster is highlighted with a red box and expanded into a detailed radial tree view. This detailed view shows a hierarchical structure with nodes connected by lines and labels. To the right of the radial tree, there is a grid of sequence motifs, each consisting of a label (e.g., SOX13, Foxq1) followed by a sequence of DNA-like characters (e.g., ACAAT, TAAAC). The entire interface is framed by a red border.

# Matrix Clusters

JASPAR 2018

Cart 0 JASPAR Blog

Home About Search Browse JASPAR CORE Browse Collections Tools RESTful API Download Data Matrix Clusters Genome Tracks

## JASPAR CORE vertebrates clustering

Home > Clustering > Detail

Radial tree

### RSAT - matrix-clustering result

Analysis: JASPAR\_2018\_vertebrates (15/09/2017 22:12)

Command

**Results Summary**

Nb Input motifs	Nb Input collections	Nb Clusters Found	Download root motifs	Complete results [zip]	Linkage method	Similarity metric	Thresholds to partition the tree
579	1	79	Download	Download	average	Ncor	Ncor = 0.4 cor = 0.6

**Logo Forest**

**Clusters Summary**

**Individual Cluster View**

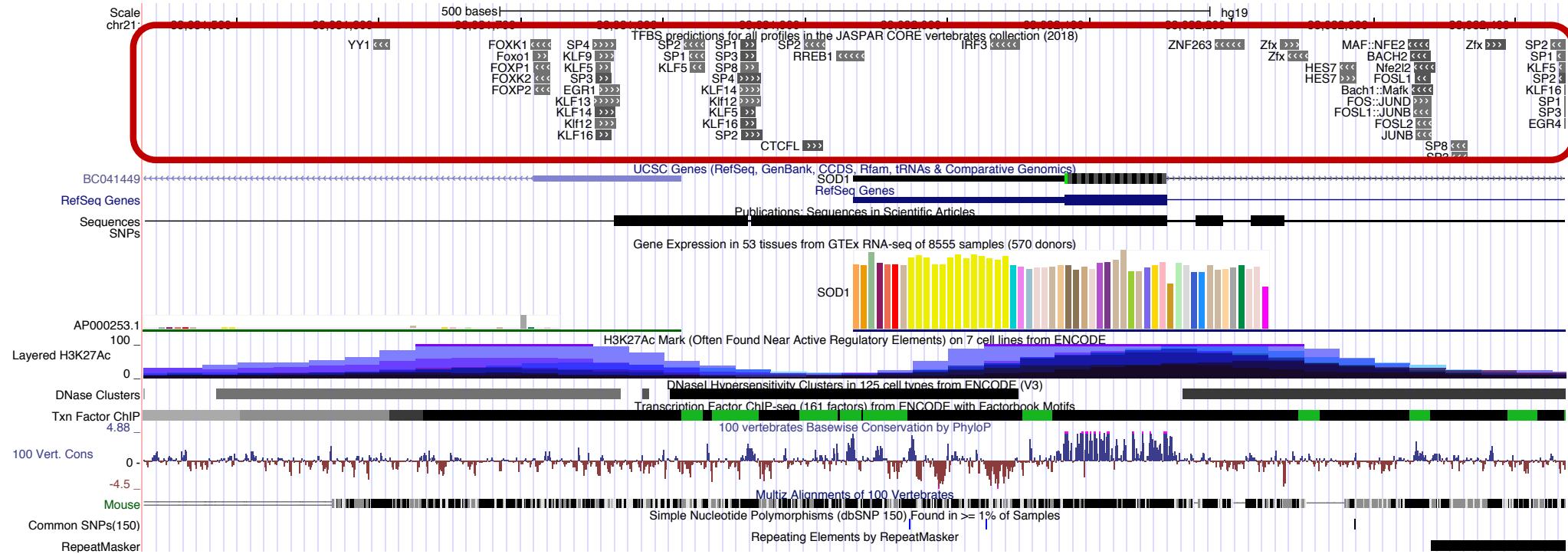
**Individual Motif View**

# Outline

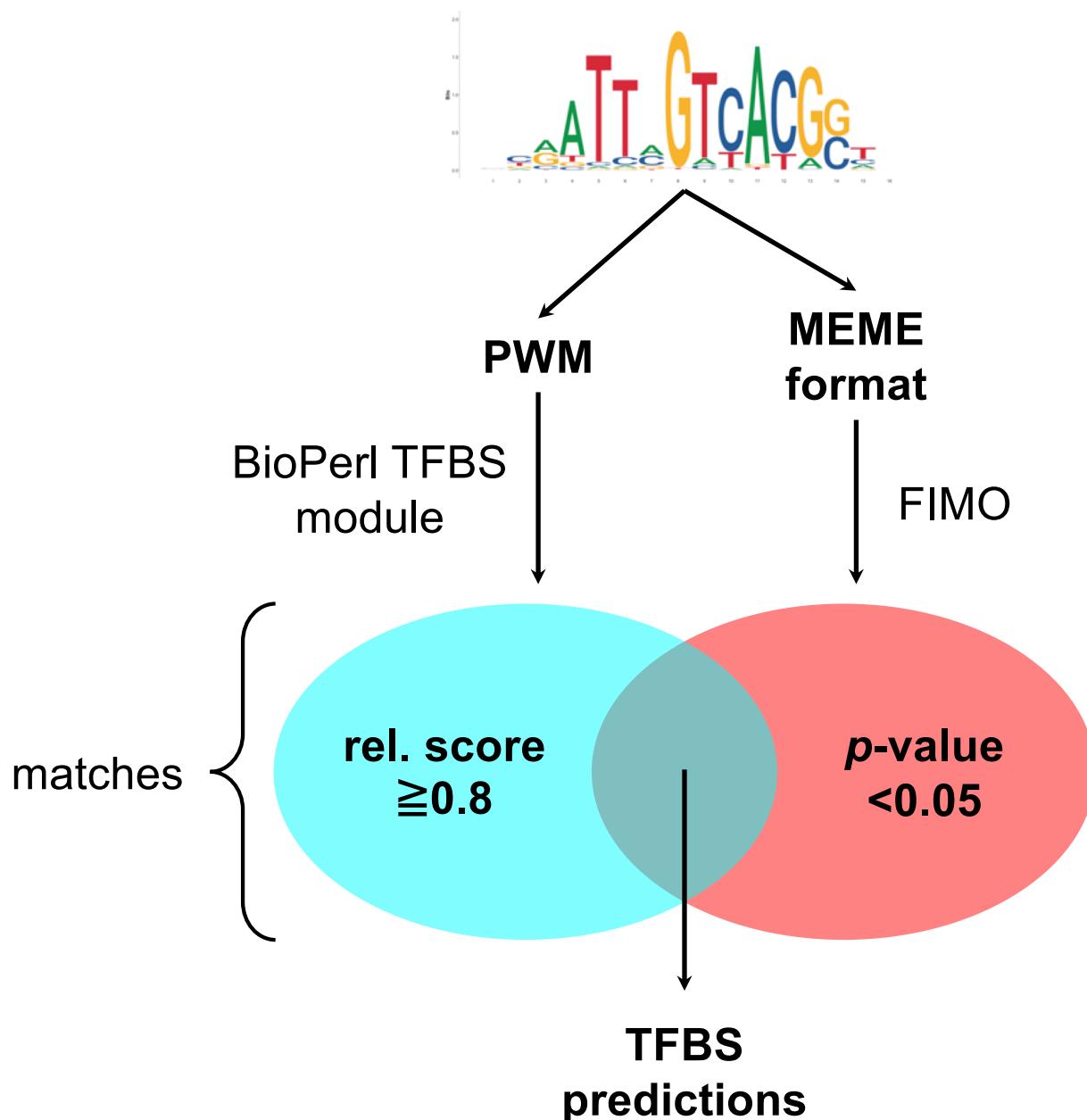
---

- Overview
- Manual Curation
- **Tools:**
  - Profile Inference
  - Matrix Clustering
  - **Genome Tracks**
- New Web Interface
- RESTful API
- Perspectives
- MANTA2

# Genome Tracks



# Genome Tracks



# Genome Tracks

---

Organism	Genome Assembly	JASPAR CORE
<i>Arabidopsis</i>	araTha1	Plants
<i>C. Elegans</i>	ce10	Nematodes
<i>Drosophila</i>	dm6	Insects
Human	hg19, hg38	Vertebrates
Mouse	mm10	Vertebrates
Yeast	sacCer3	Fungi
Zebrafish	danRer10, danRer11 (soon)	Vertebrates

**Availability:** UCSC and Ensembl browsers, Track Hub Registry

Also at: [expdata.cmmt.ubc.ca/JASPAR/downloads/UCSC\\_tracks/2018/](http://expdata.cmmt.ubc.ca/JASPAR/downloads/UCSC_tracks/2018/)

# Outline

---

- Overview
- Manual Curation
- Tools:
  - Profile Inference
  - Matrix Clustering
  - Genome Tracks
- New Web Interface
- RESTful API
- Perspectives
- MANTA2

# New Web Interface

[@ JASPAR 2018](#)

≡

Cart 0 [JASPAR Blog](#)

[Home](#)

[About](#)

[Search](#)

[Browse JASPAR CORE](#)

[Browse Collections](#)

[Tools](#)

[RESTful API](#)

[Download Data](#)

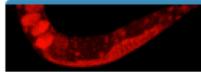
[Matrix Clusters](#)

[Genome Tracks](#)

Search JASPAR database...  
Examples: SPI1, P17676, ChIP-seq, Homo sapiens [Search Q](#) [Advanced Options](#)

Browse JASPAR CORE for six different taxonomic groups

 [Vertebrata](#)

 [Nematoda](#)

 [Insecta](#)

 [Plantae](#)

 [Fungi](#)

 [Urochordata](#)

**i** JASPAR CORE & when should it be used? [Info about other collections](#)

The JASPAR CORE contains a curated, non-redundant set of profiles, derived from published and experimentally defined transcription factor binding sites for eukaryotes. It should be used, when seeking models for specific factors or structural classes, or if experimental evidence is paramount.

 Citing JASPAR 2018 [PubMed](#) | [NAR](#) | [PDF](#)

Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 2018; 46:D260-D266, doi: [10.1093/nar/gkx1126](https://doi.org/10.1093/nar/gkx1126)

 [Profile Inference](#)  
Infer profile(s) given

 [Matrix Align](#)  
Align to a custom matrix or IUPAC

 [RESTful API](#)  
Access JASPAR database

 [Download](#)  
Batch download  
PFMs, TFFMs, sites,

# New Web Interface

[@ JASPAR 2018](#)

≡

[Home](#) > [search](#)

[Cart 0](#)

[JASPAR Blog](#)

[Home](#)

[About](#)

[Search](#) **Search Q**

[Advanced Options](#)

[Browse JASPAR CORE](#)

[Browse Collections](#)

[Tools](#)

[RESTful API](#)

[Download Data](#)

[Matrix Clusters](#)

[Genome Tracks](#)

## Search profile(s)

FOS

Examples: [SPI1](#), [P17676](#), [ChIP-seq](#), [Homo sapiens](#)

**Collection:** CORE **Taxon:** Vertebrates

**Species:** All species **Data type:** All data types

**Class:** All classes **Family:** All families

**Versions:** Latest version

**JASPAR** is an open-access database of curated, non-redundant transcription factor (TF) binding profiles stored as position frequency matrices (PFMs) and TF flexible models (TFFMs) for TFs across multiple species in six taxonomic groups.

You are using the latest 7th release ([2018](#)) of JASPAR.

[About JASPAR](#) [Profile versions](#)  
[JASPAR video tour](#) [Changelog](#)  
[Blog and News](#) [JASPAR 2016](#)  
[Contact Us](#) [JASPAR 2014](#)

[Follow @jaspar\\_db](#) 114 followers

JASPAR is brought to you by a collaborative effort of several research labs and it is licensed under [Creative Commons Attribution 4.0 International License](#).

UNIVERSITY OF COPENHAGEN   
Centre for Molecular Medicine and Therapeutics

MRC NCMM   
London Institute of Medical Sciences  
Centre for Molecular Medicine Norway

# New Web Interface

JASPAR 2018

≡

Cart 0 JASPAR Blog

Home About Search Browse JASPAR CORE Browse Collections Tools RESTful API Download Data Matrix Clusters Genome Tracks

Search profile(s)

FOS Examples: SPI1, P17676, ChIP-seq, Homo sapiens

Search Q Advanced Options

26 profile(s) found

Display 10 profiles Filter:

ID	Name	Species	Class	Family	Logo
MA0099.3	FOS::JUN	Homo sapiens	Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP)	Fos-related factors::Jun-related factors	
MA0476.1	FOS	Homo sapiens	Basic leucine zipper factors (bZIP)	Fos-related factors	
MA0477.1	FOSL1	Homo sapiens	Basic leucine zipper factors (bZIP)	Fos-related factors	
MA0605.1	Atf3	Mus musculus	Basic leucine zipper factors (bZIP)	Fos-related factors	
MA1126.1	FOS::JUN(var.2)	Homo sapiens	Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP)	Fos-related factors::Jun-related factors	

Analyze selected profiles

Please select matrix profiles on the left side to add to your cart or perform the following analysis.

Add to cart ?

You have 0 profile(s) in your cart. You can add profiles to the cart to download or perform analysis.

Add to cart View cart

Scan ?

Cluster ?

# jaspar.genereg.net/matrix/MA0476.1/

JASPAR 2018

≡

Cart 3 JASPAR Blog

Home About Search Browse JASPAR CORE Browse Collections Tools RESTful API Download Data Matrix Clusters Genome Tracks

Detailed information of matrix profile **MA0476.1**

Home > Matrix > MA0476.1

Profile summary Remove

Name: FOS

Matrix ID: MA0476.1

Class: Basic leucine zipper factors (bZIP)

Family: Fos-related factors

Collection: CORE

Taxon: Vertebrates

Species: Homo sapiens

Data Type: ChIP-seq

Validation: 17916232

Uniprot ID: P01100

Pazar TF:

TFBSshape ID: 280

TFencyclopedia IDs:

Source: ENCODE

Comment:

Sequence logo Download SVG

Frequency matrix

JASPAR TRANSFAC MEME RAW PFM

Reverse comp.

A [	7879	7475	0	0	29396	998	0	0	29396	258	4006 ]
C [	712	10177	0	0	0	14079	0	29396	0	5823	8236 ]
G [	9686	10841	0	27108	0	11206	0	0	0	1538	7897 ]
T [	11119	903	29396	2288	0	3113	29396	0	0	21777	9257 ]

Binding sites information

HTML file FASTA file BED file

TFBS profiles

TFBSshape

# New Web Interface

JASPAR 2018

Cart 0 JASPAR Blog

Home About Search Browse JASPAR CORE Browse Collections Tools RESTful API Download Data Matrix Clusters Genome Tracks

☰

## Download JASPAR data

Home > Downloads

[CORE PFMs](#) [TFFMs](#) [JASPAR collections \(PFMs\)](#) [Matrix clustering](#) [SQL dump](#) [Other data](#)

*i* Download JASPAR CORE Position Frequency Matrix (PFM) (non-redundant and redundant) in JASPAR, MEME and TRANSFAC matrix format for six taxonomic groups. More information about matrix formats can be found [here](#).

Taxonomic group	PFMs (non-redundant)	PFMs (redundant)
Vertebrates	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>
Plants	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>
Insects	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>
Nematodes	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>
Fungi	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>
Urochordates	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>
All CORE profiles	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>	<ul style="list-style-type: none"><li>Individual PFMs (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul>

# Outline

---

- Overview
- Manual Curation
- Tools:
  - Profile Inference
  - Matrix Clustering
  - Genome Tracks
- New Web Interface
- RESTful API
- Perspectives
- MANTA2

# RESTful API

 JASPAR REST API

Home

Browsable API ▾

API Overview

Documentation

Live API

API Clients

Contact Us

Welcome to

## {JASPAR RESTful API}

This API provides easy-to-use REST web interface to query/retrieve matrix profile data from the latest version of JASPAR database. The API comes with a human browsable interface and also programmatic interface, which return the results in eight different formats, including `json`, `jsonp`, `jaspar`, `meme`, `transfac`, `pfm`, `yaml` and `bed`.

Try it Now

Live API

 Read more about JASPAR and JASPAR RESTful API

- Khan, A. and Mathelier, A. JASPAR RESTful API: accessing JASPAR data from any programming language. *Bioinformatics*, 2017, doi: [10.1093/bioinformatics/btx804](https://doi.org/10.1093/bioinformatics/btx804)
- Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 2018; 46:D260–D266, doi: [10.1093/nar/gkx1126](https://doi.org/10.1093/nar/gkx1126)

A RESTful [\(API\)](#) to programmatically access the latest version of JASPAR database.

# RESTful API

JASPAR REST API

Home

Browsable API ▾

API Overview

Documentation

Live API

API Clients

Contact Us

Home / Live API

## JASPAR Live API

### collections

Show/Hide | List Operations | Expand Operations

GET /api/v1/collections/

List all the collections are available in JASPAR.

GET /api/v1/collections/{collection}/

API endpoint that returns a list of all matrix profiles based on collection name.

### infer

Show/Hide | List Operations | Expand Operations

GET /api/v1/infer/{sequence}/

Infer matrix profiles, given protien sequence

### matrix

Show/Hide | List Operations | Expand Operations

GET /api/v1/matrix/

REST API endpoint that returns a list of all matrix profiles.

GET /api/v1/matrix/{base\_id}/versions/

List matrix profile versions based on base\_id

GET /api/v1/matrix/{matrix\_id}/

Gets profile detail information

### releases

Show/Hide | List Operations | Expand Operations

GET /api/v1/releases/

REST API endpoint that returns all releases of JASPAR database.

GET /api/v1/releases/{release\_number}/

Gets JASPAR release information based on release number

### sites

Show/Hide | List Operations | Expand Operations

GET /api/v1/sites/{matrix\_id}/

List matrix profile sites based on matrix\_id

# RESTful API

Home / Live API

## infer

Show/Hide | List Operations | Expand Operations

GET

/api/v1/infer/{sequence}

Infer matrix profiles, given protein sequence

### Implementation Notes

Infer matrix profiles, given protein sequence

### Parameters

Parameter	Value	Description	Parameter Type	Data Type
sequence	IPMGTSGTTSTGLISPQVSVPVQVPGSEPDMQSQYWPRQLQ		path	string

### Response Messages

HTTP Status Code	Reason	Response Model	Headers
------------------	--------	----------------	---------

200

Try it out!

[Hide Response](#)

### Curl

```
curl -X GET --header 'Accept: application/json' 'http://jaspar.genereg.net/api/v1/infer/MQNSHSGVNQLGGVFVNNGRPLPDSTRQKIVI'
```

### Request URL

```
http://jaspar.genereg.net/api/v1/infer/MQNSHSGVNQLGGVFVNNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVA
```

### Request Headers

```
{
  "Accept": "application/json"
}
```

# RESTful API

---

```
Oriols-MacBook-Pro:JASPAR orfornes$ curl -X GET --header 'Accept: application/json' 'http://jaspar.genereg.net/api/v1/infer/MQNSHSGVNQLGGVFVNGRPLPDSTRQKIVELAHSARGPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATPEEVSKIAQYKRECPEIFAEIRDRLSEGVCCTNDNIPSVSSINRVLRNLAZEKQQMGADGMYDKLRLNGQTGSWGTRPGWYPGTSVPGQPTQDGCCQQEGGGENTNSISSNGEDSDEAQMRLQLKRKLQRNRTSFTQEQUIEALEKEFERTHYPDVFARERLAAKIDLPEARIQVWFNSNRRAKWRREEKLRNQRRQASNTPSHIPISSSFSTSVDQPIPQPTTPVSSFTSGSMLGRTDTALTNTYSALPPMPSFTMANNLPMQPPVPSQTSSYSCMLPTSPSVNGRSYDTYTPPHMQTHMNSQPMGTSGTTSTGLISPVGVSVPVQVPGSEPDMSQYWPRLLQ/'  
{"count":6,"results":[{"name":"Pax6","url":"http://jaspar.genereg.net/api/v1/matrix/MA0069.1","evalue":0.0,"matrix_id":"MA0069.1","dbd":1.0,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0069.1.svg"}, {"name":"PAX7","url":"http://jaspar.genereg.net/api/v1/matrix/MA0680.1","evalue":1.91654e-71,"matrix_id":"MA0680.1","dbd":0.7109375,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0680.1.svg"}, {"name":"PAX5","url":"http://jaspar.genereg.net/api/v1/matrix/MA0014.2","evalue":6.0162e-67,"matrix_id":"MA0014.2","dbd":0.776,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0014.2.svg"}, {"name":"PAX9","url":"http://jaspar.genereg.net/api/v1/matrix/MA0781.1","evalue":2.34098e-54,"matrix_id":"MA0781.1","dbd":0.72,"sequence_logo":"http://jaspar.genereg.net/static/logo/s/svg/MA0781.1.svg"}, {"name":"PAX1","url":"http://jaspar.genereg.net/api/v1/matrix/MA0779.1","evalue":1.01449e-52,"matrix_id":"MA0779.1","dbd":0.72,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0779.1.svg"}, {"name":"al","url":"http://jaspar.genereg.net/api/v1/matrix/MA0208.1","evalue":1.98946e-22,"matrix_id":"MA0208.1","dbd":0.7192982456140351,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0208.1.svg"}]}Oriols-MacBook-Pro:JASPAR orfornes$
```

# RESTful API

JASPAR REST API

Home

Browsable API ▾

API Overview

Documentation

Live API

API Clients

Contact Us

## Matrix Inference

This is JASPAR browsable API, which provides easy-to-use REST web interface to query/retrieve matrix profile data from JASPAR database. The API comes with a human browsable interface and also programmatic interface, which return the results in eight different formats, including `json`, `jsonp`, `jaspar`, `transfac`, `pfm`, `meme` and `yaml`, `bed` and also `api` for browsable interface.

API endpoint that infer matrix models based on protein sequence.

GET /api/v1/infer/MQNSHSGVNQLGGVFVNGRPLPSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATPEVVSKIAQYKRECPSIFAEIRDRLLSEGV

GET ▾

```
{"count":6,"results": [{"name":"Pax6","url":"http://jaspar.genereg.net/api/v1/matrix/MA0069.1","evalue":0.0,"matrix_id":"MA0069.1","dbd":1.0,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0069.1.svg"}, {"name":"PAX7","url":"http://jaspar.genereg.net/api/v1/matrix/MA0680.1","evalue":1.91654e-71,"matrix_id":"MA0680.1","dbd":0.7109375,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0680.1.svg"}, {"name":"PAX5","url":"http://jaspar.genereg.net/api/v1/matrix/MA0014.2","evalue":6.0162e-67,"matrix_id":"MA0014.2","dbd":0.776,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0014.2.svg"}, {"name":"PAX9","url":"http://jaspar.genereg.net/api/v1/matrix/MA0781.1","evalue":2.34098e-54,"matrix_id":"MA0781.1","dbd":0.72,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0781.1.svg"}, {"name":"PAX1","url":"http://jaspar.genereg.net/api/v1/matrix/MA0779.1","evalue":1.01449e-52,"matrix_id":"MA0779.1","dbd":0.72,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0779.1.svg"}, {"name":"al","url":"http://jaspar.genereg.net/api/v1/matrix/MA0208.1","evalue":1.98946e-22,"matrix_id":"MA0208.1","dbd":0.7192982456140351,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0208.1.svg"}]}  
  
{"name": "Pax6",  
 "url": "http://jaspar.genereg.net/api/v1/matrix/MA0069.1",  
 "evalue": 0.0,  
 "matrix_id": "MA0069.1",  
 "dbd": 1.0,  
 "sequence_logo": "http://jaspar.genereg.net/static/logos/svg/MA0069.1.svg"  
},  
{  
    "name": "PAX7",  
    "url": "http://jaspar.genereg.net/api/v1/matrix/MA0680.1",  
    "evalue": 1.91654e-71,  
    "matrix_id": "MA0680.1",  
    "dbd": 0.7109375,  
    "sequence_logo": "http://jaspar.genereg.net/static/logos/svg/MA0680.1.svg"  
},
```

# RESTful API

---

```
Oriols-MacBook-Pro:JASPAR ofornes$ python
Python 2.7.13 (default, Jul 18 2017, 09:17:00)
[GCC 4.2.1 Compatible Apple LLVM 8.1.0 (clang-802.0.42)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> import coreapi
>>> import urllib2
>>> from coreapi import codecs
>>>
>>> # Initialize #
... client = coreapi.Client()
>>> codec = codecs.CoreJSONCodec()
>>> seq = "MQNSHSGVNQLGGVFVNGRPLPDSTRKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATPEVVSKIAQYKRECPEIFAWEIRDRLLSEGVCTNDNIPSVSSINVRNLASEKQQMGADGMYDKLRMLNQQTGWSGTRPGWPGTSVPQPTQDGCCQQEGGGENTNSISSNGEDSDEAQMRLQLKRKLQRNRTSFTQEQUIALEKEFERTHYPDVFARERLAAKIDLPEARIQVWFNSNRRAKWRREEKLRNQRRQASNTPSHIPISSSFSTSVYQPIPQPTTVSSFTSGSMLGRTDTALTNTYSALPPMPSFTMANNLPMQPPVPSQTSSYSCMLPTSPSVNGRSYDTYTPPHMQTHMNSQPMGTSGTSTGLISPGVSVPVQVPGSEPDMSQYWPRLO"
>>> url = "http://jaspar.genereg.net/api/v1/infer/%s/" % seq
>>>
>>> # Get response w/ coreapi
... coreapi_response = client.get(url)
>>> coreapi_response = codec.encode(coreapi_response)
>>>
>>> # Get response w/ urllib2
... request = urllib2.Request(url, headers={'Content-Type': 'application/json'})
>>> f = urllib2.urlopen(request)
>>> urllib2_response = f.read()
>>> f.close()
>>>
>>> print(urllib2_response)
{"count":6,"results":[{"name":"Pax6","url":"http://jaspar.genereg.net/api/v1/matrix/MA0069.1","evaluate":0.0,"matrix_id":"MA0069.1","dbd":1.0,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0069.1.svg"}, {"name":"PAX7","url":"http://jaspar.genereg.net/api/v1/matrix/MA0680.1","evaluate":1.91654e-71,"matrix_id":"MA0680.1","dbd":0.7109375,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0680.1.svg"}, {"name":"PAX5","url":"http://jaspar.genereg.net/api/v1/matrix/MA0014.2","evaluate":6.0162e-67,"matrix_id":"MA0014.2","dbd":0.776,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0014.2.svg"}, {"name":"PA X9","url":"http://jaspar.genereg.net/api/v1/matrix/MA0781.1","evaluate":2.34098e-54,"matrix_id":"MA0781.1","dbd":0.72,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0781.1.svg"}, {"name":"PAX1","url":"http://jaspar.genereg.net/api/v1/matrix/MA0779.1","evaluate":1.01449e-52,"matrix_id":"MA0779.1","dbd":0.72,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0779.1.svg"}, {"name":"al","url":"http://jaspar.genereg.net/api/v1/matrix/MA0208.1","evaluate":1.98946e-22,"matrix_id":"MA0208.1","dbd":0.7192982456140351,"sequence_logo":"http://jaspar.genereg.net/static/logos/svg/MA0208.1.svg"}]}
>>> print(coreapi_response == urllib2_response)
True
>>>
```

# Outline

---

- Overview
- Manual Curation
- Tools:
  - Profile Inference
  - Matrix Clustering
  - Genome Tracks
- New Web Interface
- RESTful API
- Perspectives
- MANTA2

# Perspectives

---

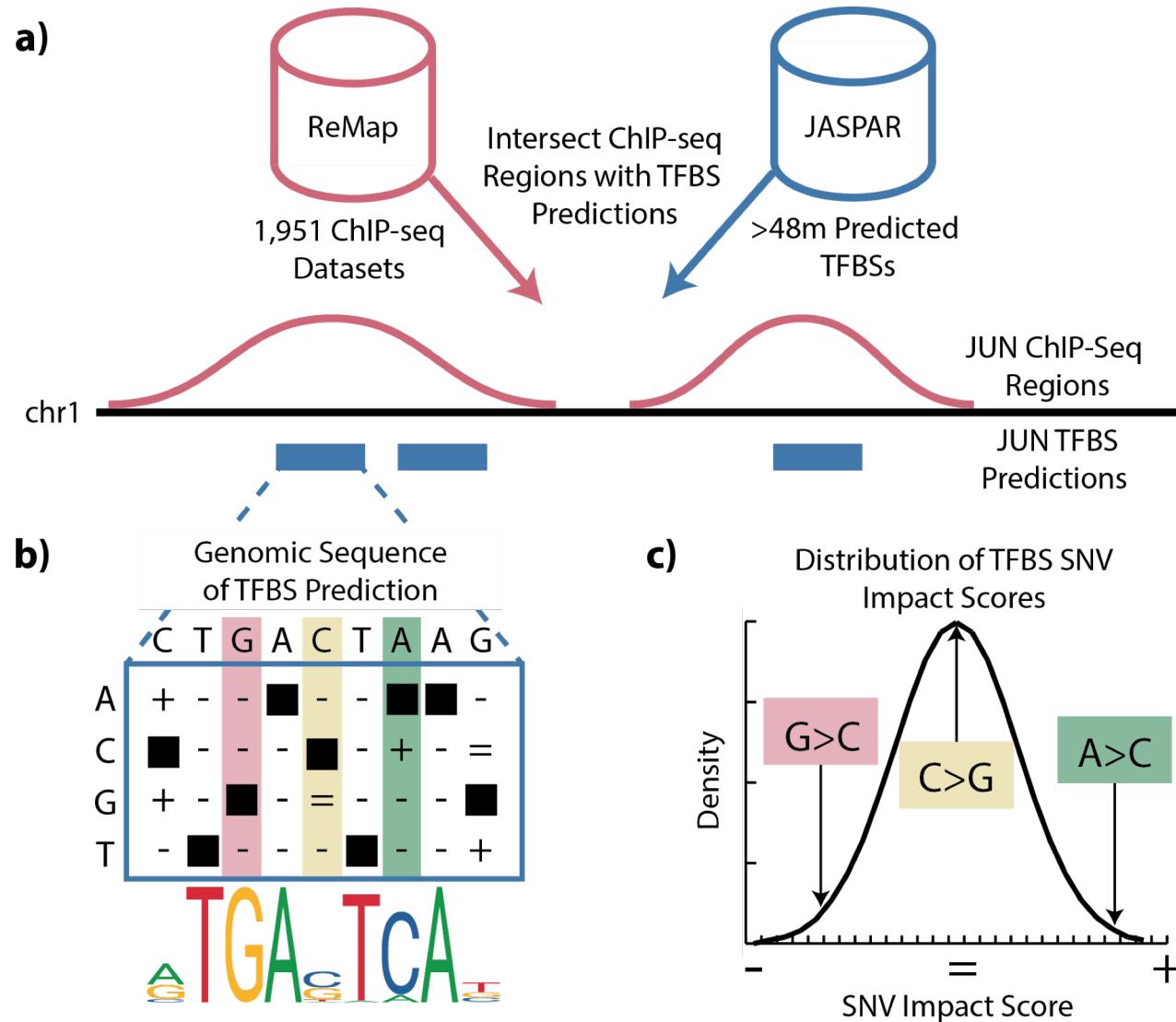
- JASPAR beta (contains profiles/functionalities to be added to the next release)
- Validation section (contains profiles that we failed to validate through literature support)
- Genome tracks at user's request?
- While we are close to profile every human TF (except for zinc fingers), we don't have an estimate on the % of covered binding modes, heterodimers, etc.

# Outline

---

- Overview
- Manual Curation
- Tools:
  - Profile Inference
  - Matrix Clustering
  - Genome Tracks
- New Web Interface
- RESTful API
- Perspectives
- **MANTA2**

# MANTA2



# MANTA2



Search



Upload

Communities

Log in

Sign up

December 1, 2017

Journal article

Open Access

## The Mongo database for the analysis of transcription factor binding site alterations (MANTA)

Fornes, Oriol; Gheorghe, Marius; Richmond, Phillip A.; Arenillas, David J.; Wasserman, Wyeth W.; Mathelier, Anthony

The MongoDB for the ANalysis of Transcription factor (TF)-binding site (TFBS) Alterations (MANTA) was originally created in 2015 to study the impact of regulatory mutations in B-cell lymphomas (Mathelier et al. 2015). The database stores TFBSs predicted in the human genome by combining ChIP-seq regions from ReMap and JASPAR profiles, as well as the potential impact scores for all possible single nucleotide variants (SNVs) at these TFBSs. The second release of the database, MANTA2, houses >48 million TFBS predictions for 225 TFs, covering 255,918,025 bp (~8%) of the human genome (hg38). Here we provide a MongoDB dump of MANTA2 comprising the files `experiments.json`, `experiments.metadata.json`, `system.indexes.json`, `tfbs_snvs.json`, and `tfbs_snvs.metadata.json`.

The MANTA2 source code is deposited in GitHub (<https://github.com/wassermanlab/MANTA2>). Instructions on the installation and usage of MANTA2 are also provided in the GitHub repository. The MANTA2 database can also be accessed via a dedicated web server at <http://manta.cmmi.ubc.ca/manta2>.

Files (6.6 GB)	
Name	Size
<a href="#">manta2_mongodb_dump.tgz</a>	6.6 GB
md5:f3501a9edff55cb7885e7f9a7f44fd9	

Indexed in

OpenAIRE

Publication date:

December 1, 2017

DOI:

DOI [10.5281/zenodo.1044747](https://doi.org/10.5281/zenodo.1044747)

License (for files):

Versions

[Version 2](#) 10.5281/zenodo.1044747

Dec 1, 2017

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.1044746](https://doi.org/10.5281/zenodo.1044746). This DOI represents all versions, and will always resolve to the latest one. [Read more](#).

Share



# github.com/wassermanlab/MANTA2

This repository Search Pull requests Issues Marketplace Explore

wassermanlab / MANTA2 Unwatch 2 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

MANTA version 2.0 Edit

Add topics

37 commits 1 branch 0 releases 2 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

darenillas Changed the link describing the GFF2 format plus did some minor wording ... Latest commit 5a4e5d9 on Nov 30, 2017

manta2 Changed the link describing the GFF2 format plus did some minor wording 2 months ago

.gitignore Updated all paths/URLs to work in the Apache mod\_wsgi environment. 2 months ago

MANIFEST.in Initial commit 3 months ago

README.md Update README.md 3 months ago

setup.py Initial commit 3 months ago

README.md

## MANTA2

The MongoDB for the ANalysis of Transcription factor (TF)-binding site (TFBS) Alterations (MANTA) was originally created in 2015 to study the impact of regulatory mutations in B-cell lymphomas (Mathelier et al. 2015). The database stores

# The JASPAR Consortium



# Acknowledgements

---

Funds from Genome Canada and the CIHR (OnTarget grants 255ONT and BOP-149430) and the Weston Brain Institute (20R74681)

WestGrid ([www.westgrid.ca](http://www.westgrid.ca)) and Compute Canada ([www.computecanada.ca](http://www.computecanada.ca)) support for genome tracks

**Special thanks to the scientific community for performing the experiments and for publicly releasing the data**

# Questions

---

# TF Coverage for Human

---

## The Human Transcription Factors

Samuel A. Lambert,<sup>1,9</sup> Arttu Jolma,<sup>2,9</sup> Laura F. Campitelli,<sup>1,9</sup> Pratyush K. Das,<sup>3</sup> Yimeng Yin,<sup>4</sup> Mihai Albu,<sup>2</sup> Xiaoting Chen,<sup>5</sup> Jussi Taipale,<sup>3,4,6,\*</sup> Timothy R. Hughes,<sup>1,2,\*</sup> and Matthew T. Weirauch<sup>5,7,8,\*</sup>

<sup>1</sup>Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

<sup>2</sup>Donnelly Centre, University of Toronto, Toronto, ON, Canada

<sup>3</sup>Genome-Scale Biology Program, University of Helsinki, Helsinki, Finland

<sup>4</sup>Division of Functional Genomics and Systems Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna, Sweden

<sup>5</sup>Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

<sup>6</sup>Department of Biochemistry, Cambridge University, Cambridge CB2 1GA, United Kingdom

<sup>7</sup>Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

<sup>8</sup>Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

<sup>9</sup>These authors contributed equally

\*Correspondence: ajt208@cam.ac.uk (J.T.), t.hughes@utoronto.ca (T.R.H.), Matthew.Weirauch@cchmc.org (M.T.W.)

<https://doi.org/10.1016/j.cell.2018.01.029>

Transcription factors (TFs) recognize specific DNA sequences to control chromatin and transcription, forming a complex system that guides expression of the genome. Despite keen interest in understanding how TFs control gene expression, it remains challenging to determine how the precise genomic binding sites of TFs are specified and how TF binding ultimately relates to regulation of transcription. This review considers how TFs are identified and functionally characterized, principally through the lens of a catalog of over 1,600 likely human TFs and binding motifs for two-thirds of them. Major classes of human TFs differ markedly in their evolutionary trajectories and expression patterns, underscoring distinct functions. TFs likewise underlie many different aspects of human physiology, disease, and variation, highlighting the importance of continued effort to understand TF-mediated gene regulation.

# Genome Tracks Usage

**Hosts (Top 10)** - [Full list](#) - [Last visit](#) - [Unresolved IP Address](#)

Hosts : 0 Known, 116 Unknown (unresolved ip) 95 Unique visitors	Pages	Hits	Bandwidth	Last visit
128.114.119.134	492	1,784	2.05 MB	31 Dec 2017 - 18:43
128.114.119.135	462	1,694	1.94 MB	31 Dec 2017 - 20:15
128.114.119.133	458	1,750	1.78 MB	31 Dec 2017 - 19:41
128.114.119.132	354	1,528	1005.95 KB	31 Dec 2017 - 17:13
128.114.119.131	299	1,210	1.09 MB	31 Dec 2017 - 12:25
129.70.40.99	117	337	748.28 KB	29 Dec 2017 - 11:38
132.249.245.79	70	231	77.98 KB	29 Dec 2017 - 02:58
193.62.192.249	49	69	8.49 KB	28 Dec 2017 - 04:24
193.62.192.251	42	58	5.72 KB	28 Dec 2017 - 04:25
193.62.192.253	42	64	8.60 KB	28 Dec 2017 - 04:25
Others	457	1,012	9334.11 GB	