

JASPAR: a comprehensive database of transcription factor binding profiles

Aziz Khan¹, Oriol Fornes², Arnaud Stigliani³, Marius Gheorghe¹, Jaime A. Castro-Mondragon¹, Robin van der Lee², Adrien Bessy³, Jeanne Chèneby⁴, Shubhada R. Kulkarni⁵, Ge Tan⁶, Damir Baranasic⁶, David J. Arenillas², Albin Sandelin⁷, Klaas Vandepoele⁵, Boris Lenhard⁶, Benoît Ballester⁴, Wyeth W. Wasserman², François Parcy³, and Anthony Mathelier¹

¹Centre for Molecular Medicine Norway (NCMM), University of Oslo, Norway

²Centre for Molecular Medicine and Therapeutics, University of British Columbia, Canada

³Univ. Grenoble Alpes, CNRS, CEA, INRA, BIG-LPCV, Grenoble, France

⁴INSERM, UMR1090 TAGC, Marseille, France

⁵Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

⁶Institute of Clinical Sciences, Imperial College London, UK

⁷The Bioinformatics Centre, University of Copenhagen, Denmark

JASPAR (<http://jaspar.genereg.net>) is an open-access database of curated and non-redundant transcription factor (TF)-binding profiles stored as position frequency matrices (PFMs) and TF flexible models (TFFMs) for TFs across species in six taxonomic groups. In the seventh release of JASPAR, we have updated and added 355 new PFMs and 316 TFFMs to the JASPAR CORE collection, representing ~30% and ~243% expansion, respectively. The JASPAR CORE collection now includes 1,404 non-redundant PFMs (579 for vertebrates, 489 for plants, 176 for fungi, 133 for insects, 26 for nematodes, and 1 for urochordata). Further, it contains 446 TFFMs (225 for vertebrates, 218 for plants and 3 for insects). Further, the 2018 release incorporates clusters of similar PFMs in each taxon and each TF class per taxon. Furthermore, we have used PFMs from JASPAR CORE collection to predict TF-binding sites in several species, which are available through UCSC and Ensembl Genome Browser track hubs. Finally, the current release comes with a new web framework with an interactive and responsive user-interface, along with a Representational State Transfer (REST) application programming interface (API) to access the JASPAR database programmatically. Additionally, we are planning to introduce community-based curation for profiles that were not added into JASPAR due to lack of experimental evidence from existing literature. We will encourage researchers to perform experiments and/or point us to literature that our curators missed in order to support these profiles. All the underlying data can be browsed and downloaded using the website, and can be retrieved programmatically using the RESTful API (<http://jaspar.genereg.net/api>) and through the R/Bioconductor package.