

Ch 3 - L 1.2

Transcription Factors

The Human Transcription Factors

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

¹Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

²Donnelly Centre, University of Toronto, Toronto, ON, Canada

³Genome-Scale Biology Program, University of Helsinki, Helsinki, Finland

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

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

⁶Department of Biochemistry, Cambridge University, Cambridge CB2 1GA, United Kingdom

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

⁸Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

⁹These 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.

How TFs recognize DNA sequences

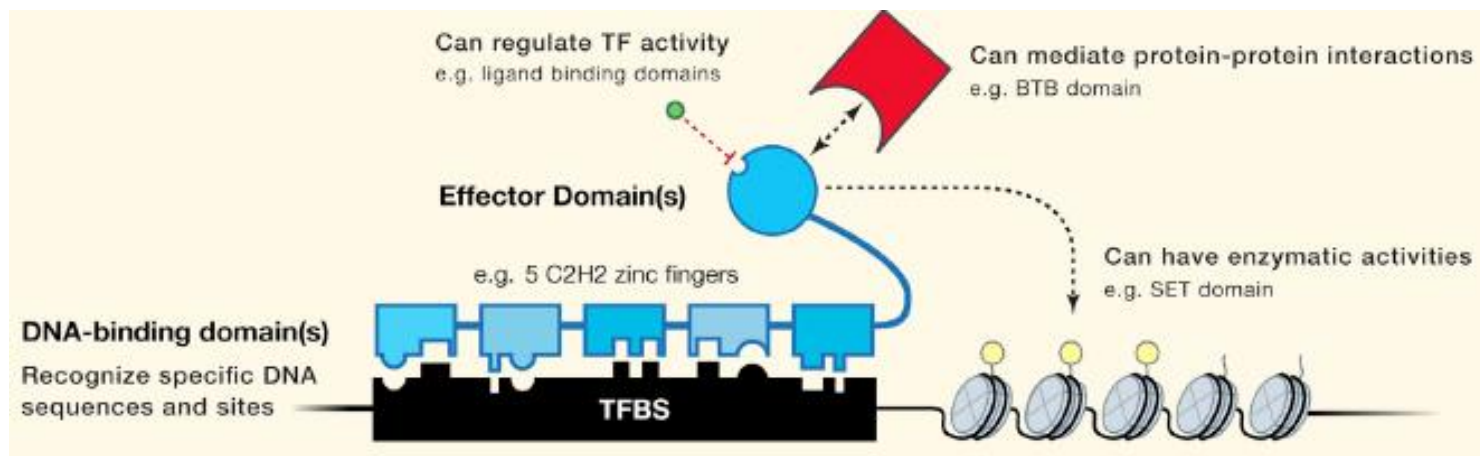
Transcription Factors

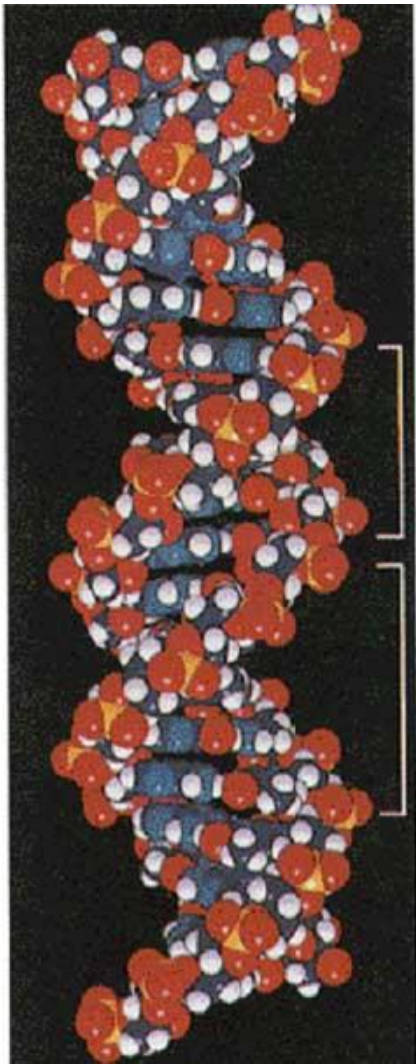
>10% of the coding potential (2,000-3,000) of the Human Genome

TFs recognize DNA motifs by multiple chemical interactions between aa residues of their **DBD** (DNA binding domain) and 4-6 bp in the major groove

TFBS for a specific TF can vary a bit → consensus sequence

TFs contain one or more transactivating domains (TA) that interact with co-regulatory proteins

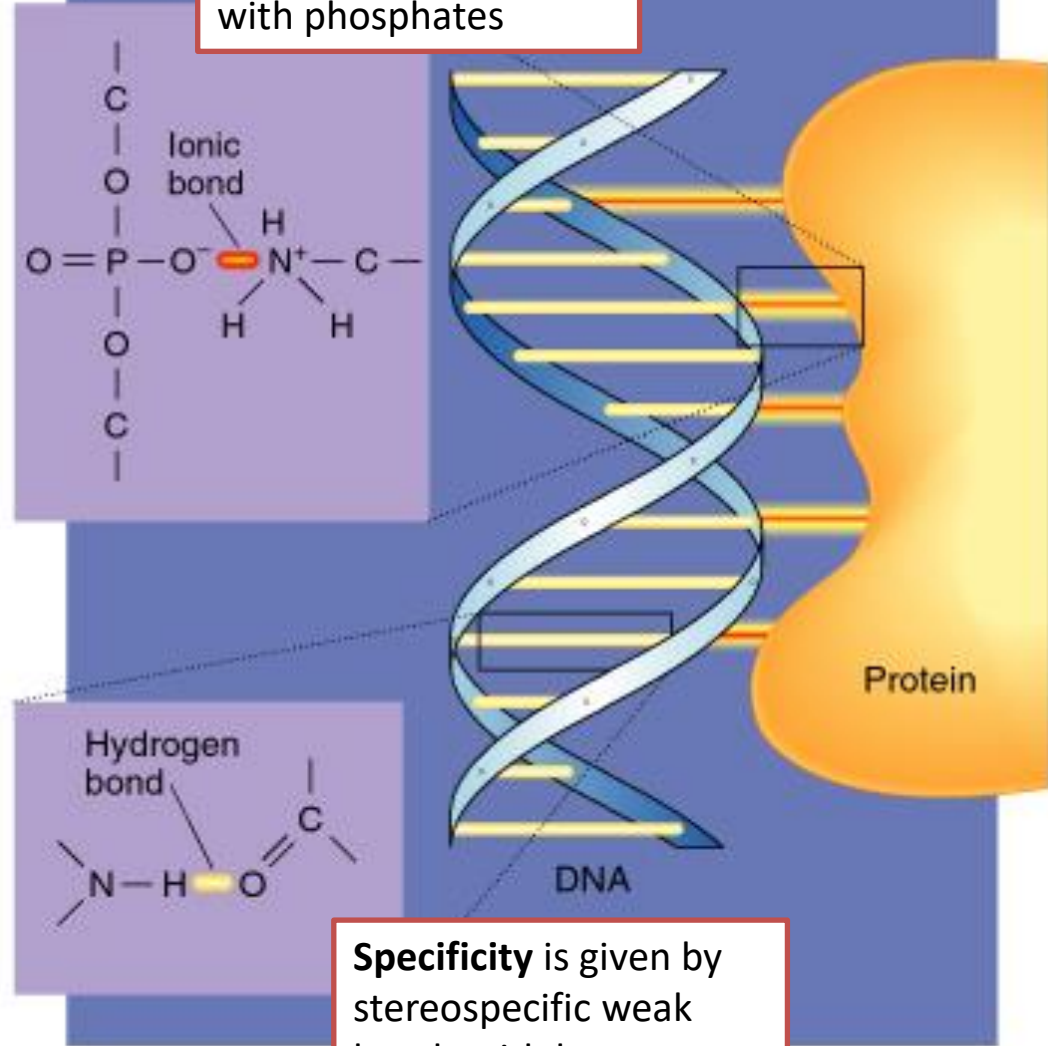




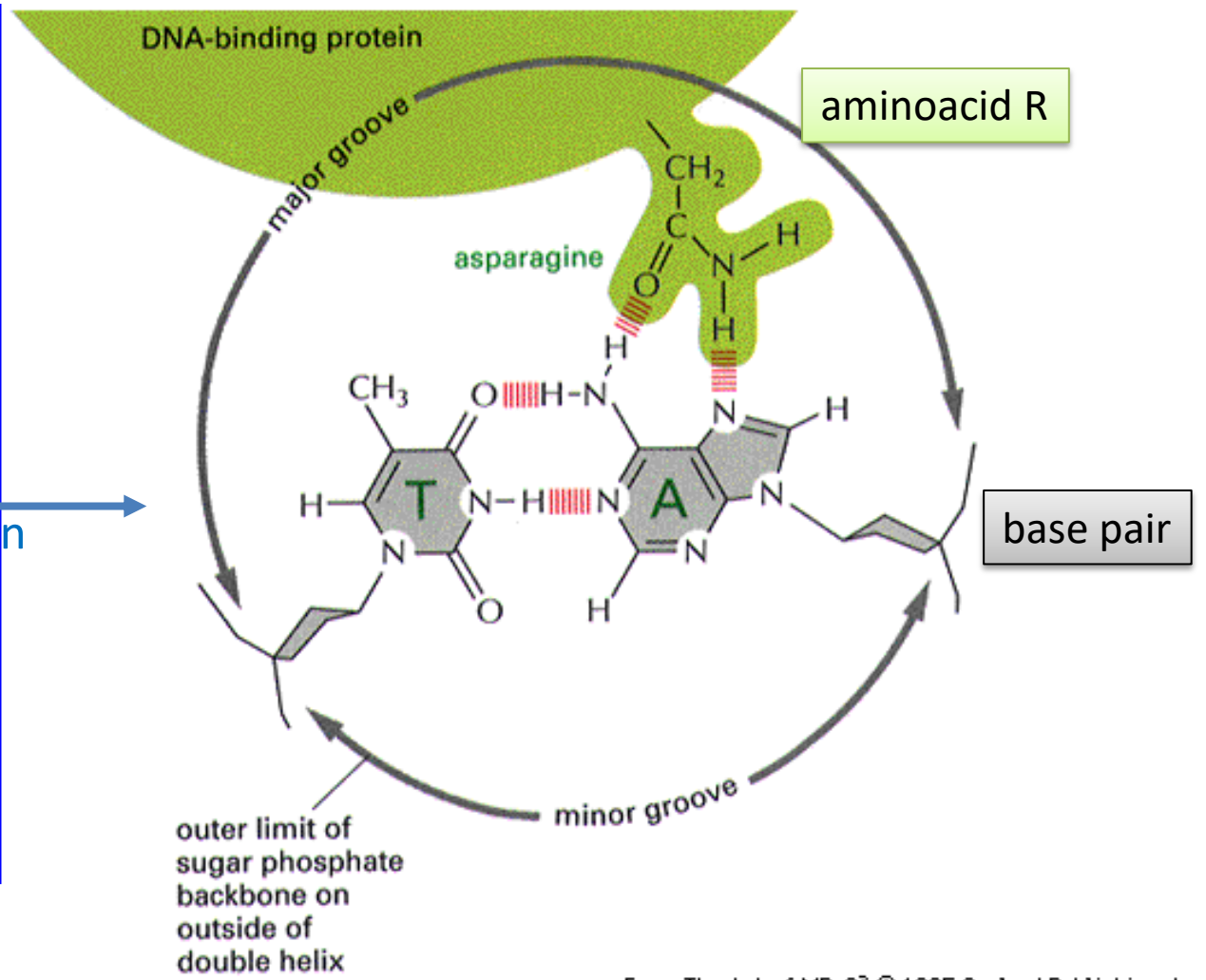
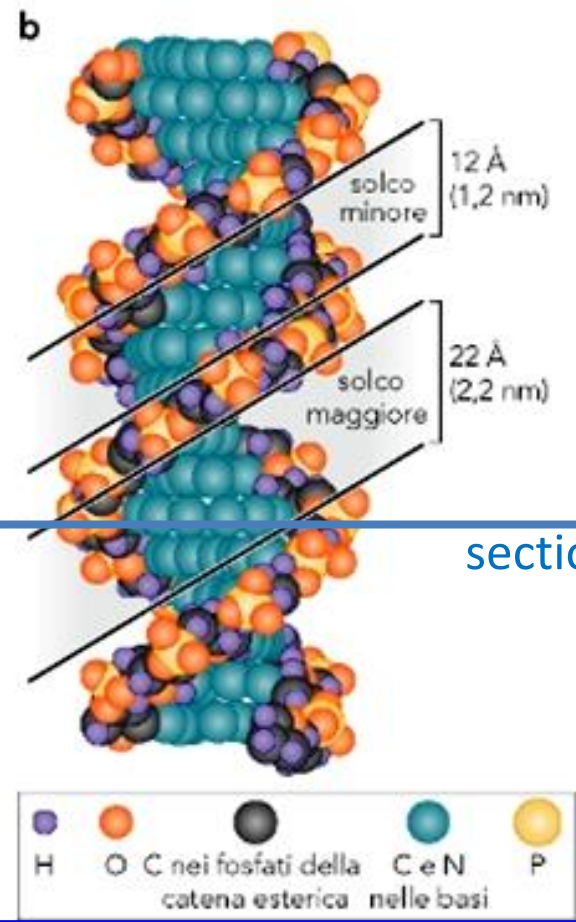
minor groove

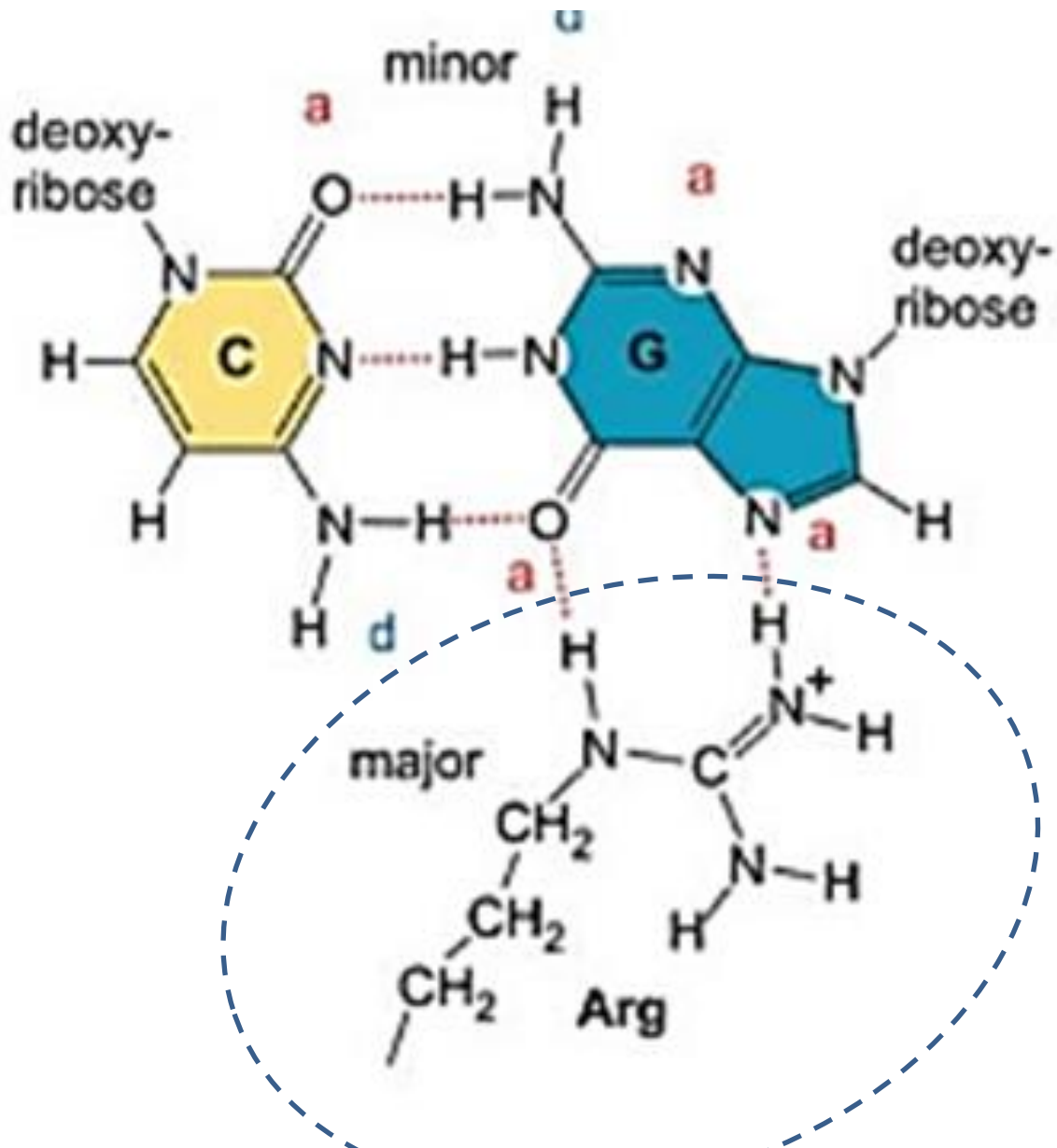
major groove

Stability is given mainly by ionic bonds with phosphates



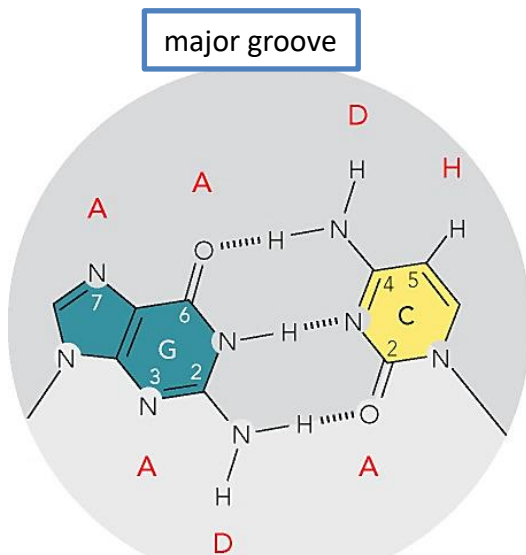
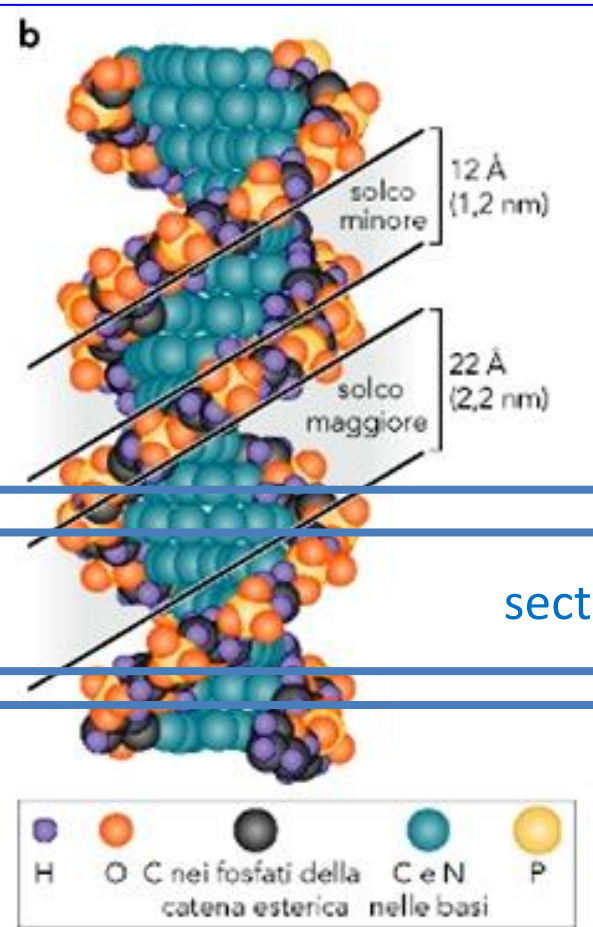
Specificity is given by stereospecific weak bonds with bases



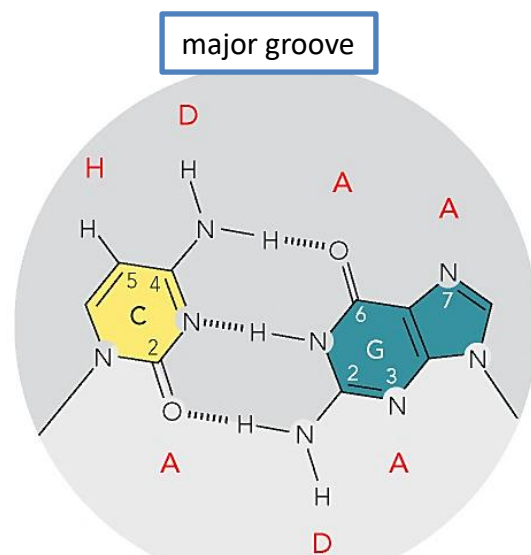


The reading code of DNA double helix

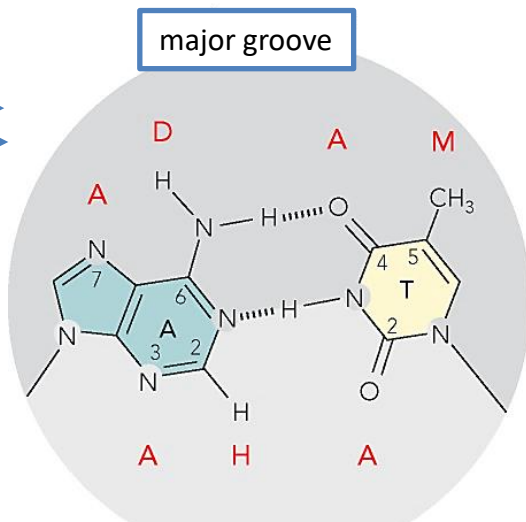
DNA-B



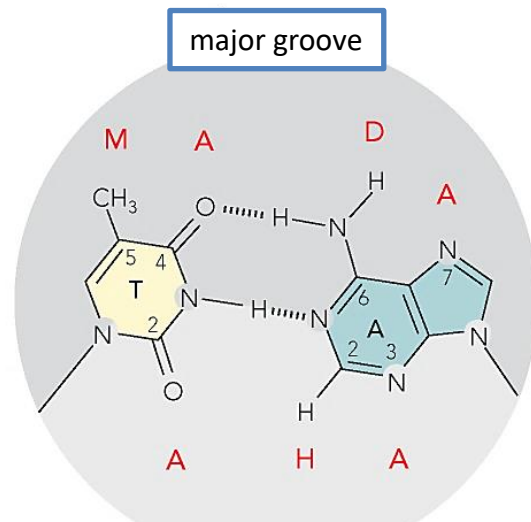
minor groove



minor groove

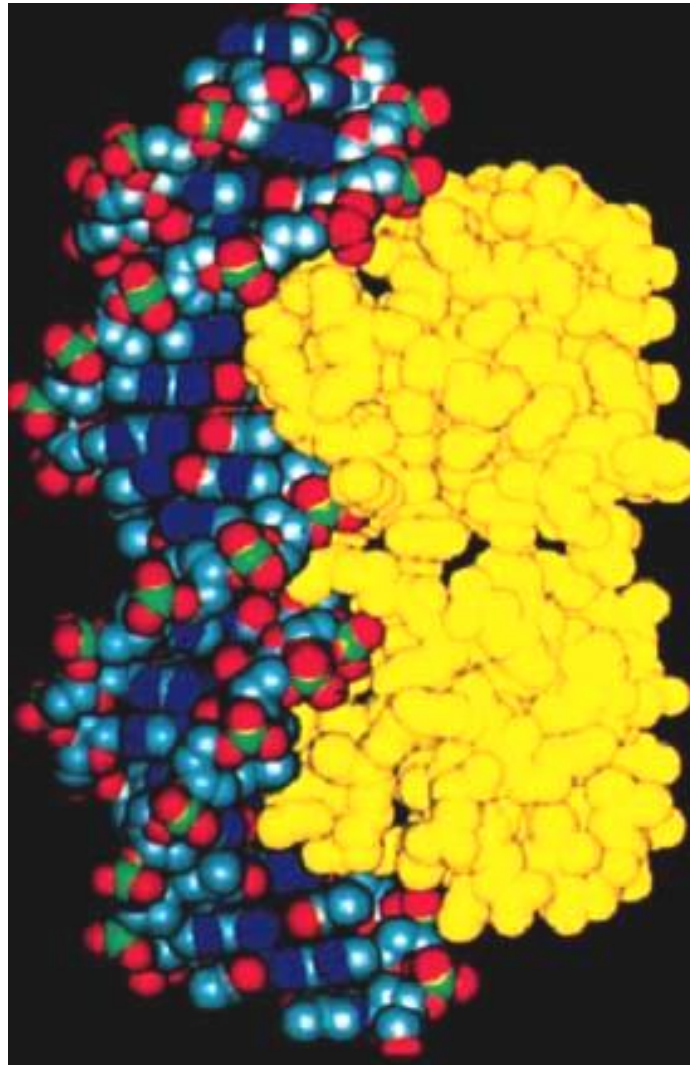


minor groove



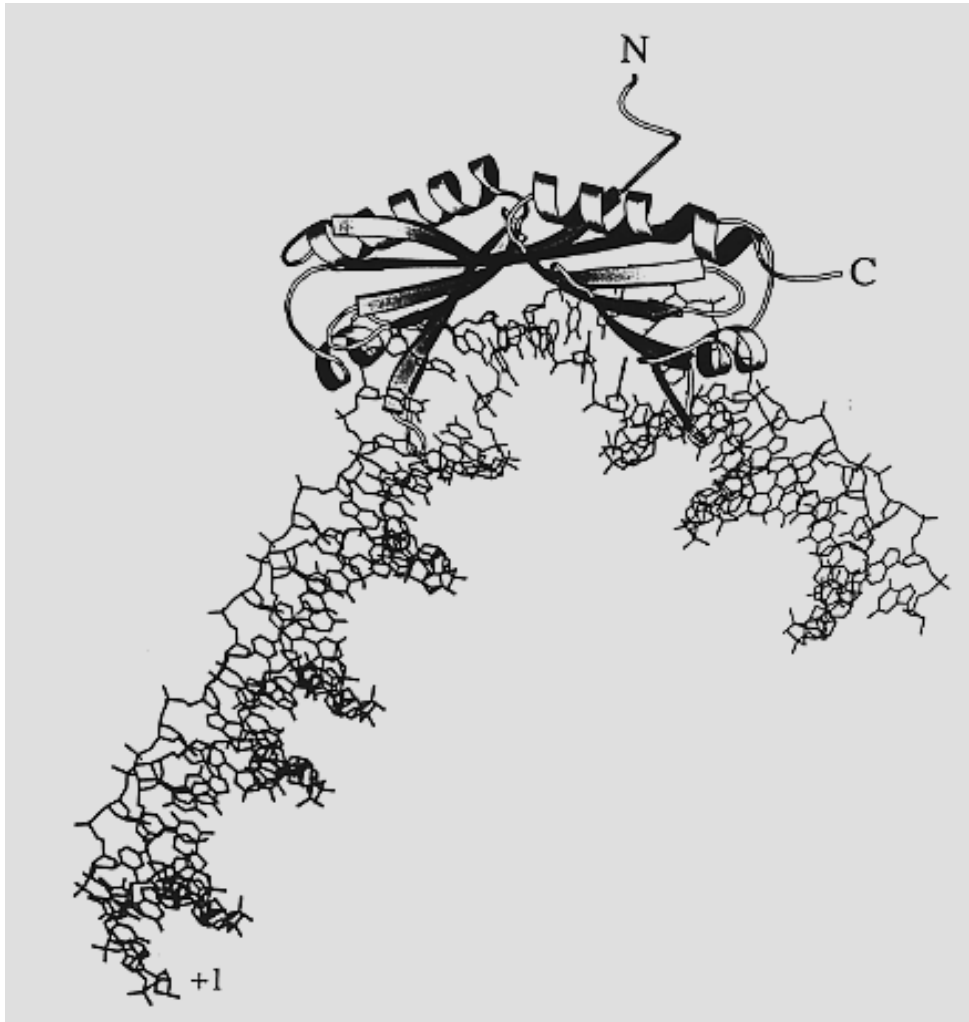
minor groove

A transcription factor protein bound to a specific nucleotide sequence of DNA, interacting with major grooves



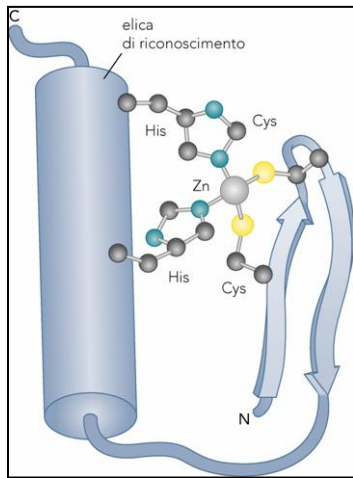
The DNA helix is bent: « **induced-fit** »

Extreme cases exist...



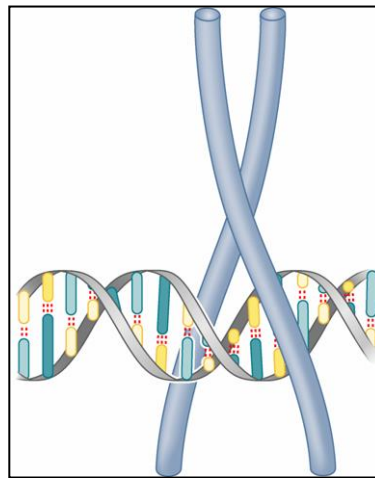
The TBP-TATA (DNA) complex

TBP recognizes its cognate sequence in the minor groove, but it is widely distorted during the interaction



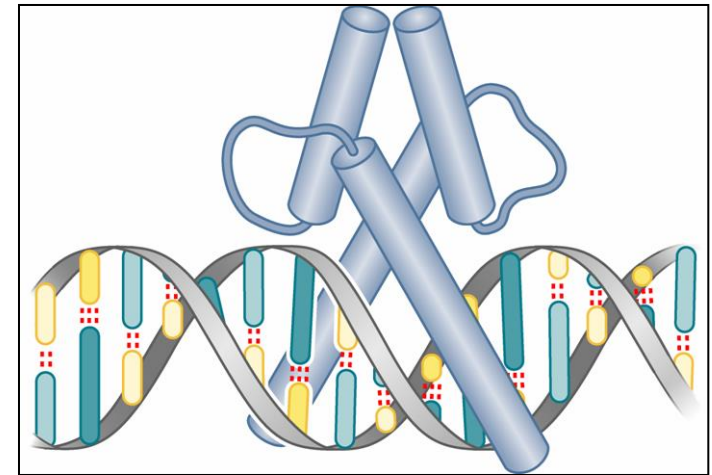
Zn-finger

GAL4, Nuclear receptors



b-ZIP

GCN4, fos-jun (AP-1),
CREB



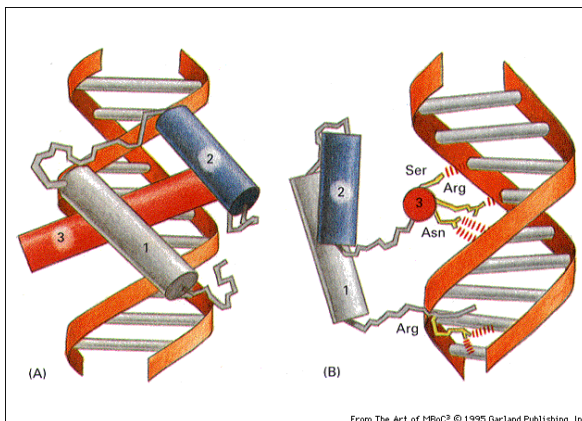
Helix-Loop-Helix (HLH)

Myc, Myo-D, Neuro-D, SREBP

Homeodomain

Hox proteins, Antennapedia

Mata



From The Art of MBoc © 1995 Garland Publishing, Inc.

Trans-activating domain classification is rather based on aminoacid composition, i.e.:

- acidic
- glutamine-rich
- glutamine/proline rich
- hydrophobic

DNA-binding Transcription Factors (regulatory factors) (TF)

GO category

Do not include:

- Coactivators or corepressors
- Enzymes
- General Transcription Factors (basal PIC components)

Do include:

- Putative proteins with similitude to known TFs
- Proteins that possess structural domains similar to DNA Binding domain (DBD) of known TFs

TF database: <http://jaspar.genereg.net/>

Structures (DBD) <http://www.rcsb.org/pdb/home/home.do>

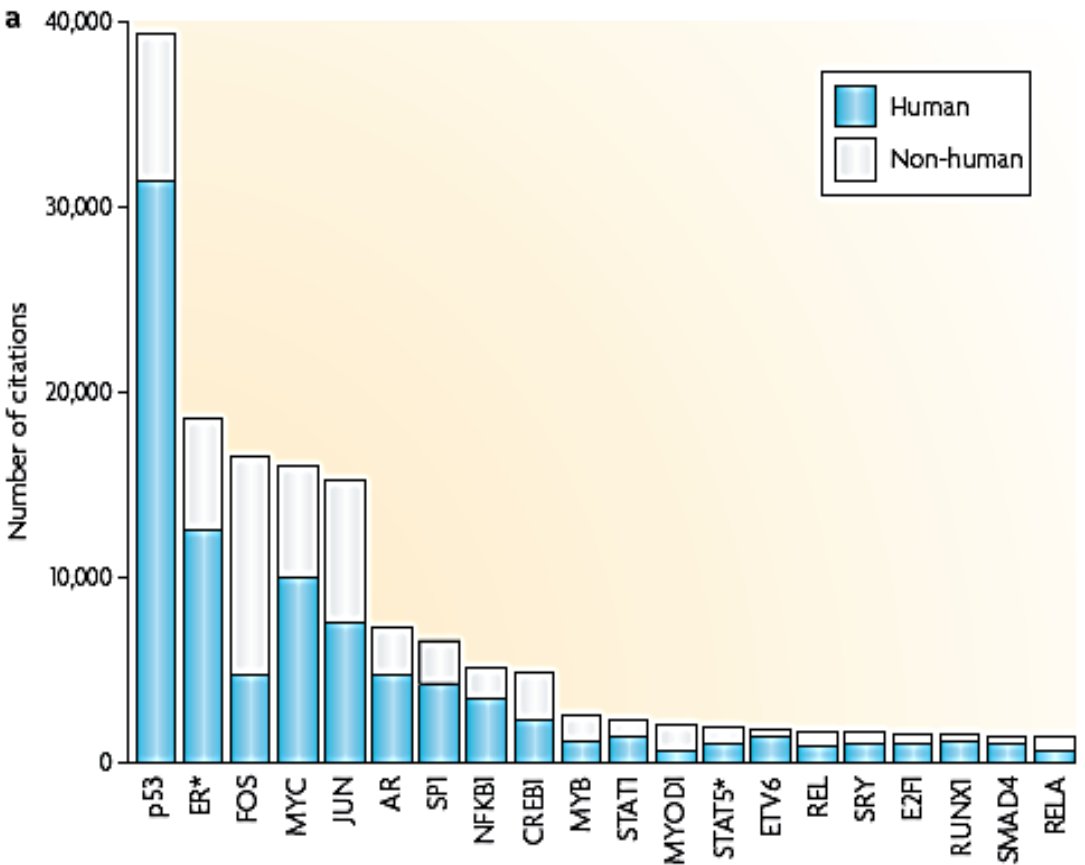
Yearly NAR database issues

<https://academic.oup.com/nar/article/48/D1/D1/5695332>

Figure 1

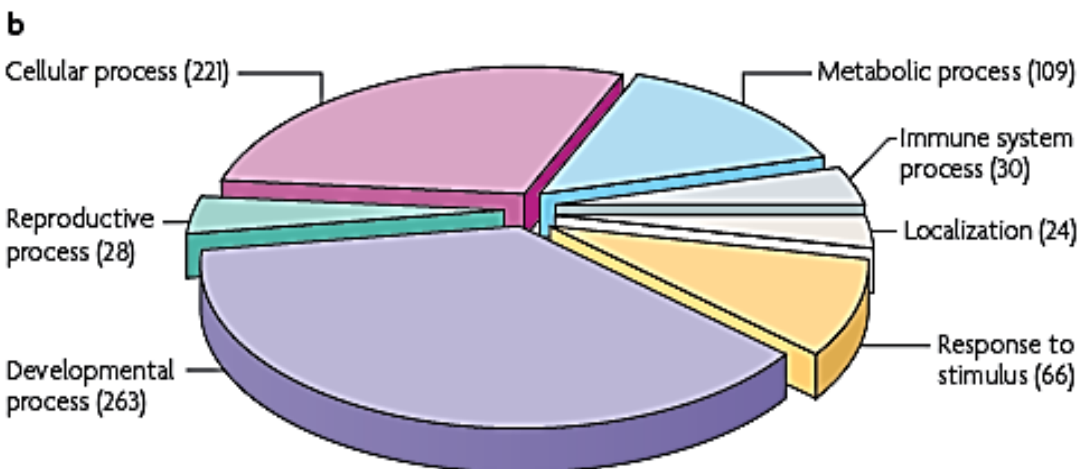
Current state **of knowledge** about transcription factors in the human genome.

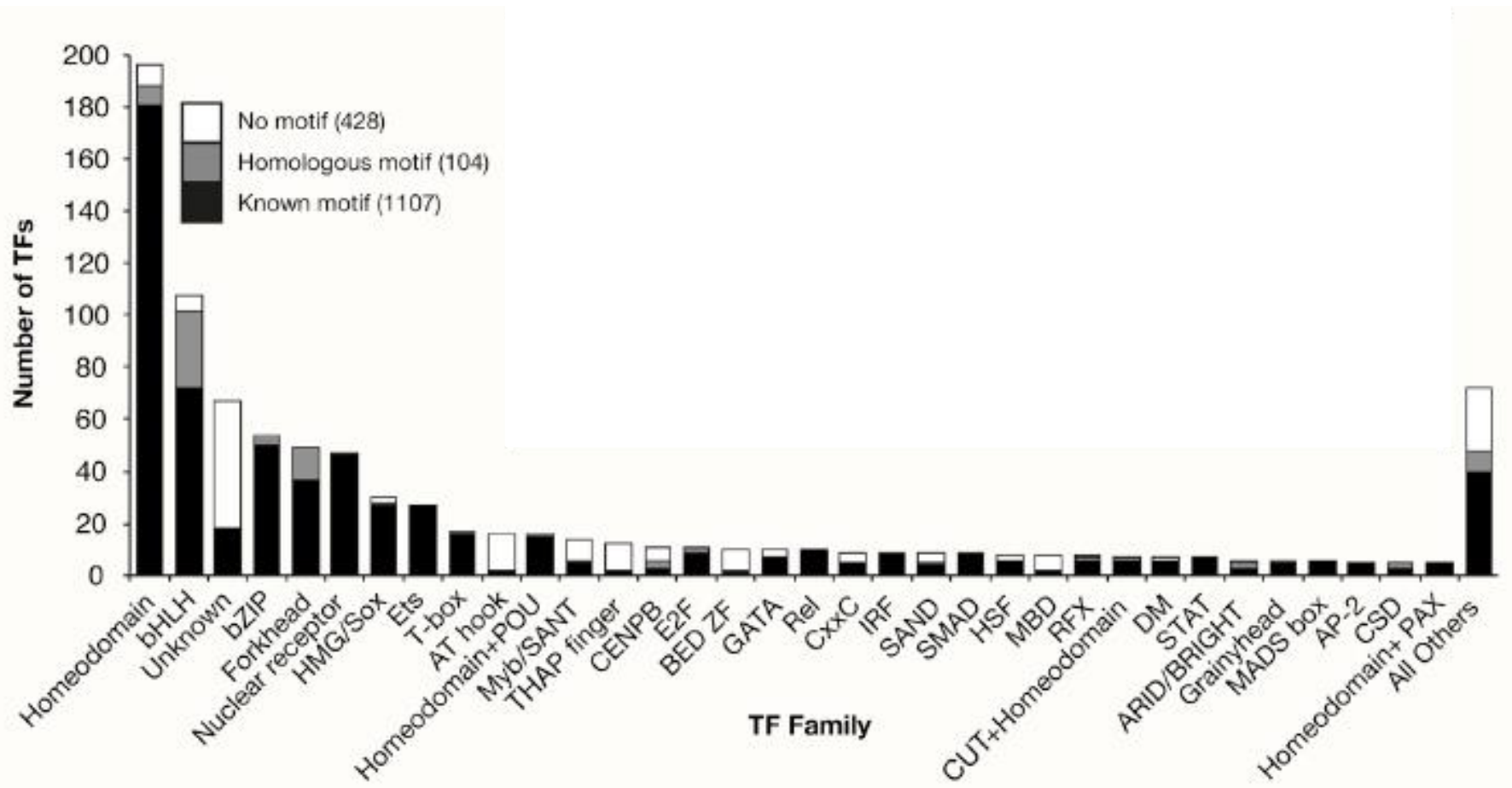
a | For the top 20 most cited transcription factors (TFs) in PubMed the number of studies performed in humans (blue bars) and in all other organisms (grey bars) is shown. ER* combines the citations for ERS1 and ERS2, which were indistinguishable in the literature search; similarly, STAT5* includes citations for both STAT5A and STAT5B.



b | summary of biological processes regulated by TFs.

Annotations were obtained from the Gene Ontology database, excluding those based only in electronic annotation. Numbers of annotated TFs are given in parentheses; each gene can be annotated with more than one function.





from Lambert 2018

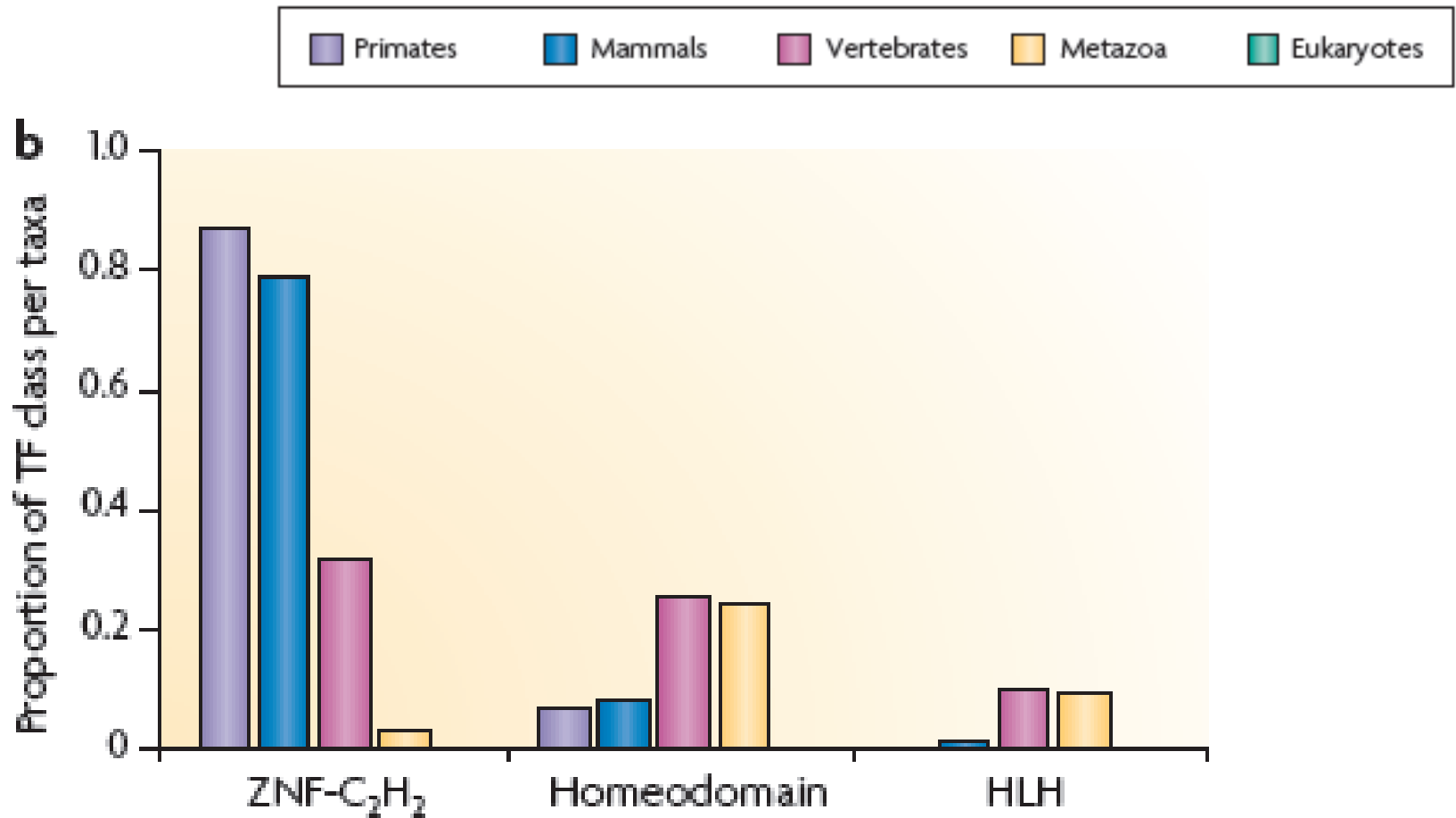
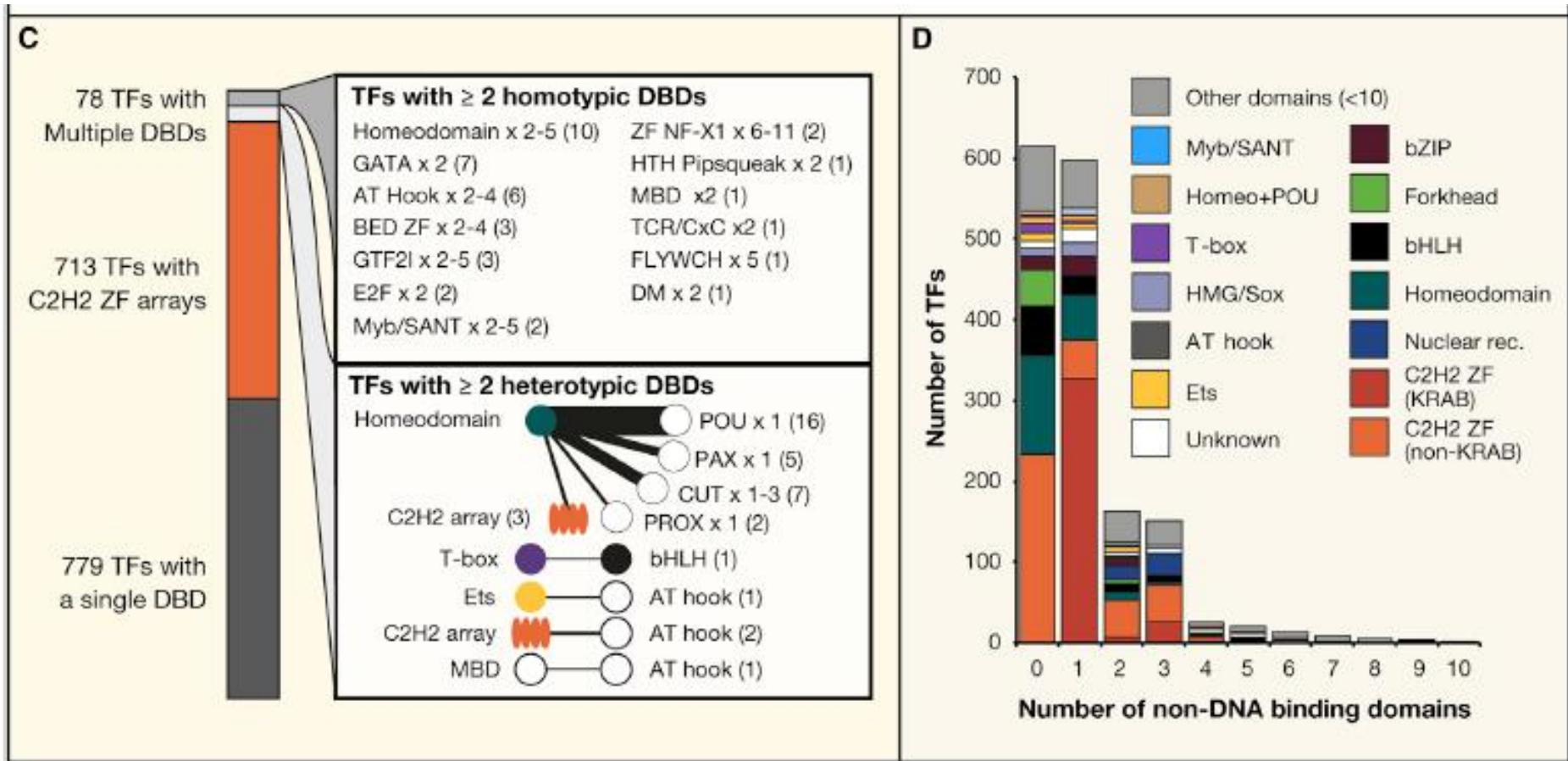
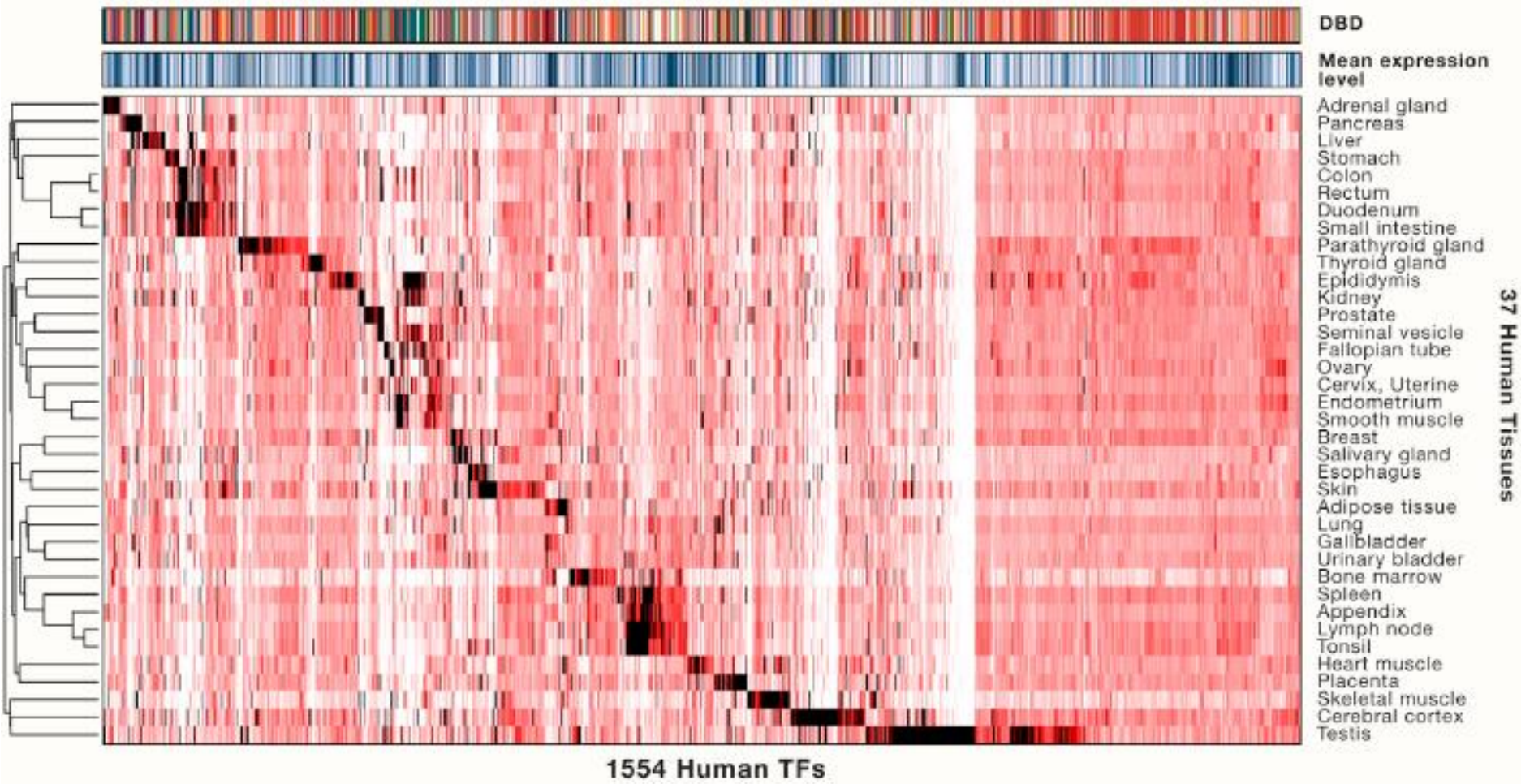


Figure 5 | **conservation of human transcription factors across 24 eukaryotic genomes.**
b | For human TFs in the three largest families, the proportion that are conserved in each taxonomic group is shown.



from Lambert 2018



tissue-specific

mostly ubiquitous