

C2H2锌指结构转录因子家族分析 ——以CTCF为例

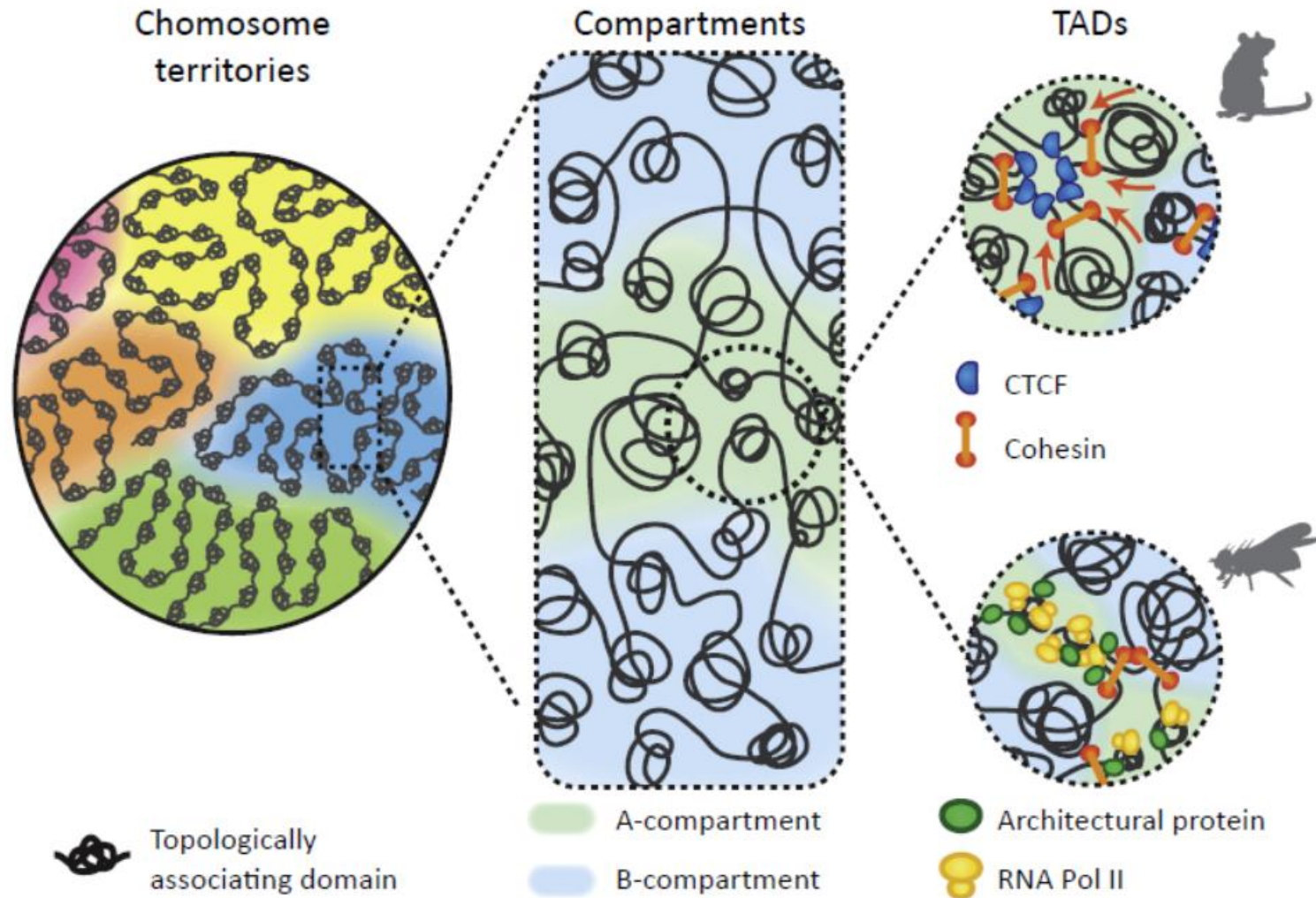
Analysis of C2H2 Zinc finger Domain Transcription Factor Family
——CTCF as a case

小组：G02

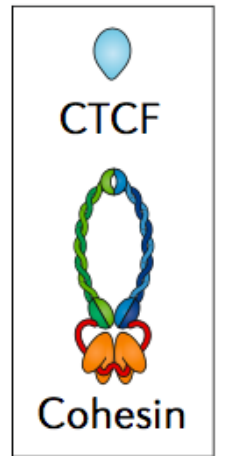
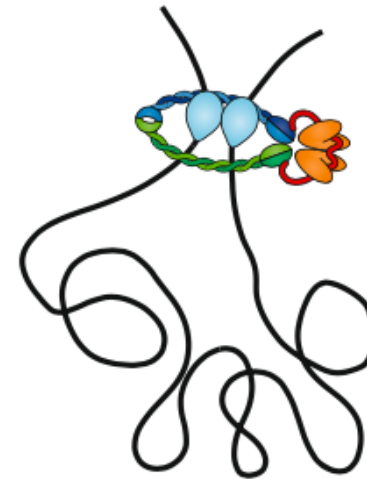
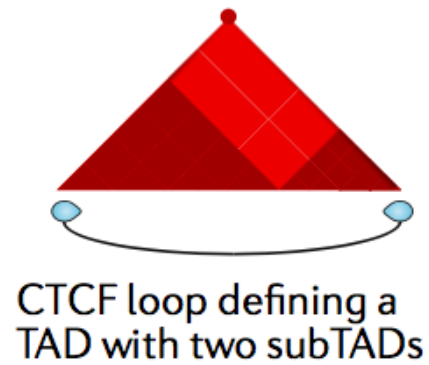
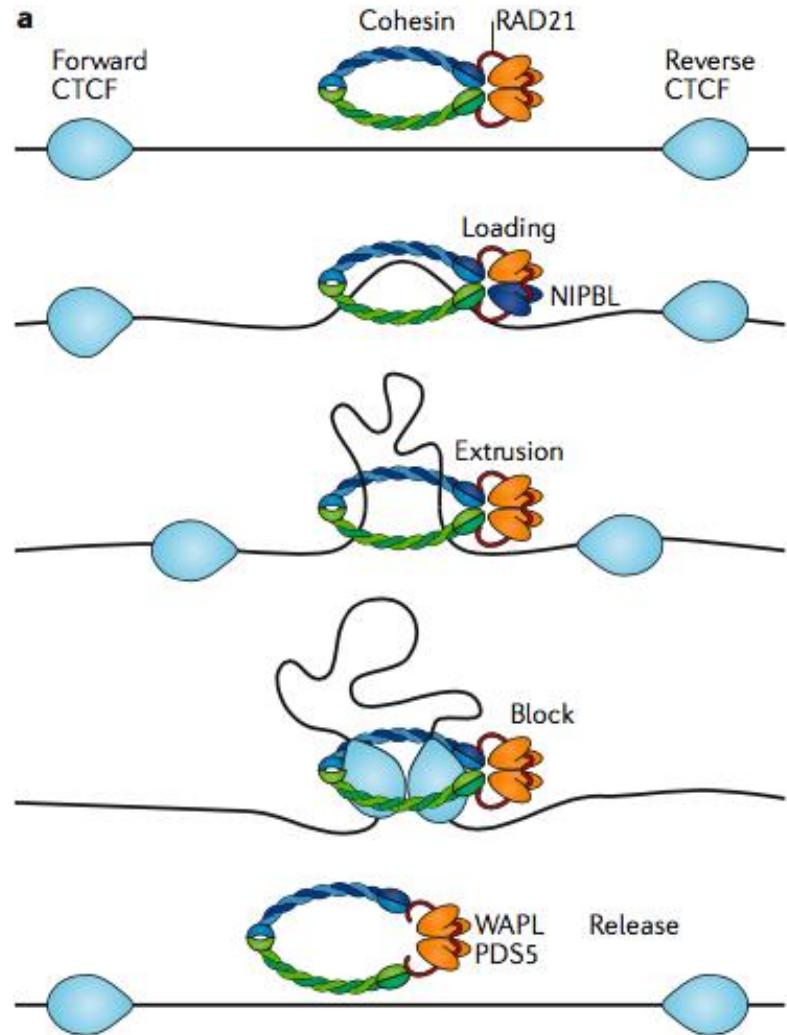
汇报人：段文嘉

组员：吕钰麟 王梦萱 崔英姿

真核生物细胞染色质结构组织

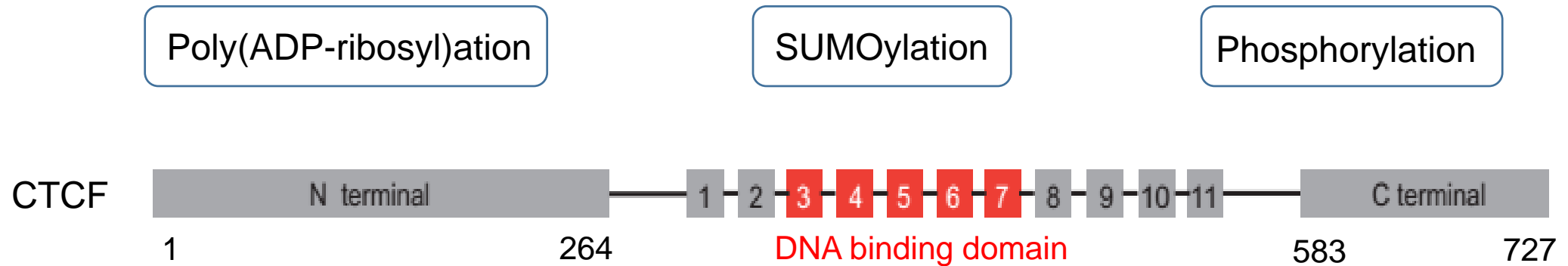


CTCF作为一种染色质结构蛋白介导染色质环的形成



M. Jordan Rowley and Victor G. Corces, 2018 *Nature reviews*

CTCF与多种蛋白结合发挥多种功能



DNA-binding proteins

- YB1
- YY1
- Kaiso
- CIITA

Chromatin proteins

- H2A
- SIN3A
- Cohesin
- Taf1/Set
- SUZ12
- CHD8

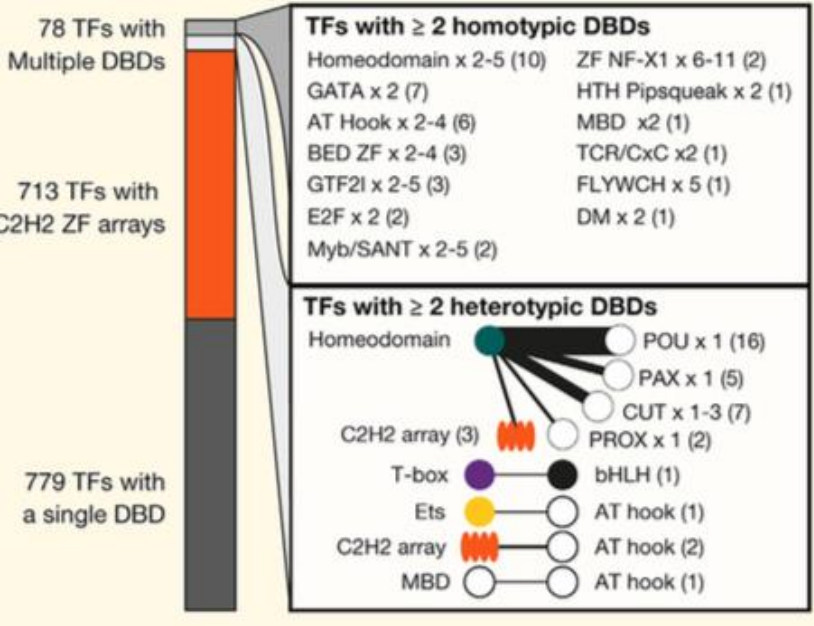
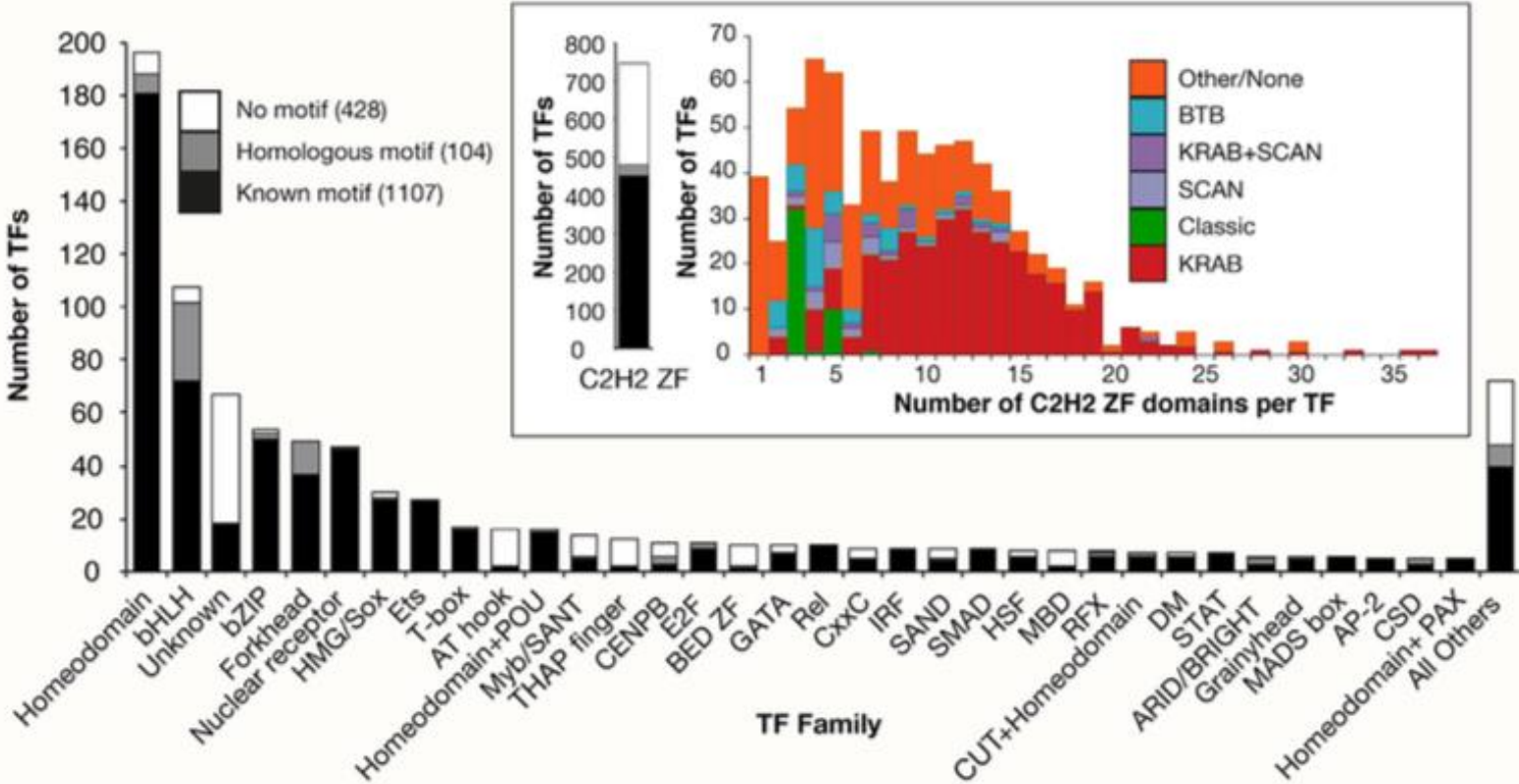
Multifunctional proteins

- PARP1
- Nucleophosmin
- Topo II

Scellaneoue proteins

- Lamin A/C
- Importins $\alpha 3$ and $\alpha 1$
- RNAP II
- CP190

C2H2转录因子家族是人类转录因子家族中最大的一类



以CTCF为例观察C2H2锌指与DNA双螺旋的结合形式



CTCF在有脊椎动物中高度保守

CTCF_HUMAN	1	MEGDAVEAIVEESETFIKGERKTYQRRREGGQEEDA----CH-----LPQNQTD	46
CTCF_MOUSE	1	MEGEAVEAIVEESETFIKGERKTYQRRREGGQEEDA----CH-----LPQNQTD	46
CTCF_RAT	1	MEGEAVEAIVEESETFIKGERKTYQRRREGGQEEDA----CH-----LPQNQTD	46
CTCF_CHICK	1	MEGEAVEAIVEESETFIKGERKTYQRRREGGQEEDA----CH-----IAPNQAD	46
Q6PF79_XENLA	1	MEGEMAEDIVEDSETFMKRKTKTYQRRREGGVDEEN----CV-----IVQSQTD	46
Q6JAA4_DANRE	1	MEGGPTEAVVEDAGDAFKAKECKTYQRRREDEEVGAELLOAAVIEQAQAEVPEPVVEAQQQ	60
		*** .* :***: :* ** *****. . : * :	
CTCF_HUMAN	47	GGEVVQDVNSSVQVMMEQLDPTLLQMKTEVMEGTVP-----EAEAAVDDTQII	96
CTCF_MOUSE	47	GGEVVQDVNSSVQVMMEQLDPTLLQMKTEVMEGTVP-----EAEAAVDDTQII	96
CTCF_RAT	47	GGEVVQDVNSSVQVMMEQLDPTLLQMKTEVMEGTVP-----EAEAAVDDTQII	96
CTCF_CHICK	47	GGEVVQDVNSGVQVMMEHLDPTLLQMKTEVMEGAVPQ-----ETEATVDDTQII	96
Q6PF79_XENLA	47	ICEVPHDVNSNVQVMMEQLDPTLLQMKTEVMEGMVSQ-----EGDPTVDDTQII	96
Q6JAA4_DANRE	61	LVESVSVNSSVDMMMETLDPALLQMKTEVMEAAVAGAPVAVAGAAHEATVTTVDDTQII	120
		* .***.**:*** ***:*****. * :*****	
CTCF_HUMAN	97	TLQVVNMEEQPINIGELQLVQVPVPTVPVATTVEELQAYENEVSKEGLAESEPMICH	156
CTCF_MOUSE	97	TLQVVNMEEQPINIGELQLVQVPVPTVPVATTVEELQAYENEVSKEGLAESEPMICH	156
CTCF_RAT	97	TLQVVNMEEQPINIGELQLVQVPVPTVPVATTVEELQAYENEVSKEGLAESEPMICH	156
CTCF_CHICK	97	TLQVVNMEEQPINLQELQLVQVPVPTVPVATTVEELQAYENEVSKGGLQEGEPMICH	156
Q6PF79_XENLA	97	TLQVVNMEEQPINLQELQLVQVP--VAVPMATTVEELQAYENEVSKVLEGEPMICH	154
Q6JAA4_DANRE	121	TLQVVNMEEQQLGLQELQLVQVPVS-AVPVTAATVEELQGTLDAT--AMPKDGEPVICH	177
		***** :.:***** :***:*** ***: : . :.***:***	
CTCF_HUMAN	157	TLPLPEGFQVVKVGANGEVETLEQGELPPQ-----EDPSWQKDPDYQ	198
CTCF_MOUSE	157	TLPLPEGFQVVKVGANGEVETLEQGELPPQ-----EDSSWQKDPDYQ	198
CTCF_RAT	157	TLPLPEGFQVVKVGANGEVETLEQGELPPQ-----EDPSWQKDPDYQ	198
CTCF_CHICK	157	TLPLPEGFQVVKVGANGEVETLEQELQPQ-----EDPNWQKDPDYQ	198
Q6PF79_XENLA	155	TLPLPEGFQVVKVGANGEVETLEQAELQPQ-----EEPGWQKDPDYV	196
Q6JAA4_DANRE	178	TLPLPEGFQVVKVGANGEVETVEQDELQPQDDQPPHQEEEEEMAEPQEDPAWSKDPDYT	237
		*****:*** ** ** * : * .*****	
CTCF_HUMAN	199	PPAKKTKKTKSKLRYTEEG-KDVDVSVYDFEEEEQEGLLSEVNAEKVVGNMPPKPTKI	257
CTCF_MOUSE	199	PPAKKTKKTKSKLRYTEEG-KDVDVSVYDFEEEEQEGLLSEVNAEKVVGNMPPKPTKI	257
CTCF_RAT	199	PPAKKTKKTKSKLRYTEEG-KDVDVSVYDFEEEEQEGLLSEVNAEKVVGNMPPKPTKI	257
CTCF_CHICK	199	PPAKKTKKTKSKLRYTEEG-KDVDVSVYDFEEEEQEGLLSEVNAEKVVGNMPPKPTKI	257
Q6PF79_XENLA	197	PPMKSKKTKKSKLRYTEEG-KDVDVSVYDFEEEEQEGLLSDVNAEKVVGNMPPKPTKI	255
Q6JAA4_DANRE	238	PPVKKVKKTKKSKLRYNTEGDKMDVSVYDFEEEEQEGLLSEVNAEKVVGNMPPKPTKI	297
		** ** ** .*****. ** **:*****:*****:*****	

CTCF_HUMAN	258	KKKGVKKT	FQCELCSYTCPRRSNLDHRHMKSH	TDERP	HKCHLCGRAFRTVTLLRNHLNHT	317			
CTCF_MOUSE	258	KKKGVKKT	FQCELCSYTCPRRSNLDHRHMKSH	TDERP	HKCHLCGRAFRTVTLLRNHLNHT	317			
CTCF_RAT	258	KKKGVKKT	FQCELCSYTCPRRSNLDHRHMKSH	TDERP	HKCHLCGRAFRTVTLLRNHLNHT	317			
CTCF_CHICK	258	KKKGVKKT	FQCELCSYTCPRRSNLDHRHMKSH	TDERP	HKCHLCGRAFRTVTLLRNHLNHT	317			
Q6PF79_XENLA	256	KKKGVKKT	FQCELCSYTCPRRSNLDHRHMKSH	TDERP	HKCHLCGRAFRTVTLLRNHLNHT	315			
Q6JAA4_DANRE	298	KKKGVKKT	FQCELCSYTCPRRSNLDHRHMKSH	TDERP	HKCHLCGRAFRTVTLLRNHLNHT	357			

CTCF_HUMAN	318	GTRP	HKCPDCDMAFVTS	GELVRHRRYKHTHEKP	FKCSMCDYASVEVSKLKRHIRSH	TGER	377		
CTCF_MOUSE	318	GTRP	HKCPDCDMAFVTS	GELVRHRRYKHTHEKP	FKCSMCDYASVEVSKLKRHIRSH	TGER	377		
CTCF_RAT	318	GTRP	HKCPDCDMAFVTS	GELVRHRRYKHTHEKP	FKCSMCDYASVEVSKLKRHIRSH	TGER	377		
CTCF_CHICK	318	GTRP	HKCPDCDMAFVTS	GELVRHRRYKHTHEKP	FKCSMCDYASVEVSKLKRHIRSH	TGER	377		
Q6PF79_XENLA	316	GTRP	HKCPDCDMAFVTS	GELVRHRRYKHTHEKP	FKCSMCDYASVEVSKLKRHIRSH	TGER	375		
Q6JAA4_DANRE	358	GTRP	HKCTDCDMAFVTS	GELVRHRRYKHTHEKP	FKCSMCDYASVEVSKLKRHIRSH	TGER	417		

CTCF_HUMAN	378	PF	FQCSLCSYASRD	TYKLRHMRTHS	GKEKPYECYI	CHARFTQSGTMKMHILQKH	TENVAKF	437	
CTCF_MOUSE	378	PF	FQCSLCSYASRD	TYKLRHMRTHS	GKEKPYECYI	CHARFTQSGTMKMHILQKH	TENVAKF	437	
CTCF_RAT	378	PF	FQCSLCSYASRD	TYKLRHMRTHS	GKEKPYECYI	CHARFTQSGTMKMHILQKH	TENVAKF	437	
CTCF_CHICK	378	PF	FQCSLCSYASRD	TYKLRHMRTHS	GKEKPYECYI	CHARFTQSGTMKMHILQKH	TENVAKF	437	
Q6PF79_XENLA	376	PF	FQCSLCSYASRD	TYKLRHMRTHS	GKEKPYECYI	CHARFTQSGTMKMHILQKH	TENVAKF	435	
Q6JAA4_DANRE	418	PF	FQCSLCSYASRD	TYKLRHMRTHS	GKEKPYECYI	CHARFTQSGTMKMHILQKH	TENVAKF	477	

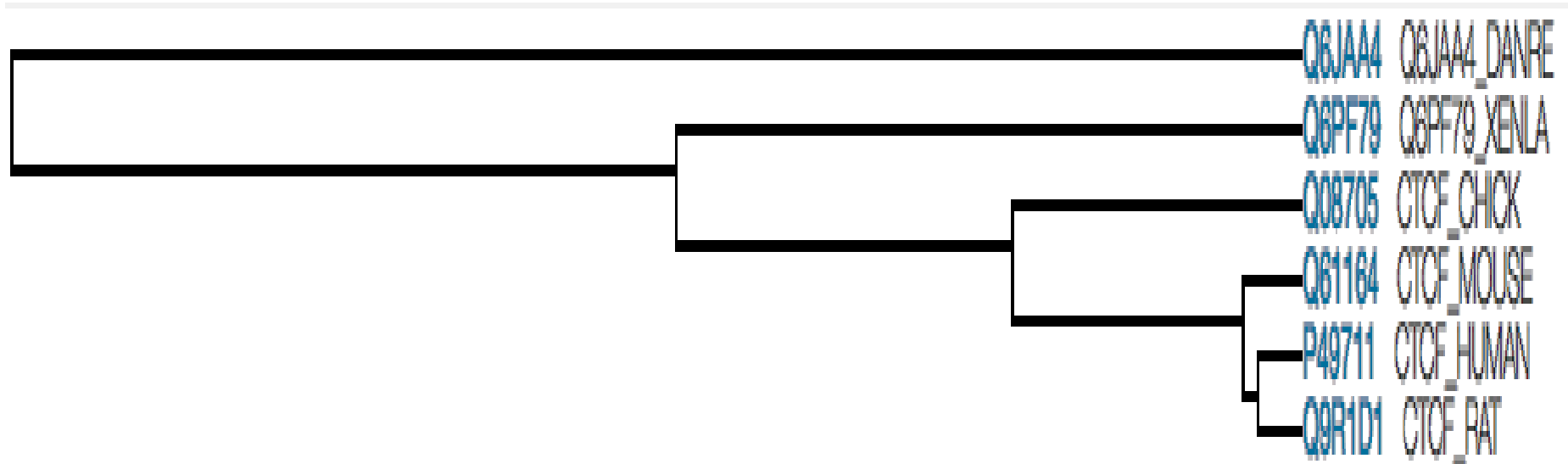
CTCF_HUMAN	438	HCP	HC	PHCDTVIARKSD	LGVHLRKQHSYIEQG	KKCRYCDAVPHERYALIQHOKSH	KNEKRFKC	497	
CTCF_MOUSE	438	HCP	HC	PHCDTVIARKSD	LGVHLRKQHSYIEQG	KKCRYCDAVPHERYALIQHOKSH	KNEKRFKC	497	
CTCF_RAT	438	HCP	HC	PHCDTVIARKSD	LGVHLRKQHSYIEQG	KKCRYCDAVPHERYALIQHOKSH	KNEKRFKC	497	
CTCF_CHICK	438	HCP	HC	PHCDTVIARKSD	LGVHLRKQHSYIEQG	KKCRYCDAVPHERYALIQHOKSH	KNEKRFKC	497	
Q6PF79_XENLA	436	HCP	HC	PHCDTVIARKSD	LGVHLRKQHSYIEQG	KKCRYCDTVFHERYALIQHOKSH	KNEKRFKC	495	
Q6JAA4_DANRE	478	HCP	HC	PHCDTVIARKSD	LGVHLRKQHSYIEQGR	KCRYCDAVPHERYALIQHOKSH	KNEKRFKC	537	

CTCF_HUMAN	498	DQ	CDYACRQERHM	MIMHKRTH	TGEKPYACSH	CDKTFRQKQLDMHF	KRYHDPNFVPAA	FVC	557
CTCF_MOUSE	498	DQ	CDYACRQERHM	MIMHKRTH	TGEKPYACSH	CDKTFRQKQLDMHF	KRYHDPNFVPAA	FVC	557
CTCF_RAT	498	DQ	CDYACRQERHM	MIMHKRTH	TGEKPYACSH	CDKTFRQKQLDMHF	KRYHDPNFVPAA	FVC	557
CTCF_CHICK	498	DQ	CDYACRQERHM	MIMHKRTH	TGEKPYACSH	CDKTFRQKQLDMHF	KRYHDPNFVPAA	FVC	557
Q6PF79_XENLA	496	DQ	CEYACRQERHM	MIMHKRTH	TGEKPYACSH	CDKTFRQKQLDMHF	KRYHDPNFVPAA	FVC	555
Q6JAA4_DANRE	538	DQ	CDYACRQERHM	MIMHKRTH	TGEKPYAC	SQCEKTFRQKQLDMHF	FRRYHDPNFVPTS	FVC	597
;**;*****;*:*****;*****.***;:***									

CTCF_HUMAN	558	SKCGKTFTRRNTMARHADNCAGPDGVEGENGGE----TKKSKRGRKRKMRSKKEDS--SD	611
CTCF_MOUSE	558	SKCGKTFTRRNTMARHADNCAGPDGVEGENGGE----TKKSKRGRKRKMRSKKEDS--SD	611
CTCF_RAT	558	SKCGKTFTRRNTMARHADNCAGPDGVEGENGGE----TKKSKRGRKRKMRSKKEDS--SD	611
CTCF_CHICK	558	SKCGKTFTRRNTMARHADNCAGPDGVEGENGGE----TKKSKRGRKRKMRSKKEDS--SD	611
Q6PF79_XENLA	556	SKCGKTFTRRNTMSRHADSC TGP DGT DGENGEEGEVIHKKGKRGRKRKMRSKKEGS--TD	613
Q6JAA4_DANRE	598	TKCGKTFTRRNTMARHAENCTGMSADGENGTP----PKRGRGGRKRKMRSRKDDDDDD	653
		:*****:***:.*: * . :**** *:: *****:*. . *	
CTCF_HUMAN	612	S-ENAEPDLDDNEDEEEPAVEIEPEPEP-----QPVTPAPPPAKKRRGRPPGRTN	660
CTCF_MOUSE	612	SEENAEPDLDDNEEEEPAVEIEPEPEPQPQP--PPPPQP VAPAPPPAKKRRGRPPGRTN	669
CTCF_RAT	612	S-ENAEPDLDDNEEEEPAVEIEPEPEPQPQPQPQPQP VAPAPPPAKKRRGRPPGRTN	670
CTCF_CHICK	612	SEENAEPDLDDNEDEEETAVEIEAEPEVS-----AEAPAPPPSKKRRGRPPGKAA	661
Q6PF79_XENLA	614	SEDNAEPDLDDDEDEDDEEEET---PVEIEADPEPEEPVSPIPPPAKKRRGRPPGKAN	670
Q6JAA4_DANRE	654	SDEHGEPDLDDIDEDEDLLEDQ--GLLDQAPPSVPI PAPAEPPIKRKRGRPPKNAP	711
		* :. .**:* ** : : : : *	
CTCF_HUMAN	661	-----QPKQNQPTAIIQVEDQNTGAIENIIVEVKKEPDAEP-----AE----GE	700
CTCF_MOUSE	670	-----QPKQNQPTAIIQVEDQNTGAIENIIVEVKKEPDAEP-----AE----GE	709
CTCF_RAT	671	-----QPKQNQPTAIIQVEDQNTGAIENIIVEVKKEPDAEP-----AE----GE	710
CTCF_CHICK	662	T-----QTKQSQPAAIIQVEDQNTGEIENIIVEVKKEPDAET-----VE----EE	702
Q6PF79_XENLA	671	Q-----A---KQNAAVIQVEDHNTRAIENIIVQVKKESDLEA-----EV----VV	708
Q6JAA4_DANRE	712	KVSPTKSITKTTTAAAIQVEDESTGAIENIIVKKEPEGTDAVVAAQPIIEEVEAVEADV	771
		. :*:* ** . . * * * * * : : *	
CTCF_HUMAN	701	EEEAQPAATDAPNGDLTPEMILSMMDR	727
CTCF_MOUSE	710	EEEAQAATTDAPNGDLTPEMILSMMDR	736
CTCF_RAT	711	EEEAQAAPADAPNGDLTPEMILSMMDR	737
CTCF_CHICK	703	-EEAQPAVVEAPNGDLTPEMILSMMDR	728
Q6PF79_XENLA	709	EAPVLT PAVEAPNGDLTPEMILSMMDR	735
Q6JAA4_DANRE	772	ETVQLTVPEAAPNGDLTPEMILSMMDR	798

Between avian and mammalian CTCF proteins, 93% of amino acids are identical. However, the identity rises to 100% for the region containing the 11 ZFs.


利用CTCF构建的系统发生树与进化树一致



利用结构域识别网站SMART分析转录因子CTCF结构



查找人类转录因子数据库HumanTFs



☰ Browse TFs by DBD family

☰ All considered proteins

☰ All TFs

☰ Download

☰ Links

☰ How to cite

☰ Contact

The Human Transcription Factors

Samuel Lambert^{1,*}, Arttu Jolma^{2,*}, Laura Campitelli^{1,*}, Pratyush Das³, Yimeng Yin⁴, Mihai Albu², Xiaoting Chen⁵, Jussi Taipale^{3,4,6,^}, Timothy R. Hughes^{1,2,^}, and Matthew T. Weirauch^{5,7,8,^}

1 Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
2 Donnelly Centre, University of Toronto, Toronto, ON, Canada
3 Genome-Scale Biology Program, University of Helsinki, Helsinki, Finland
4 Division of Functional Genomics and Systems Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna, Sweden
5 Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
6 Department of Biochemistry, Cambridge University, Cambridge, Great Britain
7 Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
8 Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

^ To whom correspondence should be addressed:
Matthew.Weirauch@cchmc.org
t.hughes@utoronto.ca
ajt208@cam.ac.uk

Abstract

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.

在HumanTFs找到具有C2H2锌指结构的其他转录因子

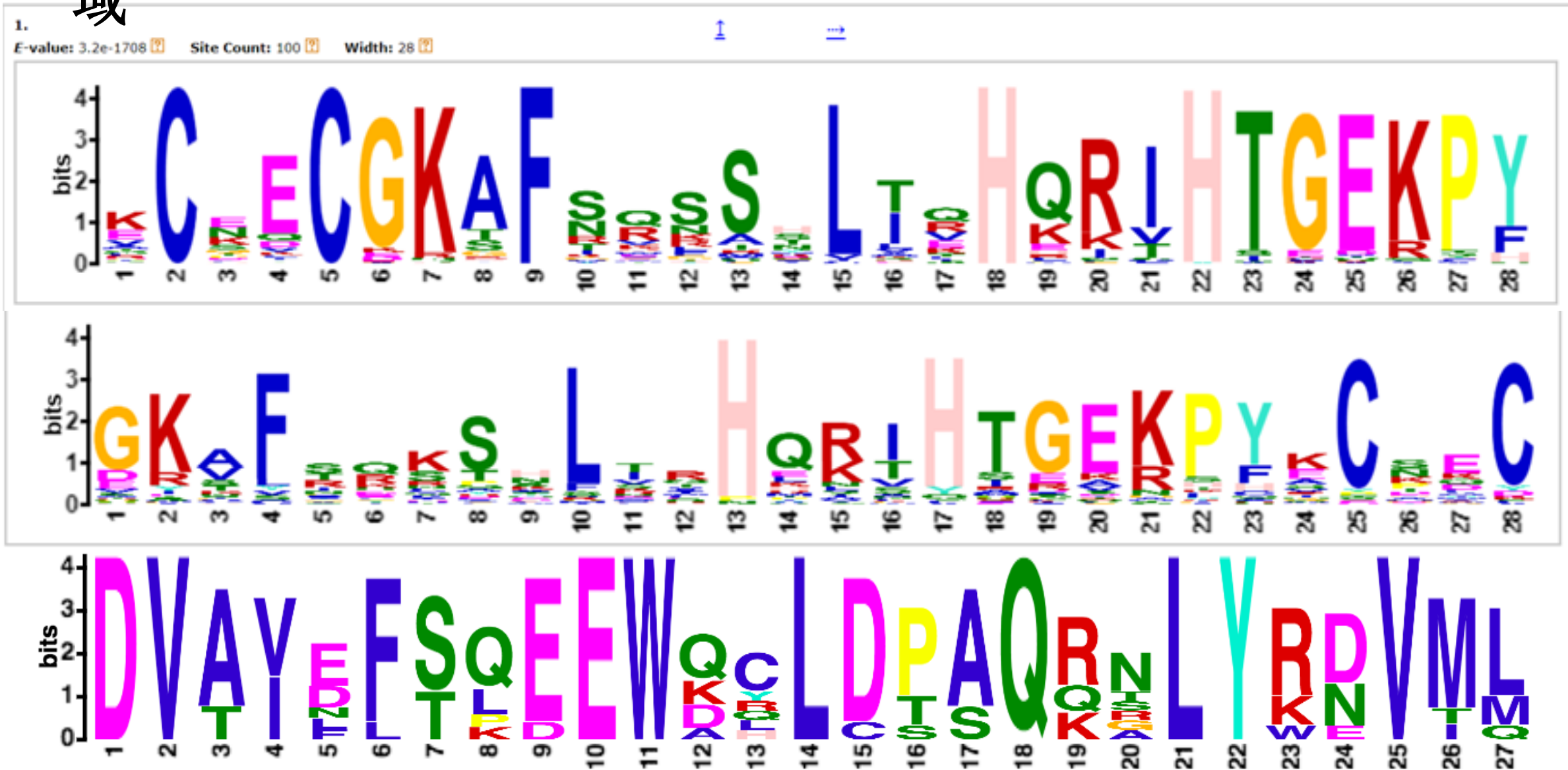
TFs and their motifs, by family

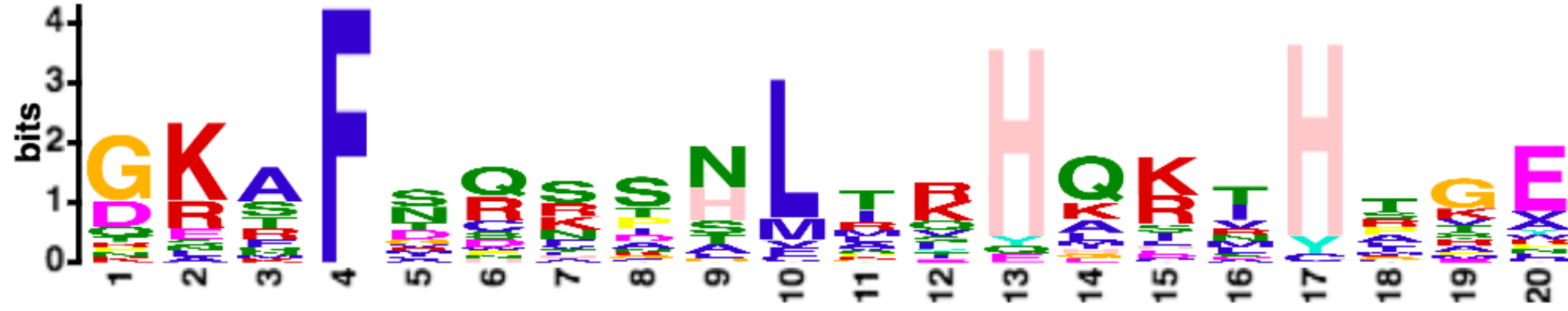
100 records per page

Search:

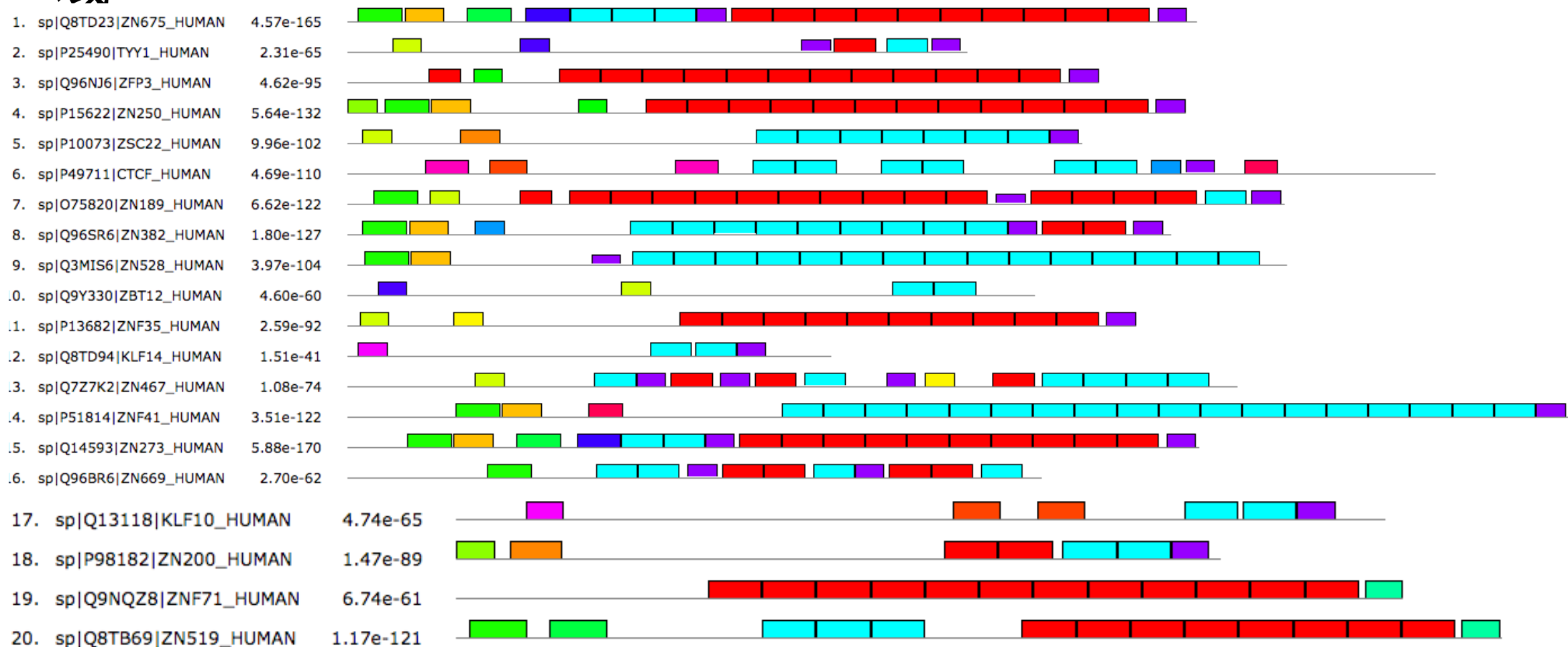
DBD Family	TF Assessment	Binding Mode	TFs	Considered Proteins
C2H2 ZF(non-KRAB)			403	403
C2H2 ZF(KRAB)			344	344

利用MEME鉴定人类C2H2家族蛋白中不同成员的保守结构域





利用MEME鉴定人类C2H2家族蛋白中不同成员的保守结构域



谢谢大家！