

# BTB-Zinc Finger Domain Protein Conservation Analysis

Group15 Members: Dongsheng Bai

Jianhang Yin

Mengjie Sun

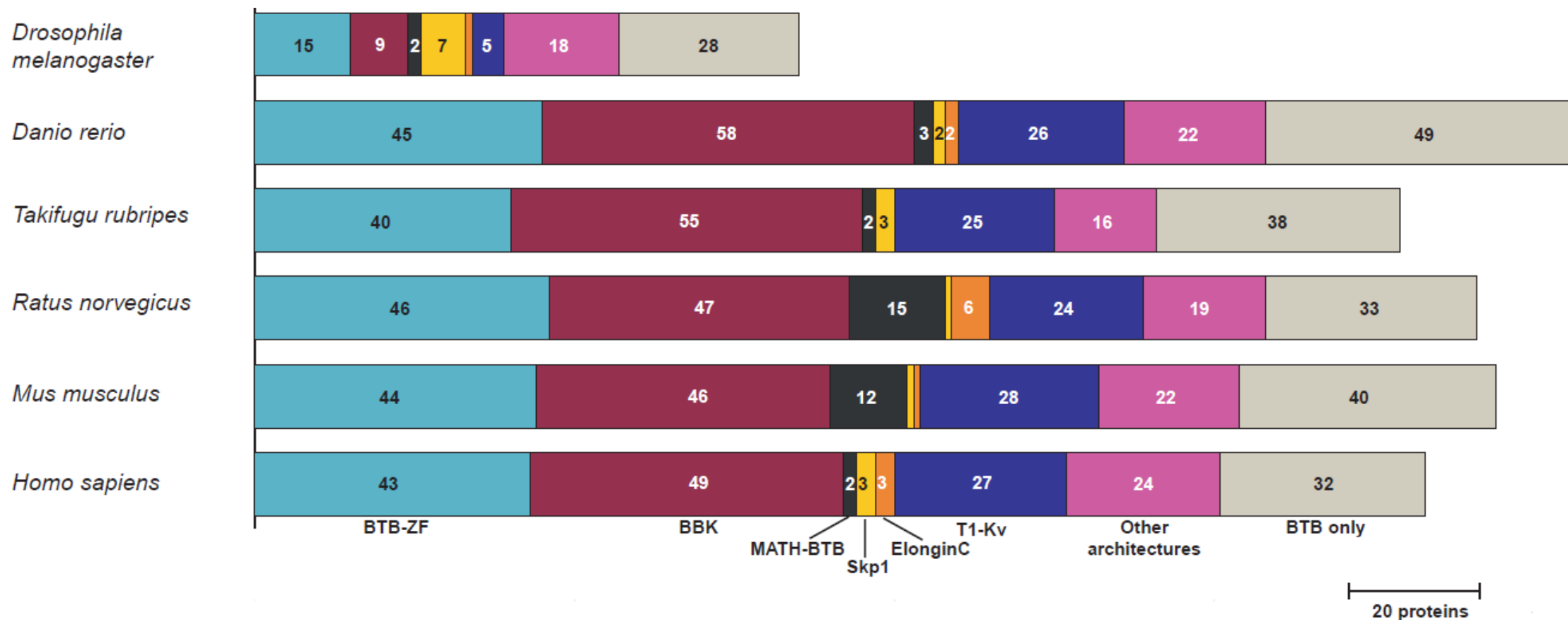
Ziqian Li

01/07/2017

# Outline:

- **Introduction of the BTB-zinc finger protein family**
- **BTB domain function and sequence conservation analysis**
- **Zinc finger motif prediction and analysis**

# BTB domain exists in many protein families



(Stogios et al., 2005)

# BTB-zinc finger protein structure



BTB domain:



annotation rule: PRU00037

Function: homodimerization domain

Zinc-finger domain:



annotation rule: PRU00042

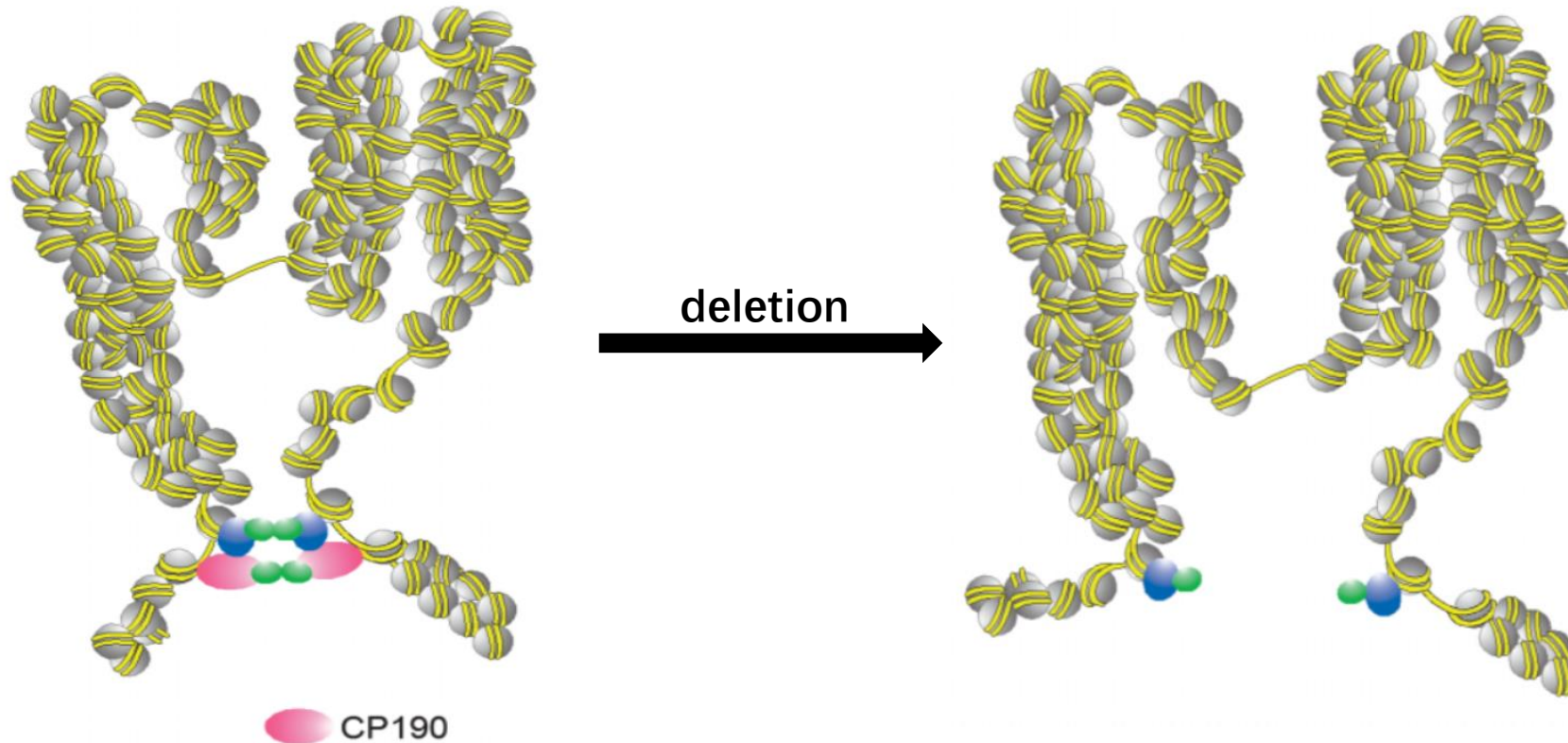
Function: DNA binding domain

(Stogios et al., 2005)

# BTB-zinc finger protein's role in Drosophila

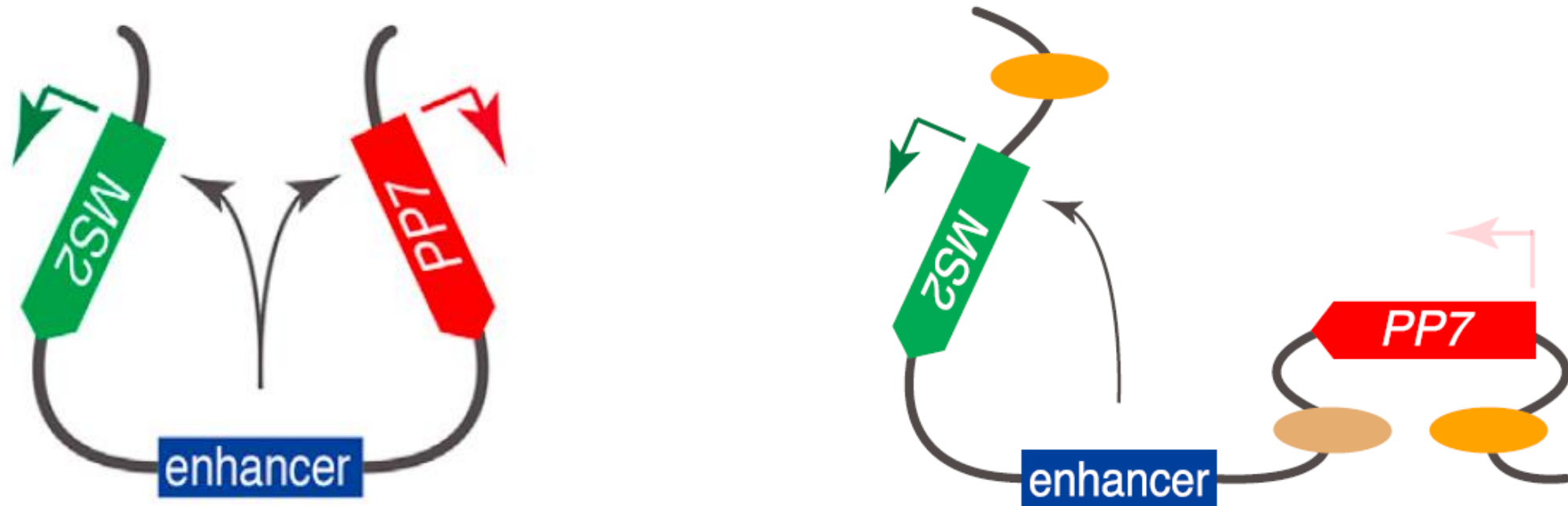
Molecular Cell, Vol. 16, 737-748, December 3, 2004, Copyright ©2004 by Cell Press

## The Centrosomal Protein CP190 Is a Component of the *gypsy* Chromatin Insulator



(Pai et al., 2004)

# Loop is important for enhancer-promoter interaction



(Fukaya et al., 2016)

# Alignment of btb domain in Drosophila

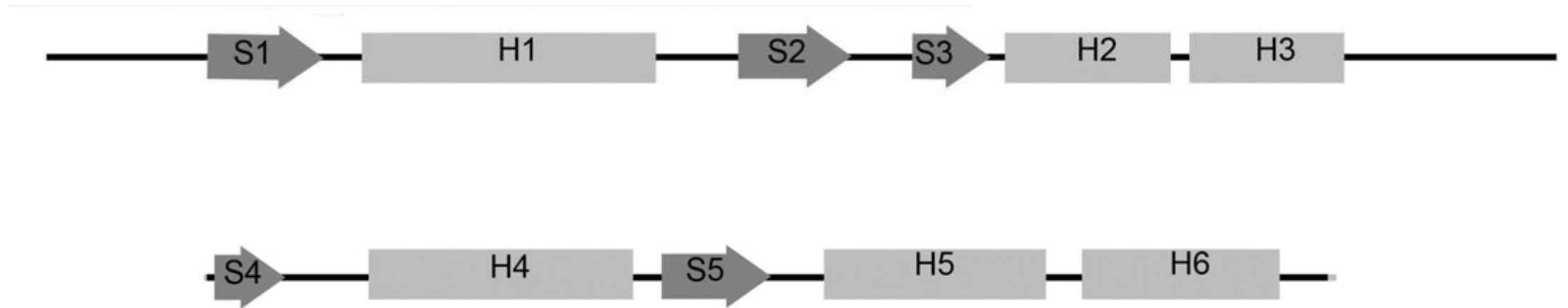
MEGA alignment :

Protein Sequences

Species*	Group Name				*									*							*	*		*		*																																	
1.	CHNMO	-	M	D	P	Q	Q	Q	F	C	L	K	W	N	S	F	S	N	L	A	I	T	F	S	N	L	F	K	S	D	L	L	A	D	V	I	L	S	C	D	G	V	V	-	F	K	A	H	K	L	I	L	A	A	C	S			
2.	FRU	Q	G	A	M	D	Q	Q	F	C	L	R	W	N	N	H	P	T	N	L	T	G	V	L	T	S	L	L	Q	R	E	A	L	C	D	V	T	L	A	C	E	G	E	T	-	V	K	A	H	Q	T	I	L	S	A	C	S		
3.	MMD4	-	M	A	D	D	E	Q	F	S	L	C	W	N	N	F	N	T	N	L	S	A	G	F	H	E	S	L	C	R	G	D	L	V	D	V	S	L	A	A	E	G	Q	I	-	V	K	A	H	R	L	V	L	S	V	C	S		
4.	BRC4	-	M	D	D	T	Q	H	F	C	L	R	W	N	N	Y	Q	S	S	I	T	S	A	F	E	N	L	R	D	D	E	A	F	V	D	V	T	L	A	C	E	G	R	S	-	I	K	A	H	R	V	V	L	S	A	C	S		
5.	ABRU	Q	Q	Q	H	Q	Q	H	Y	A	L	K	W	N	D	F	Q	S	S	I	L	S	S	F	R	H	L	R	D	E	E	D	F	V	D	V	T	L	A	C	D	E	R	S	-	F	T	A	H	K	V	V	L	S	A	C	S		
6.	GAF	S	L	P	M	N	S	L	Y	S	L	T	W	G	D	Y	G	T	S	L	V	S	A	I	Q	L	L	R	C	H	G	D	L	V	D	C	T	L	A	A	G	G	R	S	-	F	P	A	H	K	I	V	L	C	A	A	S		
7.	KEN	M	K	E	F	Q	R	M	L	M	L	Q	Y	S	K	H	G	E	C	I	L	K	E	I	G	A	A	F	R	G	E	H	P	A	D	L	T	I	V	C	E	N	K	V	K	L	H	A	H	K	L	V	L	A	A	A	S		
Species*	Group Name																											*	*	*	*																					*							
1.	CHNMO	K	K	F	A	D	L	F	E	N	T	P	T	N	G	-	Q	C	V	I	I	L	E	A	T	T	P	D	N	M	A	A	L	L	E	F	M	Y	K	G	E	V	H	V	S	Q	E	A	L	N	S	F	L	K	S	A	E	S	L
2.	FRU	P	Y	F	E	T	I	F	L	Q	N	Q	H	P	-	-	H	P	I	I	Y	L	K	D	V	R	Y	S	E	M	R	S	L	L	D	F	M	Y	K	G	E	V	N	V	G	Q	S	S	L	P	M	F	L	K	T	A	E	S	L
3.	MMD4	P	F	F	R	K	M	F	T	Q	M	P	S	N	T	-	H	A	I	V	F	L	N	N	V	S	H	S	A	L	K	D	L	I	Q	F	M	Y	C	G	E	V	N	V	K	Q	D	A	L	P	A	F	I	S	T	A	E	S	L
4.	BRC4	P	Y	F	R	E	L	L	K	S	T	P	C	K	-	-	H	P	V	I	L	L	Q	D	V	N	F	M	D	L	H	A	L	V	E	F	I	Y	H	G	E	V	N	V	H	Q	K	S	L	Q	S	F	L	K	T	A	E	V	L
5.	ABRU	P	Y	F	R	R	L	L	K	A	N	P	C	E	-	-	H	P	I	V	I	L	R	D	V	R	C	D	D	V	E	N	L	L	S	F	M	Y	N	G	E	V	N	V	S	H	E	Q	L	P	D	F	L	K	T	A	H	L	L
6.	GAF	P	F	L	L	D	L	L	K	N	T	P	C	K	-	-	H	P	V	V	M	L	A	G	V	N	A	N	D	L	E	A	L	L	E	F	V	Y	R	G	E	V	S	V	D	H	A	Q	L	P	S	L	L	Q	A	A	Q	C	L
7.	KEN	P	L	I	R	N	L	L	E	D	T	H	L	S	D	C	S	T	T	V	Y	F	P	D	V	N	A	T	Y	F	K	F	L	L	D	F	L	Y	S	G	Q	T	C	I	T	S	R	D	V	N	Y	L	H	D	L	L	L	L	L

# Secondary structure prediction of btb domain

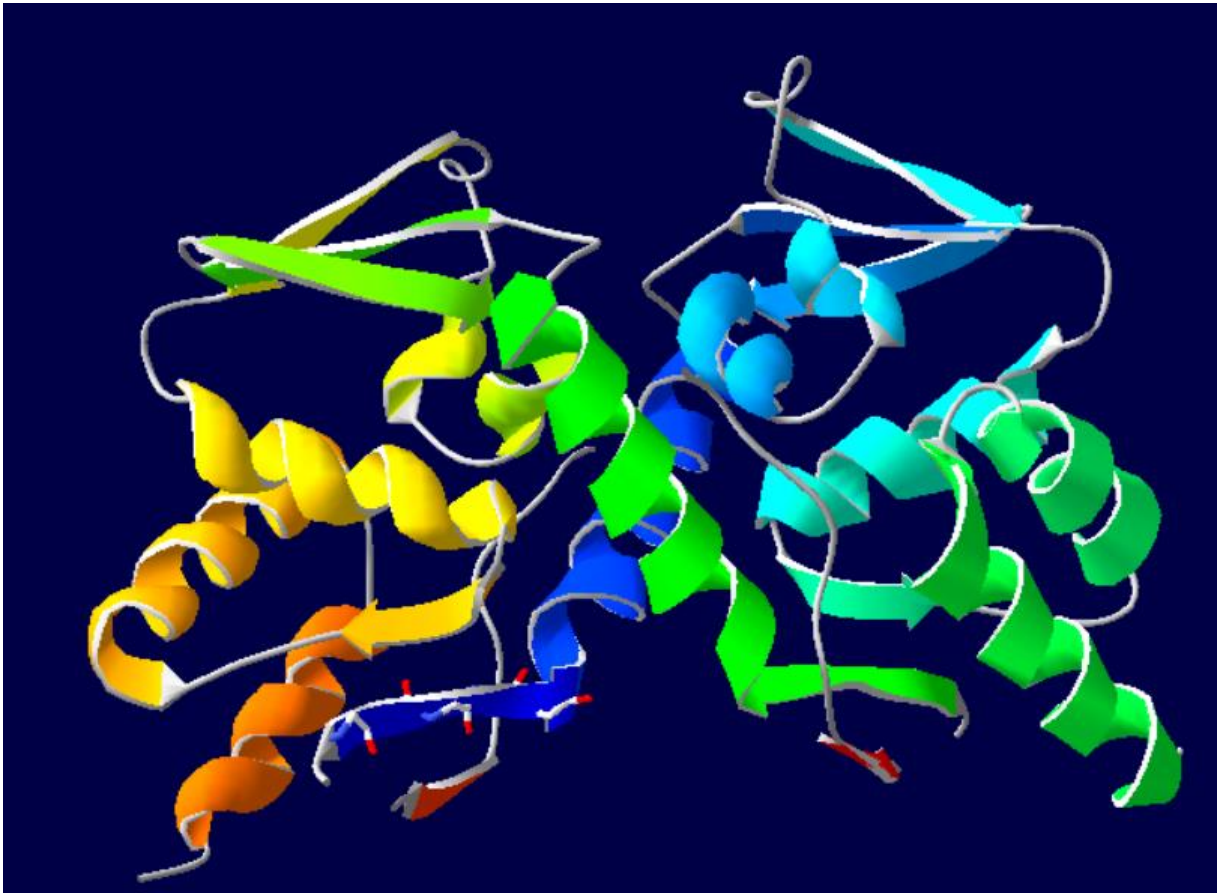
Intersection of  $\alpha$ -helix and  $\beta$ -sheet:





# The mechanism of dimerization

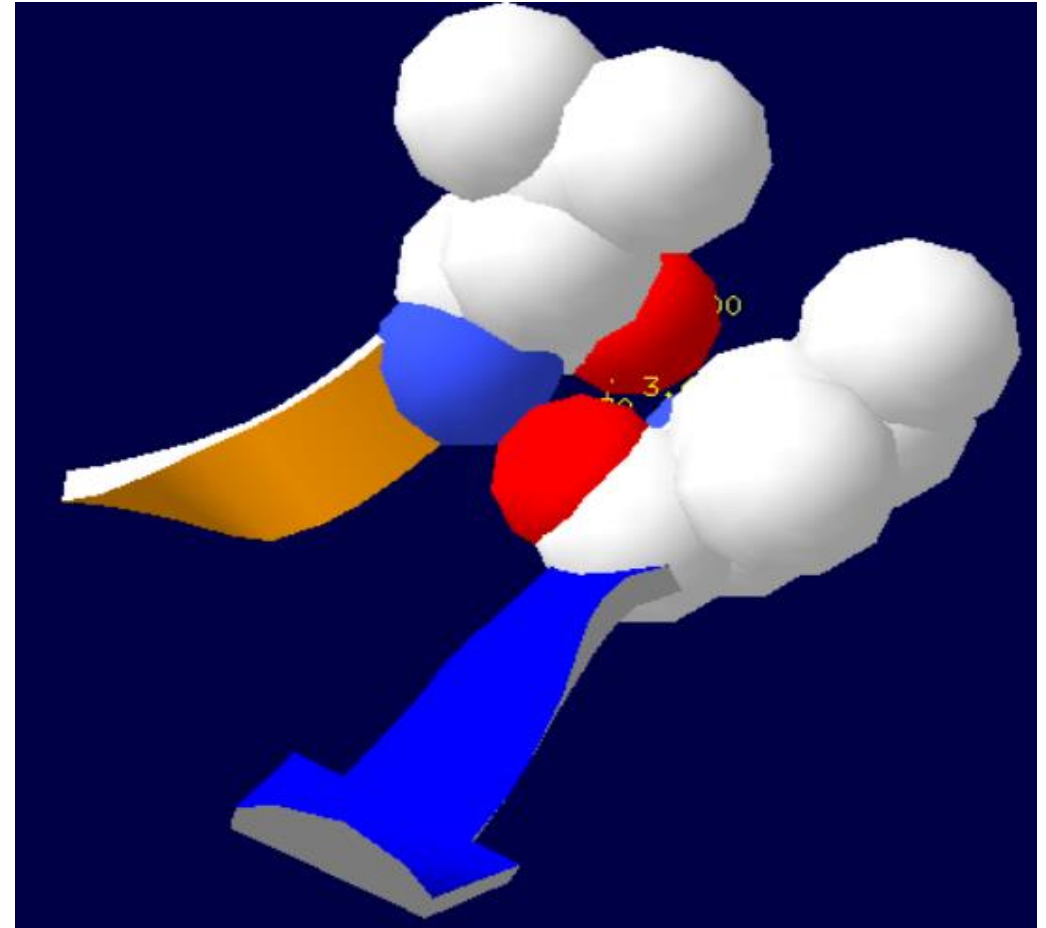
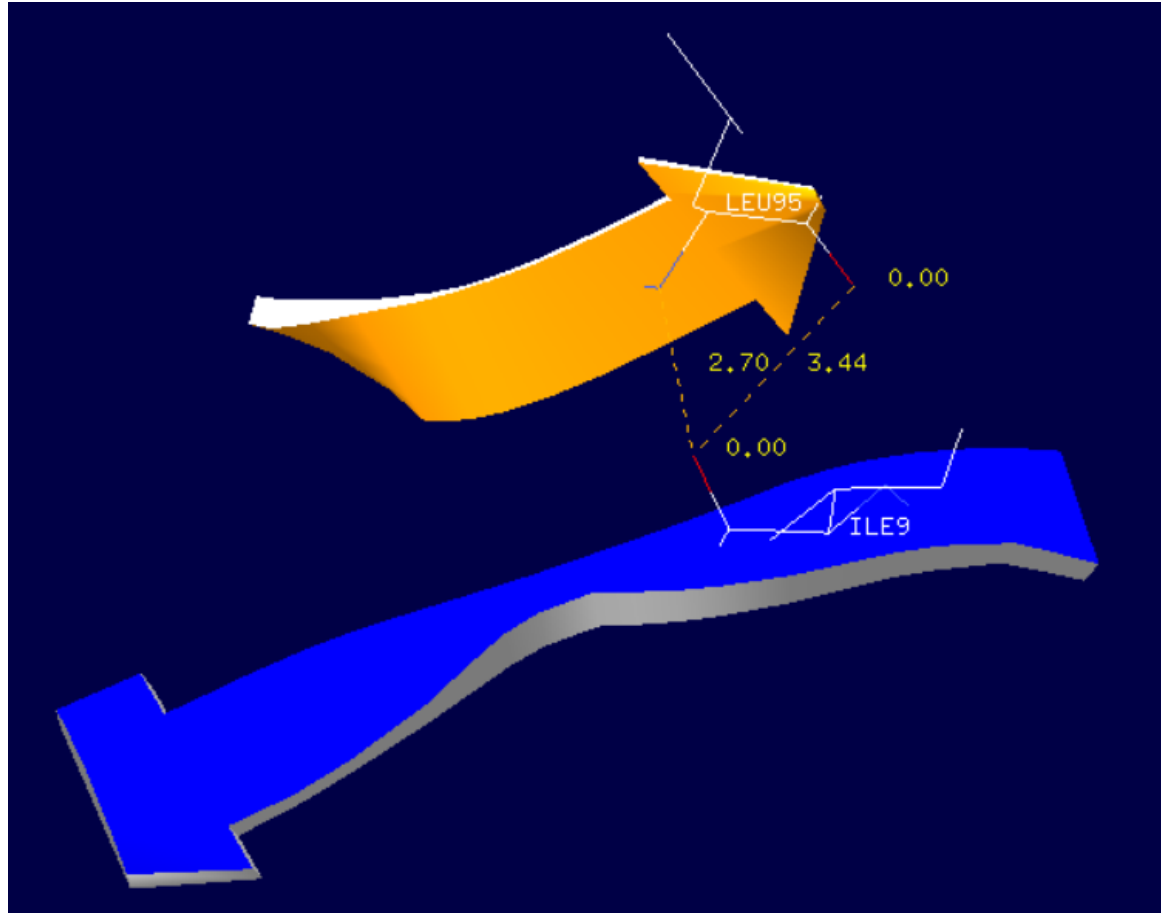
The hydrophobic residue on  $\beta$ -sheet keeps the dimerization :



$\beta 1$  from one and  $\beta 5$  from the other

# The mechanism of dimerization

Hydrophobic residue:



# tBlastn for Cp190 in vertebrates

Results: negative

In *Danio rerio*:

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Danio rerio zinc finger and BTB domain containing 12, tandem duplicate 1 (zbtb12.1), mRNA</a>	56.6	95.9	22%	5e-08	28%	<a href="#">NM_001044887.1</a>
<input type="checkbox"/>	<a href="#">Danio rerio zinc finger and BTB domain containing 12, tandem duplicate 2 (zbtb12.2), mRNA</a>	55.8	96.7	21%	1e-07	28%	<a href="#">NM_001044884.1</a>
<input type="checkbox"/>	<a href="#">Danio rerio zinc finger and BTB domain containing 18 (zbtb18), mRNA</a>	42.7	82.8	13%	0.001	34%	<a href="#">NM_001082952.1</a>

In *homo sapiens*:

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

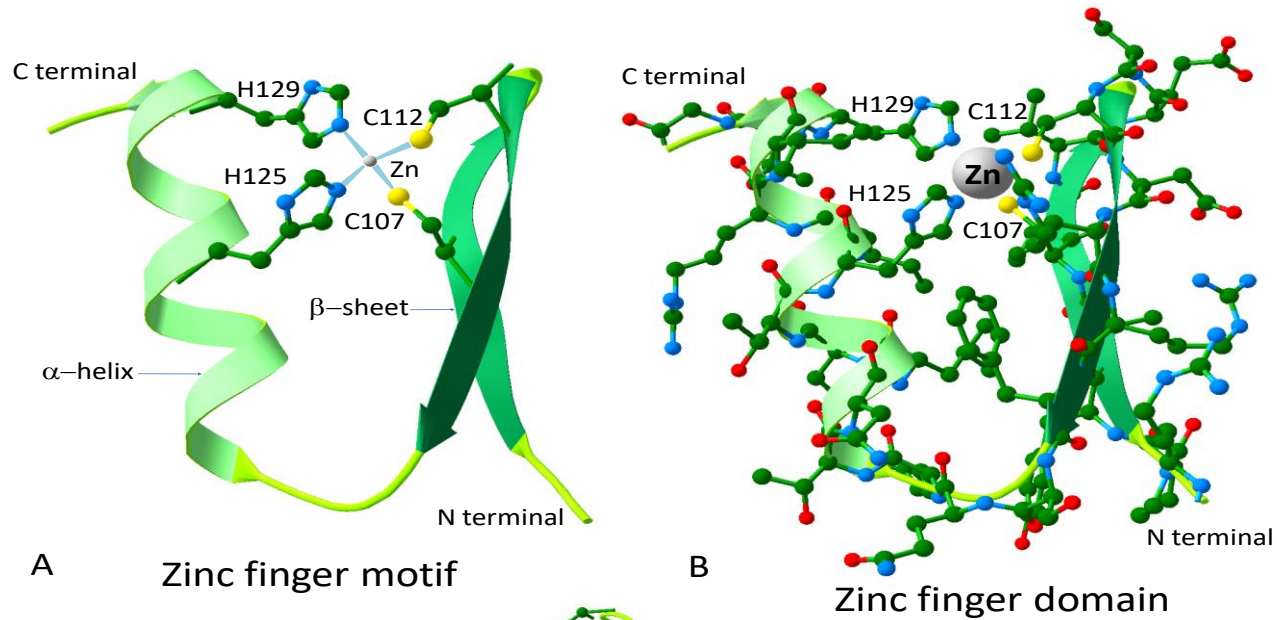
	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 264 (ZNF264), mRNA</a>	36.6	36.6	14%	0.60	24%	<a href="#">NM_003417.4</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 710 (ZNF710), mRNA</a>	38.1	38.1	13%	0.19	25%	<a href="#">NM_198526.3</a>

(<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)

# Part I summary:

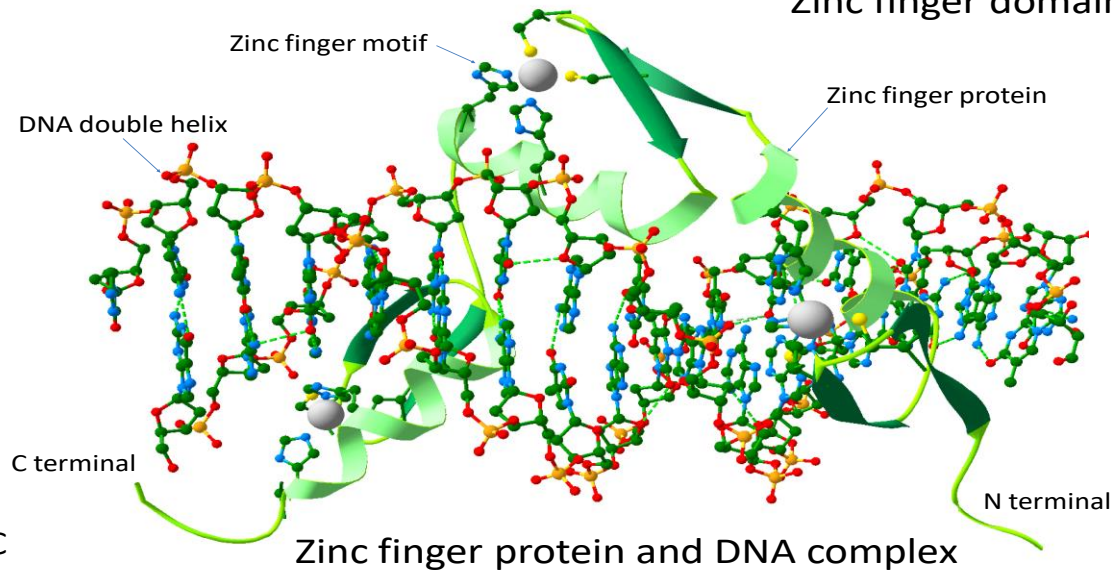
- Btb domain is not conserved in sequence in the *Drosophila*.
- Hydrophobic residues are important for dimerization.
- Btb-zinc finger proteins are not conserved in sequence between vertebrates and invertebrates.

# Structure of zinc-finger domain



A Zinc finger motif

B Zinc finger domain



C Zinc finger protein and DNA complex

(From Teacher Luo)

# Zinc-finger domain of CP190 alignment

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 236 (ZNF236), transcript variant 1, mRNA</a>	35.8	92.8	97%	0.001	33%	<a href="#">NM_001306089.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 236 (ZNF236), transcript variant 2, mRNA</a>	35.8	92.4	97%	0.001	33%	<a href="#">NM_007345.3</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger and BTB domain containing 41 (ZBTB41), transcript variant 1, mRNA</a>	35.4	35.4	97%	0.002	36%	<a href="#">NM_194314.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger and BTB domain containing 41 (ZBTB41), transcript variant 2, non-coding RNA</a>	35.0	35.0	97%	0.003	36%	<a href="#">NR_135153.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 423 (ZNF423), transcript variant 3, mRNA</a>	33.9	64.3	97%	0.006	27%	<a href="#">NM_001330533.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 423 (ZNF423), transcript variant 1, mRNA</a>	33.9	64.3	97%	0.006	27%	<a href="#">NM_015069.4</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 423 (ZNF423), transcript variant 2, mRNA</a>	33.9	64.3	97%	0.006	27%	<a href="#">NM_001271620.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger and BTB domain containing 26 (ZBTB26), transcript variant 2, mRNA</a>	32.7	59.7	97%	0.019	29%	<a href="#">NM_001304363.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger and BTB domain containing 26 (ZBTB26), transcript variant 1, mRNA</a>	32.7	59.7	97%	0.019	29%	<a href="#">NM_020924.3</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger and BTB domain containing 26 (ZBTB26), transcript variant 3, mRNA</a>	32.7	59.7	97%	0.019	29%	<a href="#">NM_001304364.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens ovo like zinc finger 2 (OVOL2), transcript variant 1, mRNA</a>	31.6	89.3	97%	0.050	31%	<a href="#">NM_021220.3</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 521 (ZNF521), transcript variant 2, mRNA</a>	31.2	56.2	97%	0.068	31%	<a href="#">NM_001308225.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens zinc finger protein 521 (ZNF521), transcript variant 1, mRNA</a>	31.2	56.2	97%	0.068	31%	<a href="#">NM_015461.2</a>

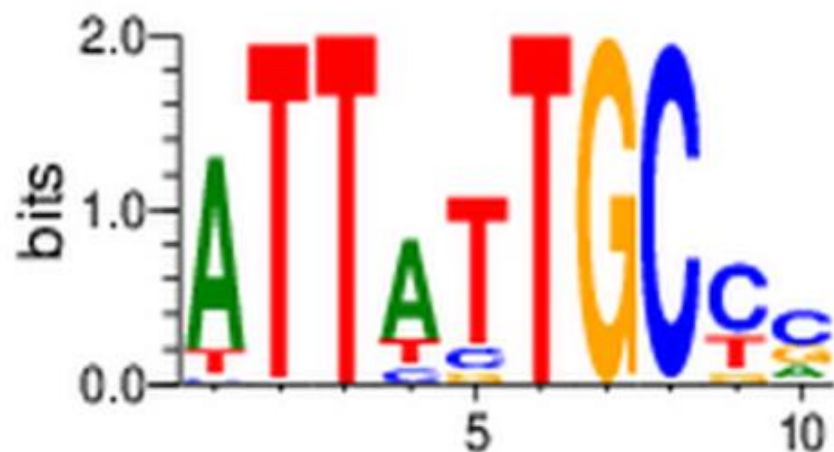
**Shortcomings: Similar motifs doesn't mean similar genes, for genes are not all conserved among the two species.**

(<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)

# Zinc-finger domain motif prediction

## DNA Sequence Logo Generator

A DNA binding site predictor for Cys<sub>2</sub>His<sub>2</sub> Zinc Finger Proteins



A candidate protein in mammalian which has btb domain and zinc-finger domain. Then we take this as an example.

Anton Persikov and Mona Singh (2014) *"De Novo Prediction of DNA-binding Specificities for Cys<sub>2</sub>His<sub>2</sub> Zinc Finger Proteins"*. NAR, 42(1): 97-108. Epub 2013 Oct 3.

# Motif mapped to the Genome

10,000 results back(mapped to the hg19)



# FIMO

Find Individual Motif Occurrences

name	start	stop	strand	score	p-value	q-value	sequence
chr1	121484880	121484889	-	4.95556	1.14E-10	0.663	TTTTCTGCCA
chr5	134260230	134260239	+	3.8	2.88E-10	0.685	ATTGGTGCGG
chr5	134262786	134262795	+	3.37778	4.48E-10	0.685	ATGACTGCGC
chr11	10531048	10531057	+	3.33333	4.73E-10	0.685	TTTCTTGCCA
chr13	108594574	108594583	+	2.73333	9.20E-10	0.747	attcctgctg
chr19	41769928	41769937	-	2.62222	9.97E-10	0.747	TTTCCTGCGC
chr22	21271467	21271476	+	2.55556	1.07E-09	0.747	TTTCCTGCCA
chr6	58777233	58777242	-	2.51111	1.15E-09	0.747	TTTTCTGCCG
chr6	37151230	37151239	+	2.26667	1.73E-09	0.747	TTTCCTGCCC
chr1	121485197	121485206	+	2.17778	1.97E-09	0.747	atttcagccg
chr6	43027215	43027224	-	2.13333	2.11E-09	0.747	TTTTCTGCCC
chr16	72127617	72127626	+	2.11111	2.23E-09	0.747	TTTCCTGCCC
chr16	33963105	33963114	+	1.95556	2.72E-09	0.747	ttagtgcca
chr8	70602577	70602586	-	1.95556	2.72E-09	0.747	ATTTCTGCCC



# Motif mapped to the Genome

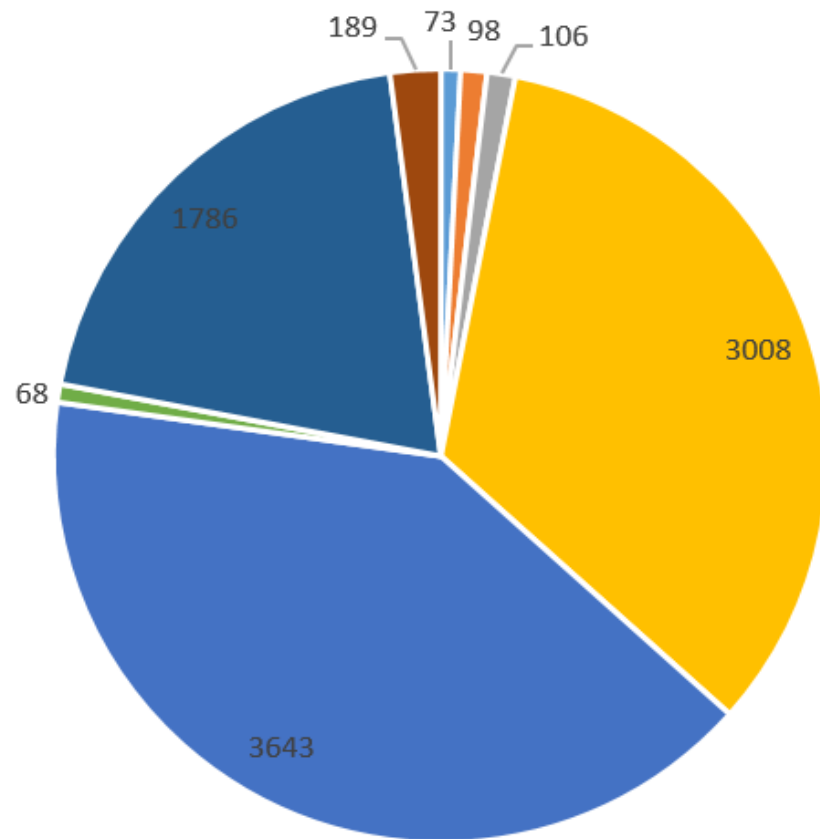
UCSC Galaxy: intersect interval. (Mapped to hg19.)

Gene Name	Annotation	Gene Type	Detailed Annotation	Distance to TSS
CPPED1	intron (NM_018340, intron)	protein-coding	intron (NM_018340, intron)	2931
LOC101928068	intron (NM_001290327, intron)	ncRNA	intron (NM_001290327, intron)	-38125
SUSD1	intron (NM_022486, intron)	protein-coding	MER89 LTR ERV1	4169
MCM3	promoter-TSS (NM_001290327, promoter)	protein-coding	promoter-TSS (NM_001290327, promoter)	-31
GABRR2	Intergenic	protein-coding	Intergenic	-6030
EDN2	Intergenic	protein-coding	Intergenic	51881
WDR66	intron (NM_001178003, intron)	protein-coding	L1M5 LINE L1	9104
HIST1H4H	Intergenic	protein-coding	Intergenic	-35819
ZNF277	promoter-TSS (NM_014706, promoter)	protein-coding	promoter-TSS (NM_014706, promoter)	120
TAS2R40	Intergenic	protein-coding	Intergenic	-6760
PSMD8	intron (NM_021185, intron)	protein-coding	CpG-13646	-11848
RCAN3AS	promoter-TSS (NM_013446, promoter)	ncRNA	promoter-TSS (NM_013446, promoter)	-209
STXBP5	promoter-TSS (NM_001121, promoter)	protein-coding	promoter-TSS (NM_001121, promoter)	-864
C2CD4A	TTS (NM_207322)	protein-coding	TTS (NM_207322)	4910
CNOT3	promoter-TSS (NM_014551, promoter)	protein-coding	promoter-TSS (NM_014551, promoter)	-585
CSNK1D	intron (NR_110578, intron)	protein-coding	intron (NR_110578, intron)	6404
PSEN2	Intergenic	protein-coding	Intergenic	-65340
MIR1208	Intergenic	ncRNA	Intergenic	203709
ACPT	Intergenic	protein-coding	Intergenic	-6717
UMAD1	promoter-TSS (NM_001306, promoter)	protein-coding	promoter-TSS (NM_001306, promoter)	-127
EXD3	intron (NR_104598, intron)	protein-coding	THE1A LTR ERV1-MaLF	20265
TMEM116	intron (NM_001294314, intron)	protein-coding	L1MC2 LINE L1	37552
LOC100506023	intron (NR_037845, intron)	ncRNA	LTR5B LTR ERVK	1993
PIP5K1B	Intergenic	protein-coding	L3 LINE CR1	-78617

(<https://usegalaxy.org/>)

# Binding site analysis

Binding site analysis

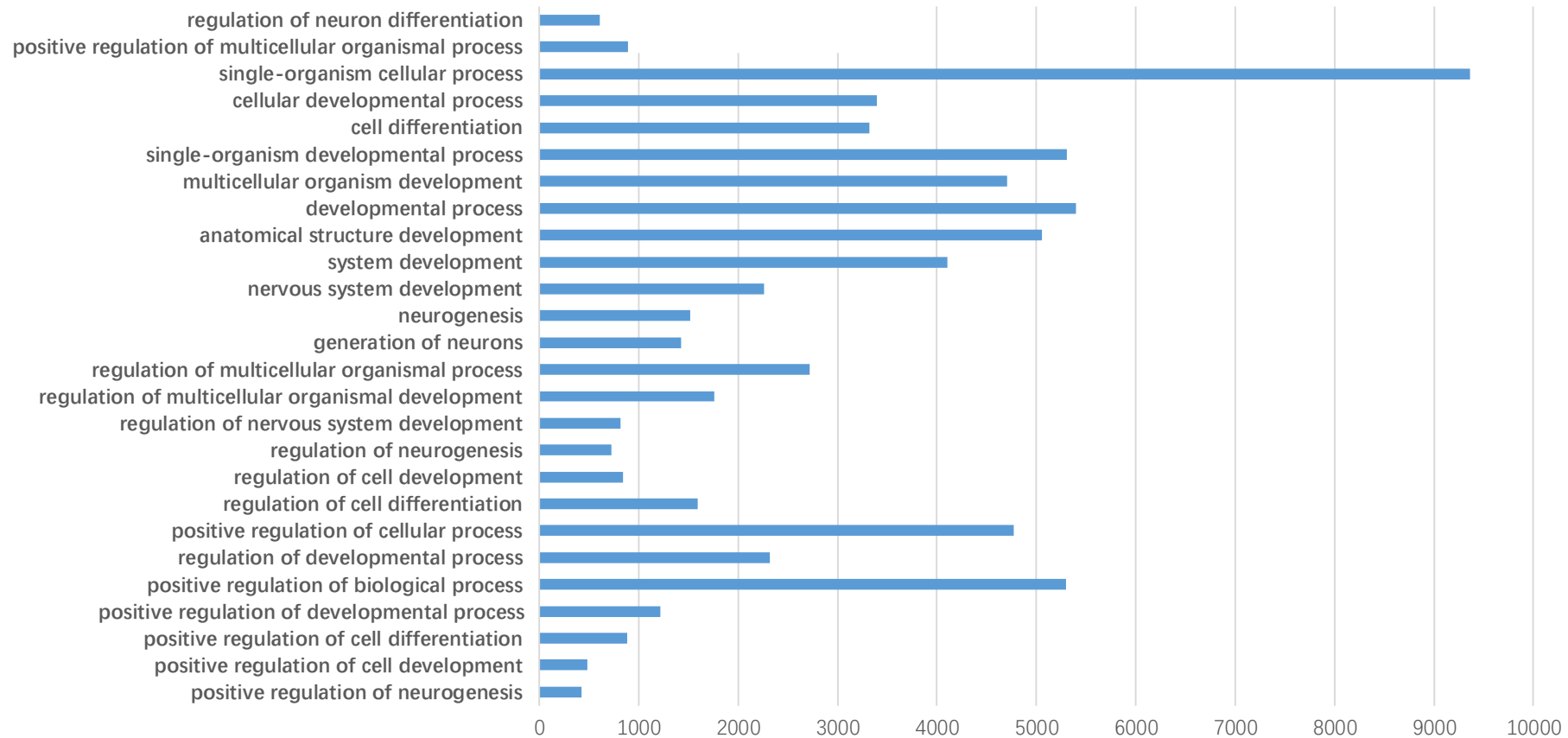


■ 3'UTR ■ 5'UTR ■ exon ■ intergenic ■ intron ■ non-coding ■ promoter ■ TSS

Promoter: directly mediate transcription

Intergenic&intron: keep chromosome structure

# Gene ontology analysis



(<http://www.geneontology.org/page/go-enrichment-analysis>)

# Conclusion:

- **Btb domain is not sequenced-conserved but conserves in structure and function.**
- **Zinc finger motif prediction combined with chip-seq reveals some clues for conservation.**

# Future plan:

- Compare our prediction with zbtb family chip-seq data.
- Point mutation of the hydrophobic residue.

# Acknowledgement:

- Jingchu Luo(Teacher)
- Xiong Ji(Rotation mentor)
- Lan Ke(teaching assistant)
- Group 15 members
- All of you

# Happy New Year!

先定一个  
能达到的小目标



比如发个science

