

Function analysis of FOXP2 based on the structure

基于结构的语言相关蛋白FOXP2功能分析



PKU16F G10

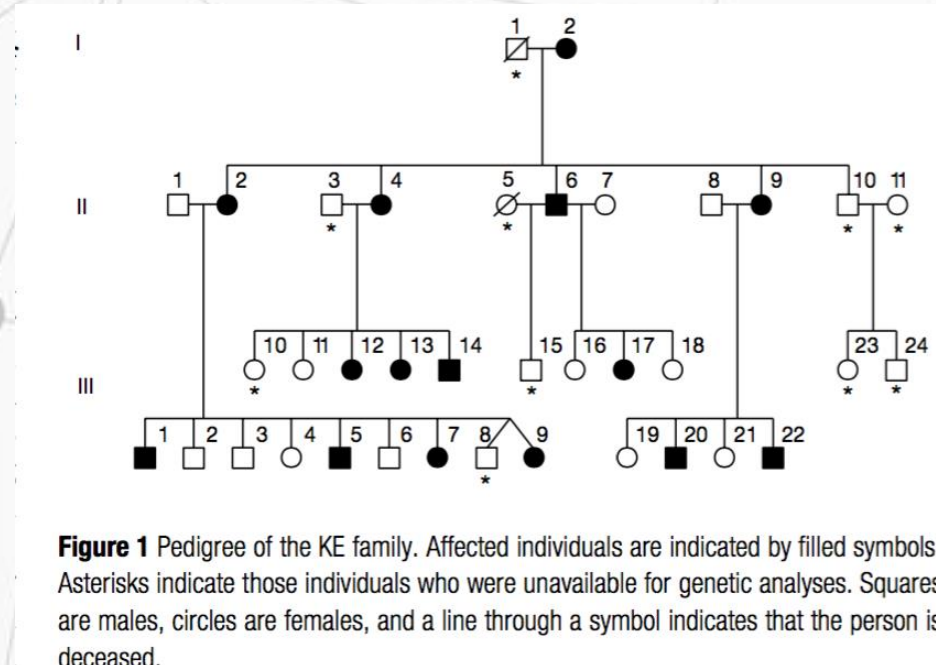
王博 张鑫 刘璐 申辉

2017/1/7

Outline

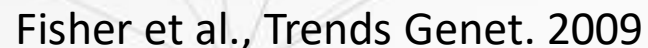
- Description of FOXP2
- Nucleotide information
- FOXP2 phylogenetic tree
- Protein information

FOXP2: Forkhead box protein P2



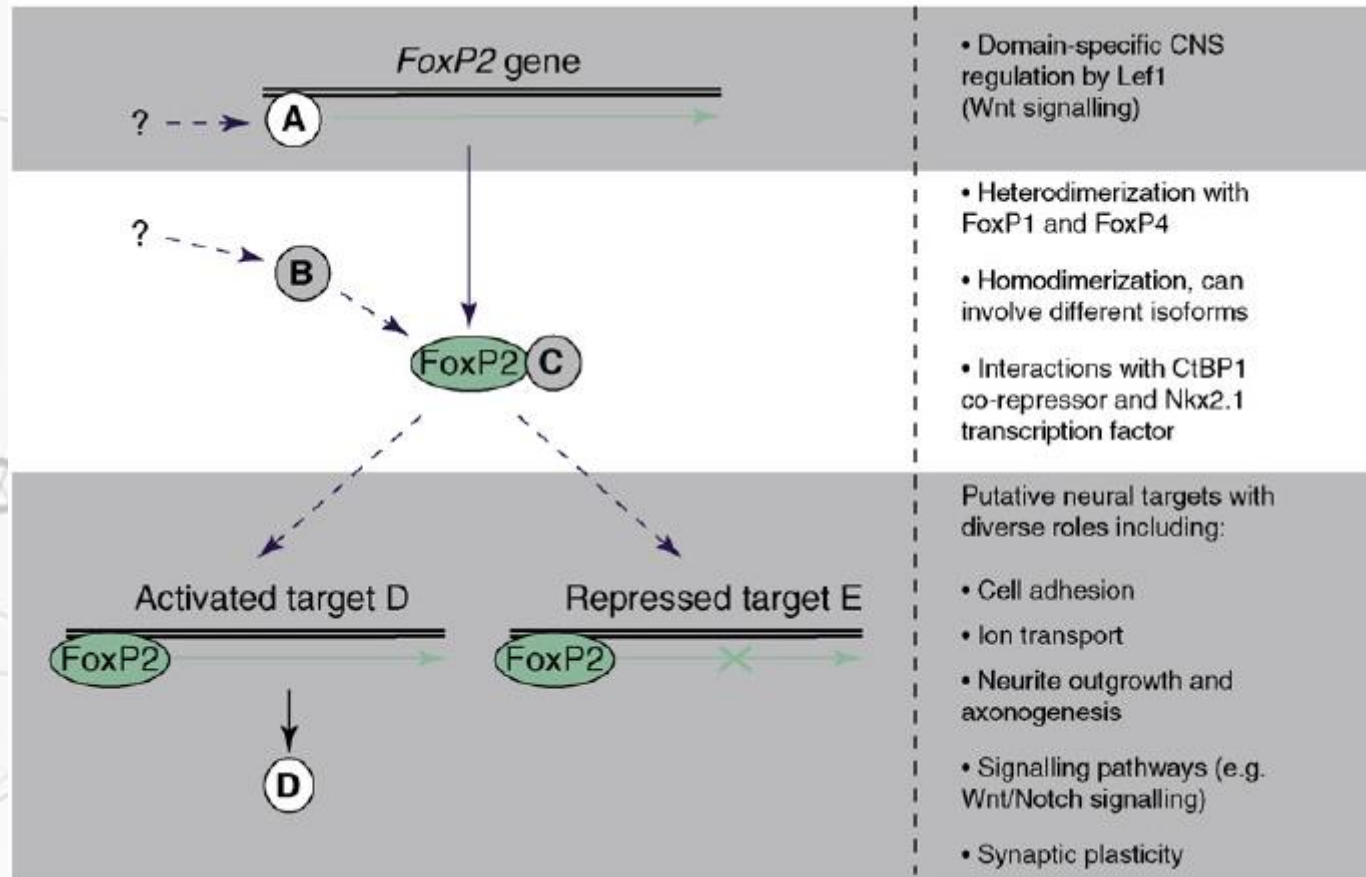
- Forkhead box protein P2 (FOXP2) is a protein that, in humans, is encoded by the FOXP2 gene, also known as CAGH44, SPCH1 or TNRC10, and is required for proper development of speech and language.
- Initially identified as the genetic factor of speech disorder in KE family, its gene is the first gene discovered associated with speech and language.
- R553H mutation in DNA binding domain

- misconception that a unique trait requires genes that are exclusive to the species and organs that exhibit it.
- The *FOXP2* gene is highly conserved in mammals.

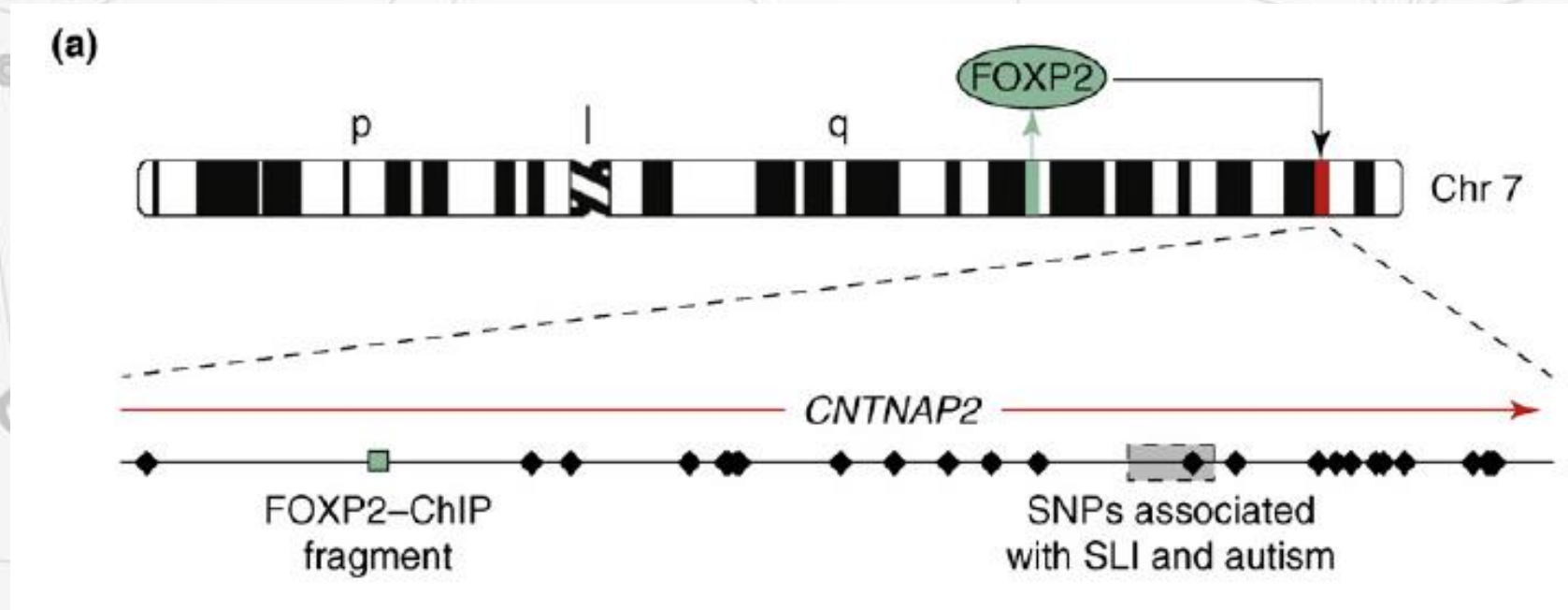


Entry point into functional genetic pathways

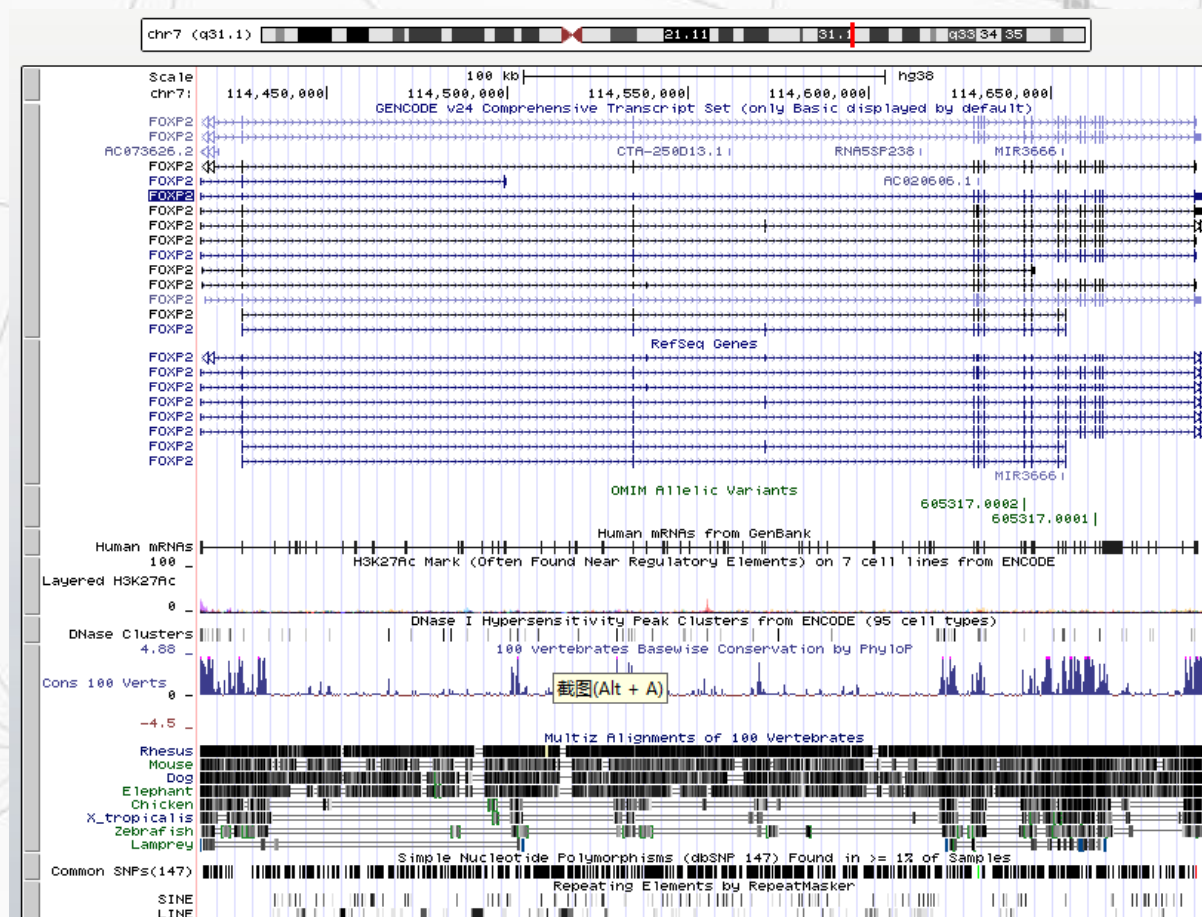
Expressed in fetal and adult brain, heart, lung and gut.



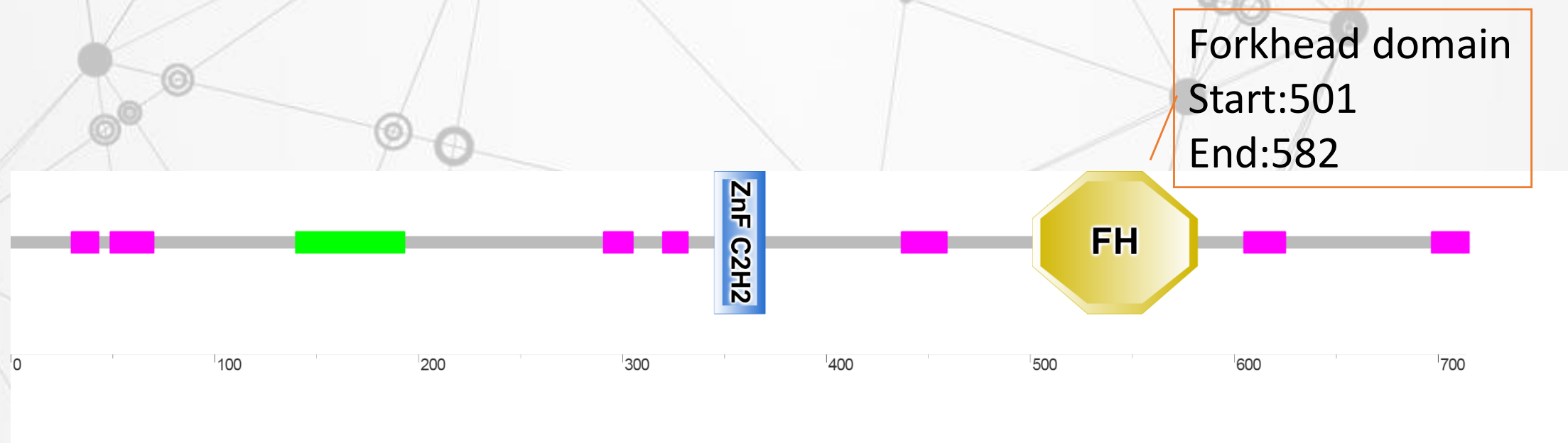
Entry point into functional genetic pathways



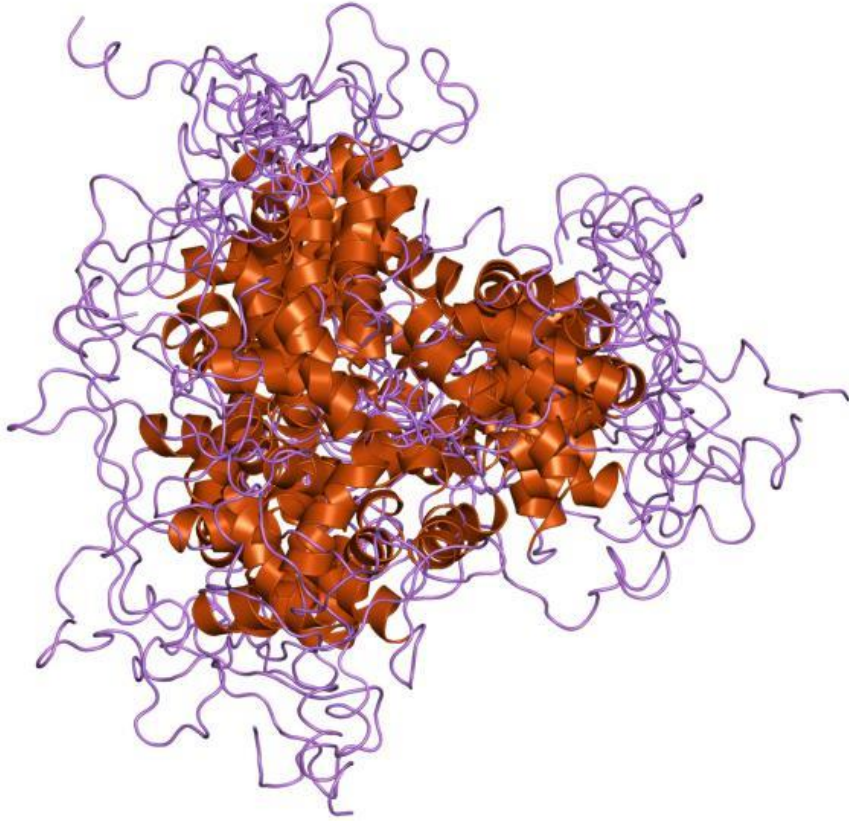
UCSC Genome Browser



SMART analysis



Forkhead domain



A type of protein domain that is often found in transcription factors and whose purpose is to bind DNA.

Involvement in early developmental decisions of cell fates during embryogenesis

SMART analysis

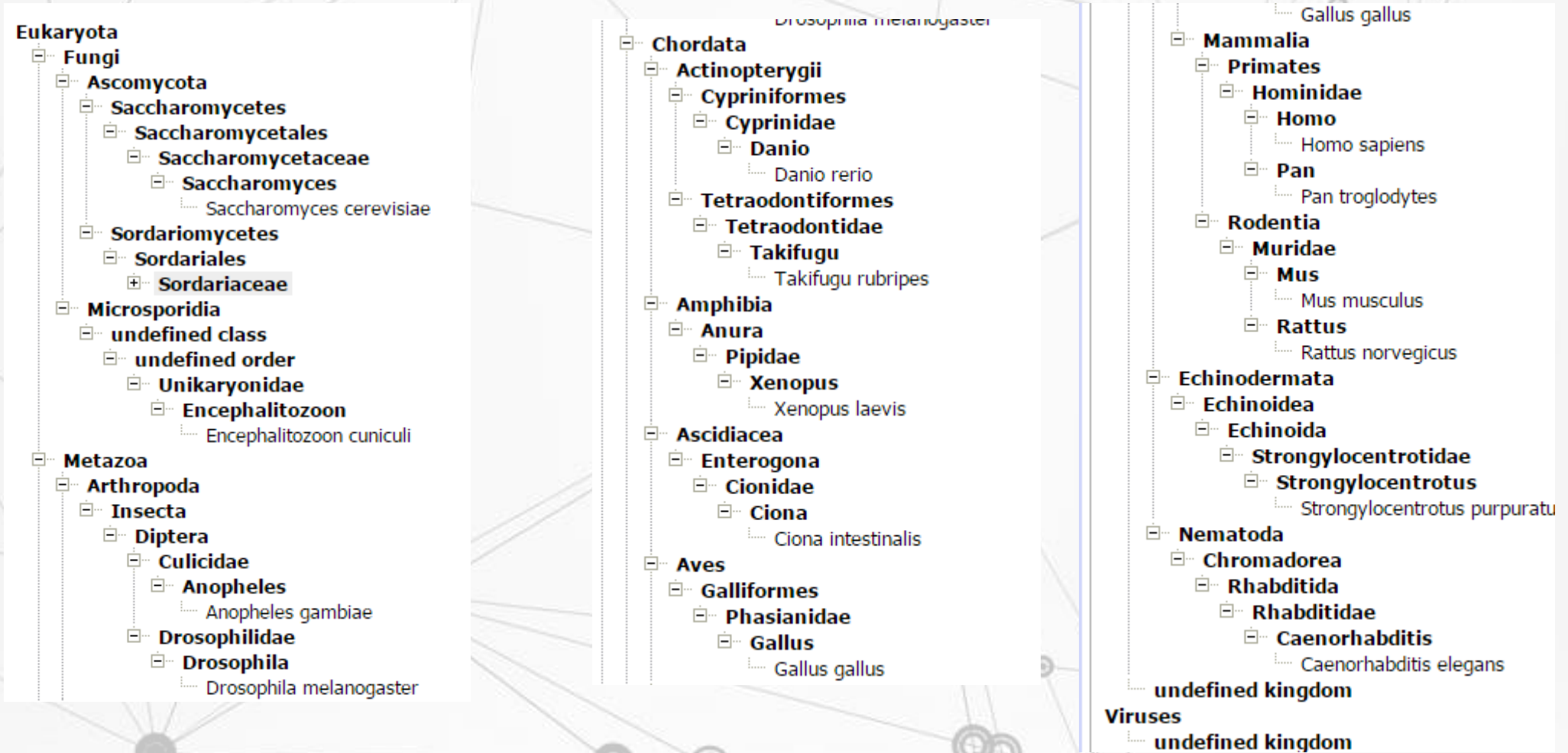
Family alignment: View **Alignment consensus sequence** **Family alignment in** FASTA format

There are **7056** FH domains in 7043 proteins in **SMART's nrdb** database.

Click on the following links for more information.

- ▶ **Evolution** (species in which this domain is found)
- ▶ **Cellular role** (predicted cellular role)
- ▶ **Literature** (relevant references for this domain)
- ▶ **Disease** (disease genes where sequence variants are found in this domain)
- ▶ **Metabolism** (metabolic pathways involving proteins which contain this domain)
- ▶ **Structure** (3D structures containing this domain)
- ▶ **Links** (links to other resources describing this domain)

SMART analysis



Blast



U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

[Sign in to NCBI](#)

BLAST® >> **blastp suite**

[Home](#)

[Recent Results](#)

[Saved Strategies](#)

[Help](#)

Standard Protein BLAST

[blastn](#) **[blastp](#)** [blastx](#) [tblastn](#) [tblastx](#)

Enter Query Sequence

BLASTP programs search protein databases using a protein query. [more...](#)

[Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) ?

[Clear](#)

Query subrange ?

AAI43868

From

To

Or, upload file

选择文件 未选择任何文件 ?

Job Title

AAI43868:FOXP2 protein [Homo sapiens]

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Database

◆ Reference proteins (refseq_protein) ?

Organism

Optional

Homo sapiens (taxid:9606)

☒ Exclude



☐ Exclude

true

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. ?

Exclude 1/7

☒ Models (XM/XP) ☐ Uncultured/environmental sample sequences Peking University

Blast

Entrez Query
Optional

[YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search [?](#)

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST

Search [database Reference proteins \(refseq_protein\)](#) using **Blastp (protein-protein BLAST)**

☒ Show results in a new window

[Algorithm parameters](#)

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

[Restore default search parameters](#)

General Parameters

Max target sequences

100 ▾

Select the maximum number of aligned sequences to display [?](#)

Short queries

☒ Automatically adjust parameters for short input sequences [?](#)

Expect threshold

10 ♦ [?](#)

Word size

6 ▾ [?](#)

Max matches in a query range

0 [?](#)

Scoring Parameters

Blast

Scoring Parameters

Matrix

BLOSUM80

Gap Costs

Existence: 10 Extension: 1

Compositional adjustments

Conditional compositional score matrix adjustment

Filters and Masking

Filter

☐ Low complexity regions

Mask

☐ Mask for lookup table only

☐ Mask lower case letters

BLAST

Search **database Reference proteins (refseq_protein)** using **Blastp (protein-protein BLAST)**

☒ Show results in a new window

BLAST is a registered trademark of the National Library of Medicine

[Support center](#) [Mailing list](#) [YouTube](#)


NCBI

National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA

[Policies and Guidelines](#) | [Contact](#)



Blast

 U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

Sign in to NCBI

BLAST® » blastp suite » RID-5R10PNWC015

Home Recent Results Saved Strategies Help

BLAST Results

i Your search is limited to records matching entrez query: all [filter] NOT(txid9606 [ORGN] OR (XP_000001:XP_999999[pacc] OR XP_000000001:XP_99999999[pacc])).

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

AAI43868:FOXP2 protein [Homo sapiens]

RID [5R10PNWC015](#) (Expires on 12-23 14:43 pm)

Query ID [AAI43868.1](#)

Description FOXP2 protein [Homo sapiens]

Molecule type amino acid

Query Length 714

Database Name refseq_protein

Description NCBI Protein Reference Sequences

Program BLASTP 2.6.0+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)**New** Analyze your query with [SmartBLAST](#)**Graphic Summary**[Show Conserved Domains](#)

Putative conserved domains have been detected, click on the image below for detailed results.

Protein Classification: FOXP-CC and FH domain-containing protein
FOXP-CC and FH domain-containing protein



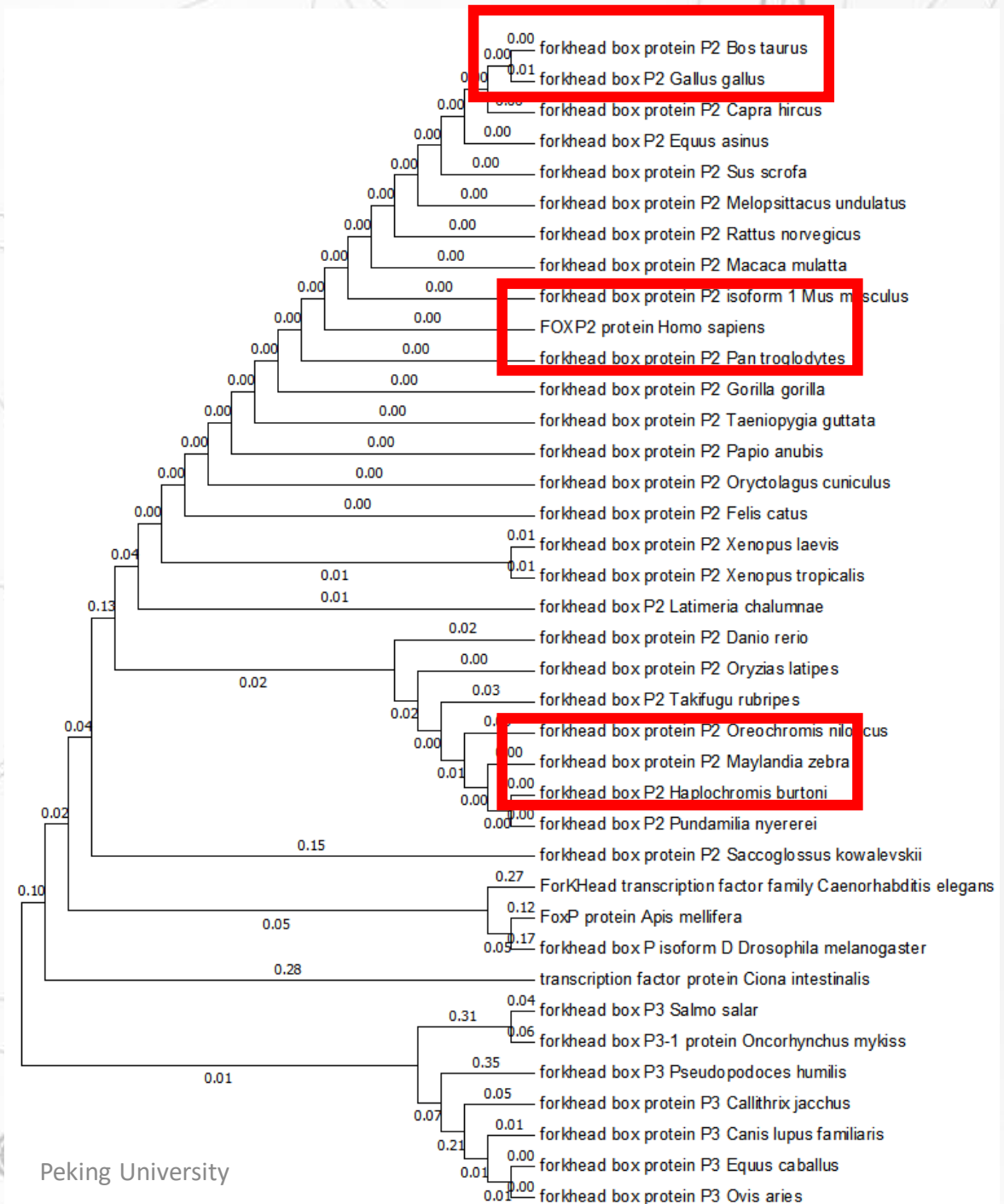
Query seq.		
Specific hits	FOXP-CC	FH
Superfamilies	FOXP-CC superf	FH superfamily

Blast



<input checked="" type="checkbox"/> forkhead box P2 [Gallus gallus]	1419	1419	100%	0.0	95%	NP_001305342.1
<input checked="" type="checkbox"/> forkhead box P2 [Latimeria chalumnae]	1036	1036	97%	0.0	89%	NP_001303811.1
<input checked="" type="checkbox"/> forkhead box protein P2 [Xenopus laevis]	1029	1029	100%	0.0	92%	NP_001089138.1
<input checked="" type="checkbox"/> forkhead box protein P2 [Xenopus tropicalis]	1021	1021	100%	0.0	91%	NP_001093722.1
<input checked="" type="checkbox"/> forkhead box protein P2 [Danio rerio]	904	904	100%	0.0	75%	NP_001025253.1
<input checked="" type="checkbox"/> forkhead box protein P2 [Oreochromis niloticus]	798	798	100%	0.0	67%	NP_001298264.1
<input checked="" type="checkbox"/> forkhead box protein P2 [Maylandia zebra]	795	795	100%	0.0	66%	NP_001297002.1
<input checked="" type="checkbox"/> forkhead box P2 [Haplochromis burtoni]	794	794	100%	0.0	66%	NP_001302467.1
<input checked="" type="checkbox"/> forkhead box P2 [Pundamilia nyererei]	792	792	100%	0.0	66%	NP_001302581.1
<input checked="" type="checkbox"/> forkhead box P2 [Takifugu rubripes]	737	737	66%	0.0	69%	NP_001292534.1
<input type="checkbox"/> forkhead box protein P1 [Bos taurus]	726	726	100%	0.0	62%	NP_001077158.1
<input type="checkbox"/> forkhead box protein P1 [Macaca mulatta]	709	709	100%	0.0	61%	NP_001253250.1
<input type="checkbox"/> forkhead box protein P4 [Taeniopygia guttata]	702	784	81%	0.0	67%	NP_001266189.1
<input type="checkbox"/> forkhead box protein P1 isoform 2 [Mus musculus]	693	693	100%	0.0	61%	NP_001184250.1
<input type="checkbox"/> forkhead box protein P1 isoform 4 [Mus musculus]	688	688	100%	0.0	59%	NP_001334274.1
<input type="checkbox"/> forkhead box protein P1 isoform 1 [Mus musculus]	681	681	100%	0.0	58%	NP_444432.1
<input checked="" type="checkbox"/> forkhead box protein P2 [Oryzias latipes]	676	676	66%	0.0	66%	NP_001121985.1
<input type="checkbox"/> forkhead box protein P4 [Danio rerio]	650	749	86%	0.0	61%	NP_001186420.1
<input type="checkbox"/> forkhead box P1 S homeolog [Xenopus laevis]	644	644	66%	0.0	64%	NP_001089002.1
<input type="checkbox"/> forkhead box protein P1 [Taeniopygia guttata]	642	642	66%	0.0	65%	NP_001070166.1
<input type="checkbox"/> forkhead box protein P1 [Rattus norvegicus]	641	641	66%	0.0	65%	NP_001029303.1
<input type="checkbox"/> forkhead box protein P1 [Gallus gallus]	639	730	86%	0.0	65%	NP_001019998.1

Mega tree

- ClustalW alignment
- Minimum evolution tree



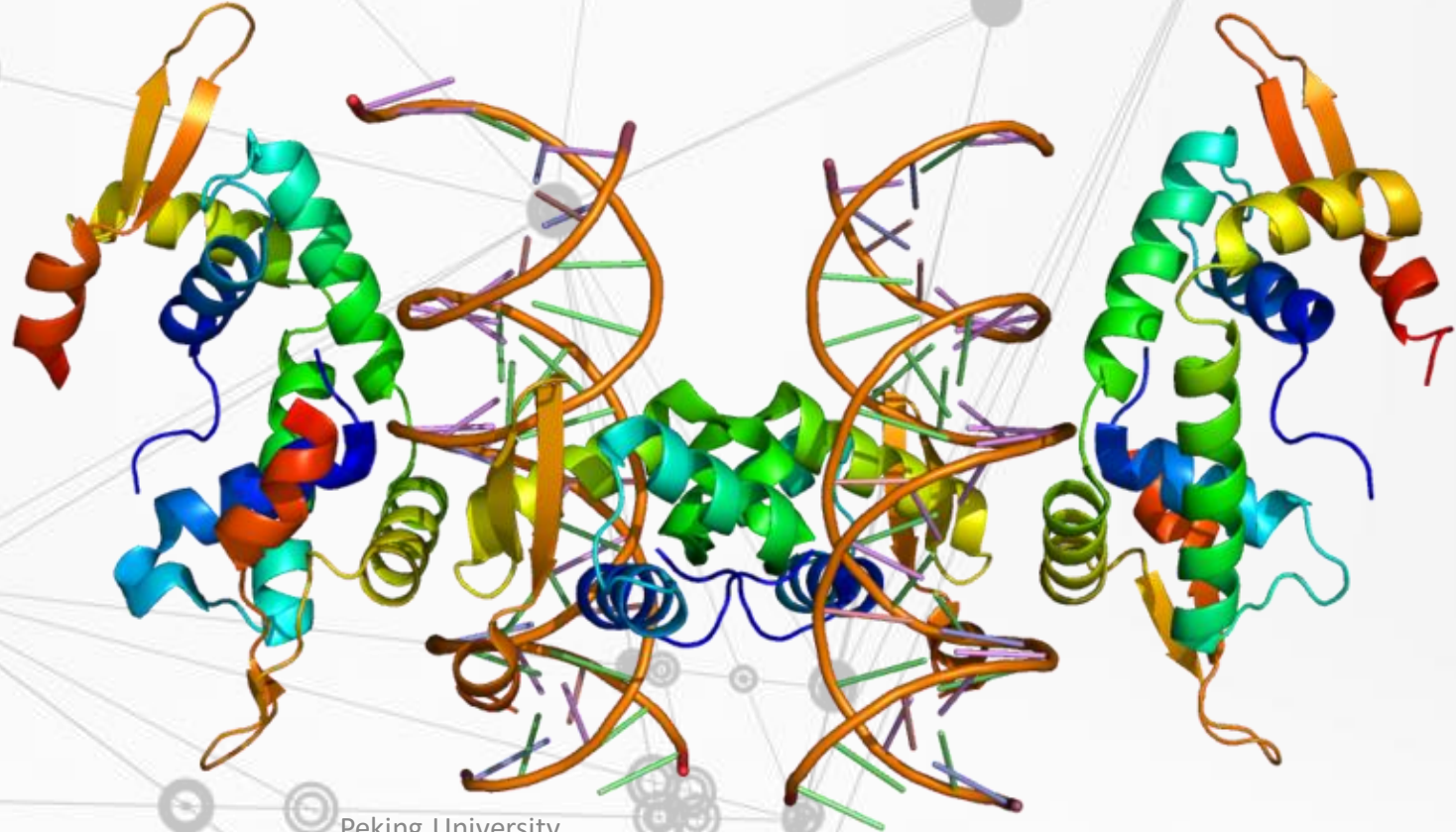
Structure of FOXP2

Regions					
Feature key	Position(s)	Description	Actions	Graphical view	Length
Zinc finger ¹	346 – 371	C2H2-type	Add BLAST		26
DNA binding ¹	504 – 594	Fork-head PROSITE-ProRule annotation	Add BLAST		91

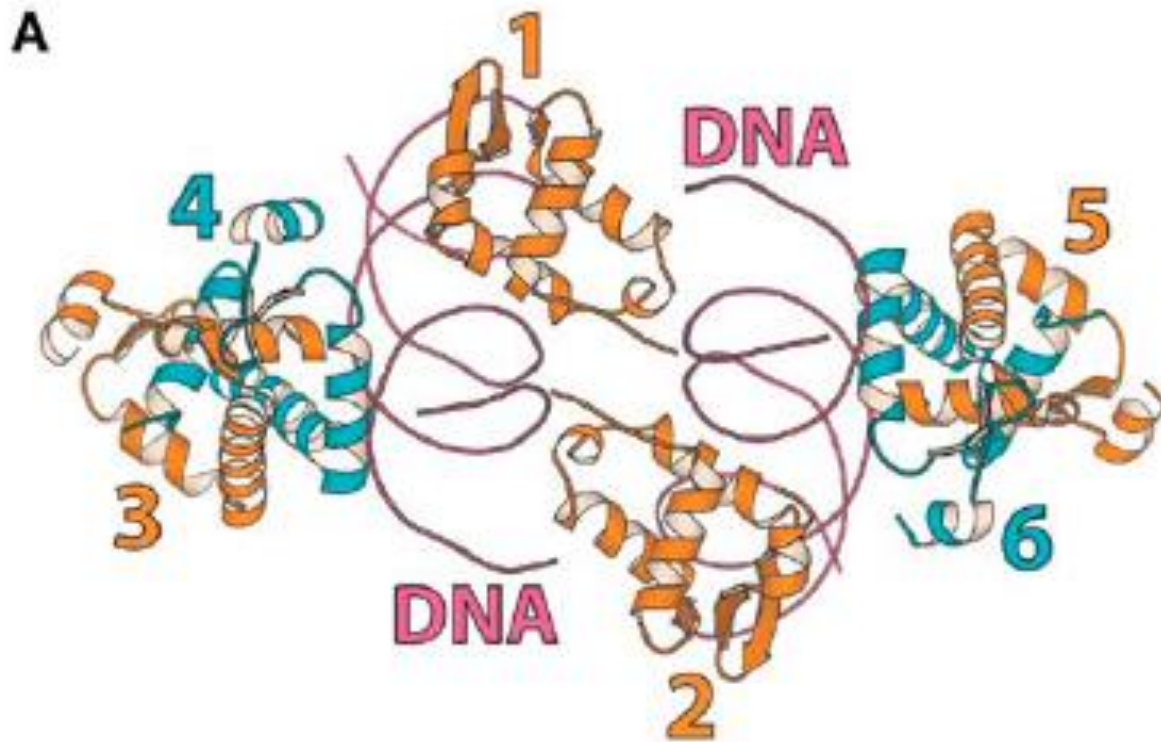
6 identical FOXP2
forkhead domains

2 monomer

2 dimer

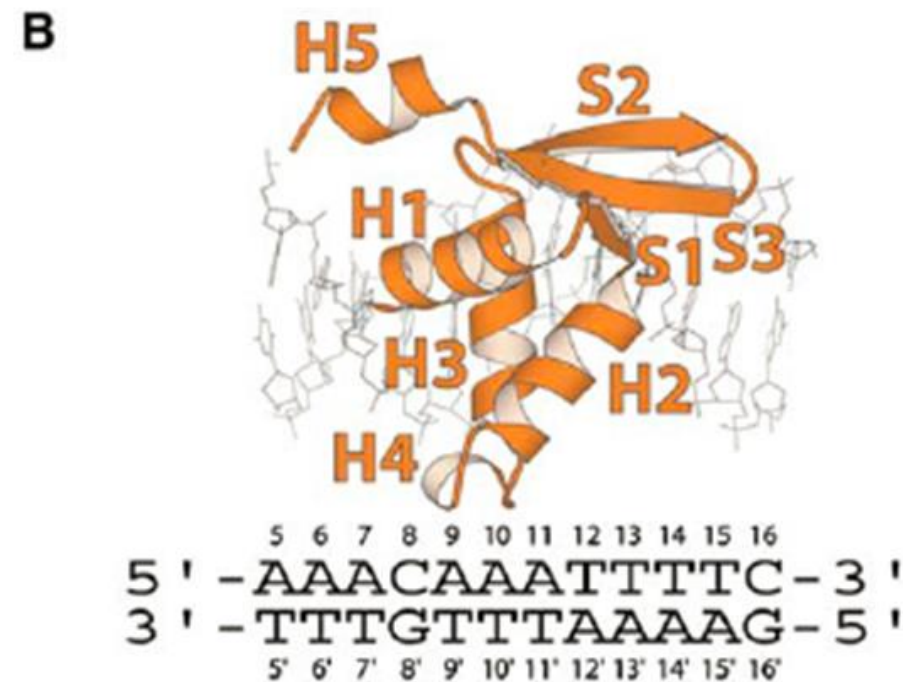


Structure of FOXP2/DNA complex



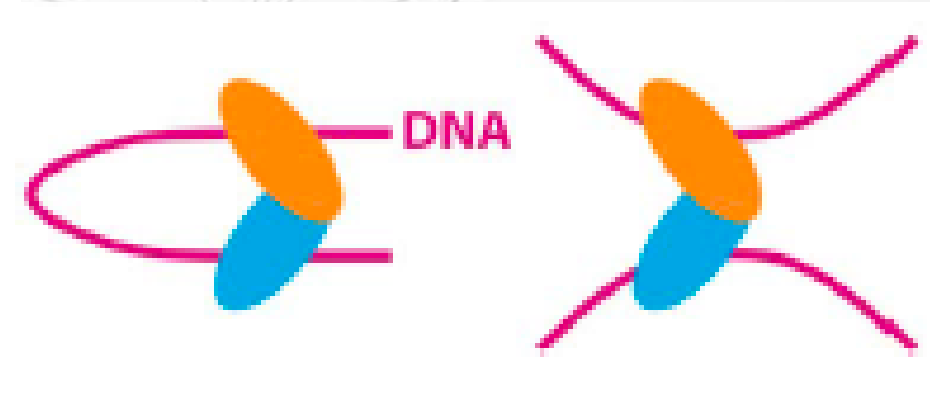
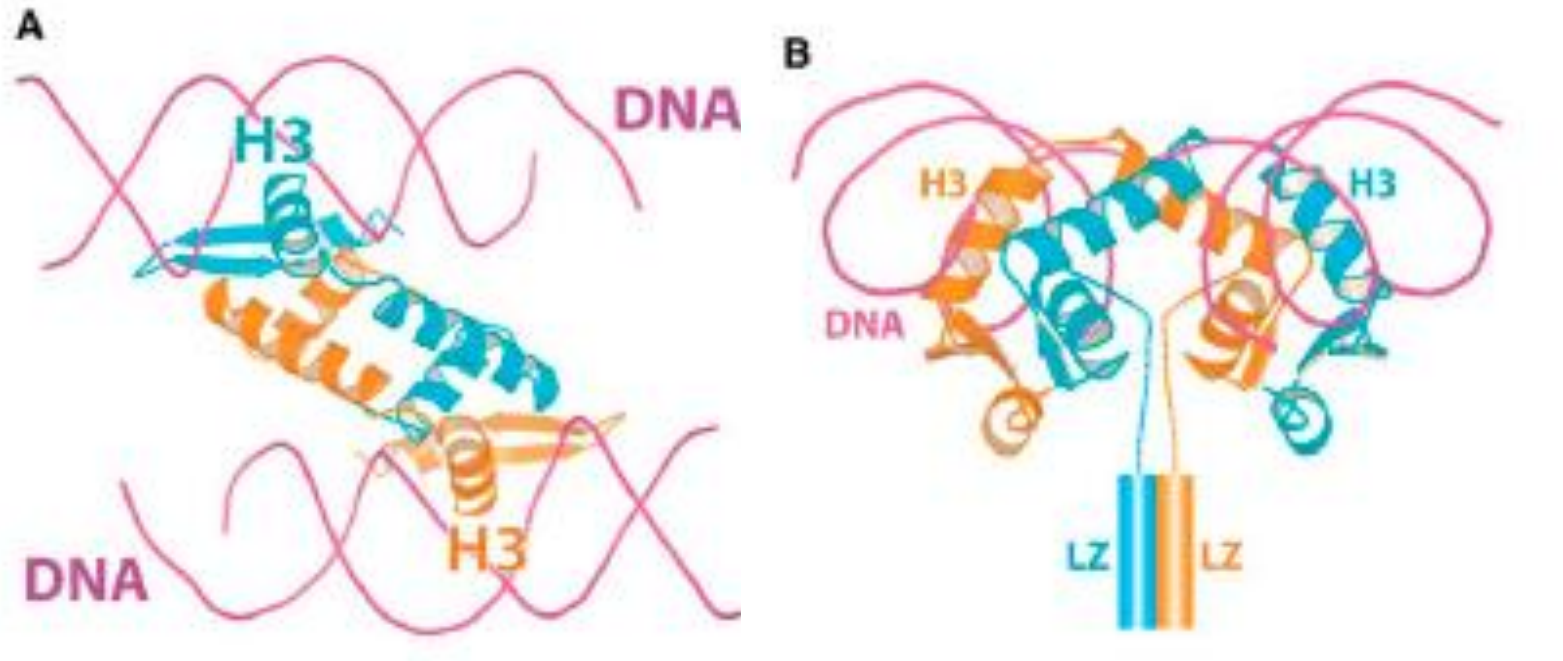
Monomer: 1, 2

Dimer: 3 & 4, 5 & 6



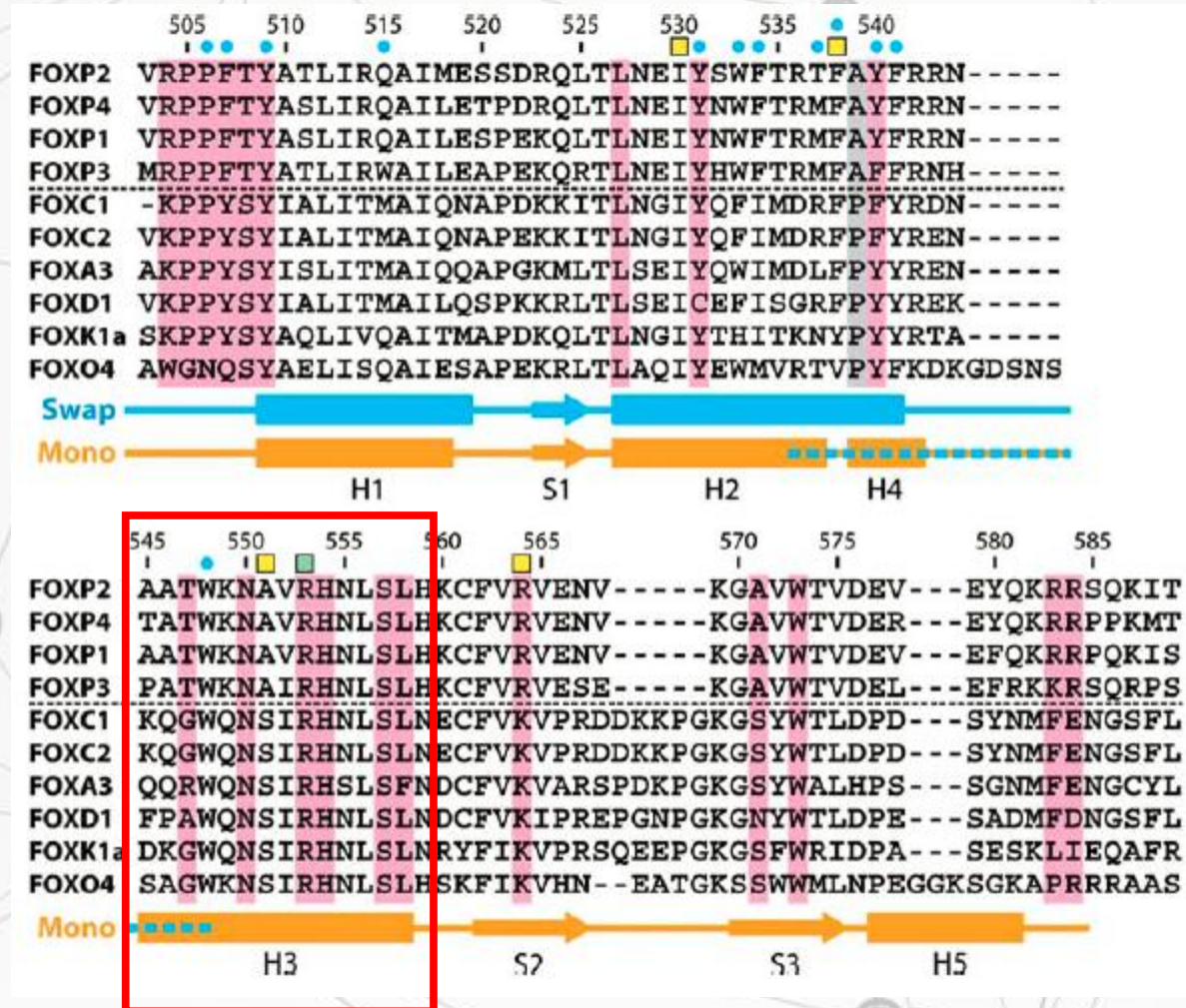
Monomer

Protein Dimerization

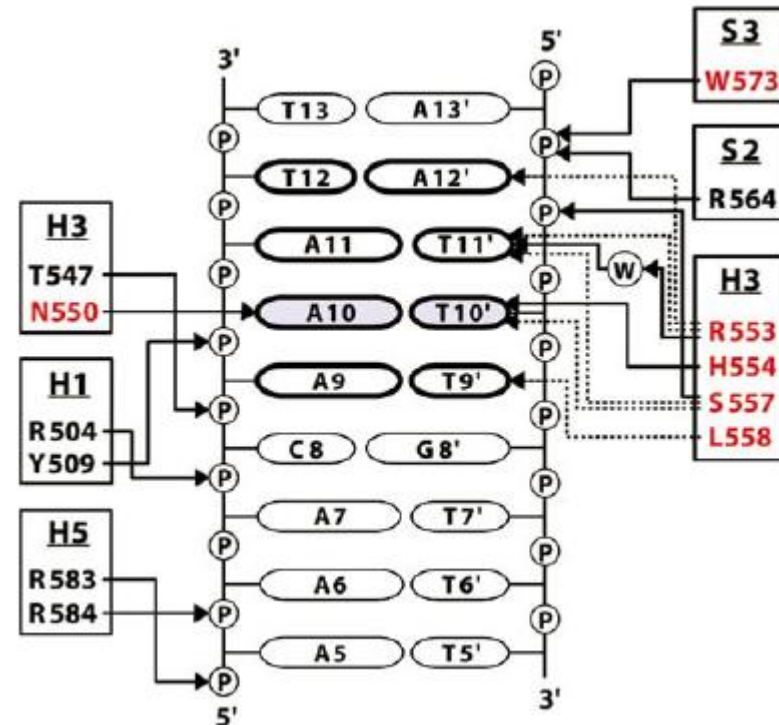
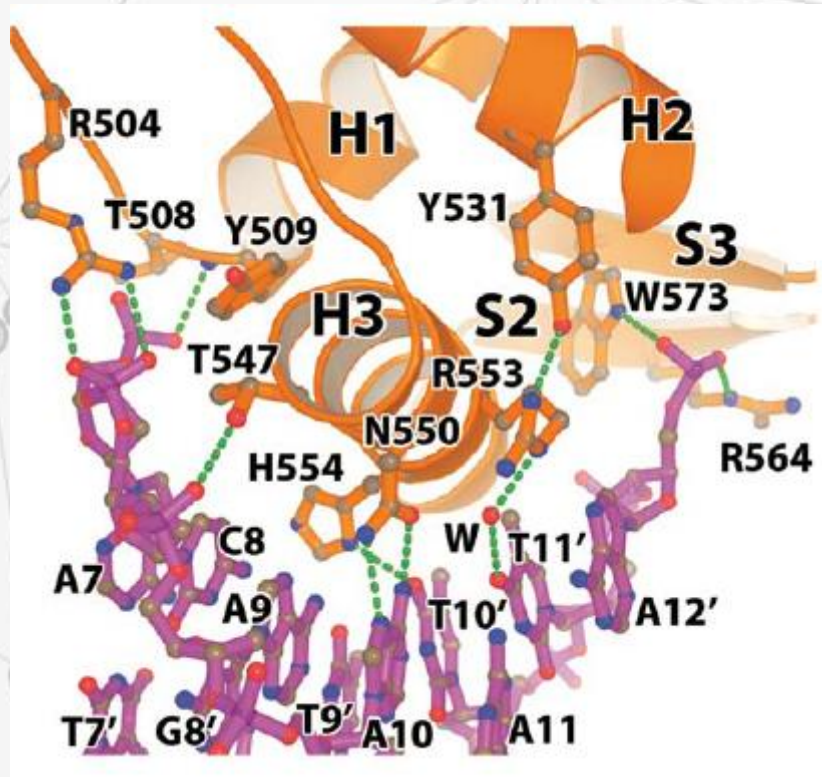


Zinc finger/Leucine zipper cooperate to regulate the protein dimerization

Conservation of the FOXP2 binding domain-H3



DNA Recognition

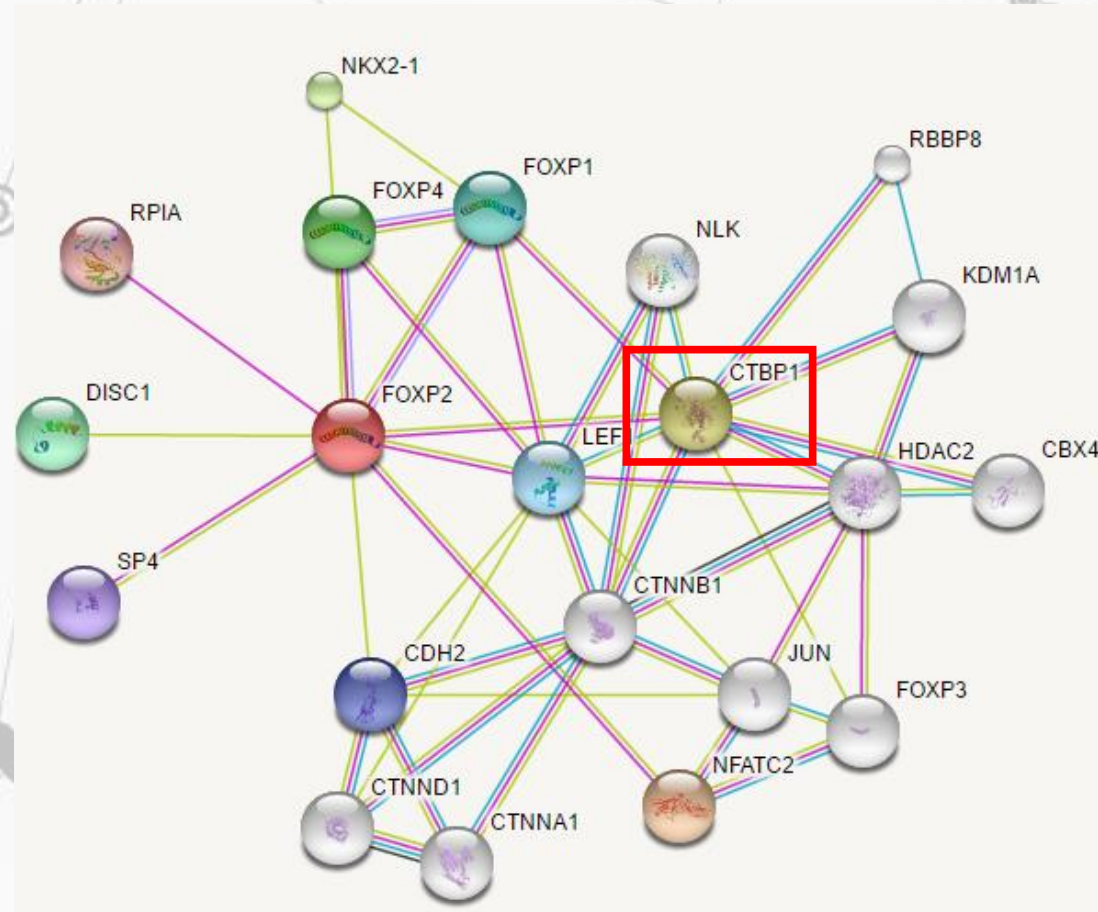


FOXP2 recognize DNA by Helix H3

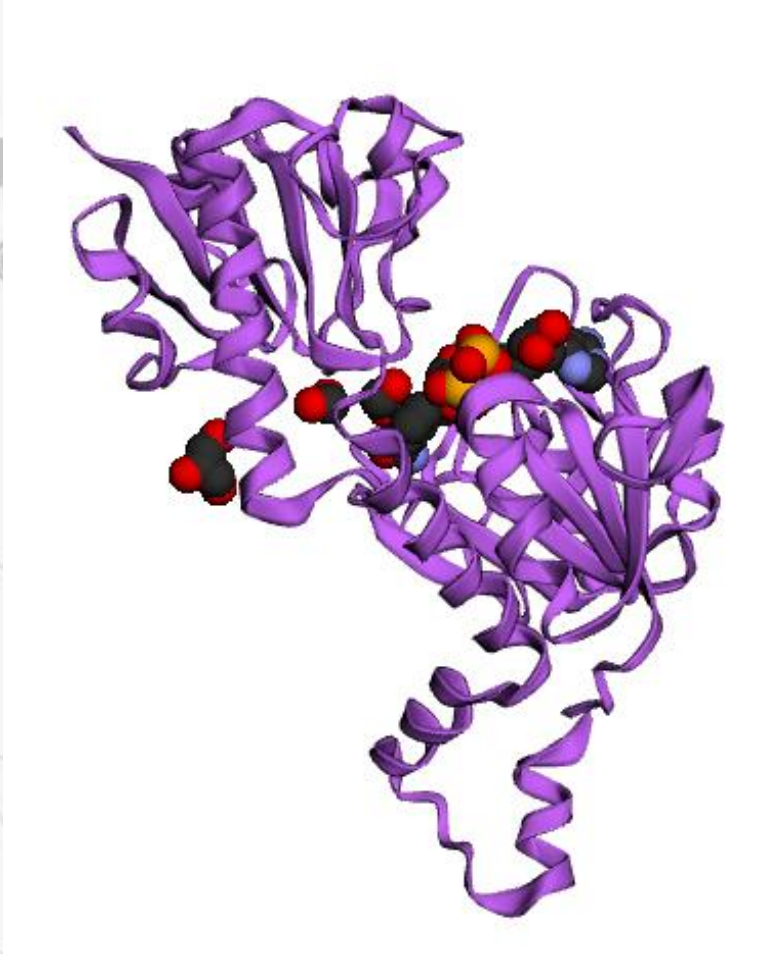
Hydrogen bond and Van der Waals' force play important roles in the recognition

James et al, Structure, 2009

Interactions of FOXP2

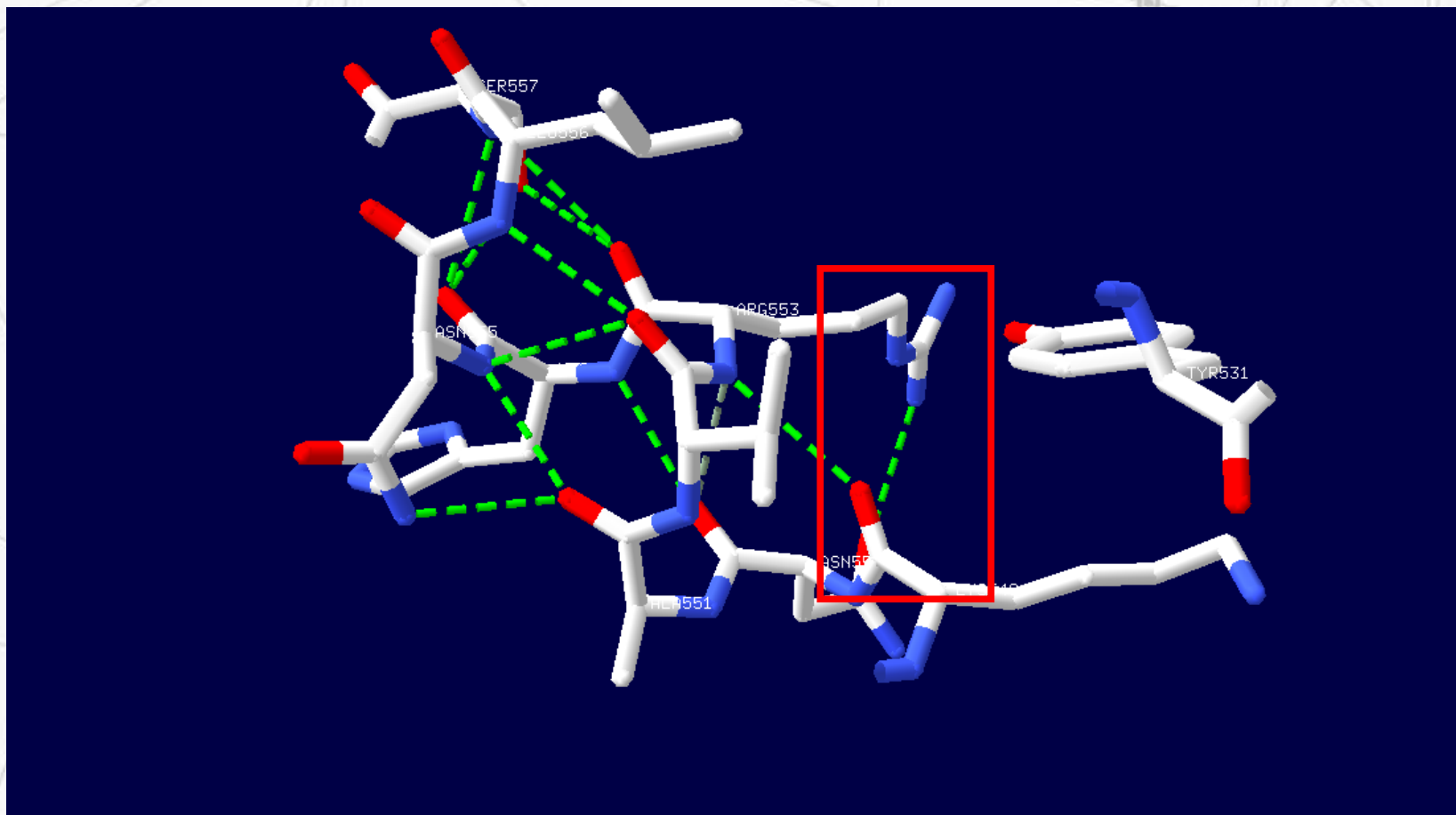


Corepressor-CTBP1

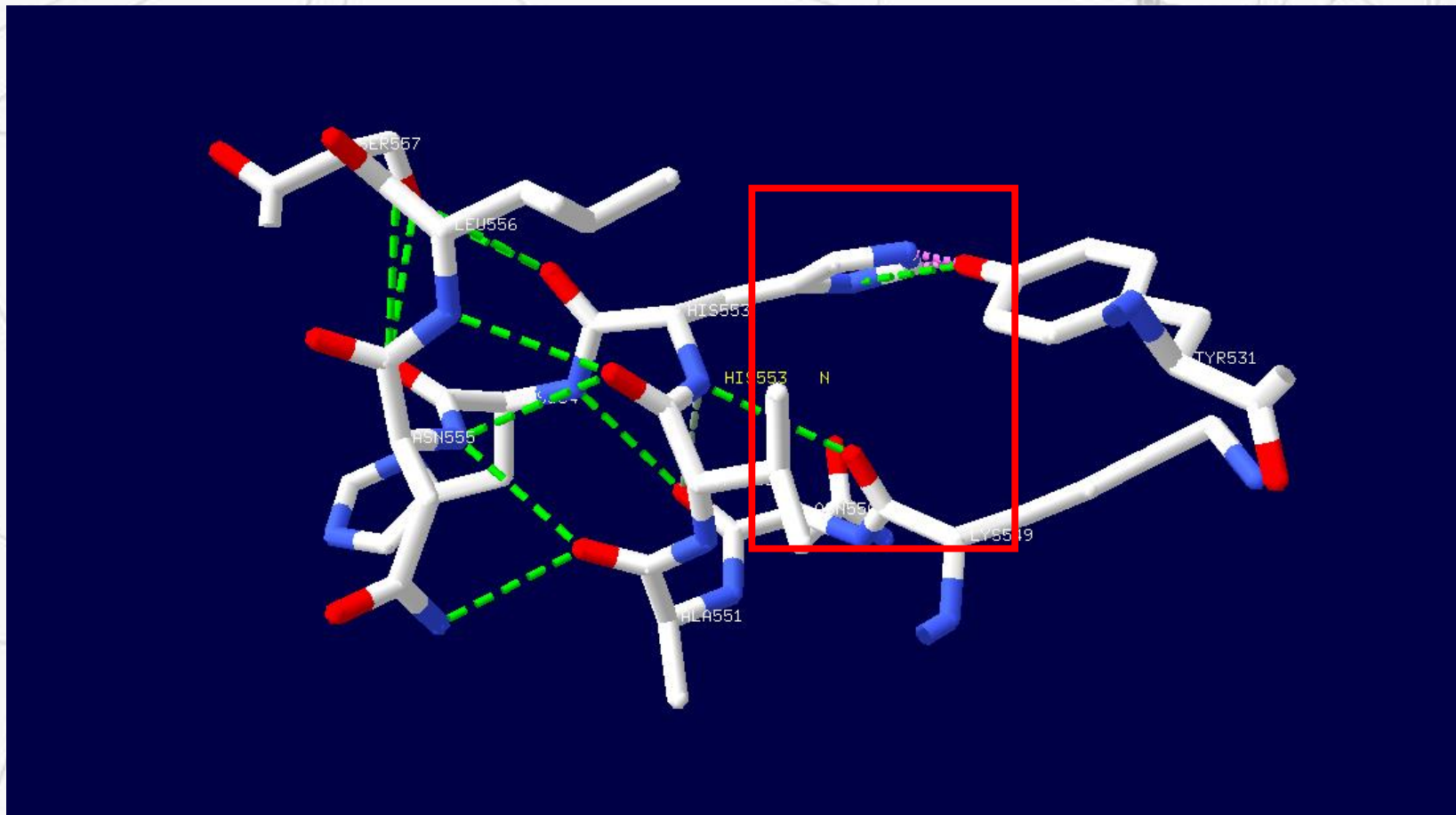


Corepressor targeting diverse transcription regulators such as GLIS2 or BCL6. Has dehydrogenase activity. Involved in controlling the equilibrium between tubular and stacked structures in the Golgi complex.

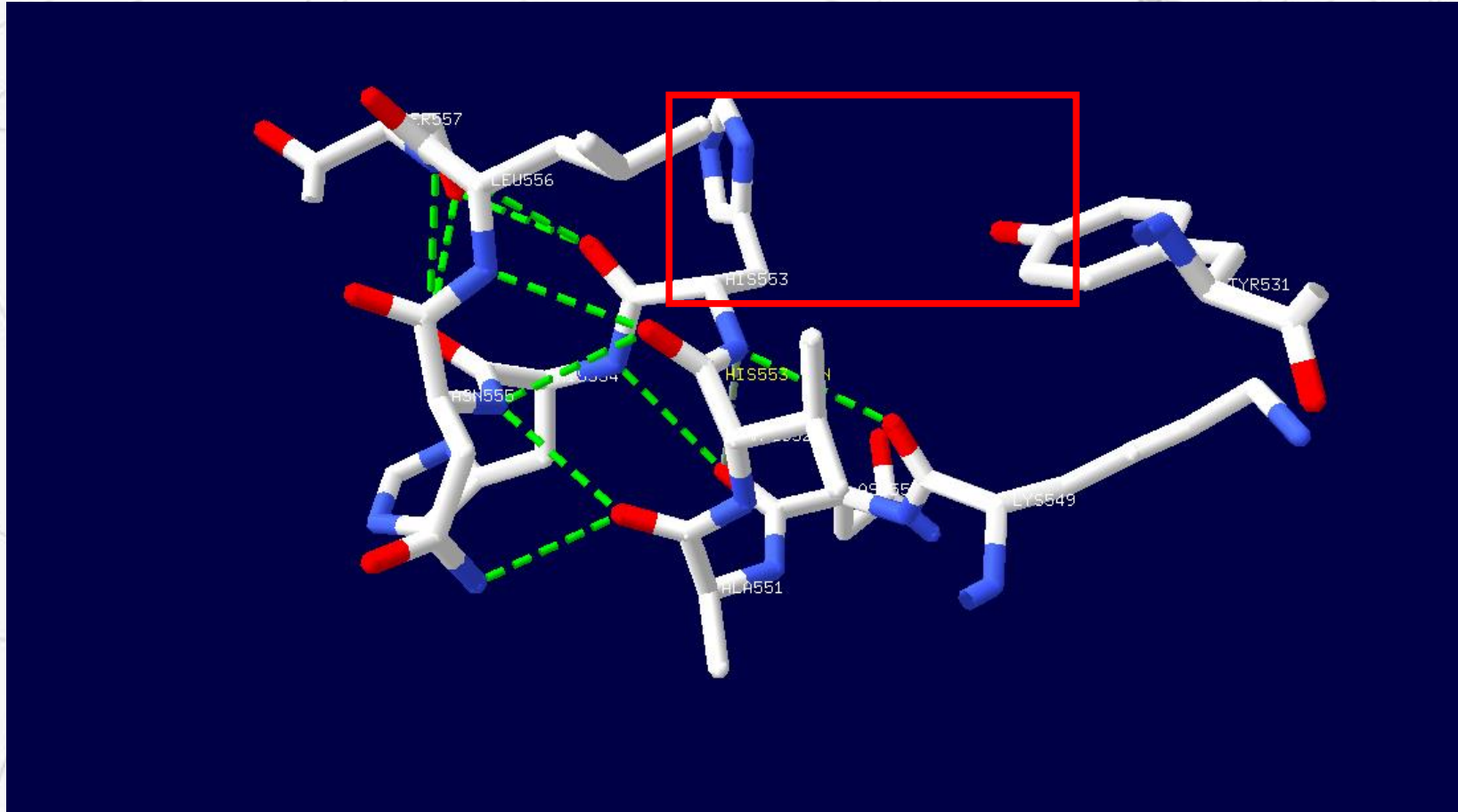
Wild Type



Mutation in R553H



Mutation in R553H



Summary

- Introduced some information about *FOXP2* gene and FOXP2 protein
- FOXP2 phylogenetic tree indicated that FOXP2 is an conserve protein in animals.
- The structure of FOXP2 binding domain figured out that FOXP2 binds DNA to work.

Acknowledgments

Prof. Luo Jingchu

TA: Ke Lan

My abc team members

Whole abc classmates

A complex network diagram with numerous nodes and connecting lines, rendered in a light gray color, serves as the background for the slide. The nodes vary in size and some contain internal patterns, while the lines represent connections between them.

Thanks for your listening!