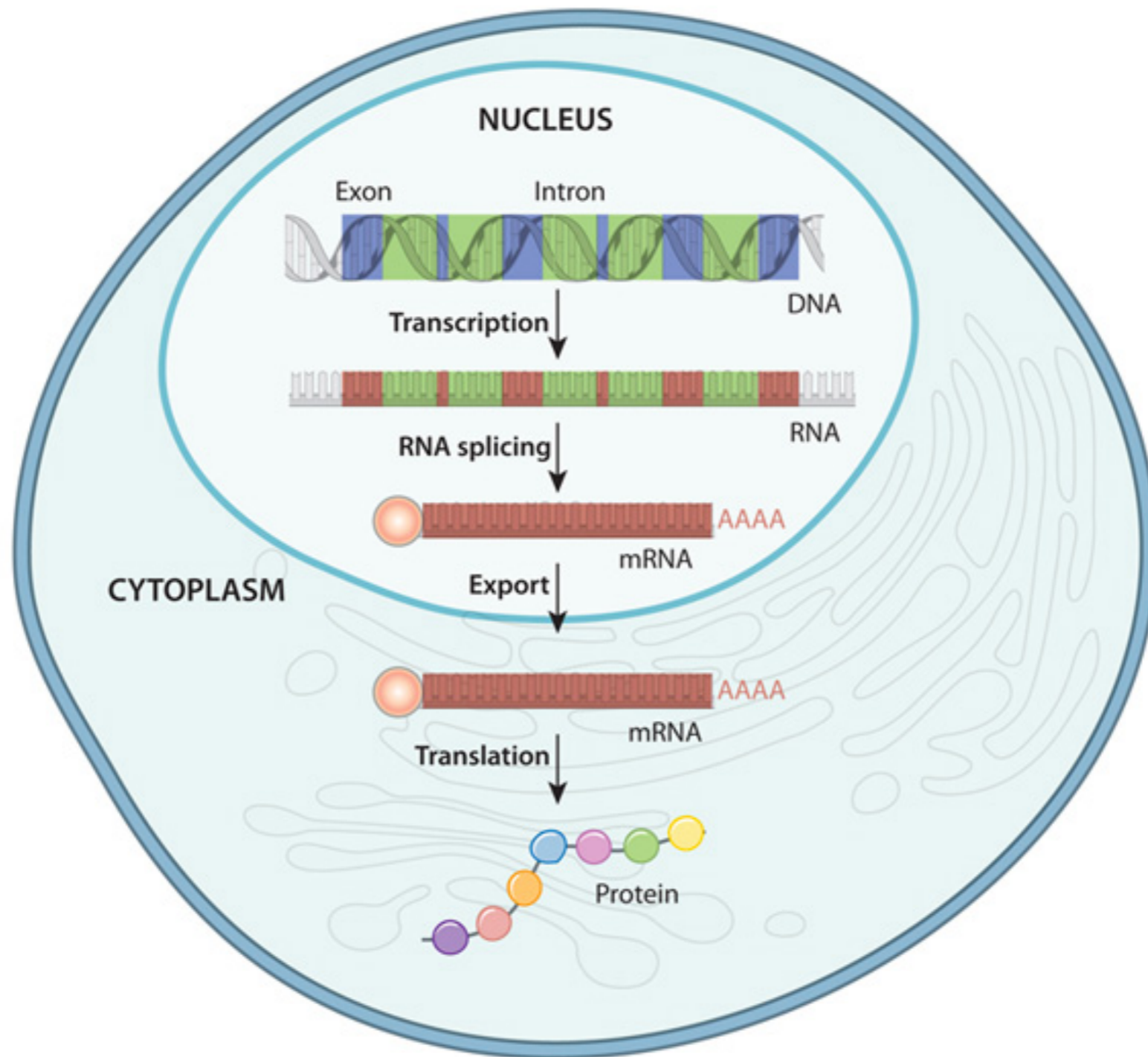


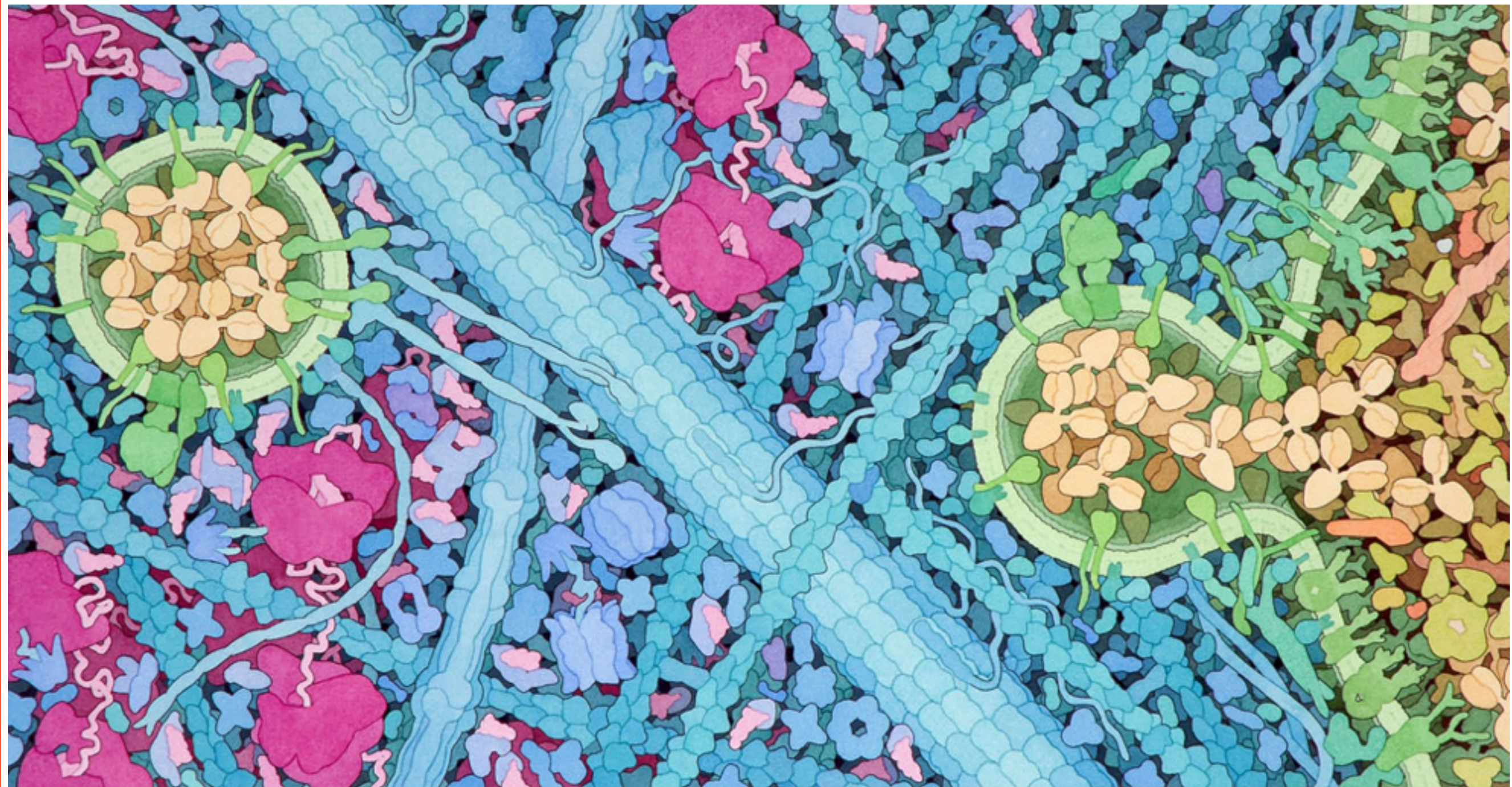
# Module 8 : Proteins, Complexes and Pathways

Hinxton, 13th May 2015

Rob Finn

# From Gene to Protein





A human B cell releases a packet (green) of antibody molecules (tan, Y-shaped) into the blood serum on the right of the image. Human serum albumin molecules appear as pale green triangles in the serum. Illustration from *The Machinery of Life* by David S. Goodsell.

# From Genes to Proteins to Functions

- How do we possibly create that image from the previous slide?
  - What do we need to know?

# From Genes to Proteins to Functions

- How do we possibly create that image from the previous slide?
  - What do we need to know?
    - Parts list
    - Which parts go together
    - Their spatial arrangement
    - The follow or path of them

```
>MySequencesOfInterest
```

```
MMQDVSSSPVSPADDSLNSNEEFPDRQQPPSGKRGGRKRRSSRRSAGGGA  
GPGGAAGGGVGGGDEPGSPAQGKRGKKSAGCGGGGGAGGGGGSSSSGGGSP  
QSYEELQTQRVMANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQTLK  
LAARYIDFLYQVLQSDDELDSKMASCSYVAHERLSYAFSVWRMEGAWSMSA  
SH
```

## **Which tools can we use?**

**PSI-BLAST**

**Phmmer, Jackhmmer**

# Why perform a homology search?

- Our ability generate sequence far outstrips our ability to functionally characterise them
  - Homology can allow you to identify a sequence that has been characterised
  - Possible to identify conserved residues between query and target

# UniProt

The screenshot shows the UniProt website interface. At the top left is the UniProt logo. A search bar contains 'UniProtKB' and an 'Advanced' search button. Below the search bar is a navigation menu with 'BLAST', 'Align', 'Retrieve/ID Mapping', 'Help', and 'Contact'. A mission statement follows: 'The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.'

The main content area is divided into several sections:

- UniProtKB**: Contains 'Swiss-Prot (547,357)' which is 'Manually annotated and reviewed', and 'TrEMBL (89,451,166)' which is 'Automatically annotated and not reviewed'.
- UniRef**: 'Sequence clusters' with a circular arrow icon.
- UniParc**: 'Sequence archive' with a database icon.
- Proteomes**: Represented by icons of a fly, a person, and a protein structure.
- Supporting data**: A grid of six categories:
 

Literature citations	Taxonomy	Subcellular locations
Cross-ref. databases	Diseases	Keywords
- News**: Includes social media icons and two news items:
  - 'Thalidomide, the pharmacological version of yin and yang | Cross-references to DEPOD, MoonProt and Proteomes' with link 'UniProt release 2015\_01'.
  - 'Higher and higher | New mouse and zebrafish variation files | Structuring of 'cofactor' annotations' with link 'UniProt release 2014\_11'.
- A 'News archive' link is located at the bottom right.



# UniProt

**P10398 - ARAF\_HUMAN**

**Protein** | Serine/threonine-protein kinase A-Raf  
**Gene** | **ARAF**  
**Organism** | *Homo sapiens (Human)*  
**Status** | Reviewed - ●●●●● - Experimental evidence at protein level<sup>i</sup>

Display None

FUNCTION  
 NAMES & TAXONOMY

**Function<sup>i</sup>**

Involved in the transduction of mitogenic signals from the cell membrane to the nucleus. May also be involved in the regulation of the cell cycle.

1 Publication

BLAST Align Format Add to basket History

UniProt-SwissProt entry

- Manually annotated, high confidence.

**Q2KJA0 - Q2KJA0\_BOVIN**

**Protein** | Submitted name: **ARAF protein**  
**Gene** | **ARAF**  
**Organism** | *Bos taurus (Bovine)*  
**Status** | Unreviewed - ●●○○○ - Experimental evidence at transcript level<sup>i</sup>

Display None

FUNCTION  
 NAMES & TAXONOMY  
 SUBCELLULAR LOCATION

**Function<sup>i</sup>**

**Catalytic activity<sup>i</sup>**  
 ATP + a protein = ADP + a phosphoprotein. SAAS annotation

BLAST Align Format Add to basket History

UniProt-TrEMBL entry

- Automatically annotated, low(er) confidence.

```
>MySequencesOfInterest
```

```
MMQDVSSSPVSPADDSLNSNEEPEPDRQQPPSGKRGGRKRRSSRRRSAGGGA  
GPGGAAGGGVGGGDEPGSPAQGKRGKKSAGCGGGGGAGGGGGSSSSGGGSP  
QSYEELQTQRVMANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQTLK  
LAARYIDFLYQVLQSDDELDSKMASCSYVAHERLSYAFSVWRMEGAWMSA  
SH
```

## **Variance at 156, I to V?**

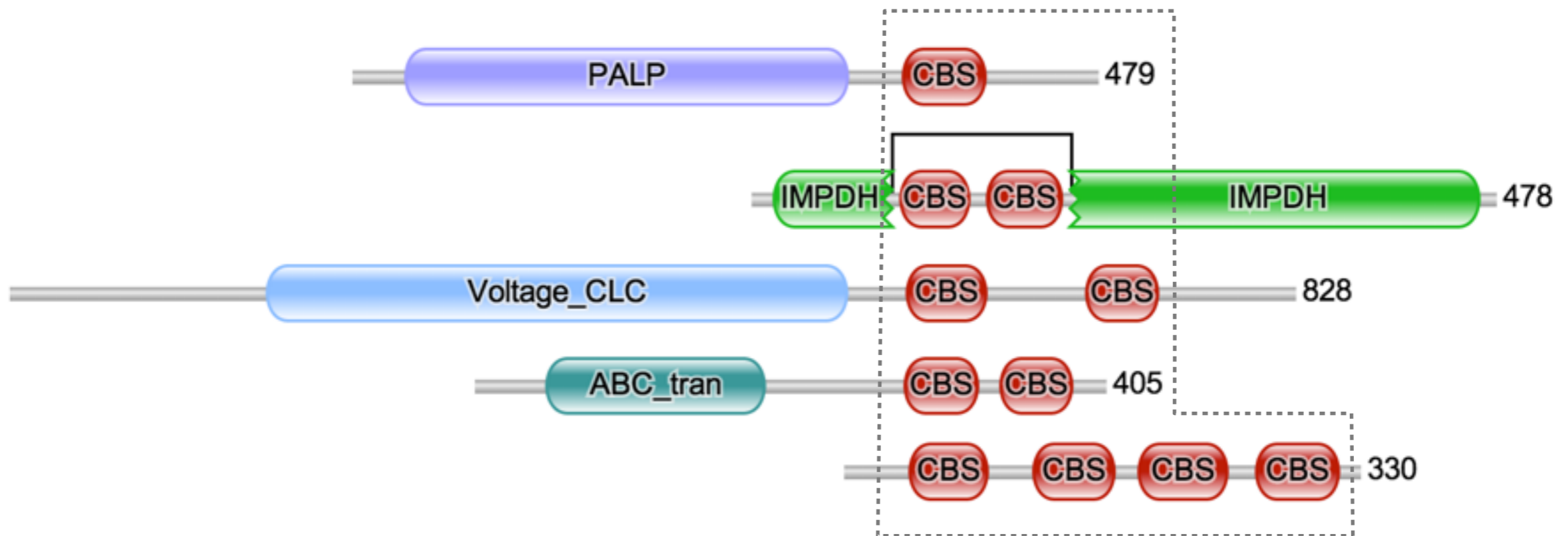
What is it's effect on the protein

# Other Homology Searches

- Compare against Pfam
  - HLH DNA binding domain
    - Description of function
    - Known structures



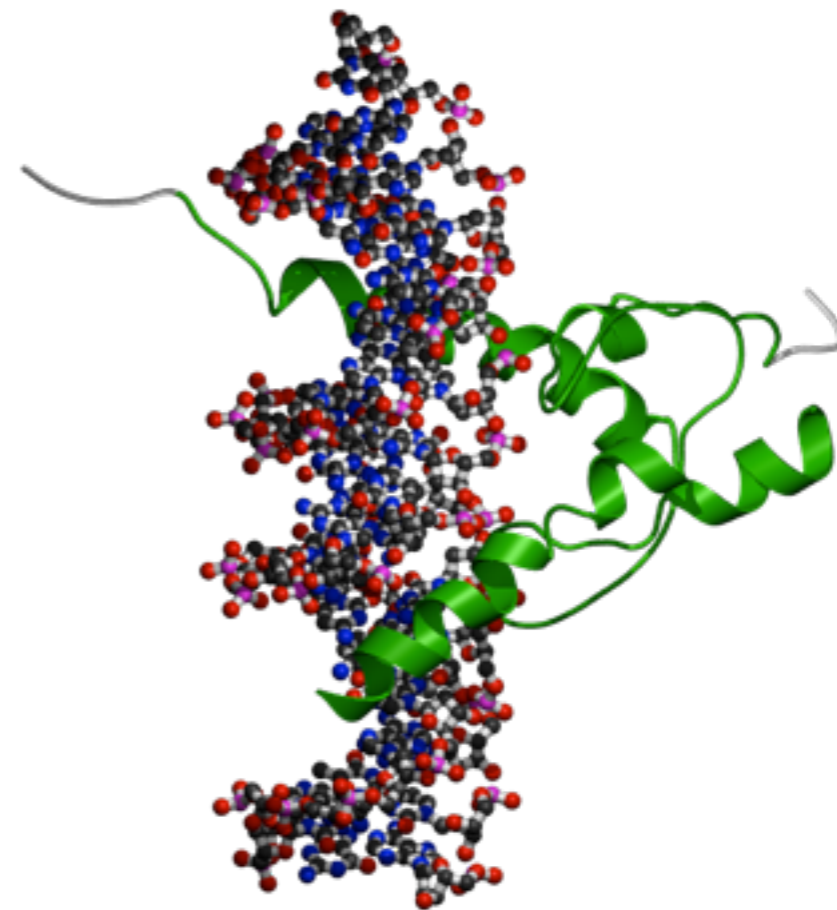
# Why Domains?



- Functional diversity
- Conservation
- Tractable drug targets

# Homology Searches

- Compare against Pfam
  - HLH DNA binding domain
    - Description of function
    - Known structures
- Structural homologs
  - DNA binding
  - Functions as a dimer



# No 3D Structure

.....\*.....\*.....\*.....\*.....\*.....  
Query 109 qrvmanvrerqrtslneafaalrkiptlp.sdklskiqtlklaaryidflyqvlqs 165  
+r+ an rer r ln a+ lrk++p + klski+tl+la +yi l ++l+s  
Target 2 RRMKANARERNRMHGLNAALDNLRKVVPCYSkTQKLSKIETLRLAKNYIWALSEILRS 59  
PP 6899\*\*\*\*\*975278\*\*\*\*\*99987

# No 3D Structure

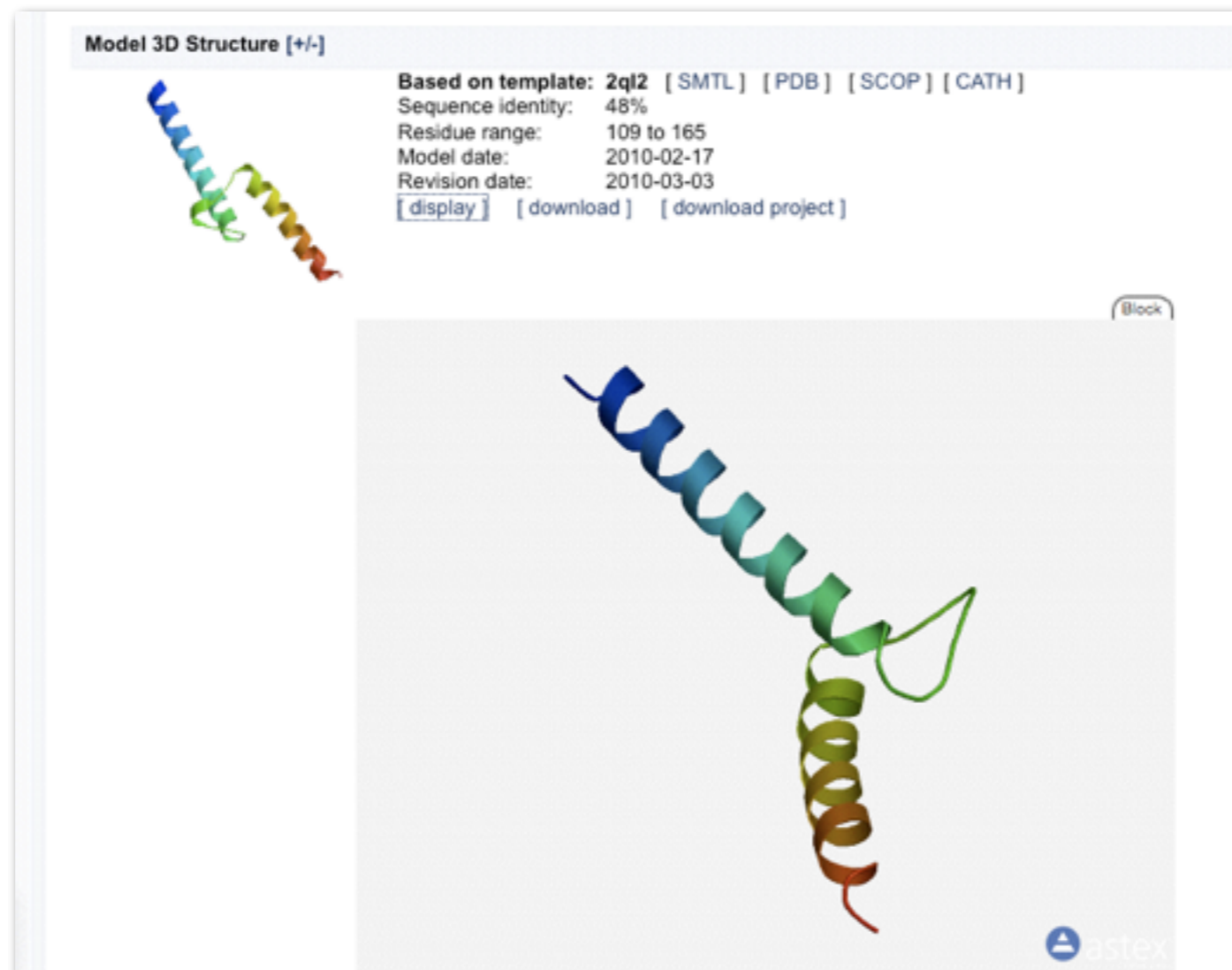
```

.....*.....*.....*.....*.....*.....
Query   109  qrvmanvrerqrtslnaafaalrkiiptlp.sdklskiqtlklaaryidflyqvlqs 165
      +r+ an rer r   ln a+  lrk++p   + klski+tl+la +yi  l ++l+s
Target  2   RRMKANARERNRMHGLNAALDNLRKVVPCYSkTQKLSKIETLRLAKNYIWALSEILRS 59
PP      6899*****975278*****99987

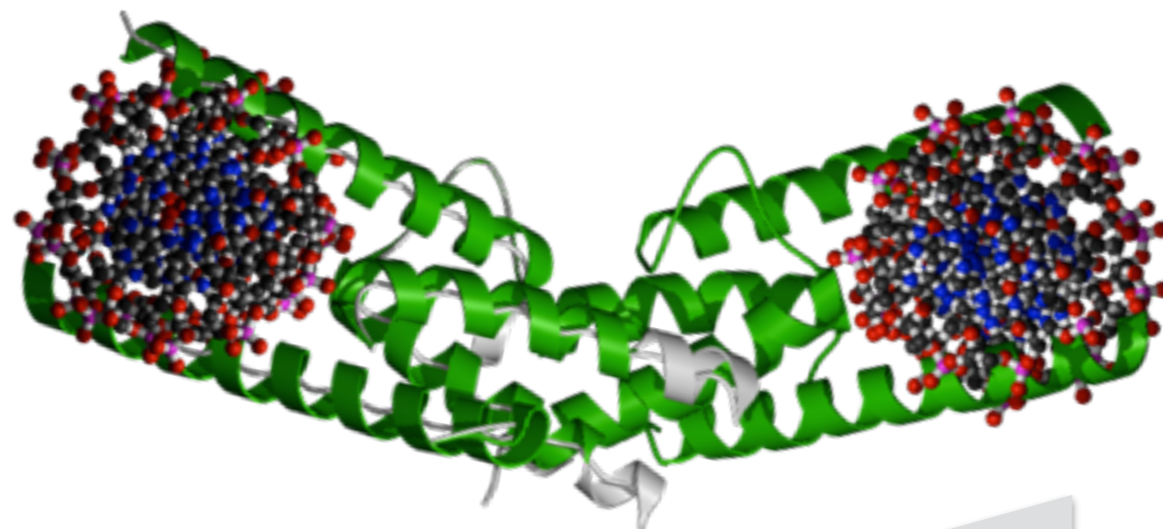
```

- SWISS-MODEL

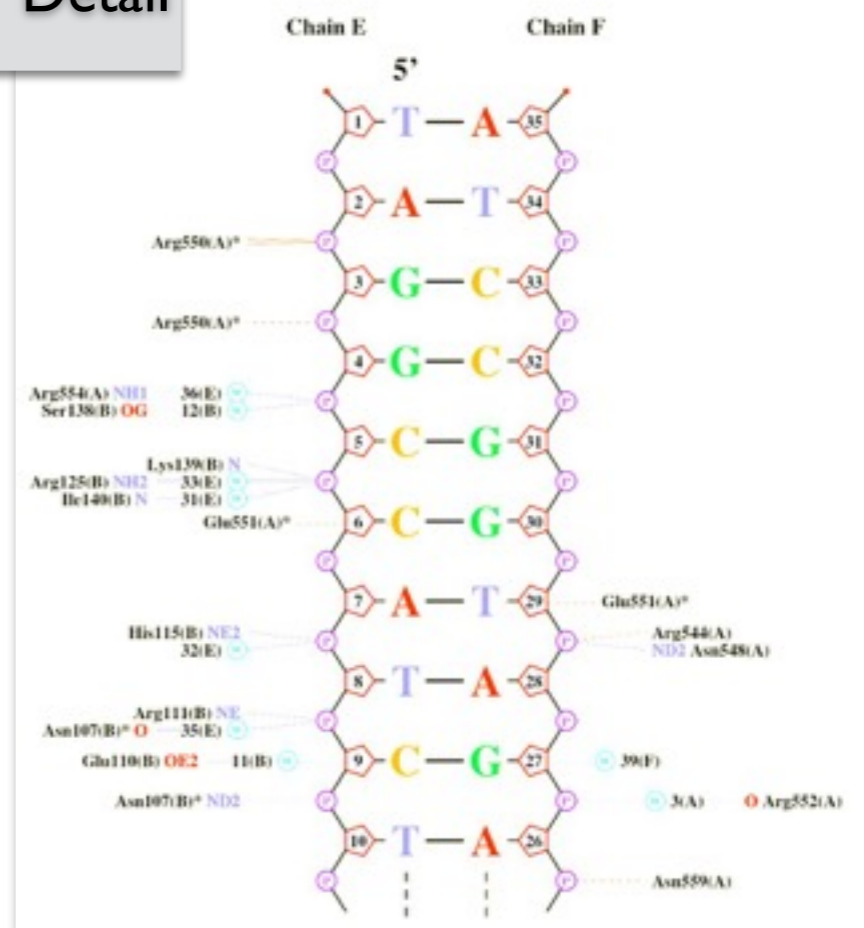
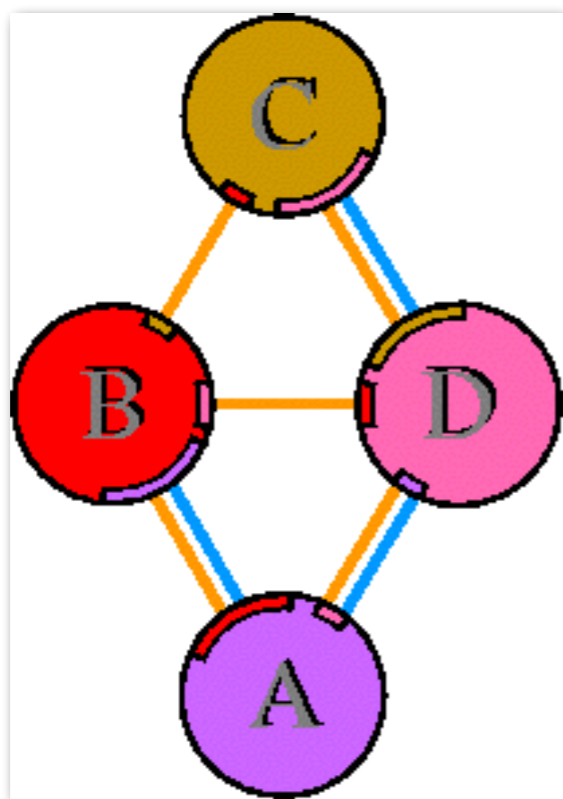
- Homology Searching and Modelling



# PDBsum - 2QL2

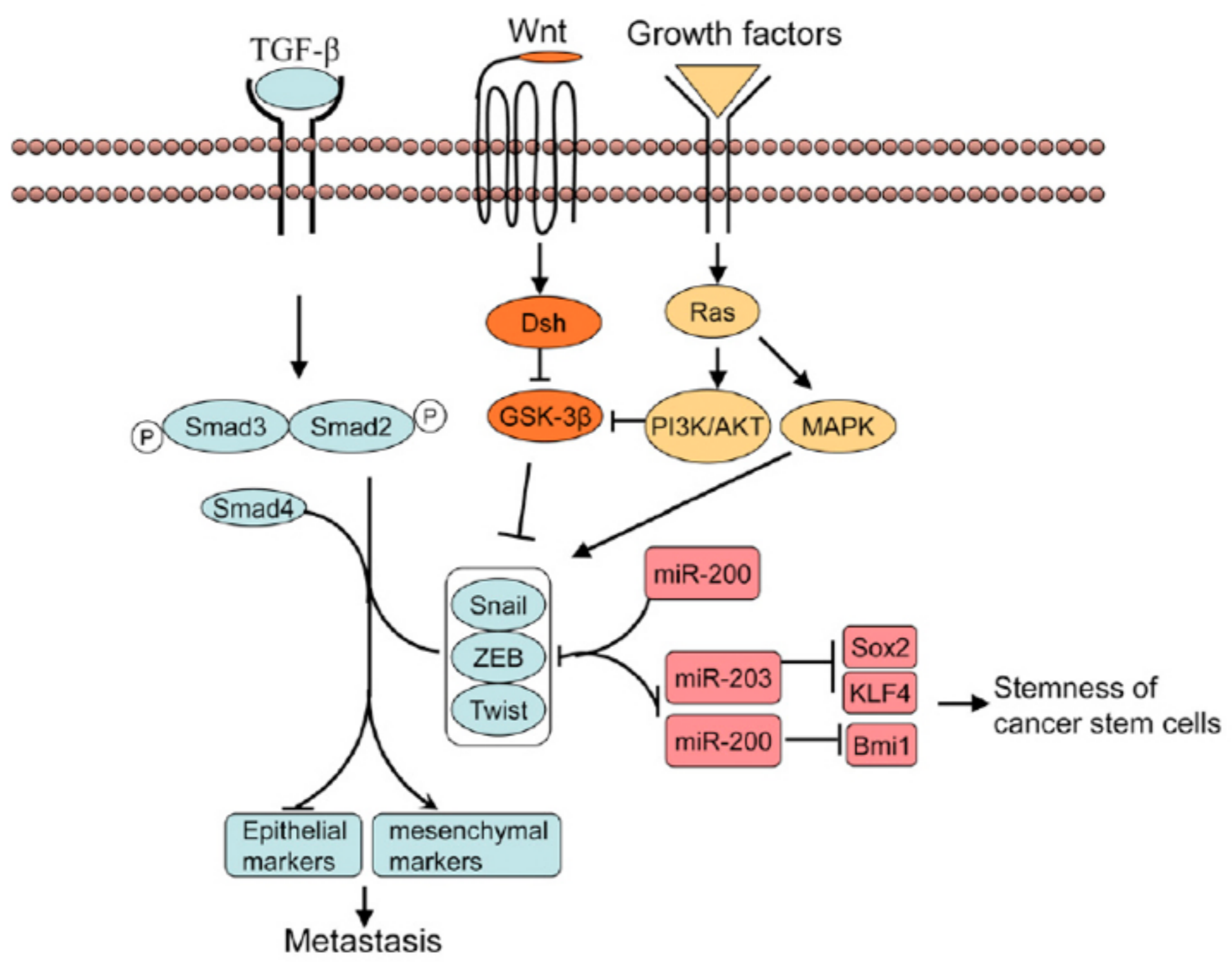


Increasing  
Detail





























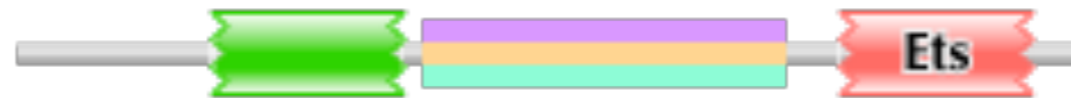
# Protein Interactions



# IntAct

Export to: <input type="text" value="Select format..."/> <input type="button" value="Export"/> <input type="button" value="Change Columns Displayed"/>												
	Name molecule A	Links molecule A	Name molecule B	Links molecule B	Aliases molecule A	Aliases molecule B	Species molecule A	Species molecule B	Publication Identifier	Interaction Detection Method	Interaction AC	Expansion Method
1	H-twist  	 	ets2_human  	 	H-twist; Class A basic helix-loop-helix protein 38; BHLHA38; [a]	ets2_human	9606	9606	IM-12059 18599945	anti tag coimmunoprecipitation	EBI-1797303 imex:IM-12059-1	
2	 	 	 	 					IM-12059 18599945	anti bait coimmunoprecipitation	EBI-1797350 imex:IM-12059-4	
3	H-twist  	 	GAP and centrosome-associated protein  	 	H-twist; Class A basic helix-loop-helix protein 38; BHLHA38; [a]	GAP and centrosome-associated protein; Rab6 GTPase-activating protein GAPConA; HSPC094; [a]	9606	9606	20195357	display technology	EBI-2689840	

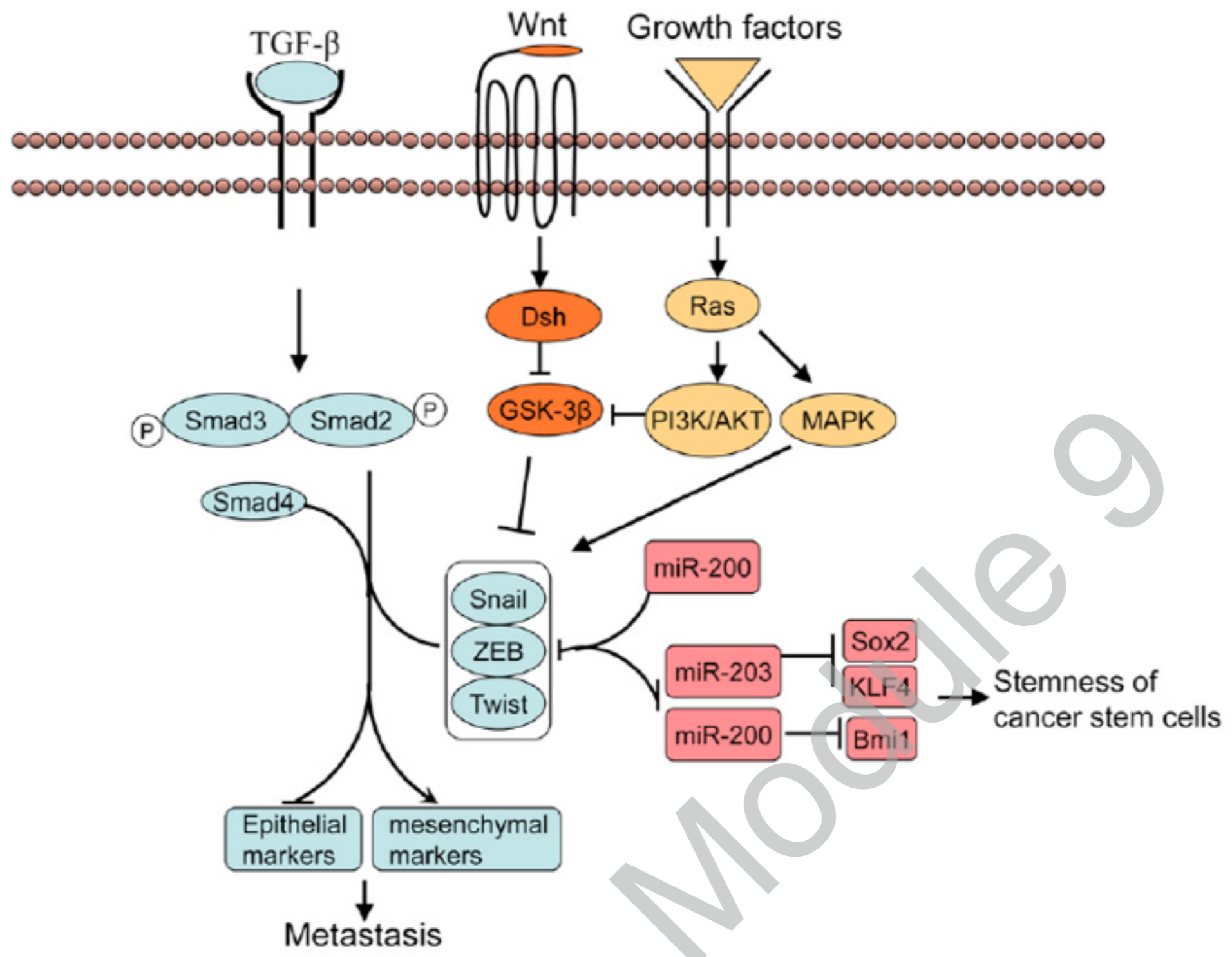
EST - Another transcription factor



Rab6 - GTPase-activating protein



# Protein Interactions



# Expression Profiles

E-MTAB-513

RNA-Seq of human individual tissues and mixture of 16 tissues (Illumina Body Map)

Organism(s): *Homo sapiens*

Reference(s): [22496456](#) [22955988](#) [23258890](#)



Gene query <sup>?</sup>   Exact match  Gene set <sup>?</sup>

Organism Part <sup>?</sup>   Specific <sup>?</sup>

Expression level cutoff <sup>?</sup>



Showing 1 of 1 genes found:



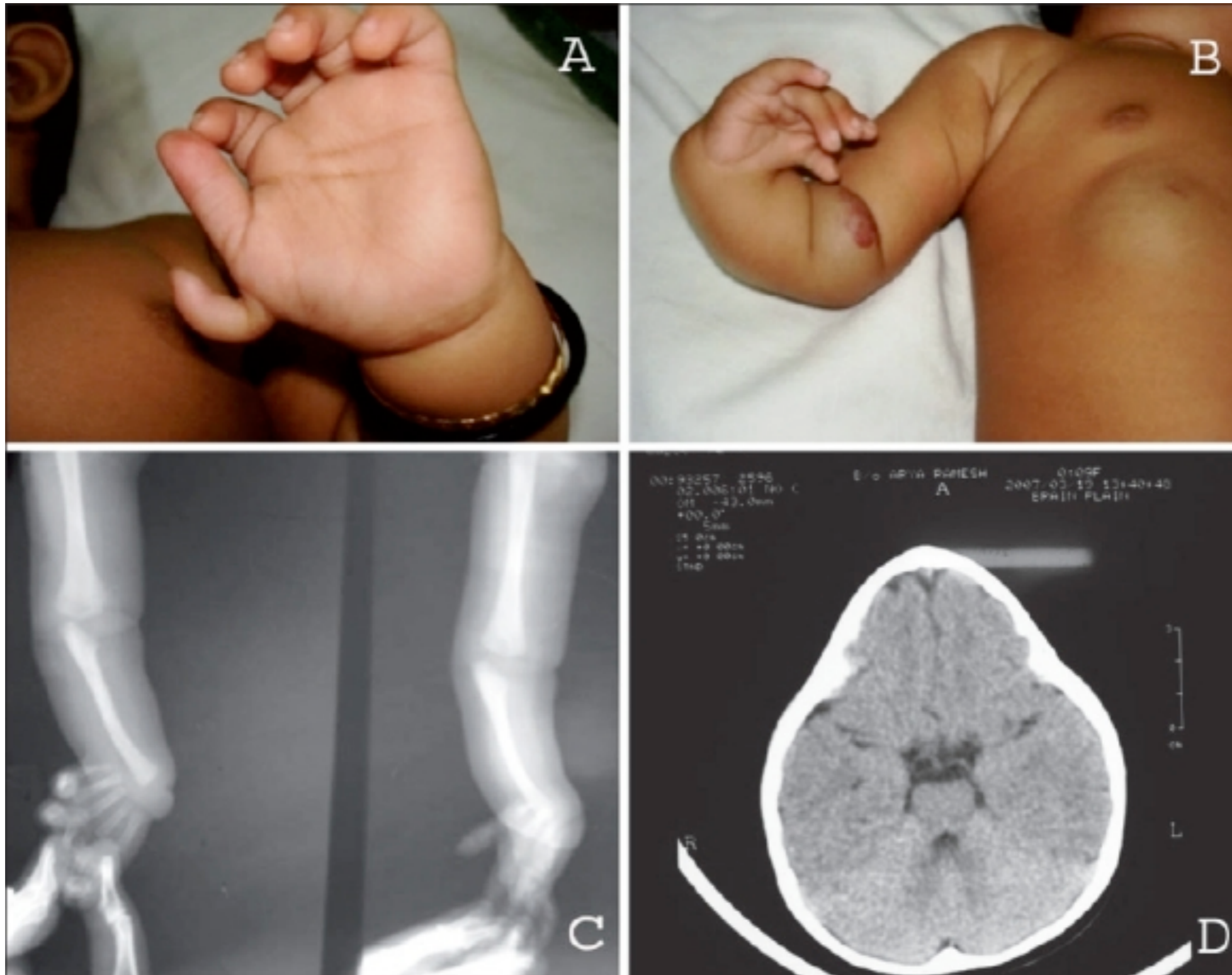
	adipose	adrenal	brain	breast	colon	heart	kidney	leukocyte	liver	lung	lymph node	ovary	prostate	skeletal muscle	testis	thyroid
Twist	High	Low	Low	High	Low	Low	Low	Low	Low	Low	Low	Low	High	Low	Low	Low

Module X

# So what do we know so far?

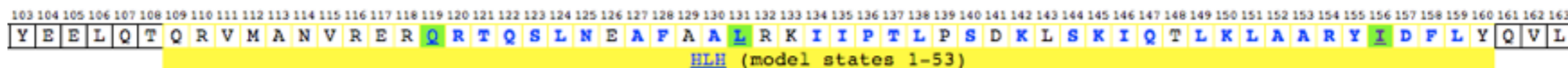
- Contains HLH and disorder
- Does not have a known structure, but homologs do
- Functions as a dimer and binds DNA and other proteins
- Which residues bind DNA and each other
- Regulators of miRNAs, with are themselves gene regulators
- Disease/Tissue expression patterns
- Signalling pathway

# Baller-Gerold Syndrome



# Baller-Gerold Syndrome Mutations

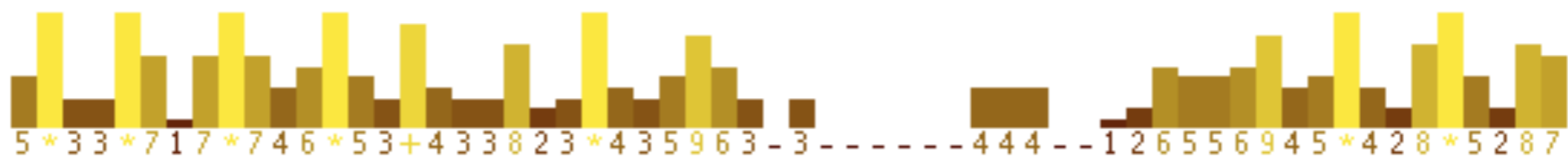
OMIM

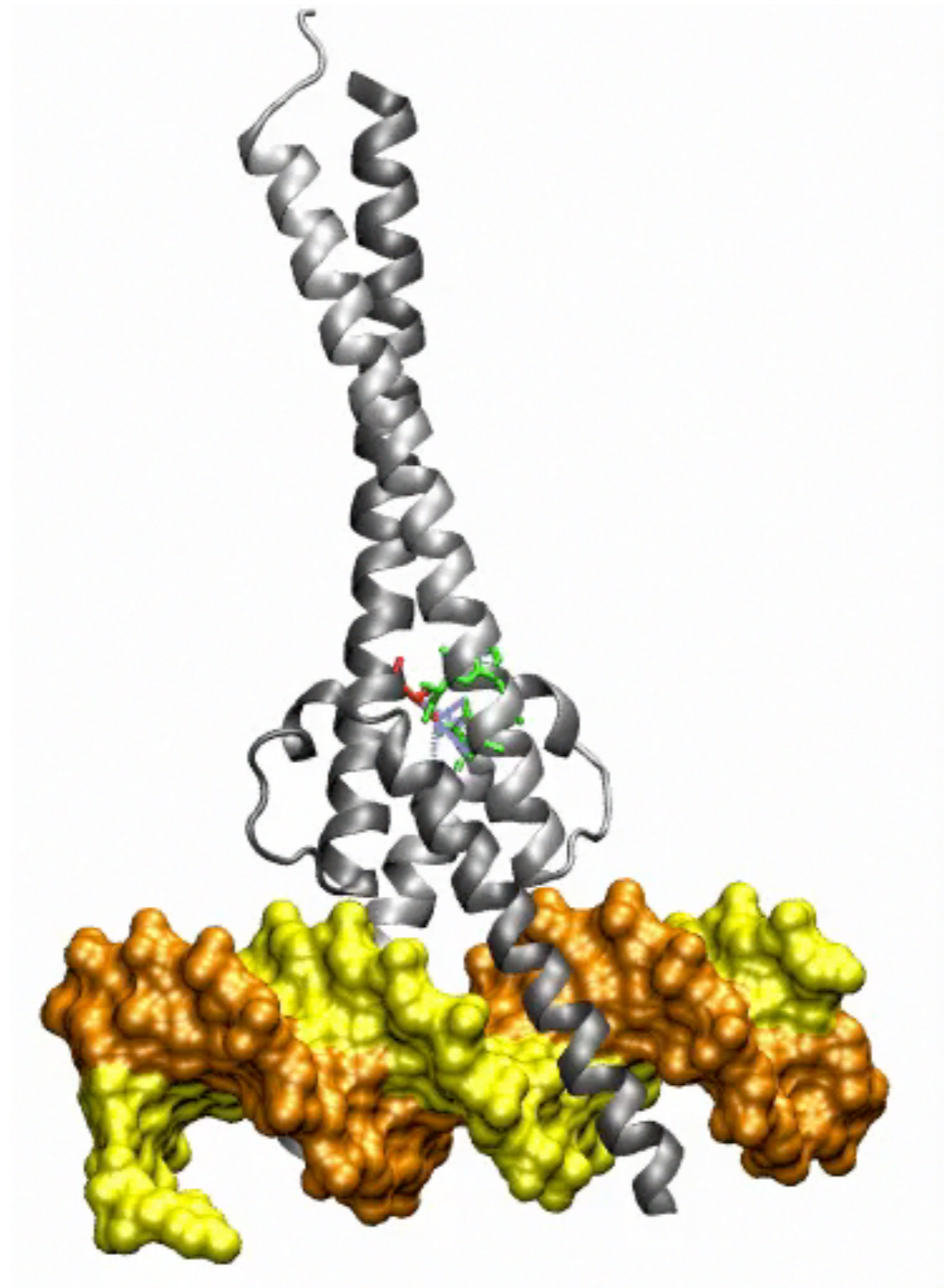


Pfam

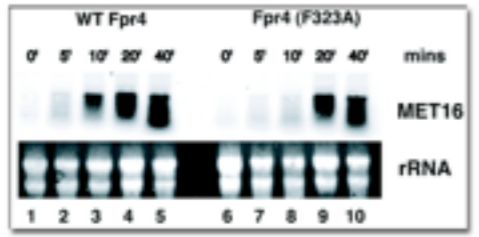


Conservation

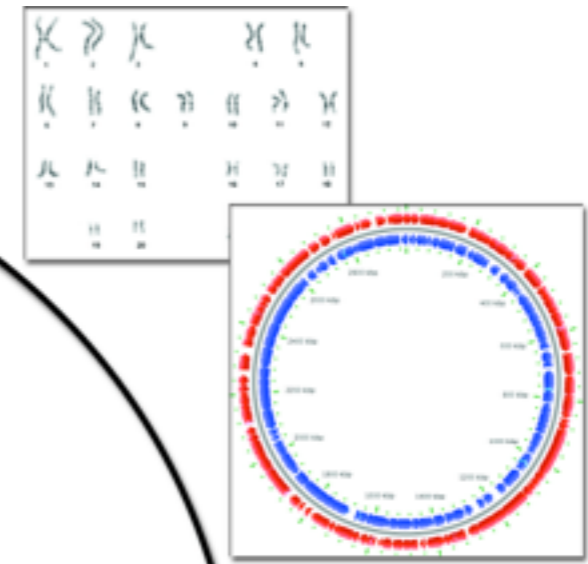








**SEQUENCE**  
10-100 Million

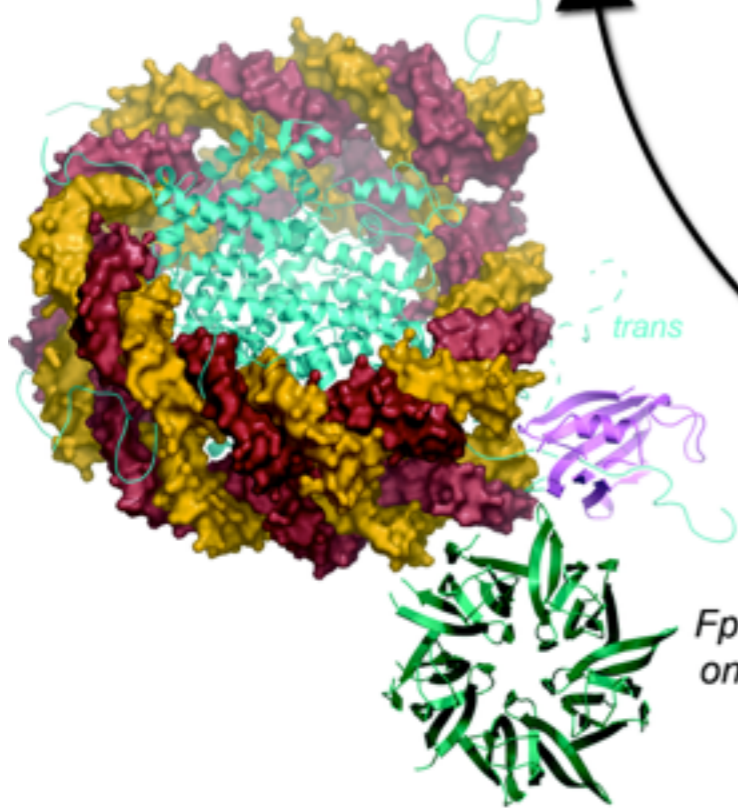


**DATABASES**



**CLASSIFICATION**  
10 Thousand Families

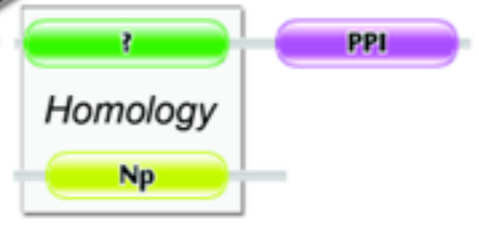
**FUNCTION**



*Fpr4 model based on nucleoplasmin*

*Model of histone modification by Fpr4*

**HYPOTHESIS**



*N-terminal domain of Fpr4 related to nucleoplasmin*



# Worked Examples