Module 8 : Proteins, Complexes and Pathways

Hinxton, 13th May 2015

Rob Finn

From Gene to Protein

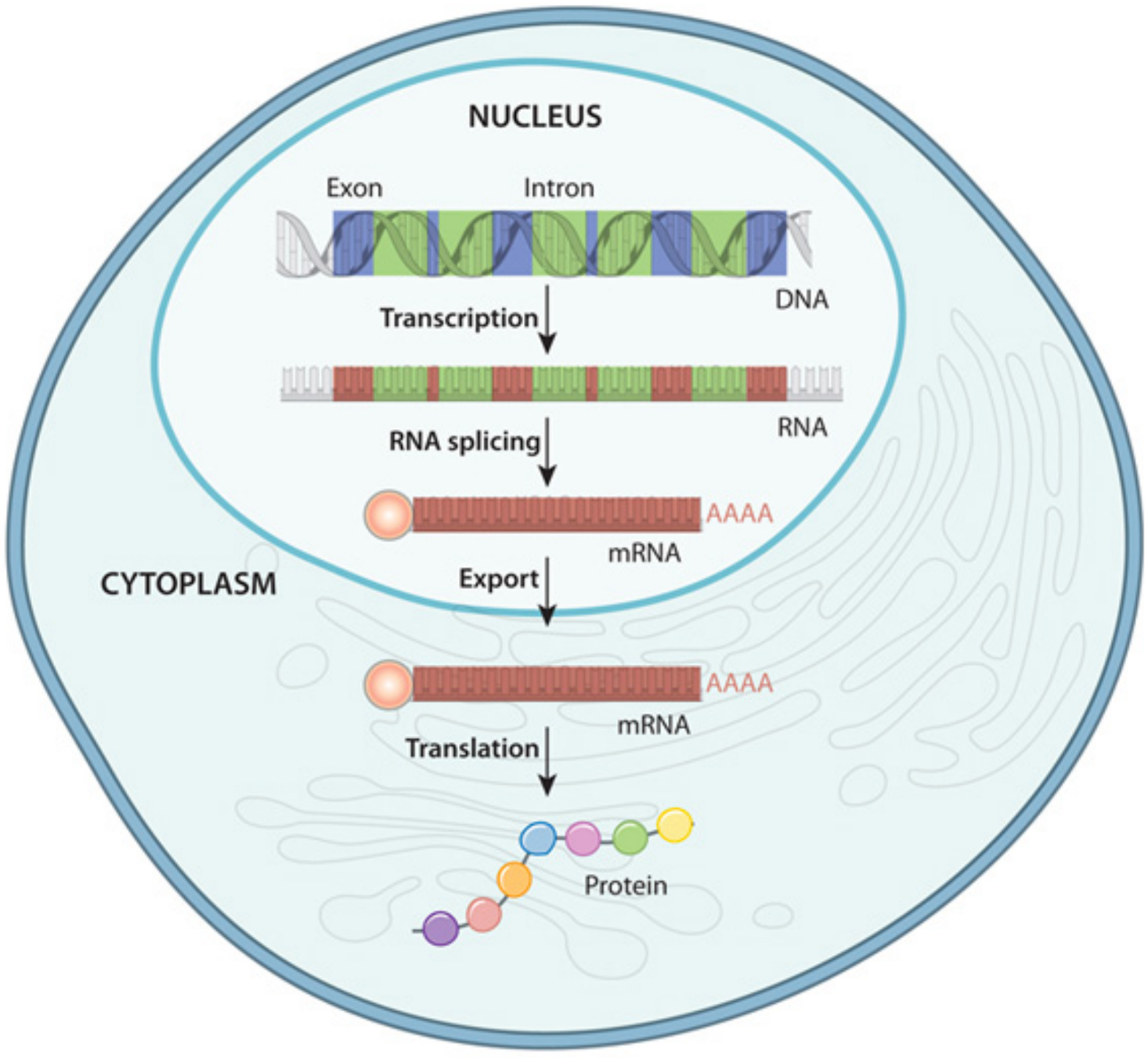
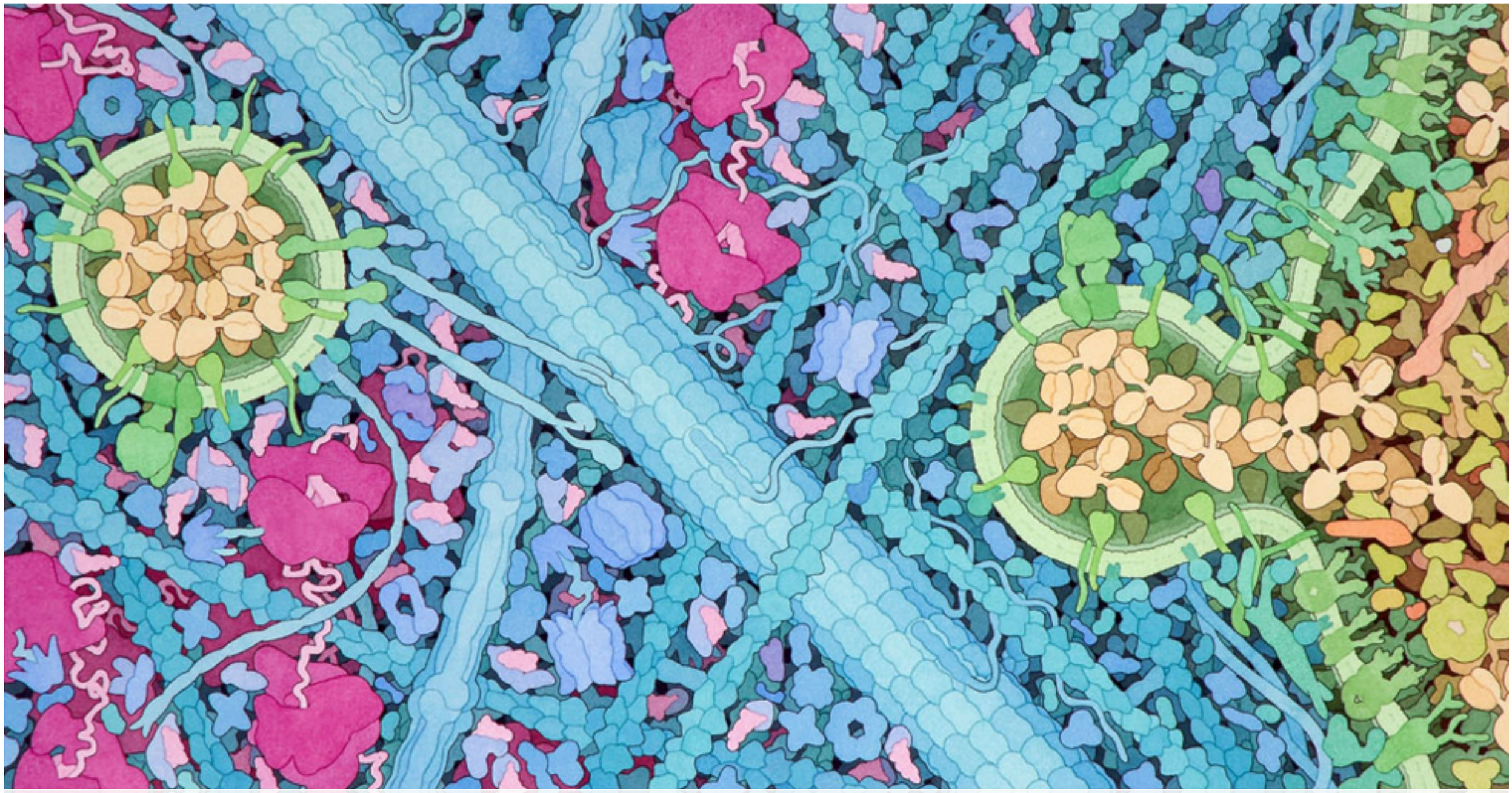


Image from http://www.nature.com/scitable/topicpage/gene-expression-14121669



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 known?

From Genes to Proteins to Functions

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

```
>MySequencesOfInterest
MMQDVSSSPVSPADDSLSNSEEEPDRQQPPSGKRGGRKRRSSRRSAGGGA
GPGGAAGGGVGGGDEPGSPAQGKRGKKSAGCGGGGGAGGGGGSSSGGGSP
QSYEELQTQRVMANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQTLK
LAARYIDFLYQVLQSDELDSKMASCSYVAHERLSYAFSVWRMEGAWSMSA
SH
```

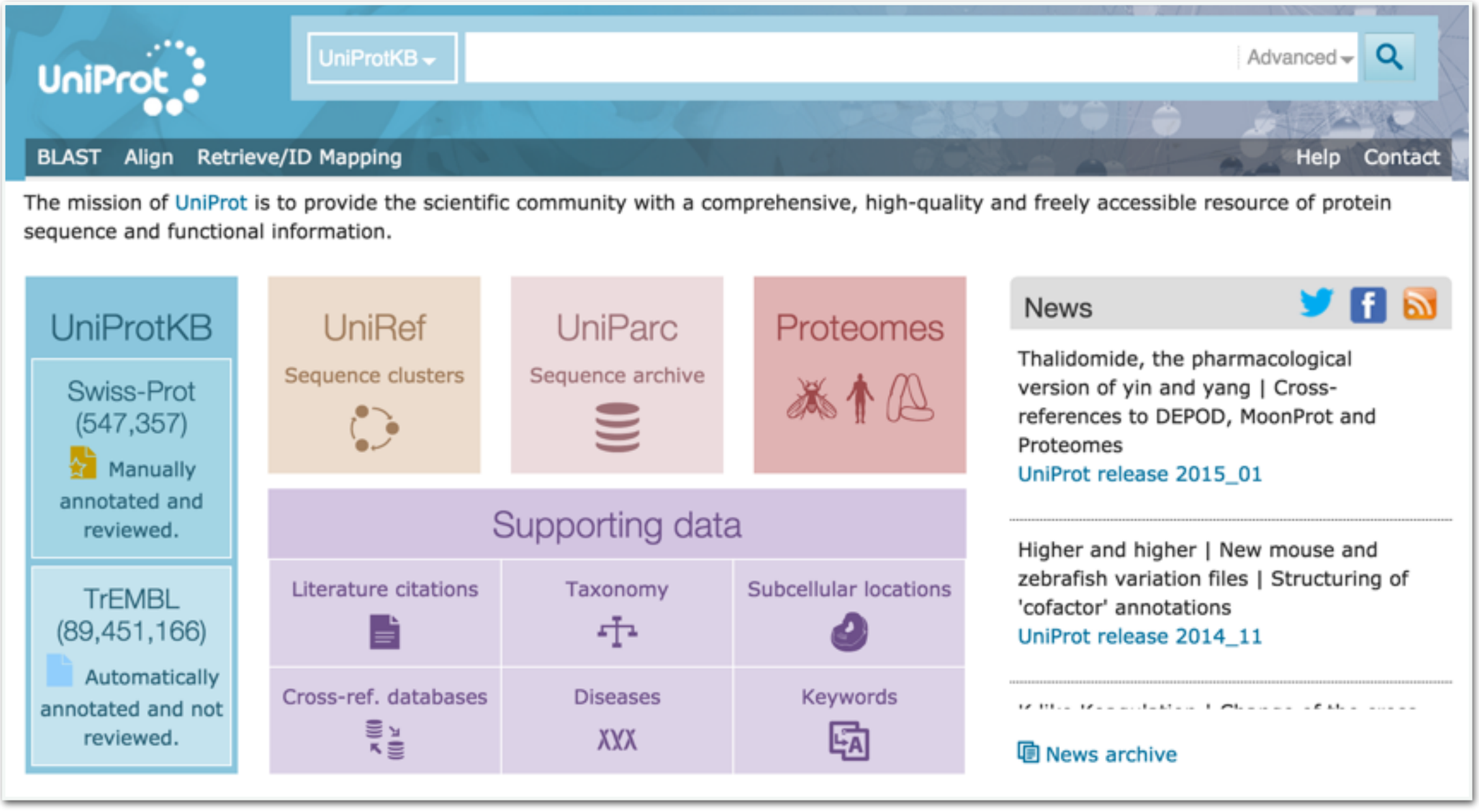
Which tools can we use?

PSI-BLAST
Phmmer, Jackhmmer

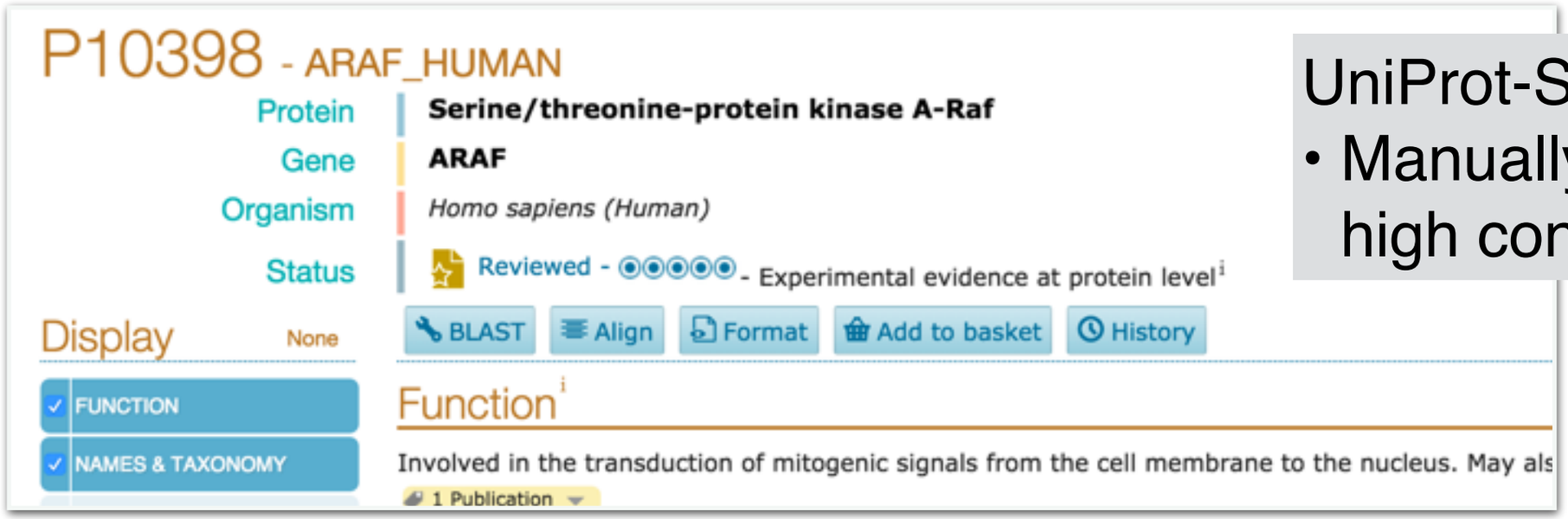
Why perform a homology search?

- 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

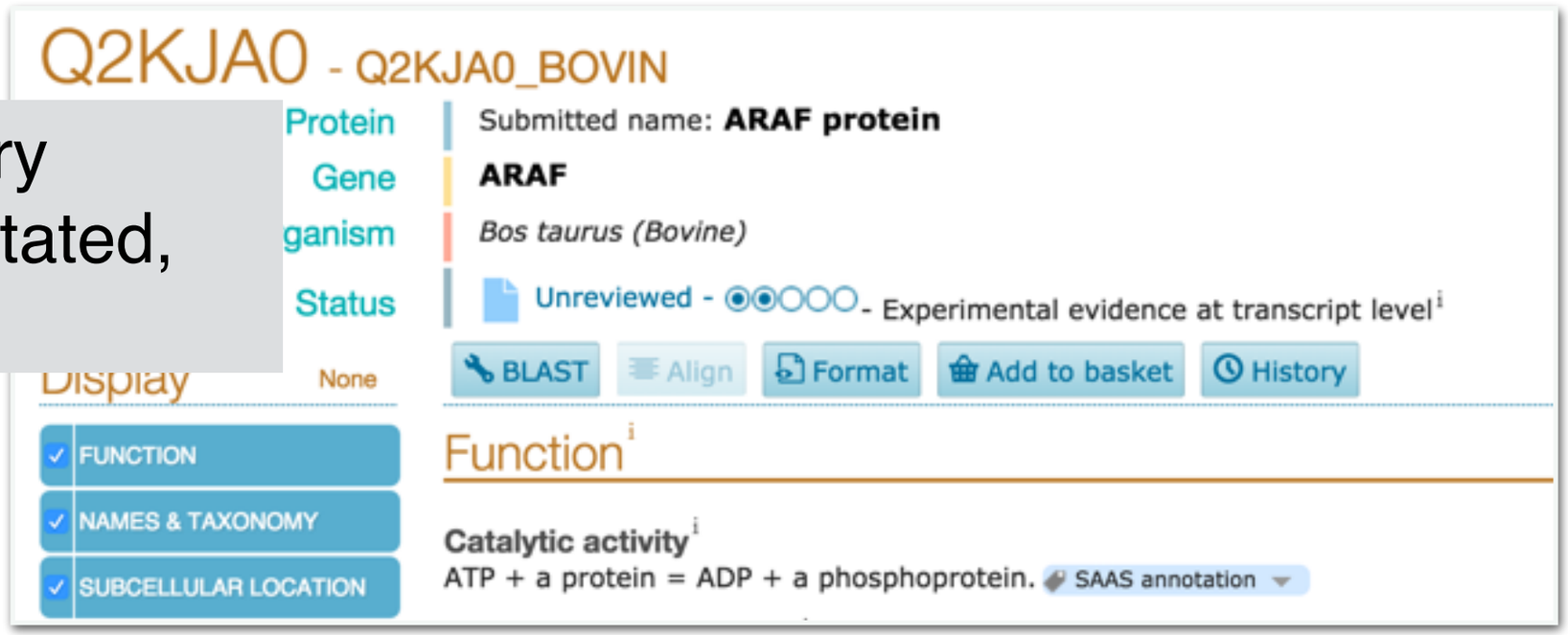
UniProt


UniProt
UniProtKB
Advanced
BLAST
Align
Retrieve/ID Mapping
Help
Contact
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.
UniProtKB
Swiss-Prot (547,357)
Manually annotated and reviewed.
TrEMBL (89,451,166)
Automatically annotated and not reviewed.
UniRef
Sequence clusters
UniParc
Sequence archive
Proteomes
Supporting data
Literature citations
Taxonomy
Subcellular locations
Cross-ref. databases
Diseases
XXX
Keywords
News
Thalidomide, the pharmacological version of yin and yang | Cross-references to DEPOD, MoonProt and Proteomes
UniProt release 2015_01
Higher and higher | New mouse and zebrafish variation files | Structuring of 'cofactor' annotations
UniProt release 2014_11
News archive

UniProt



UniProt-SwissProt entry
- Manually annotated, high confidence.



UniProt-TrEMBL entry
- Automatically annotated, low(er) confidence.

```
>MySequencesOfInterest
MMQDVSSSPVSPADDSLSNSEEEPDRQQPPSGKRGGRKRRSSRRSAGGGA
GPGGAAGGGVGGGDEPGSPAQGKRGKKSAGCGGGGGAGGGGGSSSGGGSP
QSYEELQTQRVMANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQTLK
LAARYIDFLYQVLQSDELDSKMASCSYVAHERLSYAFSVWRMEGAWSMSA
SH
```

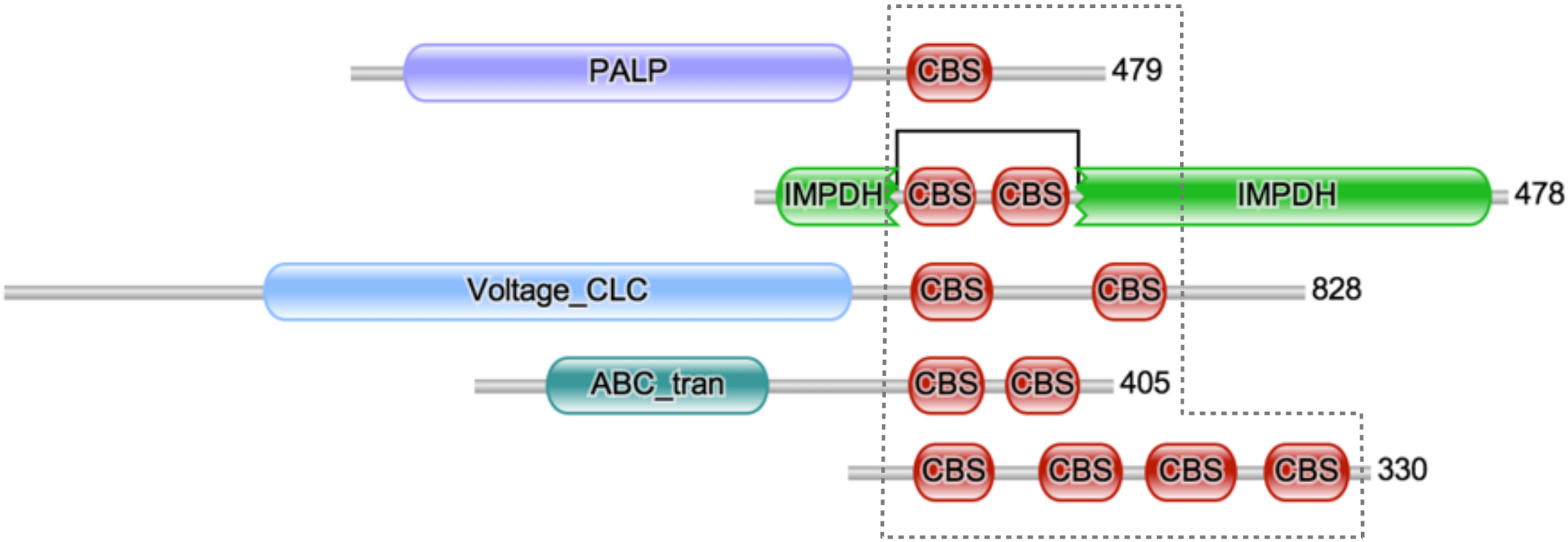
Variance at 156, I to V?

What is it's effect on the protein

Other Homology Searches

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

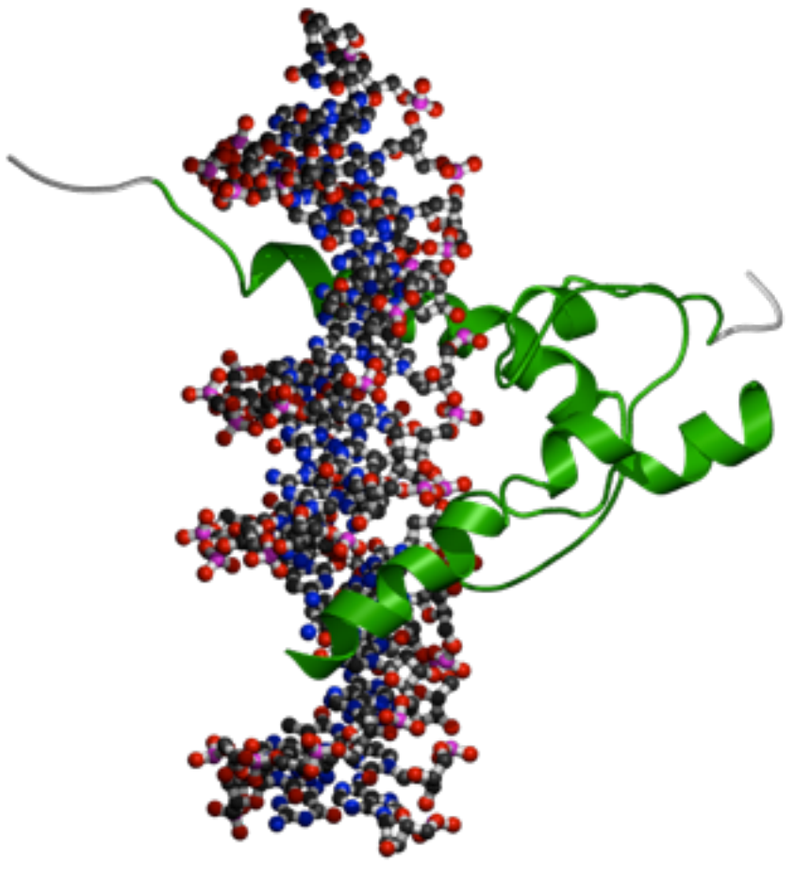
Why Domains?



- Functional diversity
- Conservation
- Tractable drug targets

Homology Searches

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



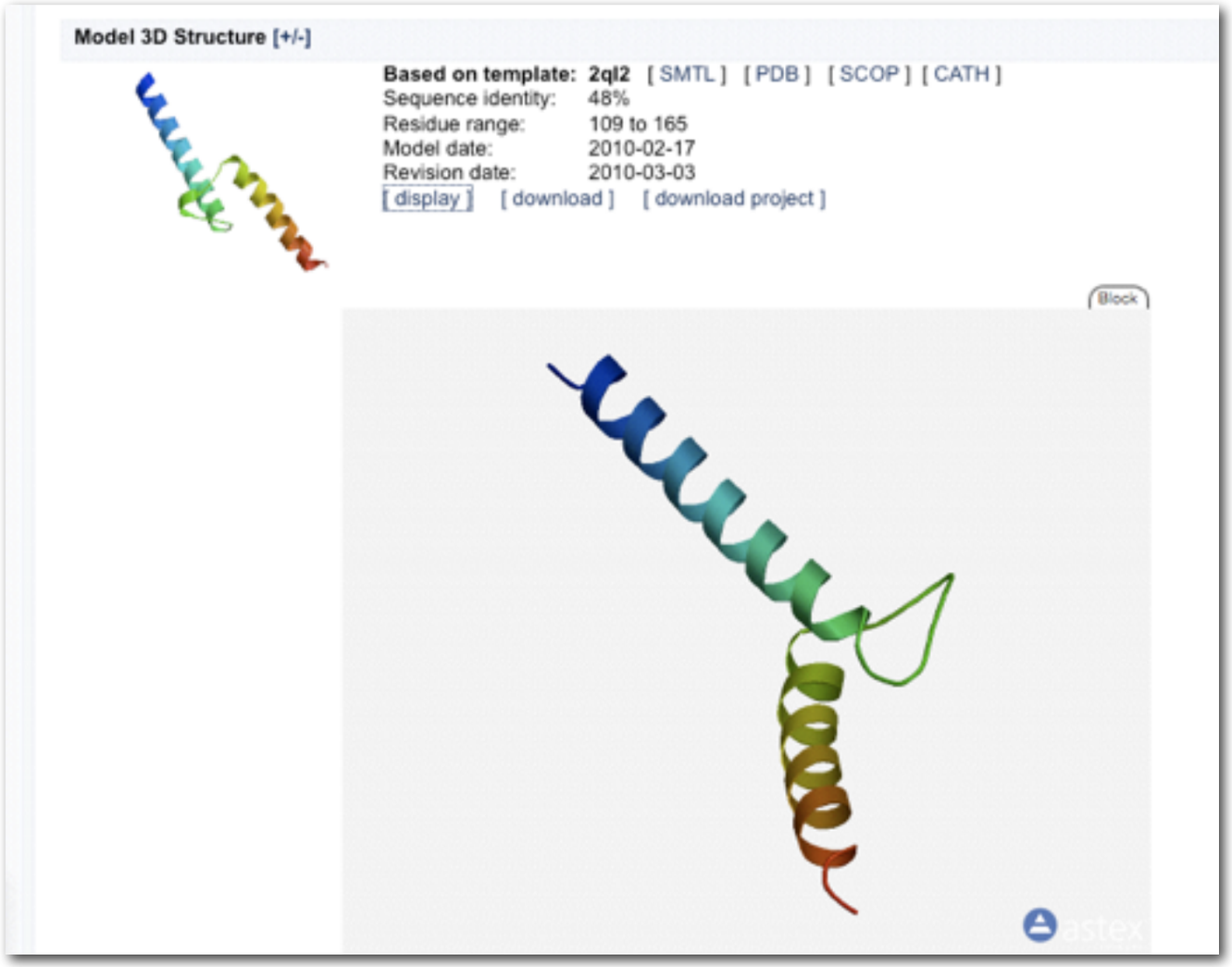
- Structural homologs
  - DNA binding
  - Functions as a dimer

No 3D Structure

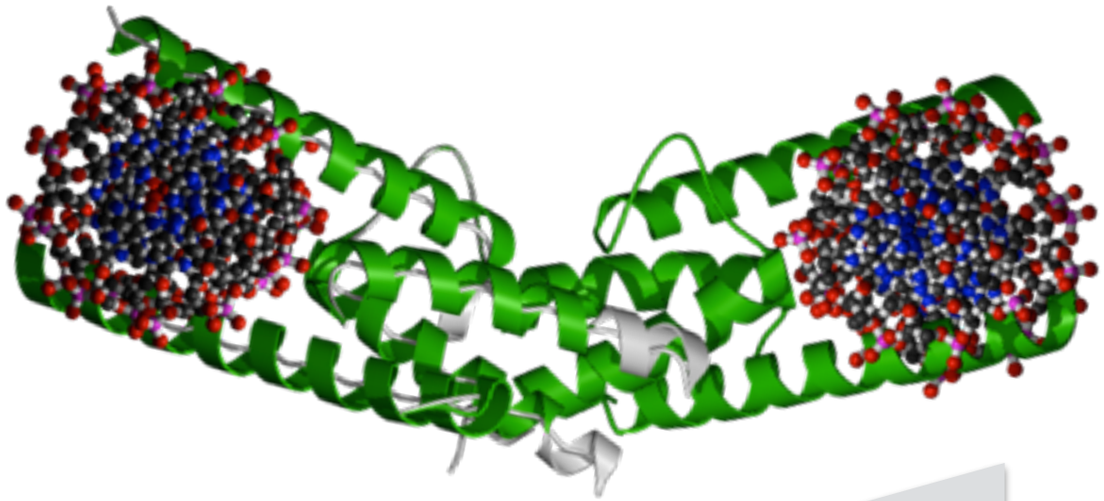
No 3D Structure

```
                 .........*.........*.........*.........*.........*.........*.........
Query      109 qrvmanvrerqrtqslneafaalrkiiptlp.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
```

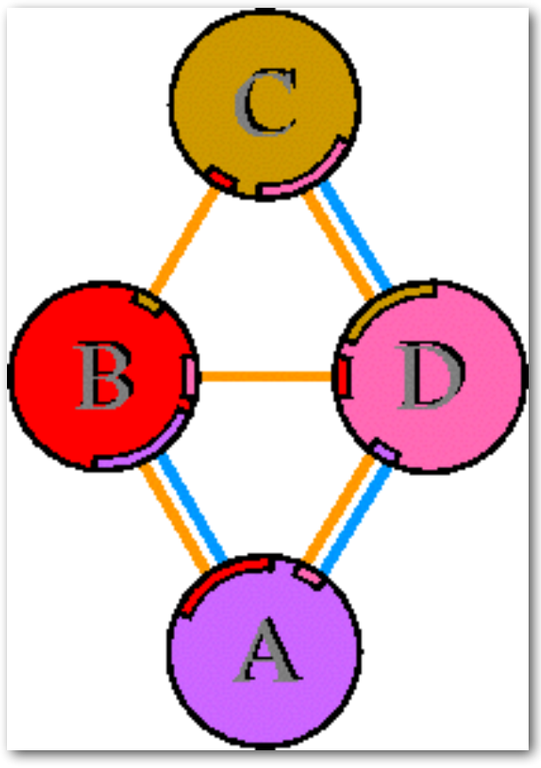
- SWISS-MODEL
  - –Homology Searching and Modelling

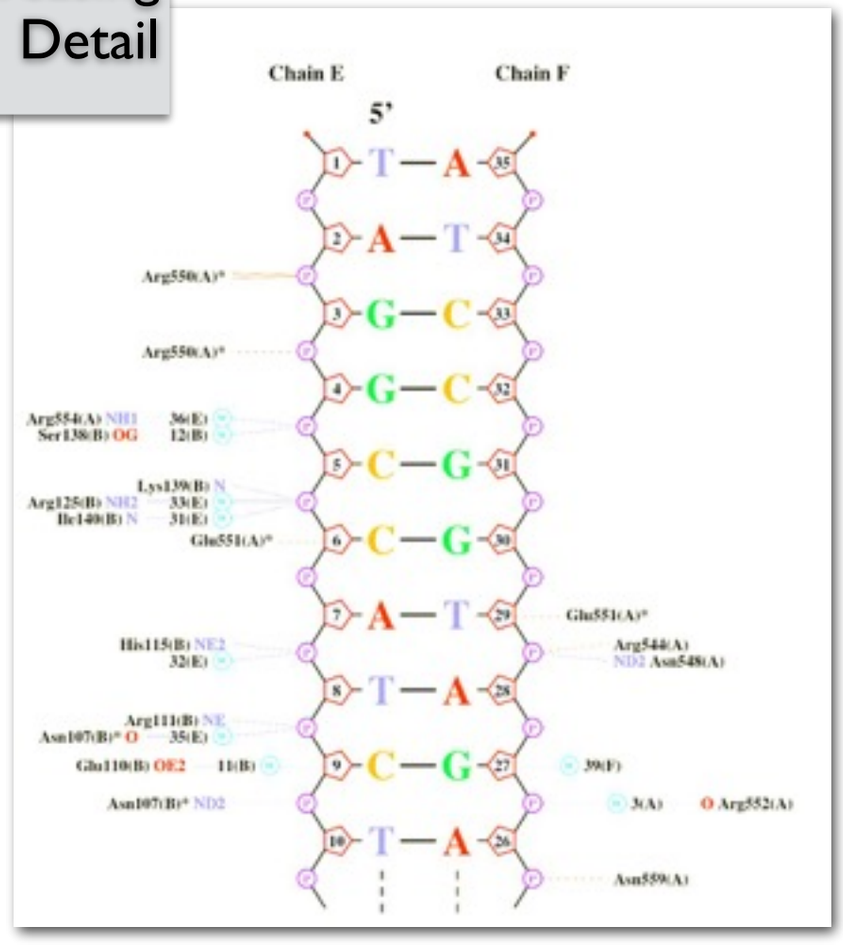


PDBsum - 2QL2

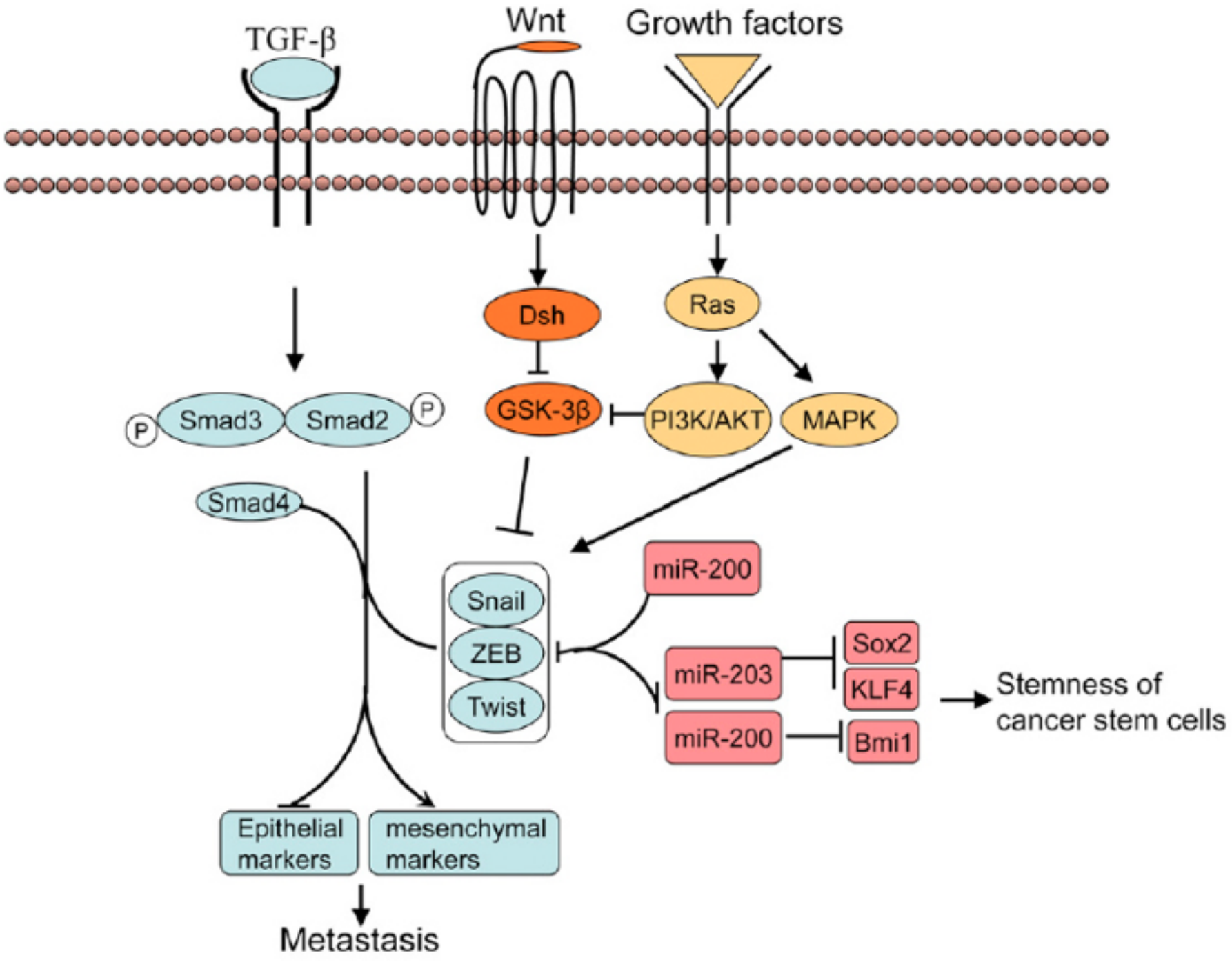


Increasing Detail





Protein Interactions



Xiao and He, JTD 2010 2:154-159

IntAct

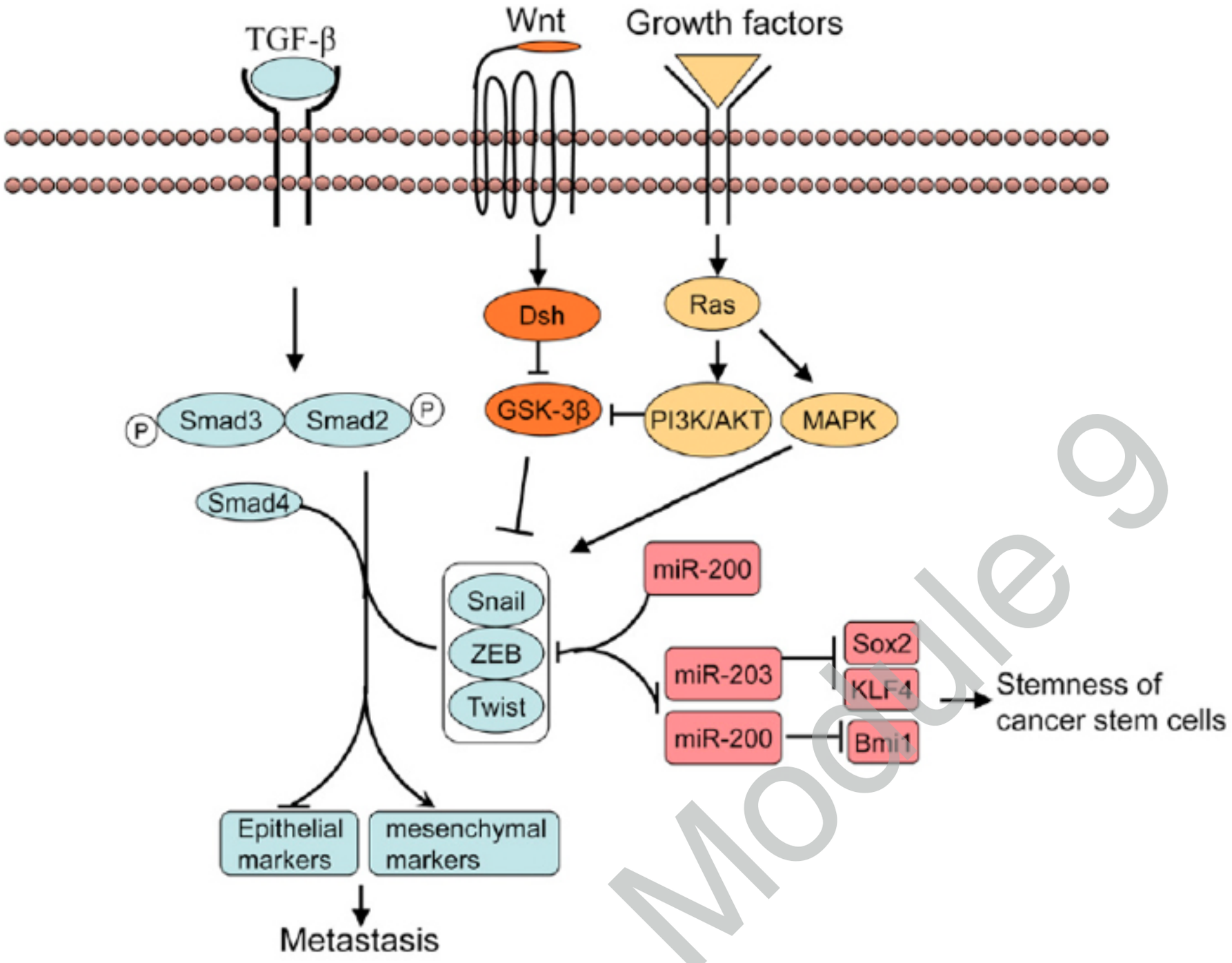
Export to: Select format... Export Change Columns Displayed

	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 PE BA UN	UniProt Dasty2	ets2_human PE PY UN	UniProt Dasty2	H-twist; Class A basic helix-loop-helix protein 38; BHLHA38; [+]	ets2_human	9606	9606	IM-12059 18598946	anti tag coimmunoprecipitation	EBI-1797303 imex:IM-12059-1	
2	PE PY UN		PE BA UN						IM-12059 18598946	anti bait coimmunoprecipitation	EBI-1797350 imex:IM-12059-4	
3	H-twist PE BA UN	UniProt Dasty2	GAP and centrosome-associated protein PE PY UN	UniProt Dasty2	H-twist; Class A basic helix-loop-helix protein 38; BHLHA38; [+]	GAP and centrosome-associated protein; Rab6 GTPase-activating protein GAPCenA; HSPC094; [+]	9606	9606	20195357	display technology	EBI-2689840	

EST - Another transcription factor

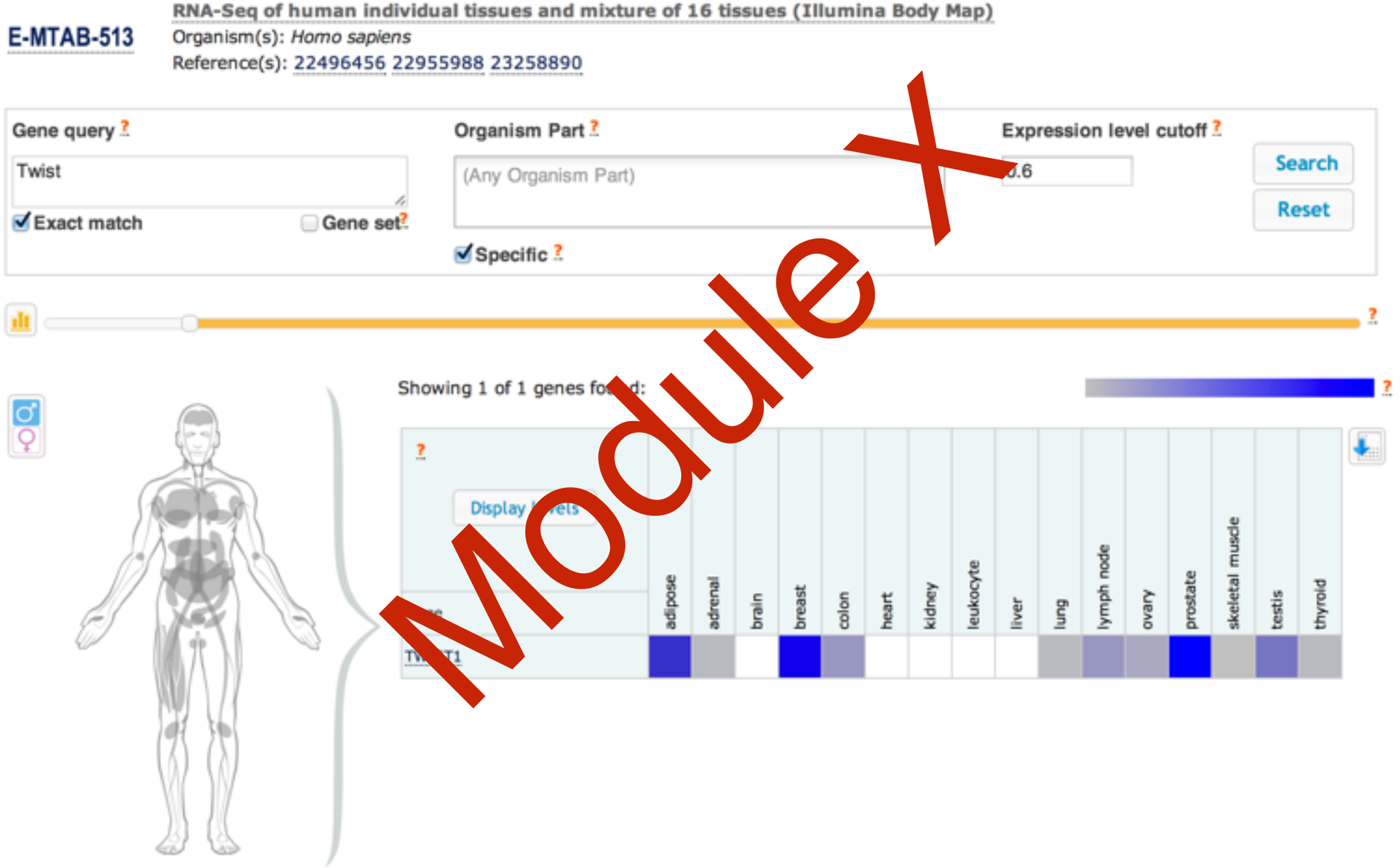
Rab6 - GTPase-activating protein

Protein Interactions



Xiao and He, JTD 2010 2:154-159

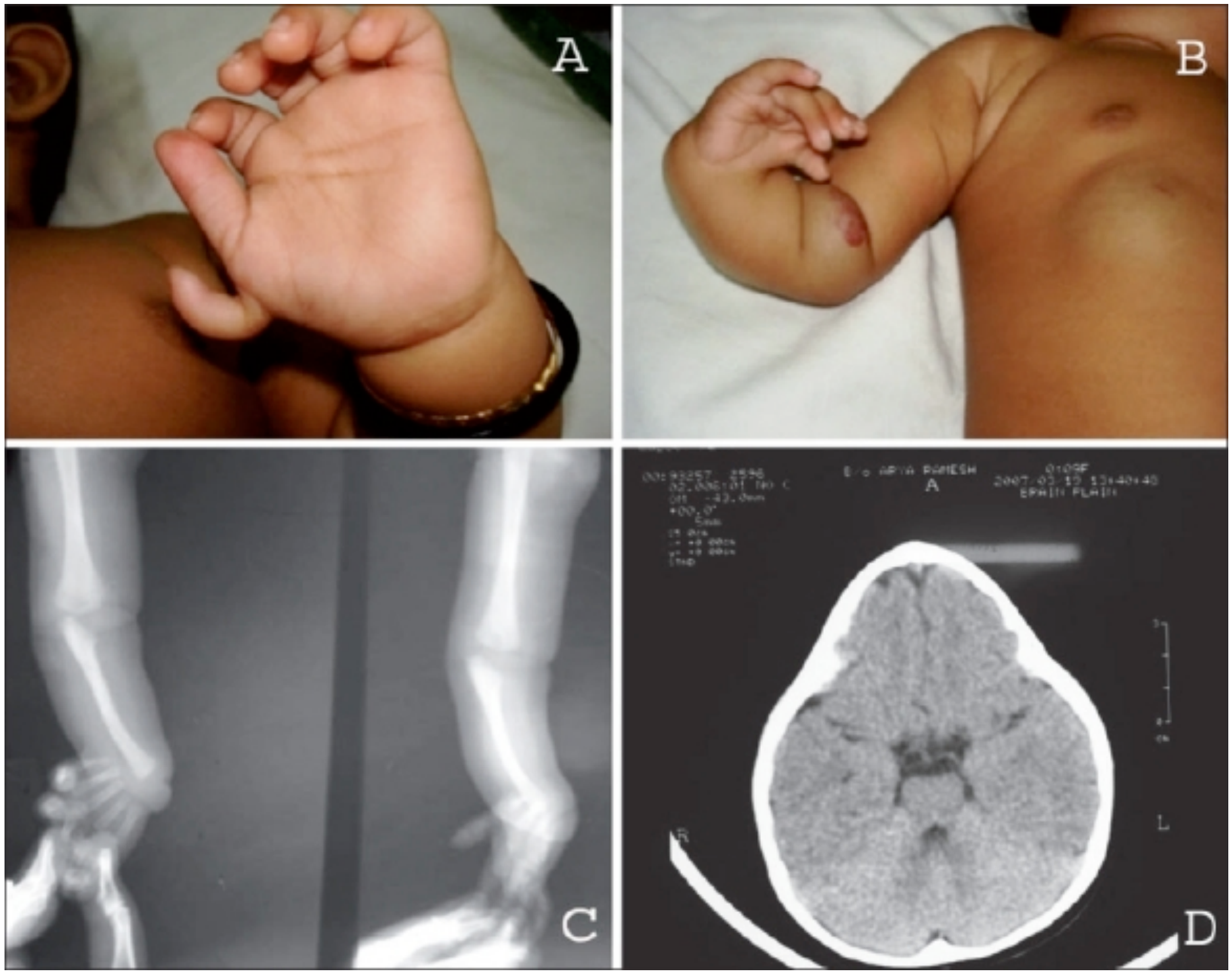
Expression Profiles



So what do we know so far?

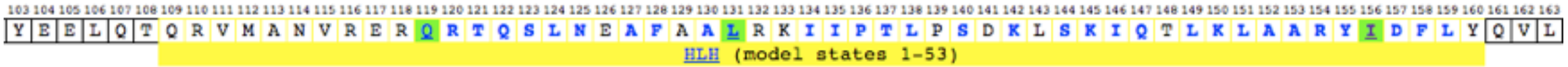
- Contains HLH and disorder
- Does not have a known structure, but homologs do
- Functions as a dimer and binds DNA and other proteins
- Which residues bind DNA and each other
- Regulators of miRNAs, with are themselves gene regulators
- Disease/Tissue expression patterns
- Signalling pathway

Baller-Gerold Syndrome

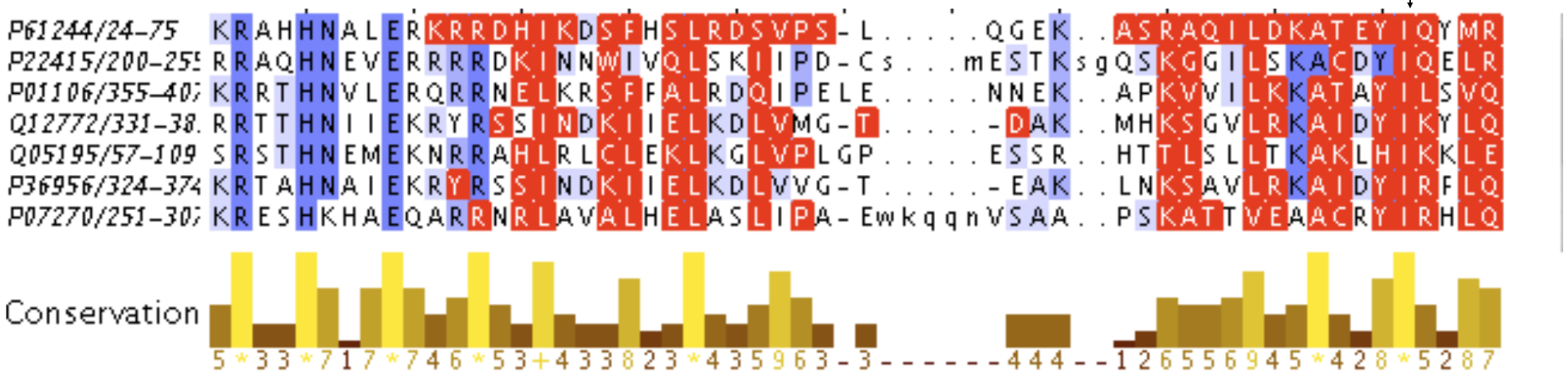


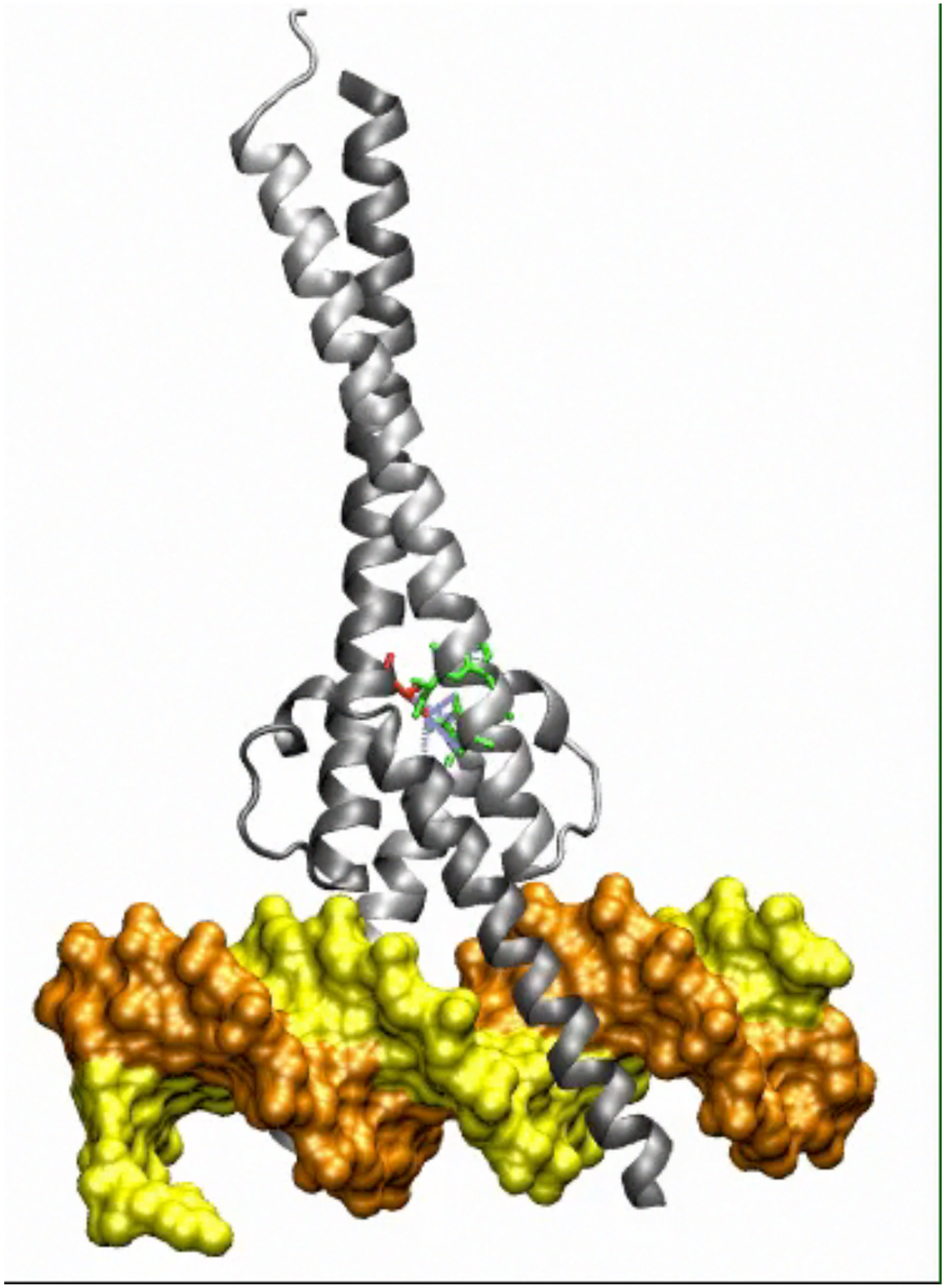
Baller-Gerold Syndrome Mutations

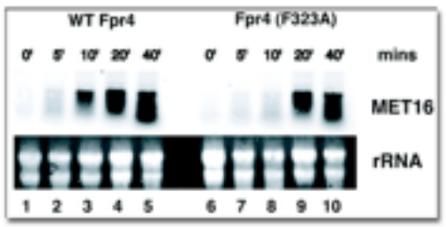
OMIM



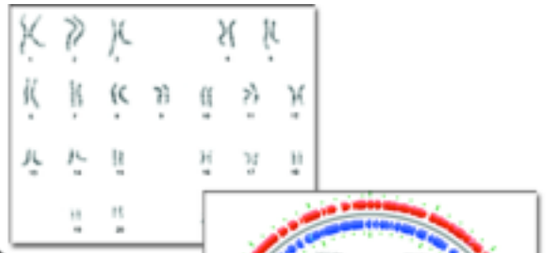
Pfam

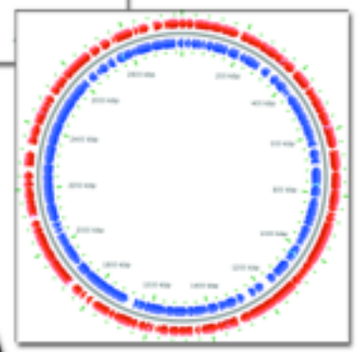




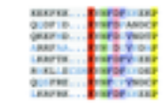


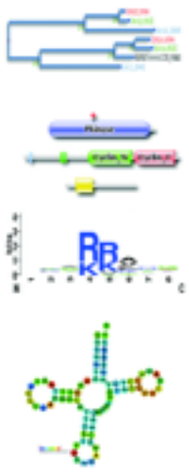
SEQUENCE
10-100 Million





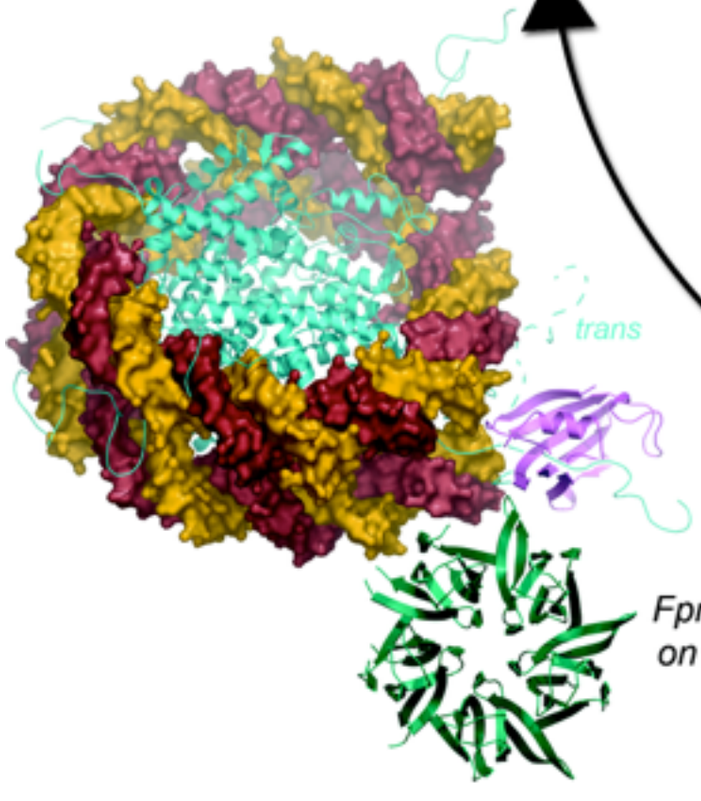
DATABASES





CLASSIFICATION
10 Thousand Families

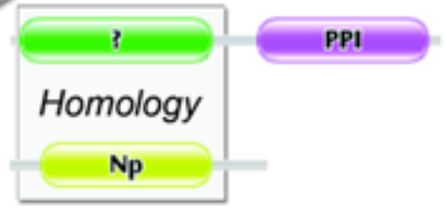
FUNCTION

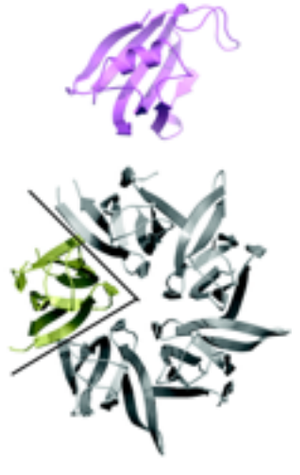


Fpr4 model based on nucleoplasmin

Model of histone modification by Fpr4

HYPOTHESIS





N-terminal domain of Fpr4 related to nucleoplasmin

Worked Examples