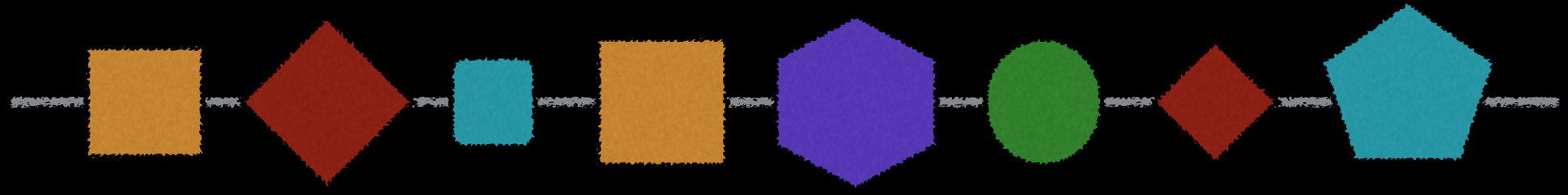
Domain & motif analysis & GO



Nicole Libert & Michaela Cline

I. Overview of motifs and domains

II. Functions of domains

III. Importance of domains

IV. Identifying domains

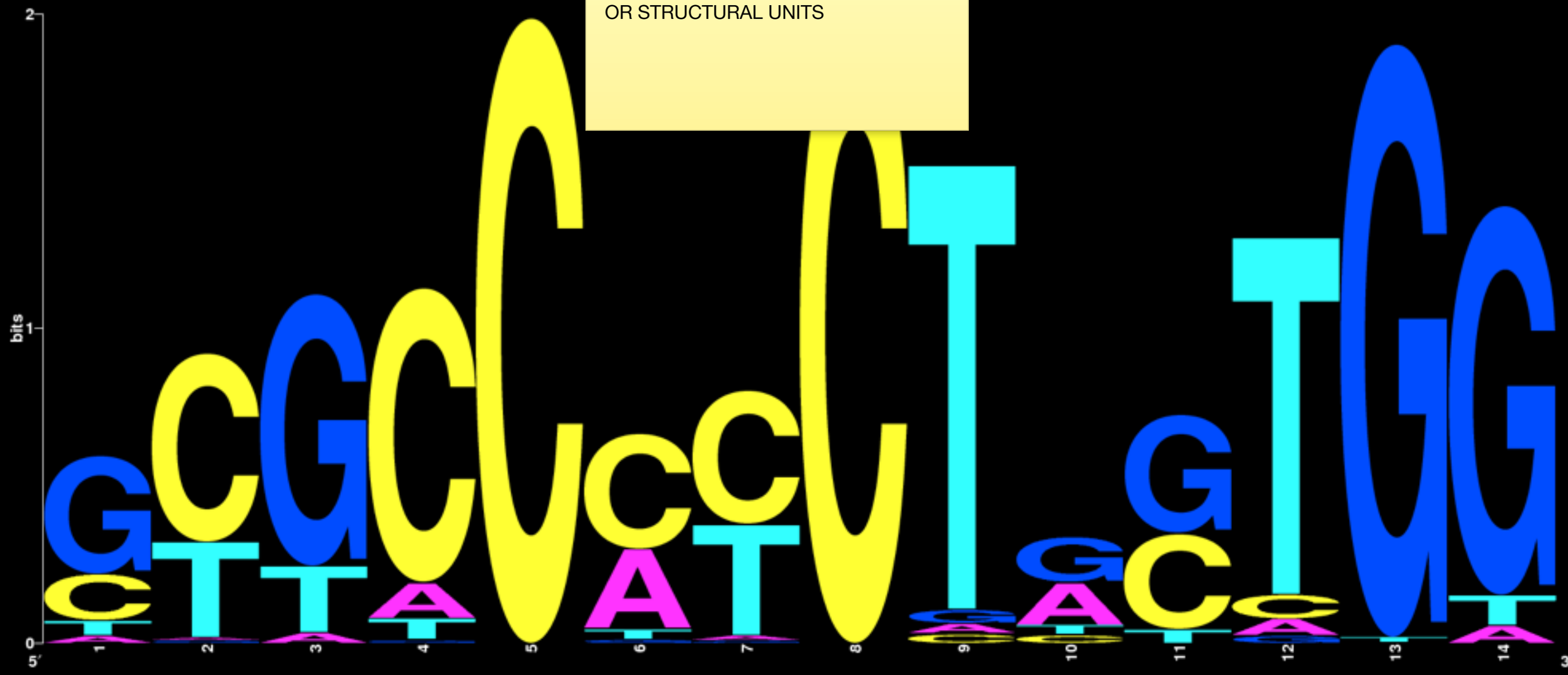
    a. Nicole's suggestions for Thursday's lab

V. Overview of GO

What are motifs?

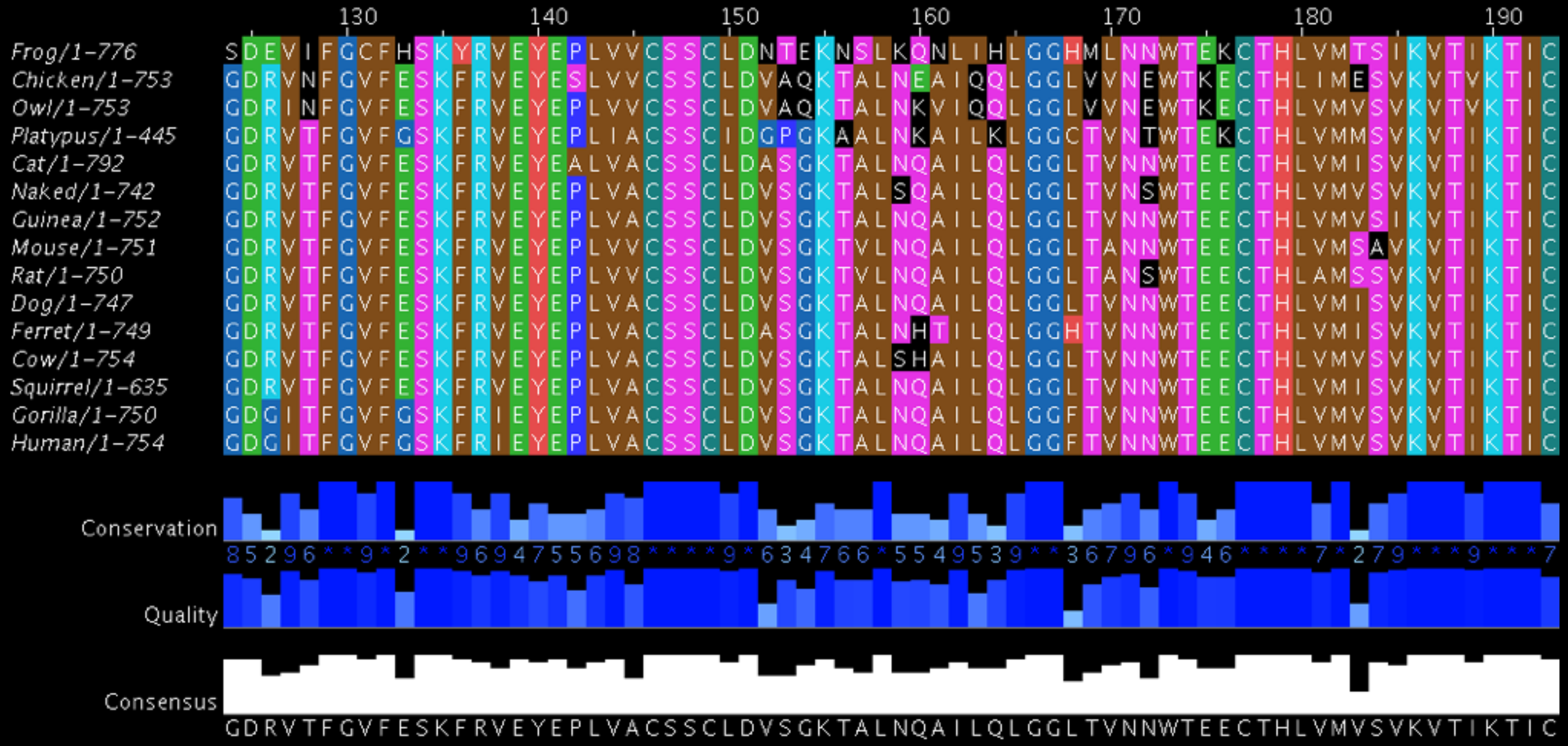
Short, conserved regions that are presumed to have a biological function

REGIONS CAN BE USED TO DESCRIBE MOLECULAR SEQUENCES OR STRUCTURAL UNITS



http://cmotifs.tchlab.org/comparison/GIMMEMOTIF/Ctcf.png

Multiple sequence alignments



Motifs can be seen within domains

Clustal Omega

What are domains?

Adaptors

Enzymes

Docking proteins

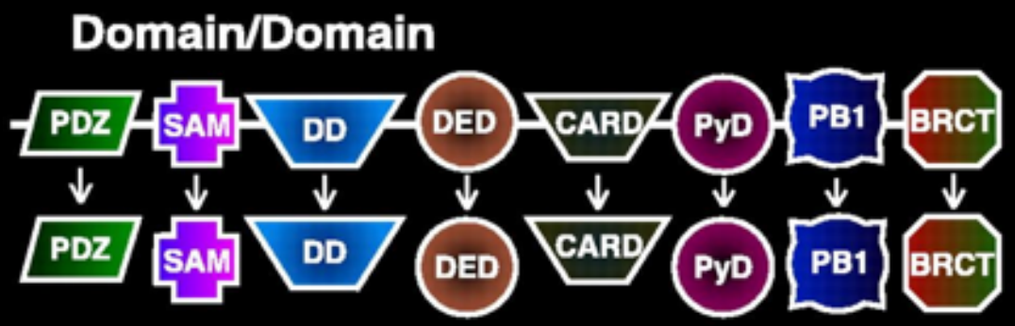
Inhibitory factors

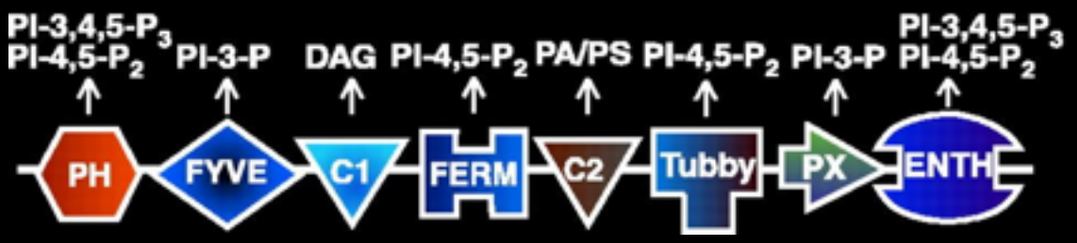
Domains are conserved structural units of a protein

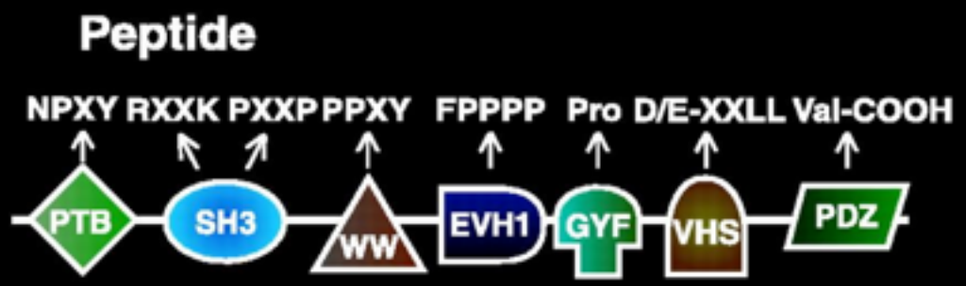
Functional and/or structural

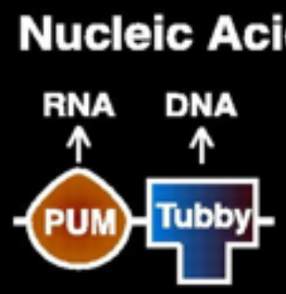
http://www.cellsignal.com/common/content/content.jsp?id=domains-sh2

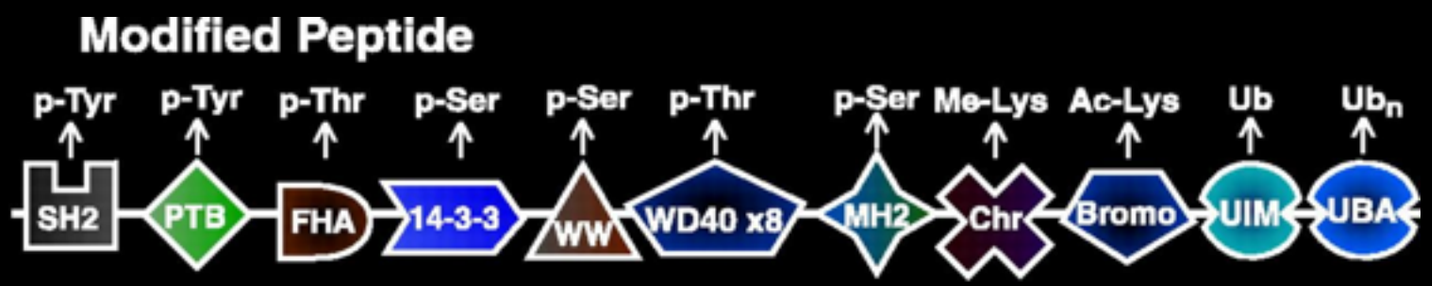
What do domains do?











Target proteins

Serve as signaling complexes

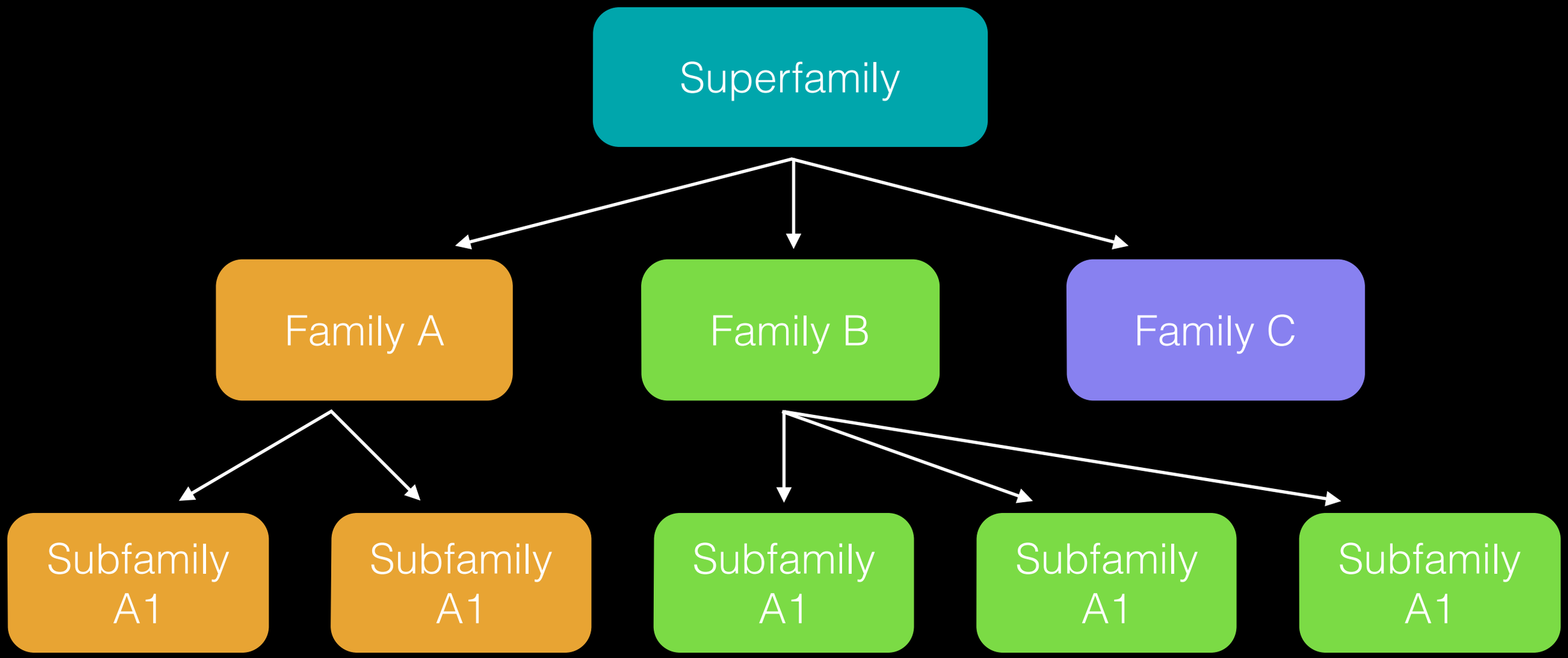
Control functions of enzymes

http://www.sciencemag.org/content/300/5618/445/F1.large.jpg

What do specific domains do?

http://pawsonlab.mshri.on.ca/index.php?option=com_content&task=view&id=30&Itemid=63

Why do domains matter?



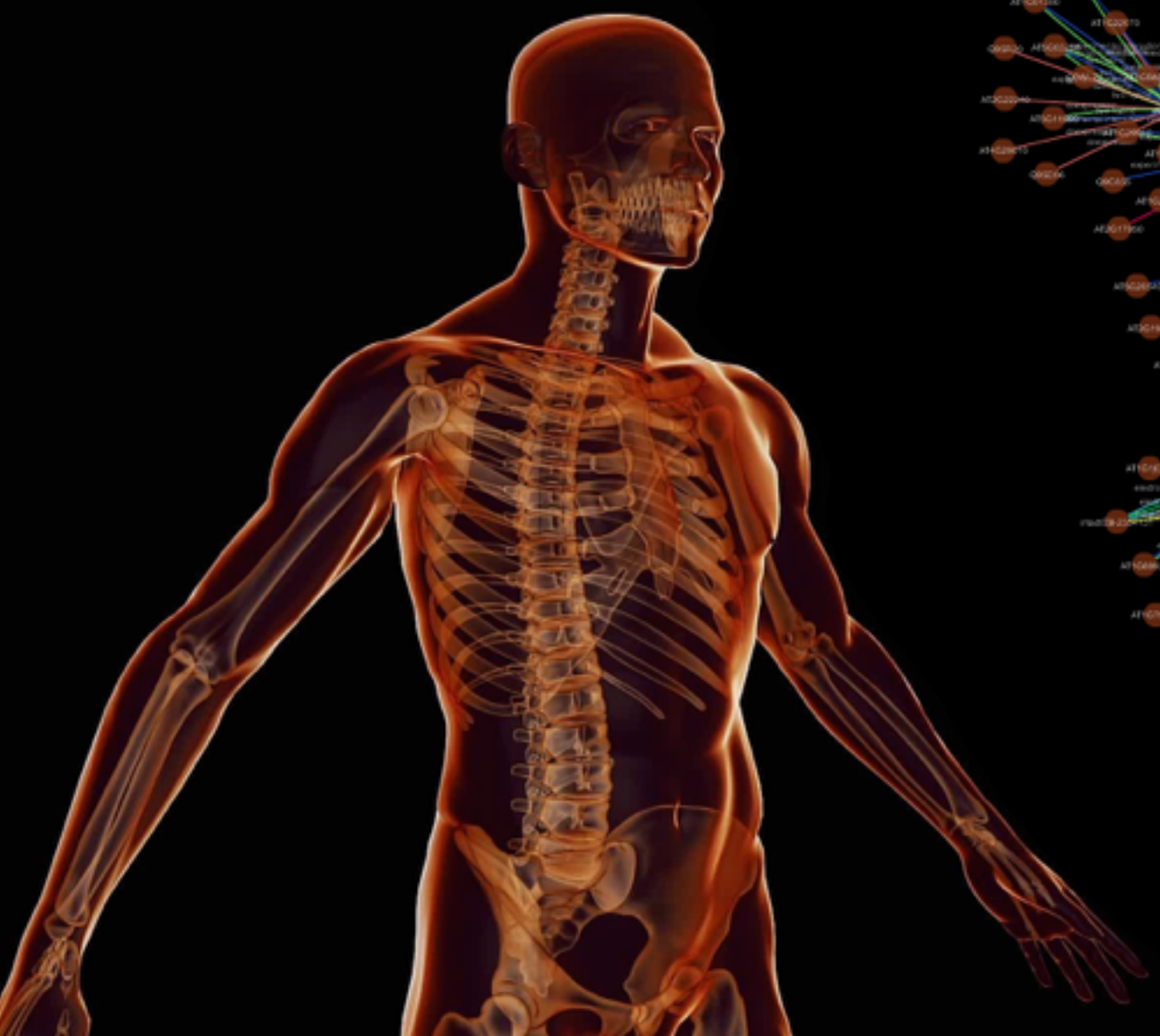
Classifications are constructed using domains as building blocks

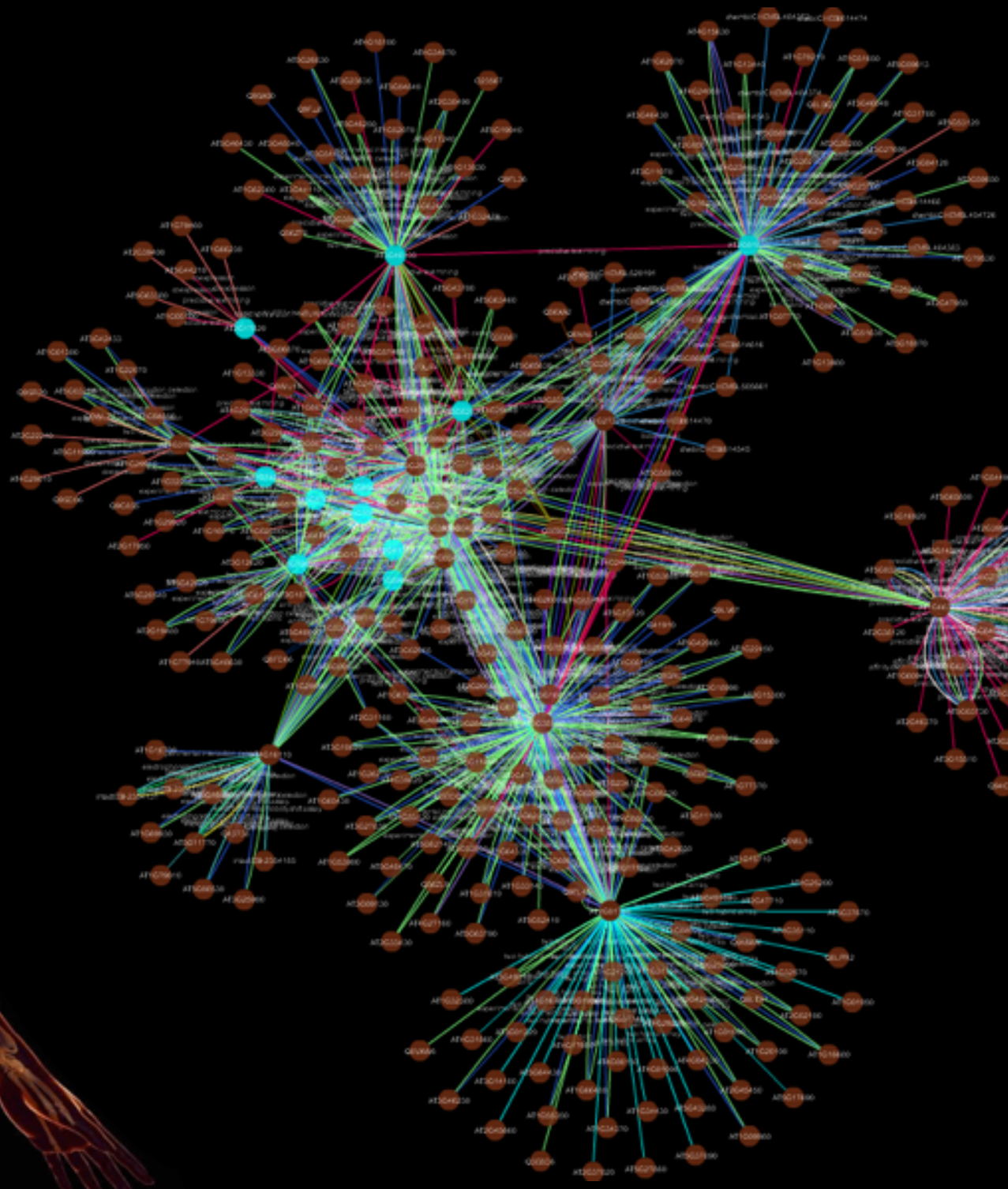
Why do domains matter?

Protein-protein interactions

Complexity of organisms

Disease??





How do we find domains?

Computers

Protein Signatures

Patterns

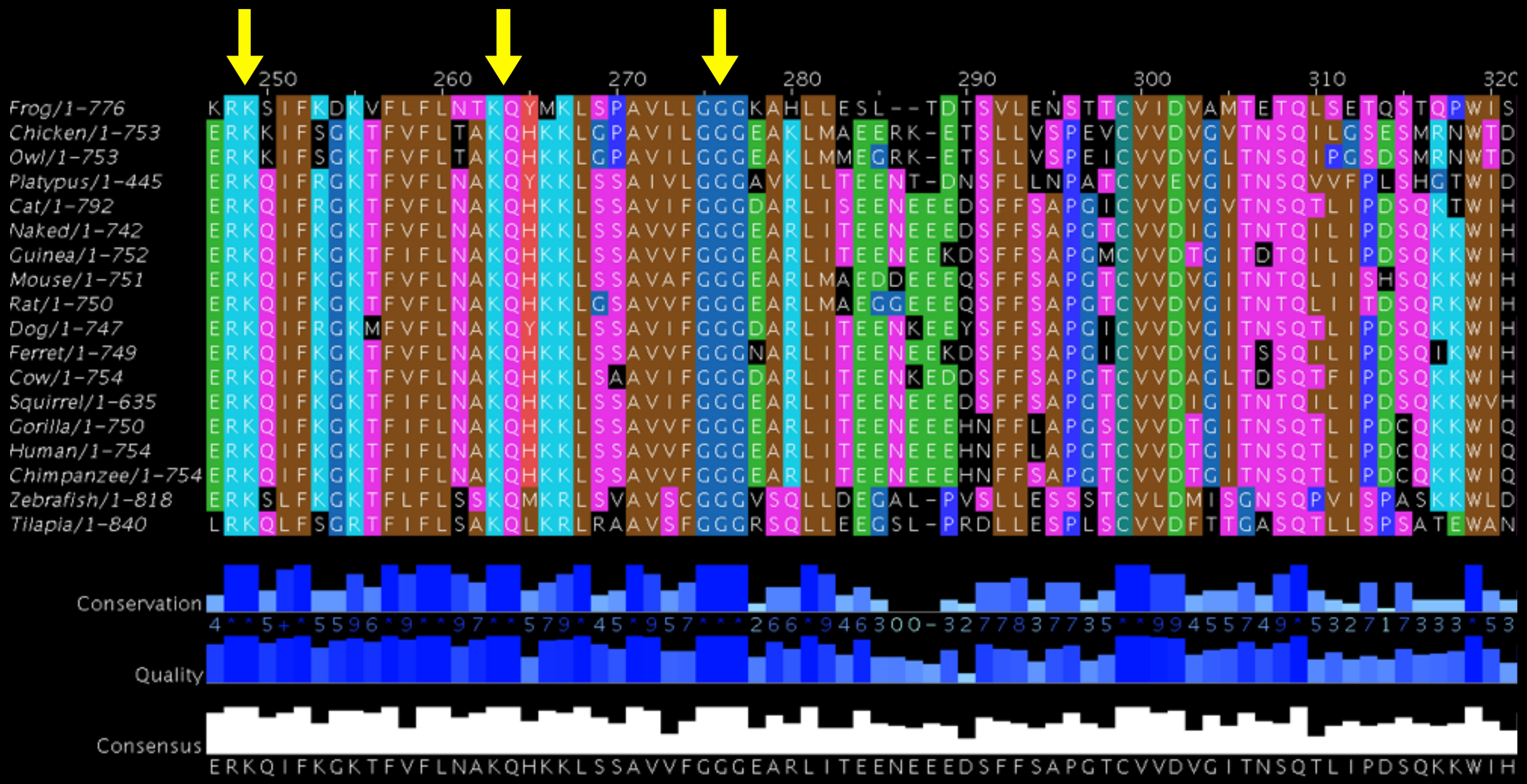
Fingerprints

Profiles

Hidden Markov models (HMMs)

http://www.wallpaper77.com/upload/DesktopWallpapers/cache/Apple-Macbook-Pro-apple-wallpapers-computer-wallpapers-1920x1200.jpg

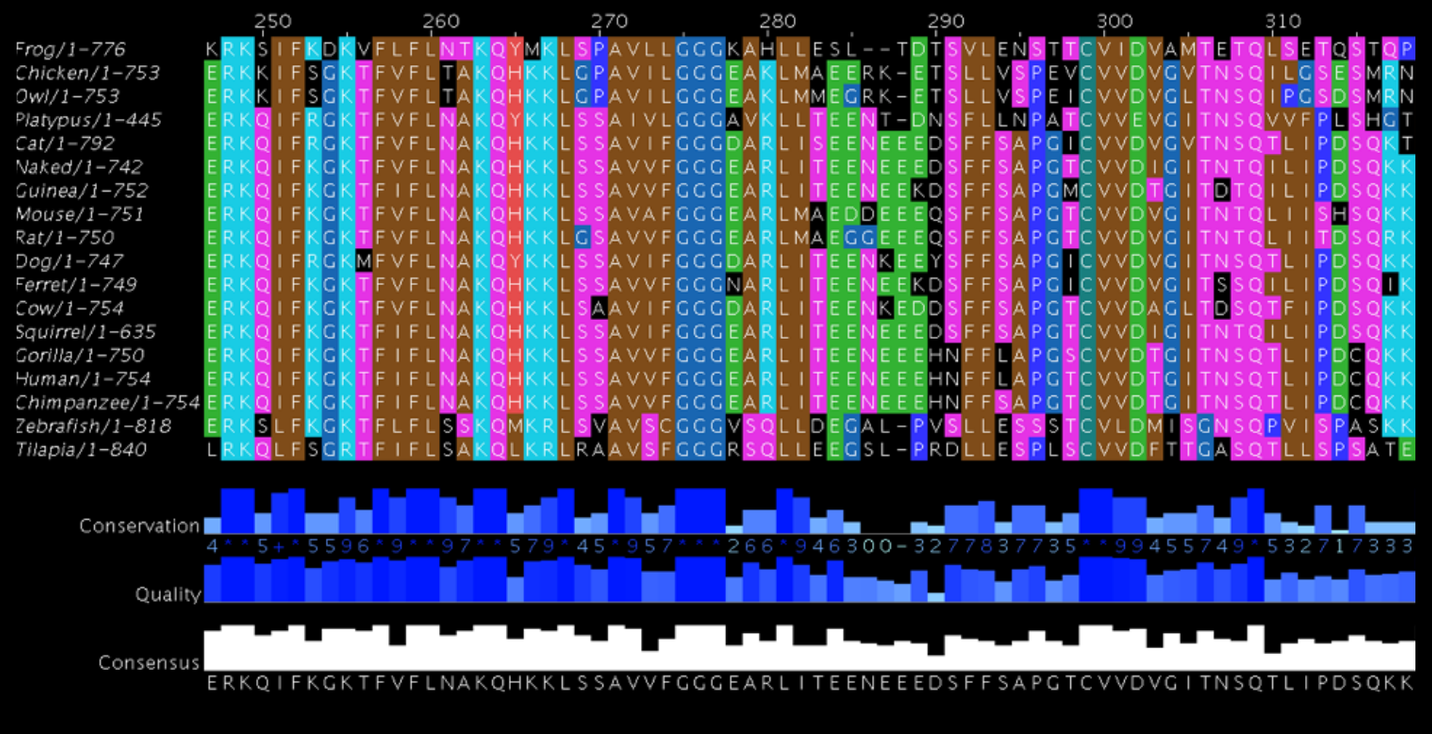
How do we find domains?



The commonality: Multiple sequence alignments

Clustal Omega

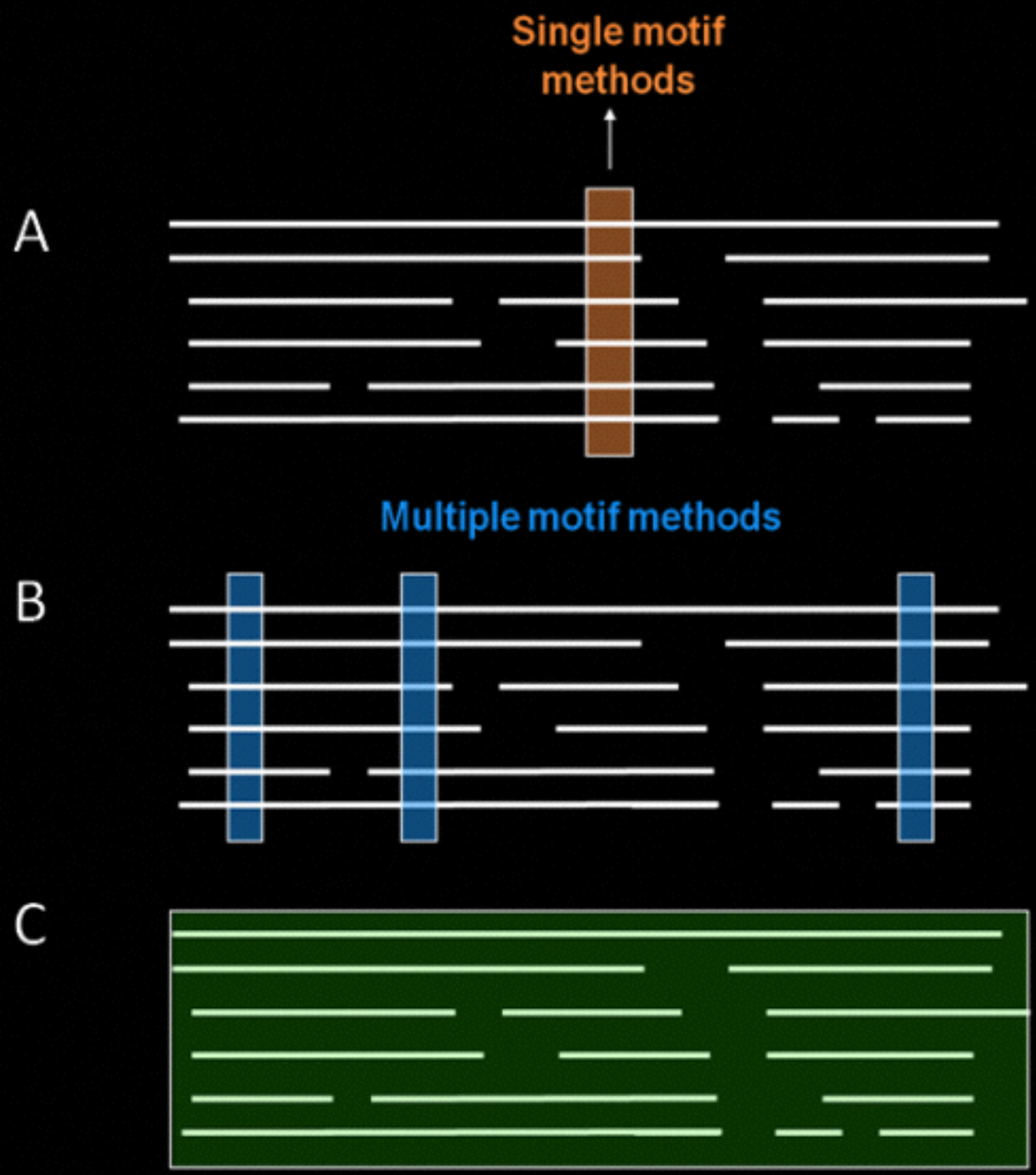
How do we find domains?



http://www.ebi.ac.uk/sites/ebi.ac.uk/files/documents/uniprot_rgb_320x146.png

http://upload.wikimedia.org/wikipedia/commons/thumb/0/07/US-NLM-NCBI-Logo.svg/2000px-US-NLM-NCBI-Logo.svg.png

How do we find domains?



Patterns

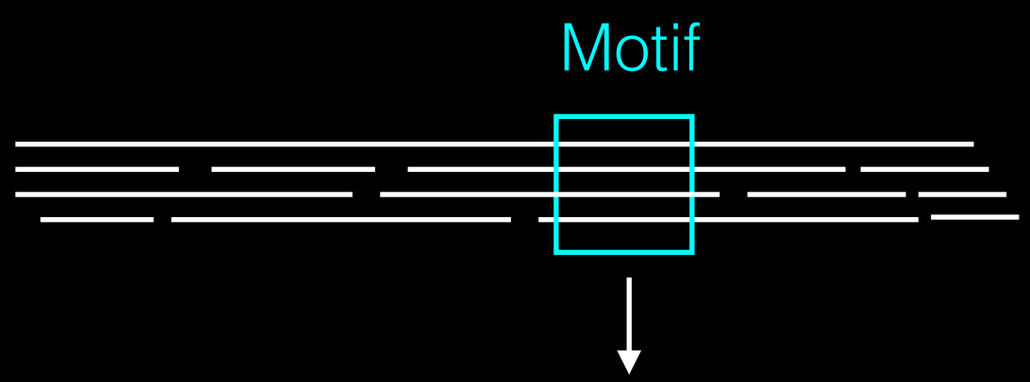
Fingerprints

Profiles
HMMs

http://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/user/84/documents/figure_12.png

Finding domains using patterns

Sequence alignment



Extract pattern sequences

Build regular expression

[AC]-x-V-x(4)-{ED}

http://prosite.expasy.org/images/prosite/prosite.gif
http://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/user/84/documents/figure13.png

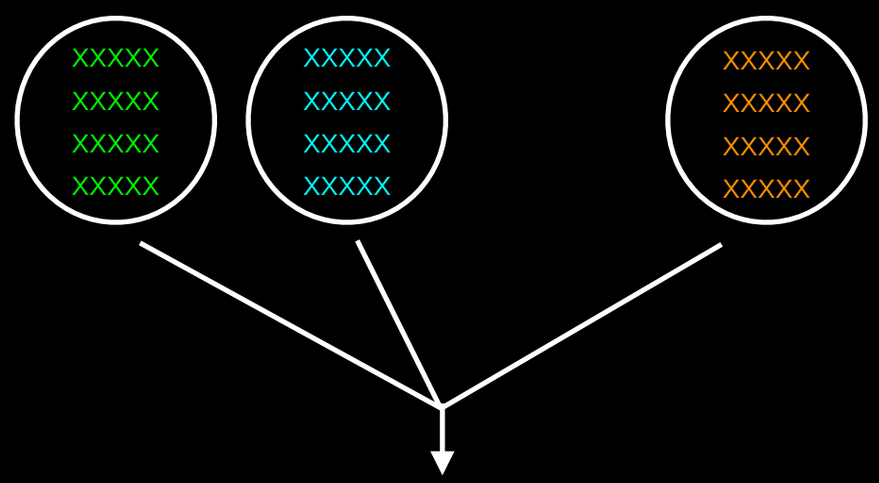
Finding domains using fingerprints

PRINTS

Sequence alignment

Define motifs

Profiles



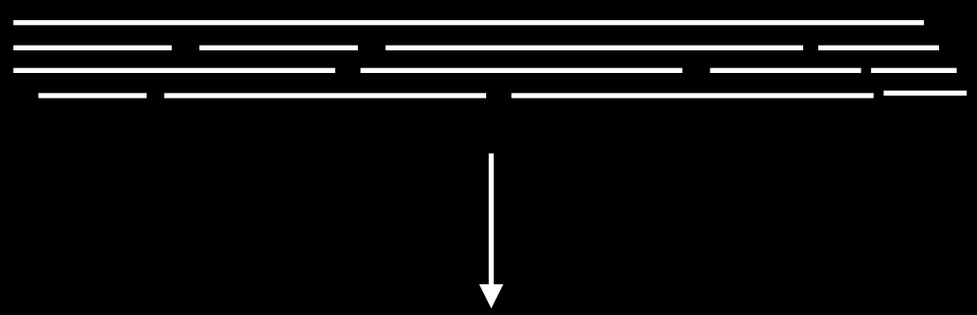
Fingerprint signatures

http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/index.php

http://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/imagecache/screenshot-carousel-img/figure15.png

Finding domains using profiles

Sequence alignment



Residue frequency at each position

```
Sequence 1:  F  K  L  L  S  H  C  L  L  V
Sequence 2:  F  K  A  F  G  Q  T  M  F  Q
Sequence 3:  Y  P  I  V  G  Q  E  L  L  G
Sequence 4:  F  P  V  V  K  E  A  I  L  K
Sequence 5:  F  K  V  L  A  A  V  I  A  D
Sequence 6:  L  E  F  I  S  E  C  I  I  Q
Sequence 7:  F  K  L  L  G  N  V  L  V  C
```

Position specific scoring matrix (PSSM)

A	-18	-10	-1	-8	8	-3	3	-10	-2	-8
C	-22	-33	-10	-19	-22	-26	22	-21	-19	-7
D	-35	0	-32	-33	-7	6	-17	-34	-31	0
E	-27	15	-25	-26	-9	23	-9	-24	-23	-1
F	60	-30	12	14	-26	-29	-15	4	12	-29
G	-30	-20	-28	-32	26	-14	-26	-33	-27	-5
H	-13	-12	-25	-22	-16	14	-22	-22	-23	-10
I	3	-27	21	25	-29	-23	-8	33	19	-23
K	-26	25	-25	-27	-6	4	-15	-27	-26	0
L	14	-20	10	27	-27	-20	-9	33	26	-21
M	3	-15	10	14	-17	-10	-9	24	12	-11
N	-22	-6	-24	-27	1	8	-15	-24	-24	-4
P	-30	24	-26	-29	14	-10	-22	-24	-26	-48
Q	-33	5	-25	-36	-9	24	-16	-17	-33	7
R	-18	6	-22	-22	-10	0	-18	-23	-22	-4
S	-22	-8	-16	-21	11	2	-1	-24	-19	-4
T	-10	-10	-6	-7	-5	-6	2	-10	-7	-11
V	0	-25	22	25	-19	-26	6	-19	16	-16
W	0	-25	-10	-19	-25	-27	-34	-20	-17	-26
Y	34	-10	-1	1	-23	-12	-19	0	0	-18

http://hamap.expasy.org/images/hamap/HAMAP_Red_Logo.png

http://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/user/84/documents/figure14.png

Finding domains using HMMs

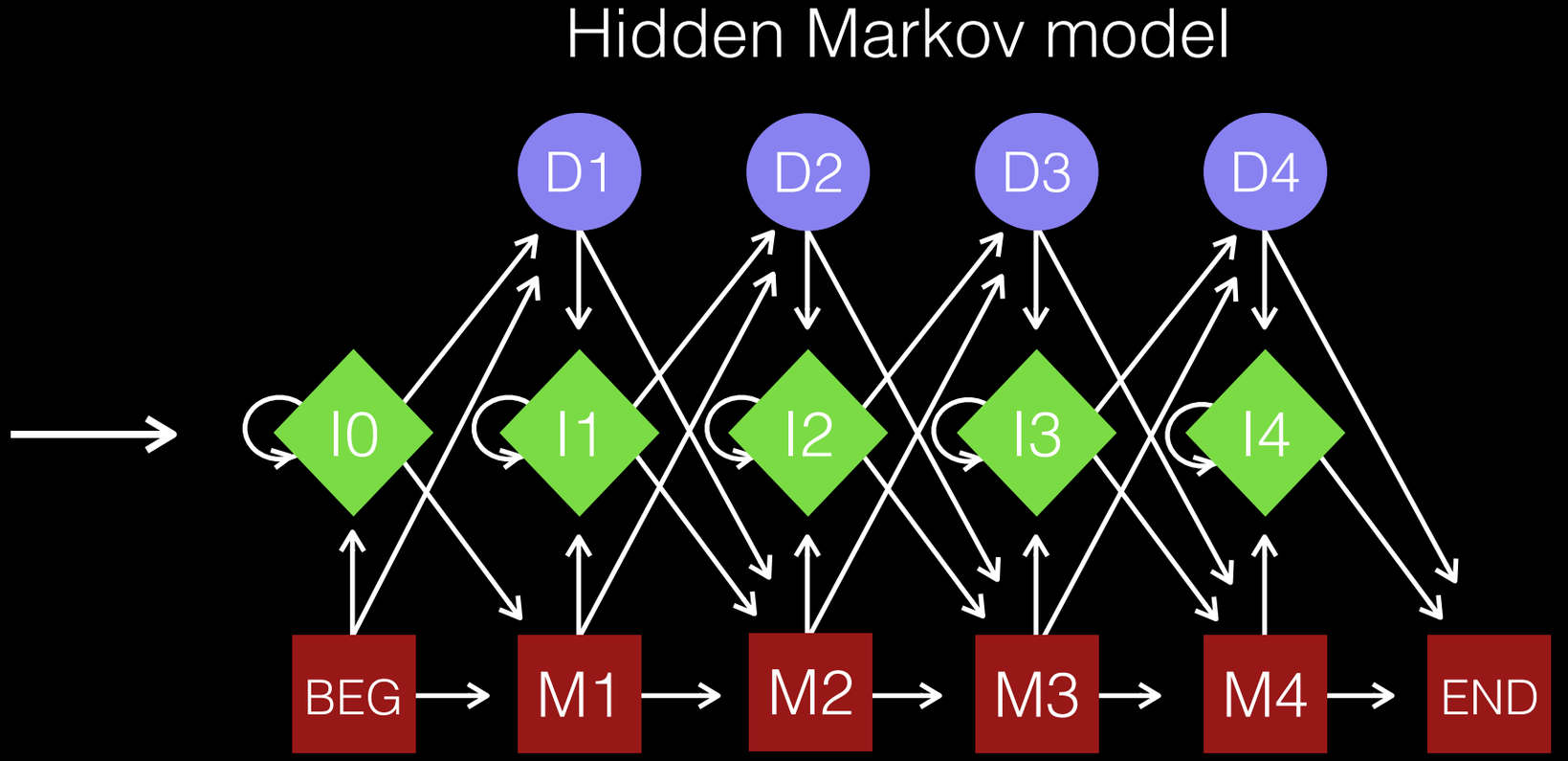
Pfam

SMART

tigr fams
tigr protein families

Sequence alignment

```
N • F L S
N • F L S
N K Y L T
Q • W — T
```



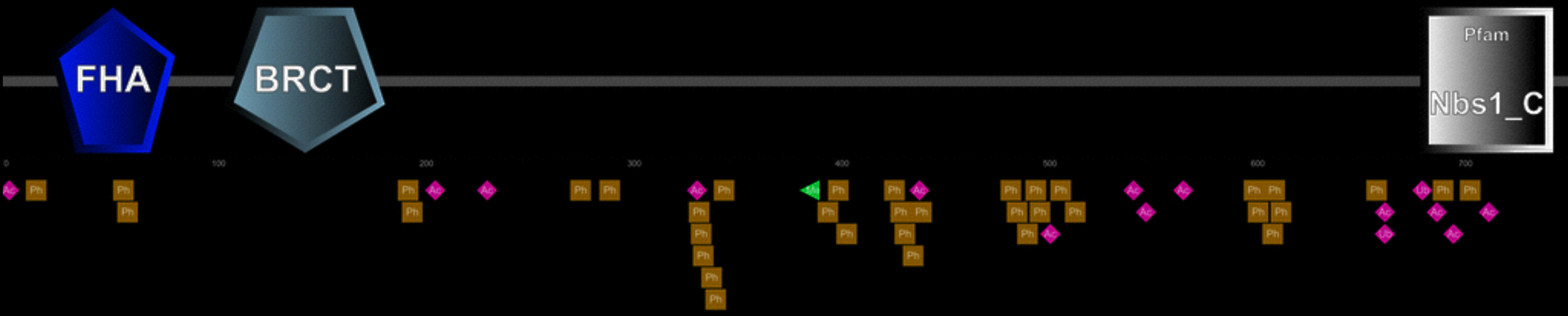
http://blast.jcvi.org/web-hmm/

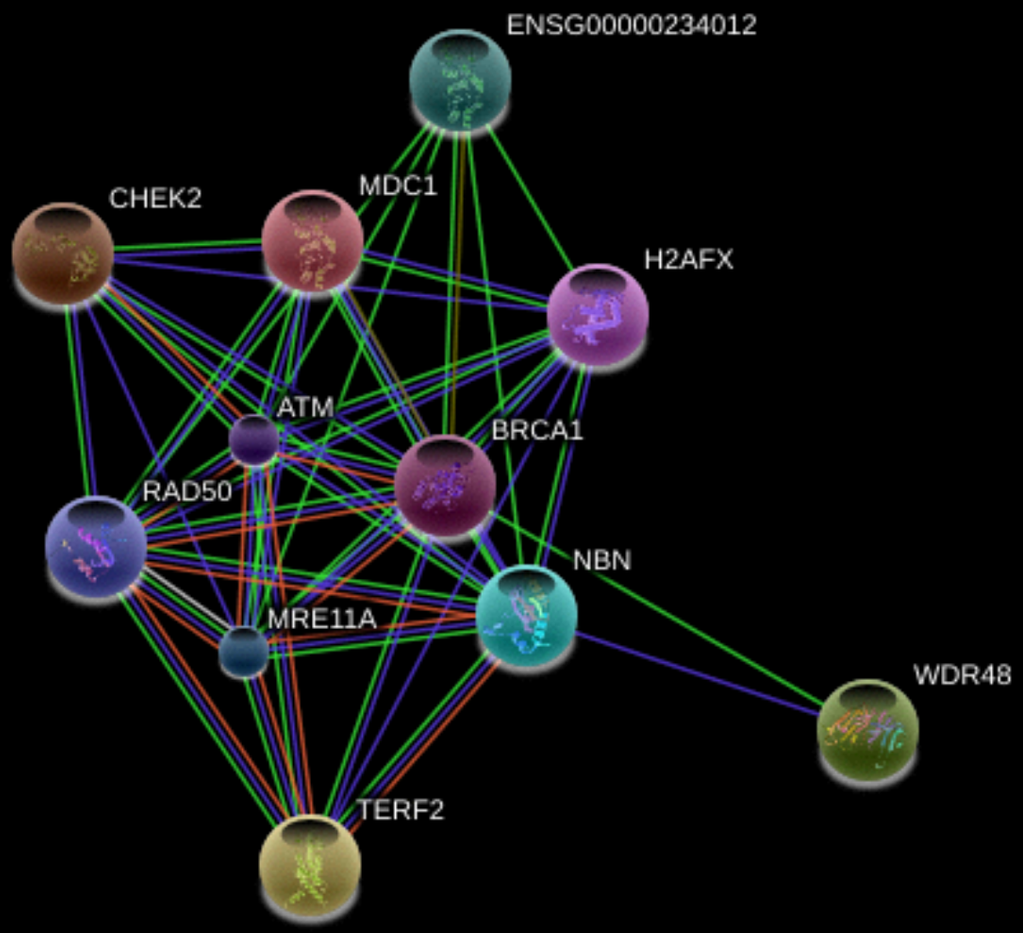
http://smart.embl-heidelberg.de/images/new_logo2.png

http://upload.wikimedia.org/wikipedia/en/0/03/Pfam_logo.gif

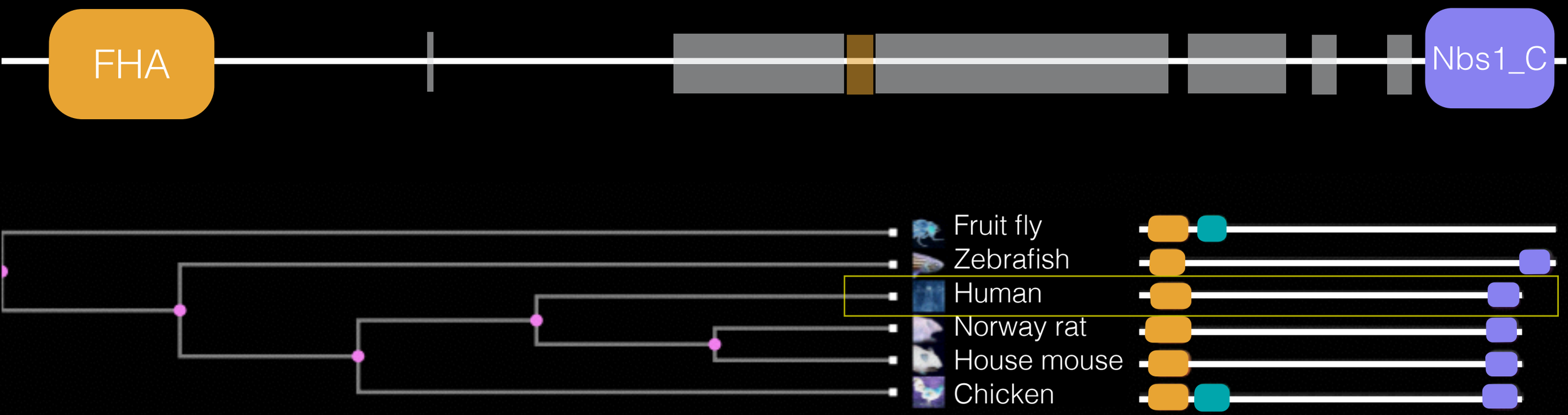
http://www.blopig.com/blog/wp-content/uploads/2014/08/HMM_example.jpg

Quick Look at SMART





Quick Look at Pfam



FHA: Binds phosphopeptides,
DNA damage response

BRCT (from SMART): Binds phospho-ser
DNA damage response

Nbs1_C: Binds Mre11 and Tel1
DNA damage response

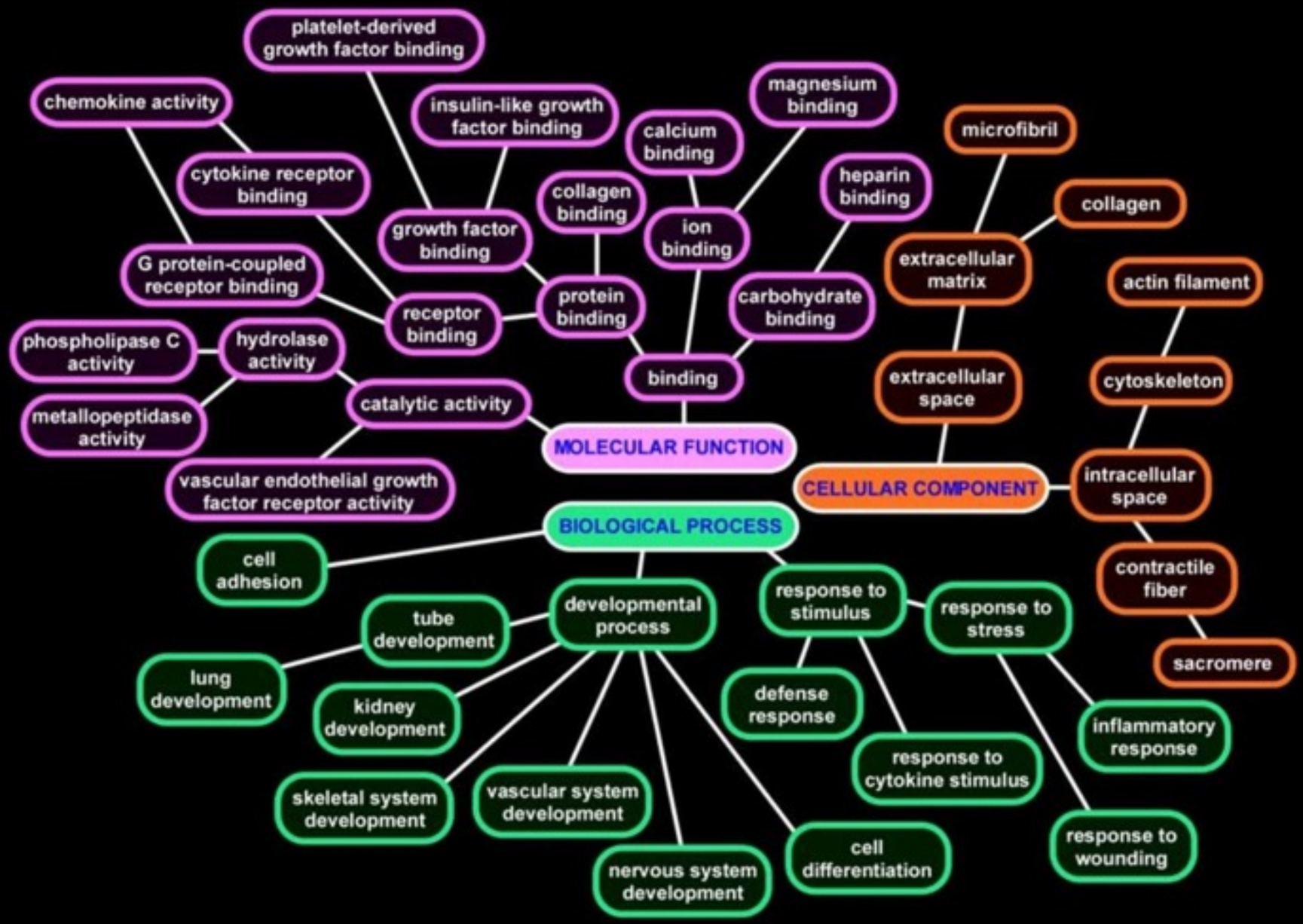
What does this tell you about my protein?

Motifs are functional groups of domains
Domains are functional groups of proteins

Learning about motifs and domains
helps us learn about proteins

What is GO?

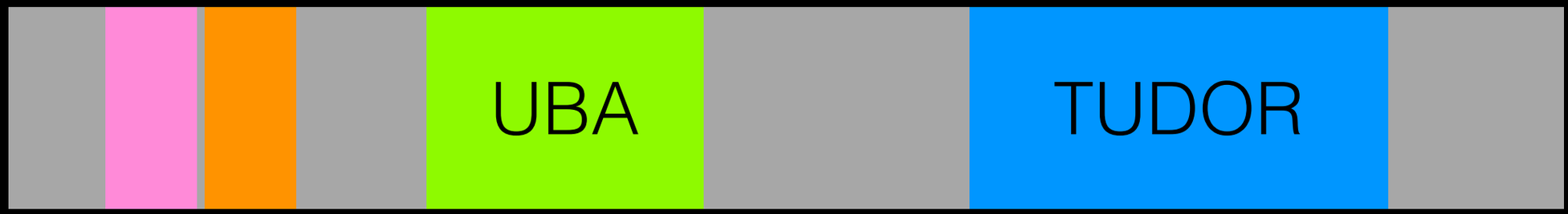
Gene ontology

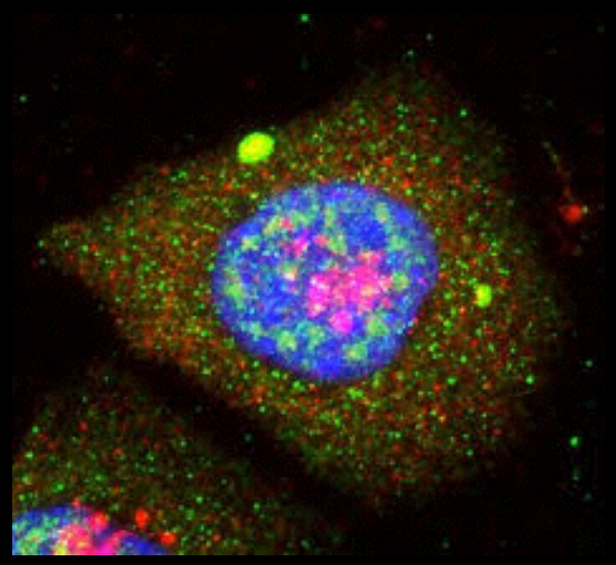


A major bioinformatics initiative to unify the representation of gene and gene product attributes across all species.

http://respiratory-research.com/content/12/1/63/figure/F4?highres=y

TDRD3, a novel Tudor domain-containing protein, localizes to cytoplasmic stress granules

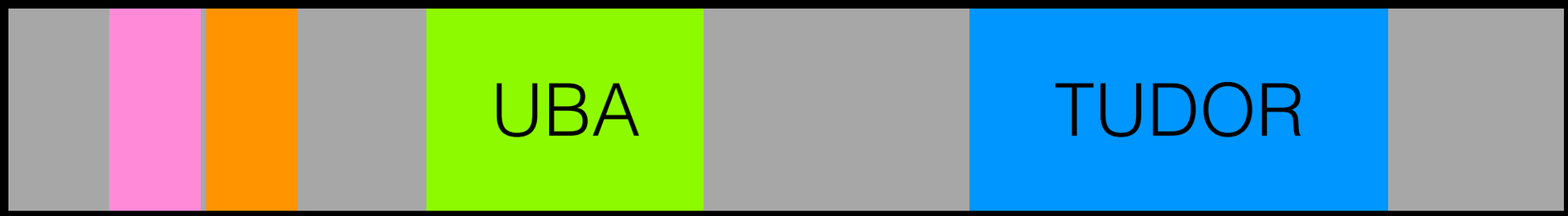




Goulet et al. (2008) Human Molecular Genetics.

Overview: TDRD3 & Tudor domain paper

TDRD3



TDRD3 domain structure & Tudor domains

Localization to stress granules (SGs)

Domain analysis

Methylated arginines important for localization

Figure 1: Domain architecture of TDRD3

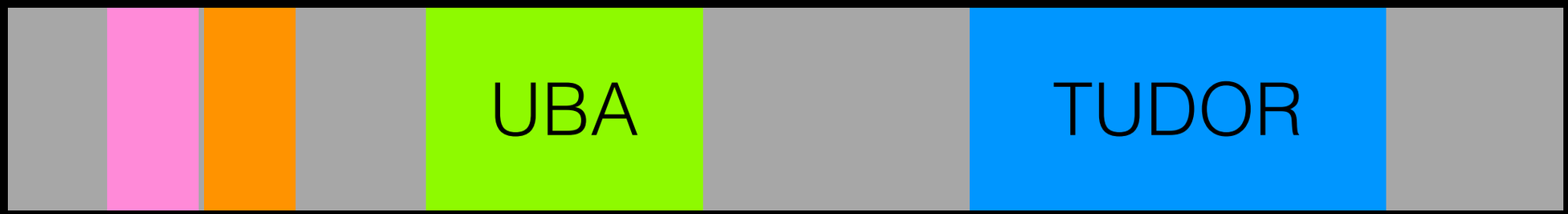
14 exons

TDRD3 domain structure

Redrawn Domain Structure of TDRD3

Structural Motifs

RNA binding



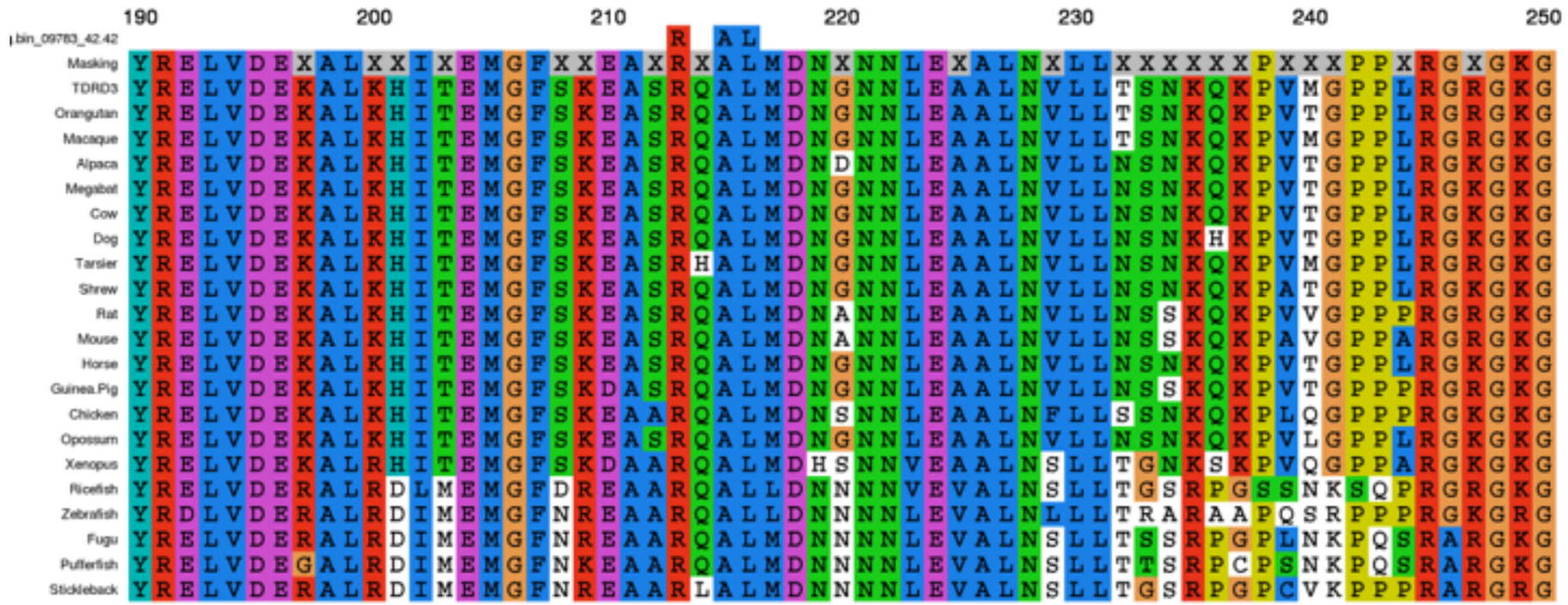
Ubq pathway

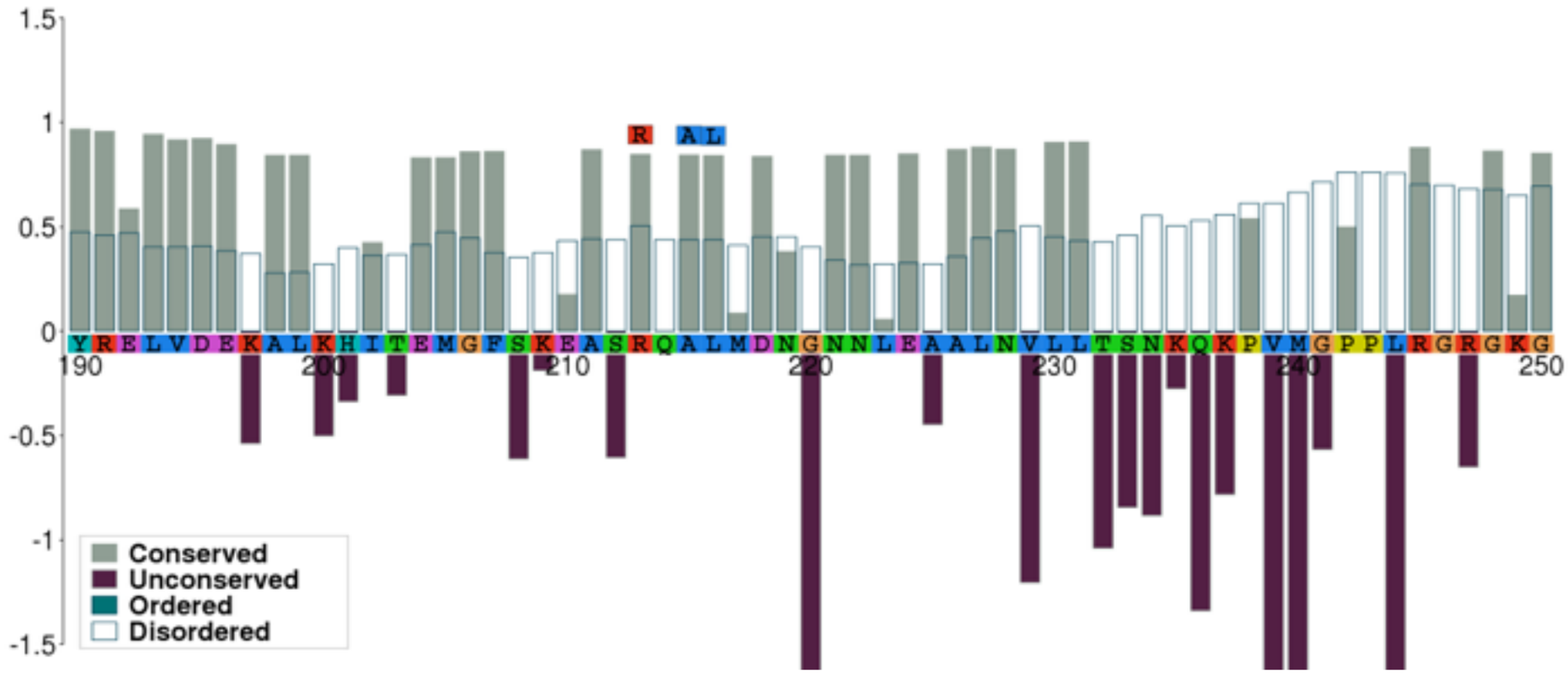
Breast cancer
Fragile X Syndrome
Severe testicular pathology

Used NCBI and PFAM

How the TUDOR domain was determined

PFAM uses HMMs





http://bioware.soton.ac.uk/slimdb/humsf09/map_pnrseq.bin_09783_42-42.TDRD3_HUMAN__Q9H7E2.00215

TDRD3 is involved in histone modification

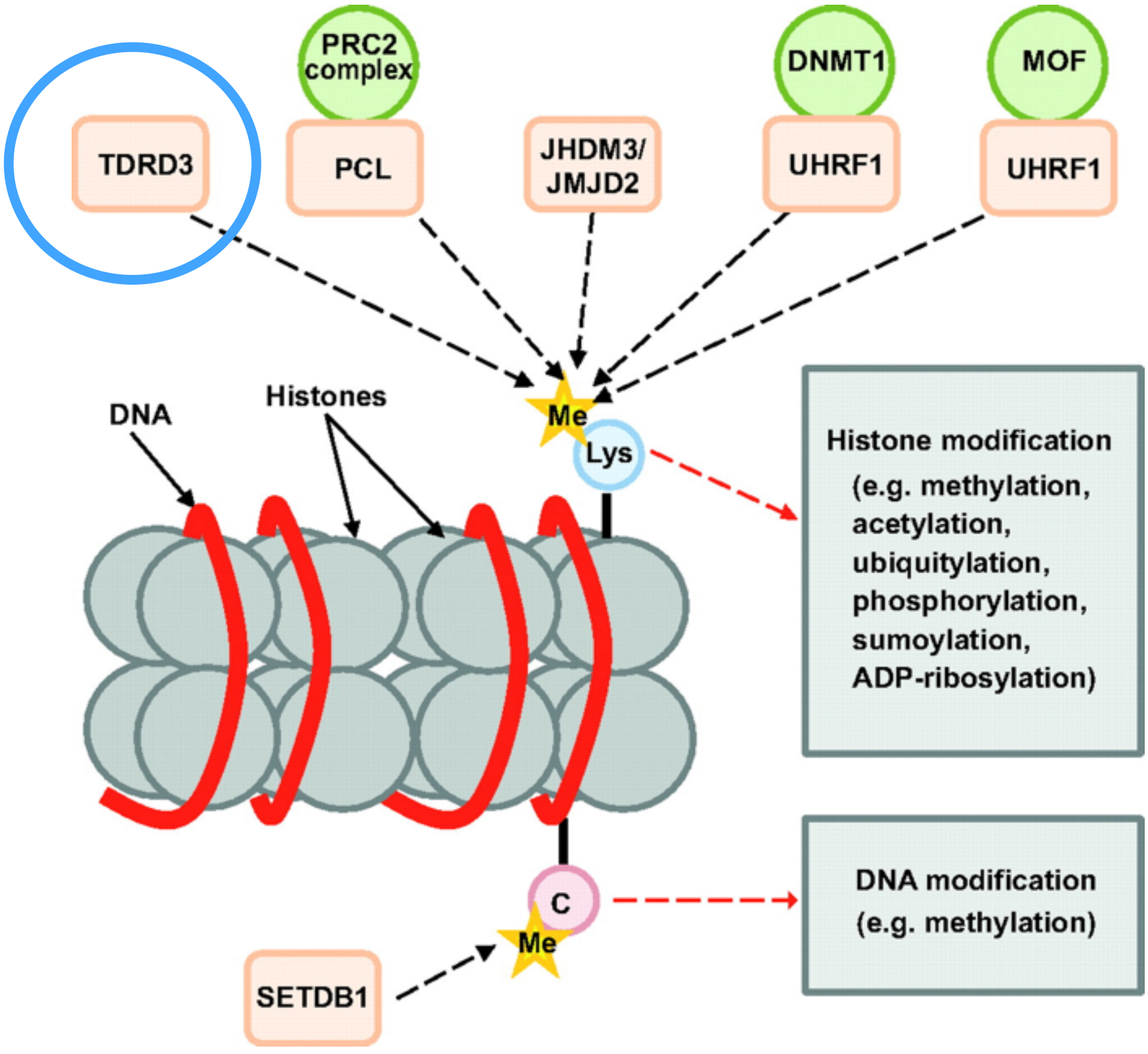
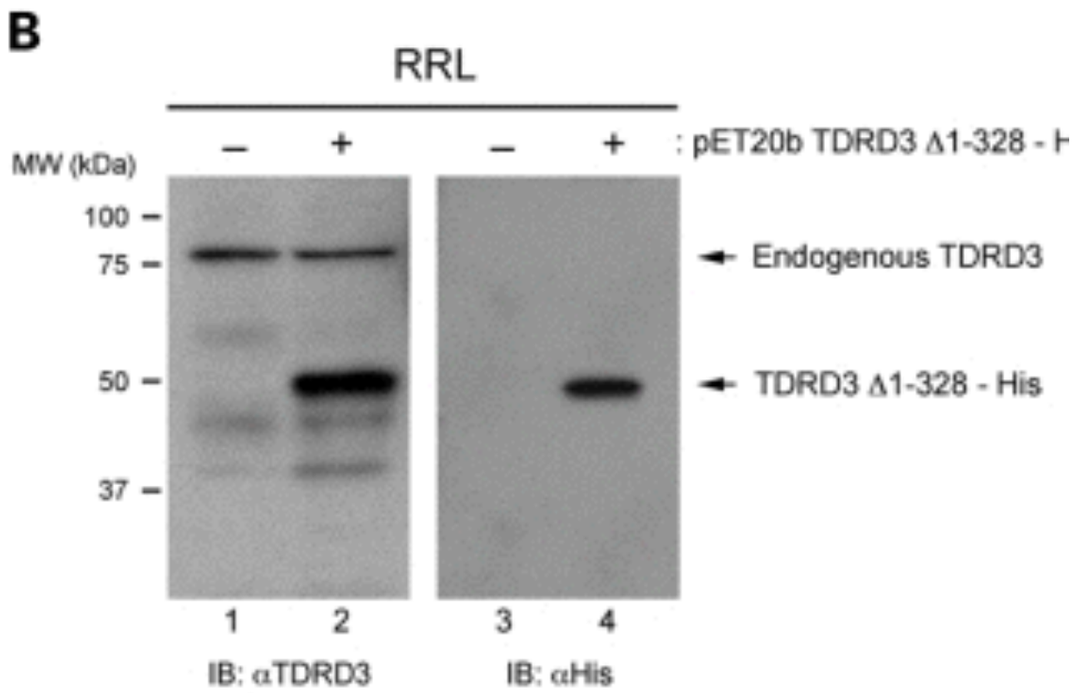
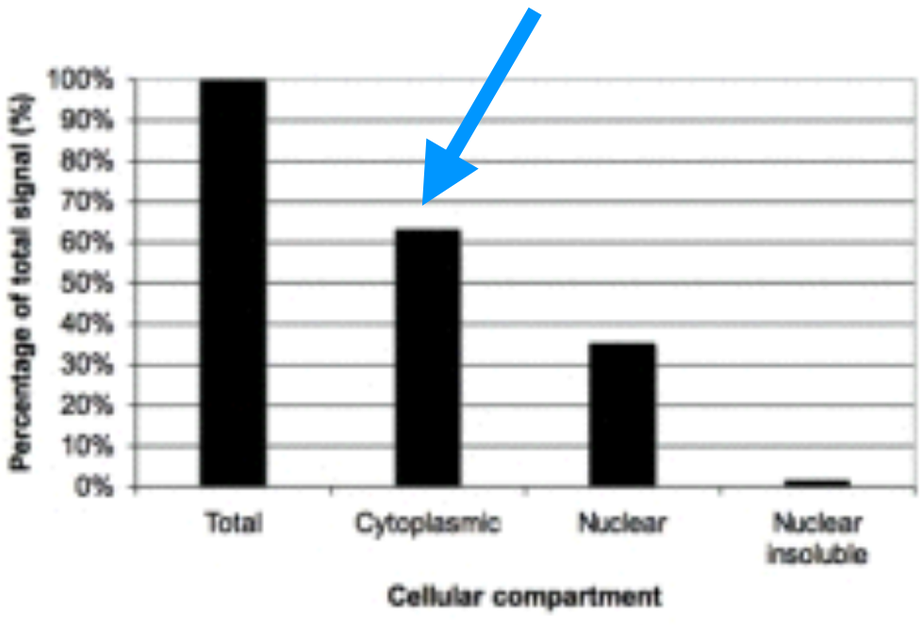
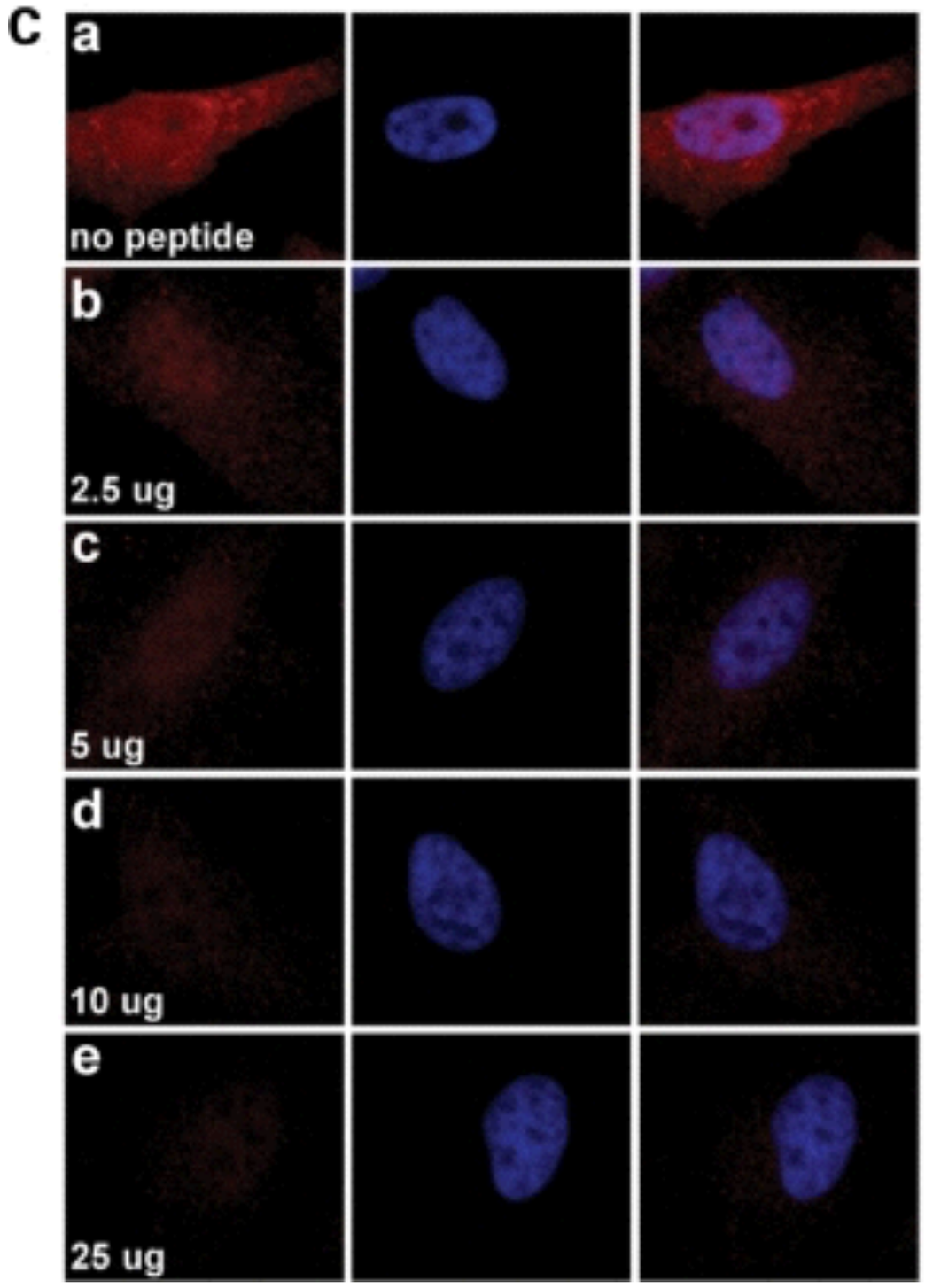


Figure 2: Where does TDRD3 localize?

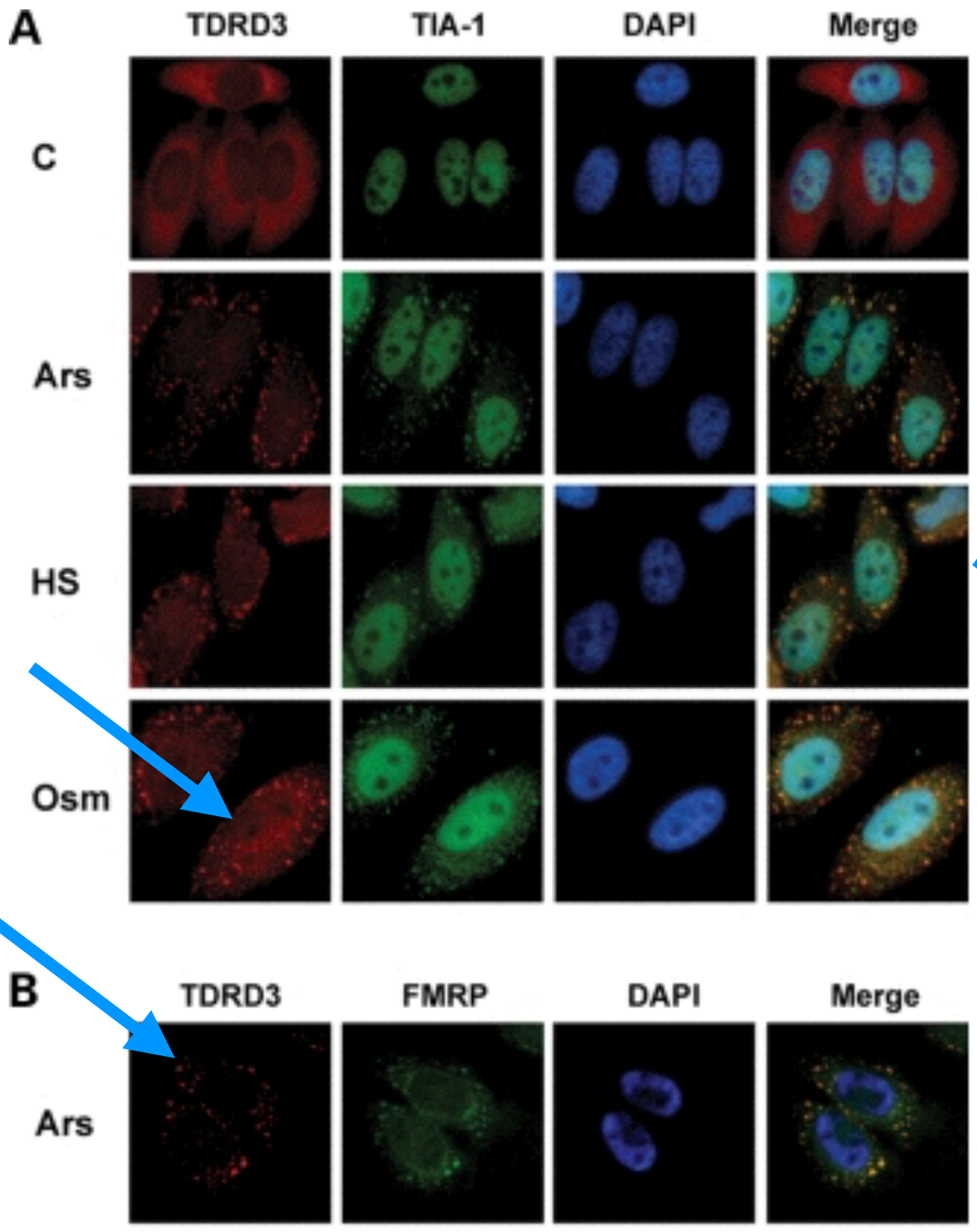


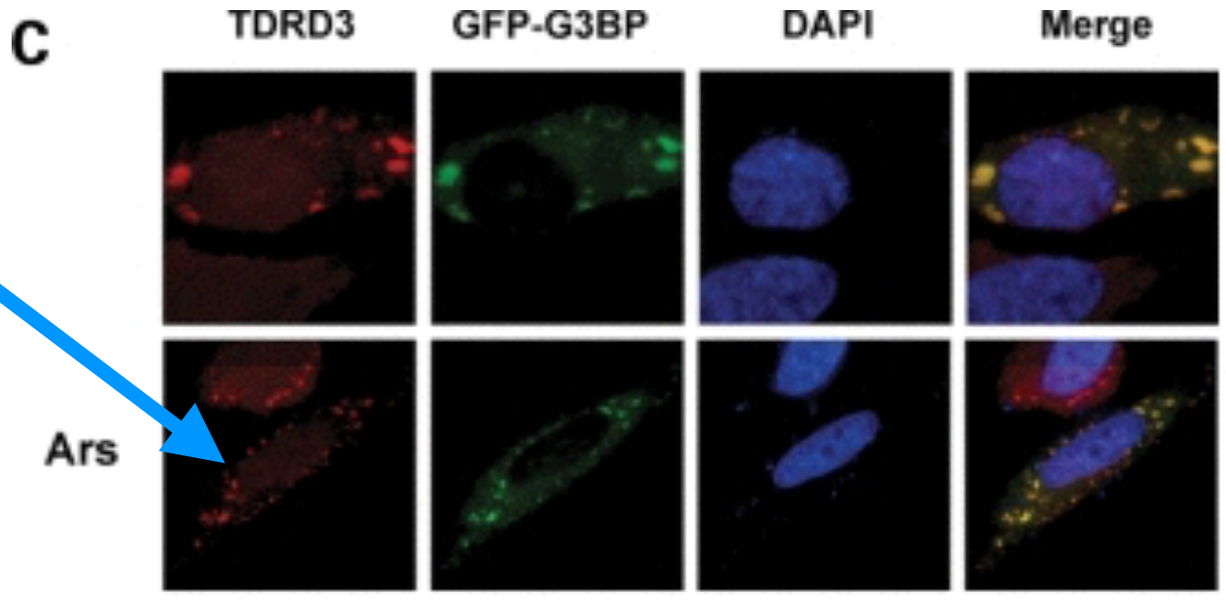




Conclusion: TDRD3 is found predominantly in the cytoplasmic compartment

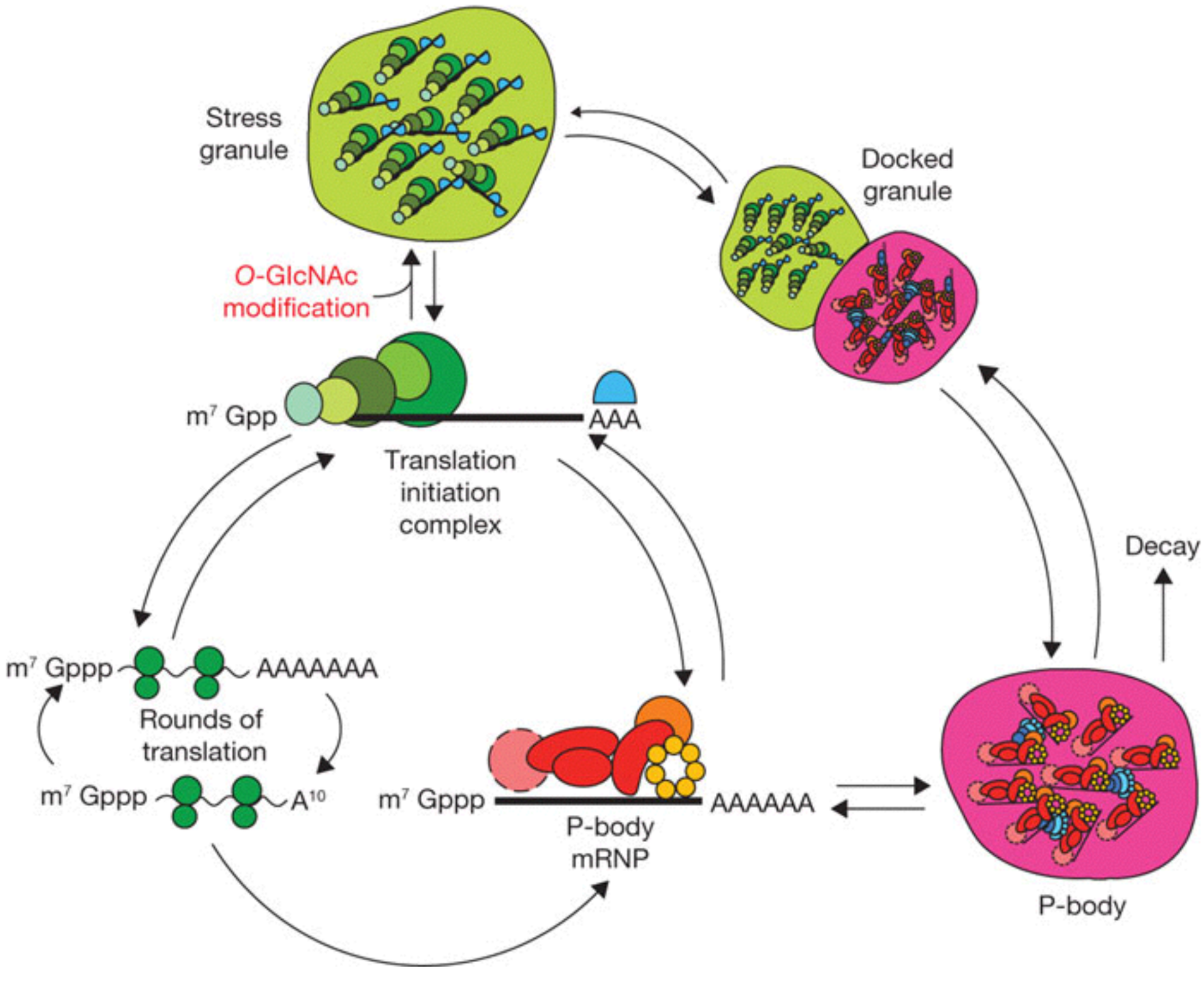
Figure 3: TDRD3 localizes to stress granules

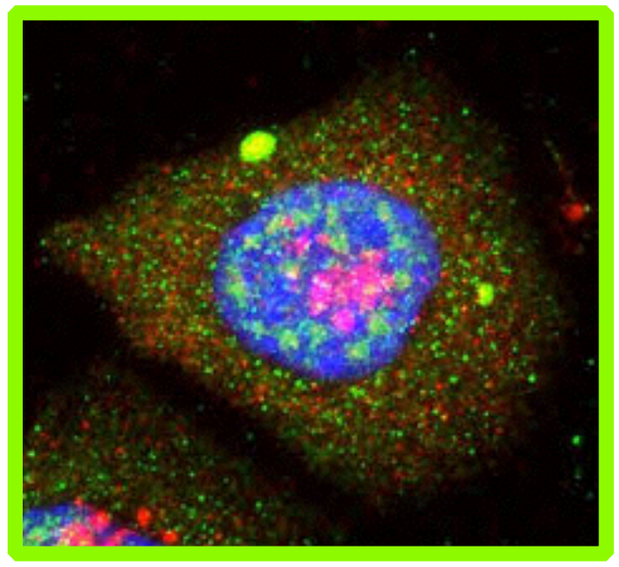




Conclusion: TDRD3 localizes to cytoplasmic stress granules during stress response

What are stress granules?

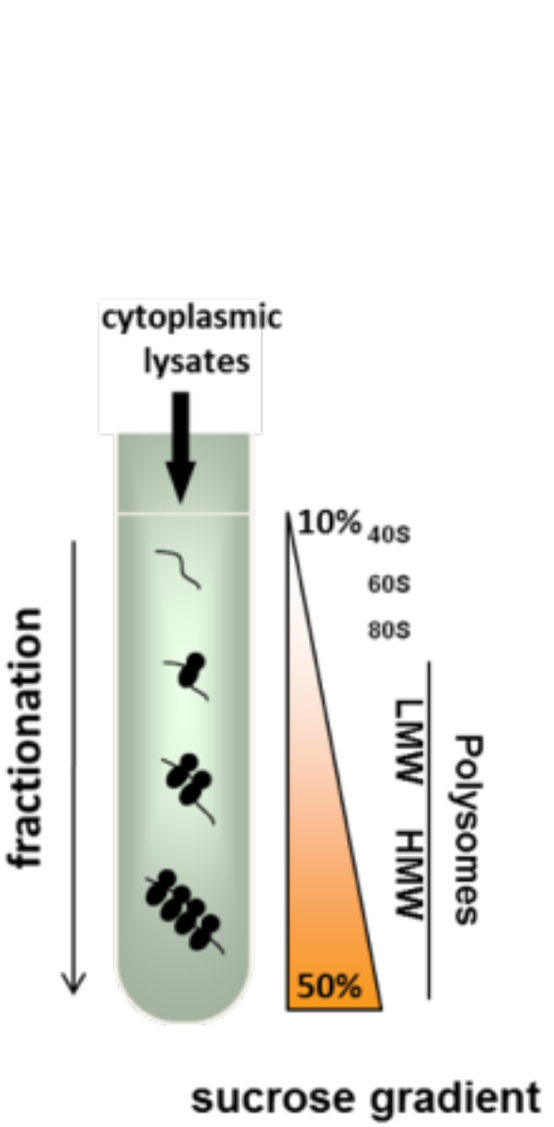


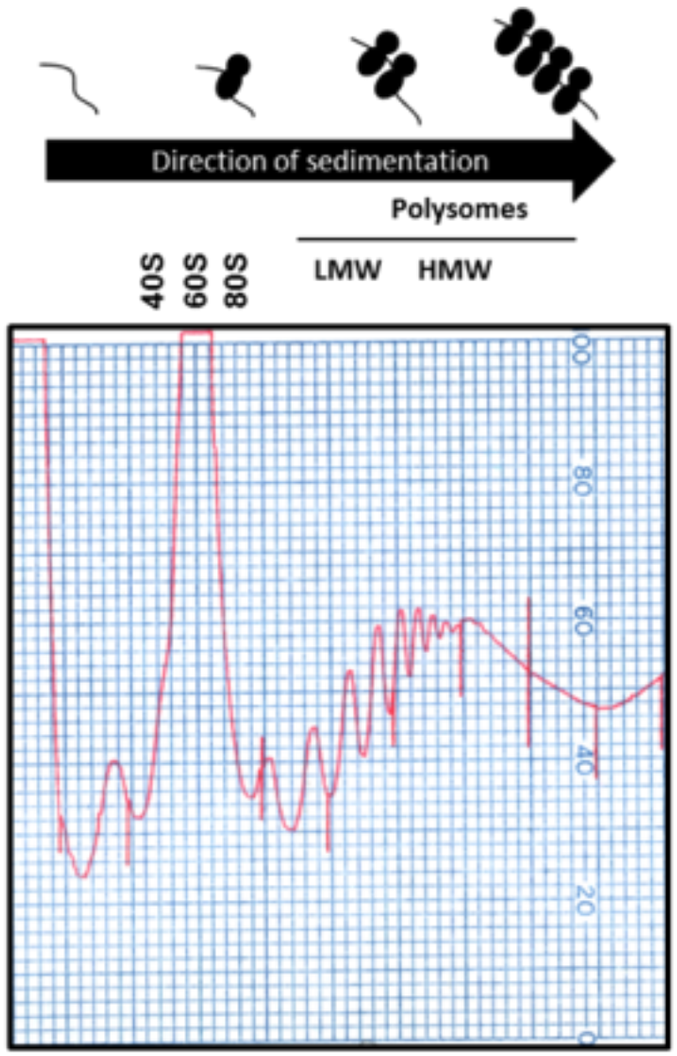


Neurodegenerative diseases:
ALS
Fragile X syndrome
Alzheimer's

Proteins can regulate proteins that regulate translation

Sedimentation analysis of polyribosomes





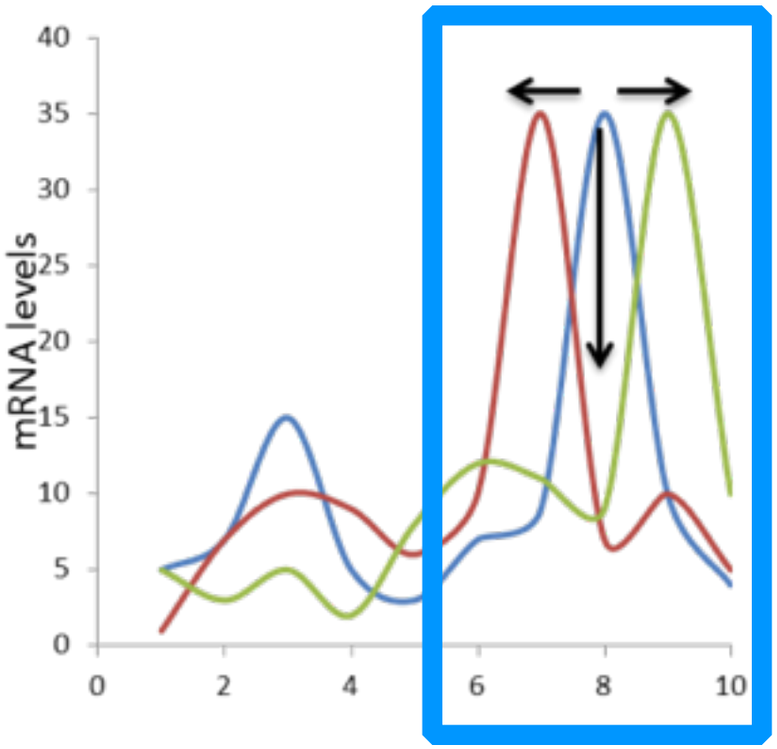
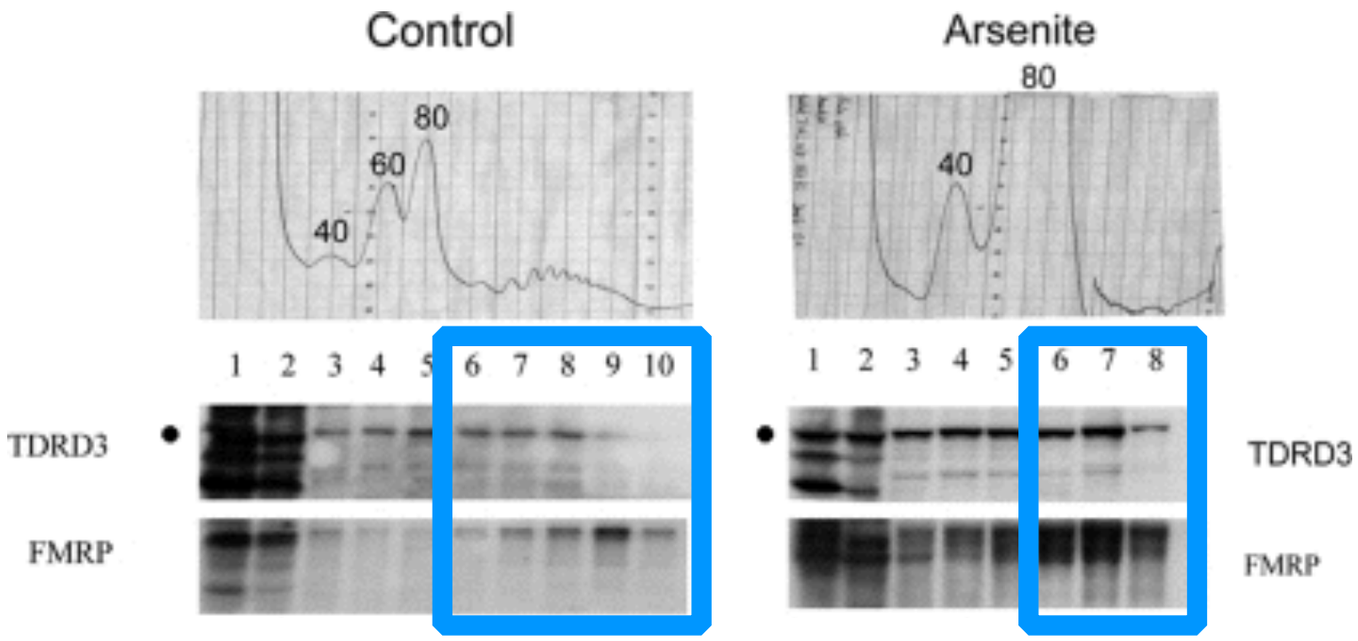
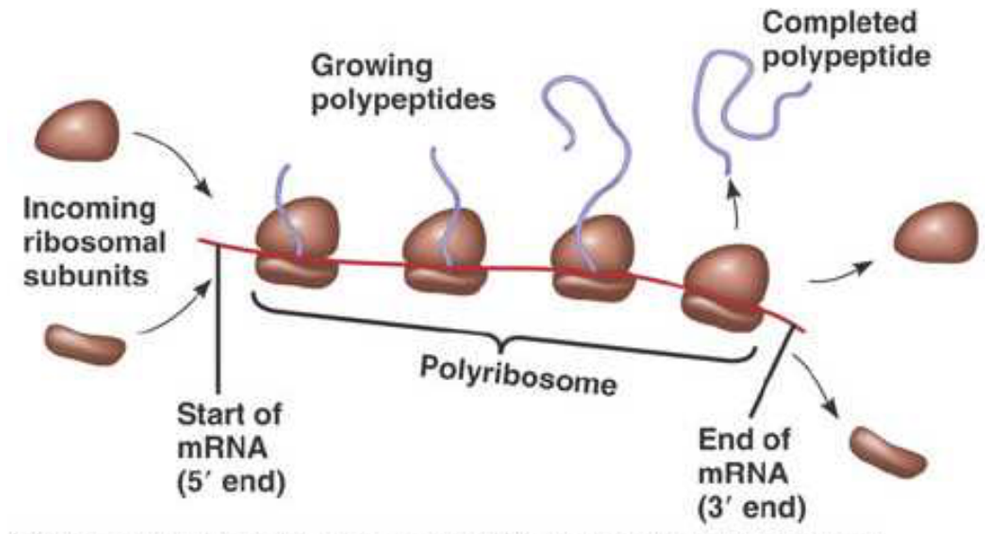
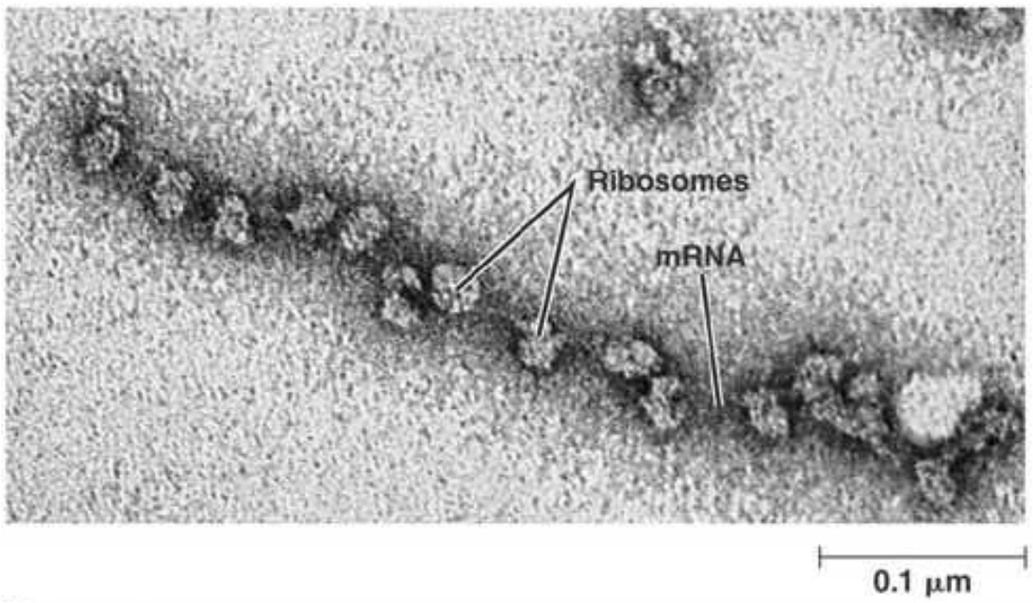


Figure 4: TDRD3 localizes to polyribosomes

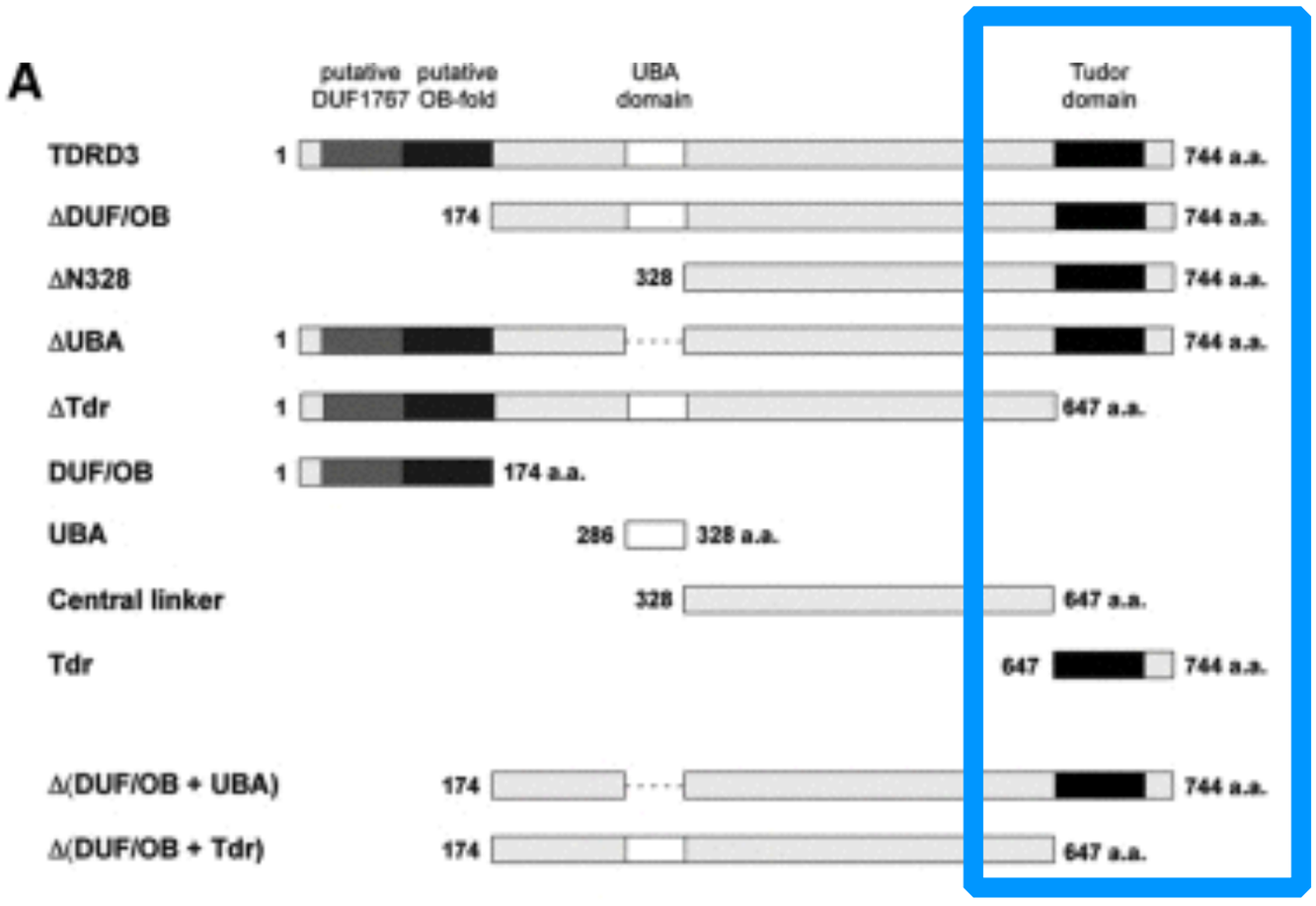


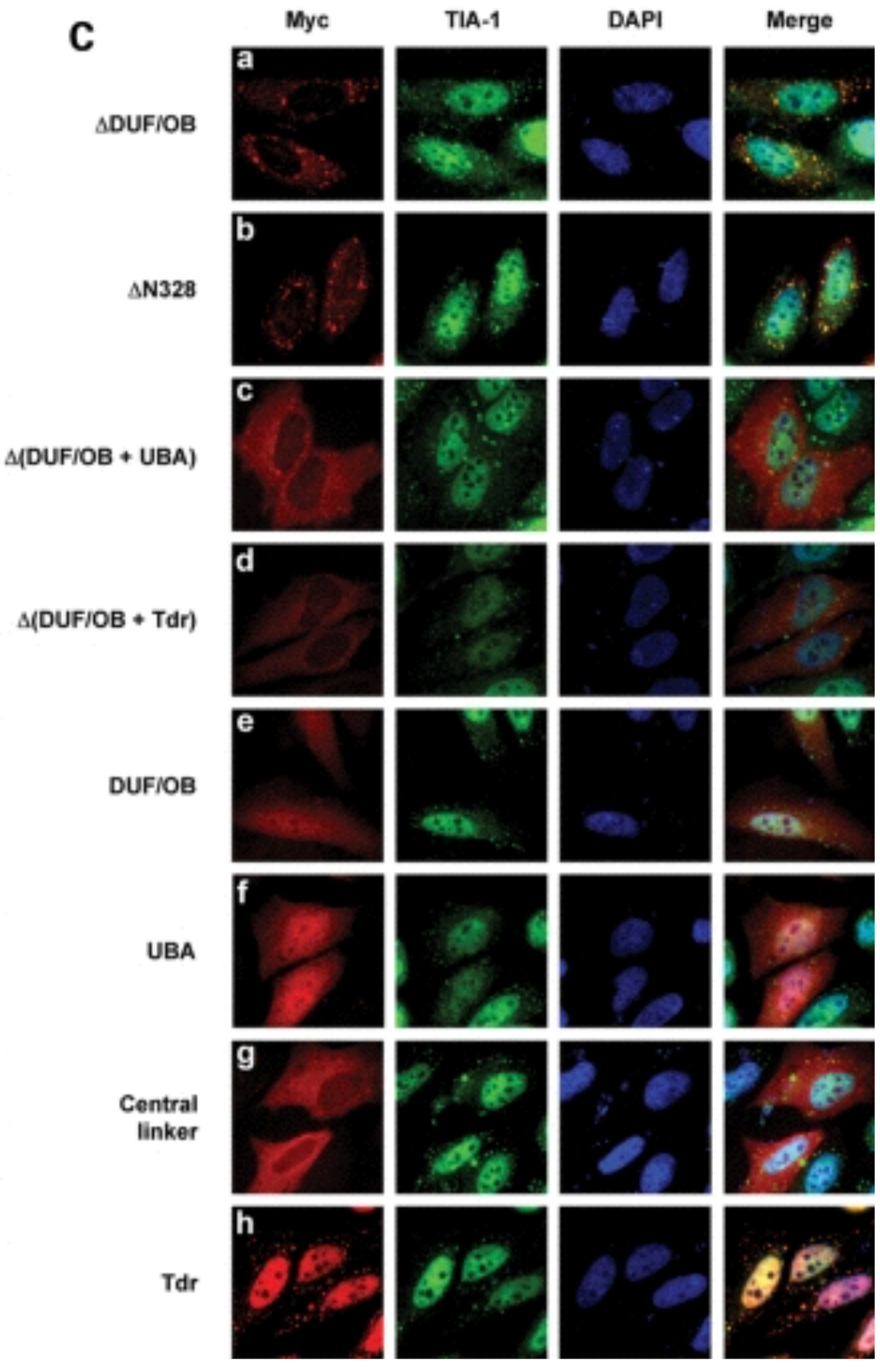




Conclusion: TDRD3 is associated with actively translating polyribosomes

Figure 5: The TUDOR domain is required for its localization to stress granules

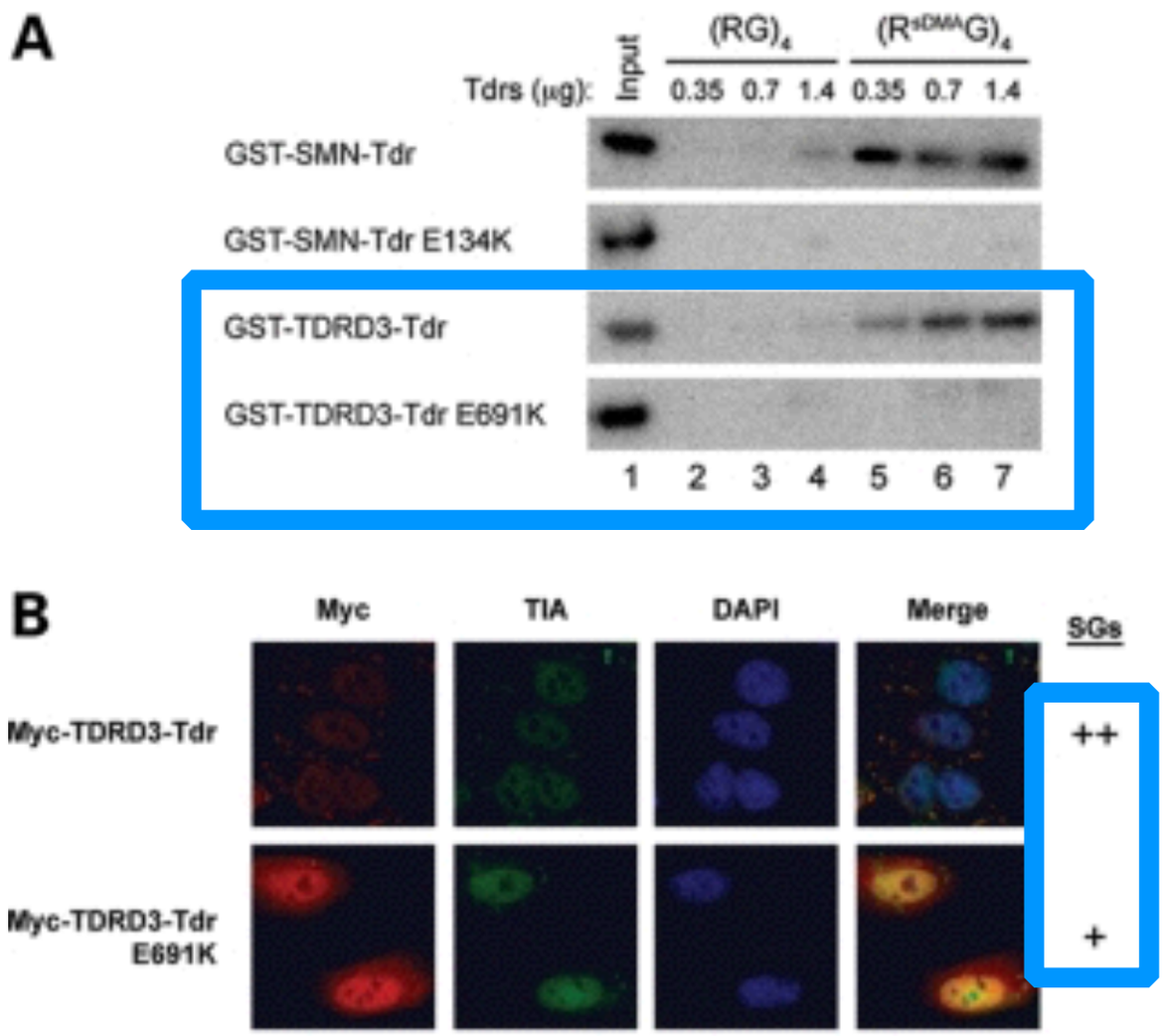




Conclusion: Tudor domain is both essential and sufficient for TDRD3 translocation to cytoplasmic stress granules

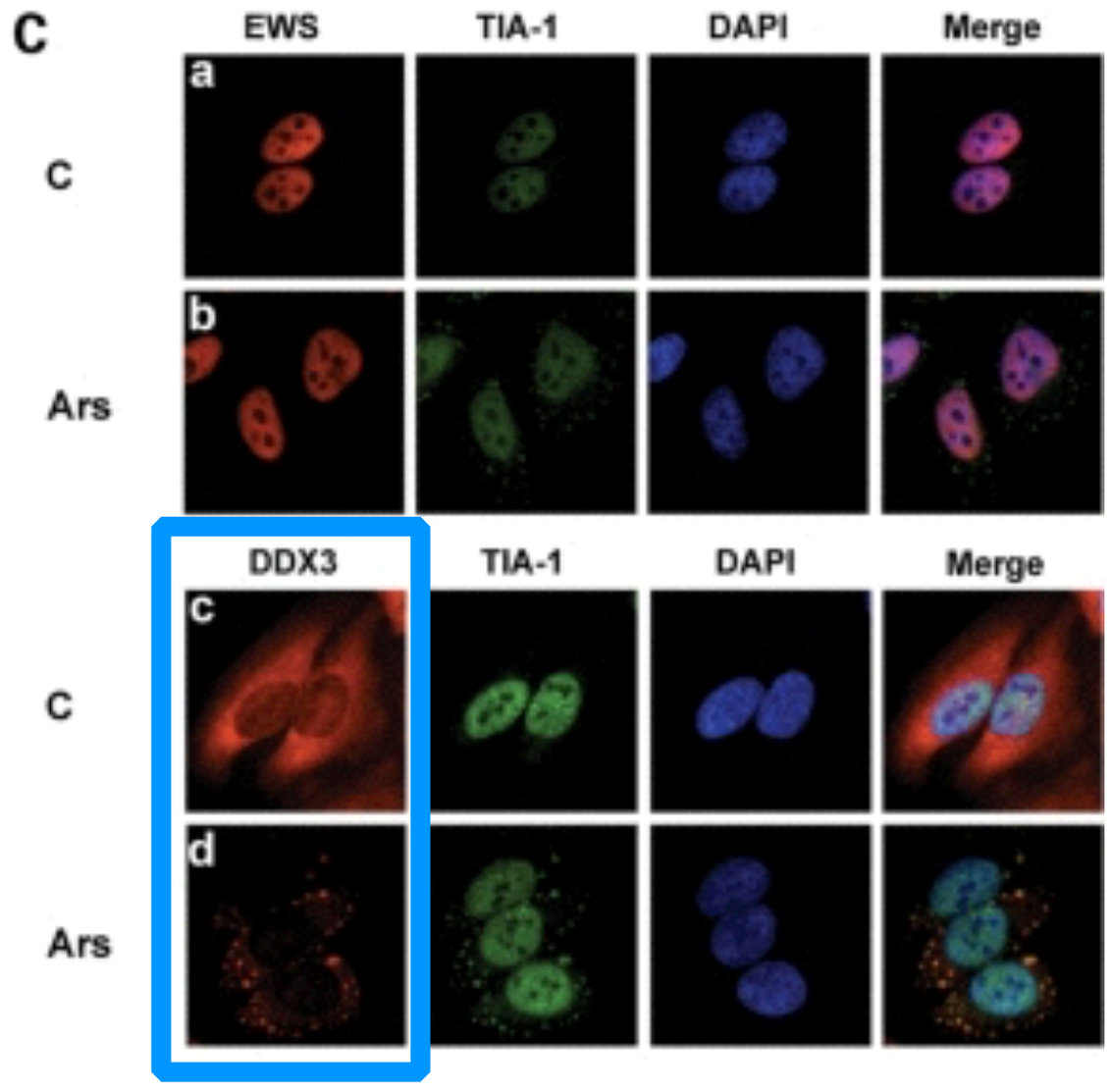
Figure 6: Direct contact with methylated arginines is important for localization to stress granules

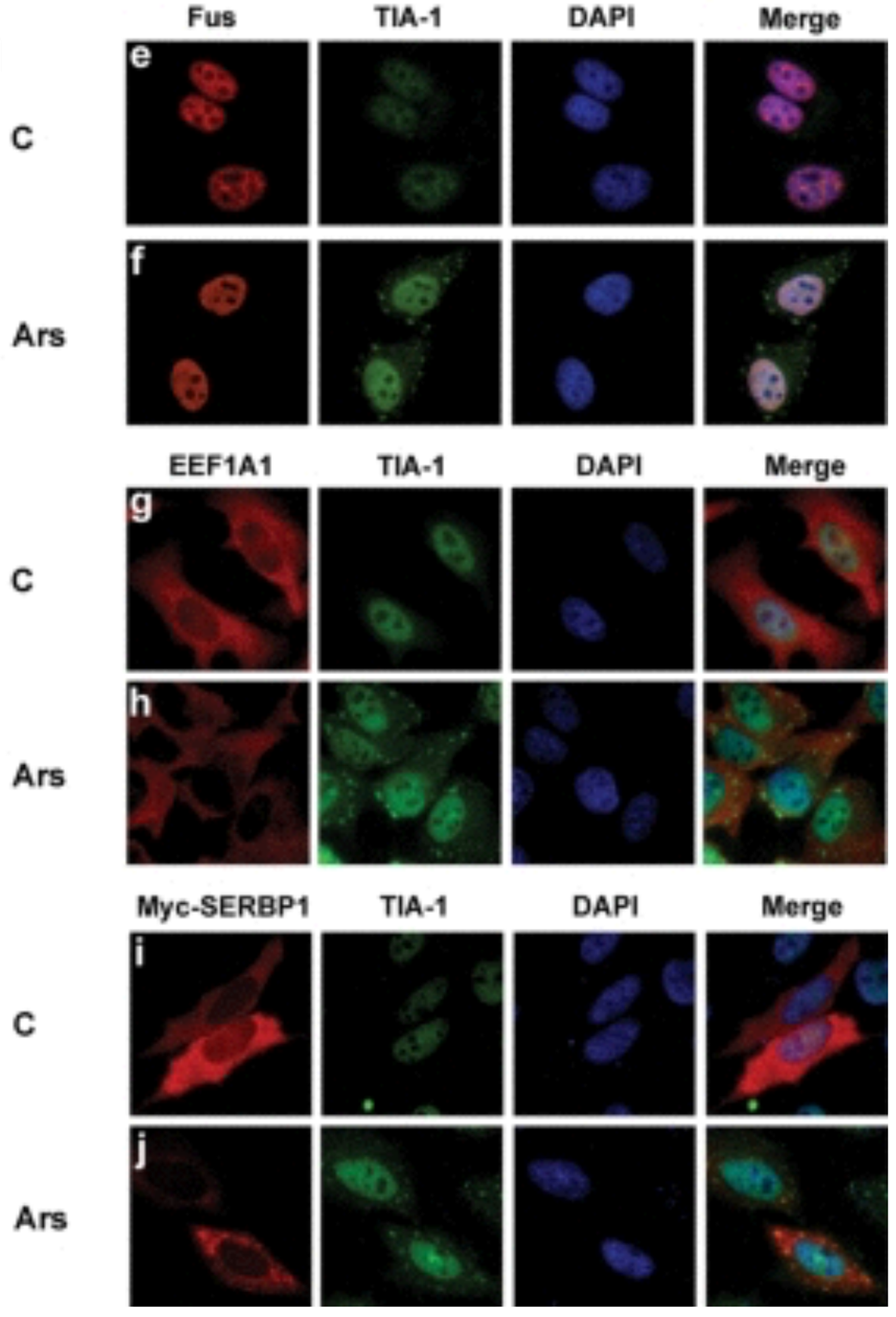
E134k and E691k abolish capacity to interact with methylated arginines



Conclusion: TDRD3's recruitment to SGs is likely mediated through specific protein–protein interactions involving methylated arginines

Figure 6 (cont'd): Direct contact with methylated arginines important for localization to stress granules

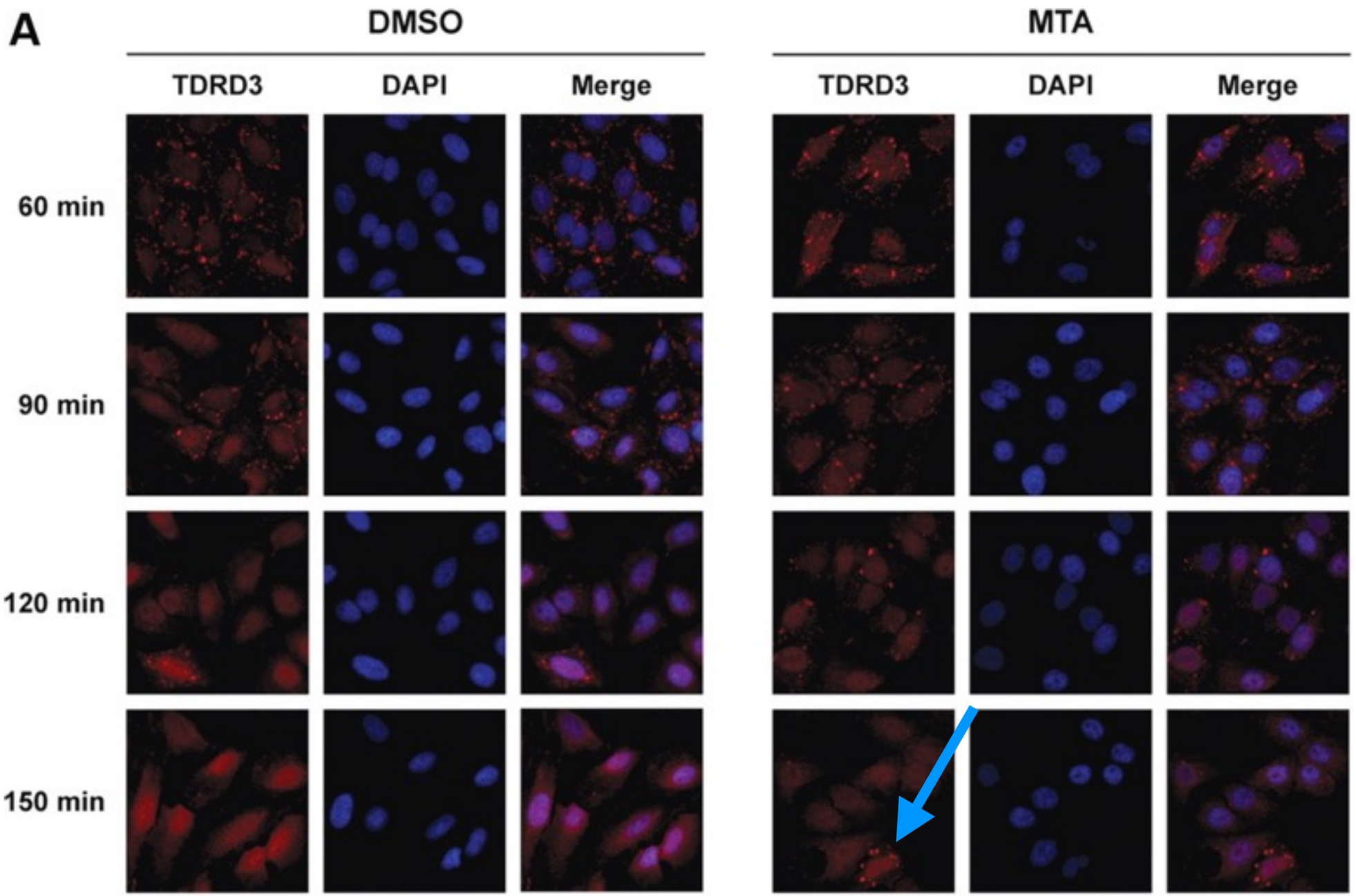




Conclusion: DDX3 as a good candidate to mediate the recruitment of TDRD3 to stress granules

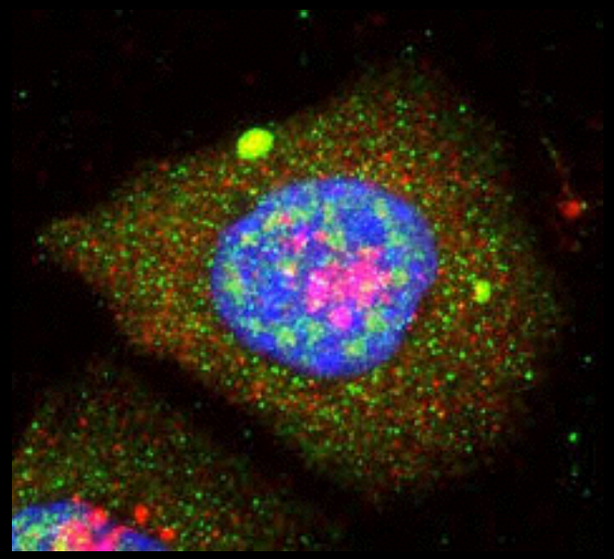
EWS, DDX3, Fus, EEF1A1, Myc-SERBP1 are proteins associated with RNA handling

Figure 7: Arginine methylation and SG dynamics

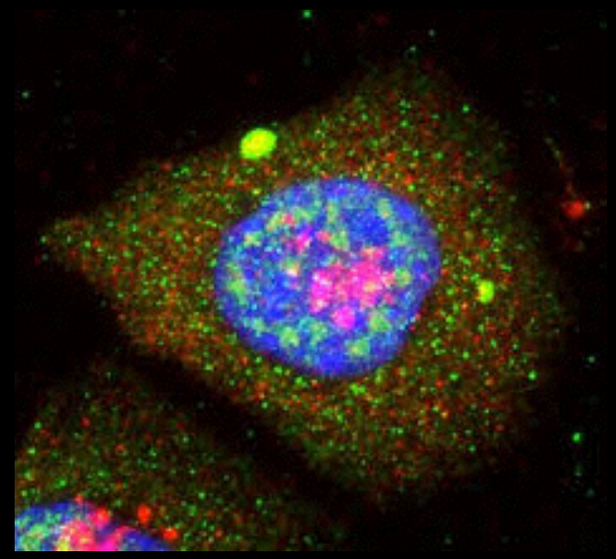


MTA inhibits methylation

Conclusion: methylation needed for regulating stress granule dynamics and composition



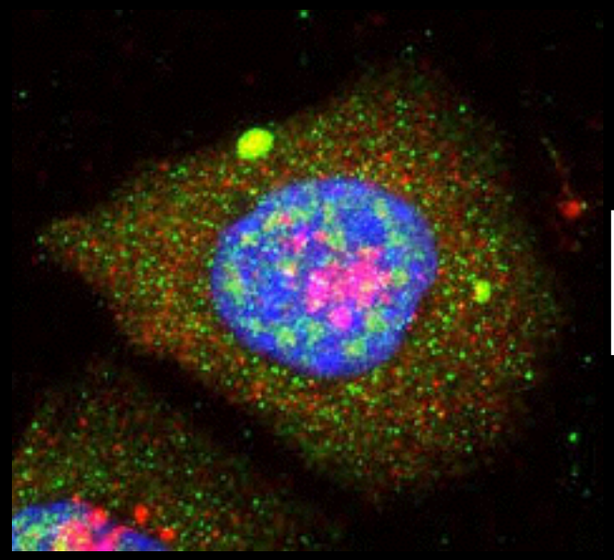
Conclusion



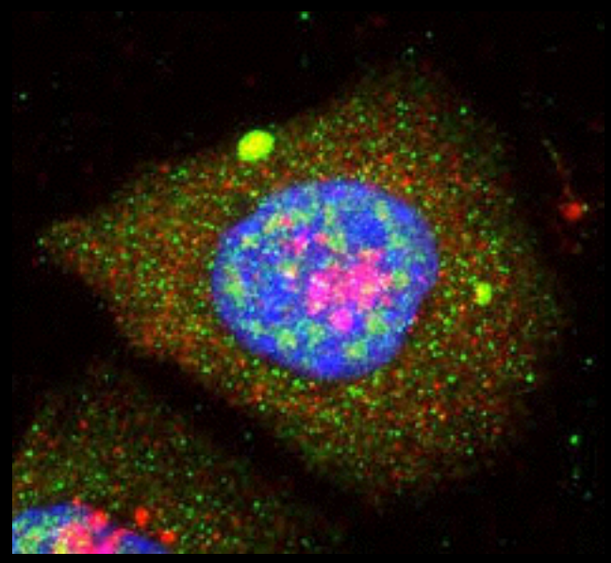
TDRD3 is found predominantly in the cytoplasmic compartment

TDRD3 was detectable in stress granules following stress stimulus

Tudor domain is both essential and sufficient for TDRD3 translocation to cytoplasmic stress granules



Future studies



Kinetics of recruitment of stress granules

Regulating translation within stress granules

Germ cell biology

Participation in disease pathways