

Bioinformatics for Biologists

Day 10: PFAM & KEGG

Complete the activity individually while working together in the group. Submit it Canvas individually.

PFAM

Which domains/families are found in the aromatic amino acid hydroxylases from human (I know, you may be tired of these by now, but let's use them one final time)?

1. Fill out the table with the information for the PFAM hits for human PAH (P00439). Clicking on the domain/family takes you to its entry.

	Family (code)	Entry type (Domain or family)	# of architectures	# of seq in seed	# of seq in full
1	PF01842	Family	177	131	27853
2	PF00351	Family	35	60	3725

2. In which species are you likely find domain/family 1 and 2, respectively?

1: E+B+A

2: E+B

3. What is the main function of the first family/domain according to the Pfam or Interpro description found under the Summary tab?

1: regulates enzymes regulated by [AA]

2: hydroxylates aromatic AA

4. For the first domain/family, which amino acid is most common in position 28 in the HMMlogo? What are shared characteristics of the 3 most common amino acids at this site?

1: I, 2: V, 3: L

All are hydrophobic

5. What is the difference between a PFAM domain and a PFAM family? See [Pfam: Help \(xfam.org\)](http://pfam.xfam.org)

Family:

A collection of related protein regions

Domain:

A structural unit

Bioinformatics for Biologists

KEGG

Go to KEGG: [KEGG: Kyoto Encyclopedia of Genes and Genomes](http://www.kegg.jp/)

Click on **KEGG PATHWAY**

Click on **Environmental Information Processing**.

Under Signal Transduction, find the link to **TGF-beta signaling pathway**

The upper left corner should look like this:



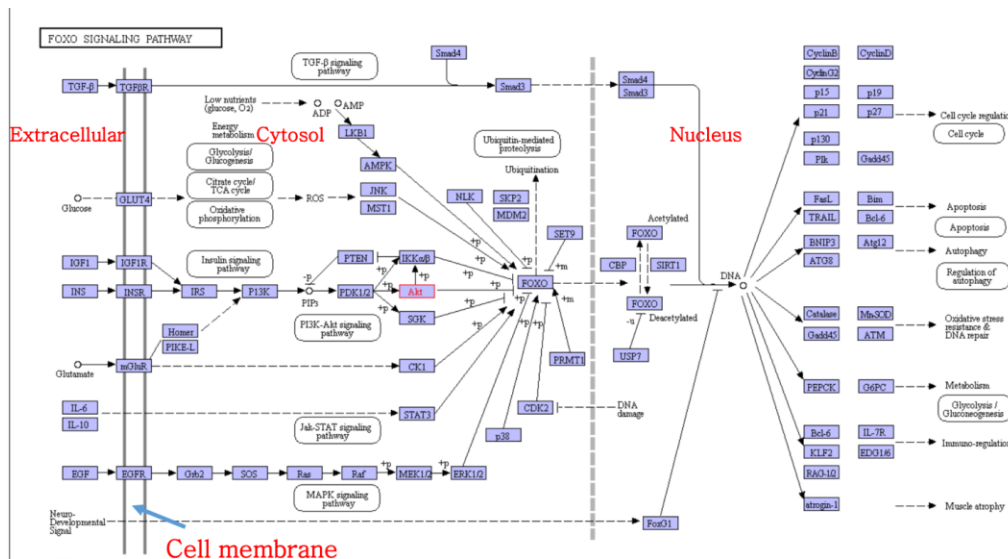
TGF-beta signaling pathway - Homo sapiens (human)

[[Pathway menu](#) | [Pathway entry](#) | [Download KGML](#) | [Show description](#) | [Image \(png\) file](#) | [Help](#)]

Change pathway type

Subcellular location

Focus on the four locations shown in red (this is an example from a different pathway than the one you are working with).



1. Where is ERK found?
 - a. cytoplasm
2. Where is BMP found?
 - a. extracellular
3. Where is the complex of TGFβ RI and TGFβ RII found?
 - a. In the cell membrane
4. Where is c-Myc found?
 - a. Nucleus

Regulation

Bioinformatics for Biologists

Use the Help tab to decode what the different symbols and lines mean.



[[Pathway menu](#) | [Pathway entry](#) | [Download KGML](#) | [Show description](#) | [Image \(png\) file](#) | [Help](#)]

5. How is BMP regulated by Noggin?
 - a. Inhibited
6. How does BMP regulate BMPRII?
 - a. Activating
7. How does Noggin indirectly regulate BMPRII?
 - a. inhibiting
8. How is SMAD1/5/8 activated by BMP RI?
 - a. By phosphorylation
9. Does SMAD2/3 inhibit SMAD6/7 or vice versa?
 - a. SMAD6/7 inhibits SMAD2/3

What's behind the button?

Click on it and find familiar information such as GeneID, accession number, sequence, GO term, and more links to other pathways that this protein is participating in etc.

10. There are several SMADs in the TGF β pathway. Assuming they are related, are they isoforms or paralogs?
 - a. Click on the boxes for each SMAD. It will show the sequence and GeneID. They have different GeneIDs so they are paralogs.
11. What proteins in the refseq_protein database are identical in sequence and have with 100% query cover to NodalRI and NodalRII, respectively? (name and accession)
 - a. Activin receptor type-1C and type 2A, respectively.
12. Are BMPRI, BMPRII, TGF β RI, TGF β RII, ActivinRI, ActivinRII, NodalRI, and NodalRII in human homologs (use BLAST to answer this question)?
 - a. YES, BLAST with one, you find the others.