**Find a Gene Assignment – BGGN213**

Jason Hsiao (A15871650)

1. Protein: Human RBP4 (Retinol Binding Protein 4 Isoform a Precursor)

Accession Number: [**NP\_001310446.1**](https://www.ncbi.nlm.nih.gov/protein/NP_001310446.1)

Species: Human

Function: “This protein belongs to the lipocalin family and is the specific carrier for retinol (vitamin A alcohol) in the blood. It delivers retinol from the liver stores to the peripheral tissues. In plasma, the RBP-retinol complex interacts with transthyretin which prevents its loss by filtration through the kidney glomeruli. A deficiency of vitamin A blocks secretion of the binding protein post translationally and results in defective delivery and supply to the epidermal cells.” [provided by RefSeq, Jul 2008].

1. BLAST Method Used: tblastn 2.14.1+

Database: Expressed sequence tags (est)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. A screenshot of a computer

   Description automatically generated

>BI546698.1 603191445F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:5262803 5', mRNA sequence

AGCGGAGGCTTGCGCGCGGTTCCCCTCCCGGTGGGCGGATTCCTGGGCAAGATGAAGTGGGTGTGGGCGC

TCTTGCTGTTGGCGGCGCTGGGCAGCGGCCGCGCGGAGCGCGACTGCCGAGTGAGCAGCTTCCGAGTCAA

GGAGAACTTCGACAAGGCTCGCTTCTCTGGGACCTGGTACGCCATGGCCAAGAAGGACCCCGAGGGCCTC

TTTCTGCAGGACAACATCGTCGCGGAGTTCTCCGTGGACGAGACCGGCCAGATGAGCGCCACAGCCAAGG

GCCGAGTCCGTCTTTTGAATAACTGGGACGTGTGCGCAGACATGGTGGGCACCTTCACAGACACCGAGGA

CCCTGCCAAGTTCAAGATGAAGTACTGGGGCGTAGCCTCCTTTCTCCAGAAAGGAAATGATGACCACTGG

ATCGTCGACACAGACTACGACACGTATGCCGTGCAGTACTCCTGCCGTCTCCTGAACCTCGATGGCACCT

GTGCTGACAGCTACTCCTTCGTGTTTTCCCGGGACCCCAACGGCCTGCCCCCAGAAGCGCAGAAGATTGT

AAGGCAGCGGCAGGAGGAGCTGTGCCTGGCCAGGCAGTACAGGCTGATCGTCCACAACGGTTACTGCGAT

GGCAGATCAGAAAGAAACCTTTTGTAGCAATATCAAGAATCTAGTTTCATCTGAGAACTTCTGATTAGCT

CTC

Translated Protein Sequence via Emboss Transeq

>BI546698.1\_1 603191445F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:5262803 5', mRNA sequence

SGGLRAVPLPVGGFLGKMKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWY

AMAKKDPEGLFLQDNIVAEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAK

FKMKYWGVASFLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLP

PEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLL\*QYQESSFI\*ELLISSX

Longest sequence without stop \* of the 6 reading frames.

A screenshot of a computer

Description automatically generated

No 100% identity with Homo sapiens, top hit is from gorilla.

1. Trimmed Alignment:

>Danio\_rerio\_RBP4\_(NP\_570995.1)

XM---LRLCIAVCVLATCWAQDCQVSNFAVQQDFNRTRYQGTWYAVAKKDPVGL

FXLLDNIVANFKVEEDGTMTATAIGRVIILNNWEMCANMFGTFEDTEDPAKFKMKYWGAA

AYXLQTGYDDHWIIDTDYDNYAIHYSCRELDEDGTCLDGYSFIFSRHPDGLRPEDQAIVT

QKKXQDICFLGKYRRVAHTGFCEAA

>Rattus\_norvegicus\_RBP4\_(NP\_037294.1)

XMEWVWALVLLAALGGGSAERDCRVSSFRVKENFDKARFSGLWYAIAKKDPEGL

FXLQDNIIAEFSVDEKGHMSATAKGRVRLLSNWEVCADMVGTFTDTEDPAKFKMKYWGVA

SFXLQRGNDDHWIIDTDYDTFALQYSCRLQNLDGTCADSYSFVFSRDPNGLTPETRRLVR

QRQXEELCLERQYRWIEHNGYCQSR

>Callithrix\_jacchus\_RBP4\_(XP\_054099088.1)

KMKWVWALLLLAVLGISRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGL

FXLQDNIIAEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVA

SFXLQKGNDDHWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIIR

QRQXEELCLARQYRLIVHNGYCDGK

>Gorilla\_RBP4\_(XP\_018890983.3)

KMKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGL

FXLQDNIVAEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVA

SFXLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVR

QRQXEELCLARQYRLIVHNGYCDGR

>Unknown\_protein\_(BI546698.1\_1)

KMKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGL

FXLQDNIVAEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVA

SFXLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVR

QRQXEELCLARQYRLIVHNGYCDGR

>Homo\_sapiens\_RBP4\_(NP\_001310446.1)

XMKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGL

FXLQDNIVAEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVA

SFXLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVR

QRQXEELCLARQYRLIVHNGYCDGR

A close up of a list

Description automatically generated

1. A screenshot of a computer screen

   Description automatically generated
2. Ran trimmed alignment of ‘unknown protein’ from (5) on NCBI’s protein blast against pdb database. Top 3 non-RBP4 hits and information are shown as follows:
3. Sander vitreus Sandercyanin Fluorescent Protein (SFP), Chain A

E-Value = 4e-08, Sequence Identity = 29.41%, PDB Identifier = 5EZ2, Method = X-ray diffraction, Resolution = 1.85 angstroms, Source Organism = Sander vitreus

1. Human Glycodelin, Chain A

E-Value = 4e-06, Sequence Identity = 24.84%, PDB Identifier = 4R0B, Method = X-ray diffraction, Resolution = 2.45 angstroms, Source Organism = Homo sapiens

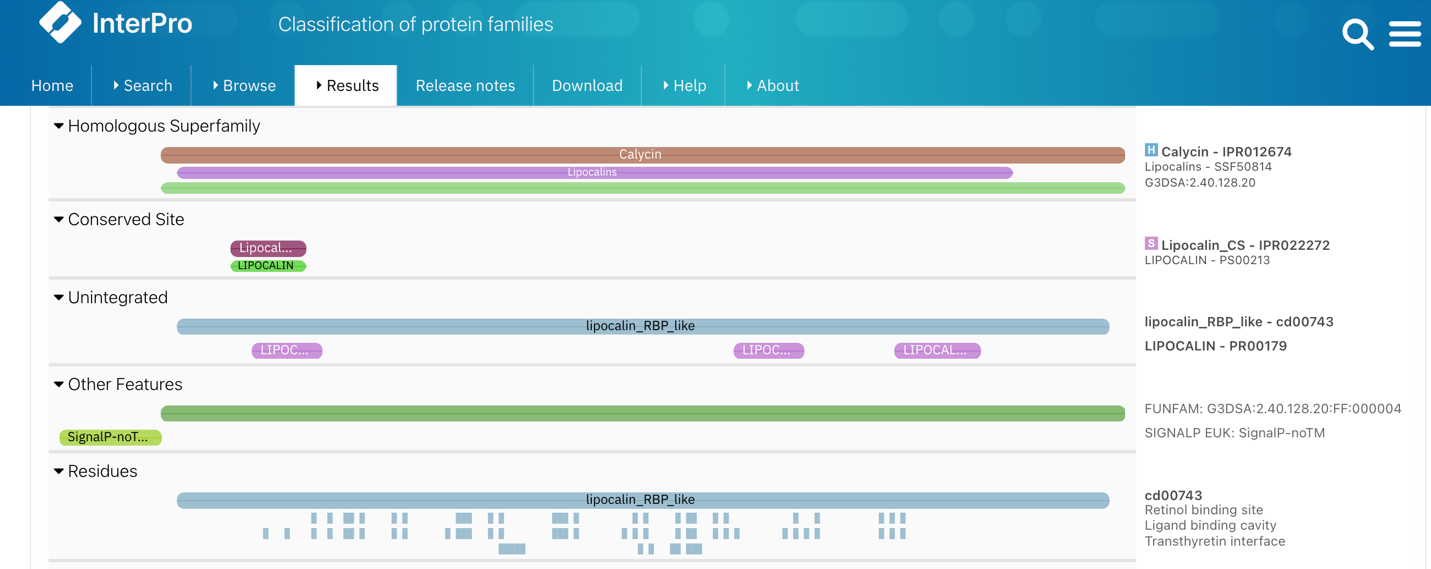
1. Human Apolipoprotein D, Chain A

E-Value = 8e-06, Sequence Identity = 27.85%, PDB Identifier = 2HZQ, Method = X-ray diffraction, Resolution = 1.8 angstroms, Source Organism = Homo sapiens

1. One of 5 of the resulting pdb files were chosen for illustration.

A structure of a protein

Description automatically generatedA structure of a protein

Description automatically generatedUsing InterPro database for conserved domains and function, I found predicted retinol binding sites of my unknown protein (conserved residues between my MSA, boxed in red) and spacefilled with spheres (bottom right figure), and colored the sequence by pLDDT score (low = blue, high = red):

The spacefill clearly shows the retinol binding site within the beta barrel of the protein structure!

1. Yes, there are target associated assays and ligand efficiency data, as shown below for the top hit (RBP) in ChEMBL:

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated