

## Help Manual for DCOMP: Database of Cancer Mutant Protein Domains

1. "Domain Genomic Positions"
2. "Significantly Mutated Domains"
3. "Mutated Domains"

### 1. "Domain Genomic Positions"

#### How to verify the protein domain chromosome positions

a) Under the "Domain genomic positions" menu, type "AMP-Binding" in the search box. The below table will appear after you click the "search" button. To verify the chromosome position, click any ensemble id where the AMP-domain is located. The red circle shows the position where the domain is located.

Ensemble_ID	Chromosome	Start	End	Strand	Domain_length
ENST00000513376.5	4	56378437	56354074	-	324
ENST00000205214.10	4	56384281	56354074	-	444
ENST00000514745.5	4	56384281	56371640	-	113
ENST00000502617.1	4	56384281	56354074	-	444
ENST00000602986.5	4	56378353	56354074	-	296
ENST00000451613.5	4	56384281	56354074	-	444
ENST00000289416.9	16	20770245	20792125	+	414

b) Click the "Show transcript table" button to display all the transcript under the location

**Human** (GRCh38.p13) ▼

Location: 4:56,338,290-56,387,423   Gene: AASDH   Transcript: AASDH-207

**Transcript-based displays**

- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
  - 3D Protein model
- Genetic Variation
  - Variant table
  - Variant image
  - Haplotypes
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
  - Supporting evidence
- ID History
  - Transcript history
  - Protein history

**Transcript: AASDH-207** ENST00000513376.5

Description: aminoadipate-semialdehyde dehydrogenase [Source:HGNC Symbol;Acc:HGNC:23993]

Gene Synonyms: ACSF4, LYS2, NRPS998

Location: **Chromosome 4: 56,338,290-56,387,423 reverse strand**

About this transcript: This transcript has 14 exons, is annotated with 25 domains and features, is associated

Gene: This transcript is a product of gene [ENSG00000157426.14](#) [Show transcript table](#)

**Summary** ⓘ

< AASDH-207 protein coding

Reverse strand

49.13 kb

Statistics: Exons: 14, Coding exons: 13, Transcript length: 3,241 bps, Translation length: 991

c) To display the Pfam domain, click the “Uniprot ID” and it will display the protein sequence information

Gene This transcript is a product of gene [ENSG00000157426.14](#) [Hide transcript table](#)

Show/hide columns (1 hidden)							
Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match
AASDH-201	<a href="#">ENST00000205214.11</a>	3581	<a href="#">1098aa</a>	Protein coding	<a href="#">CCDS35046</a>	<a href="#">Q4L235-1</a>	<a href="#">NM_181806.4</a>
AASDH-210	<a href="#">ENST00000602986.5</a>	3480	<a href="#">945aa</a>	Protein coding	<a href="#">CCDS75126</a>	<a href="#">R4GNB1</a>	-
AASDH-207	<a href="#">ENST00000513376.5</a>	3241	<a href="#">998aa</a>	Protein coding	<a href="#">CCDS68705</a>	<a href="#">Q4L235-2</a>	-
AASDH-202	<a href="#">ENST00000451613.5</a>	3182	<a href="#">857aa</a>	Protein coding	<a href="#">CCDS68706</a>	<a href="#">Q4L235-4</a>	-
AASDH-203	<a href="#">ENST00000502617.1</a>	3004	<a href="#">841aa</a>	Protein coding	<a href="#">CCDS75127</a>	<a href="#">Q4L235-3</a>	-
AASDH-208	<a href="#">ENST00000514745.5</a>	2959	<a href="#">120aa</a>	Nonsense mediated decay	-	<a href="#">D6RJA2</a>	-
AASDH-204	<a href="#">ENST00000503808.5</a>	2949	<a href="#">86aa</a>	Nonsense mediated decay	-	<a href="#">E9PH98</a>	-
AASDH-206	<a href="#">ENST00000510762.1</a>	538	No protein	Processed transcript	-	-	-
AASDH-205	<a href="#">ENST00000510012.1</a>	1222	No protein	Retained intron	-	-	-
AASDH-209	<a href="#">ENST00000514796.1</a>	404	No protein	Retained intron	-	-	-

d) From the UniProt entry “Q4L235-1”, click “View protein in Pfam” under the “Family and domain databases”

**Display** [Help video](#)

[Entry](#)

[Publications](#)

[Feature viewer](#)

[Feature table](#)

---

[None](#)

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (4+)
- Similar proteins

### Family and domain databases

Gene3D	2.130.10.10, 2 hits 3.40.50.12780, 1 hit
InterPro	<a href="#">View protein in InterPro</a> IPR036736, ACP-like_sf IPR020845, AMP-binding_CS IPR000873, AMP-dep_Synth/Lig IPR042099, AMP-dep_Synthh-like_sf IPR009081, PP-bd_ACP IPR025666, PQQ-like_dom IPR018391, PQQ_beta_propeller_repeat IPR002372, PQQ_repeat IPR011047, Quinoprotein_ADH-like_supfam IPR015943, WD40/YVTN_repeat-like_dom_sf
Pfam	<a href="#">View protein in Pfam</a> PF00501, AMP-binding, 1 hit PF00550, PP-binding, 1 hit PF13360, PQQ_2, 1 hit PF13570, PQQ_3, 1 hit
SMART	<a href="#">View protein in SMART</a> SM00564, PQQ, 6 hits
SUPFAM	SSF47336, SSF47336, 1 hit SSF50998, SSF50998, 1 hit
PROSITE	<a href="#">View protein in PROSITE</a> PS00455, AMP_BINDING, 1 hit PS50075, CARRIER, 1 hit

e) The user can view the “AMP-binding” domain architecture and the domain positions in the Pfam database summary

**Protein: ACSF4\_HUMAN (Q4L235)** 1 architecture 1 sequence

---

**Summary**

[Sequence](#)

[Structures](#)

[TreeFam](#)

**Jump to...**

enter ID/acc

**Summary**


This is the summary of UniProt entry [ACSF4\\_HUMAN](#) (Q4L235).

<b>Description:</b>	Beta-alanine-activating enzyme {ECO:0000250 UniProtKB:Q80WC9} EC=6.2.1.-
<b>Source organism:</b>	<a href="#">Homo sapiens (Human)</a> (NCBI taxonomy ID <a href="#">9606</a> )
<b>Length:</b>	1098 amino acids
<b>Reference Proteome:</b>	✓

**Please note:** when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms t some UniProt entries may be removed *after* a Pfam release, these entries will not be removed from Pfam until the *next* Pfam data release.

**Pfam domains**

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will tak the domain boundaries for each of the domains. [More...](#)

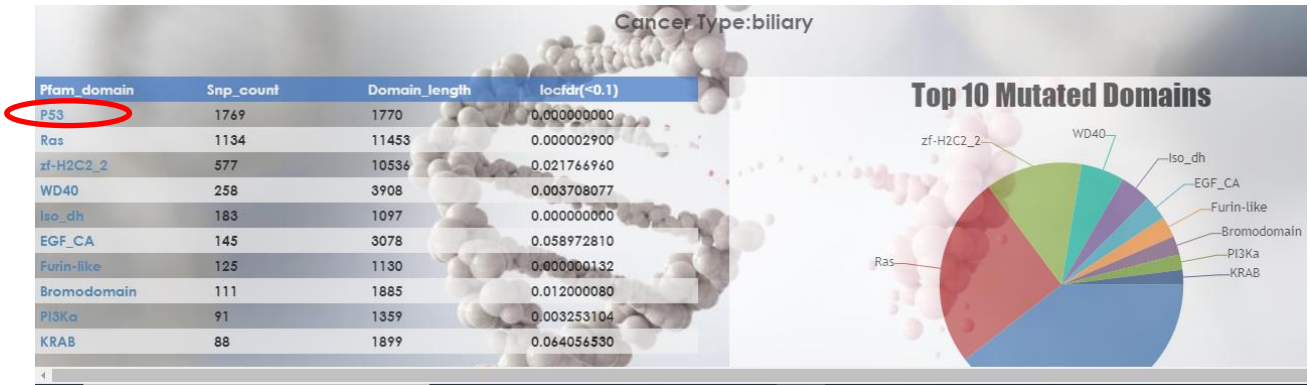


[Download](#) the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
<a href="#">Pfam</a>	<a href="#">AMP-binding</a>	7	450
low_complexity	n/a	159	173

## 2. "Significantly Mutated Domains"

a) Type "biliary" in the search box under the "Significantly mutated domains" and click the "search" button. The table will display the top 10 significantly mutated domains with the pie chart on the left side. The table is sorted based on the "locfdr" values.



b) The Pfam domain details are cross-reference in the above table, and by clicking the domain name, it will direct to the Pfam database as shown below.

### Family: P53 (PF00870)

**Summary**

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Structures

**Jump to...**

enter ID/acc

#### Summary: P53 DNA-binding domain

Pfam includes annotations and additional family information from a range of dif

Wikipedia: P53 | Pfam | InterPro

This tab holds the annotation information that is stored in the Pfam database. replaced by the Wikipedia tab.

**P53 DNA-binding domain** [Provide feedback](#)

This family contains one anomalous member, viz: Zea mays (Q6JAD8). This se contaminant within the Zea mays sampling effort.

#### Internal database links

SCOOP: [CEP1-DNA\\_bind](#)

Similarity to PfamA using HHSearch: [CEP1-DNA\\_bind](#)

#### External database links

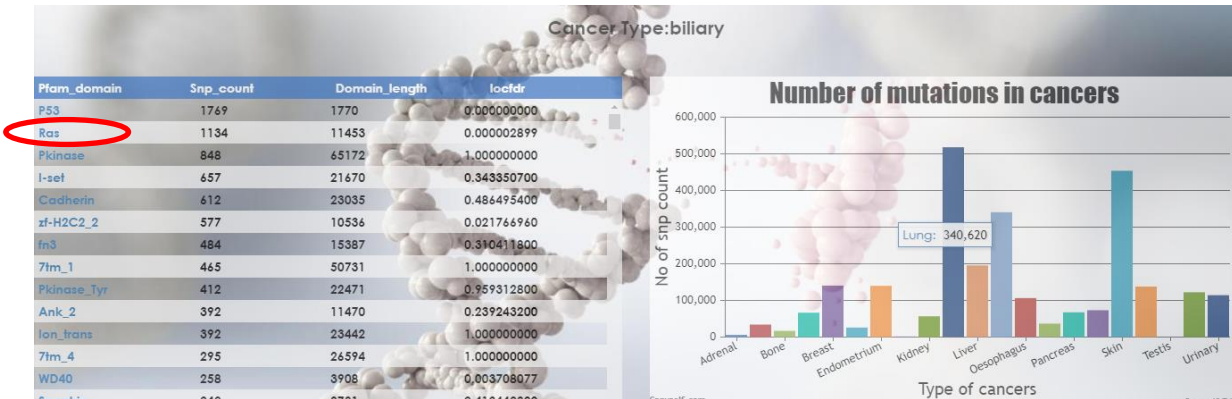
HOMSTRAD: [p53](#)

PROSITE: [PDOC00301](#)

SCOP: [1qzh](#)

### 3. "Mutated Domains"

a) Type "biliary" in the search box under the "Mutated domains" menu and click the "search" button. The table will display all the mutated domains in particular cancer. The bar chart shows the number of mutations across 21 cancers, and the table is sorted based on the "Snp\_count" values.



b) The Pfam domain details are cross-reference in the above table, and by clicking the domain name, it will direct to the Pfam database as shown below.

**Family: Ras (PF00071)**

1543 architectures | 100586 sequences | 0 interactions

**Summary: Ras family**

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

Wikipedia: Ras subfamily | **Wikipedia: Ras superfamily** | Pfam | InterPro

This is the Wikipedia entry entitled "[Ras superfamily](#)". [More...](#)

**Ras superfamily** [Edit Wikipedia article](#)

The **Ras superfamily** is a protein superfamily of small GTPases.<sup>[1]</sup> Members of the superfamily are divided into families and subfamilies based on their structure, sequence and function. The five main families are Ras, Rho, Ran, Rab and Arf GTPases.<sup>[2]</sup> The Ras family itself is further divided into 6 subfamilies: Ras, Ral, Rap, Rheb, Rad and Rit. Miro is a recent contributor to the superfamily. Each subfamily shares the common core G domain, which provides essential GTPase and nucleotide exchange activity.

The surrounding sequence helps determine the functional specificity of the small GTPase, for example the 'Insert Loop', common to the Rho subfamily, specifically contributes to binding to effector proteins such as WASP.

In general, the Ras family is responsible for cell proliferation: Rho for cell morphology, Ran for nuclear transport, and Rab and Arf for vesicle transport.<sup>[3]</sup>