FAM151A, a Menorin Orthlog, is a Kidney Tubule Transmembrane Phosphodiesterase

Thomas Atkins

atkin423@umn.edu

Summary

FAM151A is a gene that encodes a transmembrane protein that contains two domains of unknown function DUF2181. FAM151A has direct orthologs in organisms including mammals, reptiles and amphibians, bony fish, and other Eumetazoan invertebrates, but has no orthologs in birds. The FAM151 family also includes FAM151B, which has one DUF2181 domain and no transmembrane region. The FAM151 family has orthologs in nematodes and arthropods from before a gene duplication event, one of which is menorin, a *C. elegans* protein involved in dendrite morphogenesis. Sequence conservation analysis of hypothesized active sites suggests that only the first DUF2181 domain of FAM151A has biochemical function. FAM151A also contains a SNP, rs11206394, where individuals homozygous for the minor allele were found to have a 40% reduction in the odds of developing colorectal cancer, purportedly through the SNP's effect on a miRNA binding site.

The mRNA transcript of FAM151A is expressed highly in kidney, small intestine, and liver tissues, while immunohistochemical staining data indicate the FAM151A protein is only highly expressed in proximal kidney tubules. The mRNA expression pattern is speculated to be a result of HNF1, a transcription factor expressed in a similar pattern to FAM151A predicted to bind to the FAM151A promoter. Protein expression is hypothesized to be a result of competition between two proteins that bind to an unpaired conserved section of the 3' UTR of FAM151A, ZFP36 and EIF4B, which act as a degradation signal and translation initiation signal, respectively. FAM151A is underexpressed in carcinomic kidney tissue and hepatitic liver tissue, but not differentially expressed under diabetic conditions.

FAM151A is strongly predicted to be localized to the cell membrane, with the two DUF2181s residing outside of the cell. The tertiary structure of the protein is predicted with high confidence by AlphaFold2, and agrees with experimental structure of homologous domains. DUF2181 is part of the GDPD/PLCD superfamily, a class of enzymes that hydrolyze phosphodiester bonds. Other proteins, such as GDPD5 and ENPP6, are known as transmembrane phosphodiesterases acting in the kidney and brain (in which FAM151A is expressed), and bind to a glycerophosphocholine substrate, which we suggest as a potential substrate of FAM151A. In *C. elegans*, sax-7 is a known binding partner of menorin, so its human ortholog L1CAM is a potential interaction partner with proteins in the FAM151 family.

Contents

Annotated Conceptual Translation

```
agagcagaccaggcccggtggagaattaggtgctgctgggagctcctgcctcccacagga  60
ttccagctgcagggagcctcagggactctgggccgcacggagttgggggcattccccaga 120
gagcgtcgccatggtctgcagggagcagttatcaaagaatcaggtcaagtgggtgtttgc 180
          M  V  C  R  E  Q  L  S  K  N  Q  V  K  W  V  F  A   17 N-term; TMEM region

cggcattacctgtgtgtctgtggtggtcattgccgcaatagtccttgccatcaccctgcg 240
  G  I  T  C  V  S  V  V  V  I  A  A  I  V  L  A  I  T  L  R   37

gcggccaggctgtgagctggaggcctgcagccctgatgccgacatgctggactacctgct 300 ex1|ex2; rs17399297G>A
  R  P  G  C  E  L  E  A  C  S  P  D  A  D  M  L  D  Y  L  L   57

gagcctgggccagatcagccggcgagatgccttggaggtcacctggtaccacgcagccaa 360 rs147294199G>A
  S  L  G  Q  I  S  R  R  D  A  L  E  V  T  W  Y  H  A  A  N   77 DUF2181; active site

cagcaagaaagccatgacagctgccctgaacagcaacatcacagtcctggaggctgacgt 420 ex2|ex3; rs142814457C>G
  S  K  K  A  M  T  A  A  L  N  S  N  I  T  V  L  E  A  D  V   97 active sites

caatgtagaagggctcggcacagccaatgagacaggagttcccatcatggcacacccccc 480
  N  V  E  G  L  G  T  A  N  E  T  G  V  P  I  M  A  H  P  P  117 active site
                                                                   pre. N-linked gly.
cactatctacagtgacaacacactggagcagtggctggacgctgtgctgggctcttccca 540
  T  I  Y  S  D  N  T  L  E  Q  W  L  D  A  V  L  G  S  S  Q  137

aaaggcatcaaactggacttcaagaacatcaaggcagtgggcccctccctggacctcct 600 ex3|ex4
  K  G  I  K  L  D  F  K  N  I  K  A  V  G  P  S  L  D  L  L  157 active site

gcggcagctgacagaggaaggcaaagtccggcggcccatatggatcaacgctgacatctt 660
  R  Q  L  T  E  E  G  K  V  R  R  P  I  W  I  N  A  D  I  L  177

aaagggccccaacatgctcatctcaactgaggtcaatgccacacagttcctggccctggt 720 ex4|ex5
  K  G  P  N  M  L  I  S  T  E  V  N  A  T  Q  F  L  A  L  V  197 pre. N-linked gly.

ccaggagaagtatcccaaggctaccctatctccaggctggaccaccttctacatgtccac 780 rs147577669A>C
  Q  E  K  Y  P  K  A  T  L  S  P  G  W  T  T  F  Y  M  S  T  217

gtccccaaacaggacgtacacccaagccatggtggagaagatgcacgagctggtgggagg 840
  S  P  N  R  T  Y  T  Q  A  M  V  E  K  M  H  E  L  V  G  G  237 pre. N-linked gly.

agtgccccagagggtcaccttccctgtacggtcttccatggtgcgggctgcctggcccca 900
  V  P  Q  R  V  T  F  P  V  R  S  S  M  V  R  A  A  W  P  H  257

cttcagctggctgctgagccaatctgagaggtacagcctgacgctgtggcaggctgcctc 960 ex5|ex6
  F  S  W  L  L  S  Q  S  E  R  Y  S  L  T  L  W  Q  A  A  S  277 active site
```

```
ggaccccatgtcggtggaagatctgctctacgtccgggataacactgctgtccaccaagt 1020
  D  P  M  S  V  E  D  L  L  Y  V  R  D  N  T  A  V  H  Q  V  297

ctactatgacatctttgagcctctcctgtcacagttcaagcagctggccttgaatgccac 1080 ex6|ex7
  Y  Y  D  I  F  E  P  L  L  S  Q  F  K  Q  L  A  L  N  A  T  317 active site
                                                                  pre. N-linked gly.
acggaaaccaatgtactacacgggaggcagcctgatccctcttctccagctgcctgggga 1140
  R  K  P  M  Y  Y  T  G  G  S  L  I  P  L  L  Q  L  P  G  D  337

tgacggtctgaatgtggagtggctggttcctgacgtccagggcagcggtaaaacagcaac 1200
  D  G  L  N  V  E  W  L  V  P  D  V  Q  G  S  G  K  T  A  T  357 DUF2181

aatgaccctcccagacacagaaggcatgatcctgctgaacactggcctcgagggaactgt 1260 ex7|ex8
  M  T  L  P  D  T  E  G  M  I  L  L  N  T  G  L  E  G  T  V  377

ggctgaaaaccccgtgcccattgttcatactccaagtggcaacatcctgacgctggagtc 1320
  A  E  N  P  V  P  I  V  H  T  P  S  G  N  I  L  T  L  E  S  397

ctgcctgcagcagctggccacacatcccggacactggggcatccatttgcaaatagcgga 1380 rs1368883C>T
  C  L  Q  Q  L  A  T  H  P  G  H  W  G  I  H  L  Q  I  A  E  417

gcccgcagccctccggccatccctggccttgctggcacgcctctccagccttggcctctt 1440
  P  A  A  L  R  P  S  L  A  L  L  A  R  L  S  S  L  G  L  L  437

gcattggcctgtgtgggttggggccaaaatctcccacgggagttttcggtccccggcca 1500
  H  W  P  V  W  V  G  A  K  I  S  H  G  S  F  S  V  P  G  H  457

tgtggctggcagagagctgcttacagctgtggctgaggtcttcccccacgtgactgtggc 1560
  V  A  G  R  E  L  L  T  A  V  A  E  V  F  P  H  V  T  V  A  477

accaggctggcctgaggaggtgctgggcagtggctacagggaacagctgctcacagatat 1620
  P  G  W  P  E  E  V  L  G  S  G  Y  R  E  Q  L  L  T  D  M  497

gctagagttgtgccaggggctctggcaacctgtgtccttccagatgcaggccatgctgct 1680
  L  E  L  C  Q  G  L  W  Q  P  V  S  F  Q  M  Q  A  M  L  L  517

gggccacagcacagctggagccataggcaggctgctggcatcctcccccgggccaccgt 1740 rs11206394G>C
  G  H  S  T  A  G  A  I  G  R  L  L  A  S  S  P  R  A  T  V  537 rs41297135C>G

cacagtggagcacaacccagctggggcgactatgcctctgtgaggacagcattgctggc 1800 rs2289015G>A
  T  V  E  H  N  P  A  G  G  D  Y  A  S  V  R  T  A  L  L  A  557

agctagggctgtggacaggacccgagtctactacaggctaccccagggctaccacaagga 1860
  A  R  A  V  D  R  T  R  V  Y  Y  R  L  P  Q  G  Y  H  K  D  577
```

cttgctggctcatgttggtagaaactgagcacccaggggtggtgggccagcggacctcag 1920
L L A H V G R N * *586* Stop codon

ggcggaggcttcccacggggaggcaggaagaataaaggtctttggctttctcca[aaa] 1975 Poly-A signal
Poly-A tail

Key:
Bold: Conserved in all 20 orthologs
Pink: Active site as determined by Findlay et. al.
Brown: Predicted N-linked glycosylation site
Orange: SNPs
Salmon: SNPs with associated publications
Blue: exon-exon boundaries

DNA

Gene Structure

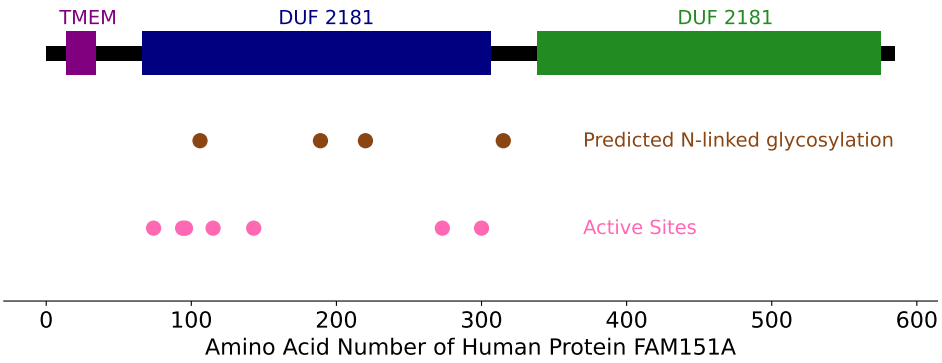


Figure 1: Schematic diagram of protein FAM151A domains and important amino acids.

Homo sapiens FAM151A (NCBI Accession: NM_176782.3/NP_788954.2) is a ~14 kbp gene located in cytogenetic band 1p32.3 that encodes a 1975 bp mRNA transcript that translates to a 585 amino acid protein.[1,2] The protein contains a 20 aa long helical transmembrane region, and two domains of unknown function DUF2181, as seen in Figure 1.[3] The gene has 8 exons, the last and longest of which composes roughly half the mRNA transcript.[4] No alternative splicings are known.[5]

Overview of Orthologs and Paralogs

FAM151 Family

Figure 2 shows 20 direct orthologs of the FAM151A protein in organisms including mammals (72%-98% identity), reptiles and amphibians (43%-50% identity), bony fish (41%-46% identity), non-vertebrate chordates (28%-30% identity), and non-chordates in Eumetazoa (21%-28% identity). The data were collected using NCBI's BLAST, TimeTree, and EMBOSS NEEDLE.[6,7,8] BLASTing and BLATing multiple FAM151A orthologs against all birds returned no results, suggesting that FAM151A is no longer present in Aves.[9]

[1]NCBI Protein (National Center for Biotechnology Information Protein Database) entry on FAM151A. https://www.ncbi.nlm.nih.gov/protein/NP_788954.2.

[2]Genecards entry on FAM151A. https://www.genecards.org/cgi-bin/carddisp.pl?gene=FAM151A.

[3]UniProt (Universal Protein Resource) entry on FAM151A. https://www.uniprot.org/uniprot/Q8WW52.

[4]NCBI Protein (National Center for Biotechnology Information Nucleotide Database) entry on FAM151A. https://www.ncbi.nlm.nih.gov/nuccore/NM_176782.3.

[5]AceView entry on FAM151A. https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&term=fam151a&submit=Go.

[6]NCBI Basic Local Alignment Search Tool. https://blast.ncbi.nlm.nih.gov/Blast.cgi.

[7]TimeTree: The Timescale of Life. http://www.timetree.org/.

[8]EMBOSS NEEDLE. https://www.ebi.ac.uk/Tools/psa/emboss_needle/.

[9]Blast-like Alignment Tool. https://genome.ucsc.edu/cgi-bin/hgBlat.

	Genus/species	Common name	Taxonomic Group	Med. Date of Divergence (mya)	Accession Number	Sequence Length	Sequence Identity	Sequence Similarity
Primata	*Homo sapiens*	Human	Hominidae	**0**	NP_788954.2	585	**100.0%**	100.0%
	Pan troglodytes	Chimpanzee	Hominidae	**6**	XP_016774503.1	585	**98.1%**	98.8%
	Papio anubis	Olive baboon	Cercopithecidae	**29**	XP_003891985.2	585	**95.2%**	96.9%
Mammalia	*Mus musculus*	Mouse	Rodentia	**89**	NP_666261.1	608	**68.6%**	79.1%
	Equus caballus	Horse	Perissodactyla	**94**	XP_001488568.4	588	**75.3%**	84.7%
	Vicugna pacos	Alpaca	Artiodactyla	**94**	XP_006200587.1	589	**72.8%**	83.5%
Reptilia	*Chrysemys picta bellii*	Painted turtle	Testudines	**318**	XP_005284924.2	590	**49.5%**	65.1%
	Alligator sinensis	Chinese alligator	Crocodilia	**318**	XP_006025880.1	585	**47.5%**	64.2%
Amphibia	*Rhinatrema bivittatum*	Two-lined caecilian	Gymnophiona	**352**	XP_029474719.1	592	**46.4%**	62.0%
	Xenopus laevis	African clawed frog	Anura	**352**	XP_018116415.1	578	**46.1%**	64.0%
	Bufo bufo	Common toad	Anura	**352**	XP_040262912.1	576	**42.9%**	60.9%
Vertabrata	*Cyprinus carpio*	Common carp	Actinopterygii	**433**	XP_042575185.1	614	**41.5%**	58.4%
	Danio rerio	Zebrafish	Actinopterygii	**433**	NP_001093565.1	599	**41.4%**	57.2%
	Rhincodon typus	Whale shark	Chondrichthyes	**465**	XP_020366386.1	600	**46.8%**	62.1%
Chordata	*Styela clava*	Stalked sea squirt	Tunicata	**603**	XP_039273176.1	597	**30.2%**	49.0%
	Ciona intestinalis	Sea squirt	Tunicata	**603**	XP_002121148.3	639	**27.9%**	45.5%
	Branchiostoma floridae	Florida lancelet	Cephalochordata	**637**	XP_035660277.1	646	**28.2%**	41.9%
Eumetazoa	*Lytechinus variegatus*	Green sea urchin	Bilateria	**627**	XP_041464769.1	544	**22.6%**	39.5%
	Stylophora pistillata	Hood coral	Cnidaria	**687**	PFX14114.1	557	**20.7%**	34.2%
	Lingula anatina	Brachiopod	Bilateria	**736**	XP_013411281.1	598	**27.6%**	47.1%

Figure 2: 20 FAM151A orthologs and related properties.

In humans, FAM151A has a processed pseudogene on Chromosome 3, ENSG00000234805.[10]

The FAM151 family also includes FAM151B, which has one DUF2181 and no transmembrane region, suggesting a different function from FAM151A.[11] In humans, FAM151B has 21%/29% sequence identity/similarity to FAM151A.[12] FAM151B has direct orthologs in all organisms for which FAM151A has orthologs. Additionally, FAM151B has direct orthologs in Aves, in contrast to FAM151A.

Menorin

Genes in the FAM151 family are homologs of the well-characterized *C. elegans* menorin (MNR-1), a dendritic branching protein involved in the creation of higher-order branches by forming a complex with sax-7.[13,14] In addition to the FAM151 family, MNR-1 has orthologs in Nematoda and Arthropoda that can be found using BLAST.

[10]GeneCards entry on ENSG00000234805. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ENSG00000234805.

[11]NCBI Protein (National Center for Biotechnology Information Protein Database) entry on FAM151B. https://www.ncbi.nlm.nih.gov/protein/NP_788954.2.

[12]EMBOSS Needle. https://www.ebi.ac.uk/Tools/psa/emboss_needle/.

[13]Findlay, A. S., McKie, L., Keighren, M., Clementson-Mobbs, S., Sanchez-Pulido, L., Wells, S., Cross, S. H., & Jackson, I. J. (2020a). Fam151b, the mouse homologue of C.elegans menorin gene, is essential for retinal function. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-019-57398-4.

[14]Salzberg, Y., Diaz-Balzac, C. A., Ramirez-Suarez, N. J., Attreed, M., Tecle, E., Desbois, M., Kaprielian, Z., & Bulow, H. E. (2013). Skin-derived cues control arborization of sensory dendrites in Caenorhabditis elegans. *Cell*, *155*(2), 308–320. https://doi.org/10.1016/j.cell.2013.08.058.

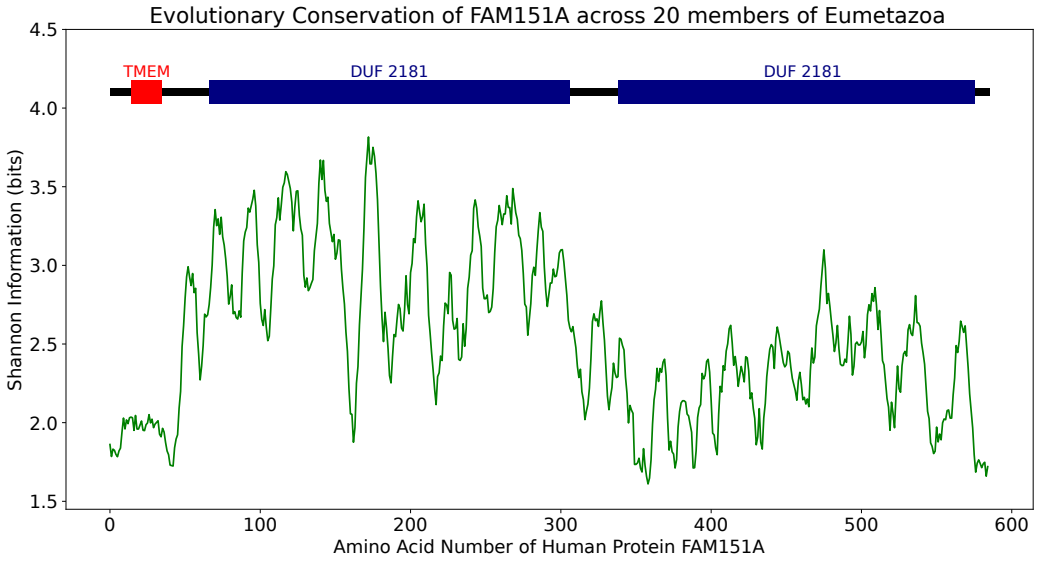


Figure 3: Shannon information content (in bits) of each amino acid in protein FAM151A aligned over 20 members of Eumetazoa (rolling average of 5 amino acids).

DUF2181 Domain

The defining characteristic of the FAM151/Menorin family of genes is the DUF2181, revealed to be part of the GDPD/PLCD (glycerophosphoryldiester phosphodiesterase/PLC-like phosphodiesterases) superfamily through homology detection, discussed in further detail in the section on the FAM151A protein.[15]

Sequence Conservation of FAM151A Orthologs

Figure 3 displays a plot of amino acid conservation across the 20 strict FAM151A orthologs in Figure 2 as measured by the Shannon information metric

$$I_b = \lg(20) - \sum_i p_{i,b} \lg p_{i,b}$$

where $p_{i,b}$ is the frequency of base i at position b.[16] In this plot, we see that the first DUF2181 is far more conserved than the second, supporting researchers' speculation that the second DUF2181 is nonfunctional.[17] Furthermore, the transmembrane region is the least conserved domain of the protein by this metric. However, our metric does not account for amino acid chemistry, suggesting that the region could still function as a transmembrane domain.

[15]PFAM Entry on GDPD. http://pfam.xfam.org/family/gdpd.

[16]Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[17]Findlay, A. S., McKie, L., Keighren, M., Clementson-Mobbs, S., Sanchez-Pulido, L., Wells, S., Cross, S. H., & Jackson, I. J. (2020b). Fam151b, the mouse homologue of c.elegans menorin gene, is essential for retinal function. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-019-57398-4.

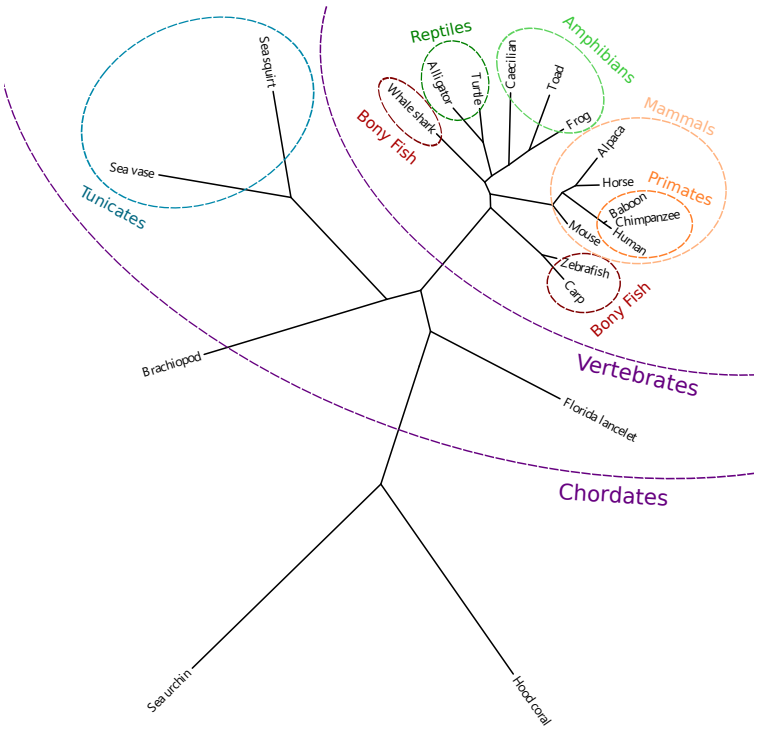


Figure 4: Unrooted phylogenetic tree displaying FAM151A ancestry.

Three multiple sequence alignments are presented in Appendix B. The first is an alignment of human protein FAM151A with the corresponding FAM151A proteins of all vertebrates listed in Figure 2 (close orthologs), the second is an alignment of human protein FAM151A with the corresponding FAM151A proteins of all invertebrates listed in Figure 2 (distant orthologs). The third is an alignment of proteins FAM151A and FAM151B in humans, mice, toads, and zebrafish, with distinguishing amino acids highlighted in red, where we see that FAM151B's DUF2181 corresponds with the first DUF2181 of FAM151A. All were created using Clustal Omega.[18] Long stretches of amino acid residues with no equivalent were found in all three non-chordates, and omitted for brevity.

Evolutionary History of FAM151A

Figures 4 displays an unrooted phylogenetic tree created from a global multiple sequence alignment of FAM151A orthologs, each containing two full DUF2181s. In general, organisms were labelled a combination of the first letter of the genus and the first two letters of the species name, a full table of organisms and labels can be found in Appendix A.

Figure 5 presents a wider evolutionary tree of FAM151A homologs (presented in a rectangular form for ease of clade distinction in Figure 6). The tree is generated from an alignment of only DUF2181s (the first if an ortholog has two), shown in full in Appendix C (the amino acid coloring scheme is described later). Labels follow the format `Species_type` where the

[18]Clustal Omega. https://www.ebi.ac.uk/Tools/msa/clustalo/.

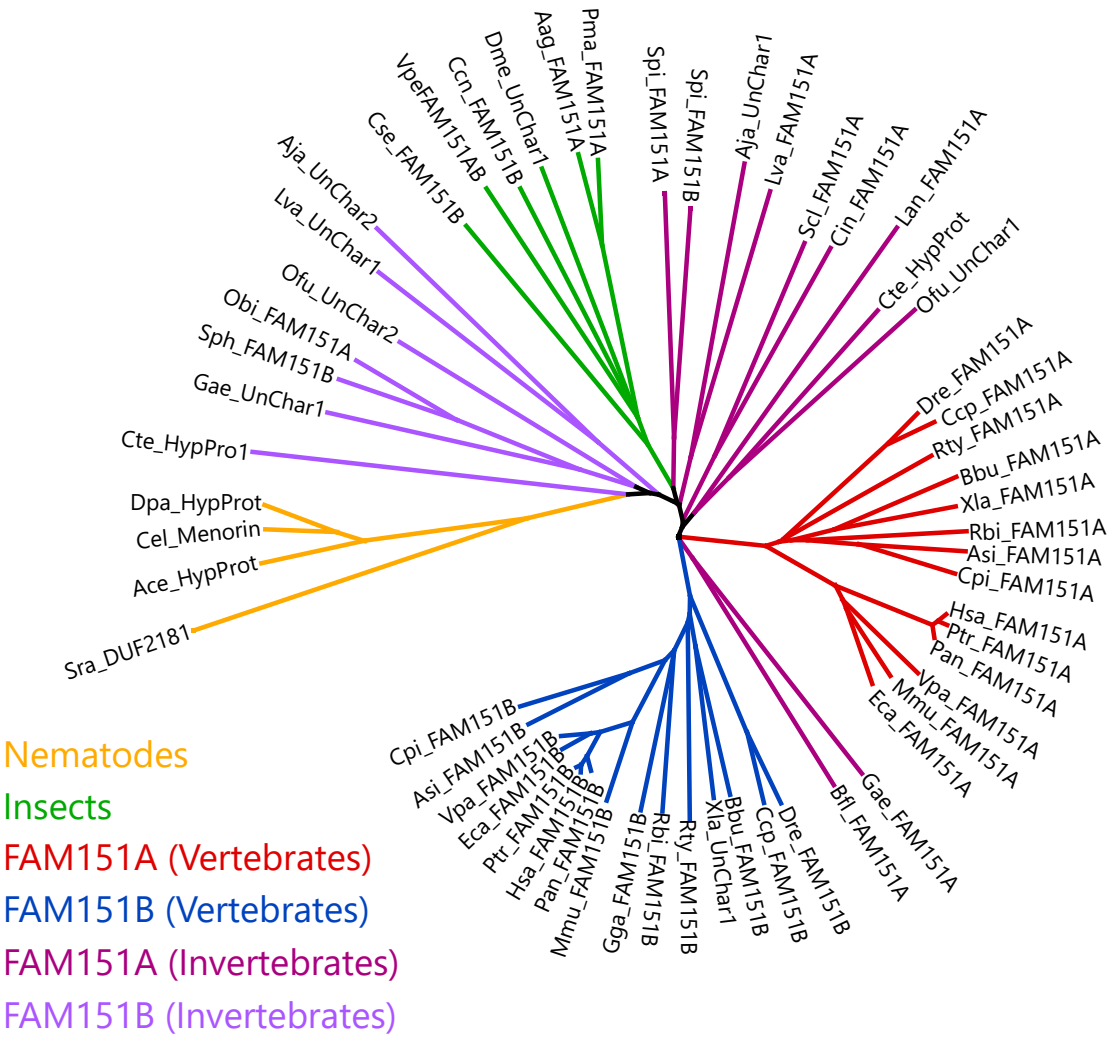


Figure 5: Unrooted phylogenetic tree displaying DUF2181 ancestry.

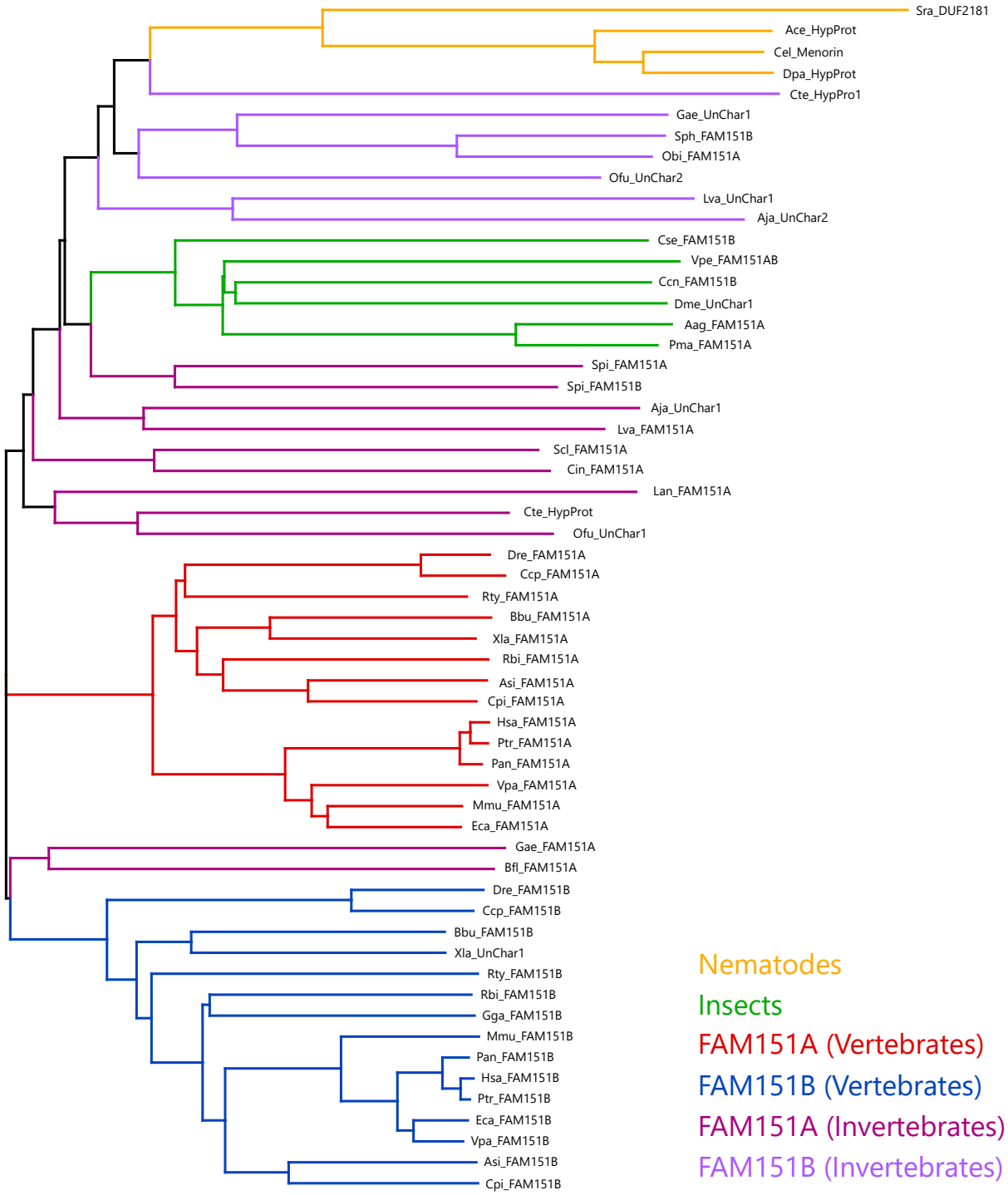


Figure 6: Unrooted phylogenetic tree displaying DUF2181 ancestry. The tree is presented in a rectangular format for ease of viewing.

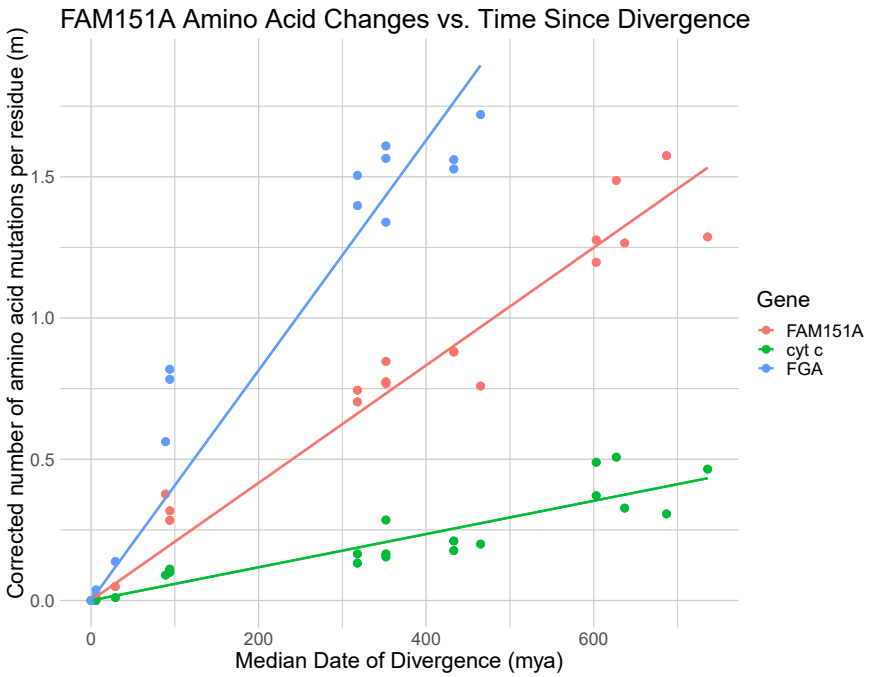


Figure 7: Graph showing mutation rate of FAM151A in comparison to mutation rates of cytochrome c and fibrinogen alpha chain.

species label can be found in Appendix A and `type` is an abbreviation of the name given to the protein containing that DUF2181 by NCBI Protein.

From this alignment, we see that vertebrate FAM151A and FAM151B orthologs cluster nicely, suggesting that the gene duplication event occurred before the emergence of vertebrates. Furthermore, we see that in non-chordates who do not belong to Nematoda or Arthoropoda, there appear to be two major groups, with organisms generally having two copies of the protein, one in each cluster. The exception is *Stylophora pistillata*, whose closely related FAM151A and FAM151B proteins provide evidence of a gene conversion event. We also see clustering of the DUF2181s of nematodes and arthropods, so we present the likely history of the gene as follows (using TimeTree for dating).[19] A gene belonging to the FAM151 family first appeared between 700 and 800 million years ago, at the latest, when arthropods and nematodes diverged from other members of Eumetazoa. Soon after this divergence, FAM151 underwent a gene duplication event, splitting into FAM151A and FAM151B before 700 mya.

Figure 7 contrasts the evolution rate of FAM151A with those of fibrinogen alpha chain and cytochrome c. FAM151A is neither evolving as quickly as fibrinogen alpha chain nor as slowly as cytochrome c, suggesting that FAM151A is likely not under evolutionary pressure to evolve quickly, nor part of a large complex that discourages mutations. From the chart, we hypothesize that FAM151B, with an m-value of 1.54, diverged from FAM151A around 900 million years ago, in the same timescale as our previous conclusion.

[19]TimeTree: The Timescale of Life. http://www.timetree.org/.

miRNA	Reference Energy (kCal/mol)	Variant Energy (kCal/mol)
hsa-miR-4706	-31.21	Not predicted
hsa-miR-4525	-25.72	-29.05
hsa-miR-4739	Not predicted	-27.86
hsa-miR-214-5p	-26.76	Not predicted

Table 1: miRNAs predicted to differentially bind to region containing rs11206394.

FAM151A Mutations

Mutation Summary

A search of dbSNP revealed 8 SNPs that encode nonsynomomous mutations, and were either reported in ClinVar or had a minor allele frequency greater than 5%.[20] These are labelled in the annotated conceptual translation. In mice, while knocking out FAM151B was associated with loss of retinal function, but there was no discernable retinal phenotype associated with a full knockout of FAM151A.[21]

Clinically Relevant SNP rs11206394

SNP rs11206394 is a missense mutation found in the FAM151 gene, where a guanine is changed into a cytosine, changing a glycine into an alanine. The minor allele ocurred with 13.7% frequency in 5008 genomes sequenced from individuals during the 1000 Genomes Project.[22] In a study examining the impact of mutations to miRNA binding sites in 3' UTRs, individuals homozygous for the minor allele were found to have a 40% reduction of odds of developing colorectal cancer ($p = 0.011$).[23]

As we hypothesize that the second DUF2181 that contains the SNP is nonfunctional, we assess the impact of the mutation on miRNA binding sites. Neither TargetScan nor miRDB found any miRNA binding sites for the FAM151A mRNA transcript.[24,25] Thus, to examine potential miRNA binding sites that could be impacted by the variant, miRanda was used to predict potential binding sites from all *Homo sapiens* miRNAs in the miRBase database in a region 28bp upstream and downstream of rs11206394 for both the reference and variant alleles.[26,27] The results are summarized in Table 1. Because the SNP is located in the 3' UTR

[20]dbSNP (Single Nucleotide Polymorphism Database) search for FAM151A. https://www.ncbi.nlm.nih.gov/snp/?term=FAM151A.

[21]Findlay, A. S., McKie, L., Keighren, M., Clementson-Mobbs, S., Sanchez-Pulido, L., Wells, S., Cross, S. H., & Jackson, I. J. (2020a). Fam151b, the mouse homologue of C.elegans menorin gene, is essential for retinal function. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-019-57398-4.

[22]dbSNP (Single Nucleotide Polymorphism Database) entry on rs11206394. https://www.ncbi.nlm.nih.gov/snp/rs11206394.

[23]Kang, B. W., Jeon, H.-S., Chae, Y. S., Lee, S. J., Park, J. S., Choi, G. S., & Kim, J. G. (2016). Impact of genetic variation in microrna-binding site on susceptibility to colorectal cancer. *Anticancer Research*, *36*(7), 3353–3361. https://ar.iiarjournals.org/content/36/7/3353.

[24]TargetScanHuman. http://www.targetscan.org/vert_80/.

[25]miRDB: MicroRNA Target Prediction Database. http://www.mirdb.org/.

[26]Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., & Marks, D. S. (2003). MicroRNA targets in Drosophila. *Genome Biology*, *5*(1), R1. https://doi.org/10.1186/gb-2003-5-1-r1.

[27]Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2018). miRBase: from microRNA sequences to

		rs2289015	
		C	T
rs11206394	C	4322	1
	G	284	401

Table 2: Table of allele frequency of in rs11206394 and rs2289015 in 5008 genomes sequenced by the Human Genome Project.

of gene ACOT11, which overlaps the last exon of FAM151A, one potential hypothesis is the the SNP affects ACOT11 expression (ACOT11 is known to be expressed in the colon).[28]

However, the discrepancy in cancer rates could also be explained by other SNPs associated with rs11206394 via linkage disequilibrium. rs11206394 has a linkage disequilibrium coefficient of 0.9971 ($p < 0.0001$) with SNP rs2289015, located 60bp downstream, also in the 3' UTR of FAM151A, as measured by LDlink using data from the 1000 Genomes project.[29] The data are shown in Table 2. This suggests either SNP could be impacting colorectal cancer rates. Yet GDPD5, discussed later as being similar to FAM151A, is also known to have a 3' UTR miRNA binding site (miR-195-5p) that increases chemosensitivity and cell apoptosis in CRC cells.[30]

function. *Nucleic Acids Research*, *47*(D1), D155–D162. https://doi.org/10.1093/nar/gky1141.

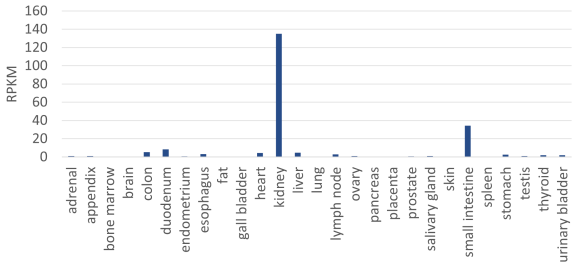
[28]The Human Protein Atlas entry on ACOT11. https://www.proteinatlas.org/ENSG00000162390-ACOT11/tissue.

[29]Machiela, M. J., & Chanock, S. J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, *31*(21), 3555–3557. https://doi.org/10.1093/bioinformatics/btv402.

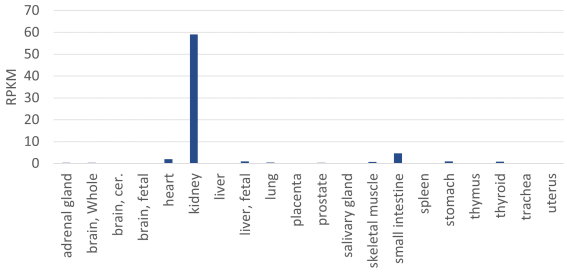
[30]Feng, C., Zhang, L., Sun, Y., Li, X., Zhan, L., Lou, Y., Wang, Y., Liu, L., & Zhang, Y. (2018). GDPD5, a target of miR-195-5p, is associated with metastasis and chemoresistance in colorectal cancer. *Biomedicine & Pharmacotherapy*, *101*, 945–952. https://doi.org/10.1016/j.biopha.2018.03.028.

RNA

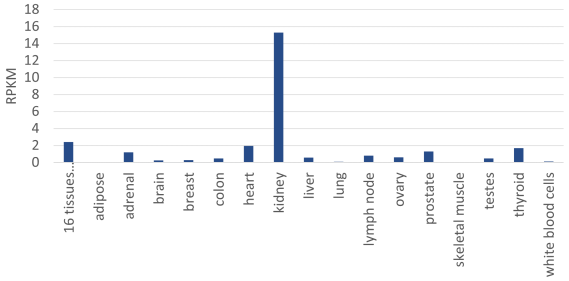
Expression Patterns of FAM151A



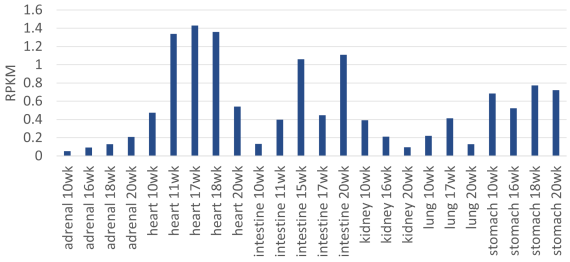
(a) HPA RNA-seq normal tissues.



(b) RNA-seq of total RNA from 20 human tissues.

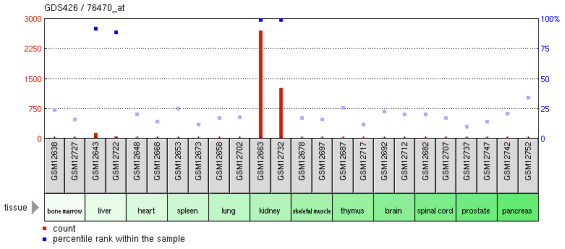


(c) Illumina bodyMap2 transcriptome.

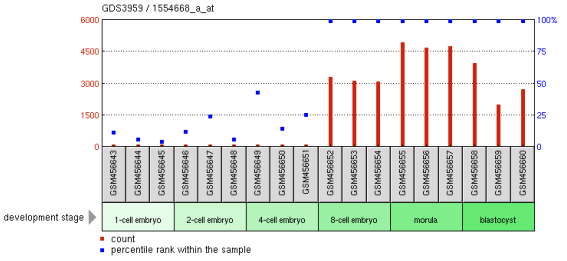


(d) Tissue-specific circular RNA induction during human fetal development.

Figure 8: FAM151A Expression Patterns from NCBI Gene.



(a) Per-Tissue expression of FAM151A.



(b) FAM151A early development expression.

Figure 9: Expression patterns of FAM151A from NCBI GEO.

Figure 8 shows expression patterns of FAM151A in four different experiments from NCBI Gene, and Figure 9 presents expression patterns across tissues and stages of development from NCBI GEO.[31,32] In Figures 8a, 8b, 8c, and 9a, we see that FAM151A is very highly expressed in adult kidney tissue and expressed at a lower level in small intestine and liver

[31]NCBI Gene entry on FAM151A. https://www.ncbi.nlm.nih.gov/gene/338094.
[32]NCBI Geo Search for FAM151A. https://www.ncbi.nlm.nih.gov/geoprofiles/?term=FAM151A.

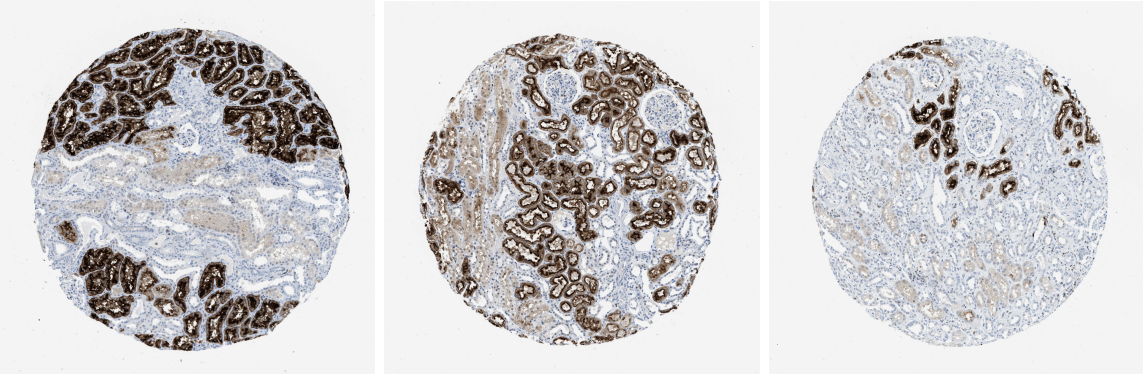


Figure 10: Immunnohistochemical staining of FAM151A in human kidney tissue reveals high expression in tubules but not in glomeruli.

tissues. Figure 8d suggests that FAM151A is not highly expressed in embryonic tissues compared to adult tissue (with reads under 2 RPKM in fetal tissues and above 15 RPKM in adult kidney tissue), while Figure 9b shows FAM151A is first significantly expressed at the stage of development when an embryo contains 8 cells.

In Figure 10, we see three slices of adult kidney tissue that have undergone H&E and DAB staining with FAM151A antibodies from the Human Protein Atlas.[33] From these, we see high expression of FAM151A in the tubules of the kidney, and low to no expression in the glomeruli. The glomerulus is primarily responsible for blood filtration, while the tubules are responsible for reabsorption of filtered substances and transport.[34] Furthermore, there is some evidence that FAM151A protein expression is localized specifically to the luminal side of proximal tubule.[35,36]

The Human Protein Atlas also presents the data shown in Figure 11, where we see that FAM151A is transcribed in the small intestine, liver, and kidney, but only translated to protein at high levels in the kidney.

Differential Expression Conditions of FAM151A

Figures 12-14 display data on differential expression of FAM151A from NCBI GEO.[37] Figure 12 displays differential mRNA expression of FAM151A in normal and cancerous kidney tissues. In the figure, we clearly see that FAM151A is more highly expressed in normal tissue than in tumor tissue, and that low expression of FAM151A consistent across studied

[33]The Human Protein Atlas entry on FAM151A. https://www.proteinatlas.org/ENSG00000162391-FAM151A.

[34]Wallace, M. A. (1998). Anatomy and physiology of the kidney. *AORN Journal*, *68*(5), 799–820. https://doi.org/https://doi.org/10.1016/S0001-2092(06)62377-6.

[35]Habuka, M., Fagerberg, L., Hallström, B. M., Kampf, C., Edlund, K., Sivertsson, Å., Yamamoto, T., Pontén, F., Uhlén, M., & Odeberg, J. (2015). The kidney transcriptome and proteome defined by transcriptomics and antibody-based profiling. *PLOS ONE*, *9*(12), 1–19. https://doi.org/10.1371/journal.pone.0116125.

[36]The Human Protein Atlas entry on FAM151A. https://www.proteinatlas.org/ENSG00000162391-FAM151A.

[37]NCBI Geo Search for FAM151A. https://www.ncbi.nlm.nih.gov/geoprofiles/?term=FAM151A.

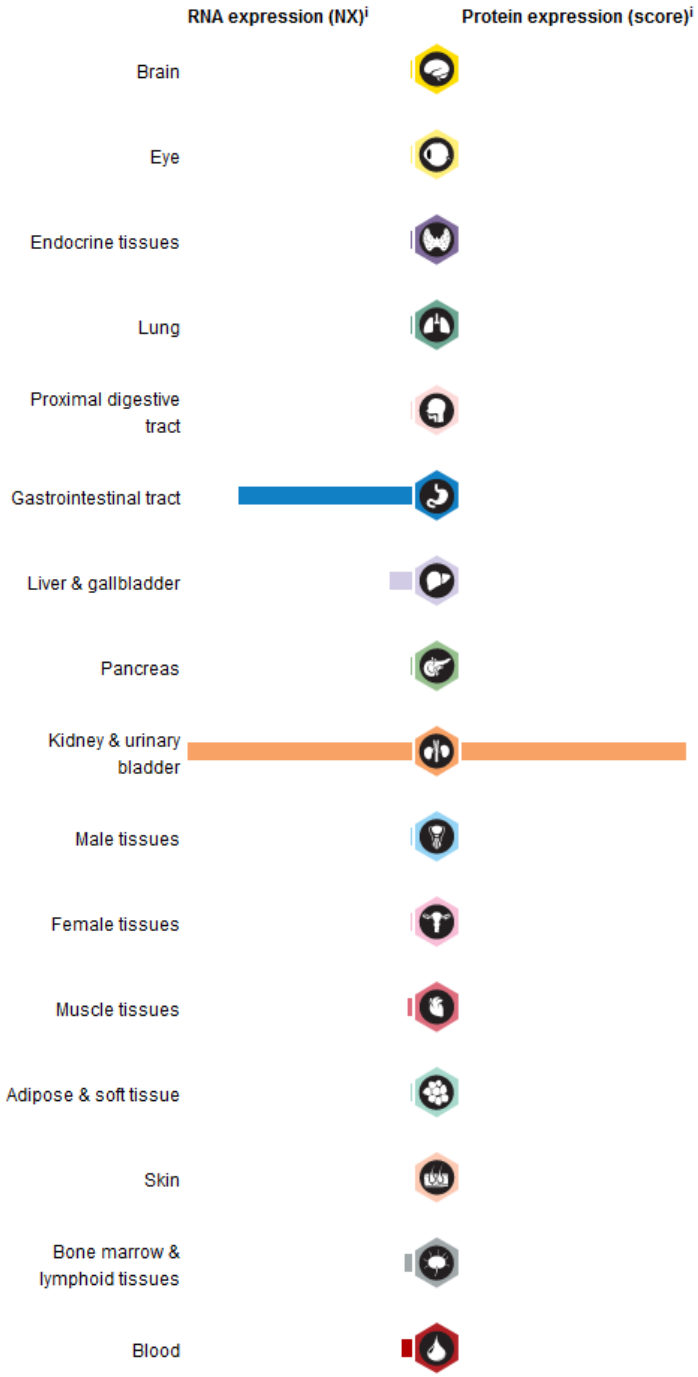


Figure 11: RNA expression and protein detection of FAM151A from the Human Protein Atlas.

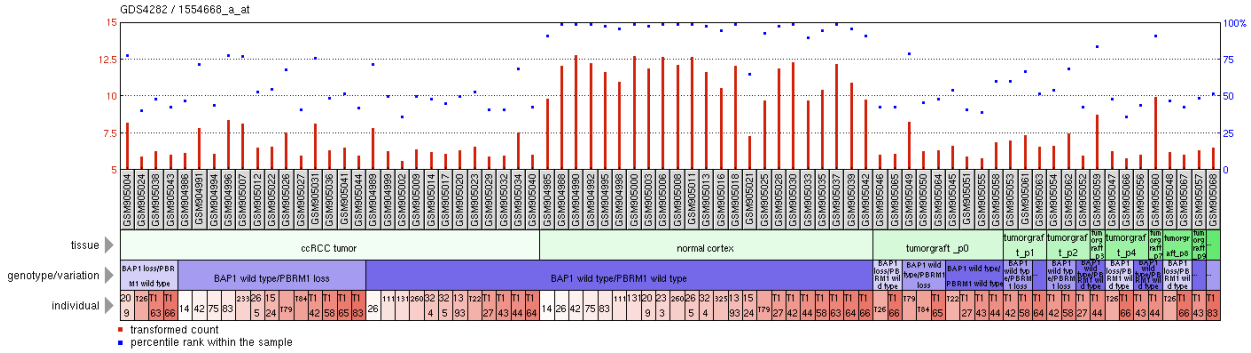


Figure 12: Expression of FAM151A in normal and renal carcinomic kidney tissue.

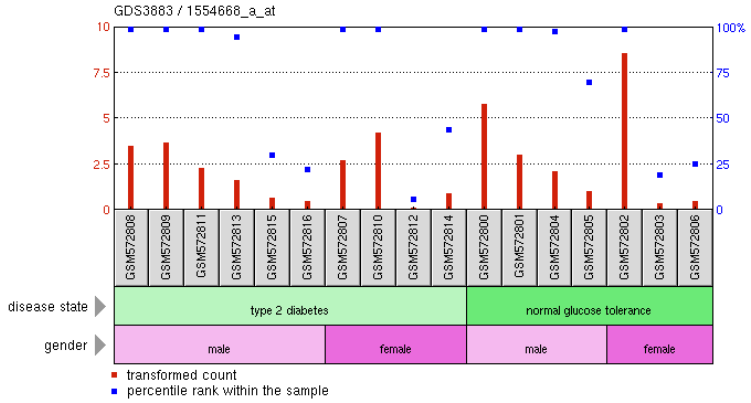


Figure 13: Expression of FAM151A in normal and Type II diabetic kidney tissue.

genotypes. This potentially suggests a causal relationship, as the FAM151A protein is present in kidney tubules.

We also investigate whether FAM151A could be associated with a prominent kidney-related condition, type 2 diabetes. Figure 13 provides no evidence that FAM151A is correlated with diabetes conditions, although it has significantly variable expression from sample to sample.

Figure 14 presents FAM151A expression in hepatitic and non-hepatitic liver tissue, and we measure lower expression in tissue samples with hepatitis. While the Human Protein Atlas suggests that the FAM151A mRNA transcript is not transcribed into protein in the liver, there still appears to be some sort of relationship between expression and alcoholic hepatitis.

Furthermore, some research has found evidence of differential expression of FAM151A in β-cells between high and low fat 20 week male mice ($p < 0.0001$).[38] However, no other evidence was found to corroborate this finding, and so we do not further investigate this thread.

[38]Miranda, M. (2021). Exploring -Cell Function and Heterogeneity in Obese SM/J Mice.

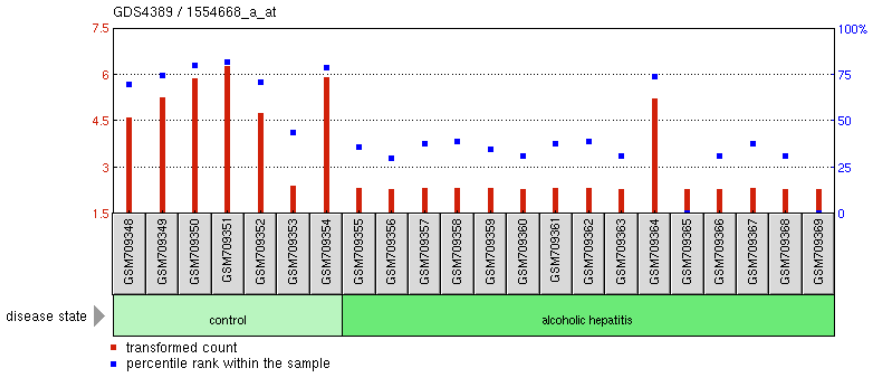


Figure 14: Expression of FAM151A in normal and alcoholic hepatitic liver tissue.

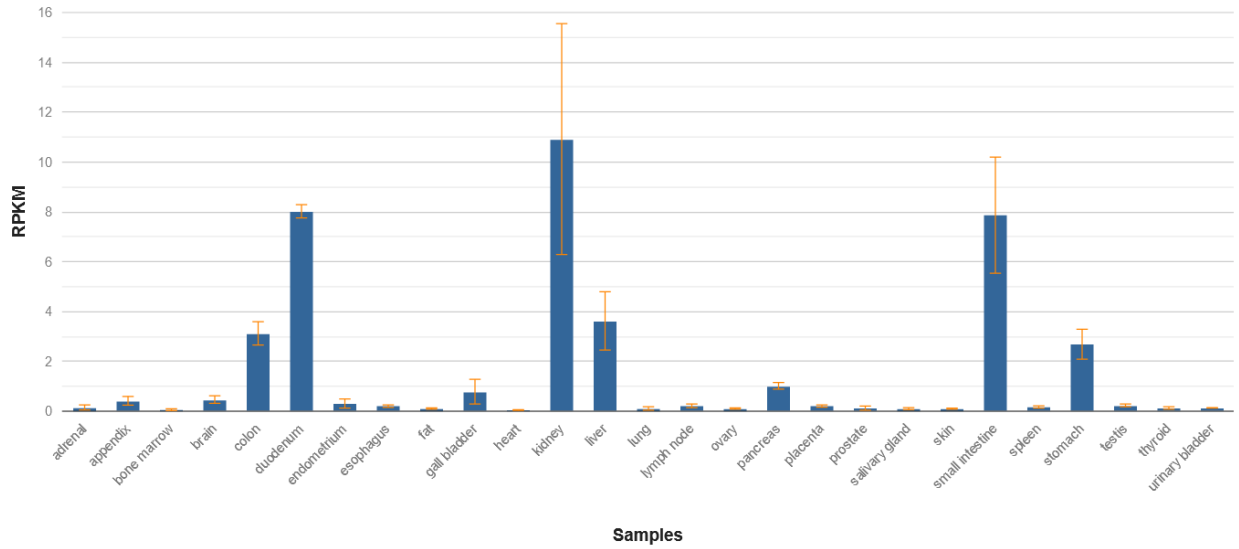


Figure 15: Expression pattern of HNF1 from NCBI Gene

FAM151A Promoter and 5' UTR Analysis

FAM151A Predicted Promoter TFs

FAM151A has one promoter region predicted by Genomatix ElDorado, a portion of which is shown in the multiple sequence alignment and TF binding site prediction annotated sequence presented in Appendix D.[39] The multiple sequence alignment was created using Clustal Omega.[40] Transcription factors binding sites were predicted using the JASPAR Core database and Genomatix Matinspector.[41,42] Transcription factors predicted by Genomatix were filtered based on matrix score as well as selective expression in the kidney.

One transcription factor, HNF1, explains much the expression pattern of FAM151A. Its binding site is conserved in all orthologs, besides mouse and rat, and located sufficiently

[39]Genomatix Software Suite. https://www.genomatix.de/solutions/genomatix-software-suite.html.
[40]Clustal Omega. https://www.ebi.ac.uk/Tools/msa/clustalo/.
[41]JASPAR Core. https://jaspar.genereg.net/.
[42]Genomatix Software Suite. https://www.genomatix.de/solutions/genomatix-software-suite.html.

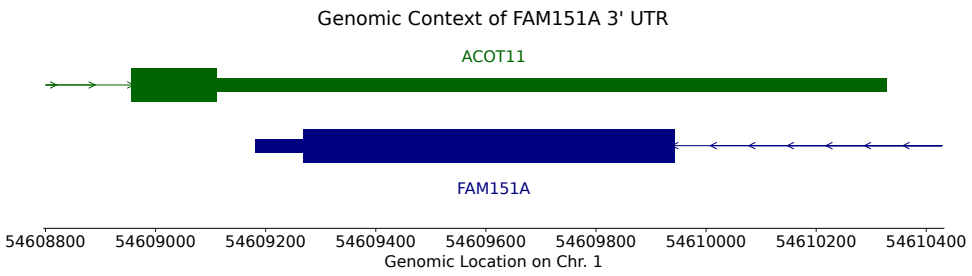


Figure 16: Genomic context of FAM151A 3' UTR.

close to the start of transcription, while sufficiently distant from a cluster of transcription factor binding sites downstream that could prevent proper binding. Furthermore, Figure 15 shows HNF1's expression pattern across tissues as reported by NCBI Gene, where we see high expression in the kidney, small intestine, and liver, consistent with the expression patterns of FAM151A reported earlier.[43]

FAM151A 3' UTR Analysis

The 3' UTR of the FAM151A mRNA transcript is 87 base pairs in length, which is unusually short. However, Figure 16 shows the region in its genomic context, where we observe that the the 3' UTR of FAM151A completely overlaps the 3' UTR of ACOT11 on the reverse strand, and that the last exon of ACOT11 lies only a few base pairs after the end of the 3' UTR of FAM151A, explaining its length.[44]

Appendix E presents an alignment of the 3' UTRs of 7 primates, where differences between the sequences and the consensus sequence are highlighted.[45] BLAST was not able to identify 3' UTRs outside of primates, likely due to the short length of the sequence.[46]

Figure 17 shows the secondary structure of the 3' UTR of FAM151A, as predicted by mFOLD, along with protein binding sites predicted by RNAPDB.[47,48] Figure 18 shows a predicted tertiary structure based on the secondary structure predicted using RNAComposer and visualized with PyMol.[49,50] We see that the most highly conserved portion (highlighted) of the UTR consists of a hairpin and an interior loop, which contains three important sites, the Poly-A signal, and predicted binding sites for ZFP36 and EIF4B. ZFP36 is a zinc finger protein that promotes degradation of the mRNA through recruitment of deadenylases and

[43]NCBI Gene entry on HNF1. https://www.ncbi.nlm.nih.gov/gene/6927.
[44]UCSC Genome Browser. https://genome.ucsc.edu/.
[45]Clustal Omega. https://www.ebi.ac.uk/Tools/msa/clustalo/.
[46]NCBI Basic Local Alignment Search Tool. https://blast.ncbi.nlm.nih.gov/Blast.cgi.
[47]Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595.
[48]RBPDB: The database of RNA-binding protein specificities. http://rbpdb.ccbr.utoronto.ca/.
[49]Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., Blazewicz, J., & Adamiak, R. W. (2012). Automated 3d structure composition for large rnas. *Nucleic Acids Research*, *40*(14), e112–e112. https://doi.org/10.1093/nar/gks339.
[50]Schrodinger, LLC. (2015). *The PyMOL molecular graphics system, version 1.8.*

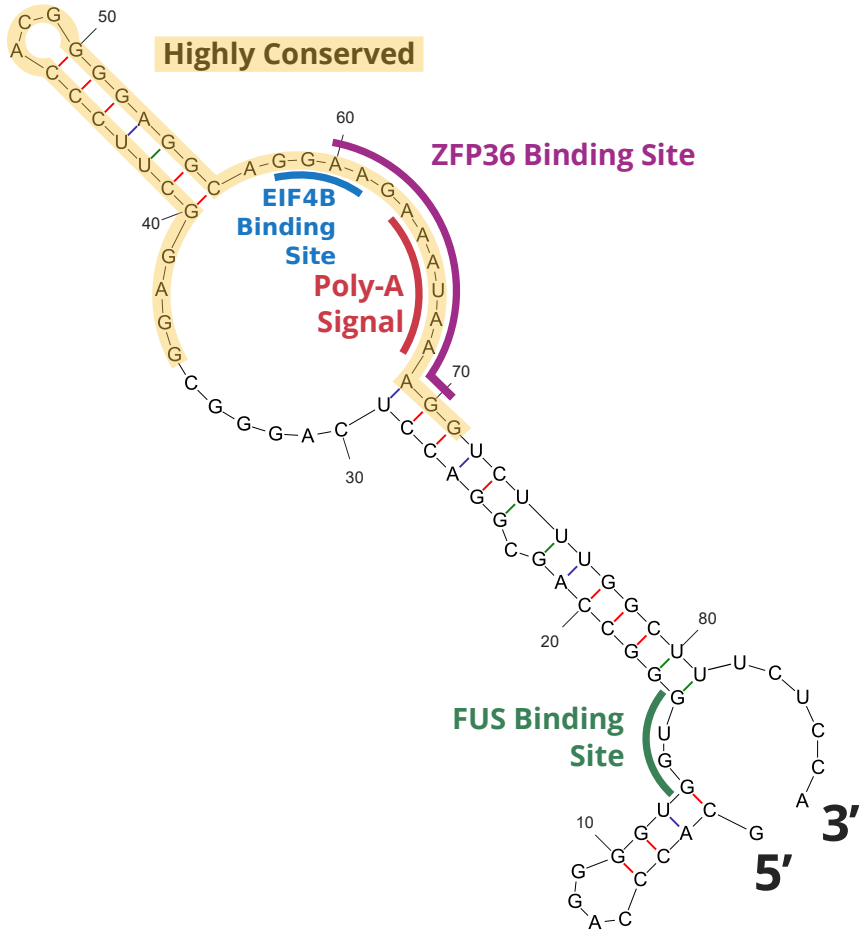


Figure 17: Annotated predicted secondary structure of FAM151A 3' UTR.

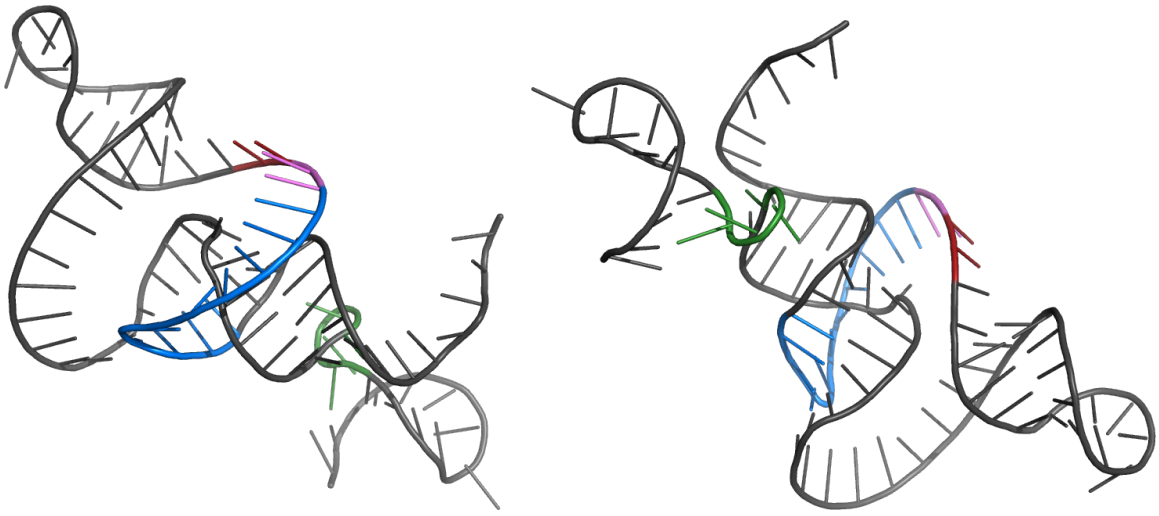


Figure 18: Two views of tertiary structure of FAM151A 3' UTR as predicted by RNAComposer. Predicted binding site of FUS shown in green, EIF4B in red, ZFP36 in blue, and the overlap between binding sites of EIF4B and ZNF36 in purple.

exoribonucleases, while EIF4B is a translation initiation factor (both are expressed ubiquitously).[51,52] Because the ZFP and EIF4B binding sites overlap by two bases, the translation of the transcript will be determined by relative expressions of the two factors.

Examining these two factors explains FAM151A's differential expression in the kidney. The Human Protein Atlas provides data on ZFP36 and EIF4B. In the glomeruli, ZFP36 is expressed while EIF4B is not, while in the tubules, EIB4B is expressed, while ZFP36 is not.[53,54] Furthermore, the cellular location of both factors is identified as cytoplasmic/membranous, which matches the expectation of regulation of FAM151A, a transmembrane protein.

Additionally, RBPDB predicts a FUS binding site close to the 5' end of the 3' UTR, which is also conserved. FUS is involved in pre-mRNA splicing and export of mRNA to the cytoplasm, but does not provide additional information to explain differential expression of FAM151A.[55]

[51]Rodriguez-Gomez, G., Paredes-Villa, A., Cervantes-Badillo, M. G., Gomez-Sonora, J. P., Jorge-Perez, J. H., Cervantes-Roldan, R., & Leon-Del-Rio, A. (2021). Tristetraprolin: A cytosolic regulator of mRNA turnover moonlighting as transcriptional corepressor of gene expression. *133*(2), 137–147. https://doi.org/10.1016/j.ymgme.2021.03.015.

[52]NCBI Gene entry on EIF4B. https://www.ncbi.nlm.nih.gov/gene/1975.

[53]Human Protein Atlas entry on ZFP36. https://www.proteinatlas.org/ENSG00000128016-ZFP36/tissue/kidney.

[54]Human Protein Atlas entry on EIF4B. https://www.proteinatlas.org/ENSG00000063046-EIF4B/tissue/kidney.

[55]NCBI Gene entry on FUS. https://www.ncbi.nlm.nih.gov/gene/2521.

Protein

Properties and Post-Translational Modifications of FAM151A

Molecular Weight of Protein FAM151A

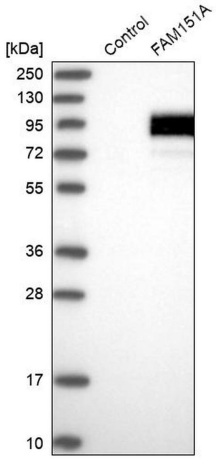


Figure 19: Western Blot of FAM151A.

Expasy predicts an isoelectric point of 6.19 and a molecular weight of 64kDa for FAM151A.[56] This does not agree with the weight of FAM151A experimentally derived from western blot experiments from ThermoFisher, which is approximately 95kDa, as shown in Figure 19.[57] This suggests that either the antibody did not properly capture FAM151A, that FAM151A undergoes significant post-translational modification, or that FAM151A was not properly separated from binding partners before Western blot.

N-Linked Glycosylation Sites of FAM151A

FAM151A is predicted to undergo N-linked glycosylation in 5 sites by NetNGlyc, ELM, and MotifScan, 4 of which are predicted by PhosphoSitePlus.[58,59,60,61] The one site not predicted by PhosphoSitePlus is not highly conserved, so it was not regarded as notable, while the other four are labelled in the annotated conceptual translation and its associated schematic diagram. This increases our confidence that the C-terminal end of FAM151A lies outside the cell, where this glycosylation occurs. Thus, phosphorylation sites were not predicted, as the major portion of FAM151A lies outside the cell. No other significant post-translational modifications were found, including disulfide bonds.

[56] Expasy: Compute pI/Mw. https://web.expasy.org/cgi-bin/compute_pi/pi_tool.

[57] ThermoFisher Antibodies for FAM151A. https://www.thermofisher.com/antibody/product/FAM151A-Antibody-Polyclonal/PA5-53502.

[58] NetNGlyc. https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0.

[59] ELM: The Eukaryotic Linear Motif resource for Functional Sites in Proteins. http://elm.eu.org/search/.

[60] MyHits Motif Scan. https://myhits.sib.swiss/cgi-bin/motif_scan.

[61] PhosphoSitePlus. https://www.phosphosite.org/homeAction.action.

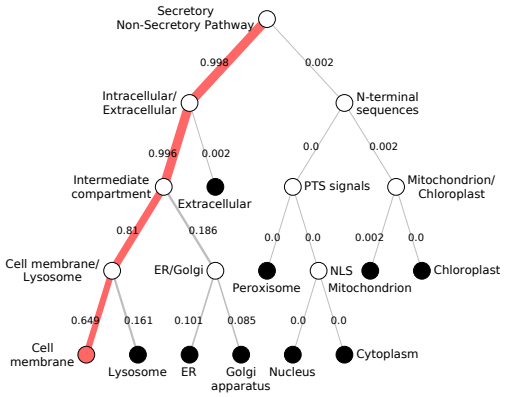


Figure 20: DeepLoc prediction graph of FAM151A subcellular localization.

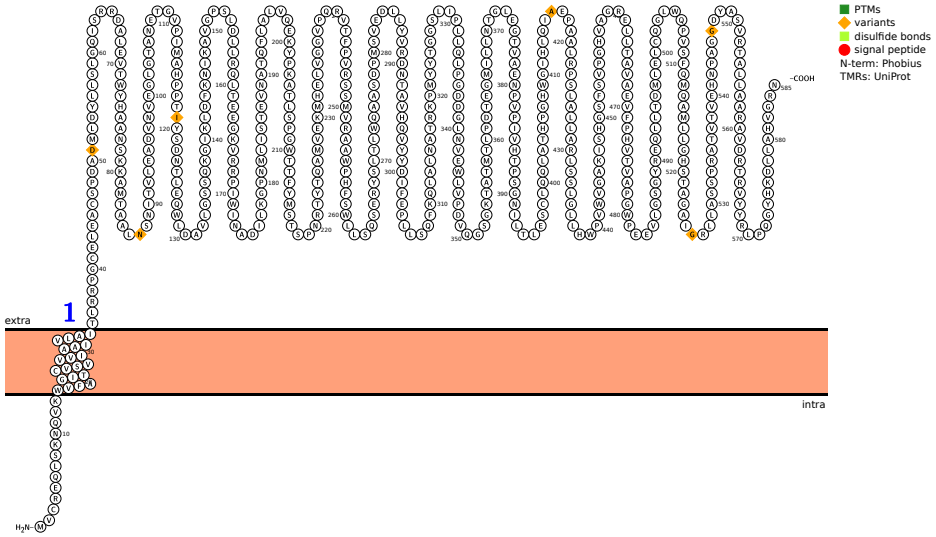


Figure 21: Protter diagram of FAM151A.

FAM151A Resides in the Cell Membrane

There exists sufficient evidence to claim that FAM151A is a transmembrane protein primarily residing outside the cell membrane. DeepLoc predicts that FAM151A is localized to the cell membrane with 43% probability, as shown in Figure 20.[62] In Figure 21, we see that Protter predicts one transmembrane region near the N-terminus of the peptide, and that the rest of the protein lies outside the cell, as does SAPS (which makes no other significant predictions).[63,64] Additionally, PSORTII predicts that FAM151A has Type II membrane

[62]DeepLoc: Prediction of eukaryotic protein subcellular localization using deep learning. https://services.healthtech.dtu.dk/service.php?DeepLoc-1.0.

[63]Omasits, U., Ahrens, C. H., Müller, S., & Wollscheid, B. (2013). Protter: Interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, *30*(6), 884–886. https://doi.org/10.1093/bioinformatics/btt607.

[64]Statistical Analysis of Protein Sequences. https://www.ebi.ac.uk/Tools/seqstats/saps/.

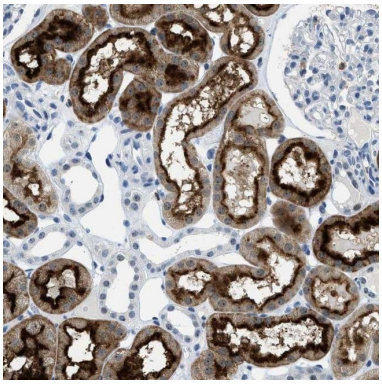


Figure 22: Antibody staining of FAM151A reveals cell membrane localization.

topology, that is, the N-terminus lies inside the membrane, and the protein is a single-pass protein (PSORT's subcellular localization prediction is not included as it does not contain an option for the cell membrane).[65] This is consistent with UniProtKB's annotation of one transmembrane region.[66] Finally, staining data of FAM151A in kidney tissue from ThermoFischer antibodies shown in Figure 22 empirically confirms the presence of FAM151A in membrane tissue.[67]

Structure of FAM151A

Figure 23 shows the tertiary structure of protein FAM151A as predicted by AlphaFold2, visualized using PyMol.[68,69] The vast majority of the structure of the two DUF2181s is predicted with very high confidence (pLDDT > 90). Furthermore, the prediction agrees with all of the previous known information on the protein: the transmembrane alpha helix is predicted correctly, the two DUF2181s are properly separated, and the structure of the domains (shown in Figure 24) is correctly predicted as a TIM barrel fold, which is known from homology between DUF2181 and bacterial glycerophosphodiester phosphodiesterases.[70]

I-TASSER did not correctly predict the tertiary structure of FAM151A, as it did not predict separation of the two main protein domains, nor the TIM barrel fold structure of the domains, as shown in Figure 25.[71] Thus, its prediction is not discussed at length here. The same reasoning applies to secondary structure prediction algorithms.

[65] PSORT II Prediction. https://psort.hgc.jp/form2.html.

[66] UniProtKB entry on FAM151A. https://www.uniprot.org/uniprot/Q8WW52.

[67] ThermoFisher Antibodies for FAM151A. https://www.thermofisher.com/antibody/product/FAM151A-Antibody-Polyclonal/PA5-53502.

[68] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[69] Schrodinger, LLC. (2015). *The PyMOL molecular graphics system, version 1.8.*

[70] Findlay, A. S., McKie, L., Keighren, M., Clementson-Mobbs, S., Sanchez-Pulido, L., Wells, S., Cross, S. H., & Jackson, I. J. (2020a). Fam151b, the mouse homologue of C.elegans menorin gene, is essential for retinal function. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-019-57398-4.

[71] Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725–738. https://doi.org/10.1038/nprot.2010.5.

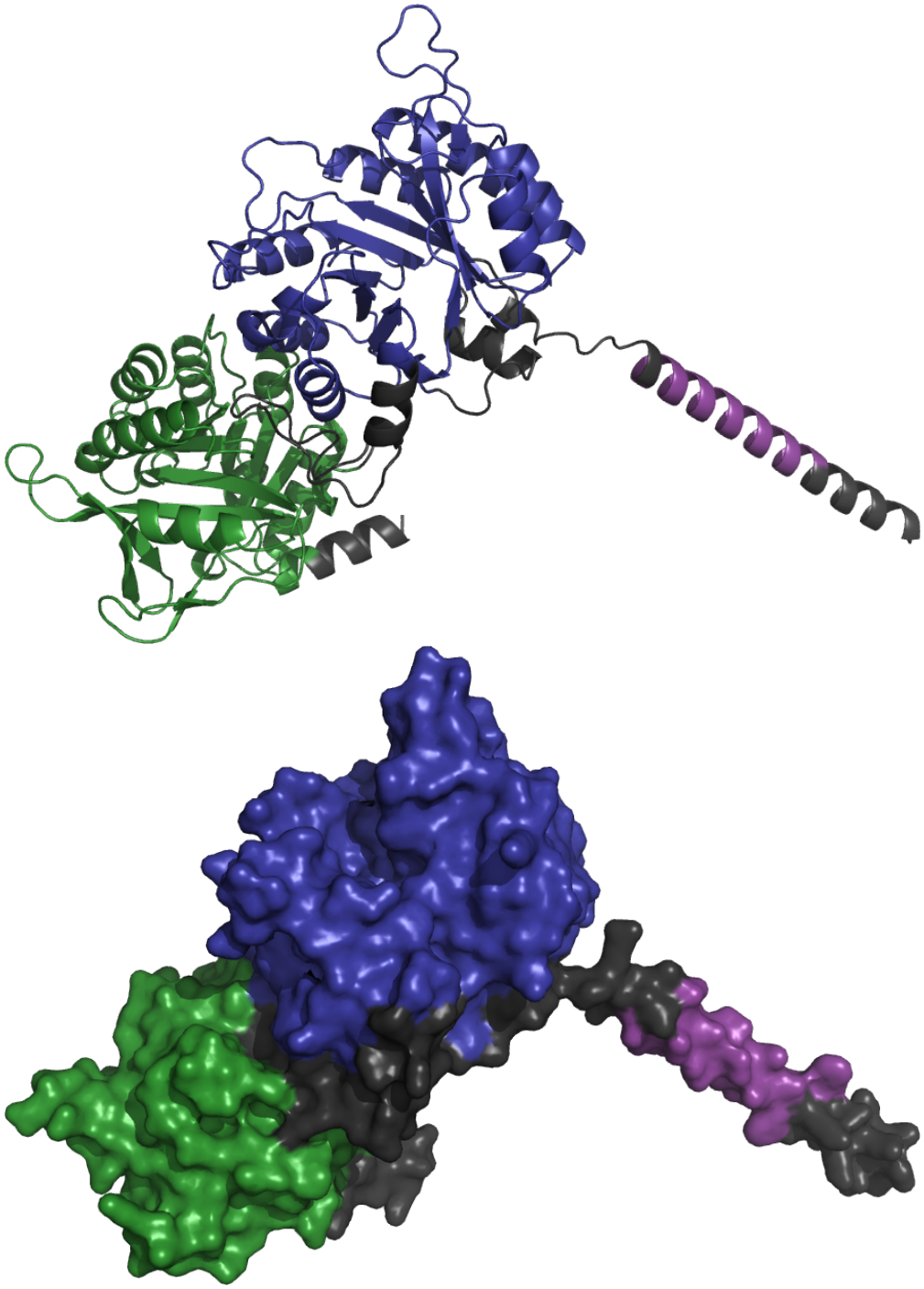


Figure 23: Tertiary structure of FAM151A as predicted by AlphaFold2. The transmembrane domain is highlighted in purple, the first DUF2181 in blue, and the second DUF2181 in green, while interdomain regions are shown in gray. Both ribbon and surface diagrams are shown.

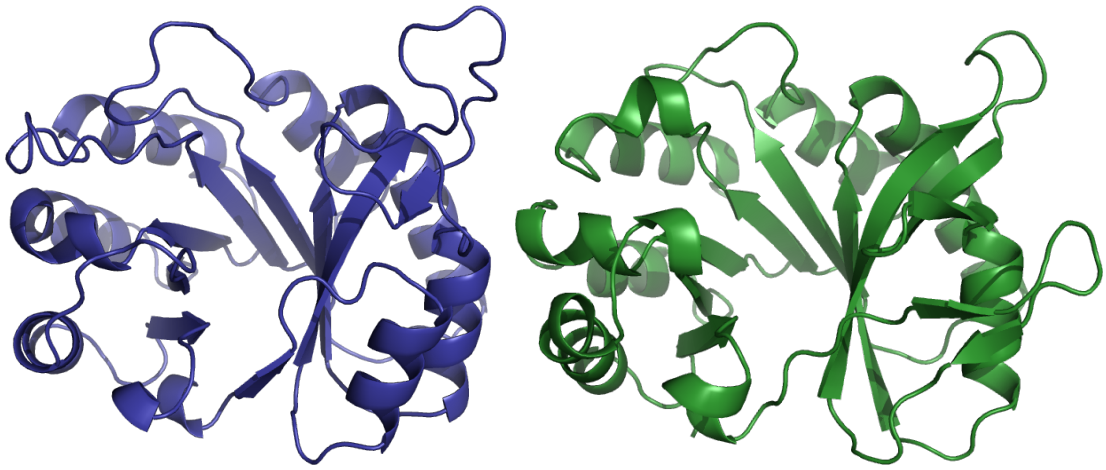


Figure 24: Tertiary structure of DUF2181s of FAM151A as predicted by AlphaFold2. The first is shown in blue on the left, the second in green on the right. Both are correctly predicted as TIM barrel folds.

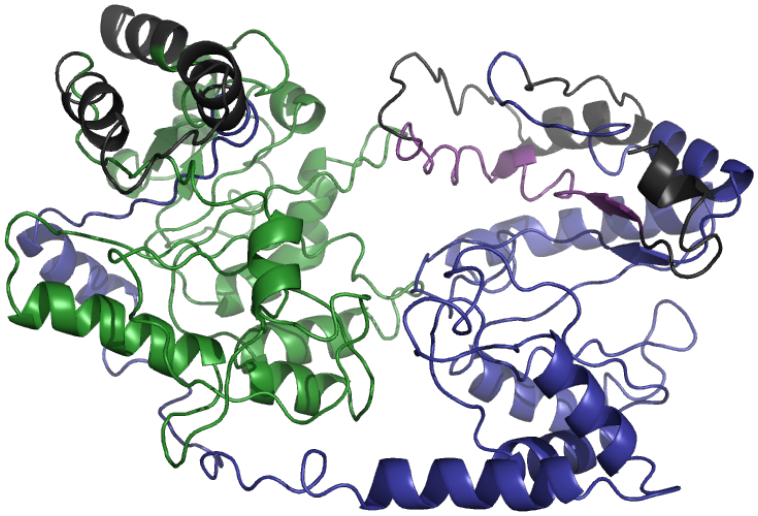


Figure 25: I-TASSER FAM151A tertiary structure prediction. Domains are colored using the same scheme as Figure 23.

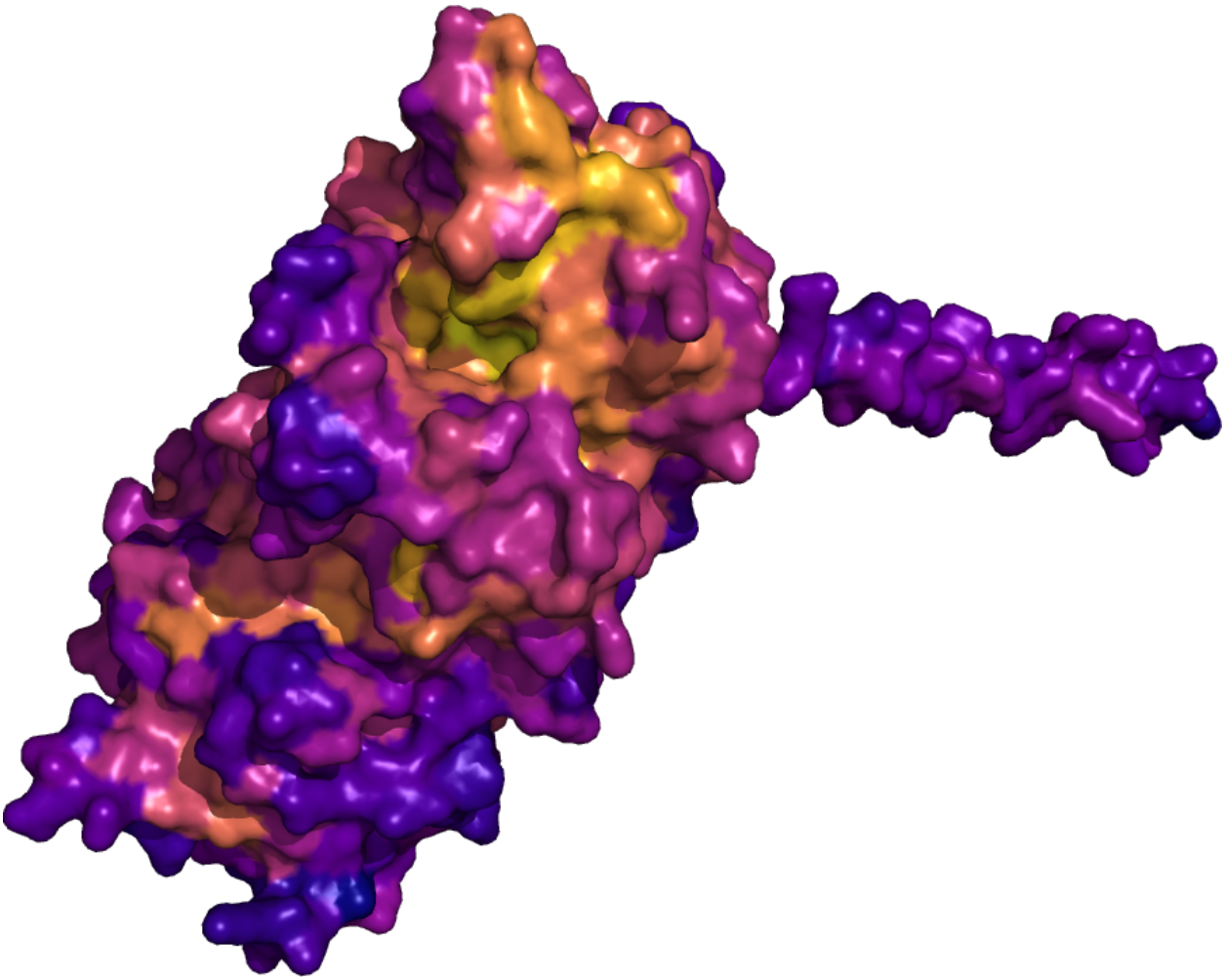


Figure 26: FAM151A tertiary structure colored by conservation. Yellows indicate highly conserved residues, purples indicate poorly conserved residues.

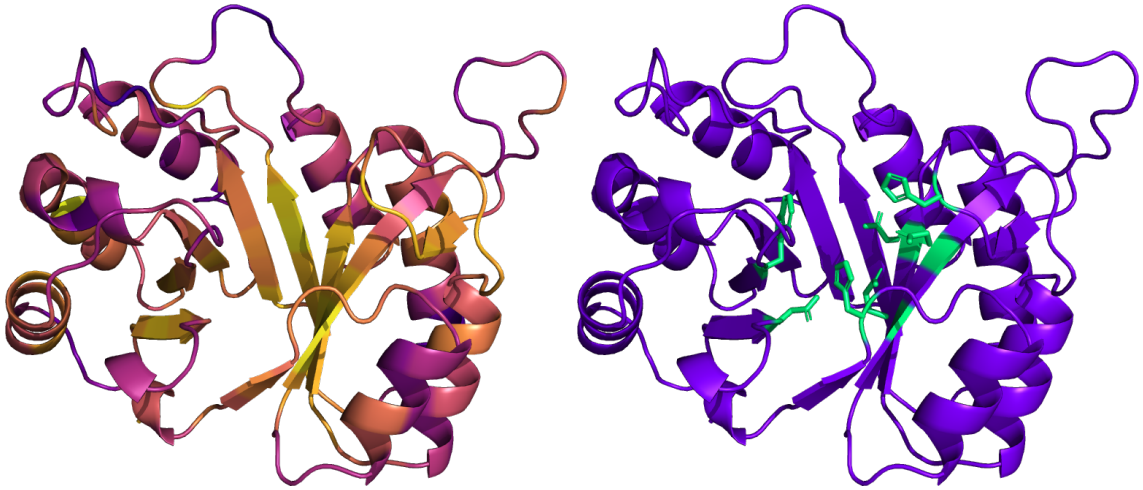


Figure 27: Side by side views of the first DUF2181. The first diagram highlights conservation, the second, published active site residues.

Figure 26 displays the predicted tertiary structure of FAM151A where each amino acid is colored according to the Shannon entropy of that residue in an MSA of 20 Eumetazoan orthologs of FAM151A (see Figures 23). This view makes it obvious that the highly conserved portions of FAM151A are concentrated in the first DUF2181.

Function of Enzymatically Active DUF2181

Active Site Residues

Appendix C presents two alignments of DUF2181s with purported active site residues highlighted. The first is an alignment of DUF2181s in the FAM151/Menorin family (described above in discussion of FAM151 phylogeny), where we see active site residues conserved in almost all orthologs. The second is an alignment of seven DUF2181/GDPD domains, aligned using Clustal Omega.[72] Of the DUF domains, two are found in FAM151A (`F151A1`,`F151A2`), one is found in FAM151B (`F151B1`), and one in menorin, a *C. elegans* homolog of the FAM151 family (`MNR1CE`). We also take a GDPD domain from humans (`GPCP1H`), *E. coli* (`GDPDEC`), and *O. iheyensis* (`GDPDOI`). In this alignment, active site residues in the GDPD family are highlighted in red.[73] We see from this alignment than the second DUF2181 in FAM151A does not contain any conserved active sites, thus we predict it to be nonfunctional. Note that the alignment contains a roughly 100 aa omission that contains no active sites for brevity.

To further validate our hypothesized FAM151A DUF2181 active sites, we plot conservation of the first domain in direct FAM151A orthologs and active site residues side by side in Figure 27. From this we see that that active sites in FAM151A are highly conserved (all were conserved in 19 or 20 of 20 orthologs), and all lie in the inside of the barrel, providing further evidence of their functionality and AlphaFold2's prediction accuracy.

[72]Clustal Omega. https://www.ebi.ac.uk/Tools/msa/clustalo/.

[73]NCBI Structure entry on Conserved Protein Domain Family GDPD. https://www.ncbi.nlm.nih.gov/Structure/cdd/cd08556.

Example Functions of GDPD/PLCD Superfamily Phosphodiesterases

Thus, to understand the enzymatic function of FAM151A, we investigate the function of the first DUF2181 as a phosphodiesterase. As mentioned above, the DUF2181 present in FAM151/Menorin family is a member of the GDPD/PLCD superfamily, so we know that the substrate of FAM151A contains a phosphodiester bond, and is most likely a glycerophosphodiester or a phospholipid.[74,75] However, phosphodiesterases bind to a large variety of substrates, so there is no obvious candidate for FAM151A's substrate.[76] Currently, the exact substrate that menorin binds to is unknown, so we cannot use this to predict the enzymatic activity of FAM151A. Thus, we examine a few representative phosphodiesterases similar in expression to FAM151A to hypothesize about possible substrates.

We first investigate GDPD5 (glycerophosphodiester phosphodiesterase domain-containing protein 5), a transmembrane phosphodiesterase involved in neuron development.[77] GDPD5 is also highly expressed in kidney tubules but not glomeruli, making it extremely similar to the FAM151/Menorin family.[78] GDPD5 (also known as GDE2) is known to cleave the glycosylphosphatidylinositol (GPI) anchor of protein RECK, but this activity is generally only attributed to six-transmembrane GDPDs, rendering it unlikely that FAM151A shares this function.[79] However, GDPD5 is also known to hydrolyze glycerophosphocholine (GPC), an organic osmolyte in the kidney, in order to maintain homeostatic sodium chloride and urea levels in the renal inner medulla.[80]

Next, we turn to ENPP6 (ectonucleotide pyrophosphatase/phosphodiesterase 6), a GPI-anchored transmembrane phosphodiesterase highly expressed in the brain and kidney proximal renal tubules.[81,82] ENPP6 hydrolyzes both α-GPC (known to be involved in Alzheimer's pathways) and β-GPC as part of the choline metabolism pathway.[83,84]

---

[74]EXPASY Enzyme Entry on Glycerophosphodiester phosphodiesterase. https://enzyme.expasy.org/EC/3.1.4.46.

[75]Kolesnikov, Y. S., Nokhrina, K. P., Kretynin, S. V., Volotovski, I. D., Martinec, J., Romanov, G. A., & Kravets, V. S. (2012). Molecular structure of phospholipase D and regulatory mechanisms of its activity in plant and animal cells. *Biochemistry (Moscow)*, *77*(1), 1–14. https://doi.org/10.1134/s0006297912010014.

[76]Corda, D., Mosca, M. G., Ohshima, N., Grauso, L., Yanaka, N., & Mariggiò, S. (2014). The emerging physiological roles of the glycerophosphodiesterase family. *The FEBS Journal*, *281*(4), 998–1016. https://doi.org/https://doi.org/10.1111/febs.12699.

[77]UniProtKB Entry on GDPD5. https://www.uniprot.org/uniprot/Q8WTR4.

[78]Human Protein Atlas Entry on GDPD5. https://www.proteinatlas.org/ENSG00000158555-GDPD5/tissue.

[79]Park, S., Lee, C., Sabharwal, P., Zhang, M., Meyers, C. L. F., & Sockanathan, S. (2013). GDE2 Promotes Neurogenesis by Glycosylphosphatidylinositol-Anchor Cleavage of RECK. *Science*, *339*(6117), 324–328. https://doi.org/10.1126/science.1231921.

[80]Gallazzini, M., Ferraris, J. D., & Burg, M. B. (2008). GDPD5 is a glycerophosphocholine phosphodiesterase that osmotically regulates the osmoprotective organic osmolyte GPC. *Proceedings of the National Academy of Sciences*, *105*(31), 11026–11031. https://doi.org/10.1073/pnas.0805496105.

[81]UniProtKB Entry on ENPP6. https://www.uniprot.org/uniprot/Q6UWR7.

[82]Human Protein Atlas Entry on ENPP6. https://www.proteinatlas.org/ENSG00000164303-ENPP6/tissue.

[83]Morita, J., Kano, K., Kato, K., Takita, H., Sakagami, H., Yamamoto, Y., Mihara, E., Ueda, H., Sato, T., Tokuyama, H., Arai, H., Asou, H., Takagi, J., Ishitani, R., Nishimasu, H., Nureki, O., & Aoki, J. (2016). Structure and biological function of ENPP6, a choline-specific glycerophosphodiester-phosphodiesterase. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep20995.

[84]Parnetti, L., Mignini, F., Tomassoni, D., Traini, E., & Amenta, F. (2007). Cholinergic precursors in

Finally, although the substrate of menorin is not known, we note experiments conducted in which knockouts of sax-7, known to act as a coligand with menorin, were exposed to cholinesterase inhibitors provide data consistent with sax-7 being involved in acetylcholine reception, along with gtl-2 (human orthologs TRMP1/3/7).[85,86] Thus, given all the evidence above, we present the most likely substrate of the biochemically active DUF2181 of FAM151A as a glycerophosphocholine, especially given its double role as neurotransmitter precursor and osmolyte.

FAM151A Interacting Proteins

Binding Partners in Orthologs

In *C. elegans*, menorin is observed to work in similar pathways as sax-7, the *C. elegans* homolog of L1CAM (L1 cell adhesion molecule) through double knockout experiments.[87] Furthermore, additional experiments have coimmunoprecipitated menorin and sax-7, stronger in the presence of DMA-1 (also involved in neuron branching), all but confirming physical interaction of menorin and sax-7, with DMA-1 potentially being a third member of the complex.[88,89] This suggests that L1CAM could have a potential interaction with FAM151A, although FAM151B's expression in the brain and more direct relationship to menorin suggests that FAM151B is a more likely candidate for interaction with L1CAM.

Predicted by Existing Databases

Two binding partners of FAM151A have been obtained by co-immunoprecipitation, CD81 (Cluster of Differentiation 81) and APP (amyloid beta precursor protein).[90,91] However, it is unlikely that either represents a meaningful interaction. While the fact that CD81 is a transmembrane protein with a ubiquitously expressed mRNA would suggest a that it is a

---

the treatment of cognitive impairment of vascular origin: Ineffective approaches or need for re-evaluation? *Journal of the Neurological Sciences*, *257*(1-2), 264–269. https://doi.org/10.1016/j.jns.2007.01.043.

[85]Dong, X., Liu, O. W., Howell, A. S., & Shen, K. (2013). An extracellular adhesion molecule complex patterns dendritic branching and morphogenesis. *Cell*, *155*(2), 296–307.

[86]Opperman, K., Moseley-Alldredge, M., Yochem, J., Bell, L., Kanayinkal, T., & Chen, L. (2014). A novel nondevelopmental role of the SAX-7/l1cam cell adhesion molecule in synaptic regulation in caenorhabditis elegans. *Genetics*, *199*(2), 497–509. https://doi.org/10.1534/genetics.114.169581.

[87]Ziegenfuss, J. S., & Grueber, W. B. (2013). SAX-7 and menorin light the path for dendrite morphogenesis. *Cell*, *155*(2), 269–271. https://doi.org/10.1016/j.cell.2013.09.029.

[88]Salzberg, Y., Díaz-Balzac, C. A., Ramirez-Suarez, N. J., Attreed, M., Tecle, E., Desbois, M., Kaprielian, Z., & Bülow, H. E. (2013). Skin-derived cues control arborization of sensory dendrites in caenorhabditis elegans. *Cell*, *155*(2), 308–320. https://doi.org/10.1016/j.cell.2013.08.058.

[89]Liu, O. W., & Shen, K. (2011). The transmembrane LRR protein DMA-1 promotes dendrite branching and growth in c. elegans. *Nature Neuroscience*, *15*(1), 57–63. https://doi.org/10.1038/nn.2978.

[90]Palor, M., Stejskal, L., Mandal, P., Lenman, A., Alberione, M. P., Kirui, J., Moeller, R., Ebner, S., Meissner, F., Gerold, G., Shepherd, A. J., & Grove, J. (2020). Cholesterol sensing by CD81 is important for hepatitis C virus entry. *Journal of Biological Chemistry*, *295*(50), 16931–16948. https://doi.org/10.1074/jbc.ra120.014761.

[91]Oláh, J., Vincze, O., Virók, D., Simon, D., Bozsó, Z., Tőkési, N., Horváth, I., Hlavanda, E., Kovács, J., Magyar, A., Szűcs, M., Orosz, F., Penke, B., & Ovádi, J. (2011). Interactions of pathological hallmark proteins. *Journal of Biological Chemistry*, *286*(39), 34088–34100. https://doi.org/10.1074/jbc.m111.243907.

strong candidate for potential interaction, the Human Protein Atlas reports no CD81 protein expression in kidney tissue.[92,93] Examining APP, it is also a transmembrane protein, yet not expressed in the kidney.[94] However, APP is highly expressed in the brain, and known to play a key role in Alzheimer's pathways, suggesting that it could be an interaction partner of FAM151B.[95]

Similarly Expressed Phosphodiesterases

As GDPD5 and ENPP6 both share similar functions to FAM151A and are expressed in similar patterns, both are plausible candidates for interaction.

[92]NCBI Gene entry on CD81. https://www.ncbi.nlm.nih.gov/gene/975.

[93]Human Protein Atlas entry on CD81. https://www.proteinatlas.org/ENSG00000110651-CD81.

[94]Human Protein Atlas entry on APP. https://www.proteinatlas.org/ENSG00000142192-APP.

[95]O'Brien, R. J., & Wong, P. C. (2011). Amyloid precursor protein processing and Alzheimer's disease. *Annual Review of Neuroscience*, *34*(1), 185–204. https://doi.org/10.1146/annurev-neuro-061010-113613.

Future Work

FAM151A Substrate Determination

To understand the function of FAM151A, the most critical piece of information that is yet undetermined is the phosphodiester bond-containing substrate to which FAM151A binds. This, the experimental procedure which should be most highly prioritized is one that would determine this substrate. Above, we postulate that the substrate of FAM151A is a glycerophosphocholine. Evidence of this function could be found by comparing the glycerophosphocholine content of urine of FAM151A knockout mice (used by Findlay et. al.) with the urine of wildtype mice.[96] Additionally, FAM151A could be purified and tested against an assay of glycerophosphocholines to determine substrate preference.

FAM151A/B Interacting Protein Experiments

Determining if members of the FAM151 family interact with other proteins to form a complex is also of significant importance. We can detect these interacting proteins through a variety of methods. To do this, we propose testing via co-immunoprecipitation. Because the interaction between menorin and sax-7 has been verified via this technique, and there exists a discrepancy between the predicted molecular weight of FAM151A and the weight of FAM151A as measured by Western blot with anti-FAM151A antibodies, it is very plausible that co-immunoprecipitation could identify a binding partner of FAM151A, if one exists.

FAM151A miRNA Binding Site Evaluation

Earlier, we described SNP rs11206394, which may impact occurrence of colorectal cancer via impact on an 3' UTR miRNA binding site, and identified a few candidate interactions. To fully determine this interaction, which could be of clinical relevance, we must experimentally verify any mRNA/miRNA interactions. This can be done by using a luciferase miRNA assay, which involves inserting the 3'-UTR of FAM151A (or ACOT11) after a luciferase or GFP, and then measuring luciferase activity (compared to a control).[97] This assay can be used to measure expression for both genotypes of the 3' UTR under many target miRNAs, potentially providing evidence of miRNA binding sites affected by the SNP.

[96]Findlay, A. S., McKie, L., Keighren, M., Clementson-Mobbs, S., Sanchez-Pulido, L., Wells, S., Cross, S. H., & Jackson, I. J. (2020a). Fam151b, the mouse homologue of C.elegans menorin gene, is essential for retinal function. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-019-57398-4.

[97]Jin, Y., Chen, Z., Liu, X., & Zhou, X. (2013). Evaluating the microrna targeting sites by luciferase reporter gene assay. In S.-Y. Ying (Ed.), *Microrna protocols* (pp. 117–127). Humana Press. https://doi.org/10.1007/978-1-62703-083-0_10.

Appendix A: Organism Key and Accession Numbers

Key	Organism	Common Name	FAM151 Copies
Aag	*Aricia agestis*	Brown argus	1
Ace	*Ancylostoma ceylanicum*	Parasitic roundworm	1
Aja	*Anneissia japonica*	Crinoid	2
Asi	*Alligator sinensis*	Chinese alligator	2
Bbu	*Bufo bufo*	Toad	2
Bfl	*Branchiostoma floridae*	Florida lanclet	2
Ccn	*Chrysoperla carnea*	Green lacewing	1
Ccp	*Cyprinus carpio*	Common carp	2
Cel	*Caenorhabditis elegans*	Nematode	1
Cin	*Ciona intestinalis*	Sea squirt	1
Cpi	*Chrysemys picta bellii*	Painted turtle	2
Cse	*Coccinella septempunctata*	Seven-spot ladybird	1
Cte	*Capitella teleta*	Annelid worm	2
Dme	*Drosophila melanogaster*	Fruit fly	1
Dpa	*Diploscapter pachys*	Nematode	1
Dre	*Danio rerio*	Zebrafish	2
Eca	*Equus caballus*	Horse	2
Gae	*Gigantopelta aegis*	Deep sea snail	2
Gga	*Gallus gallus*	Chicken	1
Hsa	*Homo sapiens*	Human	2
Lan	*Lingula anatina*	Brachiopod	2
Lva	*Lytechinus variegatus*	Green sea urchin	2
Mmu	*Mus musculus*	Mouse	2
Obi	*Octopus bimaculoides*	California two-spot octopus	1
Ofu	*Owenia fusiformis*	Polychaete worm	2
Pan	*Papio anubis*	Olive baboon	2
Pma	*Papilio machaon*	Old World swallowtail	1
Ptr	*Pan troglodytes*	Chimpanzee	2
Rbi	*Rhinatrema bivittatum*	Two-lined caecilian	2
Rty	*Rhincodon typus*	Whale shark	2
Scl	*Styela clava*	Stalked sea squirt	1
Sph	*Sepia pharaonis*	Pharaoh cuttlefish	1
Spi	*Stylophora pistillata*	Hood coral	2
Sra	*Strongyloides ratti*	Nematode parasite	1
Vpa	*Vicugna pacos*	Alpaca	2
Vpe	*Vespula pensylvanica*	Western yellowjacket	1
Xla	*Xenopus laevis*	African clawed frog	2

Table A1: Key of organism abbreviations in following diagrams.

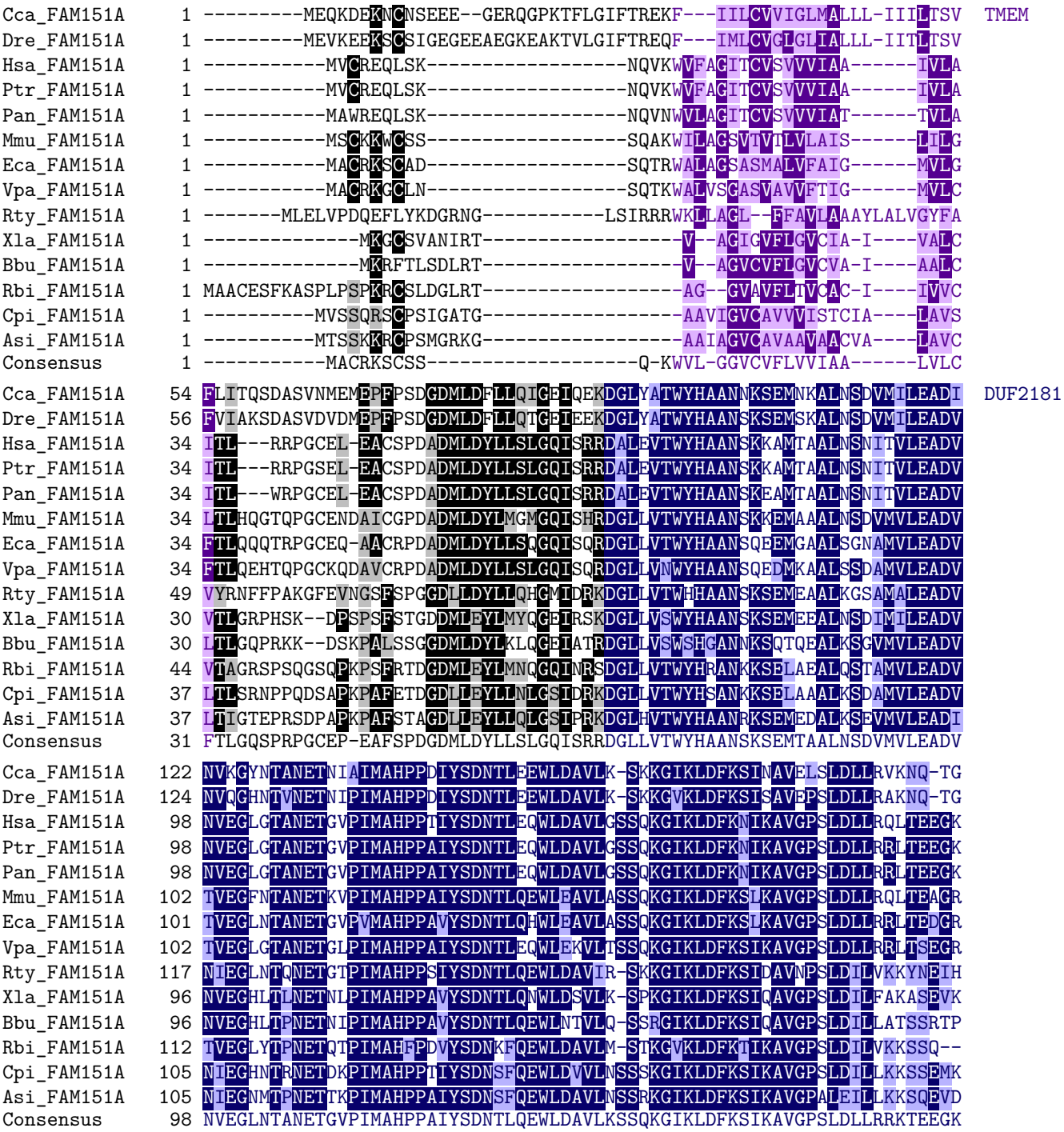
Key	Accession number 1 (FAM151A)	Accession number 2 (FAM151B/Menorin)
Aag	-	XP_041984464.1
Ace	-	EYB95295.1
Aja	XP_033116333.1	XP_033108640.1
Asi	XP_006025880.1	XP_025070544.1
Bbu	XP_040262912.1	XP_040277425.1
Bfl	XP_035660277.1	XP_035660473.1
Ccn	-	XP_044733010.1
Ccp	XP_042575185.1	XP_042579844.1
Cel	-	NP_507991.1
Cin	-	XP_002121148.3
Cpi	XP_005284924.2	XP_008165715.1
Cse	-	XP_044747787.1
Cte	ELT90991.1	ELT88790.1
Dme	-	NP_001245933.1
Dpa	-	PAV58873.1
Dre	NP_001093565.1	NP_001003531.1
Eca	XP_001488568.4	XP_023473854.1
Gae	XP_041378357.1	XP_041358246.1
Gga	-	XP_003643128.1
Hsa	NP_788954.2	NP_991111.2
Lan	XP_013411281.1	XP_013399869.1
Lva	XP_041464769.1	XP_041484609.1
Mmu	NP_666261.1	NP_001157099.1
Obi	-	XP_014778013.1
Ofu	CAC9668553.1	CAC9569733.1
Pan	XP_003891985.2	XP_017815193.1
Pma	-	XP_014369256.1
Ptr	XP_016774503.1	XP_016808561.1
Rbi	XP_029474719.1	XP_029430098.1
Rty	XP_020366386.1	XP_020386473.1
Scl	-	XP_039273176.1
Sph	-	CAE1178732.1
Spi	PFX14114.1	XP_022808211.1
Sra	-	XP_024503321.1
Vpa	XP_006200587.1	XP_031530378.1
Vpe	-	XP_043676237.1
Xla	-	XP_018116415.1

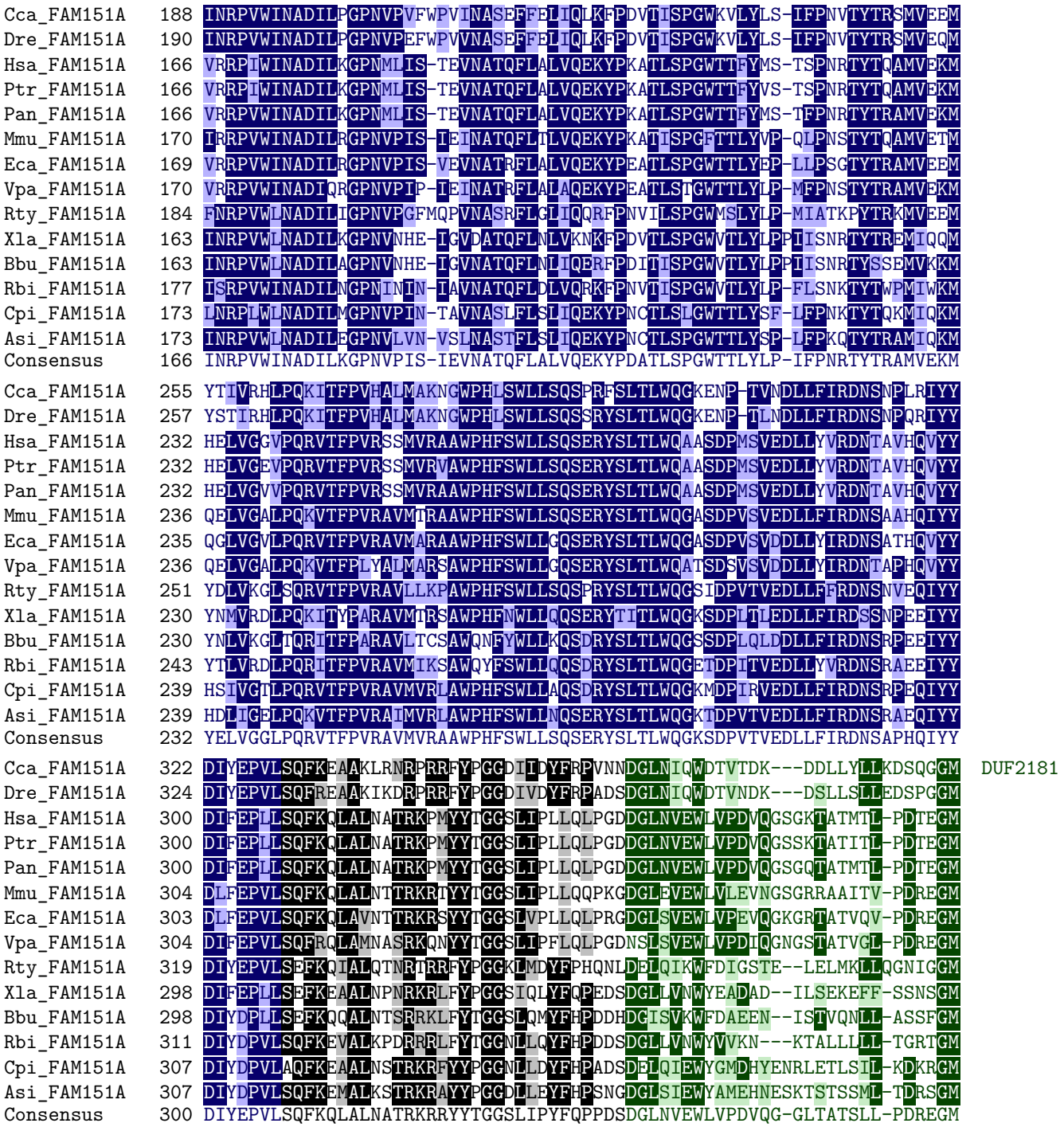
Table A2: Accession numbers of proteins found in orthologs.

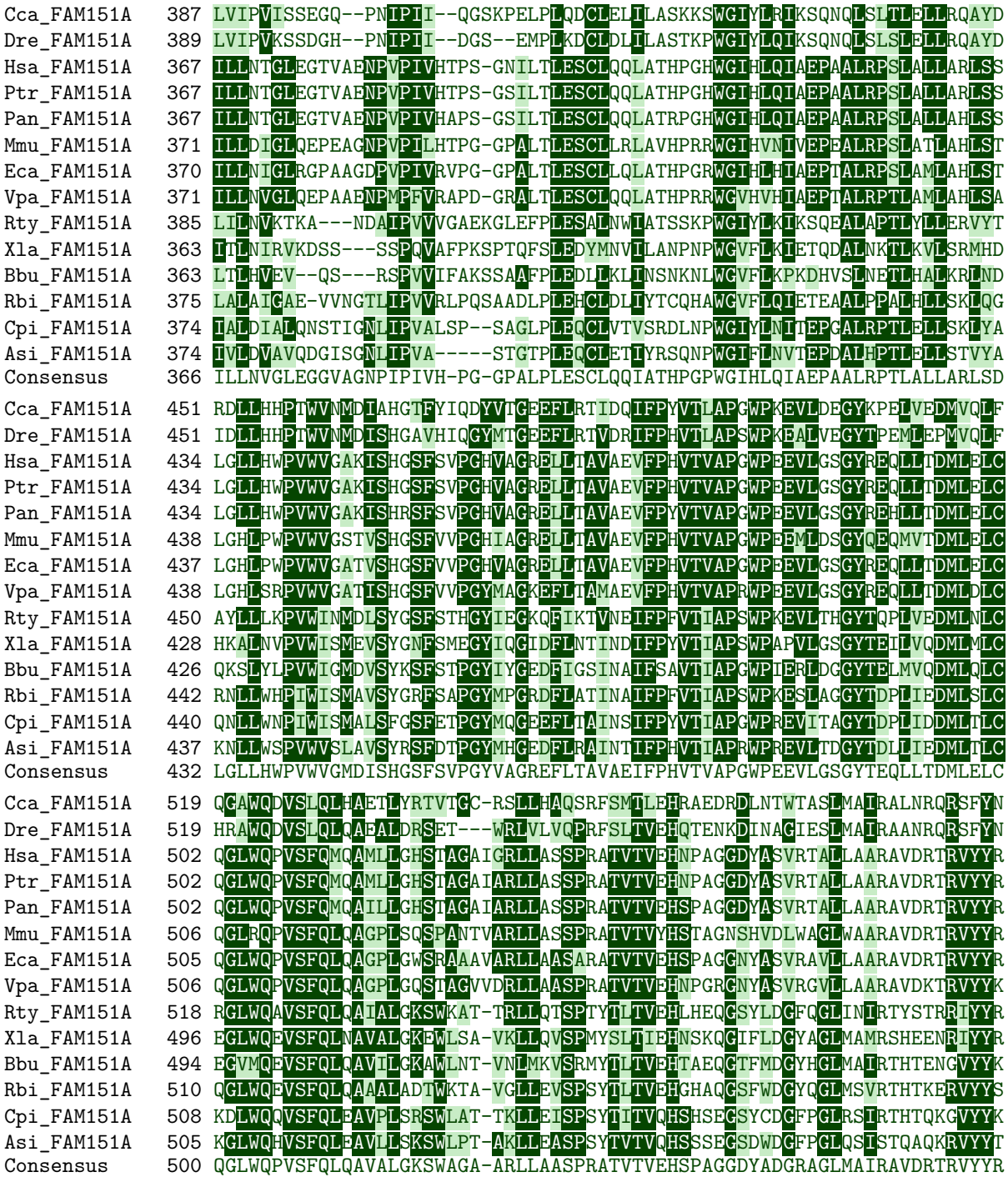
Appendix B: Multiple Sequence Alignments of FAM151A Orthologs

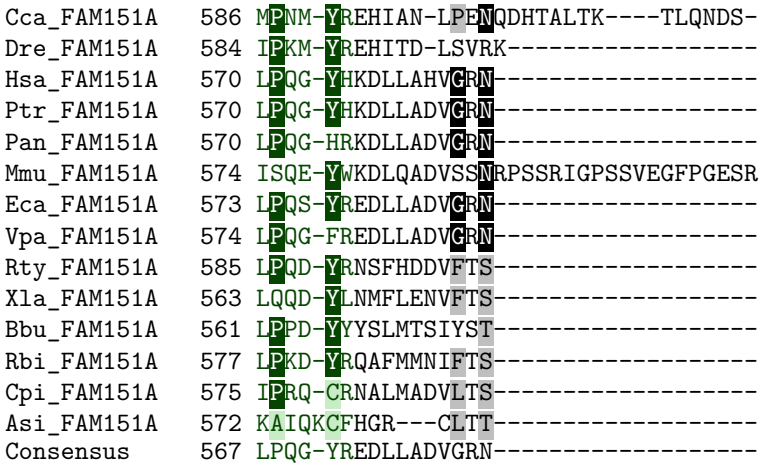
Global Alignments of FAM151A Orthologs

Strict Orthologs

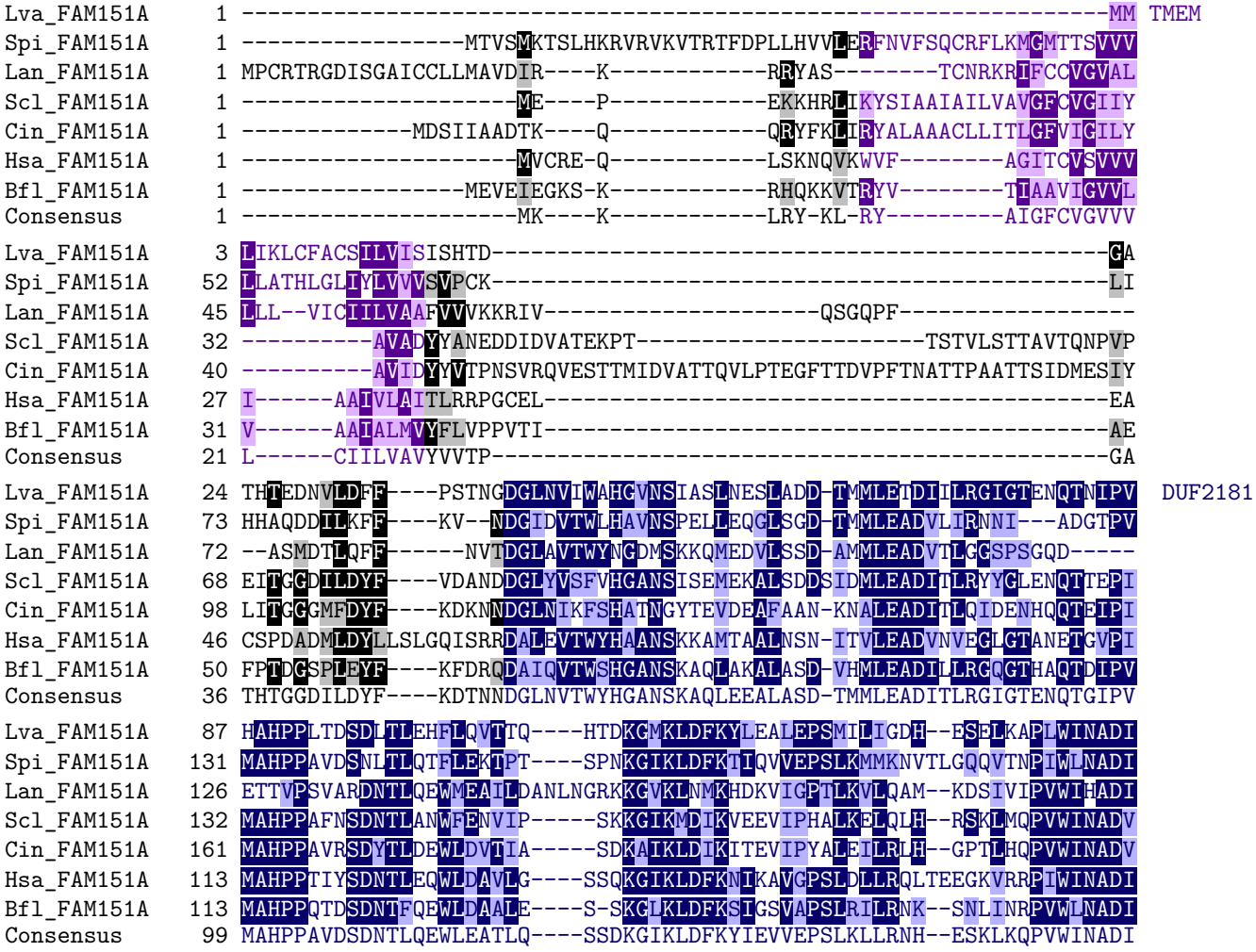


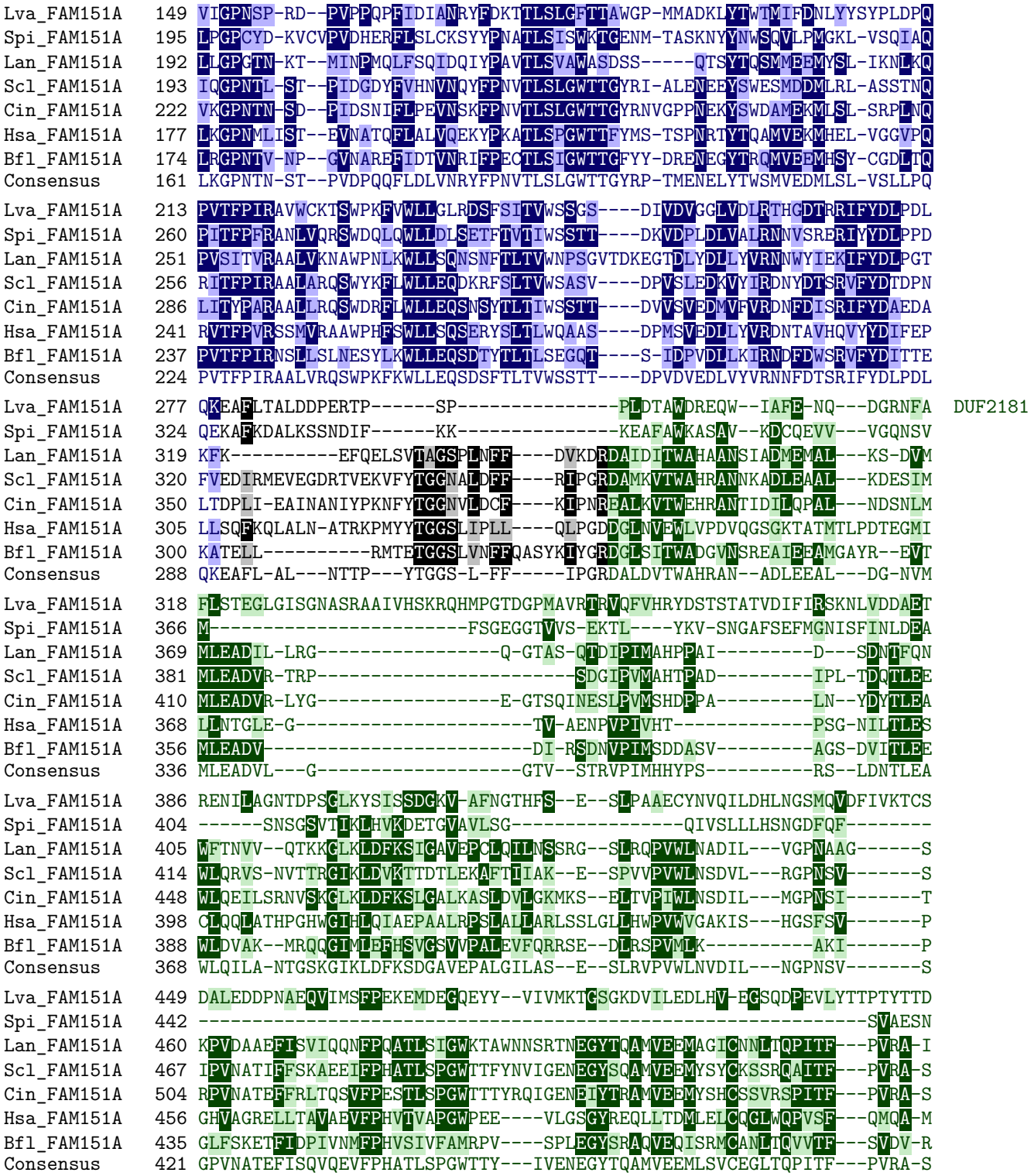

Cca_FAM151A 188 INRPVWINADILPGPNVPVFWPVINASEFFELIQLKFPDVTISPGWKVLYLS-IFPNVTYTRSMVEEM
Dre_FAM151A 190 INRPVWINADILPGPNVPEFWPVVNASEFFELIQLKFPDVTISPGWKVLYLS-IFPNVTYTRSMVEQM
Hsa_FAM151A 166 VRRPIWINADILKGPNMLIS-TEVNATQFLALVQEKYPKATLSPGWTTFYMS-TSPNRTYTQAMVEKM
Ptr_FAM151A 166 VRRPIWINADILKGPNMLIS-TEVNATQFLALVQEKYPKATLSPGWTTFYVS-TSPNRTYTQAMVEKM
Pan_FAM151A 166 VRRPVWINADILKGPNMLIS-TEVNATQFLALVQEKYPKATLSPGWTTFYMS-TFPNRTYTRAMVEKM
Mmu_FAM151A 170 IRRPVWINADILRGPNVPIS-IEINATQFLTLVQEKYPKATISPGFTTLYVP-QLPNSTYTQAMVETM
Eca_FAM151A 169 VRRPVWINADILRGPNVPIS-VEVNATRFLALVQEKYPEATLSPGWTTLYEP-LLPSGTYTRAMVEEM
Vpa_FAM151A 170 VRRPVWINADIQRGPNVPIP-IEINATRFLALAQEKYPEATLSTGWTTLYLP-MFPNSTYTRAMVEKM
Rty_FAM151A 184 FNRPVWLNADILIGPNVPGFMQPVNASRFLGLIQQRFPNVILSPGWMSLYLP-MIATKPYTRKMVEEM
Xla_FAM151A 163 INRPVWLNADILKGPNVNHE-IGVDATQFLNLVKNKFPDVTLSPGWVTLYLPPIISNRTYTREMIQQM
Bbu_FAM151A 163 INRPVWLNADILAGPNVNHE-IGVNATQFLNLIQERFPDITISPGWVTLYLPPIISNRTYSSEMVKKM
Rbi_FAM151A 177 ISRPVWINADILNGPNININ-IAVNATQFLDLVQRKFPNVTISPGWVTLYLP-FLSNKTYTWPMIWKM
Cpi_FAM151A 173 LNRPLWLNADILMGPNVPIN-TAVNASLFLSLIQEKYPNCTLSLGWTTLYSF-LFPNKTYTQKMIQKM
Asi_FAM151A 173 INRPVWLNADILEGPNVLVN-VSLNASTFLSLIQEKYPNCTLSPGWTTLYSP-LFPKQTYTRAMIQKM
Consensus 166 INRPVWINADILKGPNVPIS-IEVNATQFLALVQEKYPDATLSPGWTTLYLP-IFPNRTYTRAMVEKM
Cca_FAM151A 255 YTIVRHLPQKITFPVHALMAKNGWPHLSWLLSQSPRFSLTLWQGKENP-TVNDLLFIRDNSNPLRIYY
Dre_FAM151A 257 YSTIRHLPQKITFPVHALMAKNGWPHLSWLLSQSSRYSLTLWQGKENP-TLNDLLFIRDNSNPQRIYY
Hsa_FAM151A 232 HELVGGVPQRVTFPVRSSMVRAAWPHFSWLLSQSERYSLTLWQAASDPMSVEDLLYVRDNTAVHQVYY
Ptr_FAM151A 232 HELVGEVPQRVTFPVRSSMVRVAWPHFSWLLSQSERYSLTLWQAASDPMSVEDLLYVRDNTAVHQVYY
Pan_FAM151A 232 HELVGVVPQRVTFPVRSSMVRAAWPHFSWLLSQSERYSLTLWQAASDPMSVEDLLYVRDNTAVHQVYY
Mmu_FAM151A 236 QELVGALPQKVTFPVRAVMTRAAWPHFSWLLSQSERYSLTLWQGASDPVSVEDLLFIRDNSAAHQIYY
Eca_FAM151A 235 QGLVGVLPQRVTFPVRAVMARAAWPHFSWLLGQSERYSLTLWQGASDPVSVDDLLYIRDNSATHQVYY
Vpa_FAM151A 236 QELVGALPQKVTFPLYALMARSAWPHFSWLLGQSERYSLTLWQATSDSVSVDDLLYIRDNTAPHQVYY
Rty_FAM151A 251 YDLVKGLSQRVTFPVRAVLLKPAWPHFSWLLSQSPRYSLTLWQGSIDPVTVEDLLFFRDNSNVEQIYY
Xla_FAM151A 230 YNMVRDLPQKITYPARAVMTRSAWPHFNWLLQQSERYTITLWQGKSDPLTLEDLLFIRDSSNPEEIYY
Bbu_FAM151A 230 YNLVKGLTQRITFPARAVLTCSAWQNFYWLLKQSDRYSLTLWQGSSDPLQLDDLLFIRDNSRPEEIYY
Rbi_FAM151A 243 YTLVRDLPQRITFPVRAVMIKSAWQYFSWLLQQSDRYSLTLWQGETDPITVEDLLYVRDNSRAEEIYY
Cpi_FAM151A 239 HSIVGTLPQRVTFPVRAVMVRLAWPHFSWLLAQSDRYSLTLWQGKMDPIRVEDLLFIRDNSRPEQIYY
Asi_FAM151A 239 HDLIGELPQKVTFPVRAIMVRLAWPHFSWLLNQSERYSLTLWQGKTDPVTVEDLLFIRDNSRAEQIYY
Consensus 232 YELVGGLPQRVTFPVRAVMVRAAWPHFSWLLSQSERYSLTLWQGKSDPVTVEDLLFIRDNSAPHQIYY
Cca_FAM151A 322 DIYEPVLSQFKEAAKLRNRPRRFYPGGDIIDYFRPVNNDGLNIQWDTVTDK---DDLLYLLKDSQGGM DUF2181
Dre_FAM151A 324 DIYEPVLSQFREAAKIKDRPRRFYPGGDIVDYFRPADSDGLNIQWDTVNDK---DSLLSLLEDSPGGM
Hsa_FAM151A 300 DIFEPLLSQFKQLALNATRKPMYYTGGSLIPLLQLPGDDGLNVEWLVPDVQGSGKTATMTL-PDTEGM
Ptr_FAM151A 300 DIFEPLLSQFKQLALNATRKPMYYTGGSLIPLLQLPGDDGLNVEWLVPDVQGSSKTATITL-PDTEGM
Pan_FAM151A 300 DIFEPLLSQFKQLALNATRKPMYYTGGSLIPLLQLPGDDGLNVEWLVPDVQGSGQTATMTL-PDTEGM
Mmu_FAM151A 304 DLFEPVLSQFKQLALNTTRKRTYYTGGSLIPLLQQPKGDGLEVEWLVLEVNGSGRRAAITV-PDREGM
Eca_FAM151A 303 DLFEPVLSQFKQLAVNTTRKRSYYTGGSLVPLLQLPRGDGLSVEWLVPEVQGKGRTATVQV-PDREGM
Vpa_FAM151A 304 DIFEPVLSQFRQLAMNASRKQNYYTGGSLIPFLQLPGDNSLSVEWLVPDIQGNGSTATVGL-PDREGM
Rty_FAM151A 319 DIYEPVLSEFKQIALQTNRTRRFYPGGKLMDYFPHQNLDPLQIKWFDIGSTE--LELMKLLQGNIGGM
Xla_FAM151A 298 DIFEPLLSEFKEAALNPNRKRLFYPGGSIQLYFQPEDSDGLLVNWYEADAD--ILSEKEFF-SSNSGM
Bbu_FAM151A 298 DIYDPLLSEFKQQALNTSRRKLFYTGGSLQMYFHPDDHDGISVKWFDAEEN--ISTVQNLL-ASSFGM
Rbi_FAM151A 311 DIYDPVLSQFKEVALKPDRRRLFYTGGNLLQYFHPDDSDGLLVNWYVVKN---KTALLLLL-TGRTGM
Cpi_FAM151A 307 DIYDPVLAQFKEAALNSTRKRFYYPGGNLLDYFHPADSDPLQIEWYGMDHYENRLETLSIL-KDKRGM
Asi_FAM151A 307 DIYDPVLSQFKEMALKSTRKRAYYPGGDLLEYFHPSNGDGLSIEWYAMEHNESKTSTSSML-TDRSGM
Consensus 300 DIYEPVLSQFKQLALNATRKRRYYTGGSLIPYFQPPDSDGLNVEWLVPDVQG-GLTATSLL-PDREGM

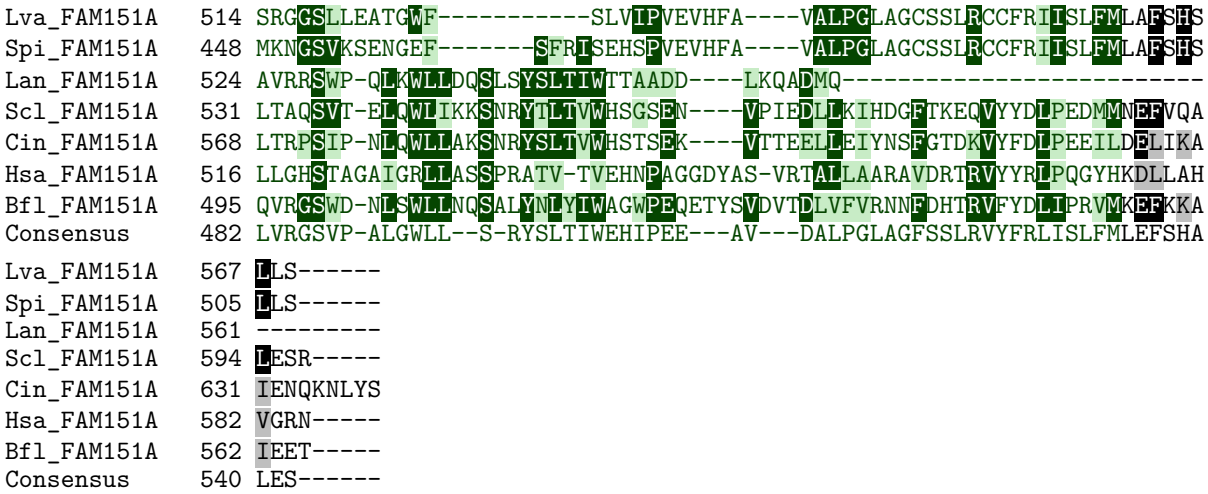

Cca_FAM151A 387 LVIPVISSEGQ--PNIPII--QGSKPELPLQDCLELILASKKSWGIYLRIKSQNQLSLTLELLRQAYD
Dre_FAM151A 389 LVIPVKSSDGH--PNIPII--DGS--EMPLKDCLDLILASTKPWGIYLQIKSQNQLSLSLELLRQAYD
Hsa_FAM151A 367 ILLNTGLEGTVAENPVPIVHTPS-GNILTLESCLQQLATHPGHWGIHLQIAEPAALRPSLALLARLSS
Ptr_FAM151A 367 ILLNTGLEGTVAENPVPIVHTPS-GSILTLESCLQQLATHPGHWGIHLQIAEPAALRPSLALLARLSS
Pan_FAM151A 367 ILLNTGLEGTVAENPVPIVHAPS-GSILTLESCLQQLATRPGHWGIHLQIAEPAALRPSLALLAHLSS
Mmu_FAM151A 371 ILLDIGLQEPEAGNPVPILHTPG-GPALTLESCLRLAVHPRRWGIHVNIVEPEALRPSLATLAHLST
Eca_FAM151A 370 ILLNIGLRGPAAGDPVPIVRVPG-GPALTLESCLLQLATHPGRWGIHLHIAEPTALRPSLAMLAHLST
Vpa_FAM151A 371 ILLNVGLQEPAAENPMPFVRAPD-GRALTLESCLQQLATHPRRWGVHVHIAEPTALRPTLAMLAHLSA
Rty_FAM151A 385 LILNVKTKA---NDAIPVVVGAEKGLEFPLESALNWIATSSKPWGIYLKIKSQEALAPTLYLLERVYT
Xla_FAM151A 363 ITLNIRVKDSS---SSPQVAFPKSPTQFSLEDYMNVILANPNPWGVFLKIETQDALNKTLKVLSRMHD
Bbu_FAM151A 363 LTLHVEV--QS---RSPVVIFAKSSAAFPLEDLLKLINSNKNLWGVFLKPKDHVSLNETLHALKRLND
Rbi_FAM151A 375 LALAIGAE-VVNGTLIPVVRLPQSAADLPLEHCLDLIYTCQHAWGVFLQIETEAALPPALHLLSKLQG
Cpi_FAM151A 374 IALDIALQNSTIGNLIPVALSP--SAGLPLEQCLVTVSRDLNPWGIYLNIIEPGALRPTLELLSKLYA
Asi_FAM151A 374 IVLDVAVQDGISGNLIPVA-----STGTPLEQCLETIYRSQNPWGIFLNVTEPDALHPTLELLSTVYA
Consensus 366 ILLNVGLEGGVAGNPIPIVH-PG-GPALPLESCLQQIATHPGPWGIHLQIAEPAALRPTLALLARLSD
Cca_FAM151A 451 RDLLHHPTWVNMDIAHGTFYIQDYVTGEEFLRTIDQIFPYVTLAPGWPKEVLDEGYKPELVEDMVQLF
Dre_FAM151A 451 IDLLHHPTWVNMDISHGAVHIQGYMTGEEFLRTVDRIFPHVTLAPSWPKEALVEGYTPEMLEPMVQLF
Hsa_FAM151A 434 LGLLHWPVWVGAKISHGSFSVPGHVAGRELLTAVAEVFPHVTVAPGWPEEVLGSGYREQLLTDMLELC
Ptr_FAM151A 434 LGLLHWPVWVGAKISHGSFSVPGHVAGRELLTAVAEVFPHVTVAPGWPEEVLGSGYREQLLTDMLELC
Pan_FAM151A 434 LGLLHWPVWVGAKISHRSFSVPGHVAGRELLTAVAEVFPYVTVAPGWPEEVLGSGYREHLLTDMLELC
Mmu_FAM151A 438 LGHLPWPVWVGSTVSHGSFVVPGHIAGRELLTAVAEVFPHVTVAPGWPEEMLDSGYQEQMVTDMLELC
Eca_FAM151A 437 LGHLPWPVWVGATVSHGSFVVPGHVAGRELLTAVAEVFPHVTVAPGWPEEVLGSGYREQLLTDMLELC
Vpa_FAM151A 438 LGHLSRPVWVGATISHGSFVVPGYMAGKEFLTAMAEVFPHVTVAPRWPEEVLGSGYREQLLTDMLDLC
Rty_FAM151A 450 AYLLLKPVWINMDLSYGSFSTHGYIEGKQFIKTVNEIFPFVTIAPSWPKEVLTHGYTQPLVEDMLNLC
Xla_FAM151A 428 HKALNVPVWISMEVSYGNFSMEGYIQGIDFLNTINDIFPYVTIAPSWPAPVLGSGYTEILVQDMLMLC
Bbu_FAM151A 426 QKSLYLPVWIGMDVSYKSFSTPGYIYGEDFIGSINAIFSAVTIAPGWPIERLDGGYTELMVQDMLQLC
Rbi_FAM151A 442 RNLLWHPIWISMAVSYGRFSAPGYMPGRDFLATINAIFPFVTIAPSWPKESLAGGYTDPLIEDMLSLC
Cpi_FAM151A 440 QNLLWNPIWISMALSFGSFETPGYMQGEEFLIAINSIFPYVTIAPGWPREVITAGYTDPLIDDMLILC
Asi_FAM151A 437 KNLLWSPVWVSLAVSYRSFDTPGYMHGEDFLRAINTIFPHVTIAPRWPREVLTDGYTDLLIEDMLILC
Consensus 432 LGLLHWPVWVGMDISHGSFSVPGYVAGREFLTAVAEIFPHVTVAPGWPEEVLGSGYTEQLLTDMLELC
Cca_FAM151A 519 QGAWQDVSLQLHAETLYRTVTGC-RSLLHAQSRFSMTLEHRAEDRDLNTWTASLMAIRALNRQRSFYN
Dre_FAM151A 519 HRAWQDVSLQLQAEALDRSET---WRLVLVQPRFSLTVEHQTENKDINAGIESLMAIRAANRQRSFYN
Hsa_FAM151A 502 QGLWQPVSFQMQAMLLGHSTAGAIGRLLASSPRATVTVEHNPAGGDYASVRTALLAARAVDRTRVYYR
Ptr_FAM151A 502 QGLWQPVSFQMQAMLLGHSTAGAIARLLASSPRATVTVEHNPAGGDYASVRTALLAARAVDRTRVYYR
Pan_FAM151A 502 QGLWQPVSFQMQAILLGHSTAGAIARLLASSPRATVTVEHSPAGGDYASVRTALLAARAVDRTRVYYR
Mmu_FAM151A 506 QGLRQPVSFQLQAGPLSQSPANTVARLLASSPRATVTVYHSTAGNSHVDLWAGLWAARAVDRTRVYYR
Eca_FAM151A 505 QGLWQPVSFQLQAGPLGWSRAAAVARLLAASARATVTVEHSPAGGNYASVRAVLLAARAVDRTRVYYR
Vpa_FAM151A 506 QGLWQPVSFQLQAGPLGQSTAGVVDRLLAASPRATVTVEHNPGRGNYASVRGVLLAARAVDKTRVYYK
Rty_FAM151A 518 RGLWQAVSFQLQAIALGKSWKAT-TRLLQTSPTYTLTVEHLHEQGSYLDGFQGLINTRTYSTRRIYYR
Xla_FAM151A 496 EGLWQEVSFQLNAVALGKEWLSA-VKLLQVSPMYSLTIEHNSKQGIFLDGYAGLMAMRSHEENRIYYR
Bbu_FAM151A 494 EGVMQEVSFQLQAVILGKAWLNT-VNLMKVSRMYTLTVEHTAEQGTFMDGYHGLMAIRTHTENGVYYK
Rbi_FAM151A 510 QGLWQEVSFQLQAAALADTWKTA-VGLLEVSPSYTLTVEHGHAQGSFWDGYQGLMSVRTHTKERVYYS
Cpi_FAM151A 508 KDLWQQVSFQLEAVPLSRSWLAT-TKLLEISPSYTITVQHSHSEGSYCDGFPGLRSIRTHTQKGVYYK
Asi_FAM151A 505 KGLWQHVSFQLEAVLLSKSWLPT-AKLLEASPSYTVTVQHSSSEGSDWDGFPGLQSISTQAQKRVYYT
Consensus 500 QGLWQPVSFQLQAVALGKSWAGA-ARLLAASPRATVTVEHSPAGGDYADGRAGLMAIRAVDRTRVYYR



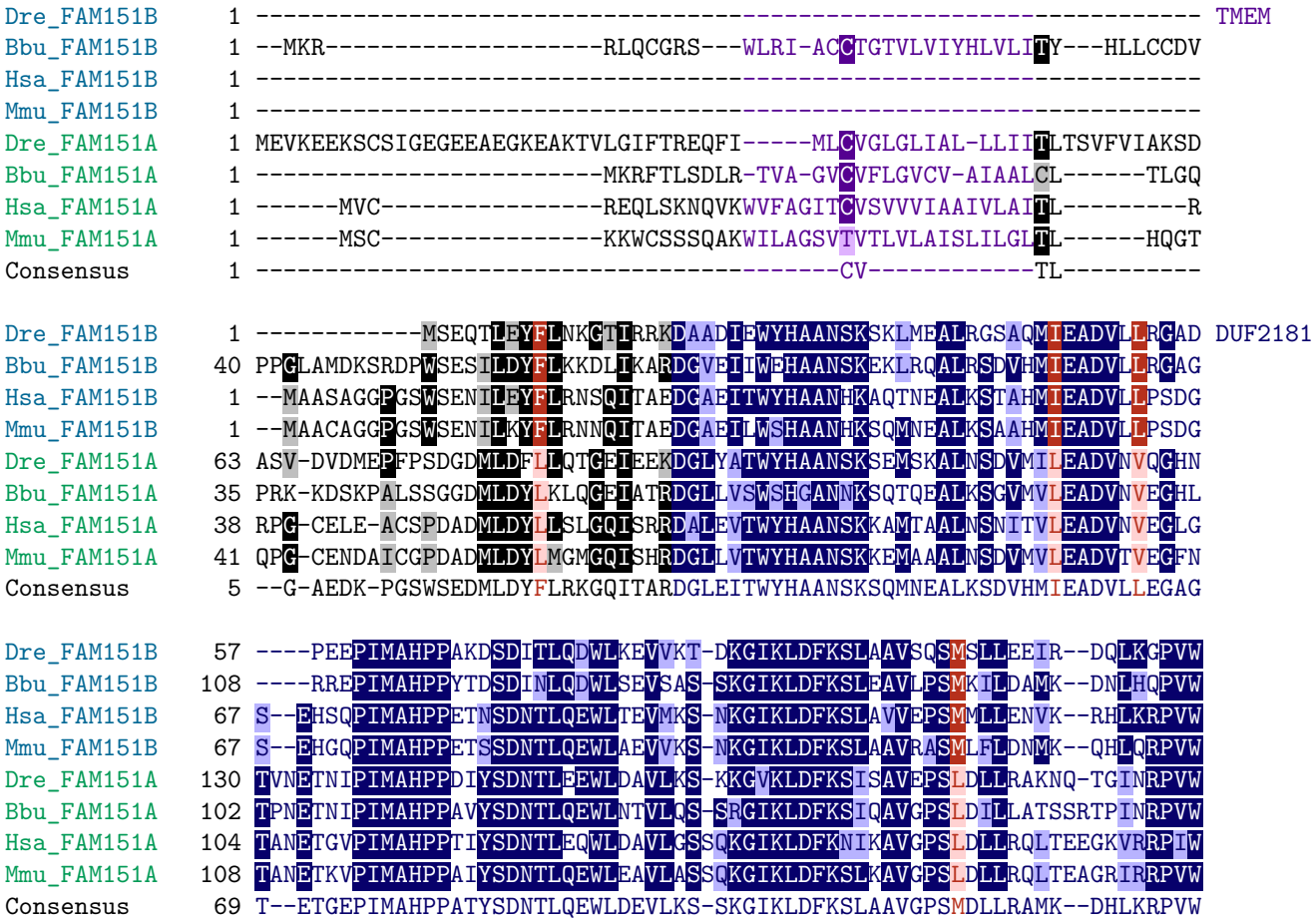
Distant Orthologs

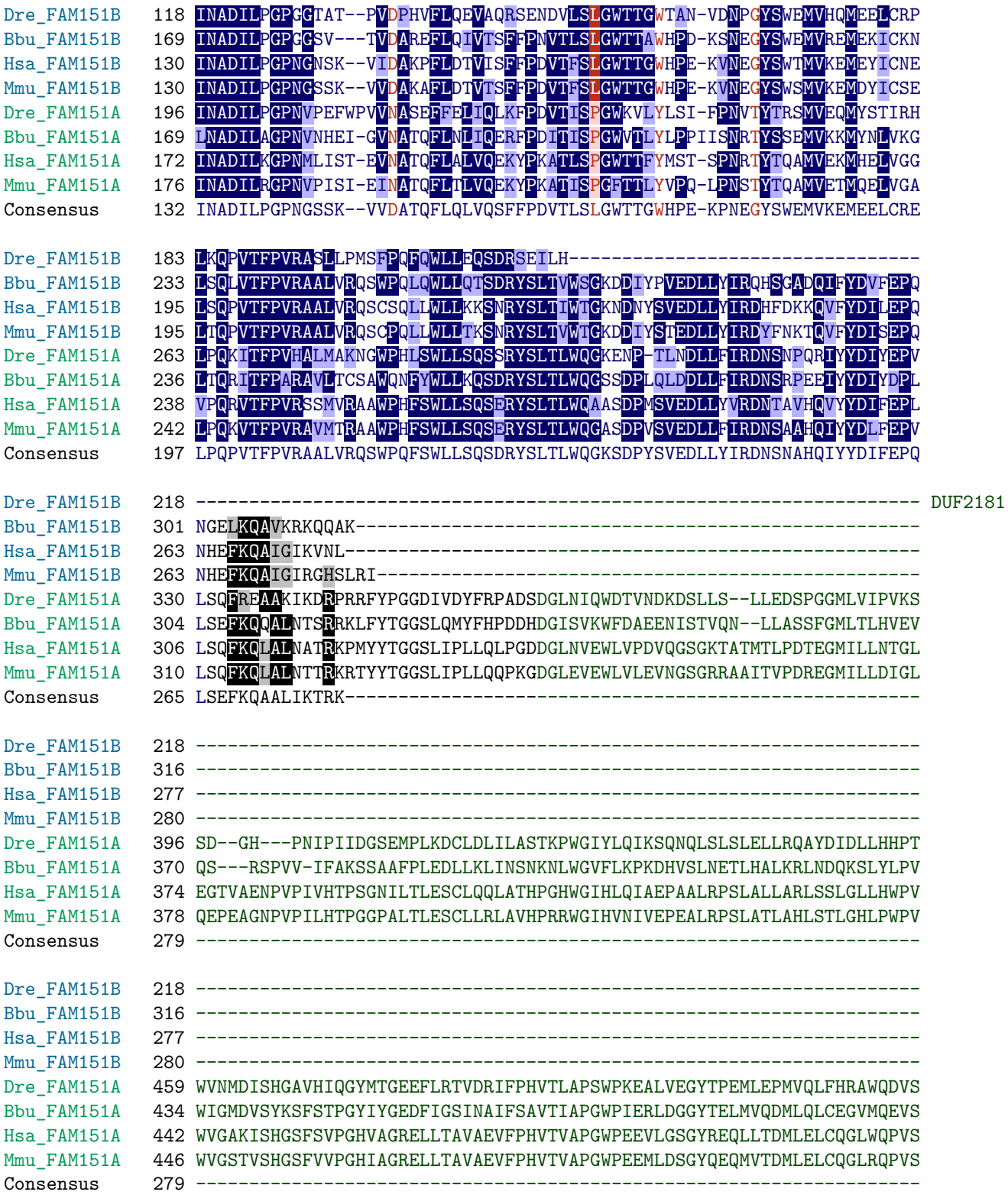



Lva_FAM151A 149 VIGPNSP-RD--PVPPQPFIDIANRYFDKTTLSLGFTTAWGP-MMADKLYTWTMIFDNLYYSYPLDPQ
Spi_FAM151A 195 LPGPCYD-KVCVPVDHERFLSLCKSYYPNATLSISWKTGENM-TASKNYYNWSQVLPMGKL-VSQIAQ
Lan_FAM151A 192 LLGPGTN-KT--MINPMQLFSQIDQIYPAVTLSVAWASDSS-----QTSYTQSMMEPMYSL-IKNLKQ
Scl_FAM151A 193 IQGPNTL-ST--PIDGDYFVHNVNQYFPNVTLSLGWTTGYRI-ALENEEYSWESMDDMLRL-ASSTNQ
Cin_FAM151A 222 VKGPNTN-SD--PIDSNIFLPEVNSKFPNVTLSLGWTTGYRNVGPPNEKYSWDAMEKMLSL-SRPLNQ
Hsa_FAM151A 177 LKGPNMLIST--EVNATQFLALVQEKYPKATLSPGWTTFYMS-TSPNRTYTQAMVEKMHEL-VGGVPQ
Bfl_FAM151A 174 LRGPNTV-NP--GVNARPFIDTVNRIFPECTLSIGWTTGFYY-DRENEGYTRQMVEPMHSY-CGDLIQ
Consensus 161 LKGPNTN-ST--PVDPQQFLDLVNRYFPNVTLSLGWTTGYRP-TMENELYTWSMVEDMLSL-VSLLPQ
Lva_FAM151A 213 PVTFPIRAVWCKTSWPKFVWLLGLRDSFSITVWSSGS----DIVDVGGLVDLRTHGDTRRIFYDLPDL
Spi_FAM151A 260 PITFPFRANLVQRSWDQLQWLLDLSETFTVTIWSSTT----DKVDPLDLVALRNNVSRERIYYDLPPD
Lan_FAM151A 251 PVSITVRAALVKNAWPNLKWLLSQNSNFTLTVWNPSGVTDKEGTDLYDLLYVRNNWYIEKIFYDLPGT
Scl_FAM151A 256 RITFPIRAALARQSWYKFLWLLEQDKRFSLTVWSASV----DPVSLEDKVYIRDNYDTSRVFYDTDPN
Cin_FAM151A 286 LITYPARAALLRQSWDRFLWLLEQSNSYTLTIWSSTT----DVVSVEDMVFVRDNFDISRIFYDAEDA
Hsa_FAM151A 241 RVTFPVRSSMVRAAWPHFSWLLSQSERYSLTLWQAAS----DPMSVEDLLYVRDNTAVHQVYYDIFEP
Bfl_FAM151A 237 PVTFPIRNSLLSLNESYLKWLLEQSDTYTLTLSEGQT----S-IDPVDLLKIRNDFDWSRVFYDITTE
Consensus 224 PVTFPIRAALVRQSWPKFKWLLEQSDSFTLTVWSSTT----DPVDVEDLVYVRNNFDTSRIFYDLPDL
Lva_FAM151A 277 QKEAFLTALDDPERTP------SP---------------PLDTAWDREQW--IAFE-NQ---DGRNFA DUF2181
Spi_FAM151A 324 QEKAFKDALKSSNDIF------KK---------------KEAFAWKASAV--KDCQEVV---VGQNSV
Lan_FAM151A 319 KFK----------EFQELSVTAGSPLNFF----DVKDRDAIDITWAHAANSIADMEMAL---KS-DVM
Scl_FAM151A 320 FVEDIRMEVEGDRTVEKVFYTGGNALDFF----RIPGRDAMKVTWAHRANNKADLEAAL---KDESIM
Cin_FAM151A 350 LTDPLI-EAINANIYPKNFYTGGNVLDCF----KIPNREALKVTWEHRANTIDILQPAL---NDSNLM
Hsa_FAM151A 305 LLSQFKQLALN-ATRKPMYYTGGSLIPLL----QLPGDDGLNVPWLVPDVQGSGKTATMTLPDTEGMI
Bfl_FAM151A 300 KATELL----------RMTETGGSLVNFFQASYKIYGRDGLSITWADGVNSREAIEEAMGAYR--EVT
Consensus 288 QKEAFL-AL---NTTP---YTGGS-L-FF-----IPGRDALDVTWAHRAN--ADLEEAL---DG-NVM
Lva_FAM151A 318 FLSTEGLGISGNASRAAIVHSKRQHMPGTDGPMAVRTRVQFVHRYDSTSTATVDIFIRSKNLVDDAET
Spi_FAM151A 366 M----------FSGEGGTVVS-EKTL----YKV-SNGAFSEFMGNISFINLDEA
Lan_FAM151A 369 MLEADIL-LRG----------Q-GTAS-QTDIPIMAHPPAI---------D---SDNTFQN
Scl_FAM151A 381 MLEADVR-TRP----------SDGIPVMAHTPAD---------IPL-TDQTLEE
Cin_FAM151A 410 MLEADVR-LYG----------E-GTSQINESLPVMSHDPPA---------LN--YDYTLEA
Hsa_FAM151A 368 LLNTGLE-G----------TV-AENPVPIVHT----------PSG-NILTLES
Bfl_FAM151A 356 MLEADV----------DI-RSDNVPIMSDDASV---------AGS-DVITLEE
Consensus 336 MLEADVL---G----------GTV--STRVPIMHHYPS---------RS--LDNTLEA
Lva_FAM151A 386 RENILAGNTDPSGLKYSISSDGKV-AFNGTHFS--E--SLPAAECYNVQILDHLNGSMQVDFIVKTCS
Spi_FAM151A 404 ------SNSGSVTIKLHVKDETGVAVLSG----------QIVSLLLHSNGDFQF--------
Lan_FAM151A 405 WFTNVV--QTKKGLKLDFKSIGAVEPCLQILNSSRG--SLRQPVWLNADIL---VGPNAAG------S
Scl_FAM151A 414 WLQRVS-NVTTRGIKLDVKTTDTLEKAFTIIAK--E--SPVVPVWLNSDVL---RGPNSV-------S
Cin_FAM151A 448 WLQEILSRNVSKGLKLDFKSLGALKASLDVLGKMKS--ELTVPIWLNSDIL---MGPNSI-------T
Hsa_FAM151A 398 CLQQLATHPGHWGIHLQIAEPAALRPSLALLARLSSLGLLHWPVWVGAKIS---HGSFSV-------P
Bfl_FAM151A 388 WLDVAK--MRQQGIMLPFHSVGSVVPALEVFQRRSE--DLRSPVMLK----------AKI-------P
Consensus 368 WLQILA-NTGSKGIKLDFKSDGAVEPALGILAS--E--SLRVPVWLNVDIL---NGPNSV-------S
Lva_FAM151A 449 DALEDDPNAEQVIMSFPEKEMDEGQEYY--VIVMKTGSGKDVILEDLHV-EGSQDPEVLYTTPTYTTD
Spi_FAM151A 442 ----------SVAESN
Lan_FAM151A 460 KPVDAAEFISVIQQNFPQATLSIGWKTAWNNSRTNEGYTQAMVEEMAGICNNLTQPITF---PVRA-I
Scl_FAM151A 467 IPVNATIGFSKAEEIFPHATLSPGWTTFYNVIGENEGYSQAMVEEMYSYCKSSRQAITF---PVRA-S
Cin_FAM151A 504 RPVNATEGFRLTQSVFPESTLSPGWTTTYRQIGENEIYTRAMVEEMYSHCSSVRSPITF---PVRA-S
Hsa_FAM151A 456 GHVAGRELLTAVAEVFPHVIVAPGWPEE----VLGSGYREQLLTDMLELCQGLWQPVSF---QMQA-M
Bfl_FAM151A 435 GLFSKETFIDPIVNMFPHVSIVFAMRPV----SPLEGYSRAQVEQISRMCANLTQVVTF---SVDV-R
Consensus 421 GPVNATEFISQVQEVFPHATLSPGWTTY---IVENEGYTQAMVEEMLSVCEGLTQPITF---PVRA-S



Alignment of FAM151A and FAM151B Orthologs

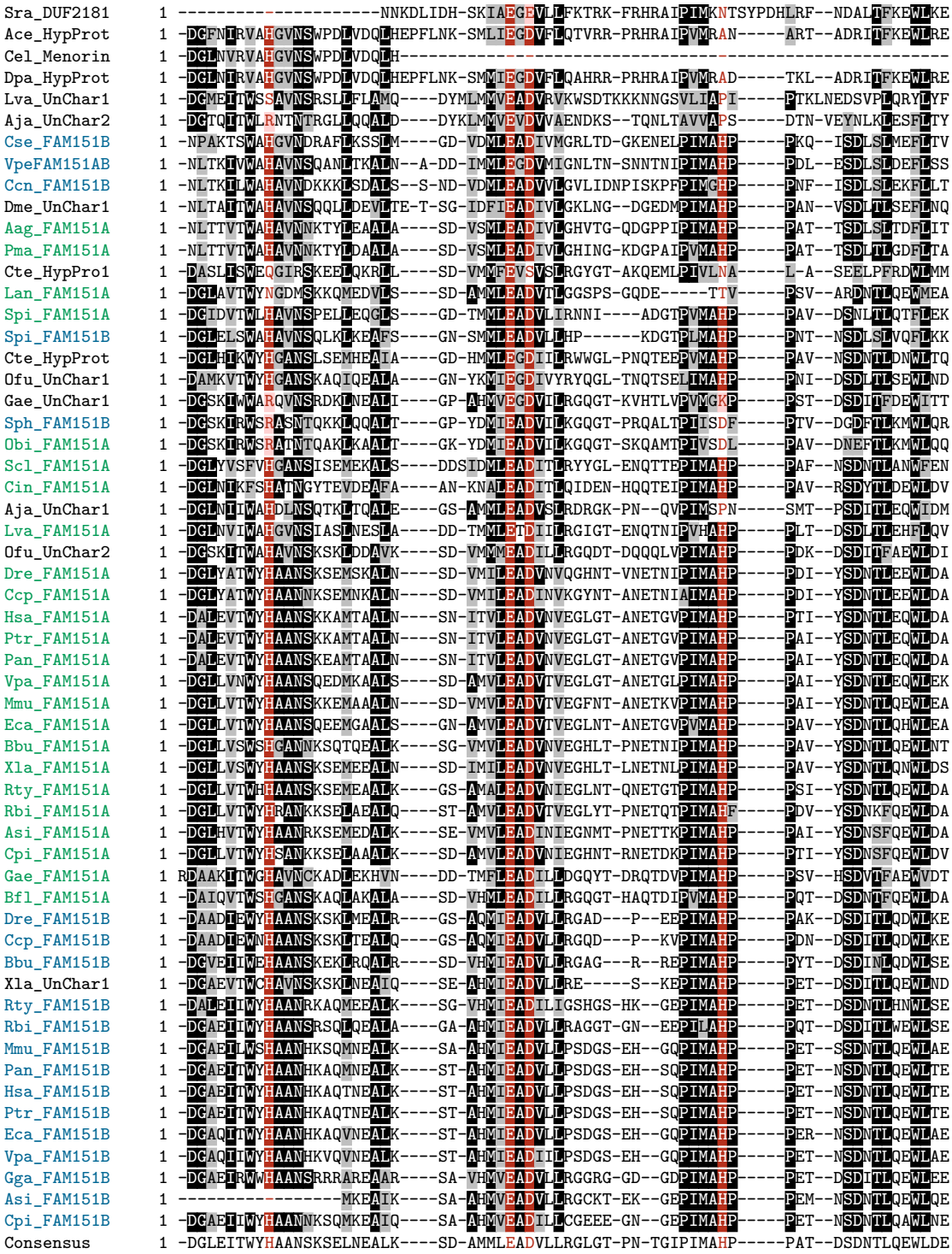


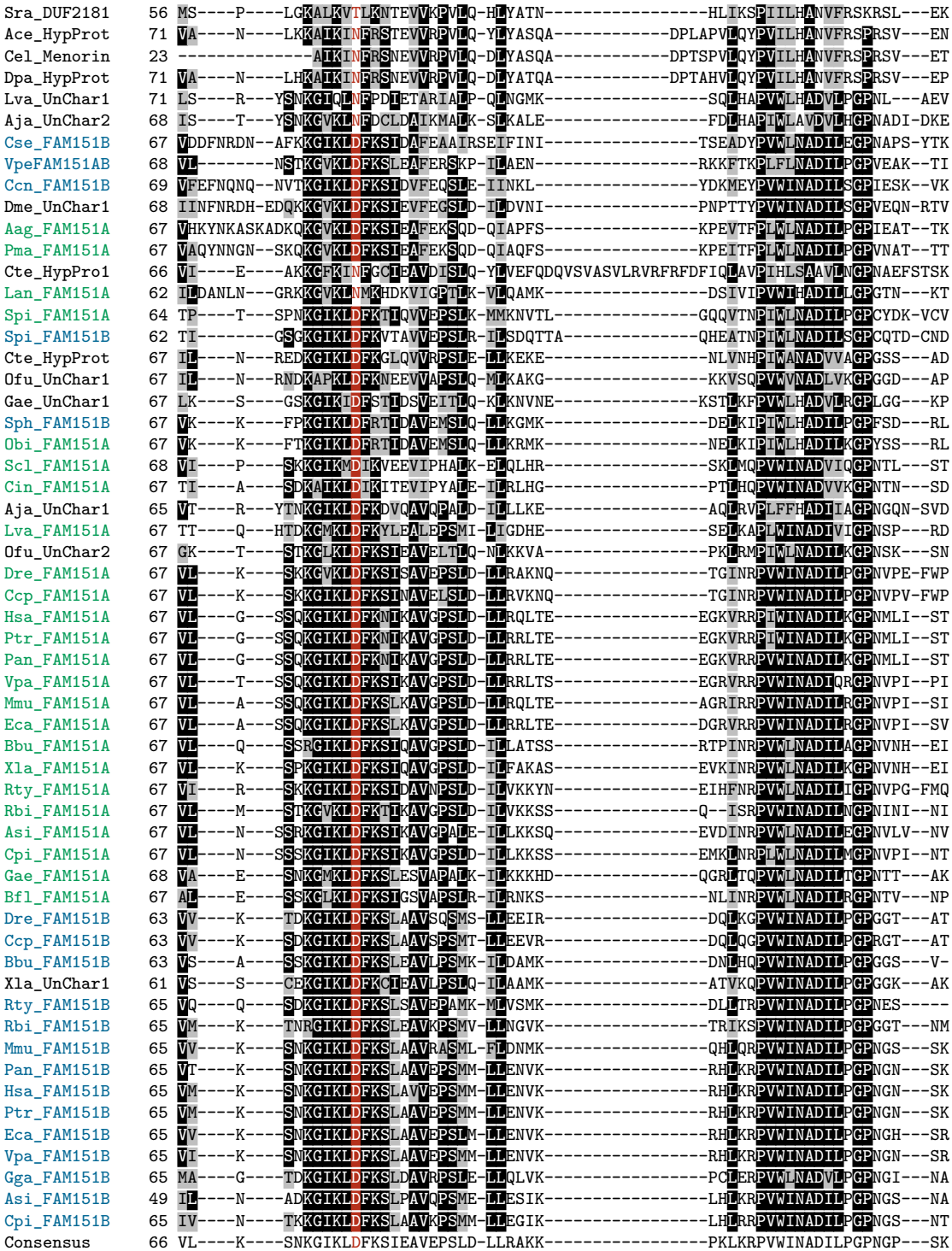

Dre_FAM151B 118 INADILPGPGGTAT--PVDPHVFLQEVAQRSENDVLSLGWTTGWTAN-VDNPGYSWEMVHQMEELCRP
Bbu_FAM151B 169 INADILPGPGGSV---TVDAREFLQIVTSFFPNVTLSLGWTTAWHPD-KSNEGYSWEMVREMEKICKN
Hsa_FAM151B 130 INADILPGPNGNSK--VIDAKPFLDTVISFFPDVTFSLGWTTGWHPE-KVNEGYSWTMVKEMEYICNE
Mmu_FAM151B 130 INADILPGPNGSSK--VVDAKAFLDTVTSFFPDVTFSLGWTTGWHPE-KVNEGYSWSMVKEMDYICSE
Dre_FAM151A 196 INADILPGPNVPEFWPVVNASEFFPLIQLKFPDVTISPGWKVLYLSI-FPNVTYTRSMVEQMYSTIRH
Bbu_FAM151A 169 LNADILAGPNVNHEI-GVNATQFLNLIQERFPDITISPGWVILYLPPIISNRTYSSEMVKKMYNLVKG
Hsa_FAM151A 172 INADILKGPNMLIST-EVNATQFLALVQEKYPKATLSPGWTIFYMST-SPNRTYTQAMVEKMHELVGG
Mmu_FAM151A 176 INADILRGPNVPISI-EINATQFLTLVQEKYPKATISPGFTTLYVPQ-LPNSTYTQAMVETMQELVGA
Consensus 132 INADILPGPNGSSK--VVDATQFLQLVQSFFPDVTLSLGWTTGWHPE-KPNEGYSWEMVKEMEELCRE
Dre_FAM151B 183 LKQPVTFPVRASLLPMSFPQFQWLLEQSDRSEILH----------
Bbu_FAM151B 233 LSQLVTFPVRAALVRQSWPQLQWLLQTSDRYSLTVWSGKDDIYPVEDLLYIRQHSGADQIFYDVFEPQ
Hsa_FAM151B 195 LSQPVTFPVRAALVRQSCSQLLWLLKKSNRYSLTIWTGKNDNYSVEDLLYIRDHFDKKQVFYDILEPQ
Mmu_FAM151B 195 LTQPVTFPVRAALVRQSCPQLLWLLTKSNRYSLTVWTGKDDIYSTEDLLYIRDYFNKTQVFYDISEPQ
Dre_FAM151A 263 LPQKITFPVHALMAKNGWPHLSWLLSQSSRYSLTLWQGKENP-TLNDLLFIRDNSNPQRIYYDIYEPV
Bbu_FAM151A 236 LTQRITFPARAVLTCSAWQNFYWLLKQSDRYSLTLWQGSSDPLQLDDLLFIRDNSRPEEIYYDIYDPL
Hsa_FAM151A 238 VPQRVTFPVRSSMVRAAWPHFSWLLSQSERYSLTLWQAASDPMSVEDLLYVRDNTAVHQVYYDIFEPL
Mmu_FAM151A 242 LPQKVTFPVRAVMTRAAWPHFSWLLSQSERYSLTLWQGASDPVSVEDLLFIRDNSAAHQIYYDLFEPV
Consensus 197 LPQPVTFPVRAALVRQSWPQFSWLLSQSDRYSLTLWQGKSDPYSVEDLLYIRDNSNAHQIYYDIFEPQ
Dre_FAM151B 218 ---------- DUF2181
Bbu_FAM151B 301 NGELKQAVKRKQQAK----------
Hsa_FAM151B 263 NHEFKQAIGIKVNL----------
Mmu_FAM151B 263 NHEFKQAIGIRGHSLRI----------
Dre_FAM151A 330 LSQFREAAKIKDRPRRFYPGGDIVDYFRPADSDGLNIQWDTVNDKDSLLS--LLEDSPGGMLVIPVKS
Bbu_FAM151A 304 LSEFKQQALNTSRRKLFYTGGSLQMYFHPDDHDGISVKWFDAEENISTVQN--LLASSFGMLTLHVEV
Hsa_FAM151A 306 LSQFKQLALNATRKPMYYTGGSLIPLLQLPGDDGLNVEWLVPDVQGSGKTATMTLPDTEGMILLNTGL
Mmu_FAM151A 310 LSQFKQLALNTTRKRTYYTGGSLIPLLQQPKGDGLEVEWLVLEVNGSGRRAAITVPDREGMILLDIGL
Consensus 265 LSEFKQAALIKTRK----------
Dre_FAM151B 218 ----------
Bbu_FAM151B 316 ----------
Hsa_FAM151B 277 ----------
Mmu_FAM151B 280 ----------
Dre_FAM151A 396 SD--GH---PNIPIIDGSEMPLKDCLDLILASTKPWGIYLQIKSQNQLSLSLELLRQAYDIDLLHHPT
Bbu_FAM151A 370 QS---RSPVV-IFAKSSAAFPLEDLLKLINSNKNLWGVFLKPKDHVSLNETLHALKRLNDQKSLYLPV
Hsa_FAM151A 374 EGTVAENPVPIVHTPSGNILTLESCLQQLATHPGHWGIHLQIAEPAALRPSLALLARLSSLGLLHWPV
Mmu_FAM151A 378 QEPEAGNPVPILHTPGGPALTLESCLLRLAVHPRRWGIHVNIVEPEALRPSLATLAHLSTLGHLPWPV
Consensus 279 ----------
Dre_FAM151B 218 ----------
Bbu_FAM151B 316 ----------
Hsa_FAM151B 277 ----------
Mmu_FAM151B 280 ----------
Dre_FAM151A 459 WVNMDISHGAVHIQGYMTGEEFLRTVDRIFPHVTLAPSWPKEALVEGYTPEMLEPMVQLFHRAWQDVS
Bbu_FAM151A 434 WIGMDVSYKSFSTPGYIYGEDFIGSINAIFSAVTIAPGWPIERLDGGYTELMVQDMLQLCEGVMQEVS
Hsa_FAM151A 442 WVGAKISHGSFSVPGHVAGRELLTAVAEVFPHVTVAPGWPEEVLGSGYREQLLTDMLELCQGLWQPVS
Mmu_FAM151A 446 WVGSTVSHGSFVVPGHIAGRELLTAVAEVFPHVTVAPGWPEEMLDSGYQEQMVTDMLELCQGLRQPVS
Consensus 279 ----------

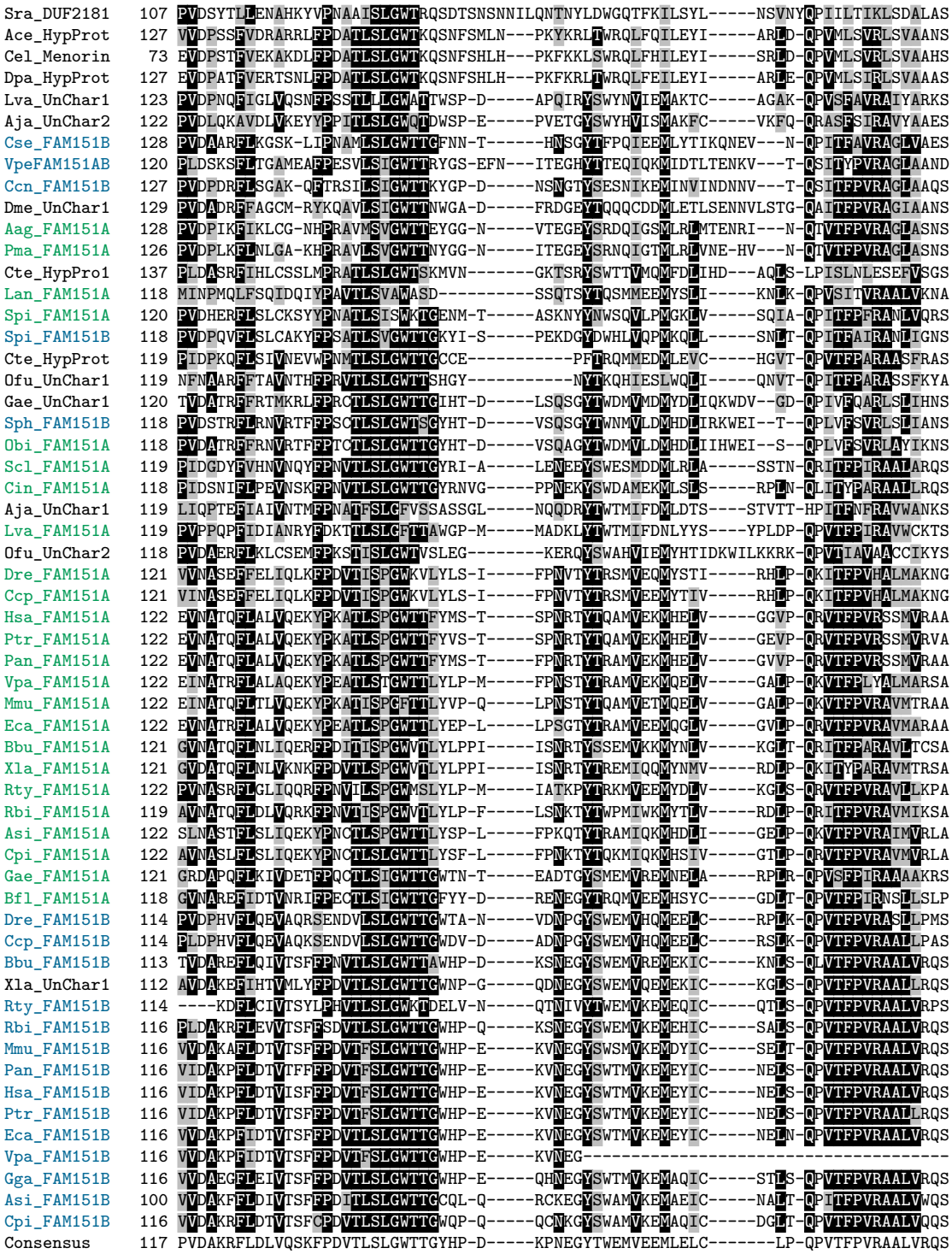
```
Dre_FAM151B   218 ------------------------------------------------------------
Bbu_FAM151B   316 ------------------------------------------------------------
Hsa_FAM151B   277 ------------------------------------------------------------
Mmu_FAM151B   280 ------------------------------------------------------------
Dre_FAM151A   527 LQLQAEALDRSE---TWRLVLVQPRFSLTVEHQTENKDINAGIESLMAIRAANRQRSFYNIPKMYREH
Bbu_FAM151A   502 FQLQAVILGKAWLNTVN-LMKVSRMYTLTVEHTAEQGTFMDGYHGLMAIRTHTENGVYYKLPPDYYYS
Hsa_FAM151A   510 FQMQAMLLGHSTAGAIGRLLASSPRATVTVEHNPAGGDYASVRTALLAARAVDRTRVYYRLPQGYHKD
Mmu_FAM151A   514 FQLQAGPLSQSPANTVARLLASSPRATVTVYHSTAGNSHVDLWAGLWAARAVDRTRVYYRISQEYWKD
Consensus     279 ------------------------------------------------------------

Dre_FAM151B   218 --------------------------
Bbu_FAM151B   316 --------------------------
Hsa_FAM151B   277 --------------------------
Mmu_FAM151B   280 --------------------------
Dre_FAM151A   592 ITDLSVRK------------------
Bbu_FAM151A   569 LMTSIYST------------------
Hsa_FAM151A   578 LLAHVGRN------------------
Mmu_FAM151A   582 LQADVSSNRPSSRIGPSSVEGFPGESR
Consensus     279 --------------------------
```

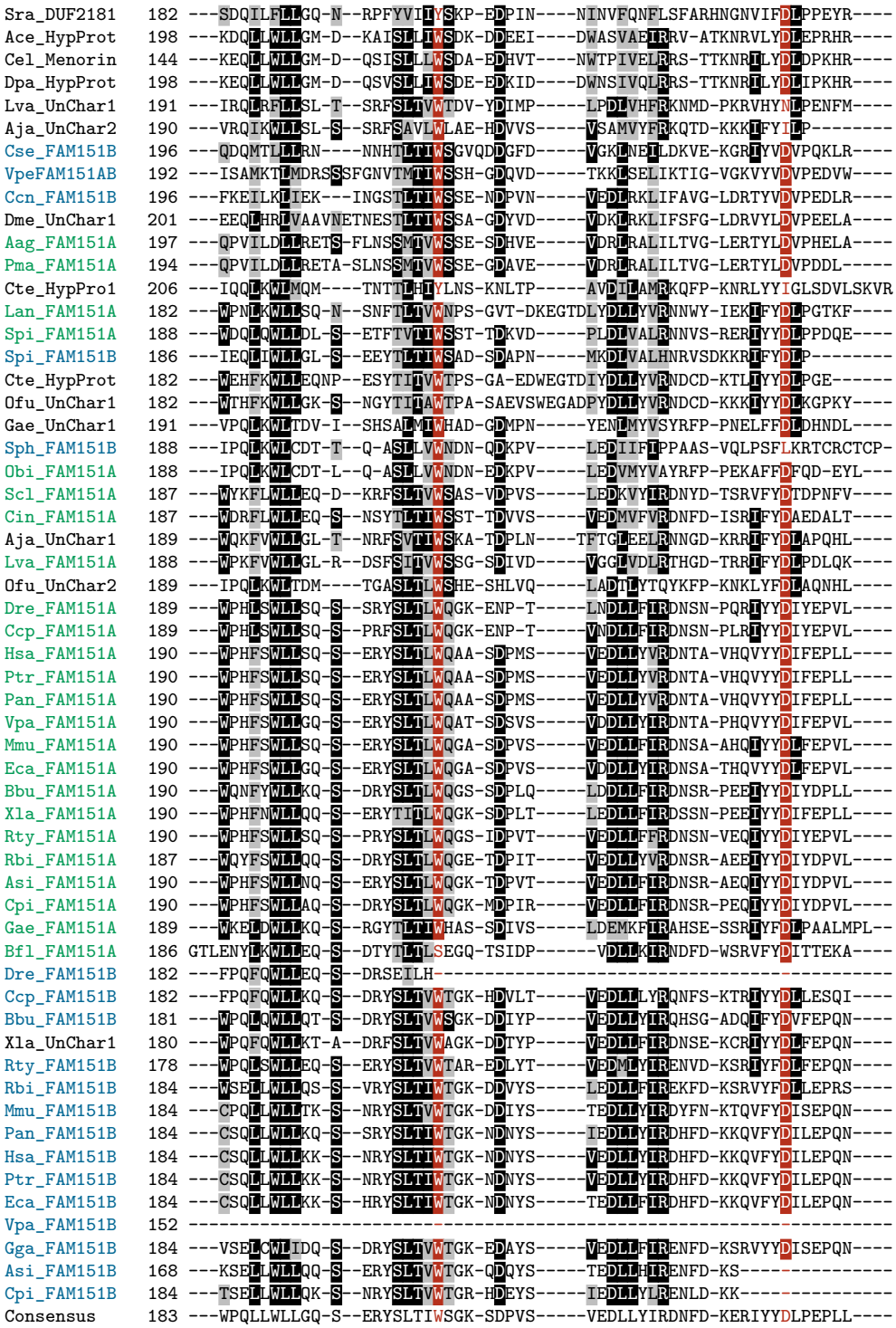
Appendix C: Alignments of DUF2181 Domains

Alignment of DUF2181 Domain in All Orthologs




Sra_DUF2181 56 MS----P----LGKALKVTLKNTEVVKPVLQ-HLYATN----------HLIKSPIILHANVFRSKRSL---EK
Ace_HypProt 71 VA----N----LKKAIKINFRSTEVVRPVLQ-YLYASQA----------DPLAPVLQYPVILHANVFRSPRSV---EN
Cel_Menorin 23 ----------AIKINFRSNEVVRPVLQ-DLYASQA----------DPTSPVLQYPVILHANVFRSPRSV---ET
Dpa_HypProt 71 VA----N----LHKAIKINFRSNEVVRPVLQ-DLYATQA----------DPTAHVLQYPVILHANVFRSPRSV---EP
Lva_UnChar1 71 LS----R---YSNKGIQLNFPDIETARIALP-QLNGMK----------SQLHAPVWLHADVLPGPNL---AEV
Aja_UnChar2 68 IS----T---YSNKGVKLNFDCLDAIKMAIK-SLKALE----------FDLHAPIWLAVDVLHGPNADI-DKE
Cse_FAM151B 67 VDDFNRDN--AFKKGIKLDFKSIDAFEAAIRSEIFINI----------TSEADYPVWLNADILEGPNAPS-YTK
VpeFAM151AB 68 VL--------NSTKGVKLDFKSLEAFERSKP-ILAEN----------RKKFTKPLFLNADILPGPVEAK--TI
Ccn_FAM151B 69 VFEFNQNQ--NVTKGIKLDFKSIDVFEQSLE-IINKL----------YDKMEYPVWINADILSGPIESK--VK
Dme_UnChar1 68 IINFNRDH-EDQKKGVKLDFKSIEVFEGSLD-ILDVNI----------PNPTTYPVWINADILSGPVEQN-RTV
Aag_FAM151A 67 VHKYNKASKADKQKGVKLDFKSIEAFEKSQD-QIAPFS----------KPEVTFPLWLNADILPGPIEAT--TK
Pma_FAM151A 67 VAQYNNGN--SKQKGVKLDFKSIEAFEKSQD-QIAQFS----------KPEITFPLWLNADILPGPVNAT--TT
Cte_HypPro1 66 VI----E----AKKGFKINFGCIEAVDISLQ-YLVEFQDQVSVASVLRVRFRFDFIQLAVPIHLSAAVLNGPNAEFSTSK
Lan_FAM151A 62 ILDANLN---GRKKGVKLNMKHDKVIGPTLK-VLQAMK----------DSIVIPVWIHADILLGPGTN---KT
Spi_FAM151A 64 TP----T---SPNKGIKLDFKTIQVVEPSLK-MMKNVTL----------GQQVTNPIWLNADILPGPCYDK-VCV
Spi_FAM151B 62 TI--------GSGKGIKLDFKVTAVVEPSIR-ILSDQTTA----------QHEATNPIWLNADILSGPCQTD-CND
Cte_HypProt 67 IL----N---REDKGIKLDFKGLQVVRPSLE-LLKEKE----------NLVNHPIWANADVVAGPGSS---AD
Ofu_UnChar1 67 IL----N---RNDKAPKLDFKNEEVVAPSLQ-MLKAKG----------KKVSQPVWVNADLVKGPGGD---AP
Gae_UnChar1 67 LK----S----GSKGIKIDFSTIDSVEITLQ-KLKNVNE----------KSTLKFPVWLHADVLRGPLGG---KP
Sph_FAM151B 67 VK----K----FPKGIKLDFRTIDAVEMSLQ-LLKGMK----------DELKIPIWLHADILPGPFSD---RL
Obi_FAM151A 67 VK----K----FTKGIKLDFRTIDAVEMSLQ-LLKRMK----------NELKIPIWLHADILKGPYSS---RL
Scl_FAM151A 68 VI----P----SKKGIKMDIKVEEVIPHALK-ELQLHR----------SKLMQPVWINADVIQGPNTL---ST
Cin_FAM151A 67 TI----A----SDKAIKLDIKITEVIPYALE-ILRLHG----------PTLHQPVWINADVVKGPNTN---SD
Aja_UnChar1 65 VT----R---YTNKGIKLDFKDVQAVQPALD-ILLLKE----------AQLRVPLFFHADIIAGPNGQN-SVD
Lva_FAM151A 67 TT----Q---HTDKGMKLDFKYLEALEPSMI-LIGDHE----------SELKAPLWINADIVIGPNSP---RD
Ofu_UnChar2 67 GK----T----STKGLKLDFKSIEAVELTLQ-NLKKVA----------PKLRMPIWLNADILKGPNSK---SN
Dre_FAM151A 67 VL----K----SKKGVKLDFKSISAVEPSLD-LLRAKNQ----------TGINRPVWINADILPGPNVPE-FWP
Ccp_FAM151A 67 VL----K----SKKGIKLDFKSINAVELSLD-LLRVKNQ----------TGINRPVWINADILPGPNVPV-FWP
Hsa_FAM151A 67 VL----G---SSQKGIKLDFKNIKAVGPSLD-LLRQLTE----------EGKVRRPIWINADILKGPNMLI--ST
Ptr_FAM151A 67 VL----G---SSQKGIKLDFKNIKAVGPSLD-LLRRLTE----------EGKVRRPIWINADILKGPNMLI--ST
Pan_FAM151A 67 VL----G---SSQKGIKLDFKNIKAVGPSLD-LLRRLTE----------EGKVRRPVWINADILKGPNMLI--ST
Vpa_FAM151A 67 VL----T---SSQKGIKLDFKSIKAVGPSLD-LLRRLTS----------EGRVRRPVWINADIQRGPNVPI--PI
Mmu_FAM151A 67 VL----A---SSQKGIKLDFKSLKAVGPSLD-LLRQLTE----------AGRIRRPVWINADILRGPNVPI--SI
Eca_FAM151A 67 VL----A---SSQKGIKLDFKSLKAVGPSLD-LLRRLTE----------DGRVRRPVWINADILRGPNVPI--SV
Bbu_FAM151A 67 VL----Q----SSRGIKLDFKSIQAVGPSLD-ILLATSS----------RTPINRPVWLNADILAGPNVNH--EI
Xla_FAM151A 67 VL----K----SPKGIKLDFKSIQAVGPSLD-ILFAKAS----------EVKINRPVWLNADILKGPNVNH--EI
Rty_FAM151A 67 VI----R----SKKGIKLDFKSIDAVNPSLD-ILVKKYN----------EIHFNRPVWLNADILIGPNVPG-FMQ
Rbi_FAM151A 67 VL----M----SIKGVKLDFKTIKAVGPSLD-ILVKKSS----------Q--ISRPVWINADILNGPNINI--NI
Asi_FAM151A 67 VL----N---SSRKGIKLDFKSIKAVGPAIE-ILLKKSQ----------EVDINRPVWLNADILEGPNVLV--NV
Cpi_FAM151A 67 VL----N---SSSKGIKLDFKSIKAVGPSLD-ILLKKSS----------EMKLNRPLWLNADILMGPNVPI--NT
Gae_FAM151A 68 VA----E----SNKGMKLDFKSLESVAPALK-ILKKKHD----------QGRLTQPVWLNADILTGPNTT---AK
Bfl_FAM151A 67 AL----E----SSKGLKLDFKSIGSVAPSLR-ILRNKS----------NLINRPVWLNADILRGPNTV---NP
Dre_FAM151B 63 VV----K----TDKGIKLDFKSLAAVSQSMS-LLEEIR----------DQLKGPVWINADILPGPGGT---AT
Ccp_FAM151B 63 VV----K----SDKGIKLDFKSLAAVSPSMT-LLEEVR----------DQLQGPVWINADILPGPRGT---AT
Bbu_FAM151B 63 VS----A----SSKGIKLDFKSLEAVLPSMK-ILDAMK----------DNLHQPVWINADILPGPGGS---V-
Xla_UnChar1 61 VS----S----CEKGIKLDFKCIEAVLPSLQ-ILAAMK----------ATVKQPVWINADILPGPGGK---AK
Rty_FAM151B 65 VQ----Q----SDKGIKLDFKSLSAVEPAMK-MLVSMK----------DLLTRPVWINADILPGPNES-----
Rbi_FAM151B 65 VM----K----TNRGIKLDFKSLEAVKPSMV-LLNGVK----------TRIKSPVWINADILPGPGGT---NM
Mmu_FAM151B 65 VV----K----SNKGIKLDFKSLAAVRASML-FLDNMK----------QHLQRPVWINADILPGPNGS---SK
Pan_FAM151B 65 VT----K----SNKGIKLDFKSLAAVEPSMM-LLENVK----------RHLKRPVWINADILPGPNGN---SK
Hsa_FAM151B 65 VM----K----SNKGIKLDFKSLAVVEPSMM-LLENVK----------RHLKRPVWINADILPGPNGN---SK
Ptr_FAM151B 65 VM----K----SNKGIKLDFKSLAAVEPSMM-LLENVK----------RHLKRPVWINADILPGPNGN---SK
Eca_FAM151B 65 VV----K----SNKGIKLDFKSLAAVEPSLM-LLENVK----------RHLKRPVWINADILPGPNGH---SR
Vpa_FAM151B 65 VI----K----SNKGIKLDFKSLAAVEPSMM-LLENVK----------RHLKRPVWINADILPGPNGN---SR
Gga_FAM151B 65 MA----G----TDKGIKLDFKSLDAVRPSLE-LLQLVK----------PQLERPVWLNADVLPGPNGI---NA
Asi_FAM151B 49 IL----N----ADKGIKLDFKSLPAVQPSME-LLESIK----------LHLKRPVWINADILPGPNGS---NA
Cpi_FAM151B 65 IV----N----TKKGIKLDFKSLAAVKPSMM-LLEGIK----------LHLRRPVWINADILPGPNGS---NT
Consensus 66 VL----K----SNKGIKLDFKSIEAVEPSLD-LLRAKK----------PKLKRPVWINADILPGPNGP---SK


Sra_DUF2181 107 PVDSYTLLENAHKYVPNAAISLGWTRQSDTSNSNNILQNTNYLDWGQTFKILSYL-----NSVNYQPIILTIKLSDALAS
Ace_HypProt 127 VVDPSSFVDRARRLFPDATLSLGWTKQSNFSMLN---PKYKRLTWRQLFQILEYI-----ARLD-QPVMLSVRLSVAANS
Cel_Menorin 73 EVDPSTFVEKAKDLFPDATLSLGWIKQSNFSHLH---PKFKKLSWRQLFHILEYI-----SRLD-QPVMLSVRLSVAAHS
Dpa_HypProt 127 EVDPATFVERTSNLFPDATLSLGWIKQSNFSHLH---PKFKRLTWRQLFEILEYI-----ARLE-QPVMLSIRLSVAAAS
Lva_UnChar1 123 PVDPNQFIGLVQSNFPSSTLLLGWATTWSP-D-----APQIRYSWYNVIEMAKTC-----AGAK-QPVSFAVRAIYARKS
Aja_UnChar2 122 PVDLQKAVDLVKEYYPPITLSLGWQTDWSP-E-----PVETGYSWYHVISMAKFC-----VKFQ-QRASFSIRAVYAAES
Cse_FAM151B 128 PVDAARFLKGSK-LIPNAMLSLGWTTGFNN-T------HNSGYTFPQIEEMLYTIKQNEV---N-QPITFAVRAGLVAES
VpeFAM151AB 120 PLDSKSFLTGAMEAFPESVLSIGWTTRYGS-EFN---ITEGHYTTEQIQKMIDTLTENKV---T-QSITYPVRAGLAAND
Ccn_FAM151B 127 PVDPDRFLSGAK-QFTRSILSIGWTTKYGP-D-----NSNGTYSESNIKEMINVINDNNV---T-QSITFPVRAGLAAQS
Dme_UnChar1 129 PVDADRFFAGCM-RYKQAVLSIGWTTNWGA-D-----FRDGEYTQQQCDDMLETLSENNVLSTG-QAITFPVRAGIAANS
Aag_FAM151A 128 PVDPIKFIKLCG-NHPRAVMSVGWTTEYGG-N-----VTEGEYSRDQIGSMLRLMTENRI---N-QTVTFPVRAGLASNS
Pma_FAM151A 126 PVDPLKFLNLGA-KHPRAVLSVGWTTNYGG-N-----ITEGEYSRNQIGTMLRLVNE-HV---N-QTVTFPVRAGLASNS
Cte_HypPro1 137 PLDASRFIHLCSSLMPRATLSLGWISKMVN-------GKTSRYSWTTVMQMFDLIHD---AQLS-LPISLNLESEFVSGS
Lan_FAM151A 118 MINPMQLFSQIDQIYPAVTLSVAWASD----------SSQTSYTQSMMEEMYSLI-----KNLK-QPVSITVRAALVKNA
Spi_FAM151A 120 PVDHERFLSLCKSYYPNATLSISWKTGENM-T-----ASKNYYNWSQVLPMGKLV-----SQIA-QPITFPFRANLVQRS
Spi_FAM151B 118 PVDPQVFLSLCAKYFPSATLSVGWTTGKYI-S-----PEKDGYDWHLVQPMKQLL-----SNLT-QPITFAIRANLIGNS
Cte_HypProt 119 PIDPKQFLSIVNEVWPNMTLSLGWTTGCCE----------PFTRQMMEDMLEVC-----HGVT-QPVTFPARAASFRAS
Ofu_UnChar1 119 NFNAARFFTAVNTHFPRVTLSLGWTTSHGY----------NYTKQHIESLWQLI-----QNVT-QPITFPARASSFKYA
Gae_UnChar1 120 TVDATRFFRTMKRLFPRCTLSLGWTTGIHT-D-----LSQSGYTWDMVMDMYDLIQKWDV--GD-QPIVFQARLSLIHNS
Sph_FAM151B 118 PVDSTRFLRNVRTFFPSCTLSLGWISGYHT-D-----VSQSGYTWNMVLDMHDLIRKWEI--T--QPLVFSVRLSLIANS
Obi_FAM151A 118 PVDATRFFRNVRTFFPTCTLSLGWTTGYHT-D-----VSQAGYTWDMVLDMHDLIIHWEI--S--QPLVFSVRLAYIKNS
Scl_FAM151A 119 PIDGDYFVHNVNQYFPNVTLSLGWTTGYRI-A-----LENEEYSWESMDDMLRLA-----SSTN-QRITFPIRAALARQS
Cin_FAM151A 118 PIDSNIFLPEVNSKFPNVTLSLGWTTGYRNVG-----PPNEKYSWDAMEKMLSLS-----RPLN-QLITYPARAALLRQS
Aja_UnChar1 119 LIQPTEFIAIVNTMFPNATFSLGFVSSASSGL-----NQQDRYTWTMIFDMLDTS----STVTT-HPITFNFRAVWANKS
Lva_FAM151A 119 PVPPQPFIDIANRYFDKTTLSLGFTTAWGP-M-----MADKLYTWTMIFDNLYYS----YPLDP-QPVTFPIRAVWCKTS
Ofu_UnChar2 118 PVDAERFLKLCSEMFPKSTISLGWTVSLEG--------KERQYSWAHVIEMYHTIDKWILKKRK-QPVTIAVAACCIKYS
Dre_FAM151A 121 VVNASEFFELIQLKFPDVTISPGWKVLYLS-I-----FPNVTYTRSMVEQMYSTI-----RHLP-QKITFPVHALMAKNG
Ccp_FAM151A 121 VINASEFFELIQLKFPDVTISPGWKVLYLS-I-----FPNVTYTRSMVEEMYTIV-----RHLP-QKITFPVHALMAKNG
Hsa_FAM151A 122 EVNATQFLALVQEKYPKATLSPGWTTFYMS-T-----SPNRTYTQAMVEKMHELV-----GGVP-QRVTFPVRSSMVRAA
Ptr_FAM151A 122 EVNATQFLALVQEKYPKATLSPGWTTFYVS-T-----SPNRTYTQAMVEKMHELV-----GEVP-QRVTFPVRSSMVRVA
Pan_FAM151A 122 EVNATQFLALVQEKYPKATLSPGWTTFYMS-T-----FPNRTYTRAMVEKMHELV-----GVVP-QRVTFPVRSSMVRAA
Vpa_FAM151A 122 EINATRFLALAQEKYPEATLSTGWTTLYLP-M-----FPNSTYTRAMVEKMQELV-----GALP-QKVTFPLYALMARSA
Mmu_FAM151A 122 EINATQFLTLVQEKYPKATISPGFTTLYVP-Q-----LPNSTYTQAMVETMQELV-----GALP-QKVTFPVRAVMTRAA
Eca_FAM151A 122 EVNATRFLALVQEKYPEATLSPGWTTLYEP-L-----LPSGTYTRAMVEEMQGLV-----GVLP-QRVTFPVRAVMARAA
Bbu_FAM151A 121 GVNATQFLNLIQERFPDITISPGWVTLYLPPI-----ISNRTYSSEMVKKMYNLV-----KGLT-QRITFPARAVLTCSA
Xla_FAM151A 121 GVDATQFLNLVKNKFPDVTLSPGWVTLYLPPI-----ISNRTYTREMIQQMYNMV-----RDLP-QKITYPARAVMTRSA
Rty_FAM151A 122 PVNASRFLGLIQQRFPNVILSPGWMSLYLP-M-----IATKPYTRKMVEEMYDLV-----KGLS-QRVTFPVRAVLLKPA
Rbi_FAM151A 119 AVNATQFLDLVQRKFPNVTISPGWVTLYLP-F-----LSNKTYTWPMIWKMYTLV-----RDLP-QRITFPVRAVMIKSA
Asi_FAM151A 122 SLNASTFLSLIQEKYPNCTLSPGWTTLYSP-L-----FPKQTYTRAMIQKMHDLI-----GELP-QKVTFPVRAIMVRLA
Cpi_FAM151A 122 AVNASLFLSLIQEKYPNCTLSLGWTTLYSF-L-----FPNKTYTQKMIQKMHSIV-----GTLP-QRVTFPVRAVMVRLA
Gae_FAM151A 121 GRDAPQFLKIVDETFPQCTLSIGWTTGWTN-T-----EADTGYSMEMVREMNELA-----RPLR-QPVSFPIRAAAAKRS
Bfl_FAM151A 118 GVNAREFIDTVNRIFPECTLSIGWTTGFYY-D-----RENEGYTRQMVEEMHSYC-----GDLT-QPVTFPIRNSILSLP
Dre_FAM151B 114 PVDPHVFLQEVAQRSENDVLSLGWTTGWTA-N-----VDNPGYSWEMVHQMEELC-----RPLK-QPVTFPVRASILPMS
Ccp_FAM151B 114 PLDPHVFLQEVAQKSENDVLSLGWTTGWDV-D-----ADNPGYSWEMVHQMEELC-----RSLK-QPVTFPVRAALLPAS
Bbu_FAM151B 113 TVDAREFLQIVTSFFPNVTLSLGWITAWHP-D-----KSNEGYSWEMVREMEKIC-----KNLS-QLVTFPVRAALVRQS
Xla_UnChar1 112 AVDAKEFIHTVMLYFPDVTLSLGWTTGWNP-G-----QDNEGYSWEMVQEMEKIC-----KGLS-QPVTFPVRAALLRQS
Rty_FAM151B 114 ----KDFLCIVTSYLPHVTLSLGWKTDELV-N-----QTNIVYTWEMVKEMEQIC-----QTLS-QPVTFPVRAALVRPS
Rbi_FAM151B 116 PLDAKRFLEVVTSFFSDVTLSLGWTTGWHP-Q-----KSNEGYSWEMVKEMEHIC-----SALS-QPVTFPVRAALVRQS
Mmu_FAM151B 116 VVDAKAFLDTVTSFFPDVTFSLGWTTGWHP-E-----KVNEGYSWSMVKEMDYIC-----SELT-QPVTFPVRAALVRQS
Pan_FAM151B 116 VIDAKPFLDTVTFFPDVTFSLGWTTGWHP-E-----KVNEGYSWTMVKEMEYIC-----NELS-QPVTFPVRAALVRQS
Hsa_FAM151B 116 VIDAKPFLDTVISFPDVTFSLGWTTGWHP-E-----KVNEGYSWTMVKEMEYIC-----NELS-QPVTFPVRAALVRQS
Ptr_FAM151B 116 VIDAKPFLDTVTSFPDVTFSLGWTTGWHP-E-----KVNEGYSWTMVKEMEYIC-----NELS-QPVTFPVRAALLRQS
Eca_FAM151B 116 VVDAKPFIDTVTSFFPDVTLSLGWTTGWHP-E-----KVNEGYSWTMVKEMEYIC-----NELN-QPVTFPVRAALVRQS
Vpa_FAM151B 116 VVDAKPFIDTVTSFFPDVTFSLGWTTGWHP-E-----KVNEG----------
Gga_FAM151B 116 VVDAEGFLEIVTSFFPDVTLSLGWTTGWHP-E-----QHNEGYSWTMVKEMAQIC-----STLS-QPVTFPVRAALVRQS
Asi_FAM151B 100 VVDAKFFLDIVTSFFPDITLSLGWTTGCQL-Q-----RCKEGYSWAMVKEMAEIC-----NALT-QPITFPVRAALVWQS
Cpi_FAM151B 116 VVDAKRFLDTVTSFCPDVTLSLGWTTGWQP-Q-----QCNKGYSWAMVKEMAQIC-----DGLT-QPVTFPVRAALVQQS
Consensus 117 PVDAKRFLDLVQSKFPDVTLSLGWTTGYHP-D-----KPNEGYTWEMVEEMLELC-------LP-QPVTFPVRAALVRQS


Sra_DUF2181 182 ---SDQILFLLGQ-N--RPFYVIIYSKP-EDPIN----NINVFQNFLSFARHNGNVIFDLPPEYR----
Ace_HypProt 198 ---KDQLLWLLGM-D--KAISLLIWSDK-DDEEI----DWASVAEIRRV-ATKNRVLYDLEPRHR----
Cel_Menorin 144 ---KEQLLWLLGM-D--QSISLLLWSDA-EDHVT----NWTPIVELRRS-TTKNRILYDLDPKHR----
Dpa_HypProt 198 ---KEQLLWLLGM-D--QSVSLLIWSDE-EDKID----DWNSIVQLRRS-TTKNRILYDLIPKHR----
Lva_UnChar1 191 ---IRQLRFLLSL-T--SRFSLTVWTDV-YDIMP-----LPDLVHFRKNMD-PKRVHYNLPENFM----
Aja_UnChar2 190 ---VRQIKWLLSL-S--SRFSAVLWLAE-HDVVS-----VSAMVYFRKQTD-KKKIFYILP--------
Cse_FAM151B 196 ---QDQMTLLLRN----NNHTLTIWSGVQDDGFD-----VGKLNEILDKVE-KGRIYVDVPQKLR----
VpeFAM151AB 192 ---ISAMKTLMDRSSSFGNVTMTIWSSH-GDQVD-----TKKLSELIKTIG-VGKVYVDVPEDVW----
Ccn_FAM151B 196 ---FKEILKLIEK---INGSTLTIWSSE-NDPVN-----VPDLRKLIFAVG-LDRTYVDVPEDLR----
Dme_UnChar1 201 ---EEQLHRLVAAVNETNESTLTIWSSA-GDYVD-----VDKLRKLIFSFG-LDRVYLDVPEELA----
Aag_FAM151A 197 ---QPVILDLLRETS-FLNSSMTVWSSE-SDHVE-----VDRLRALILTVG-LERTYLDVPHELA----
Pma_FAM151A 194 ---QPVILDLLRETA-SLNSSMTVWSSE-GDAVE-----VDRLRALILTVG-LERTYLDVPDDL-----
Cte_HypPro1 206 ---IQQLKWLMQM----TNTTLHIYLNS-KNLTP-----AVDILAMRKQFP-KNRLYYIGLSDVLSKVR
Lan_FAM151A 182 ---WPNLKWLLSQ-N--SNFTLTVWNPS-GVT-DKEGTDLYDLLYVRNNWY-IEKIFYDLPGTKF----
Spi_FAM151A 188 ---WDQLQWLLDL-S--ETFTVTIWSST-TDKVD-----PLDLVALRNNVS-RERIYYDLPPDQE----
Spi_FAM151B 186 ---IEQLIWLLGL-S--EEYTLTIWSAD-SDAPN-----MKDLVALHNRVSDKKRIFYDLP--------
Cte_HypProt 182 ---WEHFKWLLEQNP--ESYTITVWTPS-GA-EDWEGTDIYDLLYVRNDCD-KTLIYYDLPGE------
Ofu_UnChar1 182 ---WTHFKWLLGK-S--NGYTITAWTPA-SAEVSWEGADPYDLLYVRNDCD-KKKIYYDLKGPKY----
Gae_UnChar1 191 ---VPQLKWLTDV-I--SHSALMIWHAD-GDMPN-----YENLMYVSYRFP-PNELFFDLDHNDL----
Sph_FAM151B 188 ---IPQLKWLCDT-T--Q-ASLLVWNDN-QDKPV-----LPDIIFIPPAAS-VQLPSFLKRTCRCTCP-
Obi_FAM151A 188 ---IPQLKWLCDT-L--Q-ASLLVWNDN-EDKPV-----LPDVMYVAYRFP-PEKAFFDFQD-EYL---
Scl_FAM151A 187 ---WYKFLWLLEQ-D--KRFSLTVWSAS-VDPVS-----LPDKVYIRDNYD-TSRVFYDTDPNFV----
Cin_FAM151A 187 ---WDRFLWLLEQ-S--NSYTLTIWSST-TDVVS-----VPDMVFVRDNFD-ISRIFYDAEDALT----
Aja_UnChar1 189 ---WQKFVWLLGL-T--NRFSVTIWSKA-TDPLN----TFTGLEELRNNGD-KRRIFYDLAPQHL----
Lva_FAM151A 188 ---WPKFVWLLGL-R--DSFSITVWSSG-SDIVD-----VGGLVDLRTHGD-TRRIFYDLPDLQK----
Ofu_UnChar2 189 ---IPQLKWLTDM----TGASLTLWSHE-SHLVQ-----LADTLYTQYKFP-KNKLYFDLAQNHL----
Dre_FAM151A 189 ---WPHLSWLLSQ-S--SRYSLTLWQGK-ENP-T-----LNDLLFIRDNSN-PQRIYYDIYEPVL----
Ccp_FAM151A 189 ---WPHLSWLLSQ-S--PRFSLTLWQGK-ENP-T-----VNDLLFIRDNSN-PLRIYYDIYEPVL----
Hsa_FAM151A 190 ---WPHFSWLLSQ-S--ERYSLTLWQAA-SDPMS-----VPDLLYVRDNTA-VHQVYYDIFEPLL----
Ptr_FAM151A 190 ---WPHFSWLLSQ-S--ERYSLTLWQAA-SDPMS-----VPDLLYVRDNTA-VHQVYYDIFEPLL----
Pan_FAM151A 190 ---WPHFSWLLSQ-S--ERYSLTLWQAA-SDPMS-----VPDLLYVRDNTA-VHQVYYDIFEPLL----
Vpa_FAM151A 190 ---WPHFSWLLGQ-S--ERYSLTLWQAT-SDSVS-----VDDLLYIRDNTA-PHQVYYDIFEPVL----
Mmu_FAM151A 190 ---WPHFSWLLSQ-S--ERYSLTLWQGA-SDPVS-----VPDLLFIRDNSA-AHQIYYDLFEPVL----
Eca_FAM151A 190 ---WPHFSWLLGQ-S--ERYSLTLWQGA-SDPVS-----VDDLLYIRDNSA-THQVYYDLFEPVL----
Bbu_FAM151A 190 ---WQNFYWLLKQ-S--DRYSLTLWQGS-SDPLQ-----LDDLLFIRDNSR-PEEIYYDIYDPLL----
Xla_FAM151A 190 ---WPHFNWLLQQ-S--ERYTITLWQGK-SDPLT-----LPDLLFIRDSSN-PEEIYYDIFEPLL----
Rty_FAM151A 190 ---WPHFSWLLSQ-S--PRYSLTLWQGS-IDPVT-----VPDLLFFRDNSN-VEQIYYDIYEPVL----
Rbi_FAM151A 187 ---WQYFSWLLQQ-S--DRYSLTLWQGE-TDPIT-----VPDLLYVRDNSR-AEEIYYDIYDPVL----
Asi_FAM151A 190 ---WPHFSWLLNQ-S--ERYSLTLWQGK-TDPVT-----VPDLLFIRDNSR-AEQIYYDIYDPVL----
Cpi_FAM151A 190 ---WPHFSWLLAQ-S--DRYSLTLWQGK-MDPIR-----VPDLLFIRDNSR-PEQIYYDIYDPVL----
Gae_FAM151A 189 ---WKELDWLLKQ-S--RGYTLTIWHAS-SDIVS-----LDEMKFIRAHSE-SSRIYFDLPAALMPL--
Bfl_FAM151A 186 GTLENYLKWLLEQ-S--DTYTLTLSEGQ-TSIDP------VDLLKIRNDFD-WSRVFYDITTEKA----
Dre_FAM151B 182 ---FPQFQWLLEQ-S--DRSEILH----------
Ccp_FAM151B 182 ---FPQFQWLLKQ-S--DRYSLTVWTGK-HDVLT-----VPDLLLYRQNFS-KTRIYYDLLESQI----
Bbu_FAM151B 181 ---WPQLQWLLQT-S--DRYSLTVWSGK-DDIYP-----VPDLLYIRQHSG-ADQIFYDVFEPQN----
Xla_UnChar1 180 ---WPQFQWLLKT-A--DRFSLTVWAGK-DDTYP-----VPDLLFIRDNSE-KCRIYYDLFEPQN----
Rty_FAM151B 178 ---WPQLSWLLEQ-S--ERYSLTVWTAR-EDLYT-----VPDMLYIRENVD-KSRIYFDLFEPQN----
Rbi_FAM151B 184 ---WSELLWLLQS-S--VRYSLTIWTGK-DDVYS-----LPDLLFIREKFD-KSRVYFDLLEPRS----
Mmu_FAM151B 184 ---CPQLLWLLTK-S--NRYSLTVWTGK-DDIYS-----TPDLLYIRDYFN-KTQVFYDISEPQN----
Pan_FAM151B 184 ---CSQLLWLLKQ-S--SRYSLTIWTGK-NDNYS-----IPDLLYIRDHFD-KKQVFYDILEPQN----
Hsa_FAM151B 184 ---CSQLLWLLKK-S--NRYSLTIWTGK-NDNYS-----VPDLLYIRDHFD-KKQVFYDILEPQN----
Ptr_FAM151B 184 ---CSQLLWLLKK-S--NRYSLTIWTGK-NDNYS-----VPDLLYIRDHFD-KKQVFYDILEPQN----
Eca_FAM151B 184 ---CSQLLWLLKK-S--HRYSLTIWTGK-NDNYS-----TPDLLFIRDHFD-KKQVFYDILEPQN----
Vpa_FAM151B 152 ----------
Gga_FAM151B 184 ---VSELCWLIDQ-S--DRYSLTVWTGK-EDAYS-----VPDLLFIRENFD-KSRVYYDISEPQN----
Asi_FAM151B 168 ---KSELLWLLQQ-S--ERYSLTVWTGK-QDQYS-----TPDLLHIRENFD-KS----------
Cpi_FAM151B 184 ---TSELLWLLQK-S--NRYSLTVWTGR-HDEYS-----IPDLLYLRENLD-KK----------
Consensus 183 ---WPQLLWLLGQ-S--ERYSLTIWSGK-SDPVS-----VEDLLYIRDNFD-KERIYYDLPEPLL----

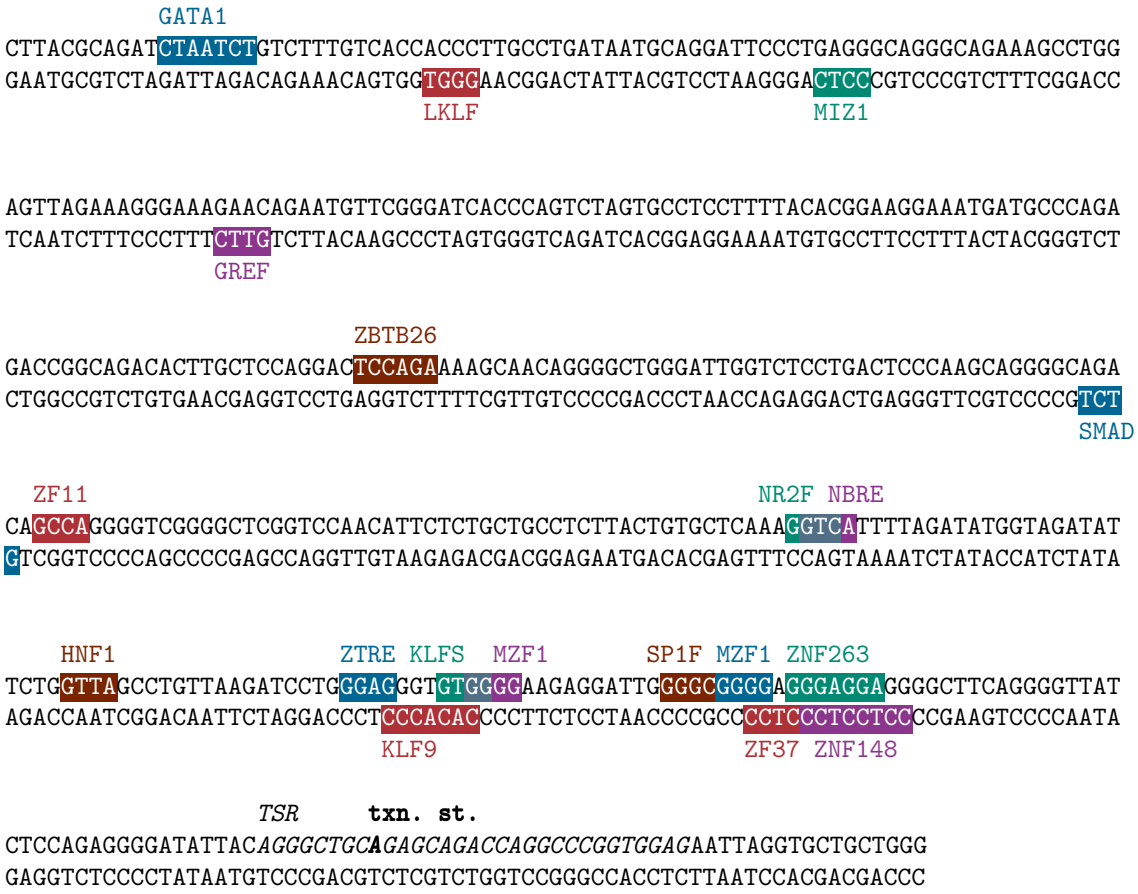
Alignment of DUF2181 with Other Members of GDPD/PLCD Superfamily

```
F151A2      DGLNVEWLVPDVQGSGKTA----------TMTLP--DTEGMILLNT--GLEGTVAENPV       45
F151A1      DALEVTWYHAANSKK--------------AMTAALNSNITVLEADVNVEGLGTANETGV       45
F151B1      DGAEITWYHAANHKA--------------QTNEALKSTAHMIEADVLLPSDG--SEHSQ       43
MNR1CE      DGLNVRVAHGVNSWPDLVD----------QLHEPFLNKSMMIEGDVFMQAHRRPRHRAV       49
GDPDEC      SNEKIVIAHRGASGYL--------PEHTLPAKAMAYAQGADYLEQDLVM------TKDDN       46
GPCP1H      -----DVGHRGAGNSTTTAQLAKVQENTIASLRNAASHGAAFVEFDVHL------SKDFV       49
GDPDOI      --MTKIIAHRGASKYA--------PENTRASFELAHQMNADGIETDVQL------SKDGI       44

F151A2      PIVHTPSG--NILTLESCL-----------------QQLAT--HPGHWGIHLQIAEPAAL       84
F151A1      PIMAHPPTIYSDNTLEQWL-----------------DAVLG---SSQKGIKLDFKNIKAV       85
F151B1      PIMAHPPETNSDNTLQEWL-----------------TEVMK----SNKGIKLDFKSLAVV       82
MNR1CE      PVMKADTKLADRITFKEWL-----------------REVAT----MNKAIKINFRSNEVV       88
GDPDEC      LVVLHDHYLDRVTDVADRFPDRARKDGRYYAIDFTLDEIKSLKFTEGFDIENGKKVQT--       104
GPCP1H      PVVYHDLTCCLTMKKK--FDADPVELFEIPVKELTFDQLQLLKLTHVTALKSKDRKESVV       107
GDPDOI      PILFHDEKIKRIMRRK--GFLQ----------DYTYNELKSMDIGSWFG--SNFIGETI-       89

...

F151A2      ----------------------SPRATVTVEHNPAGGDYASV-RTAL-----LAARAVD       225
F151A1      ----------------------SERYSLTLWQ--AASDPMSV-EDLL-----YVRDNTA       228
F151B1      ----------------------SNRYSLTIWT--GKNDNYSV-EDLL-----YIRDHFD       222
MNR1CE      ----------------------DQSISLLLWS--DAEDHVTNWTPIV-----ELRRSTT       237
GDPDEC      HMLIEETSQPGNIKLTGMVQDAQQNKLVVHPYT--VRSDKLPEYTPDVNQLYDALYNKAG       312
GPCP1H      V------HTEDLLRNPSYIQEAKAKGLVIFCWG--DDTNDP---------ENRRKLKELG       283
GDPDOI      I------HIKHRLLQPKLVQQAKTENMPLRVYT--VN--KP---------KQLELCFKYN       222

F151A2      RTRVYYRLPQGYH-------        238
F151A1      VHQVYYDIFEPLL-------        241
F151B1      KKQVFYDILEPQN-------        235
MNR1CE      KNRILYDLDPKHR-------        250
GDPDEC      VNGLFTDFPDKAVKFLN---        329
GPCP1H      VNGL----------IYDR--        291
GDPDOI      CDSVFTDVPDIAKTAYQTYL        242
```

Appendix D: FAM151A Promoter Analysis

FAM151A Promoter Diagram with Highlighted Predicted TF Binding Sites

Only the core of each transcription factor is highlighted for clarity. A full guide to each transcription factor is given in Table A3.

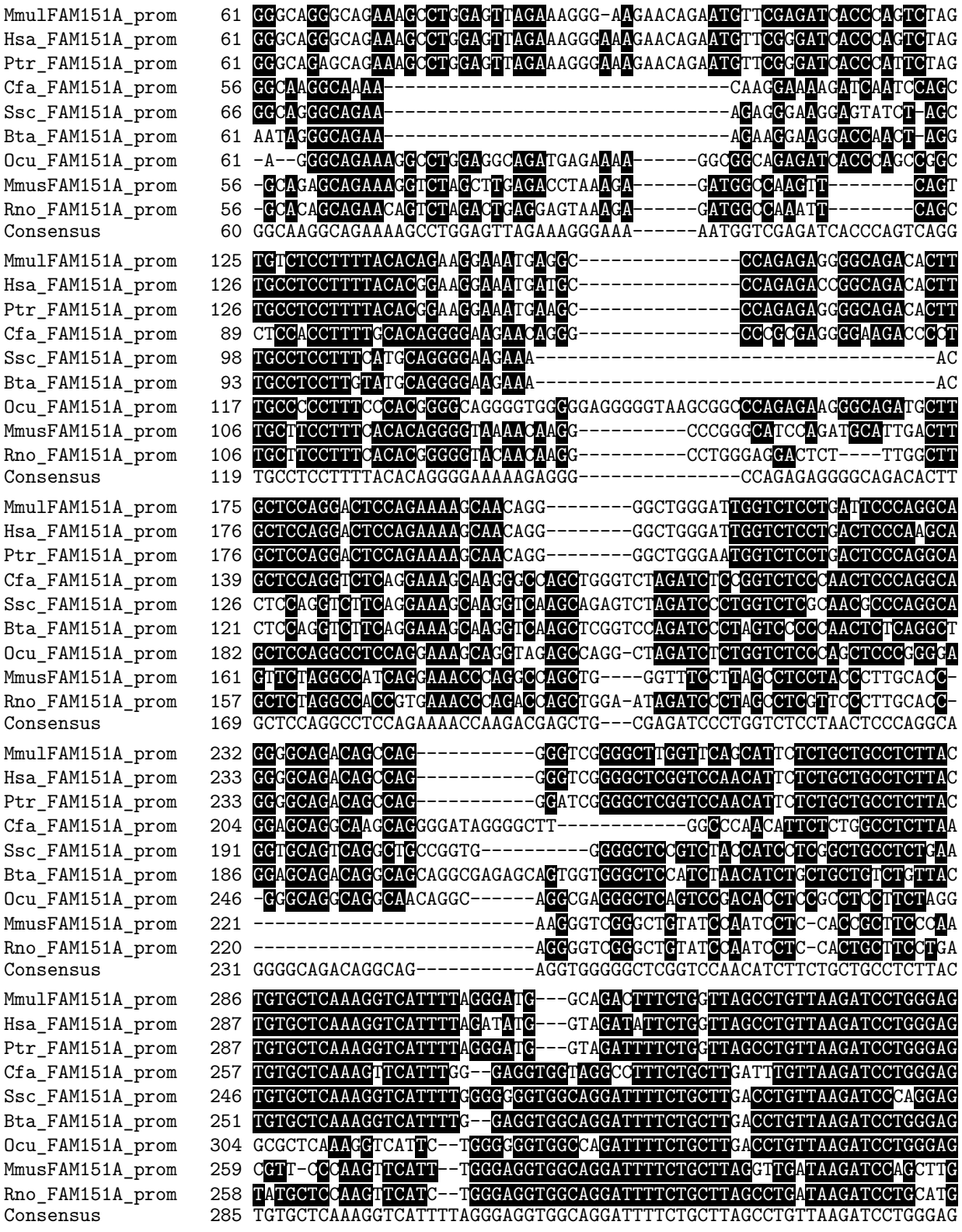


Label	Description/Full Name
GATA1	GATA binding protein 1
LKLF	Krueppel like transcription factors (Krueppel-like factor 2 (lung))
MIZ1	Myc-interacting Zn finger protein 1 (Myc-interacting Zn finger protein 1, zinc finger and BTB domain containing 17 (ZBTB17))
GREF	Glucocorticoid responsive and related elements (Androgen receptor binding site, IR3 sites)
ZBTB26	Zinc finger and BTB domain containing 26
SMAD	Vertebrate SMAD family of transcription factors (Sma- and Mad-related proteins)
ZF11	C2H2 zinc finger transcription factors 11 (Zinc finger and BTB domain containing 3)
NR2F	Nuclear receptor subfamily 2 factors (Chicken ovalbumin upstream promoter transcription factor 2, NR2F2 homodimer, DR1 sites)
NBRE	NGFI-B response elements, nur subfamily of nuclear receptors (Nuclear hormone receptor NUR77 (NR4A1))
HNF1	Hepatic Nuclear Factor 1 (Homeobox containing 1)
ZTRE	Zinc transcriptional regulatory element (3' half site of ZTRE motif)
KLF9	Kruppel like factor 9
KLFS	Krueppel like transcription factors (Gut-enriched Krueppel-like factor / KLF4)
MZF1	Myeloid zinc finger 1 factors (Myeloid zinc finger protein MZF1)
SP1F	GC-Box factors SP1/GC (Sp2 transcription factor)
ZF37	C2H2 zinc finger transcription factors 37 (Zinc finger protein 37 alpha (KOX21))
ZNF263	Zinc Finger Protein 263
ZNF148	Zinc Finger Protein 148

Table A3: Description of transcription factors predicted to bind to FAM151A promoter.

FAM151A Promoter Multiple Sequence Alignment

```
MmulFAM151A_prom      1 -----CTTATGCAGATCTAATCTGTCTTTGTCACCACCCTTGCCTGATAATACAGGATTCCCTGA
Hsa_FAM151A_prom      1 -----CTTACGCAGATCTAATCTGTCTTTGTCACCACCCTTGCCTGATAATGCAGGATTCCCTGA
Ptr_FAM151A_prom      1 -----CTTATGCAGATCTAATCTGTCTTTGTCACCACCCTTGCCTGATAATGCAGGATTCCCTGA
Cfa_FAM151A_prom      1 ----------GCAGATATAATCTGTCCTTGTCATCAACATTTGCCTGACCTTGGTGACTTCCTGG
Ssc_FAM151A_prom      1 AAGCAGATTCGCATTCTAATCCATCCACAGTCACCCACCCTTGTGTGACCTTCGTGACTCCCTGG
Bta_FAM151A_prom      1 -----CTTGGGCACATCTAATCTGTTCCTGTCACTCACCCTTGCCTGAACTTCATGACTCTCTGG
Ocu_FAM151A_prom      1 -----CTCAGGCAGATCTAATCTGTCCTTGTCACTAACCCTTGCCTGCCCGGGATTACAGGGGC
MmusFAM151A_prom      1 ----------GCAGATCTAATCTACTCTTGTCACCAACCCTTTCATGACGTAGGCTTCCCTGGA
Rno_FAM151A_prom      1 ----------GCAGATCTAATCTGCTCTTGTCACCAACCCTTTCATGACCTAGGGTTCCCTGGA
Consensus             1 -----CTTA-GCAGATCTAATCTGTCCTTGTCACCAACCCTTGCTTGACCTACGGGATTCCCTGA
```


MmulFAM151A_prom 61
Hsa_FAM151A_prom 61
Ptr_FAM151A_prom 61
Cfa_FAM151A_prom 56
Ssc_FAM151A_prom 66
Bta_FAM151A_prom 61
Ocu_FAM151A_prom 61
MmusFAM151A_prom 56
Rno_FAM151A_prom 56
Consensus 60
MmulFAM151A_prom 125
Hsa_FAM151A_prom 126
Ptr_FAM151A_prom 126
Cfa_FAM151A_prom 89
Ssc_FAM151A_prom 98
Bta_FAM151A_prom 93
Ocu_FAM151A_prom 117
MmusFAM151A_prom 106
Rno_FAM151A_prom 106
Consensus 119
MmulFAM151A_prom 175
Hsa_FAM151A_prom 176
Ptr_FAM151A_prom 176
Cfa_FAM151A_prom 139
Ssc_FAM151A_prom 126
Bta_FAM151A_prom 121
Ocu_FAM151A_prom 182
MmusFAM151A_prom 161
Rno_FAM151A_prom 157
Consensus 169
MmulFAM151A_prom 232
Hsa_FAM151A_prom 233
Ptr_FAM151A_prom 233
Cfa_FAM151A_prom 204
Ssc_FAM151A_prom 191
Bta_FAM151A_prom 186
Ocu_FAM151A_prom 246
MmusFAM151A_prom 221
Rno_FAM151A_prom 220
Consensus 231
MmulFAM151A_prom 286
Hsa_FAM151A_prom 287
Ptr_FAM151A_prom 287
Cfa_FAM151A_prom 257
Ssc_FAM151A_prom 246
Bta_FAM151A_prom 251
Ocu_FAM151A_prom 304
MmusFAM151A_prom 259
Rno_FAM151A_prom 258
Consensus 285

```
MmulFAM151A_prom   348 GGTGTGGGAAAGAGGAGTGGGGTGGGGAGGCAGGAGGGGCTTCAGGGGTTATCTCCAGTCAGAGA
Hsa_FAM151A_prom   349 GGTGTGGGGAAGAGGATTGGGGCGGGGAGGCAGGAGGGGCTTCAGGGGTTATCTCCAGAG-GGGA
Ptr_FAM151A_prom   349 GGTGTGGGGAAGAGGATTGGGGCGGGGAGGCAGGAGGGGCTTCAGGGGTTATCTCCAGAG-GGGA
Cfa_FAM151A_prom   320 GGTGTGGATAGAAGAGGAGGGTG-GGGAAGGGGGAGGAGCCTCAGAAGCTATCTCTGCTGTTGAC
Ssc_FAM151A_prom   311 AGAGAGGGAAGAGGA-GGGCGG---GGAGGTGGGAGGGGCCTCGGAGCTGTCTTAT----GGGC
Bta_FAM151A_prom   314 AGAGAGAGGAGGAGGAGCATAG---GAAGGTAGGCGGGGCCTACAAAGCTGTCTTTG----GGAC
Ocu_FAM151A_prom   367 GCGGGGGAGGGCAGGGGGGAGGAGGAGCTTCAGGAGGAGGGGATTTAGGAGAAGGGGCATTCGGG
MmusFAM151A_prom   321 GATGG-GAAGAGGAGGGCTGGGAGCAGGGCAGGGTTTCTGAGCTCTCTCCACCCAGGCAGTGGGA
Rno_FAM151A_prom   321 GGTGAGGGAAATGAGAGCTAGGAGGAGGGCAGGGTTTCTGAGCTCTCTCCACCCAGGCAGCGGGA
Consensus          350 GGTGTGGGAAAGAGGAGTGGGGAGGGGAGGGAGGAGGGGCTTCAGGAGCTATCTCCGCAG-GGGA

MmulFAM151A_prom   413 TATTACAGGGCTGCAGAGTGGACCAGACCTGGTGGAGAATTAGGTGCTGCTGGG---------
Hsa_FAM151A_prom   413 TATTACAGGGCTGCAGAGCAGACCAGGCCCGGTGGAGAATTAGGTGCTGCTGGG---------
Ptr_FAM151A_prom   413 TATTACAGGGCTGCAGAACAGACCAGGCCTGGTGGAGAATTAGGTGCTGCTGGG---------
Cfa_FAM151A_prom   384 GCTCCTAGCCTGCAGGAGCCAGAGTGCCTTGGTGGAGAACTGGATGCTACCAGGGGCTCCTG-
Ssc_FAM151A_prom   368 CATCCAGGCCCACAGGAGCAGGCCTGACTCCAA-GAGCATTGGGTGCTGCTGG----------
Bta_FAM151A_prom   372 ATTCCAGGCCTGCAGGAGCTGTCCTGCCAGGGA-AAGAATTAGGTGCTGCT------------
Ocu_FAM151A_prom   432 -TTGCAGGCGCTGCATCCCCCACCC----CATCCCAGCTGCAGCCAGCCACAGGGGCTCTGGG
MmusFAM151A_prom   385 GTGGCCAGACCTGAGTT-----------AAATGTTGCTGCA--------------------
Rno_FAM151A_prom   386 GTGGCCAGACCTGAGTT-----------AAATGCTGCTGCA--------------------
Consensus          414 TATTCCAGCCCTGAGGAGCAGACCAG-CCTGGTGGAGAATTAGGTGCTGCTGGG---------
```

Appendix E: FAM151A 3' UTR Multiple Sequence Alignment