Nucleic Acids Research, 2016 1 doi: 10.1093/nar/gkw222

FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases

Krzysztof Kuchta^{1,2,†}, Anna Muszewska^{1,3,†}, Lukasz Knizewski¹, Kamil Steczkiewicz¹, Lucjan S. Wyrwicz⁴, Krzysztof Pawlowski⁵, Leszek Rychlewski⁶ and Krzysztof Ginalski^{1,*}

¹Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw, Zwirki i Wigury 93, 02–089 Warsaw, Poland, ²College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Banacha 2C, 02–097 Warsaw, Poland, ³Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02–106 Warsaw, Poland, ⁴Laboratory of Bioinformatics and Biostatistics, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, WK Roentgena 5, 02–781 Warsaw, Poland, ⁵Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, Nowoursynowska 166, 02–787 Warsaw, Poland and ⁶BioInfoBank Institute, Limanowskiego 24A, 60–744 Poznan, Poland

Received December 16, 2015; Accepted March 22, 2016

ABSTRACT

FAM46 proteins, encoded in all known animal genomes, belong to the nucleotidyltransferase (NTase) fold superfamily. All four human FAM46 paralogs (FAM46A, FAM46B, FAM46C, FAM46D) are thought to be involved in several diseases, with FAM46C reported as a causal driver of multiple myeloma; however, their exact functions remain unknown. By using a combination of various bioinformatics analyses (e.g. domain architecture, cellular localization) and exhaustive literature and database searches (e.g. expression profiles, protein interactors), we classified FAM46 proteins as active noncanonical poly(A) polymerases, which modify cytosolic and/or nuclear RNA 3' ends. These proteins may thus regulate gene expression and probably play a critical role during cell differentiation. A detailed analysis of sequence and structure diversity of known NTases possessing PAP/OAS1 SBD domain, combined with state-of-the-art comparative modelling, allowed us to identify potential active site residues responsible for catalysis and substrate binding. We also explored the role of single point mutations found in human cancers and propose that FAM46 genes may be involved in the development of other major malignancies including lung, colorectal, hepatocellular, head and neck, urothelial, endometrial and renal papillary carcinomas and melanoma. Identification of these novel enzymes taking part in

RNA metabolism in eukaryotes may guide their further functional studies.

INTRODUCTION

Proteins adopting the nucleotidyltransferase (NTase) fold play crucial roles in various biological processes, such as RNA stabilization and degradation (e.g. RNA polyadenylation), RNA editing, DNA repair, intracellular signal transduction, somatic recombination in B cells, regulation of protein activity, antibiotic resistance and chromatin remodelling (1). Almost all known members of this large and highly diverse superfamily transfer nucleoside monophosphate (NMP) from nucleoside triphosphate (NTP) to an acceptor hydroxyl group belonging to protein, nucleic acid or small molecule. They are characterized by the presence of a common α/β -fold structure composed of a three-stranded, mixed β -sheet flanked by four α -helices. This common core corresponding to the minimal NTase fold is usually decorated by various additional structural elements and additional domains, depending on the family. Sequence analysis of distinct members of this superfamily revealed the following common sequence patterns in NTase fold domain: hG[GS], [DE]h[DE]h and h[DE]h (where h indicates a hydrophobic amino acid) that include conserved active site residues. Three conserved aspartates/glutamates are involved in the coordination of divalent ions and activation of the acceptor hydroxyl group of the substrate. Two of them (from the [DE]h[DE]h motif) are located on the second core β-strand, while the third carboxylate (from the h[DE]h motif) is placed on the structurally adjacent third β -strand. The hG[GS] pattern is placed at the beginning of a short, sec-

^{*}To whom correspondence should be addressed. Tel: +48 22 554 0800; Fax: +48 22 554 0801; Email: kginal@cent.uw.edu.pl [†]These authors contributed equally to the paper as first authors.

 $[\]ensuremath{\mathbb{C}}$ The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

ond core α -helix and has a crucial role in harbouring the substrate within the active site (1).

Human members of the NTase fold superfamily are encoded by 43 genes (1). Until now, only one group of potentially active human NTases, belonging to the so-called family-with-sequence-similarity-46 (FAM46), has not been characterized for its exact biological function. Little is known about FAM46 proteins apart from their involvement in various diseases. FAM46A, a gene preferentially expressed in the retina, was reported as a positional candidate for human retinal diseases since it maps within the RP25 locus to chromosome 6p12.1-q14.1 interval where several retinal dystrophy loci are located (2). It was also suggested that a variable number of tandem repeat polymorphisms in FAM46A may be associated with non-small cell lung cancer (3). Another FAM46 paralog, FAM46B, was identified to have lower expression in metastatic melanoma cells (United States Patent US 7615349 B2). FAM46B and FAM46C have been recently described as potential markers for refractory lupus nephritis (4) and multiple myeloma (5-10), respectively. Finally, it was reported that FAM46D is overexpressed in lung and glioblastoma tumors (11), as well as together with FAM46C, in the brain of autistic-like behaving transgenic mice (8).

Functional proteomics studies showed that FAM46A might have many potential protein interacting partners (12). One of them, ZFYVE9, detected in the yeast two-hybrid system, is involved in the recruitment of unphosphory-lated SMAD2/SMAD3 to the transforming growth factorbeta receptor (13). Another member of FAM46 family, FAM46C, is recognized as a type I interferon stimulated gene, which enhances the replication of certain viruses (14). In addition, FAM46C can be an anti-viral factor in acute infected long-tailed pygmy rice rats by Andes virus (15). It was also suggested that FAM46C is functionally related in some way to the regulation of translation (6).

To date, several studies have indicated that proteins belonging to FAM46 family might play an essential role in the cell; however, their exact functions remain unknown. In our previous work, we performed a comprehensive classification of the NTase fold proteins and assigned FAM46 proteins to this superfamily as potentially active members (1). This classification allowed us only to speculate that like other active NTases, FAM46 members may catalyze template-independent incorporation of NMP from NTP either to nucleic acid, protein or small molecule. In this study we present an in-depth bioinformatics analysis of the FAM46 family combined with an exhaustive literature and database searches and propose that the FAM46 proteins function as non-canonical poly(A) polymerases. Detailed insight into the sequence and structure diversity of NTases and their additional N- and C-terminal domains allowed us to generate a reliable 3D model for one of the family members (FAM46C) and to confidently identify the potential active site residues responsible not only for catalysis but also for substrate binding. In addition, the obtained structural model for human FAM46C sheds some new light on the molecular role of mutations found in cancer patients in the FAM46 genes. Finally, the broad sampling of sequenced genomes made it possible to track the evolutionary history of FAM46 proteins back to the origin and hypothesize that the FAM46 family members are present not only in animals but also in all four sequenced Dictyosteliidae and two Entamoeba (Amebozoa) genomes.

MATERIALS AND METHODS

Sequence searches

Four human FAM46 paralogs (FAM46A, FAM46B, FAM46C, FAM46D) were used as queries in PSI-BLAST (16) searches performed against the NCBI non-redundant (NR) protein sequence database with E-value threshold of 0.001 until profile convergence. Collected sequences were split into organism-specific sets and clustered with CD-HIT (17) in order to obtain unique sequences. All FAM46 family members were aligned using Mafft (18) with some manual adjustments. The alignment used for phylogeny reconstruction was additionally trimmed by TrimA1 (19) to eliminate poorly aligned and thus uninformative regions.

Additionally, proteins containing both NTase and Cterminal four-helical up-and-down bundle fold domains were collected with PSI-BLAST searches performed against the NCBI NR database with E-value threshold of 0.001. Sequences (PDB SEQRES) of the following structures possessing this domain context: pdbl2pbe, pdbl3c18, pdbl3jyy, pdbl3jz0, pdbl4ebs, pdbl1v4a, pdbl3k7d and pdbl1kan, were used as queries.

Analysis of gene and protein features

The architecture of the human FAM46 genes was analysed using the UCSC genome browser (20). Protein localization was predicted with BaCelLo (21), CELLO (22), WoLF PSORT (23), Euk-mPLoc 2.0 (24) and MultiLoc (25). NetNES (26) was used to detect the nuclear export signal (NES). Protein phosphorylation motifs were detected with Eukaryotic Linear Motif (ELM) (27). Gene expression patterns were analysed using the BioGPS database (28). Genes with average z-scores higher than 5 (at least in one probe set) in 'Barcode on normal tissues' dataset were considered as expressed in specific tissue/cell. 'Barcode on normal tissues' dataset provides a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray (28). The z-scores in this dataset are generated with the barcode function from the R package 'frma', which bases on barcode algorithm (29). The domain architecture was analysed using Meta-BASIC (30) and SMART (31).

Structural analysis of known NTases

Initially, known NTase fold families and structures were identified from literature (including our previous classification (1)) and various databases of catalogued protein families (PFAM (32), COG (33) and KOG (34)) and structures (PDB (35) and SCOP (36)). This initial set was then used for a comprehensive, transitive searches for all NTase fold superfamily members using our distant homology detection method Meta-BASIC (30) and Gene Relational DataBase (GRDB) system, as described in our previous work (1). Briefly, Meta-BASIC is a highly sensitive meta-profile alignment method capable of finding very distant similarity between proteins through a comparison of sequence profiles

enriched by predicted secondary structures (meta-profiles). The GRDB system includes precalculated Meta-BASIC connections between 16 230 PFAM, 4825 KOG and 4873 COG families and 38 498 representative proteins of known structure (PDB90). Each family and each structure is represented by its sequence (PDB90) or consensus sequence (PFAM, COG and KOG), sequence profile (generated with PSI-BLAST using the NCBI NR70 derivative) and secondary structure profile (predicted with PSIPRED (37)).

The structural diversity was analysed for all collected NTase superfamily structures clustered at 90% sequence identity. Structures were divided into groups based on their structural similarity using DALI (38). Structure-based alignments were generated for all considered domains (including both the conserved NTase fold and additional N-and C-terminal domains) after manually curated superposition of their structures. Secondary structures were assigned with DSSP (39).

3D model building

Potential templates were identified with Meta-BASIC and the consensus of fold recognition 3D-Jury approach (40) using human FAM46 proteins as queries. The sequenceto-structure alignment between FAM46 family members and all representative structures possessing both NTase and PAP/OAS1 SBD domains was built using a consensus alignment approach and 3D assessment (41) based on Meta-BASIC and 3D-Jury results, PSIPRED secondary structure predictions and conservation of critical active site residues and hydrophobic patterns. The 3D model of human FAM46C protein was built with MODELLER (42) using Trypanosoma brucei TUTase 4 (pdbl2ikf) (43) as a template. Finally, the side chain rotamers were optimized using SCWRL3 (44). The overall quality of the modelled structure was checked with ProSA (45). Structure visualization was carried out with Pymol (http://www.pymol.org).

Analysis of protein interactors

Human members of the NTase fold superfamily were identified from the UniProt database (46) using our transitive Meta-BASIC search strategy as described above, starting with all collected NTase fold families and structures. Proteins interacting with human NTase superfamily members were identified using the BioGRID database (version 3.4.133) (47). GO annotations (molecular function and biological process) (48) for detected interactors were taken from the UniProt database. FAM46 interacting partners were also identified with manual literature searches.

Analysis of single point mutations in cancers

Missense mutations, found in cancer patients, in FAM46 genes were collected from publications and the following databases: cBioPortal (49), ICGC (50) and IntOGen (51). The sequence conservation in FAM46 family was measured based on Jensen–Shannon divergence (JSD) (52) using created FAM46 multiple sequence alignment. The JSD quantifies the similarity between probability distributions with scores ranging from 0 to 1 (53). A background amino acid

distribution, estimated from a large sequence set, is used to approximate the distribution of amino acid sites subject to no evolutionary pressure. Positions in an alignment that are found to have amino acid distributions very different from the background distribution are proposed to be functionally important or constrained by evolution. The JSD score was computed using the score_conservation.py program (52) with default parameters, e.g. using BLOSUM62 for the background distribution. Positions in FAM46 multiple sequence alignment with more than 30% gaps were omitted from JSD computations.

Phylogeny

In order to visualize the relationships between FAM46 family members, a phylogenetic analysis was performed with PhyML3.0 (54) using the LG and JTT models, with an estimated gamma parameter and proportion of invariable sites. An approximate branch support was calculated using the aLTR (55) option implemented in PhyML. Branches with supports lower than 0.5 were collapsed. The trees were drawn using iTol (56).

RESULTS AND DISCUSSION

FAM46 family

Firstly, we identified proteins belonging to FAM46 family with an exhaustive PSI-BLAST (16) searches performed against the NCBI NR protein sequence database using all four human FAM46 paralogs (FAM46A, FAM46B, FAM46C and FAM46D) as queries. These searches quickly converged at the third iteration; however, most family members can be easily detected even with a simple BLAST search. This is a feature of compact and very conserved protein families and graphical clustering of all these sequences corroborated this observation. As many of them turned out to be variants or mutants of the same protein (e.g. there are four FAM46 genes in the human genome and 14 proteins in the NR database) we selected 868 protein sequences unique for each organism using sequence clustering at different thresholds followed by manual assessment. It should be noted that some of the detected proteins contain long deletions within conserved regions, what might be due to erroneous gene/exon prediction.

Taxonomic distribution

FAM46 proteins are present in the proteomes of all animals. Supplementary Figure S1 summarizes the taxonomic distribution of all selected 868 FAM46 proteins unique for each organism. Four FAM46 paralogs can be identified in almost all sequenced Vertebrata (with high-quality genomes), but not in Tunicata (*Ciona intestinalis* and *Oikopleura dioica*), Hemichordata (*Saccoglossus kowaleskii*), Echinodermata (*Strongylocentrotus purpuratus*) or Cephalochordata (*Branchiostoma floridae*), which encode only a single FAM46 protein. Specifically, amphibian, bird, reptile and mammal genomes harbour four distinct FAM46 genes. On the other hand, fish proteomes contain six to seven FAM46 paralogs due to lineage specific duplications followed by fast differentiation of the retained paralogs (different in each of the four analysed fishes). This evolutionary scenario has been already described in teleost fishes (57). An asymmetric acceleration of evolutionary rate in one of the paralogs after the duplication event, manifested by the high protein sequence divergence and usually leading to alignment problems in less conserved regions, was also observed in FAM46 paralogs. FAM46 family members are encoded in all sequenced animal phyla ranging from Arthropoda (Daphnia, Drosophila), Mollusca (Crassostrea gigas), Nematoda (Caenorhabditis elegans, Brugia malavi, Loa loa, Trichinella spiralis), Platyhelminthes (Schistosoma mansoni, Clonorchis sinensis), Cnidaria (Nematostella vectensis), Placozoa (Trichoplax adherens) and Porifera (Amphimedon queenslandica). FAM46 genes duplicated and diverged strongly in some Nematoda lineages leading to a variable number of paralogs in the analysed genomes. Moreover, FAM46 proteins are detectable in close metazoan relatives: Choanoflagellida (Salpingoeca sp., Monosiga brevicollis) and Ichthyosporea (Sphaeroforma arctica). Surprisingly, proteins belonging to this family can be also found in Amebozoa (four Entamoeba species, *Polysphondylium pallidum*, Acvtostelium subglobosum and three Dictyostelium species and Acanthamoeba castellanii) and one Diplomonadida (Guillardia theta). Choanoflagellida and Ichthyosporea, together with Metazoa, are a sister group of Fungi and Nucleariids. Noteworthy, FAM46 family members are absent in fungal and nucleariid genomes sequenced within the Origins of Multicellularity Project by BROAD. Amebozoa are sometimes grouped with Opisthokonta (Metazoa and Fungi) into a supertaxon Unikonta characterized by a single posterior flagellum in flagellated cells. Summarizing, the presence of FAM46 proteins in proteomes of Metazoa, Choanoflagellida and Amebozoa suggests its origin in the ancestor of Unikonta with further divergence into four distinct conserved representatives in vertebrates.

Phylogeny inference

Phylogenetic relationships were analysed both for a set of 29 representative sequences and for a set of 868 Metazoa, Choanoflagellida, Diplomonadida and Amebozoa FAM46 proteins. Entamoeba, Giardia and Dictostellids form well-separated clades with uncertain branching order (Figure 1). They are, however, clearly separated from the Metazoa-Choanoflagellida clade. *Salpingoeca rosetta* and *M. brevicollis* form a sister clade to Metazoa. Some of invertebrate FAM46 proteins display higher variability at sequence level that can lead to long branches on the phylogenetic tree. The position of basal lineages within the Metazoa is uncertain and possible involvement of long branch attraction phenomenon should be taken into account.

The evolutionary history of FAM46 in the vertebrate genomes is a story of consecutive duplications leading to four highly similar paralogs. All vertebrate genomes analysed in this study retained all four FAM46 paralogs. Surprisingly, we detected the presence of FAM46 proteins in all sequenced Amebozoa genomes.

The divergence time between Choanoflagellida and Metazoa is estimated \sim 600MYA (58,59). As FAM46 genes are present in Choanoflagellida, Metazoa and Amebozoa genomes it is possible they were already in the ancestor of

all Unikonta, and therefore also in the ancestor of Ophistokonta. The most likely scenario involves an ancient deletion in the ancestor of Fungi. This evolutionary history claims FAM46 would be a very ancient gene, what is not reflected in its encoded amino acid sequence divergence. Provided FAM46 have an ancient origin early in the Unikonts, still the presence of two FAM46 genes in the G. theta genome requires clarification. Due to cohabitation, it is plausible that the FAM46 genes in G. theta genome appeared via horizontal gene transfer from Choanoflagellida to Diplomonadida. However, the low resolution of these deep branches renders the FAM46 phylogenetic tree (Supplementary Figure S2) uninformative for HGT inference. There is significant evidence for the transfer occurring in the opposite direction (60). The mechanism underlying algae to choanoglafellate transfer is supposed to be based on phagotrophy. We have insufficient data to hypothesize about the possibility of transfer happening from choanoflagellates to algae.

Gene structure

The organization, architecture and regulation of human FAM46 genes and their homologous counterparts in others organisms seem to contribute to their functional diversification. For instance, human FAM46 genes contain 2–3 exons of which 1–2 are coding (Supplementary Table S1) and they encode up to five different transcripts. In addition, antisense transcripts have been also detected, e.g. for human FAM46A (61). Interestingly, we found that the H3K27Ac pattern in the promoter region and along the coding region is completely different for each of human FAM46 genes, what can be related to distinct nucleosome density in these chromatin regions and different expression patterns (62). Additionally, the FAM46D gene is in a repeat dense area as denoted by RepeatMasker, which might be related to the overall chromosome X repeat density.

Domain architecture

Given the high variability of human FAM46 paralogs at the gene organization level, the encoded proteins are surprisingly similar at the sequence level, including the conservation of various motifs. FAM46 proteins seem to have a common two-domain architecture composed of an α/β region (according to secondary structure predictions) followed by an α -helical region. It should be noted that the FAM46 family is a distant outlier in the NTase fold superfamily and cannot be identified with standard sequence comparison methods such as PSI-BLAST. Using a highly sensitive tool for distant homology detection, Meta-BASIC (30), we mapped FAM46 proteins, with the above threshold scores, to several 3D structures including the terminal uridylyl transferase 4 (TUTase 4) from T. brucei (pdbl2ikf) (43), 2'-5'-oligoadenylate synthase (OAS) from S. scrofa (pdbl1px5) (63), CCA-adding enzyme from A. fulgidus (pdbl4x4n) (64), cyclic AMP-GMP synthase from V. cholerae (pdbl4u0n) (65), aminoglycoside 6adenyltransferase from B. subtilis (pdbl2pbe) and nuclear factors NF90 and NF45 from M. musculus (pdbl4at7) (66). Importantly, FAM46 N-terminal α/β region has weak but



Figure 1. Phylogenetic tree of representative FAM46 protein sequences. Maximum likelihood (ML) analysis for selected 29 family members was carried out using the LG+G model. The approximate likelihood ratio test Shimodaira–Hasegawa-like (SH-like) branch supports above 0.5 are shown. Branches with support lower than 0.5 were collapsed.

evident sequence similarity to the NTase domain, which can be confirmed by several fold recognition servers. Meta-BASIC suggested that the C-terminal α -helical part may be similar either to poly(A) polymerase/2'-5'-oligoadenylate synthetase 1 substrate binding domain (PAP/OAS1 SBD) or the domain of four-helical up-and-down bundle fold (4H), however, it assigned below threshold scores to these predictions. To figure out what protein fold is adopted by the FAM46 C-terminal region, we performed a comprehensive analysis of the structural diversity of all the available NTase superfamily structures, both for their conserved NTase and additional N- and C-terminal domains (Supplementary Figure S3).

While both the PAP/OAS1 SBD and 4H domains possess four core α -helices C-terminal to NTase domain, only PAP/OAS1 SBD retains the additional (the first core) α helix located before NTase domain. According to secondary structure predictions, this helix is clearly seen in the FAM46 family members in the conserved region preceding the predicted NTase domain. In addition to a good mapping of predicted and observed core secondary structure elements, FAM46 proteins display also similar conservation of hydrophobic motifs and critical residues for NTP binding (see below) characteristic for the PAP/OAS1 SBD. In our previous studies we showed that such detailed analysis of below threshold Meta-BASIC hits usually enables identification of highly diverged superfamily members which escape detection even with advanced sequence comparison methods. For instance, using this approach we identified restriction endonuclease-like (67) and RNase H-like (68) domains in many uncharacterized and poorly annotated protein families. Finally, we found that proteins embracing both the NTase and 4H domains are mainly encoded in bacteria and rarely found in archeal genomes, with single representatives identified in eukaryotic species, including fungal Myceliophthora thermophila (gil367020986), Tribulus terrestris (gil367039397) and Rhizophagus irregularis (gil552919075), and soil-living amoeba Dictyostelium discoideum (gil66821023). This is consistent with the biological functions played by these proteins, as they participate in antibiotic resistance (e.g. Staphylococcus aureus kanamycin nucleotidyltransferase (69), Enterococcus faecium lincosamide antibiotic adenylyltransferase (70), Bacillus subtilis aminoglycoside 6-adenyltransferase) and nitrogen assimilation (e.g. Escherichia coli glutamine synthetase adenyltransferase (71)). In contrast, NTases possessing the PAP/OAS1 SBD can be widely found in eukaryotes. Altogether, results of all these analyses strongly suggest that although displaying little sequence similarity, FAM46 proteins possess PAP/OAS1 SBD consisting of the five righthanded twisted α -helices (with an α 1-NTase- α 2 α 3 α 4 α 5 topology).

In addition, we found that a few FAM46 proteins possess additional domains inserted inside the NTase domain or located at N- or C-termini (Supplementary Figure S4). It should be noted, however, that the presence of some of these additional domains may be a result of potentially incorrect gene/exon predictions.

PAP/OAS1 SBD in known NTase structures

To identify conserved PAP/OAS1 SBD residues, critical for binding NTP substrate in an NTase active site, we carried out an exhaustive sequence and structure analysis by generating the structural alignment of all the representative structures possessing both the NTase and PAP/OAS1 SBD domains (Figure 2). The PAP/OAS1 SBD specifically binds a nucleobase of the incoming NTP mainly by amino acids that provide, either directly or indirectly via water molecules, Watson-Crick hydrogen bonds. In addition, a conserved hydrophobic amino acid (e.g. V234 in poly(A) polymerase Pap1 (72) and Y212 in poly(U) polymerase Cid1 (73)), located at the beginning of the third core α -helix of PAP/OAS1 SBD, forms a flat hydrophobic surface for the incoming NTP nucleobase. Proteins containing the PAP/OAS1 SBD also possess another common residue, which is responsible for the recognition of a triphosphate moiety. Conserved lysine/arginine (e.g. K215 in Pap1) located in the second core α -helix of PAP/OAS1 SBD, together with a serine from the NTase domain hG[GS] motif, interact with NTP β - and γ -phosphate groups.

3D model of human FAM46C

Initially, we generated a sequence-to-structure alignment of FAM46 family members with all representative proteins of known structure possessing both the NTase and PAP/OAS1 SBD domains (using their structure-based alignment described above) (Figure 2). Although these structures display very little sequence similarity to the FAM46 proteins, in contrast to our previous work (1) where we focused in general on the most conserved regions of NTase fold common to all NTase superfamily members, here we were able to propose a reliable and complete sequence-to-structure alignment for all conserved regions of both domains. The alignment was guided by secondary structure predictions and conservation of (i) the NTase fold active site motifs, (ii) identified critical PAP/OAS1 SBD residues participating in substrate binding and (iii) hydrophobic patterns responsible for forming the hydrophobic core of the structure.

As a representative of FAM46 family for 3D modelling we selected human FAM46C, which is a potential biomarker for multiple myeloma. FAM46 proteins are very similar in sequence, for instance, four human paralogs share 56–75% amino acid identity within the common region encompassing both domains. In addition, the length of this region in these paralogs differs only by 1–2 residues.

The structure of *T. brucei* TUTase 4 (pdbl2ikf) (43), as assigned the highest Meta-BASIC score among the proteins possessing both NTase and PAP/OAS1 SBD domains, was used as a template to generate the 3D model of human FAM46C, based on the manually derived sequence-tostructure alignment. However, due to the lack of templates with similar insertion between the last core β -strand and the last core α -helix of NTase domain, we were unable to create a reliable model for 70 amino acids of FAM46C in

this region. Nevertheless, we can speculate that this insertion in FAM46C should fill the space usually occupied by residues responsible for binding incoming NTP nucleobase and RNA 5' end. Figure 3 presents a comparison of the FAM46C model and existing structures of non-canonical poly(A) polymerase Trf4p from Saccharomyces cerevisiae, which is a part of the Trf4p/Air2p/Mtr4p polyadenylation (TRAMP) complex (74), and the non-catalytic mitochondrial dynamic protein MiD51 from M. musculus (75). Importantly, in Trf4p, the region of 53 amino acids between the fourth and fifth core α -helices of the PAP/OAS1 SBD is crucial for binding the RNA 5' end and a nucleobase of the incoming NTP. The corresponding region in FAM46C is much shorter (only 11 amino acids) and probably is not able to form the interaction interface for a nucleobase. Therefore, it is likely that nucleobase binding residues are located within the 70 amino acids insertion between the last core β -strand and the last core α -helix of the FAM46C NTase domain. In addition, this conserved insertion, composed of the predicted two β -strands and two α -helices (with $\beta \alpha \beta \alpha$ order), may also participate in protein-protein interactions similar to the MiD51 receptor which binds the dynaminrelated protein 1 (Drp1) via a well-conserved loop located in the NTase domain (75). However, it should be noted that FAM46C, in contrast to MiD51, seems to be an active NTase; therefore, even if the insertion is responsible for protein-protein interactions, it should also play a role in substrate recognition.

Active site

Figure 4 shows a comparison of active sites of human FAM46C (model), poly(A) polymerase Pap1 from S. cerevisiae (72), poly(U) polymerase Cid1 from Schizosaccharomyces pombe (73), CCA-adding enzyme from A. fulgidus (64), 2'-5'-oligoadenylate synthetase OAS1 from S. scrofa (76) and the cyclic GMP-AMP synthase (cGAS) from M. musculus (77). FAM46 proteins probably function as active NTases because they share all the key motifs in the NTase domain responsible for catalysis and substrate binding, including the [DE]h[DE]h and h[DE]h patterns with three conserved aspartate/glutamate residues (Asp90, Asp92 and Glu166) and hG[GS] motif with Gly73 and Ser74 in human FAM46C. Although we were not able to generate a complete 3D model of human FAM46C, we were able to identify additional residues responsible for NTP binding, which are located in the conserved secondary structure elements. Comparison of active sites of experimentally solved structures showed that proteins encompassing both NTase and PAP/OAS1 SBD usually bind a nucleobase or a ribose-moiety of incoming NTP by a serine or a threonine located just before or in the last core α -helix of NTase domain (Figures 2 and 4). Therefore, it is possible that FAM46C Ser248 may bind, directly or indirectly via a water molecule, 2'-OH or/and 3'-OH hydroxyl group of a ribose-base moiety as it is observed for Thr172 in poly(U) polymerase Cid1 (73) or it can participate in a nucleobase binding similarly to Thr190 in 2'-5'-oligoadenylate synthase OAS1 (76). FAM46C shares also all the conserved residues in PAP/OAS1 SBD responsible for substrate binding. The FAM46C Leu282 probably interacts with a nucleobase of

		<pap oas<="" th=""><th>l (</th><th></th><th>MIASE</th><th></th><th></th><th></th></pap>	l (MIASE			
		ннннннн	нннннн нннннннн	аннннннннннннннн	EEEE HHHHH	EEEEEE	нннннннннннн	EEEEEE
	☐ 1q78 A Bt PAP	(34) DCLLTQKLV	-ET <mark>LK</mark> PF(6)EELQRRIL <mark>I</mark> I	LGK <mark>L</mark> NNL <mark>V</mark> KEW <mark>IR</mark> EISESK (1	L2)K <mark>IF</mark> T <mark>FGS</mark> YRLG(5) <mark>AD</mark> IDALCV (7) SDFFT <mark>SF</mark> YDK <mark>L</mark> KLQ(3) KD <mark>L</mark> RAV (5)
	201p A Sc PAP	(21) ENKLNDSLI	-OE <mark>LKKE(6)OETANRVO</mark> VI	lki <mark>l</mark> oel <mark>a</mark> orf <mark>v</mark> yevskkk (1	L2)K <mark>IFTYGS</mark> YRLG(5) <mark>SDIDTLV</mark> V(7) EDFFTVFDSLLRER(4) = -IAPV(5)
	3nyb A Sc PAP	(17) SDWLTFETK	DF (8) EETETBNOT	ISTTREAVKOLW(3) DLHVEGSVSTD (5) SDTDCVVT (9) NNLYS-LASHLKKK (4	$E = \frac{V}{V} = \frac{V}{V} = \frac{V}{V}$
	2pg1 A He DAD	(122)	TIMEE (4) ENTRIDUIT	COLTEDMANYE(2) TUDDECOCUNT (5) CDIDMET D (22) MILEO BASHDURUUU	VUCUORT (5)
	Spqi_A HS FAF	(132) 5100000	-ILLKEF (4) ENIKLKILIC	SDIEDMAAAIF(S)IVRPEGSSVNI (SS/RILSV-LGECLDHF(4	JVGVQKI(J)
	41t6_A HS PAP	(39) DHI <mark>Y</mark> TQKLI	-DAMKEE (6) EETNHKTAAT	"GKTUNTAKEMIZDASESK (1	LZ) KIETEGSYRLG (5) ADIDALCV (7) SDEFQSEFERLKHQ (3) RN <mark>L</mark> RAV (5)
	2b4v_A Tb TUTase	(17)-AV <mark>W</mark> GKA <mark>I</mark> M	-AE <mark>N</mark> NRR(4)HMFRTAIR <mark>A</mark> Ç)QQ <mark>L</mark> QGL <mark>A</mark> DKW <mark>T</mark> (3) K <mark>VY</mark> C <mark>CGS</mark> MVTY (6) <mark>SD</mark> LDLACM(12)KRTDK- <mark>L</mark> RTV <mark>I</mark> KRY(7)N- <mark>L</mark> LGL(5)
	2ikf A Tb TUTase	(25)-AVVGRSLV	-NS <mark>F</mark> KQ-(8)RHVDA <mark>T</mark> Y	YRL <mark>V</mark> LDC <mark>V</mark> AAVD(3) RLYT <mark>FGST</mark> VVY (6) <mark>SD</mark> VDFVVL(23)DILAK- <mark>L</mark> ARV <mark>I</mark> RQK(4)N- <mark>V</mark> EEV(5)
	3hil A Th TUTase	(6) KREFIRGMM	-AHYRAS (4) EHSVVIHELC	OKR <mark>VI.DIG</mark> MI.AV(3) HVELEGSHVSG (5) SDADISLT (19) KRMTR-FGKEASAM (2)EDVRYT(4)
	ACTY A Sp TUTACO	(AO) HEEFTECY	FUNET (A) VEEVEVDANT		3) FLVAFCSLESC (5) CDMDLCVL	9) TTAIO EVERTAR(2	VE-CKET(5)
Ð	4e/x_A Sp IOIase	(40) IKEFIKFCI	EVINEI (4) KEFKEKKAAI	JDILKLCHKKIS(STELVATOSILESG(0) IIADQ-FIEEDIAE(2) E-GREL(J)
ŝ	30uy_A Af CCA	(1)KVEEILE	-KA <mark>L</mark> ELV (8) KGREAEEE <mark>L</mark> F	RR <mark>L</mark> DEL <mark>G</mark> (1) E <mark>YV</mark> F <mark>VGSY</mark> ARN (6) <mark>LDIDVFL</mark> L (11)ERGLE- <mark>I</mark> GKA <mark>V</mark> L(1	.)S- <mark>Y</mark> EIR(4)
Ā	4x4n_A Af CCA	(1)KVEEILE	-KA <mark>LE</mark> LV (8) KGREAEEE <mark>L</mark> F	RR <mark>L</mark> DEL <mark>G</mark> (1)EYVFVGSYARN(6) <mark>LEID</mark> VFLL(11)ERGLE- <mark>I</mark> GKA <mark>V</mark> L(1)S- <mark>Y</mark> EI-(5)
	1px5 A Sc OAS	(9)DLD	-KF <mark>IE</mark> DH(5)CFRTQVKE <mark>A</mark> I	IDI <mark>V</mark> CRF <mark>L</mark> KER <mark>C</mark> (1	L1) K <mark>VV</mark> K <mark>GGS</mark> SGKG (6) <mark>SD</mark> ADLVVF (15)EFIQE- <mark>I</mark> RRO <mark>L</mark> EAC(10)E- <mark>V</mark> Q(8)
	4id8 A HS OAS	(10)SLD	-KETEDY (5) CERMOINHAI	IDITCGELKER(2) K <mark>VV</mark> K <mark>GGS</mark> GKG (6) SDADLVVF (15) EFTOE-TRBOLEAC (8	K - E = -(10)
			THE FERRY (A) HETCEDDEW	UNEWFUL DDLOCC (7) ALL DECOVED (A7) WEDVI TVEETV (7	
	4JIX_A SS CGAS	(6)KTQ	-TVLERV(4) HEISEAAEVV	VNWVVEHLERRLQGG(/) ALLRIGSTIER (6) NEEDVEER (47) KFRKI-IKEEIK(7) T = VERK(4)
	4k8v_C Mm cGAS	(2)DK <mark>L</mark> K	-KV <mark>L</mark> DKL(4)KDISEAAE <mark>T</mark> V	/NK <mark>V</mark> VER <mark>L</mark> LRR <mark>M</mark> QKR (7) E <mark>QL</mark> NT <mark>GS</mark> YYEH (6) <mark>N¤FDVMF</mark> K (47)KFRKI- <mark>I</mark> KEE <mark>V</mark> KE-(6)S- <mark>V</mark> EKE(4)
	4km5 A Hs cGAS	(9)SK <mark>L</mark> R	-AV <mark>LE</mark> KL(4)DDISTAAG <mark>M</mark> V	√KG <mark>V</mark> VDH <mark>L</mark> LLR <mark>LK</mark> CD(6) GLLNTGSYYEH (6) <mark>NEFD</mark> VMFK (47)KFRKI- <mark>I</mark> KEE <mark>I</mark> NDI(5)I-MKRK(4)
	4lev A Mm cGAS	(7)KLK	-KV <mark>L</mark> DKL(4)KDISEAAE <mark>T</mark> V	VNK <mark>VVERL</mark> LRRMOKR (7) EOLNTGSYYEH (6) NEEDVMEK (47)KFRKI- <mark>I</mark> KEE <mark>V</mark> KEI(5	$S = \frac{V}{V} EKE(4)$
	ACCAS A HE CCAS	(16) 9KTP	AVLEKT (7) STAACMU	VKCWUDHILL RIKCD (6) CLINTCSYVEH (6) NEEDVMEK (46) SKEPKTIKEETNDI (5	T = MKR = (5)
	4000 A HS COAS	(10)SRLK	AVDERL (7) STAAGHV	TRUE DIBLERENCO (40) SKERKLIKEEINDI (S	() I = MKK = (J)
	4u0n_A VC CGAS	(6) FHQYYTNRNL)GL <mark>M</mark> GK-(5)EEKNNLKA <mark>L</mark> F	KIIIRLRTRDVFEEAKGIA (S	53) REWTQGSEQID (TT) MDTDDGLA(I4)SLLIL <mark>LV</mark> DAS <mark>L</mark> KSL(/)K-FEAK(I)
	4u03_A Vc cGAS	(6) FHQYYTNRND)GL <mark>M</mark> GK-(5)EEKNNLKA <mark>L</mark> F	RKI <mark>I</mark> RLR <mark>T</mark> RDV <mark>FE</mark> EAKGIA (53)R <mark>FW</mark> T <mark>QGS</mark> FQYD(11) <mark>MDID</mark> D <mark>GT</mark> Y(14)SLLIL <mark>LV</mark> DAS <mark>L</mark> KSL(7)K- <mark>F</mark> EAK(1)
	4at7 A Mm NF45	(23) ETS <mark>F</mark> SEALL	-KR <mark>N</mark> QDL(4)AEQASILS <mark>L</mark> V	VTK <mark>I</mark> NNV <mark>I</mark> DNL <mark>I</mark> VAP(8) EVROVGSYKKG (6) NVADLVVI (7) EAVAA-LGNKVVES (10)T-MLTN(1)
	4at7 B Mm NF90	(8) FVNDDRHVM	AKHSSV(4) EELEAVONMU	USHTERALKAVSDWIDEOF (37) GVMBVGLVAKG (6) <mark>LDLELVL</mark> L (7) ALLOK-VADNIATO (9	D = TLOS(3)
6	Am5d A Sc Utp22	(99) SNIFKIOTD	FITEOU(A) RUVI KVERET		13) DISTICSPALK (21) VVI AV_I TUUI I TI (10	
Ę	Amou A Se Ocp22	(99) SNI <mark>I</mark> RLQID		SIRCET DEL (3) DISLIGSFALK (
1	4m5d_A_SC_Utp22	(988)EIT	-KKEHNE (17) SSEENLKKSE	TODTIKTIBOW(7) SILPVGSAF-R (IP) ELODAIPE (18) TAFLLK <mark>I</mark> QEE <mark>L</mark> SAN (4) R- <mark>S</mark> FFS (II)
	4nxt_A Hs MiD51	(15) RMS <mark>L</mark> QEK <mark>L</mark> L	-TY <mark>YR</mark> NR (5) GEQARAKQ <mark>A</mark> A	AVD <mark>I</mark> CAE <mark>L</mark> RSF <mark>LR</mark> AKLP(5) D <mark>MY</mark> L <mark>SGS</mark> LYDD (6)DH <mark>I</mark> QLIVP(56)DTFEK <mark>VV</mark> AGS <mark>I</mark> N(12	:)R- <mark>P</mark> AP-(4)
	4woy A Mm MiD49	TFOERLL	-aferkh(5)ahvtlako <mark>l</mark> a	AGD <mark>I</mark> ALE <mark>L</mark> OAY <mark>L</mark> RSKFP(5) ALVPGGPLYDG (6) EH <mark>V</mark> RLLAP (56)ELLRK <mark>AL</mark> SAS <mark>V</mark> (13)W- <mark>P</mark> DV-(2)
				- · ·				
		ннннннн	нинини ессенинин	иннининини	EEEEE EEEE	EEEEE	ннннннннннн	EFFE
	Q96IP4 Hs FAM46A	(67)-WE <mark>Q</mark> VQR <mark>L</mark> D	-GI <mark>L</mark> SET (10) TLELQPSL <mark>I</mark> V	/KV <mark>V</mark> RRR <mark>L</mark> AEK(5) D <mark>V</mark> RL <mark>NGS</mark> AASH (10) KDLDLIFC (9)QTVKD <mark>VV</mark> LDC <mark>L</mark> LDF(21	.) KM <mark>V</mark> KVC(3)
	Q96A09 Hs FAM46B	(53)-WP <mark>Q</mark> VKR <mark>L</mark> D	-AL <mark>L</mark> SEP (10) TLSVQPRQ <mark>I</mark> V	VQV <mark>V</mark> RST <mark>L</mark> EEQ(5) SVRLHGSAASH (10) K <mark>DLD</mark> LVFR (9) QLTKA <mark>VV</mark> LAC <mark>L</mark> LDF (21)KL <mark>V</mark> KVC(3)
	O5VWP2 Hs FAM46C	(18) - WDOVSBLH	-EVITEV (10) TLETTIKDIV	VOTVRSBLEEA(5) DVRLNGSAAGH (10) KDLDLTEH (9) OLVEDVVLCSLUNF (21) KLVKVC (3)
	OSNEKS He FAMAGD	(10) - WDOVITUD	OVIDEV (10) TMEVERDIT		5) DARTNCSVASY (9) OVVKDAVIDCIIDE (21) KINKVC(3)
	QONERO HS FAM40D	(104) 00001110	-QVLDEV(10) IMEVRFRD11		J) DARLINGSVASI (3) QVVKDAVLDCLLDF (21	
	Q9NVV4 Hs TUTI	(184) S <mark>1</mark> DDQ <mark>L</mark> N	-TL <mark>L</mark> KEF(4)ENTKLRYL <mark>T</mark> (JSL <mark>I</mark> EDM <mark>A</mark> AAY <mark>F</mark> (3) IVRPEGSSVNT (5) CDLDMFLD (33)KILSV- <mark>L</mark> GEC <mark>L</mark> DHF(4)VG <mark>V</mark> QKI(5)
	Q6PIY7 Hs TUT2	(152)KDK <mark>L</mark> SQQ <mark>I</mark> L	-EL <mark>F</mark> ETC(4)SDLKKKEL <mark>C</mark> F	RTQ <mark>L</mark> QRE <mark>I</mark> QLL <mark>F</mark> (3)RLFLVGSSLNG(5) <mark>SD</mark> GDLCLV(20)TLVHK- <mark>H</mark> FCTRLSG(2)E-RPQL(5)
	O8NDF8 Hs TUT3	(116) VVGLHEEIS	-DF(8)EEEKMRME <mark>V</mark> V	√NR <mark>I</mark> ESV <mark>I</mark> KEL <mark>W</mark> (3) DVOIFGSFKTG (5) <mark>SDIDLVV</mark> F(5) LPLWT-LEEALRKH (6)S- <mark>V</mark> KVL(5)
	05TAX3 Hs TUT4	(947) RETLDLVCK	-RCEDEL (4) SEOHNREOTI	LIG <mark>LEKETOKEY(</mark>	4) BLCLEGSSKNG (5) SDLDTCMT (8) KLNCK-ETTENLAK (9	N = TLPT(5)
	057097 40 7075		DF(Q)FFAAMDDFWU		3) DVOT FCSF STC (5) PRIOT -TROATPRU(6	S = T K V T (-5)
	Q5XG67 H5 1015	(210) 100 11111	DF (0) EEAAMARE	/RRIETVVRDDW(S) DVQ1FGSFS1G(J)FFEQL-BEQALKKII(0	
	Q9H6E5 HS TUT6	(155) AAD <mark>V</mark> GAQ <mark>M</mark> I	-KL <mark>V</mark> GLR(4)AERQLRSL <mark>V</mark> V	VALMOEVELEEE.	3) VVHPEGSSINS (2) CDTDTETD (1	U8)KEEKA-EGAA <mark>m</mark> lel(8	GVI(5)
			OOVEDD (A) TEDOTDEUTE				OLOX DOLL DOTEDT TO CO	INTTOTICES
	Q5VYS8 Hs TUT7	(997)LNI <mark>L</mark> DQV <mark>C</mark> I	-QC <mark>IR</mark> DF(4)IEDQAREH <mark>I</mark> F	RQN <mark>L</mark> ESF <mark>I</mark> RQD <mark>F</mark> (3) K <mark>LS</mark> L <mark>FGS</mark> KNG (5) <mark>SDLDVCM</mark> T (8)GLDCV-RTIEELAR(9)N- <mark>I</mark> LFI(5)
	Q5VYS8 Hs TUT7	(997) LNI <mark>L</mark> DQVCI	-QCIKDF(4)IEDQAREHIF	RQN <mark>L</mark> ESF <mark>I</mark> RQD <mark>F</mark> (3) KLSLFGSSKNG (5) SOLOVCMT (8)GLDCV-RTIEELAR(9	>
	Q5VYS8 Hs TUT7	(997)LNILDQVCI NTa EEEEEEEE	-QC <mark>IK</mark> DF(4)IEDQAREH <mark>I</mark> F I SE	RQNLESFIRQDE(-> (1H HHHHHHHHHHHHHHHH	3) KLSLFCSSKNG (5) SDLDVCMT (PAP/OAS1	8)GLDCV-RTIEELAR(9	> ІНННННН (1 55)
	Q5VYS8 Hs TUT7	(997)LNILDQVCI NTa EEEEEEE EEEE PVIKL-C-F(2)IET	-QC <mark>IK</mark> DF(4)IEDQAREH <mark>I</mark> F ISE EEEE HHHHHHHHHH IIIFA(29)IN <mark>G</mark> CRV <mark>U</mark> DEILF	RQNLESFIRQDE(->	3) KLSLFCSSKNG (HHHHHHHH R (12) G <mark>V</mark> SWAMLVAR	5) SDLDVCMT(PAP/OAS1 HHHH HHHHH [CQL(5)ASTLV	8)GLDCV-RTIEELAR(9 HHHH HHHHHHHHHH HK <mark>FF</mark> (57)VSTRMV <mark>M</mark> VEE <mark>F</mark> K	> HHHHHHH QGLA <mark>IT</mark> D(165)
	Q5VYS8 Hs TUT7	(997)LNILDQVCI EEEEEEEE EEEE PVIKL-C-F(2)IEI PIIKI-K-F(2)ISI	Se EEEE HHHHHHHHHHH LLFA(29)LNGCRVMDEILE	RQNLESFIRQDE(->	3) KLSLFCSSKNG (HHHHHHHH R (12) G <mark>V</mark> SWAMLVAR R (12) GVAWAMLVAR	5) SOLOVCMT(PAP/OAS1 HHHH HHHHH FCQL(5)ASTLV ICQL(5)SAVIL	8)GLDCV-RTIEELAR(9 HHHH HHHHHHHHH HKFF(57)VSTRMYMVEER NRFF(57)ESTKKVILQEFV	> HHHHHH QGLAITD(165) RGVQITN(210)
	Q5VYS8 Hs TUT7	(997)LNILDQVCI 	Se EEEE HHHHHHHHHHH ILFA (29) LNGCRVTDEILE LICA (29) LNGTRVTDEILE VSFE (3) GIEAAKLIRE	RQNLESFIRQDF	3) KLSLFCSSKNG (HHHHHHHHH R (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) GESIICLVFS	5) SOLOVCMT(PAP/OAS1 HHHH HHHHH ICQL(5)ASTLV ICQL(5)SAVIL FLHM(14)LGVLL	8) GLDCV-RTIEELAR(9 HHHH HHHHHHHHH HKFF(57) VSTRMVMVEEFK NRFF(57) ESTKKVILQEFV IEFF(58)NIRDIKKAFA	,
	Q5VYS8 Hs TUT7	(997) LNI LOQVCI 	ISE	RQNLESFIRQDE(+ HHHHHHHHHHHHHH HL(5)NFRLTIRAIKLMAKI SL(5)VFRIAIRAIKLMAQI W(5)GIRELVLIVKOELMA W(5)GIRELVLIVKOELMA (Y 5)RVFALVFSRCAN;	3) KLSLFCSSKNG (H HHHHHHHH R (12) GVSWAMLVAR R (12) GVAWAMLVAR A (12) GFSIICLVFS A (12) GFSIICLVFS A (12) NSLTMWUF	5) SOLOVEMT (PAP/OAS1 HHHH HHHHH ICQL (5) ASTLV ICQL (5) SAVIL FLHM (14) LGVLL FLOR (46) LELLL	HHHH HHHHHHHHH HKFF (57) VSTRMVWYEEK NRFF (57) ESTKKVILQEFV IEFF (58)NIRDIKKARA KFFF (24) OSOLOKEVDIAR	> HHHHHHH QGLAITD(165) RGVQITN(210) GAFDLLT(26) ESAWILO(28)
	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3pq1_A Hs PAP 41+6_A Hs PAP	(997)LNILDQVCI EEEEEEEE EEEE PVIKI-C-F(2)ISI PIIKI-K-F(2)ISI PIIKF-V-E(4)IHI PLVKF-S-H(4)FQC VVKF-S-H(4)FQC	SEEEE HHHHHHHHHH LICA(29) LNGCRVNDELLE LICA(29) LNGCRVNDELLE LICA(29) - GIEAAKLIRE LITN(1) RIALISSELLY VVEA(29) LNGCPVNDETLE	-> [(HHHHHHHHHHHHH HL (5)NFRLTIRAIKLMAKI SL (5)VFRLAIRAIKLMACI SW (5)GLRELVLIVKOFLH/ (5)FVRALVFSVRCMAR/ H/ (5)FVLTIRAKIKAR/	3) KLSLFCSSKNG (HHHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GPSIICLVFS (13) NFSLTMWVIF (12) GVAWMUVAP	5) SOLOVEMT (PAP/OAS1 HHH HHHH PCQL (5) ASTLV ICQL (5) SAVIL THM (14) LGVLL FLQR (46) LELLL PCQL (5) ASTLV	HHH HHHHHHHH HKFF(57) VSTRWWVEEK NRFF(57) STKKVILQEV LEFF(58)NIRDIKKARA KEFF(24) QSQLQKEVDLAR HKFF(67) NSTRVWVEF	> HHHHHHH QGLAITD(165) RGVQITN(210) GAFDLIT(26) ESAWILQ(28) OGLAVTD(160)
	Q5VYS8 Hs TUT7 1q78 A Bt PAP 201p A Sc PAP 3nyb A Sc PAP 3pq1 A Hs PAP 41t6 A Hs PAP	(997)LNILDQVCI EEEEEEEE EEEE PVTKL-C-F(2)IEI PIKKI-K-F(2)ISI PIKF-V-E(4)IHI PUVRF-S-H(4)FOO PVIKF-E-F(2)IEI PUVRF-S-H(4)FOO PVIKF-E-F(2)IEI	ISE ILEA (29) LNGCRVIDELLE LLCA (29) LNGCRVIDELLE LLCA (29) LNGCRVIDELLE VSFE (3) GIEAAKLIRE LITIN (1) RIALISSELLYI LVFR (29) LNGCRVIDELLE VSFE (29) LNGCRVIDELLE	->	3) KLSLFCSSKNG (HHHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GPSIICLVFS A (13) NFSLTMWVIF R (12) GVSWAMLVAR	5) SOLOVCM T (PAP/OAS1 HHH HHHH ICQL (5) ASTLV ICQL (5) ASTLV	HHHH HHHHHHHHH HKFF (57) VSTRMVMVEEK NRFF (57) SSTRKVILOEV IEFF (58) NIRDIKKAPA KKFF (24) QSQLQKEVULAR HKFF (57) TSTRTVMVEEK	
	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3pq1_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase	(997) LNI LDQVCI EEEEEEEE EEEE PVIKL-C-F(2) IEI PIIKI-K-F(2) ISI PIIKE-V-E(4) IHI PLVRF-S-H(4) FQO PVIKF-E-F(2) IEI PVVKLRF-A(110) YRM	SEEEE HHHHHHHHHH LLEA(29)INGCRVTDEIL JUCA(29)INGCRVTDEIL VSEE(3)-GIEAAKLIKY LTTN(1)RIALTSSELIY LVFA(29)INGCRVTDEIL SISFVGYGVKNSYLIRF	-> [(HHHHHHHHHHHHHHH HL(5)NFRLTLRAIKLMAKI EL(5)VFRIALRAIKLMAQI 3%(5)GLRELVLIVKOPLH/ IY(5)RVRALVFSVRCWAR/ IL(5)TFRLTLRAVKLMAKI HL(6)ARHHTAMAVKAMGK/	3) KLSLFCSSKNG (HHHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GPSIICLVFS A (13) NFSLTMMVIF R (12) GVSWAMLVAR A (12) SVAVTVMFIY		8) GLDCV-RTIEELAR (9 HHHH HHHHHHHHH HKFF (57) VSTRNVWVEER NRFF (57) ESTKKVILQEFV IEFF (58)NIRDIKKAFA KEFF (52) JSTRTVWVEER HKFF (57) JSTRTVWVEER HGFF (64) PAKFQLVKQEFI	> HHHHHH QGLAITD(165) RGVQIIN(210) GAFDLIT(26) LESAWILQ(28) QGLAVID(160) RAAQCME(30)
	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3pq1_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase 2ikf_A Tb TUTase	(997) LNILDQVCI	ILFA (2) INGCRVTDEILF ILFA (2) INGCRVTDEILF ILFA (2) INGCRVTDEILF ILFA (2) INGCRVTDEILF USFE (3) - GIRAKLIRR ITTN (1) RIALTSSELLYI IVFA (2) INGCRVTDEILF ISFVGYGVKNSYLIRF ITTAY (1) RNGVRNSALLRA	RQNLESFIRQDE(H HHHHHHHHHHHHHH HL(5)NFRLTIRAIKLMAKE SL(5)VFRIAIRAIKLMAKE SW(5)GJRELVLIVKOPLH IV(5)RVRALVFSVRCMARA HL(5)TFRLTIRAVKLMAKE HY(6)AARHTAMAVKAMKE HY(5)CORWLSMSIKKANKE HY(5)CORWLSMSIKKANKE H(5)CORWLSMSIKKANKE H)(5)CORWLSMS	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GSI ICLVFS A (13) NESLTMNVIF R (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) SYAVTVMFIY Q (13) SYAVTVMFIY Q (13) SYAVTVMFIY Q (13) SYAVTVMFIY	5) SOLOVCM T (PAP/OAS1 HHH HHHHH TQQL (5) ASTLV TQQL (5) ASTLV TQQL (5) ASTLV PLQR (46) LELLL PLQR (46) LELLL TQQL (5) ASTLV (LLV (34) LGRLL (LLQ (35) LETQV	HHHH HHHHHHHHH HKFF (57) VSTRMVMVEEK NRFF (57) ESTKKVILQEEV LFF (58)	
	Q5VYS8 Hs TUT7 lq78_A Bt PAP 201p_A Sc PAP 3pq1_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase 2ikf_A Tb TUTase 3hj1 A Tb TUTase	(997)LNILDQVCI EEEEEEEE EEE PVIKL-C-F(2)ISI PIIKT-K-F(2)ISI PIIKT-V-E(4)IHI PVRF-S-H(4)FQO PVIKF-S-F(2)IEI PVVKLRF-A(110)YRW PVVRV-K-G(3)VDP PVVP-T-D(4)IHO	SEETE HHHHHHHHHHH LICA (29) LNGCRVTDEILE VSFE (3)GIEAAKLIRE LUTN (1) RTALTSSELLYI LVFA (29) LNGCRVTDEILE LVTA (29) LNGCRVTDEILE LVFA (29) LNGCRVTDEILE ISFVGYGVKNSVLIRE JTAY (1) RNGVRNSALLRE VSIG (1) LGGVENSKILCZ	RQNLESFIRQDE(HHHHHHHHHHHHHHH HL (5)NFRLTIRAIKLMAKI SL (5)VFRIAIRAIKLMAQI SW (5)GLRELVLIVKOELH IY (5)VFRIAILWSVCMARI HL (5)TFRLTIRAVKLMAKI HY (6)ARHTMAVKAMGKI YY (5)PCRULSMSIKRWSK(IY (6)FCALIHLVKAMGKI I (6)FCALIHLVKAMGKI	3) KLSLFCSSKNG (H HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSIICLVFS (12) GVSWAMLVAR (12) GVSWAMLVAR (12) SVGFNLWVVY (12) SVGFNLWVVY (12) SJTVTWALM	5) SOLOVCM T (PAP/OAS1 HHHH HHHHH ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ILLV (34) LGRLL ICQL (35) LGTQV ILQE (48) VFFCL	HHHH HHHHHHHHH HKFF (57) VSTRMVWVEER NRFF (57) ESTKKVILQEFV IEFF (58)NIRDIKKAR KKFF (24) QSLQKYVDLAR HKFF (57) TSTRTVMVEER HCFF (64) PAKFQLWKQEEL LDFL (65) ASRVRHLQQERN	> HHHHHH QGLAITD(165) RGVQITN(210) GAFDLT(26) QGLAVID(160) RAAQCME(30) KARDTAL(4) KARDTAL(4)
	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3pq1_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase 2ikf_A Tb TUTase 3hj1_A Tb TUTase 4e7x_A Sp TUTase	(997) LNI LOQVGI	ISE ILEFA (29) LNGCKVDEILF ILFA (29) LNGCKVDEILF LICA (29) LNGCKVDEILF LTTN (1) RIALTSSELLYI LVFA (29) LNGCKVDEILF ISFV CGVKNSVLIRF ITAY (1) RNGVRNSALLRP VSIG (1) IGGVENSKILCZ VSIG (1) IGGVENSKILCZ	RQNLESFIRQDE(H HHHHHHHHHHHHHH HL (5)NFRLTIRAIKLMAKI EL (5)VFRIAIRAIKLMAKI EV (5)CHELVLIVKOPLH IY (5)RVRALVFSVRCMAR/ HL (5)TFRLTIRAVKLMAKI Y (6)ARHTMAVKAGKX YY (5)PCRWLSMSIKRWSK Y (5)FCAYIHLWANGKX Y (5)FCAYIHLWANGKX Y (5)FCAYIHLWANGKX	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GSI ICLVFS A (13) NFSLTMMVIF R (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) SVGFNLMVVY A (12) SFVTTMALM R (12) SFVTTMALM R (12) SFVTTMALM	5) SOLUVCM T (PAP/OAS1 HHH HHHH PQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV PLQR (46) LELLL PLQR (46) LELLL PLQR (46) LELLL (LLQ (35) LGTQV /LQC (48) VFFCL (LSSLL (LLGLL) LGSLL	HHHH HHHHHHHHH HKFF (57) VSTRMVMVEEPK NRFF (57) ESTKKVILQEP LFF (58) – MIRDIKKAPA KEFF (24) QSQLQKFVDLAR HKFF (57) TSTRTVMVEEPK HKFF (57) PAKFQLWKQEFL LDFL (65) PLKRDFLRRHDE DRFA (95) ASRVRHLQEP NGFF (67) SSGLYRIEGEM	
Ve	Q5VYS8 Hs TUT7 1q78 A Bt PAP 201p-A Sc PAP 3pq1 A Hs PAP 41t6-A Hs PAP 2b4v-A Tb TUTase 2ikf A Tb TUTase 3hj1-A Tb TUTase 4e7x_A Sp TUTase 30uv-A Af CCA	(997) LNI LDQVCI EEEEEEEE EEE PVIKL-C-F(2) IEI PIIKF-V-E(4) IHI PURR-S-H(4) FQO PVIKF-E-F(2) IEI PVVKLRF-A(110) YRD PVVQF-T-D(4) IHO PVVQF-T-D(4) IHO PIIKL-TSD(8) FQO PVVQ-V-V(2) VEV	SEEEE HHHHHHHHH ILFA (29) LNGCRVIDEILH LICA (29) LNGCRVIDEILH VSFE (3) GIEAAKLIRE LITN (1) RIALTSSELLYI LVFA (29) LNGCRVIDEILH LVFA (29) LNGCRVIDEILH ISFVGYGVKNSYLIRA VSIG (1) IGGVENSKILCA VSIG (1) IGGVENSKILCA UGRN (1) RIAIHMILLSS	RQNLESFIRQDE(H HHHHHHHHHHHHHHH HL(5)NFRLTIRAIKLAAQ SW(5)GLRELVLIVKQPLM IV(5)RVRALVFSVRCMAR/ HL(5)TFRLTIRAVKLMAKI IV(6)ARHTAMAVKMGK/ V(5)PCRMLSMSIKRNSK/ V(5)PCRMLSMSIKRNSK/ V(5)PCRMLSMSIKRNSK/ V(5)PCRMLSMSIKRNSK/ V(5)PCRMLSMSIKRNSK/ SPC)CRENDVRLKA- SPC)CRENDVRLKA-	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GPSIICLVFS A (12) GSSWAMLVAR A (12) GVSWAMLVAR A (12) SGVSWAMLVAR A (12) SGVSWAMLVAR A (12) SGVSWAMLVAR A (12) SGVVLWVLY A (12) SGVVLWVLY A (12) SGCVLWVLY A (12) S	5) SOLUVCM T (PAP/OAS1 HHH HHHHH FCQL (5) ASTLV (CQL (5) ASTLV (CQL (5) ASTLV FLMM (14) LGVLL FCQL (5) ASTLV (FLQ (36) LGFLV (FLQ (35) LGFLV /LQC (48) VFFCL (IIH (43) LGSLU (GSE	HINH HHHHHHHHH HKFF (57) VSTRWVMVEEK NRFF (57) SSTRKVIJQE PV IEFF (58) NIRDIKKAPA HKFF (24) QSQLQKEVVDLAR HKFF (24) QSQLQKEVVDLAR HKFF (57) TSTRTVMVEER HGFF (64) PAKPQLVKQET LDFL (65) PLKRPTLRRHLE QRFA (95) ASRVRHIQEPN HGFF (67) SSGLYRIRGE PV HGFF (67) SSGLYRIRGE	
Active	Q5VYS8 Hs TUT7 1q78 A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3pq1_A Hs PAP 4lt6 A Hs PAP 2b4v_A Tb TUTase 2lkf_A Tb TUTase 3hj_A Tb TUTase 4e7x_A Sp TUTase 3ouy_A Af CCA	(997) LNI LDQVCI EEEEEEEE EEEE PVIKL-C-F(2) IEI PIIKI-K-F(2) ISI PIIKI-K-F(2) ISI PIIKI-K-F(2) ISI PVIKP-S-H(4) FQO PVIKP-F-F(2) IEI PVVKU-K-G(3) VDI PVVQ-T-D(4) IHQ PVVQ-T-D(4) IHQ PVIKI-TSD(8) FQO PYUHC-V-V(2) VEV	ISECTION (1) IEDUAREHEI IEEEE HHHHHHHHH IEFA (29) LNGCRVTDEILE DICA (29) LNGCRVTDEILE USFE (3) - GIRAALIRR UTTN (1) RIALTSSELLY UVFA (29) LNGCRVTDEILE ISFVGYGVKNSLIRR VSIG (1) IGGVENSKILCR VSIG (1) RIGVENSKILCR VSIG (1) RIALHNTLLSS VVPG (13) DRTPFHKKNLES	RQNLESFIRQDE(H HHHHHHHHHHHHH HL (5) NFRLTIRAIKLMAKI SU (5) UFRIAIRAIKLMAKI SU (5) UFRIAIRAIKLMAKI IY (5) RVHALVFSVRCMARI HL (5) TFRLTIRAVKLMAKI Y (6) ARHTMAVKAMKK Y (5) PCRWLSMSIKRMSK AI (6) FYGAYIHLVKAMKK Y (5) PCRWLSMSIKRMSK AI (6) FYGAYIHLVKAMKK SU (5) CRUNEVRLIKG	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSLTMWIF (12) GVSWAMLVAR (12) GVSWAMLVAR (12) SVGFNLMVIF (12) SVGFNLMVY (13) SVGFNLMVY (13) SVGFNLMVY (14) GVLCELLIVF (14) GVLCELLIVF	5) SOLUVENT (PAP/OAS1 HHH HHHHH ICQL(5) ASTLV ICQL(5) ASTLV ICQL(5) ASTLV ICQL(5) ASTLV ICQL(4) LEILL ICQL(4) LEILL ILQ(35) LGTQV ILQ(48) VFFOU ILIH(43) LGSLL ICC	HHH HHHHHHHH HKFF (57) VSTRWWVEER NRFF (57) SSTKKVILQEV LFF (58)NIRDIKKAPA KEFF (24) QSQLQKFVDLRK HKFF (57) TSTRTVMVEER KGFF (64) PAKFQLWKQERL LDFL (65) PLKRDFLRRHLE QRFA (95) ASRVRHLQQE PN HGFF (67) SSGLYRIRGEP LETV (42) LDNLARFVHLGR	> HHHHHHH QGLAITD(165) GAVDITN(210) GAVDIT(210) CGLAVID(160) RAAQCME(30) KARDTAL(4) RIREMLI(14) RIREMLI(43) EFFMEDE (200)
Active	Q5VYS8 Hs TUT7 1q78 A Bt PAP 201p A Sc PAP 3nyb A Sc PAP 3pq1 A Hs PAP 4lt6 A Hs PAP 2lkf A Tb TUTase 3hj1 A Tb TUTase 4e7x A Sp TUTase 30uy A Af CCA 4x4n A Af CCA	(997) LNI LDQVCI EEEEEEEE EEE PVIKI-C-F(2) IEI PIIKI-K-F(2) IEI PIIKT-K-F(2) ISI PIKF-S-H(4) FQO PVIKF-E-F(2) IEI PVVKV-K-G(3) VDE PVVQF-T-D(4) IHO PIKL-TSD(8) FQO PVVQF-V-V(2) VEVE PVVHG-V-V(2) VEVE	SCERDF (4) IEDOAREHE ILGA (29) LNGCRVDEILF LICA (29) LNGCRVDEILF LICA (29) LNGCRVDEILF LICA (29) LNGCRVDEILF LITN (1) RIALTSSELLYI LVFA (29) LNGCRVDEILF ISFVGYGVKNSYLIRF ISFVGYGVKNSYLIRF USIG (1) RGVENSKLLC2 UGEN (1) RLAIHNILLS VVPC (13) DRTPFHHKWLEC VVPC (13) DRTPFHHKWLEC	RQNLESFIRQDE(HHHHHHHHHHHHHHH HL(5)NFRLTIRAIRLMAKE EL(5)VFRIAIRAIRLMACE EW(5)GLRELVLIVKQFIH (5)VFRIAIRAIRLMAKE HI(5)TFRLTIRAVKLMAKE HI(6)AARHTAMAVKAWGKY 41(6)FYGAYIHLWKAMGKY 41(6)FYGAYIHLWKAMGKY 41(6)FYGAYIHLWKAMGKY 41(6)FYGAYIHLWKAMGKY 41(6)FYGAYIHLWKAMGKY 41(6)FYGAYIHCKG R(3)-KENEVRLIKGELKY	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GFSIICLVFS A (12) GSWAMLVAR A (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) SMAVTVMFIY Q (13) SMGFNLMVVY A (12) SMGFNLMVVY A (12) SMGYVLMVLY C (18) GMLCELLIVF A (14) GMLCELLIVF	5) SOLUVCM T (PAP/OAS1 HHH HHHHH TCQL (5) ASTLV TCQL (5) ASTLV TCQL (6) LELLL TCQL (6) LELLL TCQL (5) ASTLV (LLQ (35) LGTQV VLQC (48) VFFCL LI H (43) LGSL CGSF CGSF	HHH HHHHHHHHH HKFF (57) VSTRWVMVEEK NRFF (57) SSTRKVILOEV IEFF (58) NIRDIKKAPA KKFF (24) QSQLQKEVULAR HKFF (57) TSTRTVMVEEK HGFF (64) PAKFQLVKQEFI LOFL (65) PLKRDFIRRHE QRFA (95) ASRVHLQEN HGFF (67) SSGLYRIRGEM HGFF (67) SSGLYRIRGEM HGFF (42) LDNLAREVHLQR	
Active	Q5VYS8 Hs TUT7 1q78 A Bt PAP 2o1p A Sc PAP 3pq1_A Hs PAP 4lt6 A Hs PAP 2b4v_A Tb TUTase 2hf_A Tb TUTase 3hj1_A Tb TUTase 4e7x_A Sp TUTase 3cuy_A Af CCA 4x4n_A Af CCA 1px5_A Sc OAS	(997) LNI LDQVGI EEEEEEEE EEEE PVIKL-C-F(2) ISI PIIKI-K-F(2) ISI PIIKT-V-E(4) IHI PVVRF-S-H(4) FQO PVVRF-S-F(2) IEI PVVKLRF-A(110) YRW PVVRV-K-G(3) VDP PVVPC-T-D(4) IHO PIKL-TSD(8) FQO PYVHG-V-V(2) VEW -ALSF-V-L(8) VEP	SEEEE HHHHHHHHHH LIEA(29)LNGCRVTDEILL LICA(29)LNGCRVTDEILL LICA(29)LNGCRVTDEILL USEE(3)-GIEAAKLIRR LITTN(1)RIALTSSELLYI LITY(29)LNGCRVTDEILL LIFVGYGVKNSVLIRR USE(1)IGGVENSKLIC2 VSIG(1)RIGVENSKLIC2 VSIG(1)RIAHHMLLLS VVPC(13)DRTPFHHKWLEC VUPC(13)DRTPFHHKWLEC VUPC(13)DRTPFHHKWLEC VUPC(13)DRTPFHHKWLEC VUPC(13)DRTPFHHKWLEC VUPC(13)DRTPFHHKWLEC	RQNLESFIRQDE(H HHHHHHHHHHHHH HL(5)NFFLTLRAIKLAAK EL(5)VFRIALRAIKLAAK EV(5)CFLEVUIVKOPLH IV(5)RVRALVFSVRCAAR HL(5)FFLTLRAVKLAKT IV(6)AARHTAMAVKAOKK AI(6)FYGAYIHLVKAOKK AI(6)FYGAYIHLVKAOKK AI(6)FYGAYIHLVKAOKK SV(5)RLKPWULLVKHAK SR(2)CKENEVRLIKG=LK 3)-(3)-KENEVRLIKG=LK (3)KLKSLIRLVKHWQ1	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GSWAMLVAR (12) GSSITMVVIF (12) GVSWAMLVAR (12) SSITMVVIF (12) GVSWAMLVAR (12) SGVTVMFIY (12) SGVTVMFIY (12) SGVTVTMALM (12) SGVTVTMALM (12) SGVLWLIVF (12) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF	5) SOLUVCM T (PAP/OAS1 HHH HHHHH TCQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV TCQL (5) ASTLV ILUX (34) LGRLL ILUX (34) LGRLL ILUX (35) LGRQV ILUX (35) LGRQV I	HHH HHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) SSTKKVILQEY IEFF(58)NIRDIKKAPA HKFF(24) QSQLQKEVDLAR HKFF(57) TSTRTVMVEEK HKFF(57) TSTRTVMVEER HKFF(57) SSGLYRIRGEM HCFF(67) SSGLYRIRGEM LETV(42) LDNLAREVHLCR TVL(53) THSWQRLAQEAR	> HHHHHHH QGLATTD (165) 'RGVQITN (210) GAFDLIT (26) (ESAWILQ (28) (C) RAAQCME (30) (KARDUAL (4) IRLREMLI (19) IAASRLIN (43) LEFMEAP- (192) UEFMEAP- (192) VWLGY (20)
Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 4lt6_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase 3hj1_A Tb TUTase 3ouy_A Af CCA 4x4n_A Af CCA 1px5_A Sc OAS 4ig8_A Hs OAS	(997) LNILDQVCI NT EEEEEEEE EEEE PVTKL-C-F(2)ISI PIKKT-K-F(2)ISI PIKKF-V-E(4)IHI PUVRF-S-H(4)FQO PVTKV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVVVVVV-K-G(3)VDDI PVVRVV-K-G(3)VDDI PVVRVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV	ILEA (29) LNGCRVNDEILH ILEA (29) LNGCRVNDEILH ILEA (29) LNGCRVNDEILH ILEA (29) LNGCRVNDEILH ILTN (1) RIALTSSELLYI LVTN (29) LNGCRVNDEILH ITN (1) RIALTSSELLYI ISFVCYGVNNSVLIRF VSIG (1) IGGVENSKLLCZ VIGEN (1) RIA HINTLLSS VVPC (13) DRIPFHHKWLEG VLPA (37) -FIELORDFLRN VLPA (37) -FIELORDFLRN	RQNLESFIRQDE(H HHHHHHHHHHHHHH HL(5)NFRLTLRAIKLMAKE EL(5)VFRIAIRAIKLMAKE EW(5)GLEEVLIVKOPLH IY(5)RVRALVFSVRCMARJ IL(5)TFRLTLRAVKLMAKE Y(6)ARHTAMAVKAMCKA Y(6)FVGAYIHLVRAMCKA Y(5)FCRWLSMSIKRNSK(A)(6)FVGAYIHLVRAMCKA Y(5)RLKPWLLVRHMAKE R(2)GKENEVRLLKGE R(3)-KENEVRLLKGELKA (3)KLKSLIRLVKHWYQ) R(2)KLKSLIRLWKHYQ) R(2)KLKSLIRLWKHYQ)	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GSI ICLVFS A (13) NESLTMMVIF R (12) GSI ICLVFS A (13) NESLTMMVIF R (12) GVSWAMLVAR A (12) STVTIMALM R (12) STVTIMALM R (12) STVTIMALM R (12) STVTIMALM R (12) GLCELLIVF (12) GLCELLIVF F (11) GYALELLTVY R (10) GYAL (10) GYALELTVY R (10) GYALETVY R	5) SOLUVCM T (PAP/OAS1 HHH HHHHH PCQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV PLQR (46) LELLL PLQR (46) LELLL PLQR (46) LELLL ICQL (5) ASTLV (LLV (34) LGRLL LLQ (35) LGTOV VLQE (48) VFFCL (LLH (43) LGSLL ICQSF ICQ	HHHH HHHHHHHHH HKFF (57) VSTRMVMVEE FK (57) ESTKKVILQEFV URFF (57) ESTKKVILQEFV LFF (58)	> HHHHHHH QGLAITD (165) RGVQITN (210) (GAFDLIT (26) ESAWILQ (28) QGLAVTD (160) RAAQCME (30) KARDTAL (4) RAAQCME (30) KARDTAL (4) KARDTAL (4) KARDT
Active	Q5VYS8 Hs TUT7 lq78 A Bt PAP 201p-A Sc PAP 3pq1 A Hs PAP 4lt6 A Hs PAP 2b4v. A Tb TUTase 3hj1 A Tb TUTase 3hj1 A Tb TUTase 4e7x. A Sp TUTase 30uy A Af CCA 1px5 A Sc OAS 4ig8 A Hs OAS 4j1x A Ss cGAS	(997) LNILDQVCI EEEEEEE EEE PVIKL-C-F(2) ISI PIIKF-V-E(4) IHI PUVRF-S-H(4) FQO PVVKR-S-H(4) FQO PVVKR-C(3) VDP PVVR-C(3) VDP PVVR-C-D(4) IHO PIKL-TSD(8) FQO PVVR-V-V(2) VEW -ALSF-V-L(8) VEP -ALSF-V-L(8) VEP -ALSF-V-L(8) VEP -ALSF-V-L(8) VEP -ALSF-V-L(8) VEP	SEEEE HHHHHHHHHH ILFA (29) LNGCRVTDEILL LICA (29) LNGCRVTDEILL UCA (29) LNGCRVTDEILL VSEE (3)GIEAAKLIRR LITN (1) RIALTSSELLTY LITY (29) LNGCRVTDEILL ISFVGYGVKNSYLIRR LITN (1) RIALHMALLSS VVEC (13) DRTPFHHKWLEC VVEC (13) DRTPFHHKWLEC VVEC (13) DRTPFHHKWLEC VLPA (37) -FTELQRDFLRN VLPA (37) -FTELQRDF	RQNLESFIRQDE(H HHHHHHHHHHHHHH HL(5)NFRLTIRAIKLAAQ EU(5)VFRIALRAIKLAAQ EW(5)GLRELVLIVKOPLH IY(5)VFRALVFSVRCMAR/ HL(5)TFRLTIRAVKLMAKI IY(6)PTGAYIHLVKAWGK/ 3Y(5)PCWLSMSIKRNSK AI(6)FYGAYIHLVKAWGK/ 3Y(5)RLKPWLLVKHWGK/ BR(2)GKENEVRLKG-LK/ F(3)-KENEVRLKG-LK/ JR(3)KLKSLIRLVKHWYQ) 2R(2)KLKSLIRLVKHWYQ) 2R(2)KLKSLIRLVKHWYQ) 21(18)CKRECLKLMKYLES	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GPSIICLVFS A (12) GSSTLTMWIF R (12) GVSWAMLVAR A (12) STVTTMALM R (12) STVTMAL	5) SOLOVCM T (PAP/OAS1 HHH HHHHH FCQL (5) ASTLV (CQL (5) ASTLV (CQL (5) ASTLV FLMM (14) LGVLL FCQL (5) ASTLV (CQL (5) ASTLV (LQ (35) LGFLV (LQ (35) LGFLV (LQ (35) LGFLV (LQ (35) LGFLV (LQ (35) LGFLV (CGSF (CGSF WEQ (8) TAQGE WER (8) TAQGE WER (8) TAQGE (CTQ (11) LECCF	HIHH HHHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) SSTKVULQEFV LGFF(58)NIRDIKKAPA KEFF(24) QSQLQKEVDLAR HKFF(57) TSTRTVWVEEK HGFF(64) PAKFQLVKQET LDFL(65) PLKRDFLRRHLE QFA(95) SSGLYRIRGEFM HGFF(67) SSGLYRIRGEFM HGFF(67) SSGLYRIRGEFM LETV(42) LDNLAREVHLCR QTVL(53) THSWQRLAQEAR RTVL(53) FKGNGLAQEAR RTVL(53) KPSKEFLSKOIE	UPIC 10 UPIC 10 UPI
Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 2dyt_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase 3hj1_A Tb TUTase 4e7x_A Sp TUTase 4e7x_A Sp TUTase 4cx_A A Af CCA 4x4n_A Af CCA 1px5_A Sc GAS 4j1x_A Ss GAS 4k8v_C Mm cGAS	(997) LNILDQVCI NTC EEEEEEEE EEEE PVTKI-C-F(2) ISI PIIKT-K-F(2) ISI PIIKF-V-E(4) IHI PUVRF-S-H(4) FQO PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVR-V-C(2) VEVI PVVR-V-V(2) VEVI PYVHG-V-V(2) VEVI PYVHG-V-L(8) VEPI PAUTI-L-I(5) ISV	ISCALDE (4) IEDUAREHE IEEEE HHHHHHHHH ILFA (29) LNGTRVDEILE LICA (29) LNGTRVDEILE LITN (1) RIALTSSELLYI LVFA (29) LNGTRVDEILE ISFV CYGVRNSVLIRF ITAY (1) RNGVRNSALLRP VSIG (1) IGGVENSKILG7 VIGRV (1) RIA HINTLLSS VVPC (13) DRIPFHHKWLEG VVPC (13) DRIPFHHKWLEG VVPC (13) DRIPFHHKWLEG VUPC (13) DRIPFHHKWLEG VUPC (13) DRIPFHKWLEG VLPA (37) -FTELQRDFLRN VLPA (37) -FTELQRDFLRN VLPA (55) FSHTEKN VLA (55) FSHTEKN	<pre>RQNLESFIRQDE(H HHHHHHHHHHHHHHHHHHHHHHHL(5)NFRLTIRAIKLMAKF EL(5)VFRIAIRAKLMAKF EL(5)VFRIAIRAKLMAKF EV(5)RVRALVFSVRCMAR/ IY(5)TRLTIRAVKLMAKF IY(6)ARHTMMAVKANGK/ AY(5)FCRWLSMSIKRWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWSK(AY(5)FCRWLSWZ(AY(5)FCRWLSWSK(AY(5)FCRWLS</pre>	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GS I CLUFS A (13) NFSLTMMVIF R (12) GVSWAMLVAR A (12) SVTVTWAIY Q (13) SVGFNLMVVY A (12) SVVTWFIY R (14) GVLCELLIVF A (14)	5) SOLUVCM T (PAP/OAS1 HHH HHHH PQQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV PLQR (46) LELLL PLQR (46) LELLL [CQL (5) ASTLV (LV (34) LGRLL (LQ (35) LGTQV /LQE (48) VFFCL (GSF [G	HHHH HHHHHHHHH HKFF (57) VSTRMVMVEEPK NRFF (57) ESTKKVILQEPV IEFF (58) – NIRDIKKAPA KEFF (24) QSQLQKPVDLAR HKFF (57) TSTRTVMVEEFK ILDFL (55) PLKRDFLRRHDE DRFA (95) ASRVRHLQEEN LETV (42) LDNLARFVHLQF LETV (42) LDNLARFVHLQF RTVL (53) PKSWRPLAQEAR RTVL (53) PKSWRPLAQEAR DNCV (29) KPSKEFLSKOLE	> HHHHHHH QGLAITD(165) RGVQITN(210) (GAFDLIT(26) (ESAWILQ(28) QGLAVTD(160) (RAAQCME(30) (KARDTAL(4) (RAAQCME(30) (KARDTAL(4) (ASARLIN(43) (EFMEAP-(192) (EFMEAP-(208) (WLGY(20) (YEENNG-(7)) (YEENNG-(7))
Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 201p_A Sc PAP 3pyD_A Sc PAP 3pyD_A Hs PAP 41t6_A Hs PAP 2b4v_A Tb TUTase 3hj1_A Tb TUTase 3hj1_A Tb TUTase 4e7x_A Sp TUTase 30uy_A Af CCA 1px5_A Sc OAS 4ig8_A Hs OAS 4j1x_A Ss cGAS 4kwc_C Mm cGAS	(997) LNI LDQVCI	SEEEE HHHHHHHHHH ILFA (29) LNGCRVTDEILL ILFA (29) LNGCRVTDEILL UCA (29) LNGCRVTDEILL VSEE (3)GIEAAKLIRR UTTN (1) RIALTSSELLTN UVFA (29) LNGCRVTDEILL ISFVGYGVKNSYLIRR ITAY (1) RNGVRNSLLRR UTAY (1) RNGVRNSLLRR UTAY (1) RNGVRNSLLRR UTAY (1) RNGVRNSLLRR UVFA (3) DRTPFHHKWLEC VUFC (3) DRTPFHHKWLEC VUFC (3) DRTPFHHKWLEC VUFA (37) -FTELQRDFLRN VUFA (37) -FTELQRDFLRN ULFA (37) -FTELQRDFLRN ULFA (37) -FTELQRDFLRN ULFA (35) FSHTEKN ULFA (55) FSHTEKN	<pre>RQNLESFIRQDE(H HHHHHHHHHHHHHH HL(5)NFRITIRAIKLAAK EL(5)VFRIAIRAIKLAAQ EW(5)GIRELVLIVKQFIH TV(5)VFRIAIRAIKLAAK HL(5)TFRITIRAVKLAAK HL(6)ARHTAMAVKANGK/ YV(5)PCNLSMSIKRNSK AI(6)FYGAYIHLVKANGK/ SV(5)PCNLSMSIKRNSK AI(6)FYGAYIHLVKANGK/ SV(5)PCNLSMSIKAG SR(3)-KENEVRLIKG SR(3)-KENEVRLIKG SR(2)GKENEVRLIKG SR(2)KLKSIIRLVKHNYO D(18)CKKECIKLMKYLEG (10)CKKECIKLMKYLEG</pre>	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GPSIICLVFS A (12) GSWAMLVAR A (12) GSWAMLVAR A (12) GVSWAMLVAR A (12) SYGFNLMVVY A (12) SFUTTMALM R (12) SFUTTMALM R (12) SFUTTMALM R (12) SFUTTMALM C (12) SFUTTMA	5) SOLUVCM T (PAP/OAS1 HHH HHHHH ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (6) LELLL ICQL (5) ASTLV ILQ (34) LGRLL ILQ (35) LGRDV /LQC (48) VFFCL ILLQ (35) LGRDV /LQC (48) VFFCL ILLQ (35) LGRDV /LQC (8) TAQGF (GSF WER (8) TAQGF (CTQ (11) LECCF WTQ (11) LSCCF WTQ (11) LGCP	HINH HHHHHHHHH HKFF (57) VSTRWVMVEEK NRFF (57) SSTRKVILOEV IEFF (58) NIRDIKKAPA KKFF (24) QSQLQKEVDLAR HKFF (57) TSTRTVMVEEK HGFF (64) PAKFQLVKQEFI LOFL (65) PLKRDFLRRHLE QRFA (95) ASRVRHLQUEPN HGFF (67) SSGLYRIRGEM LETV (42) LDNLAREVHLCR QTVL (53) THSWQRLAQEAR RTVL (53) FKSKEFLSKKLE DKLL (29) KFSKEFLSKKLE DKLL (29) KFSKEFLSKKLE	UPL ()) UPL (
Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 2b4v_A Tb TUTase 2hf_A Tb TUTase 3b11_A Tb TUTase 4e7x_A Sp TUTase 3ouy_A Af CCA 4x4n_A Af CCA 1px5_A Sc OAS 4ig8_A Hs OAS 4j1x_A Ss CGAS 4k8v_C Mm CGAS 4k8v_C Mm CGAS	(997) LNI LDQVGI NTC EEEEEEEE EEEE PVIKI-C-F(2) IEII PIKR-V-F(2) ISII PIKR-V-E(4) IHI PVKR-S-H(4) FQO PVKF-E-F(2) IEII PVWR-F-A(110) YRM PVVQ-T-D(4) IHO PVVQ-T-D(4) IHO PVVQ-T-D(4) IHO PVVQ-T-D(2) VEVI PYVHG-V-V(2) VEVI PYVHG-V-U(2) VEVI PAUTA-LI(5) ISV PAVTI-L-I(5) ISV PAVTI-L-I(5) ISV PAVTI-L-I(5) ISV	ILEA (29) INGCRUTELLA LEFA (29) INGCRVTDEILL LICA (29) INGCRVTDEILL ULYA (29) INGCRVTDEILL UYA (29) INGCRVTDEILL ISTV GYGVNSYLIRH ISTV GYGVNSYLIRH ITAY (1) RNGVRNSALLRR VSIG (1) IGGVENSKILCR USTG (1) RALAHNTLLSS VVPC (13) DRTPFHHKWLEG VVPC (13) DRTPFHKWLEG VVPC (13) DRTPFHKWLEG ILLA (55) FSHIEK ILLA (55) FSHIEK	RQNLESFIRQDE(H HHHHHHHHHHHH HL (5)NFRLTIRAIKLMAKI EU (5)VFRIAIRAIKLMAKI EU (5)VFRIAIRAIKLMAKI EV (5)GRUFALVFSVRCMAR/ HL (5)TFRLTIRAVKLMAKI Y (5)FRVRLVFSVRCMAR/ HL (5)TFRLTIRAVKLMAKI Y (5)FCRWLSMSIKRWSK(AY (5)FCRWLSWSK(AY (5)F	3) KLSLFCSSKNG (HHHHHHHH R(12) GVSWAMLVAR R(12) GSI ICLVFS A(12) GSI ICLVFS A(12) GSI ICLVFS A(12) GSI ICLVFS A(12) SVSWAMLVAR A(12) SVSWAMLVAR A(12) SVSWAWLVAR A(12) SVSWAWLVAR A(12) SVSWAWLVAR A(12) SVSWAWLFFH A(14) GYLCELLVF A(14) GYLCELLIVF A(14) GYLCELLIVF A(14) GYLCELLIVF A(14) GYLCELLIVF A(14) GYLCELLIVF A(14) GYLCELLIVF A(14) GYLCELLIVF A(14) GYLCELLIVF A(14) GYLCELLVF A(14) GYLCELLVF A(14) GYLCELLVF A(14) GYLCELLVF A(15) SWAWKTAFFH A(15) SWAWKTAFFH A(15) SWAWKTAFFH A(15) SWAWKTAFFH A(15) SWAWKTAFFH A(15) SWAWKTAFFH	5) SOLUVCM T (PAP/OAS1 HHH HHHH PCQL (5) ASTLV ICQL (5) ASTLV (ILV (34) LGRLL (ILV (34) LGRLL (ILV (35) LGRUV /ILQ (48) VFFCL (ILV (34) LGSLL (GSF ICQ	HHHH HHHHHHHHH HKFF (57) VSTRMVMVEER NRFF (57) ESTKKVILQEV LFF (58) – NIRDIKKAPA KEFF (24) QSQLQKFVDLAR HKFF (57) TSTRTVMVEER LDFL (55) PLKRDFLRHLE DRPA (95) ASRVRHIQQEN LGFV (42) LDNLARFVHLQR LETV (42) LDNLARFVHLQR LETV (42) LDNLARFVHLQR RTVL (53) PKSNPLAQEAR RTVL (53) PKSNPLAQEAR DNCV (29) KRSKEFLSKVE DNCV (29) KRSKEFLSKVE DNCV (29) KRSKEFLSKVE DNCV (29) KRSKEFLSKVE	> HHHHHHH QGLAITD(165) RGVQITN(210) GGAFDLIT(26) ESAWIIQ(28) QGLAVID(160) RAAQCME(30) KARDTAL(4) RLREMII(192) EFMEAP-(192) EFMEAP-(192) EFMEAP-(200) WHEGY-(20) VYERNNG-(7) VYERNNG-(7)
Active	Q5VYS8 Hs TUT7 1q78 A Bt PAP 201p A Sc PAP 3nyb A Sc PAP 3nyb A Sc PAP 4lt6 A Hs PAP 4lt6 A Hs PAP 2lkf A Tb TUTase 3hj1 A Tb TUTase 4e7x A Sp TUTase 4e7x A Sp TUTase 4c7x A Sc OAS 4lya A Sc CAS 4lya A Sc CAS 4lya A Sc CAS 4kw C Mm cGAS	(997) LNI LDQVGI EEEEEEEE EEE PVIKI-C-F(2) IEI PIIKT-K-F(2) IEI PIIKT-K-F(2) IEI PIKF-S-H(4) FQO PVIKF-E-F(2) IEI PVVRURF-A(110) YRM PVVRV-K-G(3) VMD PVVQF-T-D(4) IHO PVVRV-K-G(3) VMD PVVQF-T-D(4) IHO PVVRV-V-V(2) VEV PVVQF-V-V(2) VEV PVVHC-V-V(2) VEV PVHG-V-V(2) VEV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV	SCERDF (4) IEDOAREHEF EEEE HHHHHHHHHH ILFA (29) LNGCRVDDEILF LICA (29) LNGCRVDDEILF LICA (29) LNGCRVDDEILF USFE (3) GIRAAKLIRR LTTN (1) RIALTSSELLYI LVFA (29) LNGCRVTDEILF IFVGYGVKNSYLIRF IFV (1) RIGVENSKLLC2 USFVGYGVKNSYLIRF UTAY (1) RIGVENSKLLC2 USFV (1) RIGVENSKLC2 USFV (1) RIGVENSKLC2 RIGVENSKLC2 USFV (1) RIGVENSKLC2 R	<pre>RQNLESFIRQDE(H HHHHHHHHHHHHHHHHHHHHHHHHHH</pre>	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GFSIICLVFS A (12) GSICLVFS A (12) GSWAMLVAR A (12) GSWAMLVAR A (12) SYAVTVMFIY Q (13) SYGFNLWVY A (12) SYGFNLWVY	5) SOLUVCM T (PAP/OAS1 HHH HHHHH TCQL (5) ASTLV TCQL (5) ASTLV TCQL (5) ASTLV TCQL (6) LELLL TCQL (6) LELLL TCQL (6) LELLL TCQL (5) LGTQV (LLQ (35) LGTQV (LLQ (35) LGTQV (LQ (48) VFFCL (LLQ (5) LGTQV (LQ (5) LGTQV (LQ (5) ASTLV (5) ASTLV (CQ (1) LGCCF WTQ (11) LSSCF	HHH HHHHHHHHH HKFF (57) VSTRMVMVEEK NRFF (57) VSTRMVMVEEK NRFF (57) SSTRKVILOET VEFF (58) NIRDLKKAPA KKFF (24) QSQLQKEVDLAR HKFF (57) TSTRTVMVEEK HGFF (64) PAKFQLVKQEFI LDFL (55) PLKRDFLRRHE QRFA (95) SSRVHLQEEN HGFF (67) SSGLYRIRGEM HGFF (67) SSGLYRIRGEM HGFF (67) SSGLYRIRGEM HGFF (57) TSGLYRLQEA DHCV (2) LDNLAREVHLCR TVL (53) THSWQRLAQEAR DNCV (29) KPSKEFLSKKIE DNCU (29) KRSKEFLSKKIE DNCU (29) RKSKEFLSKKIE	> HHHHHHH QGLAITD(165) RGVOITN(210) QGAFDLIT(26) ESAWILQ(28) QGLAVTD(160) RAAQCME(30) RAAQCME
Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyd_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase 3hj1_A Tb TUTase 4e7x_A Sp TUTase 4e7x_A Sp TUTase 4e7x_A Sc CAS 4ig8_A Hs CCA 4ig8_A Hs CAS 4kw_C Mm cGAS 4kw_CAS 4bcGAS	(997) LNI LDQVGI NTC EEEEEEEE EEEE PVIKI-C-F(2) IEI PIKR-V-E(2) ISI PIKR-V-E(4) IHI PUKR-S-H(4) FQO PVKRF-E-F(2) IEI PVWKRF-A(110) YRM PVVROF-T-D(4) IHO PVVROF-T-D(4) IHO PVVROF-T-D(4) IHO PVVROF-T-D(8) FQO PYHG-V-V(2) VEW PYHG-V-V(2) VEW PYHG-V-U(8) VEP ALSF-V-L(8) VEP ALSF-V-L(8) VEP PAVTL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(5) ISW	ILEAC ILEACA ILEAC ILEACA ILEAC ILEACA ILEAC ILEACA ILEACA ILEACA ILE	RQNLESFIRQDE(H HHHHHHHHHHHH HL (5) NFRLTIRAIKLMAKI EL (5) VFRITIRAIKLMAKI EU (5) VFRITIRAIKLMAKI EV (5) GURALVFSVRCMARJ HL (5) TFRLTIRAVKLMAKI Y (5) RVRALVFSVRCMARJ HL (5) TFRLTIRAVKLMAKI Y (5) RVRALVFSVRCMARJ Y (5) RVRALVFSVRCMARJ Y (5) RVRALVFSVRCMARJ SV (5) RURAVKANGKA AY (5) PCRWLSMSIKRWSKA AY (5) PCRWLSKA AY (3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GSI ICLVFS (12) GSI ICLVFS (13) NSSLTMWVIF (12) GSI ICLVFS (13) NSSLTMWVIF (12) SSUTMVIF (12) SSUTMVIF (12) SSUTMVIF (12) SSUTMVIF (12) SSUTTTMALM (12) SSUTTTMALM (12	PAP/OAS1 HHH HHHH PCQL (5) ASTLV ICQL (5) ASTLV ICQL (5) SAVIL TLQR (46) LELLL PLQR (46) LELLL ICQL (5) ASTLV ILQ (46) LELLL ICQL (5) ASTLV ILQ (48) VFCQ ILQE (48) VFCQ	HHHH HHHHHHHH HKFF (57) VSTRWWVEER NRFF (57) VSTRWWVEER VRFF (57) STRWWVEER KEFF (24) QSQLQKFVDLAR KEFF (24) QSQLQKFVDLAR KFF (57) TSTRVWVEFL LDFL (65) FLKRDFLRHLE QRFA (95) ASRVRHLQEFN LGFV (42) LDNLARFVHLQR VL (53) THSWQRLAQEAR CVU (53) THSWQRLAQEAR CVU (53) THSWQRLAQEAR CVU (53) THSWQRLAQEAR DNCV (29) KRSKEFLSKXE DNCV (29) KRSKEFLSKXE DNCU (29) KRSKEFLSKXE DNCU (29) KRSKEFLSKXE DNCU (29) KRSKEFLSKXE DNCU (29) KRSKEFLSKXE	> HHHHHHH QGLATD(165) 'RGVQITN(210) (GAFDLIT(26) (GAFDLIT(26) (GAFDLIT(26) (GARNIQ(210) (GA
Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2olp_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 4lt6_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase 3hj1_A Tb TUTase 3ouy_A Af CCA 1px5_A Sc OAS 4ig8_A Hs OAS 4j1x_A Ss CGAS 4km5_A Ms CGAS 4km5_A Ms CGAS 4u0n^A Vc CGAS	(997) LNI DQVGI NTT EEEEEEEE EEEE PUTKI-C-F(2) IEIT PIKKF-V-E(2) IEIT PIKKF-N-E(2) IEIT PUVKPS-E-F(2) IEIT PUVKURF-E-F(2) IEIT PUVKURF-E-F(2) IEIT PUVKURF-E-F(2) IEIT PUVKURF-E-F(2) IEIT PUVKURF-E-F(2) IEIT PUVKU-F-E-F(2) IEIT PUVKU-F-E-F(2) IEIT PUVKU-F-E-F(2) IEIT PUVKU-F-E-F(2) IEIT PUVKU-F-E-F(2) IEIT PUVKU-F-E-F(2) IEIT PAVTI-L-I(5) ISWT PAVTI-L-I(5) ISWT PAVTI-L-I(7) IEIT	SCEADF (4) IEDQAREHE ILFA (29) LNGCRVDEILF ILFA (29) LNGCRVDEILF LICA (29) LNGCRVDEILF LICA (29) LNGCRVDEILF LTTN (1) RIALTSSELLYI LVFA (29) LNGCRVDEILF ISFVGY GVKNSYLIRF ITAY (1) RIGVENSKLLC2 USFV (1) RIGVENSKLLC2 USFV (1) RIGVENSKLLC2 VUPC (13) DENPFHHKWLEC VUPC (13) DENPFHHKWLEC VUPC (13) DENPFHHKWLEC VUPC (13) DENPFHHKWLEC VUPC (13) DENPFHHKWLEC VUPC (13) DENPFHHKWLEC VIPA (37) -FTELORDFLK IILA (55)FSHTEKK IILA (55)FSHTEKK	RQNLESFIRQDE(HHHHHHHHHHHHHHH HL(5)NFRLTLRAIKLMAKI EL(5)VFRIALRAIKLMAKI EW(5)GLRELVLIVKQPLH IV(5)VFRIALRAIKLMAKI HV(6)AARHTAMAVKAMKKI HV(6)AARHTAMAVKAMKKI YV(5)FCRWLSMSIKRMSK(4)(6)FYCAYIHLUKAMKA HV(6)FYCAYIHLUKAMKA SY(5)RLKPWVLLVKHWAKI SY(5)RLKPWVLLVKHWAKI SY(5)RLKPWVLLVKHWAKI SY(5)RLKPWVLLKHWVLE (1)SCKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLLE (118)CKKECLKLMKYLE (118)CKKECLKLMKYLE (118)CKKECLKLMKYLE (118)CKKUCKFKARKE (118)CKKUC	3) KLSLFCSSKNG (HHHHHHHH R(12) GVSWAMLVAR A(12) GSI ICLVFS A(12) STATTMVIF Q(13) SGENLMVVI A(12) STATTMALM R(12) STATTMALM Q(14) STATTH Q(15) SHVKTAFFH Q(15) SHMATTN		HHH HHHHHHHHH HKFF (57) VSTRMVMVEEK NRFF (57) ESTKKVILQEP LFF (58) - MIRDIKKAPA KEFF (24) QSQLQKFVDLAR HKFF (57) TSTRTVMVEEK HGFF (64) PAKFQLVKQEPI LDFL (65) PLKRDFLRRHE QRFA (95) ASRVRHLQQEN LGFV (42) LDNLARFVHLCR ETV (42) LDNLARFVHLCR ETV (42) LDNLARFVHLCR TVL (53) FKSMRQLAQEAR RTVL (53) FKSMRQLAQEAR RTVL (53) FKSKEFLSKAE DKLL (29) KRSKEFLSKAE DKLL (29) KRSKEFLSKAE DKLL (29) KRSKEFLTKQE DKLL (29) KRSKEFLTKQE DKLL (29) KRSKEFLTKQE DKLL (29) RKSKEFLTKQE	> IHHHHHH QGLAITD(165) RGVQITN(210) QGAFDLIT(26) ESAWILQ(28) QGLAVTD(160) KARDTAL(4) RAAQCME(30) KARDTAL(4) KA
Active	Q5VYS8 Hs TUT7 $1q78_A$ Bt PAP $2o1p_A$ Sc PAP $3nyb_A$ Sc PAP $3pq1_A$ Hs PAP $41t6_A$ Hs PAP $2b4v_A$ Tb TUTase $2ikf_A$ Tb TUTase $3o1y_A$ Af CCA $4e7x_A$ Sp TUTase $4e7x_A$ Sp TUTase $4e7x_A$ Sc OAS $4ig8_A$ Hs OAS $4ig8_A$ Hs OAS $4ig8_A$ Hs OAS $4k8v_C$ Mm cGAS $4k8v_C$ Am cGAS $4k6S_A$ Hs cGAS $4u03_A$ Vc CGAS	(997) LNI DQVGI NTC EEEEEEEE EEEE PVIKI-C-F(2) IEI PIKI-K-F(2) ISI PIKF-V-E(4) IHI PUVRE-S-H(4) FQO PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(1) YEN PVVRF-E-A(110) YEN PVVRF-E-F(2) IEI PVVR-K-G(3) VDP PVVR-K-G(3) VDP PVK-K-G(3) VDP PVK-K-K-G(3) VDP PVK-K-K-G(3) VDP PVK-K-K-G(3) VDP PVK-K-K-G(3) VDP PVK-K-K-G(3) VDP PVK-K-K-G(3) VDP PVK-K-K-G(3) VDP PVK-K-K-G(3) VDP PVK-K-K-K-K-K-K-K-K-K-K-K-K-K-K-K-K-K-K-	ACCEADF (4) TEDUAREHEF TERE HHHHHHHHHH LICA (29) LNGCRVTDEILE LICA (29) LNGCRVTDEILE LICA (29) LNGCRVTDEILE USEC (3)GIEAAKLIRE USEC (3)GIEAAKLIRE VICA (29) LNGCRVTDEILE ISEVGYGVKNSLILRE VICA (1) IGGVENSKILCE VICA (3) -FIELGROFIKS VUCA (3) -FIELGROFIKS VUCA (3) -FIELGROFIKS VICA (3) -FIELGROFIKS ULA (55)FSHIEKS IILA (55)	RQNLESFIRQDE(++ HHHHHHHHHHHHH HL (5) NFRLTIRAIKLMAKI SU (5) VFRLTIRAIKLMAKI SU (5) VFRLTIRAIKLMAKI SU (5) VFRLVISVRCAA2 HL (5) TFRLTIRAVKLMAKI Y (5) RVHALVFSVRCAA2 HL (5) FORWLSWKINKNSK A1 (6) FYGAYIHLVKANGKX A1 (6) FYGAYILLKGFIK (1) (8) CRKECIKLMKYLLEX (1) (8) CRKECIKLMKYLLEX (1) (8) CRKECIKLMKYLLEX (6) HLRKVQFFKANGKD 25 (6) HLRKVQFFKANGDD	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GS I I CLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) GS I I CLVFS (13) NSSLTMWVIF (12) SG VTWFIY (13) SVGFNLMVVY (13) SVGFNLMVVY (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (15) SHVKTAFFH (15) SHVKTAFFH (15) SHVKTAFFH (15) SSLWATVNA (18) SISLMAATVNA (18) SISLMAATVNA (18) SISLMAATVNA	PAP/OAS1 HHH HHHH PCQL(5)ASTLV CCQL(5)ASTLV CCQL(5)ASTLV CCQL(5)ASTLV CCQL(5)ASTLV TCQR(46)LELLL PCQR(46)LELLL CCQC(48)VFFCL (LLV(34)LCRLL (LLV(34)LCRLL (CSF WEQ(6)TAQGF WEQ(6)TAQGF WEQ(6)TAQGF WEQ(6)TAQGF WEQ(6)TAQGF WEQ(11)LESCF (CTQ(11)LGLCF WTQ(11)LGLCF WTQ(11)LGLCF WTQ(11)LGLCF ILDS(7)LCFTM	HHH HHHHHHHH HKFF(57) VSTRWWVEER NRFF(57) SSTKKVILQEFV LFF(58)NIRDIKKAPA KEFF(24) QSQLQKFVDLRK HKFF(57) TSTRTVWVEER KGPF(64) PAKPQLWKQEL LDFL(65) PLKRDFLRRHLE QRFA(95) ASRVRHIQQEFN HGPF(67) SSGLYRIRGEPM LETV(42) LDNLARFVHLQR PVU(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR DNCV(29) KSKEFLSKVE DNCV(29) KSKEFLSKVE DNCU(29) KSKEFLSKVE DKLL(29) KSKEFLSKKE DKLL(29) KSKEFLSKKE DKLL(29) KSKEFLSKKE DKLL(29) KSKEFLSKKE DKLL(29) KSKEFLSKKE DKLL(29) KSKEFLTKQEE KIIA(29) GPERMDIMSKE	> HHHHHH QGLATD(165) RGVQITN(210) GAFDLIT(26) GAFDLIT(26) RAAQCME(30) KARDTAL(4) RLREMLI(19) RAAQCME(30) KMRDTAL(43) LEFMEAP-(192) WWLGY(20) WWLGY(20) WWLGY(20) YWERNMG-(5) YERNNG-(7) YERNNG-(7) YERNNG-(7) YERNNG-(7) YERNNG-(7) YERNNG-(7) YERNNG-(7)
Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2olp_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 4lt6_A Hs PAP 4lt6_A Hs PAP 2b4v_A Tb TUTase 3hj1_A Tb TUTase 3ouy_A Af CCA 1px5_A Sc OAS 4ig8_A Hs OAS 4j1x_A Ss CGAS 4kw_C Mm CGAS 4kw_C Ams CGAS 4ley_A Mm CGAS 4u03_A Vc CGAS 4u03_A Vc CGAS	(997) LNI LDQVGI NT EEEEEEEE EEE PVTKL-C-F(2) IEI PIKK-K-F(2) IEI PIKF-V-E(4) IHI PUKF-S-H(4) FOO PVKV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-V-V(2) VEV PVHC-V-V(2) VEV PVHC-V-V(2) VEV PALSF-V-L(8) VED PAVTL-L-I(5) ISV PAVTL-L-I(4) THI TCGRI-K-I(4) THI	SCEADF (4) IEDUAREHEF IEEEE HHHHHHHHH ILEA (29) INGCRVMDEILF ILCA (29) INGCRVMDEILF ILTA (29) INGCRVMDEILF ITTN (1) RIALTSSELLYI IVFA (29) INGCRVMDEILF ITTN (1) RIALTSSELLYI ISFVCYGVKNSYLIRF ITTA (1) RNGVRNSALLRP VSIG (1) IGCVENSKLICZ VUPC (13) DRIPFHHKWLEC VUPC (13) DRIPFHHKWLEC VUPC (13) DRIPFHHKWLEC VUPC (13) DRIPFHHKWLEC VUPC (13) DRIPFHHKWLEC VUPC (13) DRIPFHKWLEC IIIA (55)FSHIEK IILA	RQNLESFIRQDE(H HHHHHHHHHHHHHH HL(5)NFRLTLRAIKLMAKI EL(5)VFRIAIRAIKLMAKI EX(5)VFRIAIRAIKLMAKI EX(5)VFRALVFSVRCMARI HL(5)TFRLTLRAVKLMAKK HY(6)AARHTAMAVKAMKA AV(5)PCRWLSMSIKKANKA AV(5)PCRWLSMSIKKANKA KAI(6)FYCAYIHUKANKA SY(5)RLKPWILLVKHMAKI SY(5)RLKPWILLVKHMAKI SY(5)RLKPWILLVKHMAKI SY(5)RLKSLIRLVKHWYD EX(2)KKSLIRLVKHWYD DI(18)CRKECIKLMKYLLE (118)CRKECIKLMKYLE (118)CRKECIK (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE (118)CRKECIKLMKYLE	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GSI CLUFS A (13) NESLTMMVIF R (12) GSI CLUFS A (13) NESLTMMVIF R (12) GVSWAMLVAR A (12) GSVTVTWAIY Q (13) SGCFNLMVVY A (12) STVTTMALM R (12) SYVTWLVY Q (13) SGCFNLMVVY A (12) STVTTMALM R (12) SYVTWLY Q (13) SGCFLMVVY A (12) SYVTTAFFH Q (12) SYVKTAFFH Q (15) SYVKTAFFH Q (15) SYVKTAFFH Q (15) SYVKTAFFH Q (15) SYVKTAFFH Q (15) SYVKTAFFH Q (15) SYLVKTAFFH Q (15) SYLVKTAFFH Q (15) SYLVKTAFFH Q (15) SYLVKTAFFH Q (15) SYLVKTAFFH Q (15) SSLMAATVN A (8) STSLMAATVN A (8) STSLMAATVN	PAP/OAS1 HHH HHHH PQL(5)ASTLV [QL(5)ASTLV [QL(5)ASTLV [QL(6)LELLL PLQR(46)LELLL PLQR(46)LELLL [CQL(5)LETQV /LQE(48)VFFCL (LLV(3)LGRLL (LLV(3)LGRLL (LLV(3)LGRLL (LLV(3)LGRLL (LLV(3)LGRLL (LLV(3)LGRLL (LLV(3)LGRL (LLV(3	HHHH HHHHHHHHH HKFF (57) VSTRMVMVEER NRFF (57) ESTKKVILQEP VIEFF (58) – MIRDIKKAPA KEFF (24) QSQLQKPVDLAR HKFF (57) TSTRTVMVEER HKFF (57) PAKFQLWKQEL LDFL (65) PLKDFLRRHLE QRFA (95) ASRVRHIQQEN HGFF (61) SSGLYRIEGEM LETV (42) LDNLARFVHLCR LETV (42) LDNLARFVHLCR LETV (42) LDNLARFVHLCR UTV (53) FKSWCHAQEAR RTVL (53) FKSWCHAQEAR RTVL (53) FKSWFLSKNE DKLL (29) KSKEFLSKNE DKLL (29) KSKEFLSKNE DKLL (29) KSKEFLSKNE DKU (29) KSKEFLSKNE DKL (29) GPREMDIMSKLE KIIA (29) GPREMDIMSKLE	> HHHHHHH QGLAITD (165) RGVQITN (216) GAFDLIT (26) ESAWILQ (28) QGLAVTD (162) RAAQCME (30) KARDTAL (4) RAAQCME (30) KARDTAL (4) KARDTAL (
Active	Q5VYS8 Hs TUT7 1q78 A Bt PAP 2o1p A Sc PAP 3nyb A Sc PAP 3pq1 A Hs PAP 4lt6 A Hs PAP 2b4v A Tb TUTase 2lkf A Tb TUTase 2lkf A Tb TUTase 4e7x A Sp TUTase 4e7x A Sp TUTase 4e7x A Sc CAS 4ig8 A Hs CAS 4ig8 A Hs CAS 4k8v C Mm cGAS 4k8v C Mm cGAS 4k8v C GAS 4u07 A Vc CGAS 4u07 A Vc CGAS 4u07 A Vc CGAS 4u07 A Vc CGAS	(997) LNI LDQVGI EEEEEEEE EEE PVIKL-C-F(2) IEI PIIKF-V-E(4) IHI PIVRF-S-H(4) FQO PVVKF-E-F(2) IEI PVVKRF-A(110) YRM PVVQF-T-D(4) IHO PVVQF-T-D(4) IHO PVVG-V-V(2) VEM ALSF-V-L(8) VEP -ALSF-V-L(8) VEP -ALSF-V-L(8) VEP PAVTL-L-I(5) ISM PAVTL-L-I(5) ISM PAVTL-L-I(5) ISM PAVTL-L-I(3) ISM PAVTL-L-I(3) ISM TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI	SEEEE HHHHHHHHHH LERA(29)LNGCRVTDEILL LICA(29)LNGCRVTDEILL LICA(29)LNGCRVTDEILL VSEE(3)-GIEAAKLIRR USE(3)-GIEAAKLIRR USE(1)IGCVENSKILCP USFQ-1-GYGVKNSLIRR VSIG(1)IGGVENSKILCP VSIG(1)IRLAHMALLLSS VVEC(13)DRTPFHHKWLEG VLPA(37)-FTELORDFLK VLPA(37)-FTELORDFLK VLPA(37)-FTELORDFLK UILA(55)FSHIEKE IILA(55)FSHIEKE IILA(55)FSHIEKE IILA(55)FSHIEKE VILA(55)FSHIEKE VILA(55)FSHIEKE VILA(55)FSHIEKE VILA(55)FSHIEKE VILA(55)FSHIEKE VILA(55)FSHIEKE VILA(55)FSHIEKE VENY(27)PKIVEOWFNI GLIIT(23)LAAIRBARNFEE	RQNLESFIRQDE(H HHHHHHHHHHHH HL (5) NFRLTIRAIKLMAKI SU(5) UFRIALRAKLMAKI SU(5) UFRIALRAKLMAKI SU(5) SURALVISVRCMARI HL (5) TFRLTIRAVKLMAKI Y(6) ARHTMAVKLMAKI Y(6) ARHTMAVKLMAKI Y(5) RUKSURLVKANGKA AI (6) FYGAYI HLVKANGKA AI (6) FYGAYI HLVKANGKA SU(5) RUKSURLVKANGKA AI (6) FYGAYI HLVKANGKA AI (6) FYGAYI HLVKANGKA SU(5) RUKSURLVKANGKA AI (6) FYGAYI HLVKANGKA SU(5) RUKSURLVKANGKA SU(5) RUKSURLVKANGKA HI (6) CKECIKLMKYLLEG SI (18) CRKECIKLMKYLLEG SI (6) HLRKVCRFWKANGKA SI (6) HLRKVCRFWKANGKA SI (6) HLRKVCRFWKANGKA SI (6) HLRKVCRFWKANGKA SI (4) TVKVLI RLLKDLRII	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GS II CLVFS (12) GVSWAMLVAR (12) GS II CLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) SFVTMALM (12) SFVTMALM (12) SFVTMALM (12) SFVTTMALM (12) SFVTTMALM (13) SFVTTMALM (13) SFVTTMALM (14) SFVTTMALM (15) SFVTTMALM (12) SFVTTMALM (12) SFVTTMALM (13) STMATVN (13) SFVTTMALM (14) SFVTTMALM (15) SFVTTMALM (15) SFVTTMALM (15) SFVTTMALM (15) SFVTTMALM (12) SFVTTMALM	PAP/OAS1 HHH HHHHH ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ILU (34) LGRLL ILU (34) LGRLL ILU (34) LGRLL ILU (35) LGRQV ILU (35) LGRQV ILU (35) LGRQV MWER (8) TAQGF WWER (8) TAQGF ILU LSCCF ICTQ (11) LSCCF ICTQ (11) LSCCF ICTQ (11) LGLCF ILDS (7) LGETM WWN (8) LNVAY	HHH HHHHHHHH HKFF(57) VSTRWWVVEEK NRFF(57) SSTKKVILQEY HKFF(57) SSTKKVILQEY HKFF(57) TSTRVWVEEK HKFF(57) TSTRVWVEEK HKFF(57) TSTRVWVEEK HKFF(57) TSTRVMVEEK HKFF(57) SSGLYRIKGEM HKFF(57) SSGLYRIKGEM HKFF(57) SSGLYRIKGEM HKFF(57) SSGLYRIKGEM HKFF(53) THSWQRIAQEAR TVL(53) THSWRIA TVL(53) THSWRIA TVL(53) TVL TVL TVL TVL TVL TVL TVL TVL	> HHHHHHH QGLATD(165) 'RGVQITN(210) (GAFDLIT(26) (GAFDLIT(26) (CARAQCME(30) 'KARDCAL(4) (CARAQCME(30) 'KARDCAL(4) (CARAQCME(30) 'KARDCAL(4) 'SEFMEAP-(192) 'AASRLIN(43) 'EFMEAP-(200) 'WLGY-(20) 'AWLNY-(20) 'YERNNG-(7)
Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 2b4v_A Tb TUTase 2b4v_A Tb TUTase 3bj1_A Tb TUTase 3cj1_A Tb TUTase 3cj1_A Tb TUTase 3cj1_A Tb TUTase 3cj1_A Tb TUTase 3cj1_A Tb TUTase 4x4n_A Af CCA 1px5_A Sc CAS 4ig8_A Hs OAS 4ig8_A Hs OAS 4ig8_A Hs CGAS 4km5_A Hs CGAS 4km5_A Hs CGAS 4u03_A Vc CGAS 4u03_A Vc CGAS 4at7_B Mm NF45 4at7_B Mm NF45	(997) LNI LDQVGI NT EEEEEEEEEEE PVTKL-C-F(2) IEI PIKK-K-F(2) ISI PIKF-V-E(4) IHI PUKR-S-H(4) FQO PVKV-K-G(3) VDD PVVKV-K-G(3) VDD PVVRV-K-G(3) VDD PVVR-K-1(5) ISV PAVTL-L-I(5) ISVD PAVTL-L-I(3) ISVD PAVTL-L-I(3) ISVD PAVTL-L-I(3) ISVD PAVTL-L-I(3) ISVD PAVTL-L-I(4) THI TCGRI-K-I(4) THI TGFEI-S-S(2) ATVF	SCEADF (4) IEDUAREHEF IEEEE HHHHHHHHH ILEA (29) LNGCRVTDEILF LICA (29) LNGCRVTDEILF LICA (29) LNGCRVTDEILF LITN (1) RIALTSSELLYI LVFA (29) LNGCRVTDEILF ISFV CGVRNSVLIRF ITAY (1) RNGVRNSALLRP VSIG (1) IGGVENSKILCZ VUPC (13) DRTPFHHKWLEC VUPC (13) DRTPFHHKWLEC VUPC (13) DRTPFHHKWLEC VUPC (13) DRTPFHHKWLEC VUPC (13) DRTPFHHKWLEC VUPC (13) DRTPFHHKWLEC VUPC (13) DRTPFHHKWLEC IIIA (55) FSHIEKF IIIA (55) FSHIEKF IIIA (55) FSHIEKF IIIA (55) FSHIEKF VUPM (27)PKIVEDWFNI CHT (23) LAA INFERARWFEC IHLT (34) LASLRHARWFEC	RQNLESFIRQDE(H HHHHHHHHHHHHHH HL (5)NFRLTIRAIKLMAKI EL (5)VFRIAIRAIKLMAKI EL (5)VFRIAIRAIKLMAKI EV (5)GVRALVFSVRCMARJ HL (5)TFRLTIRAVKLMAKI Y (6)AAHHTMAVKAGKZ AY (5)FCRWLSMSIKRWSKQ AY (5)FCRWLSKU HL (18)CRECIKLMKYLEQ I (18)CRKECIKLMKYLEQ I (18)CRKECIKLMKYLEQ	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GSI CLUFS A (13) NSSLTMMVIF R (12) GSI CLUFS A (13) NSSLTMMVIF R (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) SVVTWFIY Q (13) SVGFNLMVVY A (12) SVVTWFIY R (12) SVVTWFIY R (12) SVVTWFIY R (12) SVVTWFIY R (12) SVVTWFIY R (12) SVVTWFIY R (12) SVVTWFIY Q (15) SVVTWFIFH Q (15) SVVTTAFFH Q (15) SVVTTAFFH Q (15) SVTWTAFFH Q (15) SVSTWTAFFH Q (15) SVSTWTTAFFH Q (16) SVSTWTTAFFH Q (17) CSTSTWTTAFFH Q (PAP/OAS1 HHH HHHH PQQL(5)ASTLV [QQL(5)ASTLV [QQL(5)ASTLV [QQL(5)ASTLV [QQL(6)LELLL PLQR(46)LELLL [CQL(5)ASTLV (LV(34)LGRLL (LLQ(35)LGTQV /LQE(48)VFFCL (LLQ(35)LGTQV /LQE(48)VFFCL (LLQ(35)LGTQV /LQE(48)VFFCL (LLQ(35)LGTQV /LQE(8)TAQGF WEQ(8)TAQGF WEQ(8)TAQGF WEQ(8)TAQGF WEQ(8)TAQGF (CTQ(11)LECCF (WTQ(11)LSSCF (CTQ(11)LSCF (CTQ(11)LSCF (T	HHH HHHHHHHH HKFF (57) VSTRMVMVEER NRFF (57) ESTKKVILQEP NRFF (57) ESTKKVILQE HKFF (57) TSTRTVMVEER KEFF (24) QSQLQKPVDLAR HKFF (57) TSTRTVMVEER LDFL (65) PLKRDFLRRHLE DRFA (95) ASRVRHLQER NGFF (67) SSGLYRIEGEM LETV (42) LDNLARFVHLQR NGFF (67) SSGLYRIEGEM LETV (42) LDNLARFVHLQR RTVL (53) PKGWRQLAQEAR TVL (53) PKGWRQLAQEAR TVL (53) PKGWRQLAQEAR TVL (53) PKGWRQLAQEAR TVL (53) PKGWRQLAQEAR DNCV (29) KRSKEFLSKIE DNCV (29) KRSKEFLSKIE DNCV (29) KRSKEFLSKIE DNCV (29) KRSKEFLSKIE DNCV (29) GPREMDIMSKLE KIIA (29) GPREMDIMSKLE KIIA (29) GPREMDIMSKLE RRCL (32) EQQDMVGYTAQT	> HHHHHHH QGLATD (165) RGVQITN (210) GAFDLIT (26) ESAWILQ (28) QGLAVTD (160) KARDTAL (4) RAAQCWE (30) KARDTAL (4) KARDTAL (4)
tive Active	Q5VYS8 Hs TUT7 1q78 A Bt PAP 201p A Sc PAP 3nyb A Sc PAP 3pq1 A Hs PAP 41t6 A Hs PAP 2b4v A Tb TUTase 30uy A Af CCA 4e7x A Sp TUTase 4e7x A Sp TUTase 4e7x A Sp TUTase 4e7x A Sc OAS 4ig8 A Hs OAS 4ig8 A Hs OAS 4ig8 A Hs CAS 4k8v C Mm cGAS 4k8v C Mm cGAS 4key A Mm cGAS 4u03 A Vc cGAS 4u03 A Vc cGAS 4u03 A Vc cGAS 4u7 A Mm NF45 4at7 B Mm NF90	(997) LNI LDQVGI EEEEEEE EEE PVIKL-C-F(2) ISI PIIKF-V-E(4) IHI PIKF-S-F(2) ISI PVKK-S-F(2) ISI PVKK-S-F(2) ISI PVKK-C-F(2) IEI PVKK-C-F(2) IEI PVKK-C-F(2) IEI PVK-C-F(2) IEI PVVQF-T-D(4) IHO PVKQ-V-V(2) VEW PVVQ-V-V(2) VEW PVHG-V-V(2) VEW PAVTL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(3) ISW TCGRI-K-I(4) THI TGGEI-S-S(2) ATW AATVT-K-N(5) ISIT	SEEEE HHHHHHHHH ILFA (29) LNGCRVTDEILL LICA (29) LNGCRVTDEILL LICA (29) LNGCRVTDEILL VSEE (3)GIEAAKLIRE LTTN (1) RIALTSSELLYI LVFA (29) LNGCRVTDEILL ISFVGYGVKNSVLIRE ISFVGYGVKNSVLIRE VSIG (1) IGGVENSKILCP USIG (1) RIGVENSKILCP VUEA (37) -FTELORDFIK VUEA (37) -FTELORDFIK VUEA (37) -FTELORDFIK VUEA (37) -FTELORDFIK VUEA (37) -FTELORDFIK ILLA (55) FSHIEKE ILLA (55) FSHIEKE ILLA (55) FSHIEKE ILLA (55) FSHIEKE VIEA (27) -PSHIEWE VIEA (27) -PSHIEWE VEMY (63) -PSUEDWFNI CLLT (23) LAAIRHARWFEE ILLG (42) SHENYLKYLY	RQNLESFIRQDE I HHHHHHHHHHHHH LL (5) NFRLTIRAIKLMACI SU (5) GURELWLIVCPLH SU (5) SURELWLIVCPLH SU (5) SURELWLIVCPLH H1 (5) NFRLTIRAIKLMACI SU (5) SURELWLIVCPLH H1 (5) NFRLTIRAIKLMACI H2 (5) RURALWFSWRCMAR H2 (5) SURSAIKANGSK A1 (6) FYGAYI HLVKANGKU Y (5) SUKEWLINKANKI SY (5) RUKEWLINKANKI SR (2) GKENEVRLIKGFLKI	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (12) GVSWAMLVAR (12) GSI ICLVFS (12) GVSWAMLVAR (12) SGUTNTWAI (12) SGUTNTWAI (12) SGUTNTMALM (12) SGUTNTMALM (12) SGUTNTMALM (12) SGUCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (15) SHVKTAFH (15) SHVKTAFH (15) SHVKTAFFH (15)	PAP/OAS1 HHH HHHHH ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ILQ (46) LELLL ICQL (5) ASTLV ILV (34) LGRLL ILV (34) LGRLL ILV (35) LGRQV ILQ (48) VFFCL ILI (35) LGRQV ILQ (48) VFFCL ILI LGCF WTQ (11) LSCF ICTQ (11) LSCF ICTQ (11) LSCF ICTQ (11) LGLCF WTQ (11) LSCF ICTQ (11) LGLCF ILDS (7) LGETM WMN (8) LNVAY SIGT (6) AGEAL LNG (13)SS	HHH HHHHHHHH HKFF(57) VSTRWWVVEEK NRFF(57) SSTKKVILQEFV LFF(58)NIRDIKKAPA KEFF(24) QSQLQKFVDLAR HKFF(57) TSTRTVMVEEK HKFF(57) TSTRTVMVEEK LOFI(65) PLKRDFLRRHLE QRFA(95) ASRVRHIQQEN LFTV(42) LDNLARFVHLQE TVL(53) THSWQRLAQEAR TVL(53) TTVL TVL(53) THSWQRLAQEAR TVL(53) TTVL TVL(53) TVL TVL(53) TVL TVL TVL TVL TVL TVL TVL TVL	> HHHHHHH QGLATD(165) 'RGVQITN(210) (GAFDLIT(26) (GAFDLIT(26) (CARAQCME(30) 'KARDCAL(4) (CARAQCME(30) 'KARDCAL(4) (CARAQCME(30) 'KARDCAL(4) 'LEFMEAP-(192) 'VHCRNAC-(20) 'AWLNY(20) 'YERNNAC-(7) 'YERNAC-(7) 'YERNAC-(7)'
active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78_A Bt PAP \\ 201p_A Sc PAP \\ 3nyb_A Sc PAP \\ 3nyb_A Sc PAP \\ 4lt6_A Hs PAP \\ 4lt6_A Hs PAP \\ 2b4v_A Tb TUTase \\ 2ikf_A Tb TUTase \\ 4c7x_A Sp TUTase \\ 4c7x_A Sp TUTase \\ 4c7x_A Sc OAS \\ 4ig8_A Hs OAS \\ 4kw_C Am cGAS \\ 4kw_C Am cGAS \\ 4kw_C Am cGAS \\ 4kw_A Hs CGAS \\ 4u01_A Vc cGAS \\ 4u01_A Vc cGAS \\ 4at7_A Mm NF45 \\ 4at7_B Mm NF90 \\ 4m5d_A Sc Utp22 \\ 4m5d_A Sc Utp2 \\ 4m5d_A$	(997) LNI DQVGI NT EEEEEEEE EEE PVTKL-C-F(2) ISI PIKK-V-F(2) ISI PIKF-V-E(4) IHI PVKR-S-H(4) FQO PVKR-S-H(4) FQO PVKR-C-(3) VDD PVVRV-K-G(3) VDD PVVR-V-V(2) VEV PVKL-V-V(2) VEV PVKL-V-V(2) VEV PVKL-V-V(2) VEV PVKL-V-V(2) VEV PVKL-V-V(3) VED PVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(4) THI TCGRI-K-I(4) THI TGFEI-S-S(2) ATV PIKNT-K-N(5) ISI PVTNT-L-T(3) YG	SCEADF (4) IEDUAREHE IEEEE HHHHHHHHH IEEA (29) LNGCRWDEILE LICA (29) LNGCRWDEILE LICA (29) LNGCRWDEILE LITN (1) RIALTSSELLYI LVFA (29) LNGCRWDEILE ISFV CYGVRNSVLIRF ITAY (1) RNGVRNSALLRP VSIG (1) IGGVENSKLICZ VGC (13) DRIPFHHKWLEG VUPC (13) DRIPFHHKWLEG VUPC (13) DRIPFHHKWLEG VUPC (13) DRIPFHHKWLEG VUPC (13) DRIPFHHKWLEG VUPC (13) DRIPFHKWLEG VUPC (13) DRIPFHKWLEG ILLA (55) FSHIEKE ILLA (55) FSHIEKE ILLA (55) FSHIEKE ILLA (55) FSHIEKE VIPM (27) FKIVEDWFNI VENY (27) FKIVEDWFNI VENY (23) FKIVEDWFNI VENY (23) FKIVEDWFNI VENY (23) FKIVEDWFNI VENY (23) FKIVENWFNI ILLT (24) LASLRHAKWFQP ILLG (42) STHENWLKYUPH	<pre>RQNLESFIRQDE(H HHHHHHHHHHHHHH HL(5)NFRLTIRAIKLMAKI EL(5)VFRIAIRAIKLMAKI EL(5)VFRIAIRAIKLMAKI EV(5)GVFRLVIFVRCVARJ HL(5)TFRLTIRAVKLMAKK Y(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSMSIKRWSK(AY(5)PCRWLSWSK(AV(5)PCRWLSWSK(AV(5)PCRWLSWSK(AV(5)PCRWLSWSK(AV(5)PCRWLSWSK(AV(5)PCRWSK(AV(5)P</pre>	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR A (12) GSI ICLVFS A (13) NFSLTMMVIF R (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) GVSVTWFIY Q (13) SYGFNLMVVY A (12) GVVTWFIY Q (13) SYGFNLMVVY A (12) GVVTWFIY A (12) GVLCELLIVF A (14) GVLCELLIVF A (15) SVVKTAFFH Q (15) SVVKTAFFH A (15) SISLMAATVN A (15) SISLMAATVN A (15) SISLMAATVN A (15) SISLMAATVN A (15) GVLCELLCEK A (17) FSFTLMAA A (15) FTLMAA A (15) FTLMAA	5) SULVCM T (PAP/OAS1 HHH HHHH PQQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV PLQR (46) LELLL PLQR (46) LELLL PLQR (46) LELLL (CQL (5) ASTLV (LLV (34) LGRLL (LLQ (35) LGTQV /LQE (48) VFFCL (LLQ (35) LGTQV /LQE (48) VFFCL (LLQ (35) LGRLV (LLQ (35) LGRLV (LLQ (35) LGRLV (LLQ (35) LGRLV WFR (8) TAQGF WFR (8) TAQGF WFR (8) TAQGF (CTQ (11) LGRLCF (MTQ (11) LGRLCF (MTQ (11) LGRLCF (TQ (11) LGRLCF (LDS (7) LGRLVN MMN (8) LGRLVN SIGT (6) AGEAL LNG (13)SS SVVD (9) LENGE	B)GLDCV-RTIEELAR(9) HHHH HKFF(57)ESTKKVILQEPK NRFF(57)ESTKKVILQEPK NRFF(57)ESTKKVILQEPK HKFF(57)TSTRTVMVEEFK HKFF(57)TSTRTVMVEEFK HKFF(57)TSTRTVMVEEFK HKFF(57)TSTRTVMVEEFK LDFL(55)PLKRDFLRRHDE DRFA(95)ASRVRHLQEFN LDFL(65)PLKRDFLRRHDE DRFA(95)ASRVRHLQEFK LETV(42)LDNLARFVHLCR LETV(42)LDNLARFVHLCR DNCV(29)KRSKEFLSKLE DNCV(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE RCL(31)RQQEDIMSKEFLSKE RCVL(31)RQCEDITOSAC YOLF(63)QILKEYAGETIR	> HHHHHHH QGLATD(165) RGVQITN(210) GAFDLIT(26) ESAWILQ(28) QGLAVTD(160) KARDTAL(4) KARDT
Inactive Active	Q5VYS8 Hs TUT7 1q78 A Bt PAP 201p A Sc PAP 3nyb A Sc PAP 3pq1 A Hs PAP 41t6 A Hs PAP 2b4v A Tb TUTase 2ikf A Tb TUTase 4c7x A Sp TUTase 4c7	(997) LNI DQVGI FEEEEEEE EEEE PVIKI-C-F(2) IEI PIIKF-V-E(4) IHI PIKF-S-F(2) IEI PVKKLRF-A(110) YKM PVKQF-C-F(2) IEI PVVKQF-C-G(3) VKM PVVQF-T-D(4) IHO PVKR-S-G(3) VKM PVVQF-T-D(4) IHO PVKQF-V-V(2) VEW PVVQF-V-V(2) VEW PVHG-V-V(2) VEW PAVL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(3) ISW TCGRI-K-I(4) THI TGGRI-K-I(4) THI TGGRI-K-I(4) THI TGFELS-S(1) FSIN VTLRT-C-C(1) FSIN VTLRT-L-T(3) YEF	SEEZE HHHHHHHHHH ILFA (29) LNGCRVTDEILL LICA (29) LNGCRVTDEILL LICA (29) LNGCRVTDEILL USEC (3)GIEAAKLIRE LITN (1) RIALTSSELLYI UVFA (29) LNGCRVTDEILL ISFVGYGVKNSVLIRE ISFVGYGVKNSVLIRE ISFVGYGVKNSVLIRE USIG (1) RIGVENSKILC2 USIG (2) STORTHKKLES USIG (2) STORTHKLES USIG (2) STORTHK	RQNLESFIRQDE(H HHHHHHHHHHHH HL (5) NFRLTIRAIKLMAKI SW (5) GURELWIIVKOPLH HL (5) NFRLTIRAIKLMAKI SW (5) GURELWIIVKOPLH HL (5) TFRLTIRAVKLMAKI HL (5) TFRLTIRAVKLMAKI HC (5) FCRWLSMSIKRMSK AY (5) PCRWLSMSIKRMSK AY (5) PCRWLSMSIKRMSK AY (5) PCRWLSWSIKRMSK AY (5) PCRWLSWSIKRMSK HI (6) PCRWLSWSIKRMSK HI (6) PCRWLSWSIKRMSK AY (5) PCRWLSWSIKRMSK HI (8) CRECIKLMSYLLEG I (8)	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) SSLTMWVIF (12) SGVTVTMALM (12) SGVTVTMALM (12) SGVTVTMALM (12) SGVTVTMALM (12) SGVTVTMALM (12) SGVVTVMFIF (11) QALELLTVY (16) GYLCELLIVF (11) QALELLTVY (10) QALELLTVY (10) QALELLTVY (10) QALELLTVY (10) GYALELLTVY (10) SHVKTAFFH (15) SYVKTAFFH (15) SYVKTAFFH (15) SYVKTAFFH (15) SYVKTAFFH (15) SYNKTAFFH (15)	PAP/OAS1 HHH HHHHH TCQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV TCQL (5) ASTLV TCQL (5) ASTLV ILV (34) LCRLL ILV (34) LCRLL ILV (34) LCRLL ILV (35) LCGVV ILV (34) LCRLL ILV (35) LCGVV ILV (35) LCGVV ILV (35) LCGVV WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF ILCCC ILL LCCCC ILCCCC ILL LCCCCC ILCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	HHH HHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) SSTKVILQEY HKFF(57) SSTKVILQEY HKFF(57) TSTKVWVEEK HKFF(57) TSTRVWVEEK HKFF(57) TSTRVWVEEK HKFF(57) TSTRVMVEEK LDFL(65) FLKRDFURHLE QRFA(95) ASRVRHLQEN HCFF(67) SSGLYRIRGEM HCFF(67) SSGLYRIRGEM LETV(42) LDNLAREVHLCR QTVL(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR RTVL(53) THSWQRLAQEAR RTVL(29) KSKEFLSKVE DNCV(29) KRSKEFLSKVE DNCV(29) KRSKEFLSKVE DNCV(29) KRSKEFLSKVE DNCV(29) KRSKEFLSKVE DNCV(29) CRSKEFLTKQTE KIIA(29) GPREMDLMSKE RRCL(32) EQQMVCYTAOT RRVL(31) RQQEADTOSAQ QLF(63) QILKEYAGETLR	> HHHHHHH QGLATD(165) 'RGVQITN(210) GGAFDLIT(26) (GAFDLIT(26) (CARAQCME(30) RAAQCME(30) (RAAQCME(30) (RA
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78 \\ A Bt PAP \\ 201p \\ A Sc PAP \\ 3nyb \\ A To TUTase \\ 21kf \\ A Tb TUTase \\ 21kf \\ A Tb TUTase \\ 427x \\ A Sp TUTase \\ A Sp$	(997) LNI LDQVGI PUTKI-C-F(2) IEI PUTKI-C-F(2) IEI PIKK-V-F(2) ISI PIKF-V-E(4) IHI PUVR-S-H(4) FQO PVVRV-F-F(2) IEI PVVRV-F-F(2) IEI PVVRV-F-F(2) IEI PVVRV-F-F(2) IEI PVVRU-F-F(2) IEI PVVRU-F-F(2) IEI PVVRU-F-F(2) IEI PVVRU-V-V(2) VEV PVVRU-V-V(2) VEV PVVRU-V-V(2) VEV PVVRU-V-V(2) ISV PVVRU-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV TCGRI-K-I(4) THI TCGRI-K-I(4) THI TGFEI-S-S(2) ATV AAIVI-K-N(5) LSI PILRI-S-C(1) FI	SCERDF(+) TEDUAREHER EEEEE HIHHHHHHHHH LLER(29) LNGTRVDEILE LLCA(29) LNGTRVDEILE LTTN(1) RTALTSSELLY1 LTTN(1) RTALTSSELLY1 LTTN(1) RTALTSSELLY1 LTAY(1) RNGVRNSALLRR VSEG(1) LGGVENSVLICR USEG(1) RGVENSVLIRF USEG(1) RATHNTLLSS VUPC(13) DRTPFHHKWLEG VUPC(13) DRTPFHHKWLEG VUPC(13) DRTPFHHKWLEG VUPC(13) DRTPFHSULLSS VIPA(37) -FTELQRDFLRN TILA(55) FSHIEKE ILLA(55) FSHIEKE ILLA(55) FSHIEKE VIPA(37) -PKIVEDWFNI VPMY(27)PKIVEDWFNI VPMY(27)PKIVEDWFNI VPMY(33)PKIVEDWFNI VPMY(34) LASLRHAKWFQA ILLIG(42) STHENKYLKYLW ILLIG(42) STHENKYLKYLW VLLG(42) STHENKYLKYLW VLLG(25) LRPAETA	RQNLESFIRQDE(H HHHHHHHHHHHHH HL (5)NFRLTIRAIKLMAKI EU (5)VFRIAIRAIKLMAKI EU (5)VFRIAIRAIKLMAKI EU (5)VFRIAIRAIKLMAKI HL (5)TFRLTIRAVKLMAKI Y (5)AVRALVFSVRCWARJ HL (5)TFRLTIRAVKLMAKI Y (5)PCRWLSMSIKRWSK AY (5)FCRWLSMSIKRWSK AY (18)CKECLKLMKYLLSC EI (18)CRKECLKLMKYLLSC EI (18)CRKECLKLMKYLLSC EI (18)CRKECLKLMKYLLSC EI (18)CRKECLKLMKYLLSC EI (18)CRKECLKLMKYLLSC EI (18)CRKECLKLMKYLLSC S (6)HLRKVCRFMSMRDJ S (6)HLRKVCRFMSMRDJ S (6)HLRKVCRFMSMRDJ S (5)SCVIVIRVIRDLCTI AY (5)SCVIVIRVIRCICTI AY (5)SCVIVIRVIRCICTI	3) KLSLFCSSKNG (HHHHHHHH R(12) GVSWAMLVAR R(12) GSI CLUFS A(12) GSI CLUFS A(12) GSI CLUFS A(12) GSI CLUFS A(12) GSUTVTWAIY R(12) SGVTVTWAIY R(12) SGVTVTWAIY R(12) SGVTVTWAIY R(12) SGVTVTWAIY R(12) SGVLCELLVF R(14) GYLCELLIVF R(14) GYLCELLIVF R(14) GYLCELLIVF R(14) GYLCELLIVF R(14) GYLCELLIVF R(14) GYLCELLIVF R(15) SYVKTAFFH R(15) SYVKTAFFH R(15) SYVKTAFFH R(15) SSTMAATVN R(15) SSTMAATVN	PAP/OAS1 HHH HHHH PQQL (5) ASTLV [CQL (5) ASTLV [CQL (5) ASTLV [CQL (5) ASTLV [CQL (6) LGVL PLQR (46) LELLL PLQR (46) LELLL [CQL (5) ASTLV (LU (34) LGRLL (LU (35) LGTQV /LQE (48) VFFCL (LU (35) LGTQV /LQE (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (11) LGLCF (CTQ (11) LGLCF LLDS (7) LGETM LDS (7) LGETM (10) LGC (7) LGETM LDS (7) LGETM (10) LGC (7) LGETM LDS (7) LGETM (10) LGC (7) LGC (7) LGETM (10) LGC (7) LGC (7) LGETM (10) LGC (7) LGC (B)GLDCV-RTIEELAR(9) HHHH HKFF(57)USTRMVMVEER NRFF(57)ESTKKVILGEV IEFF(58)-NIRDEKKAPA KEFF(24)QSQLQKEVDLAR HKFF(57)TSTRTVMVEER LDFL(55)PLKRDFLRRHDE DRFA(95)ASRVRHIQEEN LDFL(65)PLKRDFLRRHDE QRFA(95)ASRVRHIQEEN LETV(42)LDNLAREVHLOR LETV(42)LDNLAREVHLOR DNCV(29)KSKEFLSKQEA DNCV(29)KSKEFLSKVE DNCV(29)LOPREMDIMSKDE KIIA(29)GPREMDIMSKDE KIIA(29)GPREMDIMSKDE KIIA(29)LEVAGETING RRVL(31)RQ0REDITOSAQ QOLF(63)QILKEYAGETING RVL(30)PEEIDELGYT	> HHHHHHH QGLAITD(165) RGVQITN(210) GGAFDLIT(26) ESAWIIQ(210) GGAFULT(26) ESAWIIQ(210) KARDTAL(4) RAAQCME(30) KARDTAL(4) RAAQCME(30) KARDTAL(4) SEFMEAP-(192) EFMEAP-(192) EFMEAP-(20) AWINY(20) AWI
Inactive Active	Q5VYS8 Hs TUT7 $1q78_A$ Bt PAP $201p_A$ Sc PAP $3nyb_A$ Sc PAP $3nyb_A$ Sc PAP $3pq1_A$ Hs PAP $41t6_A$ Hs PAP $41t6_A$ Hs PAP $2b4v_A$ Tb TUTase $3hj1_A$ Tb TUTase $3ouy_A$ Af CCA $1px5_A$ Sc OAS $4ig8_A$ Hs CGAS $4km5_A$ Hs CGAS $4u03_A$ Vc CGAS $4u03_A$ Vc CGAS $4u03_A$ Vc CGAS $4at7_A$ Mm NF45 $4at7_B$ Mm NF45 $4at7_A$ Hs MiD51 $4w0y_A$ Mm MiD49	(997) LNI DQVGI FEEEEEEE EEE PVIKI-C-F(2) IEI PIIKT-K-F(2) IEI PIIKT-K-F(2) ISI PIKF-S-H(4) FQO PVKP-F-F(2) IEI PVKLRF-A(110) YRM PVVQF-T-D(4) IHO PVVQF-T-D(4) IHO PVVQF-T-D(4) IHO PVVQF-V-V(2) VEW PVVG-V-V(2) VEW PVHG-V-V(2) VEW PAVFL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(7) ISW PAVTL-L-I(7) ISW PAVTL-L-T(7) IST PIRI-S-C(17) FST VTINI-L-T(7) YEFT EEL-L-L(8) FTLF	SECENDE (4) IEDUAREHE SEEEE ILFA (29) LNGCRVTDEILE LICA (29) LNGCRVTDEILE UCA (29) LNGCRVTDEILE VSEE (3)GIEAAKLIRE UTFA (29) LNGCRVTDEILE USF (29) LNGCRVTDEILE ISFVGYGVKNSYLIRE UTA (1) RNGVRNSALLERE VSTG (1) IGGVENSKILCR IGFN (1) RLAIHMALLSS VVPC (3) DRTPFHHKWLEC VVPC (3) DRTPFHHKWLEC VUPA (37) -FRELORDFLRN ULA (35)FSHTEKY IILA (55)FSHTEKY IILA (55)FSHTEKY VPMY (33) -PKU-DØWFNI VPMY (33) -PKU-DØWFNI VPMY (33) -PKU-DØWFNI VILLG (42) STHENYLKYLKY ILLG (52) LRAKWF02 ILLG (52) LRAKWF02 ILLG (42) STHENYLKYLKY ILLG (42) STHENYLKYLKY ILLG (42) STHENYLKYLKY IVMY (27) LYPVETA	RQNLESFIRQDE(HHHHHHHHHHHHHHH HL(5)NFRLTLRAIKLMAKI EL(5)VFRIALRAIKLMAKI EW(5)GLRELVLIVKQPLH IV(5)VFRIALRAIKLMAKI HV(6)AARHTAMAVKAMKKI HV(6)AARHTAMAVKAMKKI HV(6)FYCAYIHLUKAMKA HV(6)FYCAYIHLUKAMKA SY(5)FRLKPMVLLVKHWAKI SY(5)RLKPMVLLVKHWAKI SY(5)RLKPMVLLVKHWAKI SY(5)RLKPMVLLVKHWAKI SY(5)RLKPMVLLVKHWAKI SY(5)RLKPMVLLVKHWAKI SY(5)RLKDIKLKKYLLK CHANG SY(5)RLKSLIRLVKHWYD SY(6)RLKLKKYLLK I(18)CRKCLKLMKYLLE SY(6)HLKVCRFWKARDJ SY(6)HLKVCRFWKARDJ SY(6)HLKVCRFWKARDJ SY(6)HLKVCRFWKARDJ SY(6)FYCALLKDLRLI SY(7)SFVEATVLGRLMLQ CT(5)SFVEATVLGRLMLQ CT(9)FYSPVRLEKRALD SY(10)CRSLCLKILKAICKS R(10)TRRLLLLCGTCR	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) STVTWAIW (12) STVTWAIW (12) STVTWAIW (12) STVTTWAIM (12) STVTTAIM (12) STTTAIM (12) STVTTAIM (12) STVTTAIM (12) STVTTAIM (12) STTTAIM (12) STTTAIM (13) STTTAIM (13) STTTAIM (14) STTTAIM (15) STTTAI	PAP/OAS1 HHH HHHH TCQL (5) ASTLV ICQL (5) ASTLV ITHM (14) LGVLL IT, HM (14) LGVLL IT, CQL (5) ASTLV ILV (34) LGRLL (1, 14) LGVLL (1, 14) LGSLL (2, 14) LGRL (1, 14) LGSL (2, 14) LGSCF (4, 14) LGSCF (4, 14) LGSCF (4, 14) LSSCF (4, 14) LSSCF (5, 14) LSSCF	HHH HHHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) SSTKKVILQEFV LFF(58)NIRDIKKAPA KEFF(24) QSQLQKEVDLAR HKFF(57) TSTRTVWVEEK HGFF(64) PAKFQLVKQEL LDFL(65) PLKRDFLRRHLE QFA(95) ASRVRHIQQEN HGFF(67) SSGLYRIRGEM HGFF(67) SSGLYRIRGEM HGFF(67) SSGLYRIRGEM HGFF(67) SSGLYRIRGEM HGFF(67) SSGLYRIRGEM HGFF(53) THSWQRLAQEAR RTVL(53) FKGWRQLAQEAR RTVL(53) THSWQRLAQEAR RTVL(53) HSWGRLAQEAR RTVL(53) HSWGRLAQEAR RTVL(29) RSKEFLSKNE DKL(29) RSKEFLSKNE DKL(29) RSKEFLSKNE DKL(29) RSKEFLSKNE DKL(29) GPREMDIMSKE RKIIA(29) GPREMDIMSKE RKIIA(29) GPREMDIMSKE RRVL(31) RQQREDITQSAQ RVU(31) RQQREDITQSAQ RVU(35) DFIATRITALAK KIPL(30) PEEIDELCYTY	> HHHHHHH QGLATD(165) GAFDLT(26) GAFDLT
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78 \\ A Bt PAP \\ 2o1p \\ A Sc PAP \\ 3nyb \\ A Sc PAP \\ 3nyb \\ A Sc PAP \\ 3nyd \\ A B PAP \\ 41t6 \\ A Hs PAP \\ 41t6 \\ A Hs PAP \\ 2b4v \\ A Tb TUTase \\ 3ny1 \\ A Tb TUTase \\ 3ny1 \\ A Tb TUTase \\ 47x \\ A St TUTase \\ 47x \\ A St TUTase \\ 47x \\ A St CCA \\ 4x4n \\ A A f CCA \\ 4x4n \\ A Af CCA \\ 4x4n \\ A Af CCA \\ 4x6 \\ A Sc CAS \\ 41g8 \\ A Hs CAS \\ A Hs CAS$	(997) LNI DQVGI NT EEEEEEEEEEEEE PVTKI-C-F(2) IEII PIKK-V-F(2) ISII PIKF-V-E(4) IHI PVWRIF-EF(2) ISII PVWRIF-EF(2) ISII PVWRUF-G(3) VDD PVVQF-T-D(4) IHO PVVQF-T-D(4) IHO PVVR-V-G(2) VED PVVR-V-V(2) VED PVVR-V-V(2) VED PVVR-V-U(2) VED PVVR-V-U(2) VED PVVR-V-U(2) VED PVVR-V-U(2) VED PVVR-V-U(2) VED PVVR-V-L(8) VED PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV TCGRI-K-I(4) THI TCGRI-K-I(4) THI TGFEI-S-S(2) ATV AAVTV-K-N(5) LSII PIRI-S-C(17) FSID PIRI-S-C(17) FSID PIRI-EEU(7) LFIT -EEEL-L-L(8) FTF	LEEEE HHHHHHHHHH LEEEE HHHHHHHHHH LLFA (29) LNGCRVTDEILE LLFA (29) LNGCRVTDEILE LUTA (29) LNGCRVTDEILE LYSE (3)GIEAAKLIRE LYTA (29) LNGCRVTDEILE LYTA (29) LNGCRVTDEILE LYTA (29) LNGCRVTDEILE USFE (3)GIEAAKLIRE USTA (1) RNGVRNSALLRE VSIG (1) IGGVENSKILCP USIG (1) RATHNALLSS VUPC (13) DRTPFHHKWLEC VUPC (13) DRTPFHHKWLEC VUPC (13) DRTPFHHKWLEC VIPA (37) -FTELQRDFLRN IILA (55)FSHIEK IILA (55)FSHIEK IILA (55)FSHIEK IILA (55)FSHIEK IILA (55)FSHIEK IILA (55)FSHIEK IILA (31) LASLARAWFDE YEMY (23) -ARIARWFE THLT (34) LASLARAWFE THLIG (42) STHENYLKYLY ERVL (30) TAKTLASVRHT FEPS (25)LRPATP YUMY (27)LYPETP	RQNLESFIRQDE(H HHHHHHHHHHHHH HL (5) NFRLTLRAIKLMAKI EU (5) VFRITLRAIKLMAKI EU (5) VFRITLRAIKLMAKI EU (5) VFRITLRAIKLMAKI EU (5) STRLTLRAKKMAKI HL (5) TFRLTLRAKKMAKI Y (5) ARHTMAVKANGKA AY (5) FCRWLSMSIKRWSKG AY (18) CRECLKLMSYLLEG AY (15) SCUIVIRURDLTT AY (5) SCUIVIRURDLTT AY (10) CRSLCLKIIKATCKS AY (10) TRRLLLLCGTCR	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GSI ICLVFS (12) GSI ICLVFS (13) NFSLTMWVIF (12) GSI ICLVFS (13) NFSLTMWVIF (12) SSUTMWVIF (12) SGVVTWFIY (13) SVGFNLMVVY (12) SGVVTWFIY (13) SVGFNLMVVY (12) SVGVLMVLY (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (15) SYHVKTAFFH (15) SGUTNTIFH (16) SGULMAATVN (8) SGULMAAT	PAP/OAS1 HHH HHHH PCQL (5) ASTLV ICQL (5) ASTLV ICQL (5) SAVIL TLQR (46) LELLL PCQR (46) LELLL PCQR (46) LELLL ICQL (5) ASTLV ILQC (48) VFCQ ILQC (48) VFCQ ILQC (48) VFCQ ILQC (48) VFCQ ILQC (48) VFCQ ILQC (48) VFCQ VFQ (11) LECCF VFQ (11) LECCF VFTQ (11) LECCF VFTQ (11) LECCF VFTQ (11) LECCF VFTQ (11) LECCF VFTQ (11) LECCF VFTQ (12) LEBT VMN (8) LNVAY SIGT (6) ACEAL LNGG (13)SS SFVD (9) LENGF LAQC (7)MLA LGEE (8) LGERF	HHHH HHHHHHHH HKFE (57) VSTRMYWVEER NRFF (57) VSTRMYWVEER NRFF (57) STRMYWVEER KEFF (24) QSQLQKFVDLAR KEFF (24) QSQLQKFVDLAR HKFF (57) TSTRTMVVEEK NGFF (64) PAKFQLWKOEL LDFL (65) PLKRDFLRCHLE QRFA (95) ASRVRHLQEFN LGFV (42) LDNLARFVHLQR TVU (53) THSWQRLAQEAR CVU (53) THSWQRLAQEAR RTVL (53) THSWQRLAQEAR RTVL (53) THSWQRLAQEAR DNCV (29) KRSKEFLSKNE DNCV (29) KRSKEFLSKNE DNCV (29) KRSKEFLSKNE DNCV (29) KRSKEFLSKNE DNCV (29) KRSKEFLSKNE DNCV (29) KRSKEFLTKQE KIIA (29) GPREMDIMSKLE KIIA (29) GPREMDIMSKLE KIIA (29) GPREMDIMSKLE KIIA (29) GPREMDIMSKLE KIIA (29) LFLATRATALAK DRFL (30) FEEIDELCYTY LQAL (27) EEEIDDIGYVY	> HHHHHHH QGLATD(165) RGVQITN(210) GGAFDLT(26) ESAWILQ(210) GGAFDLT(26) ESAWILQ(28) RAAQCME(30) KARDTAL(4) RLREMLI(19) RLREMLI(19) RLREMLI(19) RLREMLI(19) RLREMLI(19) RLREMC(7) YERNNG-(7)
Inactive Active	Q5VYS8 Hs TUT7 $1q78_A$ Bt PAP $201p_A$ Sc PAP $3nyb_A$ Sc PAP $3nyb_A$ Sc PAP $3pq1_A$ Hs PAP $41t6_A$ Hs PAP $41t6_A$ Tb TUTase $3hj1_A$ Tb TUTase $3ouy_A$ Af CCA $1px5_A$ Sc OAS $4ig8_A$ Hs OAS $4ig8_A$ Hs OAS $4ig8_A$ Hs CGAS $4kw1_A$ Af CCA $1px5_A$ Sc CGAS $4kw2_C$ Mm CGAS $4kw2_C$ Mm CGAS $4u03_A$ Vc CGAS $4u03_A$ Vc CGAS $4u03_A$ Vc CGAS $4u03_A$ Vc CGAS $4u03_A$ Vc CGAS $4u03_A$ Vc CGAS $4u7_A$ Mm NF45 $4a7_B$ Mm NF90 $4m5d_A$ Sc Utp22 $4m5d_A$ Sc Utp22 $4m5d_A$ Sc Mm $4woy_A$ Mm MiD49	(997) LNI LDQVGI PUTKL-C-F(2) IEI PUTKL-C-F(2) IEI PIKK-K-F(2) ISI PIKF-V-E(4) IHI PUVR-S-H(4) FOO PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-V-V(2) VEV PVVR0-V-V(2) VEV PVVR0-V-V(2) VEV PVVR0-V-V(2) VEV PVVR0-V-V(2) VEV PVVR0-V-V(2) VEV PVVR0-V-V(2) VEV PVVR0-V-V(2) VEV PVVR0-V-V(2) VEV PVVR0-V-V(3) VEV PVVR0-V-V(3) VEV PVVR0-V-V(3) VEV PVT1-L-I(5) ISV PAVT1-L-I(5) ISV PAVT1-L-I(7) ISV PAVT1-L-I(7) ISV PINI-L-T(7) YET VTINI-L-T(8) FTIF EEEEEEE EEE	SCENDF(+) IEDUAREHER IEEEE ILEA(29) INGCRVDEILF ILEA(29) INGCRVDEILF ILEA(29) INGCRVDEILF ILTN(1) RIALTSSELLYI ITYN(1) RIALTSSELLYI VIPC(13) DRIPFHHKWLEC VUPC(13) DRIPFHHKWLEC VIPA(37) - FIELORDFLKI ILLA (55)FSHTEKI VILA (55)FSHTEKI VPMY (63)PKIVEOWFNI VPMY (63)PKIVEOWFNI VILMY (27)LYNVEWYEN ILLI (42) STHENWLKYLYKI VPMY (23)LYNVEYN VILMY (27)LYNVETA	RQNLESFIRQDE HHHHHHHHHHHHHHHHHHHH HL (5) NFRLTLRAIKLMAKI SU(5) OVERIALRAIKLMAKI SU(5) OVERIALRAIKLMAKI SU(5) OVERIALRAIKLMAKI YU(5) SURALVFSVRCMARI YU(5) FORMLSMSIKKANGKA YU(5) FORMLSMSIKANGKA YU(5) FORMLSMSIKANGKA YU(5) FORMLSMSIKANGKA YU(2) FORMLSMAKANGKA YU(2) FORMLSMAKANGKA YU(18) CRECIKLMKYLLEG	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR R (12) GSI CLUFS A (13) NSSLTMMVIF R (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) GVSWAMLVAR A (12) GVVTWHIY Q (13) SGVTVTWHIY Q (12) SGVTVTWALM R (12) SGVVTWHIY Q (12) SGVVTWHIY Q (12) SGVVTWHIY Q (12) SGVVTWHIY Q (12) SGVVTWHIY Q (12) SGVVTWHIY Q (12) SGVVTAFFH Q (12) SGVVTTAFFH Q (12) SGVTTAFFH Q		8)GLDCV-RTIEELAR(9) HHHH HKFF(57)ESTKKVILQEEV NRFF(57)ESTKKVILQEV IRFF(58)-NIRGEN KEFF(24)QSQLQKFVDLAR HKFF(57)TSTRTVMVEEFK HKFF(57)PAKFQLWKQEL LDFL(55)PLKRDFLRRHLE QRFA(95)ASRVRHLQQEN RGF(64)PAKFQLWKQEL LDFL(65)PLKRDFLRRHLE QRFA(95)ASRVRHLQQEN RTVL(42)LDNLARFVHLQR QTVL(53)PKGWRQLAQEAR RTVL(53)PKSWFLSKNE DNCV(29)KRSKEFLSKNE DNCV(29)KRSKEFLSKNE DNCV(29)KRSKEFLSKNE DNCV(29)KRSKEFLSKNE DNCV(29)KRSKEFLSKNE DNCV(29)KRSKEFLSKNE DNCV(29)KRSKEFLSKNE DNCV(29)KRSKEFLSKNE NCV(29)KRSKEFLSKNE NCV(29)KRSKEFLSKNE DNCV(29)KRSKEFLSKNE NCV(29)KRSKEFLSKNE NCV(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE KIIA(29)GPREMDIMSKLE	> HHHHHHH QGLATD(165) GAPDLLT(26) GAPDLLT(26) GAPDLLT(26) GAPDLLT(26) GAPDLLT(26) GAPDLLT(26) GAPDLLT(26) GAPDLLT(26) GAPDLAC(40) G
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78 & A Bt PAP \\ 2c1p & A Sc PAP \\ 3nyb & A Sc PAP \\ 3nyb & A Sc PAP \\ 3nyb & A Sc PAP \\ 1t6 & A Hs PAP \\ 1t6 & A Hs PAP \\ 2b4v & A Tb TUTase \\ 3hj1 & A Tb TUTase \\ 3cuy & A Af CA \\ 4cxAn & A Af CCA \\ 4x4n & A Af CCA \\ 4x4n & A Af CCA \\ 4x6 & A Sp TUTase \\ 3cuy & A Af CA \\ 4x6 & A Sc Af A \\ 4x7 & A Mm NF90 \\ 4x7 & A Mm MiD49 \\ 0961P4 Hs FAM6A \\ 000 & A Mm CAB \\ 000 & A Mm Mid49 \\ 000 & A Mm Mage \\ 000 & A Mm Mid4 \\ 00 $	(997) LNI DQVGI NTT EEEEEEEE EEEE PVIKI-C-F(2) IEI PIKR-V-E(2) ISI PIKR-V-E(4) IHI PVKRF-S-H(4) FQO PVKRF-E-F(2) IEI PVWRF-FA(110) YRM PVVQC-T-D(4) IHO PVVRC-T-D(4) IHO PVVRC-T-D(4) IHO PVVRC-T-D(4) IHO PVVRC-V(2) VEW PYHG-V-V(2) VEW PYHG-V-V(2) VEW PYHG-V-U(8) VEP PAUTL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(7) FST VILNI-L-T(7) YGP ITL-E-V(7) FTF EEEEEEE EEE	SCERDF (4) IEDUAREHER SEEEE SHAHAHAHAHAHA SEEEE SEEEE SEEEE SEEEE SEGEA SEEEE SEGEA SEGEA SEEEE SEGEA VER (3) SEGEA VER (3) SEGEA SEGEA SEGEA SEGEA VER (3) SEGEA SEGEA SEGEA SEGEA SEGEA SEGEA SEGEA SEGEA SEGEA <th><pre>RQNLESFIRQDE(</pre></th> <th>3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GSI ICLVFS (12) GSI ICLVFS (13) NSSITMVIF (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSITMVVIF (12) SVVVMFIY (13) SVGFNLMVVY (13) SVGFNLMVVY (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (15) SYHVKTAFFH (15) SYHVKTAFFH (15) SYHVKTAFFH (15) SYHVKTAFFH (15) SYNKTAFFH (15) SYNKTAFFH (16) SISLMAATVN (17) TFFTILMAA (18) SSITNVILH (17) TFFTILMAA (17) TGFORMEN</th> <th>PAP/OAS1 HHH HHHH PCQL (5) ASTLV CCQL (5) ASTLV CCQL (5) ASTLV CCQL (5) ASTLV CCQL (5) ASTLV TCQR (46) LEILL CCQL (5) ASTLV (1) C(34) LGRLL (1) LQ (35) LGTOV (1) LQ (34) LGRLL (1) LQ (35) LGTOV (1) LQ (34) LGRLL (1) LG (3) LGTOF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (11) LGLCF (CTQ (11) LGLCF (CTQ (11) LGLCF (CTQ (11) LGCF WTQ (11) LGCF WTQ (11) LGCF WTQ (11) LGCF MVN (8) LNVAY SIGT (6) AGEAL LNG (13)SS FFVD (9) LENGF LAQC (7)MLA LGEE (8) LGEFF HHH HHH</th> <th>HHH HHHHHHHH HKFF (57) VSTRWWVEER NRFF (57) VSTRWWVEER NRFF (57) SSTKVILQEY HKFF (57) SSTKVILQEY HKFF (57) TSTRVWVER KEFF (24) QSQLQKFVDLAR HKFF (57) TSTRVWVER HKFF (57) TSTRVWVER HKFF (57) SSGLYRIRGEM LCTV (42) LDNLARFVHLQE DRF (67) SSGLYRIRGEM LETV (42) LDNLARFVHLQE DNCV (29) KSKEFLSK LETV (42) LDNLARFVHLQ DNCV (29) KSKEFLSK DNCV (29) KSKEFLSK DNCV (29) KSKEFLSK DNCV (29) KSKEFLSK EXIL (29) GPREMDIMSK KIIA (29) GPREMDIMSK KIIA (29) GPREMDIMSK KIIA (29) GPREMDIMSK KIIA (29) GPREMDIMSK KIIA (29) LPIATRITALAK NORL (30) PEEIDELGYTY LQAL (27) EEEIDDIGYUY HHHH HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</th> <th>> HHHHHH QGLATD(165) RGVQITN(210) GAFDLIT(26) GAFDLIT(26) GAFDLIT(26) RAAQCME(30) KARDTAL(4) RAAQCME(30) KARDTAL(4) RLREMLI(19) RAACME(43) KMRDTAL(43) KMRMC4(43) KMRMC4(43) KMRMC4(50) KMMC4(50) KMMC4(50)</th>	<pre>RQNLESFIRQDE(</pre>	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GSI ICLVFS (12) GSI ICLVFS (13) NSSITMVIF (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSITMVVIF (12) SVVVMFIY (13) SVGFNLMVVY (13) SVGFNLMVVY (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (15) SYHVKTAFFH (15) SYHVKTAFFH (15) SYHVKTAFFH (15) SYHVKTAFFH (15) SYNKTAFFH (15) SYNKTAFFH (16) SISLMAATVN (17) TFFTILMAA (18) SSITNVILH (17) TFFTILMAA (17) TGFORMEN	PAP/OAS1 HHH HHHH PCQL (5) ASTLV CCQL (5) ASTLV CCQL (5) ASTLV CCQL (5) ASTLV CCQL (5) ASTLV TCQR (46) LEILL CCQL (5) ASTLV (1) C(34) LGRLL (1) LQ (35) LGTOV (1) LQ (34) LGRLL (1) LQ (35) LGTOV (1) LQ (34) LGRLL (1) LG (3) LGTOF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF WEQ (11) LGLCF (CTQ (11) LGLCF (CTQ (11) LGLCF (CTQ (11) LGCF WTQ (11) LGCF WTQ (11) LGCF WTQ (11) LGCF MVN (8) LNVAY SIGT (6) AGEAL LNG (13)SS FFVD (9) LENGF LAQC (7)MLA LGEE (8) LGEFF HHH HHH	HHH HHHHHHHH HKFF (57) VSTRWWVEER NRFF (57) VSTRWWVEER NRFF (57) SSTKVILQEY HKFF (57) SSTKVILQEY HKFF (57) TSTRVWVER KEFF (24) QSQLQKFVDLAR HKFF (57) TSTRVWVER HKFF (57) TSTRVWVER HKFF (57) SSGLYRIRGEM LCTV (42) LDNLARFVHLQE DRF (67) SSGLYRIRGEM LETV (42) LDNLARFVHLQE DNCV (29) KSKEFLSK LETV (42) LDNLARFVHLQ DNCV (29) KSKEFLSK DNCV (29) KSKEFLSK DNCV (29) KSKEFLSK DNCV (29) KSKEFLSK EXIL (29) GPREMDIMSK KIIA (29) GPREMDIMSK KIIA (29) GPREMDIMSK KIIA (29) GPREMDIMSK KIIA (29) GPREMDIMSK KIIA (29) LPIATRITALAK NORL (30) PEEIDELGYTY LQAL (27) EEEIDDIGYUY HHHH HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	> HHHHHH QGLATD(165) RGVQITN(210) GAFDLIT(26) GAFDLIT(26) GAFDLIT(26) RAAQCME(30) KARDTAL(4) RAAQCME(30) KARDTAL(4) RLREMLI(19) RAACME(43) KMRDTAL(43) KMRMC4(43) KMRMC4(43) KMRMC4(50) KMMC4(50)
Inactive	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 2b4v_A Tb TUTase 2b4v_A Tb TUTase 3hj1_A Tb TUTase 3hj1_A Tb TUTase 3ouy_A Af CCA 4x4n_A Af CCA 1px5_A Sc OAS 4ig8_A Hs OAS 4j1x_A Ss CGAS 4kw_C Mm CGAS 4kw_C Ams CGAS 4kw_C Ams CGAS 4kw_C Ams CGAS 4u03_A Vc CGAS 4u5_B Mm NF45 4at7_B Mm NF45 4at7_A Bm NF45 4at7_A Hs MiD51 4woy_A Mm MiD49 Og6IP4 Hs FAM46A	(997) LNI LDQVGI NT EEEEEEEEEEEEE PVTKL-C-F(2) IEI PIKK-K-F(2) IEI PIKK-K-F(2) IEI PIKK-K-F(2) IEI PVKR-K-F(2) IEI PVKR-K-G(3) VDD PVVR-K-G(3) VDD PVVR-K-C(3) VDD PVVR-K-K-G(3) VDD PVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-K-M(5) ISI PIRT-S-C(17) FSI VTINI-L-T(3) YGF K-ICI-E-V(7) IFI EEL-L-L(8) FTLP EEEEEEE EEE PRWSL-I-S(6) KNW	SCEREDF (4) IEDUAREHEF IEEEE HHHHHHHHH ILEA (29) LNGCRVJDEILF ILGA (29) LNGCRVJDEILF ILGA (29) LNGCRVJDEILF ITN (1) RIALTSSELLYI LVFA (29) LNGCRVJDEILF ISFV CGVKNSVLIRF ITAY (1) RNGVRNSALLRP VSIG (1) IGGVENSKILCP VIG (1) RNGVRNSALLRP VIG (1) RNGVRNSALLRP VIG (1) RNGVRNSALLRP VIG (1) RNGVRNSALLRP VIG (1) GOFFHKWLCC VPC (13) DRNFHHKWLCC VPC (13) DRNFHHKWLCC VIC (13) DRNFHHKWLCC VIC (13) DRNFHHKWLCC VIC (13) DRNFHKWLCC VIC (14) LASLA VIC (14) LASLA VIC (14) LASLA VIC (14) CASLA VIC (14) CASLA VIC (14) COLLACYCLLVE HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	<pre>RQNLESFIRQDE(H HHHHHHHHHHHHHHHHHHHHHHHHHHHHH</pre>	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR R (12) GSI CLUFS A (13) NSSLTMMVIF R (12) GVSWAMLVAR A (12) GVSWAML A (12) GVSWAMLVAR A (12) GVSWAML A (12) GVS	PAP/OAS1 HHH HHHH PQQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (6) LELLL PLQR (46) LELLL [CQL (5) ASTLV (LQ (35) LGTOV /LQC (48) VFFCL (LLQ (35) LGTOV /LQC (48) VFFCL (LLQ (35) LGTOV /LQC (8) TAQGF WEQ (8) TAQGF WEQ (8) TAQGF (CTQ (11) LGCCF (WTQ (11) LSCCF (WTQ (11) LSCCF (WTQ (11) LSCCF (WTQ (11) LSCCF (CTQ (11) LGCCF (UTQ (11) LGCCF (UTQ (11) LGCCF (UTQ (11) LGCCF (UTQ (11) LGCCF (UTQ (11) LGCCF (UTQ (11) LSCCF (UTQ (11) LGCCF (UTQ (11) LG	HHH HHHHHHHH HKFF (57) VSTRMVMVEER NRFF (57) ESTKKVILQEP NRFF (57) ESTKKVILQEP HKFF (57) TSTRTVMVEER KEFF (24) QSQLQKPVDLAR HKFF (57) TSTRTVMVEER HKFF (57) TSTRTVMVEER LDFL (65) PLKRDFLRRHLE DRFA (95) ASRVRHLQEEN LDFL (42) LDNLARFVHLQE LETV (42) LDNLARFVHLQE RTVL (53) PKSWEFLSKIE DNCV (29) KSKEFLSKIE DNCV (29) KSKEFLSKIE DNCV (29) KSKEFLSKIE DNCV (29) KSKEFLSKIE DNCV (29) KSKEFLSKIE DNCV (29) KSKEFLSKIE DNCV (29) KSKEFLSKIE NCV (29) GPREMDIMSKLE KIIA (29) GPREM	> HHHHHHH QGLATD(165) RGVQITN(210) GAFDLIT(26) ESAWILQ(28) QGLAVTD(160) KARDTAL(4) KARDT
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78 & A Bt PAP \\ 2c1p & A Sc PAP \\ 3nyb & A Sc PAP \\ 3nyb & A Sc PAP \\ 3nyb & A Sc PAP \\ 1t6 & A Hs PAP \\ 1t6 & A Hs PAP \\ 2b4v & A Tb TUTase \\ 3hj1 & A Tb TUTase \\ 3cuy & A Af CCA \\ 4cxAn & A Af CCA \\ 4cxAn & A Af CCA \\ 4xdn & A Af CA \\ 4xdn & A Af \\ 4xdn &$	(997) LNI DQVGI NTT EEEEEEEE EEEE PVIKI-C-F(2) IEI PIKR-V-E(4) IHI PIKR-V-E(4) IHI PVKRF-S-H(4) FQO PVKRF-S-F(2) IEI PVKRF-S-F(2) IEI PVKRF-S-F(3) VEP PVKRF-S-F(3) VEP PVKRF-S-100 YEN PVVR-V-V(2) VEW PVVR-V-V(2) VEW PVHG-V-V(2) VEW PVHG-V-V(2) VEW PVHG-V-V(2) VEW PVHG-V-V(2) VEW PVHG-V-V(2) VEW PVTL-L-I(3) ISV PAVTL-L-I(3) ISV TCGRI-K-I(4) THI TCGRI-K-I(4) THI PILRI-S-S(2) ATVK AAIVT-K-N(5) LSTI PILRI-S-S(17) FSI VILNI-L-T(3) YGP UTL-E-V(7) FSI VILNI-S(6) KNVE PRWSL-I-S(6) KNVE	SCEREF(+) TEDUAREHER SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEGE SEEGE SEEGE SEEGE SEEGE SEEGE SEEGE SEEGE SEGEGE SEEGE SEGEGE SEGEGE SEGEGE VEQUER SIGE VIPC (13) DENPEHHKWLEG VUPC (13) DENPEHHKWLEG VUPC (13) DENPEHHKWLEG VUPA (37) - FELORDELKE SILLA (55) FSHIEKE SILLA (55) FSHIEKE SILLA (55) FSHIEKE VPMY (63) - PEX'VEOWFNIC SILLG (2) SHEMALKYLY SEEE SHEMK HAWEPEP VILMU (27)	RQNLESFIRQDE HHHHHHHHHHHHHHHHH HHHHHHHHHHHHHH HL (5) NFRLTIRAIKLMACI SW (5) GURELVLIVKOPLH SW (5) STRLEVNIVKOPLH IY (5) NFRLTIRAIKLMACI IY (5) NFRLTIRAIKLMACI IY (5) NFRLTIRAIKLMAKI IY (5) NFRLTIRAIKLMAKI IY (5) STRLTIRAVKLMAKI SY (5) SURMANAIKANGKA AY (5) PORWLSMSIKRNSKA AY (5) PORWLSMSIKRNSKA SY (5) RUKPWULUKAHAKI SY (2) SURNEVRLIKGFIKA SY (2) SURNEVRLIKGFIKA SY (2) SURNEVRIKGEN AY (18) CRKECIKLMKYLLEG YI (18) CRKECIKLMKYLLEG SY (5) SURVULKKARDI SY (5) SURVULKLER SY (10) CRSLCIKLIKAYLLEG SY (10) CRSLCIKLIKAYLLEG SY (10) CRSLCIKLIKARDI SY (10)	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GS I CLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) GS I CLVFS (13) NSSLTMWVIF (12) SSUTWIVAR (12) SSUTWIVAR (12) SGVTWINFIY (13) SGCFNLTWVY (13) SGCFNLTVY (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (15) SHVKTAFFH (15) SUMATVNN (8) SILMAATVNN (8) SULMAATVNN (8) SULMAATVNN (8) SULMAATVNN (8) SULMAATVNN (8) SULMAATVNN (9) SULMAATVNN (17) TEFTILMAA (17) TEFTILMAA (17) UFEORTTLER (17) COQURTESS (17) UFEORTTLER	PAP/OAS1 HHH HHHHH ICQL (5) ASTLV ICQL (46) VFICL ICQL (46) VFICL ICQL (48) VFICL ICQL (48) VFICL ICL (48) VFICL ICC (48) VFICL IC	HHH HHHHHHHH HKFF(57) VSTRWWVEER NRFF(57) SSTKVILQEY LFF(58)NIRDIKKAPA KEFF(24) QSQLQKFVDLAR KEFF(24) QSQLQKFVDLAR HKFF(57) TSTRTVWVEER KFF(57) TSTRTVWVEER LOFI(65) PLKRDFLRRHLE QRFA(95) ASRVRHIQQEYN HCFF(67) SSGLYRIRGEM LETV(42) LDNLARFVHLQR UTV(42) LDNLARFVHLQR TVV(42) LDNLARFVHLQR TVV(53) FKGRQLQOEA LETV(42) LDNLARFVHLQR TVV(53) FKGRQLQOEA DNCV(29) KRSKEFLSKQTE DKLL(29) RKSKEFLSKQTE DKLL(29) RKSKEFLSKXE DNCV(29) KRSKEFLSKXE DNCV(29) KRSKEFLSKXE DNCV(29) KRSKEFLSKXE DNCV(29) KRSKEFLSKXE DNCV(29) KRSKEFLSKXE DNCV(29) KRSKEFLSKXE DNCV(29) KRSKEFLSKXE DNCV(29) KRSKEFLSKXE DNCV(29) KRSKEFLSKZE DNCV(29) KRSKEFLSKZE KIL (29) EFL KIL (29) EFL KIL (29) KRSKEFLSKZE KIL (29) EFL KIL (29) KRSKEFLSKZE KIL (29)	> HHHHHHH QGLATD(165) RGVQITN(210) GGAPDLT(26) GGAPDLT(26) (QGLAVTD(160) RAAQCME(30) (XARDTAL(4)) RLREMII(19) IAASRLIN(43) LEFMEAP-(192) LEFMEAP-(192) LEFMEAP-(208) VWLCY(20) AWLNY(20) AWLNY(20) AWLNY(20) AWLNY(20) IVERNNG-(7) YERNNG-(7)
Inactive Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 2b4v_A Tb TUTase 2b4v_A Tb TUTase 3hj1_A Tb TUTase 3hj1_A Tb TUTase 30uy_A Af CCA 4x4n_A Af CCA 1px5_A Sc OAS 4ig8_A Hs OAS 4j1x_A Ss CGAS 4k8v_C Mm CGAS 4k8v_C Mm CGAS 4k8v_C Ams CGAS 4k03_A Vc CGAS 4u03_A Vc CGAS 4u03_A Vc CGAS 4u03_A Vc CGAS 4u03_A Vc CGAS 4u03_A Vc CGAS 4u17_A Mm NF45 4at7_B Mm NF90 4m5d_A Sc Utp22 4msd_A Sc Utp22 4msd_A Sm46B Q96IP4 Hs FAM46B Q5VWP2 Hs FAM46B	(997) LNI LDQVGI PUTKL-C-F(2) IEI PUTKL-C-F(2) IEI PIKK-K-F(2) ISI PIKF-V-E(4) IHI PUVR-S-H(4) FQO PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-L-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(5) ISV PVVRV-K-I(5) ISV PVVRV-K-	SCERDF (4) IEDUAREHER SEEEE STEPA SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEET SEET SEET SEET SEET SEET SEET SEET SET	RQNLESFIRQDE H HHHHHHHHHHHHHHHHHHHH L(5)NFRLTIRAIKLMAKI EU(5)GURELVLIVGVCHARI IY(5)RVRALVFSVRCMARI IY(18)CRECLKLMKYLLEG IY(18)CRECLKLMKYLEG IY(19)FYPVRLEKRKRUD	3) KLSLFCSSKNG (H HHHHHHH R (12) GVSWAMLVAR R (12) GVSWAMLVAR R (12) GSI CLUFS A (13) NFSLTMMVIF R (12) GVSWAMLVAR A (12) GVSWAML A (12)	5) SULVCM T(PAP/OAS1 HHH HHHH PCQL(5) ASTLV ICQL(5) SAVIL ICQL(5) SAVIL PLQR(46) LELLL PLQR(46) LELLL PLQR(46) LELLL ICQL(5) ASTLV ILQ(35) LGTQV ILQE(48) VFFCL (ILV(34) LGSLL (ILV(34) LGSLL (ILV(34) LGSLL ICQC(11) LGSLF WER(8) TAQGF WER(8) TAQGF WER(8) TAQGF WER(8) TAQGF WER(8) TAQGF ILDS(7) LGETM ILDS(7)	HHH HHHHHHHH HKFF (57) VSTRMVMVEER NRFF (57) ESTKKVIJQEP NRFF (57) ESTKKVIJQEP HKFF (57) TSTRTVMVEER KEFF (24) QSQLQKPVDLAR HKFF (57) TSTRTVMVEER HKFF (57) TSTRTVMVEER LDFL (55) PLKRDFLRRHDE DRFA (95) ASRVRHIQQEN LGFV (42) LDNLARFVHLQR NGFF (57) SSGLYRIRGEM LETV (42) LDNLARFVHLQR RTVL (53) PKSWEFLSKLE DNCV (29) KRSKEFLSKLE DNCV (29) GPREMDIMSKLE KIIA (29) GP	> HHHHHHH QGLATD(165) RGVQITN(210) GAFDLIT(26) ESAWILQ(28) QGLAVTD(160) KARDTAL(4) RAAQCME(30) KARDTAL(4) KARDTAL(4
Inactive Active	Q5VYS8 Hs TUT7 $1q78_A$ Bt PAP $201p_A$ Sc PAP $3nyb_A$ Sc PAP $3pq1_A$ Hs PAP $41t6_A$ Hs PAP $2b4v_A$ Tb TUTase $2ikf_A$ Tb TUTase $30uy_A$ Af CCA $4e7x_A$ Sp TUTase $4e7x_A$ Sp TUTase $4e7x_A$ Sc TUTase $4e7x_A$ Sc TUTase $4e7x_A$ Sc CAS $4ig8_A$ Hs CAS $4ig8_A$ Hs CAS $4k8v_C$ Mm cGAS $4k8v_C$ Mm cGAS $4k8v_C$ Mm cGAS $4k0n_A$ Vc cGAS $4u03_A$ Vc cGAS $4u547_B$ Mm NF90 $4m5d_A$ Sc Utp22 $4m5d_A$ Sc Utp22 $4mxt_A$ Hs MiD51 $4woy_A$ Mm MiD49 Q96IP4 Hs FAM46A Q96A09 Hs FAM46B Q5VWP2 Hs FAM46C Q8NER8 Hs FAM46B	(997) LNI DQVGI NTT EEEEEEEE EEEE PVIKL-C-F(2) IEI PIIKI-K-F(2) ISI PIIKF-V-E(4) IHI PLVRF-S-H(4) FQO PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-E-F(2) IEI PVVR-V-V(2) VEW PVVR-V-V(2) VEW PVVR-U-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(5) ISW PAVTL-L-I(3) ISW PAVTL-L-I(4) ISW PAVTL-L	SEEEE HHHHHHHHHH LEFA (29) LNGCRVTDEILL LICA (29) LNGCRVTDEILL LICA (29) LNGCRVTDEILL LICA (29) LNGCRVTDEILL LTYR (29) LNGCRVTDEILL LTYN (1) RTALTSSELLYI LYFA (29) LNGCRVTDEILL LTYN (1) RTALTSSELLYI LTYN (1) RTALTSSELLYI USFE (1) IGGVENSKILCP USTG (1) RGGVENSKILCP USTG (1) RGVENSKILCP VUCA (37) -FTELORDFIKN VUPC (3) DETPFHHKWLEC ULA (55) FSHTEKY IILA (55) FSHTEKY IILG (42) STHEW/LANEWFEH HULG (42) STH	RQNLESFIRQDE HHHHHHHHHHHHHHHH L(5)NFRLTIRAIKLMACI SU(5)GLEEWLIVKOPLH HL(5)NFRLTIRAIKLMACI SU(5)GLEWLIVKOPLH HL(5)NFRLTIRAIKLMACI SU(5)GLEWLIVKOPLH HL(5)NFRLTIRAIKLMACI SU(5)GLEWLIVKOPLH HL(5)TFRLTIRAVKLMAKI Y(6)ARHTMAVKANGKA AY(5)PCRWLSMSIRRNSK AI(6)FYGAYIHLWANGKA Y(5)RIKPWULLVKHNAKI SR(2)GKENEVRLIKGFLKA Y(2)GKENEVRLIKGFLKA Y(16)CKECIKLMKYLLEG Y(18)CKECIKLMKYLLEG Y(18)CKECIKLMKYLLEG Y(18)CKECIKLMKYLLEG S(6)HLRKVCRFWKANRDD S(6)HLRKVCRFWKANRDD S(6)HLRKVCRFWKANRDD S(6)HLRKVCRFWKANRDD S(6)HLRKVCRFWKANRDD S(6)HLRKVCRFWKANRDD S(7)SYEDATULRLIND S(7)SYEDATULRLIND S(7)SYEDATURE S(7)SYEDATURE S(10)CRSLCKLINACIKNON S(4)TYRNCSI S(7)SYEDAULALGRYMCSI S(7)SYEDAULALGRYMCSI S(7)SYEDAULALGRYMCSI S(7)SYEDAULALGRYMCSI S(7)SYEDAULALGRYMCSI	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (12) GVSWAMLVAR (12) GSI ICLVFS (12) GVSWAMLVAR (12) SFUTINALM (12) SFUTINA	5) SULVCM T(PAP/OAS1 HHH HHHHH ICQL(5) ASTLV ICQL(5) ASTLV ICQL(5) ASTLV ICQL(5) ASTLV ICQL(5) ASTLV ILU(34) LGRLL ICQL(35) LGRQV ILU(34) LGRLL ILU(35) LGRQV ILU(35) LGRQV ILU(35) LGRQV ILU(35) LGRQV WER(8) TAQGF WER(8) TAQGF WER(8) TAQGF WER(8) TAQGF WER(8) TAQGF ILU(1) LSCF ICTQ(11) LSCF ILDS(7) LGETM WMN(8) LNVAY SIGT(6) AGEAL LGEE(8) LGERF HHH HHHHH ILAN(8) KYDYI ILNN(8) KYDYI	HHH HHHHHHHH HKFF(57) VSTRWWVEEK NRFF(57) SSTKVILQEY HKFF(57) SSTKVILQEY KKFF(24) QSQLQKFVDLRA KKFF(24) QSQLQKFVDLRA HKFF(57) TSTRTVWVEEK HKFF(57) TSTRTVWVEEK HKFF(57) SSGLYRIRGEM LFTV(42) LDNLARFVHCQEN LFTV(42) LDNLARFVHCQEN TVL(53) THSWQRLAQEAR TVL(53) THSWQRLAQEAR TVL(29) KRSKEFLSKIE DNCV(29) KRSKEFLSKIE DNCV(29) KRSKEFLSKIE DNCV(29) KRSKEFLSKIE TVL(29) GPREMDIMSKE RRCL(32) EQQDMVGYTAGT RRVL(31) RQCREDTOSAQ VLF(63) QILKEYAGETLR LKVL(95) LPIATRITALAR DRFL(30) PEEIDDGGYTY HHH HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	> HHHHHHH QGLATD(165) 'RGVQITN(210) (GAFDLIT(26) (GAFDLIT(26) (CAPDLIT(26) (CA
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78_A Bt PAP \\ 201p_A Sc PAP \\ 3nyb_A Sc PAP \\ 3nyb_A Sc PAP \\ 3nyb_A Sc PAP \\ 1t6_A Hs PAP \\ 41t6_A Hs PAP \\ 41t6_A Tb TUTase \\ 21kf_A Tb TUTase \\ 427x_A Sp TUTase \\ 427x_A Sc OAS \\ 41g8_A HS OAS \\ 41g8_A AS CAS \\ 41g8_A AS \\ 41$	(997) LNI LDQVGI PUTKI-C-F(2) IEI PUTKI-C-F(2) IEI PITKT-K-F(2) ISI PITKF-V-E(4) IHI PUVR-S-H(4) FQO PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVR-V-V(2) VEV PVRV-V-V(2) VEV PVRV-V-V(2) VEV PVR-V-V(2) VEV PVR-V-V-V(2) VEV PVR-V-V-V(2) VEV PVR-V-V-V(2) VEV PVR-V-V-V(2) VEV PVR-V-V-V(2) VEV PVR-V-V-V-V(2) VEV PVR-V-V-V-V(2) VEV PVR-V-V-V-V(2) VEV PVR-V-V-V-V(2) VEV PVR-V-V-V-V(2) VEV PVR-V-V-V-V-V(2) VEV PVR-V-V-V-V-V(2) VEV PVR-V-V-V-V(2) VEV PVR-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V-	SCERDF (4) IEDUAREHER SEEEE SHHHHHHHHHH SEEEE SEEEE SEETE SEETE SETT SEETE SETT	<pre>RQNLESFIRQDE</pre>	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVSS (13) NFSLTMMVIF (12) GSI ICLVSS (13) NFSLTMMVIF (12) GVSWAMLVAR (12) GVVTWFIY (13) SYGFNLMVVY (12) SFVTTMALM (12) SGVVTWFIY (13) SYGFNLMVVY (12) SFVTTMALM (12) SGVVLWLYF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (15) SYVKTAFFH (15) SYNKTAFFH (15) SYVKTAFFH (15) SYVKTAFFH (15) SYNKTAFFH (15)	PAP/OAS1 HHH HHHH PQQL (5) ASTLV [CQL (5) ASTLV [CQL (5) ASTLV [CQL (5) ASTLV [CQL (5) ASTLV [CQL (6) LGULL PLQR (46) LELLL [CQL (5) ASTLV [CQL (6) ASTLV [CQL (1) LGSLF [CQL (6) AGEAL LNG (13)SS [CQL (7)MLA GEEE (8) LGERF HHH HHHH [CQN (8) KYEYL [LEA (9) RYACL [CQN (8) KYDYL [LHN (8) KY	B)GLDCV-RTIEELAR(9) HHHH HKFF(57)ESTKKVILGEP NRFF(57)ESTKKVILGEP NRFF(57)ESTKKVILGEP LDFF(58)-NIRDEKAPA KEFF(24)QSQLQKEVDLAR HKFF(57)TSTRTVMVEEF LDFL(55)PLKRDFLRHLDE DRFA(95)ASRVHLQEEN LDFL(65)PLKRDFLRHLDE QRFA(95)ASRVHLQEEN LETV(42)LDNLAREVHLQR LETV(42)LDNLAREVHLQEA CVL(53)PKGWRQLAQEAR DNCV(29)KRSKEFLSKVE DNCV(29)KRSKEFLSKVE DNCV(29)KRSKEFLSKVE DNCV(29)KRSKEFLSKVE DNCV(29)KRSKEFLSKVE DNCV(29)KRSKEFLSKVE DNCV(29)KRSKEFLSKVE DNCV(29)KRSKEFLSKVE DNCV(29)LRSKEFLSKVE DNCV(29)LRSKEFLSKVE DNCV(29)LRSKEFLSKVE DNCV(29)LRSKEFLSKVE DNCV(29)LRSKEFLSKVE DNCV(29)LARSKEFLSKVE	> HHHHHHH QGLATD(165) RGVQITN(210) GAFDLIT(26) ESAWIIQ(28) QGLAVID(160) RAAQCME(30) KARDTAL(4) RAAQCME(30) KARDTAL(4) RLREHI(1(192) EFMEAP-(192) EFMEAP-(192) EFMEAP-(20) AWINY(20) AWINY(20
Inactive Active	Q5VYS8 Hs TUT7 $1q78_A$ Bt PAP $201p_A$ Sc PAP $3nyb_A$ Sc PAP $3pq1_A$ Hs PAP $41t6_A$ Hs PAP $2b4v_A$ Tb TUTase $2ikf_A$ Tb TUTase $30uy_A$ Af CCA $4e7x_A$ Sp TUTase $4e7x_A$ Sp TUTase $4e7x_A$ Sp TUTase $4e7x_A$ Sc OAS $4ig8_A$ Hs OAS $4ig8_A$ Hs OAS $4ig8_A$ Hs CAS $4k8v_C$ Mm cGAS $4k8v_C$ Mm cGAS $4k60_A$ Hs cGAS $4u03_A$ Vc cGAS $4u07_A$ Vc cGAS $4u07_A$ Vc cGAS $4u07_A$ Vc cGAS $4u07_A$ Vc cGAS $4u07_A$ Vc cGAS $4u7_B$ Mm NF90 $4m5d_A$ Sc Utp22 $4mxt_A$ Hs M1D51 $4my_A$ Mm MiD49 Q96IP4 Hs FAM46B Q5VWP2 Hs FAM46B Q8NEV8 Hs FAM46B Q8NVV4 Hs TUT1 Q6PUY H= TUT2	(997) LNI DQVGI NTT EEEEEEEE EEEE PVIKL-C-F(2) IEI PIKR-V-E(4) IHI PLVRF-S-H(4) FQO PVIKF-E-F(2) IEI PVVRF-E-F(2) IEI PVVRF-AC(10) YRW PVVRF-AC(10) YRW PVVRF-AC(10) YRW PVVRC-V-C(2) IEI PVVR-V-C(2) VEW PVVRC-V-V(2) VEW PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V PVVRC-V-V	SEEZE HHHHHHHHHH LIFA (29) LNGCRVDDELL LIFA (29) LNGCRVDDELL LICA (29) LNGCRVDDELL LICA (29) LNGCRVDELL LICA (29) LNGCRVDELL LITN (1) RIALTSSELLY LIFY (29) LNGCRVDELL LIFY (1) RIALTSSELLY LIFY (1) RIGCRVDELL SUPF (29) LNGCRVDELL IGFN (1) RIGVENSALLRS VUFA (29) LNGCRVDELL UIFX (1) RIGVENSALLRS VUFA (37) -FTELORDFLK VUFA (37) -FTELORDFLK VUFA (37) -FTELORDFLK IILA (55)FSHTEK VPMY (3) -PK'VEOWFNI IILG (22) STHEN/LKYLKY VPMY (30) TAKYLASVENT FVLG	<pre>RQNLESFIRQDE(H HHHHHHHHHHHHH HL(5)NFRLTIRAIKLMAKI SV(5)UFRLAIRAIKLMAKI SV(5)UFRLAIRAIKLMAKI VS)SVFRALVFSVRCMARI HL(5)TFRLTIRAVKLMKKI HL(5)TFRLTIRAVKLMKKI VS)SVFRALVFSVRCMARI KAY(5)PCRWLSWKIKKMKK AI(6)FYGAYHLVKAMGKX AI(6)FYGAYHLVKAMGKX AI(6)FYGAYHLVKAMGKX AI(6)FYGAYHLVKAMGKX AI(6)FYGAYHLVKAMGKX AI(6)FYGAYHLVKAMGKX AI(6)FYGAYHLVKAMGKX AI(7)CKCLKLMKYLLEG I(18)CKKCLKLMKYLLEG I(18)CKECLKLMKYLMKYLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYLLEG I(18)CKECLKLMKYL</pre>	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (12) GVSWAMLVAR (12) GSI ICLVFS (12) GVSWAMLVAR (12) SGI TWINAR (12) SGI TWINAR (12) SGI TWITMALM (12) SGI TWITMALM (12) SGI CELLIVF (11) GYLCELLIVF (11) GYLCELLIVF (12) SYHVKTAFFH (15) SYHVKTAFFH (12) SYHVKTAFFH (13) SIMAATVN (13) GSLMAATVN (13) GYLCELLIK (13) GYLCELLIK (13) GYLCELLIK (14) GYLCELLIK (15) SYHVKTAFFH (15) SYHVKTAFFH (12) GYCKTES (11) GYCKTES	PAP/OAS1 HHH HHHHH TCQL (5) ASTLV TCQL (5) ASTLV TCQL (5) ASTLV TCQL (5) ASTLV TCQL (5) ASTLV TCQL (5) ASTLV TLW (14) LGVLL TCQL (46) LELLL (11) (43) LGSLL (11) (43) LGSLL (11) (43) LGSLL (11) (43) LGSLL (11) LGCCF WEQ (8) TAQGF WEQ (11) LGCCF WTQ (11) LG	HHH HHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) SSTKVILQEY NRFF(57) SSTKVILQEY KKFF(57) SSTKVILQEY HKFF(57) TSTRVMVEEK KKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK NGFF(61) SSGLYRIRGEM LFTV(42) LDNLAREVHLCR 2TVL(53) FLKRDFURHLE LTV(42) LDNLAREVHLCR 2TVL(53) THSWQRLAQEAR RTVL(53) HSWQRLAQEAR RTVL(53) HSWQRLAQEAR RTVL(53) HSWRFLSKIE DNCV(29) KRSKEFLSKIE DNCV(29) KRSKEFLSKIE DNCV(29) KRSKEFLSKIE DNCV(29) KRSKEFLSKIE DNCV(29) KRSKEFLSKIE DNCV(29) KRSKEFLSKIE DNCV(29) CPRSMDIMSKE RRCL(32) EQQMVCYTAOT RRVL(31) RQREDTOSAC QLF(63) QILKEYACETTR LKVL(95) LPIATRITALAR DRFL(30) PEEIDELGYTUY HHH HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	> HHHHHHH QGLATD(165) 'RGVQITN(210) (GAFDLIT(26) (GAFDLIT(26) (CAPDLT(26)
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78 \\ A Bt PAP \\ 201p \\ A Sc PAP \\ 3nyb \\ A Sc PAP \\ A Sc PAP$	(997) LNI DQVGI NT EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	SCEND (4) IEDUAREHE SEEEE SEEE SEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEEEE SEEEEEEE SEEEEEEEE SEEEEEEEEE SEEEEEEE SEEEEEEEEEEE SEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	RQNLESFIRQDE(H HHHHHHHHHHHHH HL (5) NFRLTIRAIKLMAKI EU (5) VFRIAIRAIKLMAKI EU (5) VFRIAIRAIKLMAKI EU (5) VFRIAIRAIKLMAKI EU (5) TFRLTIRAVKLMAKI Y (5) RVRALVFSVRCMARI HL (5) TFRLTIRAVKLMAKI Y (6) ARHTMAVKAGKX AY (5) FCRWLSMSIKRWSK AY (6) FYGAYIHTMAKE GR (2) GKENEVRLIKG GR (3) -KENEVRLIKG-LK AY (16) CRECLKLMKYLLE (2) KLKSLIRLVKHWYO DI (18) CRKECLKLMKYLLE CI (18) CRKECLKLMKYLE CI (18) CRKECLKLMK CI (18) CR	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GSI CLUFS (12) GSI CLUFS (13) NFSLTMMVIF (12) GSI CLUFS (13) NFSLTMMVIF (12) SGVTVTVALM (12) SGTTVTALM (12) SGTTVTALM (12) SGTTVTTALM (12) SGTTVTA (12) SGTTVTTALM (12) SGTTVTTALM (12) SGTTVTTALM (12) SGTTVTTALM (13) NSLTMVTF (12) SGTTMVTF (12) SGTTMTF (12) SGTTMTF	PAP/OAS1 HHH HHHH PCQL (5) ASTLV ICQL (6) LELLL PCQR (46) LELLL ICQL (5) ASTLV ICQL (6) LGQL ILQC (5) ASTLV ILQC (46) LELLL ICQL (5) ASTLV ILQC (46) LELLL ICQL (5) ASTLV ILQC (46) LELLL ICQL (5) ASTLV ICQC (46) LGQL ICQL (5) ASTLV ICQC (11) LGLCF ICQC (10) LENCF ICQC (10) LENCF ICQ	HHH HHHHHHHH HKFF (57) VSTRMVMVEER NRFF (57) ESTKKVILGEV NRFF (57) ESTKKVILGEV IFF (58) - NIRDIKKAPA KEFF (24) QSQLQKEVDLAR HKFF (57) TSTRTVMVEEK HKFF (57) TSTRTVMVEEK LDFL (55) PLKRDFLRHLE QRPA (95) ASRVRHIQUEN LGFY (42) LDNLAREVHLQR LETV (42) LDNLAREVHLQR LETV (42) LDNLAREVHLQR UETV (42) LDNLAREVHLQR NCV (29) KSKEFISKQE DNCV (29) KSKEFI HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	> HHHHHHH QGLAITD(165) RGVQITN(210) GGAFDLIT(26) ESAWILQ(210) GGAFDLIT(26) ESAWILQ(210) GGLAVID(160) RAAQCME(30) KARDTAL(4) RLREMLI(14) RLREMLI(14) RLREMLI(14) RLREMLI(14) RLREMLI(14) RLREML(14) RLREML(14) RLREML(14) RLREML(14) RLREML(14) RLREMLS(49) RLPEILS(49) RL
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78 & A & Bt & PAP \\ 2c1p & A & Sc & PAP \\ 3nyb & A & Sc & PAP \\ 3pq1 & A & Hs & PAP \\ 4lt6 & A & Hs & PAP \\ 2b4v & A & Tb & TUTase \\ 2ikf & A & Tb & TUTase \\ 3ouy & A & f & CCA \\ 4ig8 & A & Hs & OAS \\ 4ig8 & A & Hs & OAS \\ 4ig8 & A & Hs & CAS \\ 4ig8 & A & Hs & MIS1 \\ 4ig8 & A & Hs & MIS$	(997) LNI DQVGI FEEEEEEE EFF PVIKI-C-F(2) IEII PIIKF-V-E(2) IEII PIIKF-V-E(2) IEII PUVRS-E-F(2) IEII PVVKQE-F-F(2) IEII PVVKQE-T-D(4) IHO PVVQE-T-D(4) IHO PVVQE-T-D(4) IHO PVVQE-T-D(4) IHO PVVQE-T-D(5) ISV PVVT-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(6) ISV PIRE-S-S(6) KNV DRWSL-I-S(6) KNV DRWSL-I-S(6) KNV DEWSL-I-S(6) KNV PUVKF-S-H(4) FQO PIVKF-S-H(4) FQO PIVKF-S-H(5) F PIVKF-S-H(5) F PIVKF-S-H(5) F PIVKF-S-H(5) F PIVKF-S-H(SEEEE HHHHHHHHHH SEEEE HHHHHHHHHH SIGC (29) LNGCRVDDELLE SIGC (29) LNGCRVDELLE SIGT (29) LNGCRVDELLE SIGT (1) GCVLNSKLLCR SIGT (1) GCVLNSKLLCR SIGT (1) GCVLNSKLLCR VUPC (3) DENPFHHKWLEC VUPC (3) DENPFHHKWLEC VUPA (37) -FNELORDFLKN VILA (37) -FNELORDFLKN SUPC (13) DENPFHHKWLEC SUPC (3) DENPFHHKWLEC SUPC (3) DENPFHHKWLEC SUPC (3) DENPFHKWLEC SUPC (2) STHENYLWEDWFNI SUPC (2)FSHTEKN SULTA (55)FSHTEKN SULTA (55)FSHTEKN SULTA (2) STHENYLWEDWFNI SULTA (2) STHENYLWEDWFNI SULTA (2) STHENYLWEDWFNI SULTA (2) STHENYLWELVYLWE SULTA (2) STHENYLWELVYLWE <tr< th=""><th>RQNLESFIRQDE HHHHHHHHHHHHHHHHHHHHHHHHHHL 1.5. NFFLTIRAIKLAAKI 20.5. SUPRIALRAIKLAAKI 21.5. NFFLTIRAIKLAAKI 22.5. SUPRIALRAIKLAAKI 23.5. SUPRIALRAIKLAAKI 23.5. SUPRIALRAIKLAAKI 23.5. SUPRIALRAIKLAKI 25.5. SUPRIALRAIKLAKI 25.5. SUPRIALRAIKLAKI 25.5. SUPRIALRAIKLAKI 25.5. SUPRIALLAKANGKA 24.5. SUPRIALIKANGKA 25.5. SUPRIALIKANGKA 25.5. SUPRIALIKAIKANGKA 26.5. SUPRIALIKAIKANGKA 27.5. SUPRIALIKAIKANG 28.6. SUPRIALIKAIKANG 29.6. SUPRIALIKAIKANG 29.6. SUPRIALIKAIKANG 21.18. CRKECIKLAKYILLEG 21.18. CRKECIKLAKYILLEG 21.18. CRKECIKLAKYILLEG 25. SUPRIALIKURENTIA 26. SUPRIALIKURENTIA 27.5. SUPRIALIKAICANGENTIA 28.6. SUPRIALIKURENTIA 29.6. SUPRIALIKURENTIA 29.6. SUPRIALIKURENTIA 20.6. SUPRIALIKURENTIA 27.6. SUPRIALIKURENTIA 28.6. SUPRIALIKURENTIA 29. SUPRIALIKURENTIA 20.6. SUPRIALIKURENTIA <th>3) KLSLFCSEKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) SSUTWVVF (12) SGVVTWFIY (13) SGFNLMVVY (12) SGVVTMFIY (13) GVCELLIVF (14) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF (15) SHVKTAFFH (15) SUSL (17) TGEQCRE (17) TGECORKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (12) SUSLFMVIF (12) SSLTMVIF (12) SSLTMVIF (12) SSLTMVIF</th><th>PAP/OAS1 HHH HHHH TCQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ILV (34) LCRLL III (3) LGTQL III (3) CTQL (1) LGCF WTQ (1) LGCF WTQ (1) LGCF WTQ (1) LGCCF WTQ (1) LGCCF WTQ (1) LGCCF WTQ (1) LGCCF IIIS (7) LGETM WMN (8) LNVAY SIGT (6) AGEA LNG (13)SS PFVD (9) LENGF EAQE (7)MLA GEEE (8) LGERF HHH HHHHH IIQN (8) KYDYL IIQN (8) KYDYL IIQN (4) LELLL IIQN (8) KYDYL IIQC (4) LGDLL IIQC (12) YGVL</th><th>HHH HHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) VSTRNVWVEEK NRFF(57) SSTKVILQEY HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) SSGLYRIRGEM LDFL(55) PLKBRFLQEN LDFL(55) PLKBRFLRHLE QRFA(95) ASTVHLQEN LDFL(53) THSWQRLAQEN RTVL(53) THSWQRLAQEN THSWR THAN T</th><th>> HHHHHHH QGLATD(165) RGVQITN(210) GGAPDLT(26) GAPDLT(26) GAPDLT(26) CSADQCME(30) RAAQCME(30) RAA</th></th></tr<>	RQNLESFIRQDE HHHHHHHHHHHHHHHHHHHHHHHHHHL 1.5. NFFLTIRAIKLAAKI 20.5. SUPRIALRAIKLAAKI 21.5. NFFLTIRAIKLAAKI 22.5. SUPRIALRAIKLAAKI 23.5. SUPRIALRAIKLAAKI 23.5. SUPRIALRAIKLAAKI 23.5. SUPRIALRAIKLAKI 25.5. SUPRIALRAIKLAKI 25.5. SUPRIALRAIKLAKI 25.5. SUPRIALRAIKLAKI 25.5. SUPRIALLAKANGKA 24.5. SUPRIALIKANGKA 25.5. SUPRIALIKANGKA 25.5. SUPRIALIKAIKANGKA 26.5. SUPRIALIKAIKANGKA 27.5. SUPRIALIKAIKANG 28.6. SUPRIALIKAIKANG 29.6. SUPRIALIKAIKANG 29.6. SUPRIALIKAIKANG 21.18. CRKECIKLAKYILLEG 21.18. CRKECIKLAKYILLEG 21.18. CRKECIKLAKYILLEG 25. SUPRIALIKURENTIA 26. SUPRIALIKURENTIA 27.5. SUPRIALIKAICANGENTIA 28.6. SUPRIALIKURENTIA 29.6. SUPRIALIKURENTIA 29.6. SUPRIALIKURENTIA 20.6. SUPRIALIKURENTIA 27.6. SUPRIALIKURENTIA 28.6. SUPRIALIKURENTIA 29. SUPRIALIKURENTIA 20.6. SUPRIALIKURENTIA <th>3) KLSLFCSEKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) SSUTWVVF (12) SGVVTWFIY (13) SGFNLMVVY (12) SGVVTMFIY (13) GVCELLIVF (14) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF (15) SHVKTAFFH (15) SUSL (17) TGEQCRE (17) TGECORKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (12) SUSLFMVIF (12) SSLTMVIF (12) SSLTMVIF (12) SSLTMVIF</th> <th>PAP/OAS1 HHH HHHH TCQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ILV (34) LCRLL III (3) LGTQL III (3) CTQL (1) LGCF WTQ (1) LGCF WTQ (1) LGCF WTQ (1) LGCCF WTQ (1) LGCCF WTQ (1) LGCCF WTQ (1) LGCCF IIIS (7) LGETM WMN (8) LNVAY SIGT (6) AGEA LNG (13)SS PFVD (9) LENGF EAQE (7)MLA GEEE (8) LGERF HHH HHHHH IIQN (8) KYDYL IIQN (8) KYDYL IIQN (4) LELLL IIQN (8) KYDYL IIQC (4) LGDLL IIQC (12) YGVL</th> <th>HHH HHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) VSTRNVWVEEK NRFF(57) SSTKVILQEY HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) SSGLYRIRGEM LDFL(55) PLKBRFLQEN LDFL(55) PLKBRFLRHLE QRFA(95) ASTVHLQEN LDFL(53) THSWQRLAQEN RTVL(53) THSWQRLAQEN THSWR THAN T</th> <th>> HHHHHHH QGLATD(165) RGVQITN(210) GGAPDLT(26) GAPDLT(26) GAPDLT(26) CSADQCME(30) RAAQCME(30) RAA</th>	3) KLSLFCSEKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) SSUTWVVF (12) SGVVTWFIY (13) SGFNLMVVY (12) SGVVTMFIY (13) GVCELLIVF (14) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF (14) GLCELLIVF (15) SHVKTAFFH (15) SUSL (17) TGEQCRE (17) TGECORKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (17) TGEQCKIES (12) SUSLFMVIF (12) SSLTMVIF (12) SSLTMVIF (12) SSLTMVIF	PAP/OAS1 HHH HHHH TCQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ICQL (5) ASTLV ILV (34) LCRLL III (3) LGTQL III (3) CTQL (1) LGCF WTQ (1) LGCF WTQ (1) LGCF WTQ (1) LGCCF WTQ (1) LGCCF WTQ (1) LGCCF WTQ (1) LGCCF IIIS (7) LGETM WMN (8) LNVAY SIGT (6) AGEA LNG (13)SS PFVD (9) LENGF EAQE (7)MLA GEEE (8) LGERF HHH HHHHH IIQN (8) KYDYL IIQN (8) KYDYL IIQN (4) LELLL IIQN (8) KYDYL IIQC (4) LGDLL IIQC (12) YGVL	HHH HHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) VSTRNVWVEEK NRFF(57) SSTKVILQEY HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) SSGLYRIRGEM LDFL(55) PLKBRFLQEN LDFL(55) PLKBRFLRHLE QRFA(95) ASTVHLQEN LDFL(53) THSWQRLAQEN RTVL(53) THSWQRLAQEN THSWR THAN T	> HHHHHHH QGLATD(165) RGVQITN(210) GGAPDLT(26) GAPDLT(26) GAPDLT(26) CSADQCME(30) RAAQCME(30) RAA
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78 & A Bt PAP \\ 2o1p & A Sc PAP \\ 3nyb & A Sc PAP \\ 1t6 & A Hs PAP \\ 41t6 & A Hs PAP \\ 41t6 & A Hs PAP \\ 2b4v & A Tb TUTase \\ 3hj1 & A Tb TUTase \\ 3ouy & A f CCA \\ 4x4n & A Af CCA \\ 4x4n & A Af CCA \\ 4x4n & A Af CCA \\ 4x6 & A Sc UTAse \\ 4x6 & A Sc OAS \\ 41g8 & A Hs CAS \\ 4k8v & C Mm CGAS \\ 4c0 & A Hs CGAS \\ 4u0n & A Vc CGAS \\ 4u0n & A Vc CGAS \\ 4u0n & A Vc CGAS \\ 4u0 & A Vc CGAS \\ 4at7 & B Mm NF90 \\ 4w0 & A Mm NF90 \\ 4w0 & A Mm NF91 $	(997) LNI DQVGI NT EEEEEEEE EEE PVTKI-C-F(2) IEI PIIKT-K-F(2) ISI PIIKF-V-E(4) IHI PVKR-S-H(4) FQO PVKR-F-F(2) IEI PVWR-F-F(2) IEI PAVT-L-I(3) ISW PAVTL-L-I(3) ISW PAVTL-L-I(4) THI TCGRI-K-I(4) THI TCGRI-K-I(4) THI TCFRI-S-C(17) FSI PILF-S-C(17) FSI PILF-S-C(5) KNW DRWSL-I-S(6) KNW DRWSL-I-S(6) KNW DCWSL-I-S(6) KNW PUKF-RDK(3) VEP PIKK-ERR(3) LEG	SECEND (4) IEDUAREH IERE (29) LNGCRVTDEILL LLFA (29) LNGCRVTDEILL DILFA (29) LNGCRVTDEILL SEE (3)GIEAAKLIRE UYFA (29) LNGCRVTDEILL LTTN (1) RIALTSSELLY DVFA (29) LNGCRVTDEILL SIG (1) IGGVENSKLLCR VSIG (1) RAGVNSALLRE VSIG (1) RAGVNSALLRE VSIG (1) RAGVNSALLRE VUFA (37) -FTELQROFIK IILA (55)FSHIEK IILA (51) LASLRARWFEE YPMY (27)PKIVEDWFNI LILT (33) LASLRARWFEE THUT (34) LASLRARWFEE VPMY (27)	RQNLESFIRQDE HHHHHHHHHHHHHH HHHHHHHHHHHHH HL (5) NFRLTIRAIKLMAQI FW (5) GURELVLIVKQPLH YU (5) NFRLTIRAIKLMAQI YU (5) NFRLTIRAIKLMAQI YU (5) RVRALVFSVRCMARI HL (5) TFRLTIRAVKLMAKI YU (5) RVRALVFSVRCMARI HU (5) TFRLTIRAVKLMAKI YU (6) ARHTMAVKANGKA AY (5) PCRWLSMSIKRWSKQ YU (5) RVRALVFSVRCMARI YU (6) ARHTMAVKANGKA YU (5) RVRALVFSVRCMARI YU (6) ARHTMAVYRANGKA YU (5) RVRALVKANKYLEQ YU (8) CRKECLKLMKYLLEQ YU (8) CRKECLKLMKYLLEQ YU (8) CRKECLKLMKYLLEQ YU (8) CRKECLKLMKYLLEQ YU (8) CRKDCLKLMKYLLEQ YU (8) CRKDCLKLMKYLLEQ YU (8) CRKDLKMARYLLEQ YU (9) FSPVWREKRADD YU (9) FSVVWREKRADD YU (9) CRVVREKRADD YU (10) CRSLCLKLIKYLLEQ YU (9) SVVWREKRADD YU (9) FSVVWREKRADD YU (9) CRVVREKRADD YU (9) CRVVREKRADD YU (9) CRVVLVREVNCSI YU (9) SRVVVSEVNKARAND YU (9) SRVVVSEVNKAR	3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (13) NFSLTMVVIF (12) GVSWAMLVAR (12) GSI ICLVFS (13) NFSLTMVVIF (12) SVVTVMIY (13) SVGFNLMVVY (12) SVVTVMIVY (13) SVGFNLMVVY (12) SVGVLMVLY (13) GVLCELLIVF (14) GVLCELLIVF (14) GVLCELLIVF (14) GVLCELLIVF (15) SVHVKTAFFH (15) SUHVKTAFFH (15) SUHVKTAFFH (15) SUHVKTAFFH (15) SUHVKTAFFH (15) SUHVKTAFFH (15) SUFUKTAFFH (15) SUFUKTAFFH (17) TFFILMAA (17) TFFILMAA (17) TFFILMAA (17) TFGQQRKIES (13) NFSLTMVVIH (12) SVSLFLMVVIH (12) SVSLFLMVIH (12) SVSLFLMV	PAP/OAS1 HHH HHHH PCQL (5) ASTLV CCQL (5) ASTLV PCQR (46) LELLL PCQR (46) LELLL PCQR (46) LELLL PCQR (46) LELLC PCQR (48) VFQC (11) LGSLC (11) LGSLC (11) LGSCF (27Q (11) LGCCF WTQ (11) LGSCF (27Q (11) LGCCF WTQ (11) LGSCF (27Q (11) LGCCF (11) LGCCF (11) LGCCF (11) LGCCF (11) LGCCF (12) CTQ (11) LGCCF WTQ (11) LGSCF (27Q (11) LGCCF (11) LGCCF (12) CTQ (12) LGETM WMN (8) LNVAY STGT (6) ACEAL LNG (13)SS FVD (9) LENGF HHH HHH HH HH HHH HHH HH HH HHH HH	HHH HHHHHHHH HKF (57) VSTRMYWYEER NRFF (57) VSTRMYWYEER NRFF (57) STRMYWYEER KEFF (24) QSQLQKFYDLR KEFF (24) QSQLQKFYDLR KEFF (57) TSTRMYWYEEK NGFF (64) PAKFQLWKOEL LDFL (65) PLKRDFLRCHLE QRFA (95) ASRVRHLQEFN LDFL (65) PLKRDFLRCHLE QRFA (95) ASRVRHLQEFN LETV (42) LDNLARFYHLQR TVU (53) THSWQRLAQEAR CYU (53) THSWCR THY (53) THSWCR THY (53) THSWCR THY TH TA	> HHHHHHH QGLATTD(165) RGVQITN(210) GGAFDLT(26) ESAWILQ(28) QGLAVTD(160) RAAQCME(30) KARDTAL(4) RLREMLI(19) RLREMLI(19) RLREMLI(19) RLREMLI(19) RLREMLI(19) RLREMAP-(20) YERNNG-(7) IVURLUS(49) RLPEILS(49) RLPE
Inactive Active	Q5VYS8 Hs TUT7 $\begin{bmatrix} 1q78 & A & Bt & PAP \\ 2c1p & A & Sc & PAP \\ 3nyb & A & Sc & PAP \\ 3pq1 & A & Hs & PAP \\ 4lt6 & A & Hs & PAP \\ 2b4v & A & Tb & TUTase \\ 3hj1 & A & Tb & TUTase \\ 3hj1 & A & Tb & TUTase \\ 4c7x & A & Sp & TUTase \\ 4c7x & A & Sp & TUTase \\ 4c7x & A & Sc & OAS \\ 4ig8 & A & Hs & CGAS \\ 4ig8 & A & Hs & MF45 \\ 4ig8 & A & Hs & MID51 \\ 4ig8 $	(997) LNI DQVGI NT EEEEEEE EEE PVTKL-CF(2) IEI PIKKT-KF(2) IEI PIKF-V-E(4) IHI PUVRS-SH(4) FOO PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVRV-V-V(2) VEV PVVRV-V-V(2) VEV PVVRO-V-V(2) VEV PVVRO-V-V(3) VEV PVVT-L-I(5) ISV PAVTL-L-I(5) IST PIVRT-S-C(17) FSI VTINI-L-T(3) YGE EEEEEEE EEE DRWSI-I-S(6) KNV DRWSI-I-S(6) KNV DRWSI-I-S(6) KNV DRWSI-I-S(6) KNV DRWSI-I-S(6) KNV PUVRF-SH(4) FOO PIVKF-ENR(3) VEE PIIKL-T-D(4) VKV	SEEEE HHHHHHHHHH SEEEE HHHHHHHHHH SIGC (29) LNGCRVDDELLE SIGC (29) LNGCRVDDELLE SIGC (29) LNGCRVDDELLE SIGC (29) LNGCRVDDELLE SIGC (29) LNGCRVDELLE SIGC (29) LNGCRVDELLE SIGC (29) LNGCRVDELLE SIGT (29) LNGCRVDELLE SIGT (1) GGVENSKLLCR SIGT (1) GGVENSKLLCR SIGT (1) GGVENSKLLCR VUPC (3) DENPFHHKWLEC VUPC (3) DENPFHHKWLEC VUPA (37) -FNELORDFLRN VUPA (37) -FNELORDFLRN SILLA (55) FSHTEKN SILLA (52) FSHTEKN SILLA (52) FSHTEKN VPMY (63) - PK'VEOWFNI	RQNLESFIRQDE HHHHHHHHHHHHHHHHHHHHHHHHHH L(5) NFFLTLRAIKLAAKI EU(5) SUFRIALRAIKLAAKI EU(5) VFRIALRAIKLAAKI EU(5) SUFRIALRAIKLAAKI EU(5) SUFRIALRAIKLAAKI EU(5) SUFRIALRAIKLAAKI HL(5) FFLTIRAVKLAKI HL(5) FFLTIRAVKLAKI Y(6) AARHTAMAVKAOGKA AX(5) PCRWLSMSIRRASKA A1(6) FYGAYIHLVKANGKA Y(5) SUKEMULLVKHMAKI SR(2) CKENEVRLIKG= R(2) CKENEVRLIKG= R(3) - ENEVRLIKG= R(3) - CKECLKLAKYLLEG YI (18) CRKECLKLAKYLLEG YI (18) CRKECLKLAKYLLEG YI (18) CRKECLKLAKYLLEG SI (6) HLRKVCRFWARDD SI (6) HLRKVCRFWARDD R(10) TRRELLIKOLRIKA XR (10) CRSLCIKLIKAICHALD XR (10) CRSLCIKLIKAICKNCSI XG (4) - SDEIKTLORYMCSI XG (5) -STOVARLORYMCSI YI (5) SVVEDUVINKARASI	3) KLSLFCSEKNG (HHHHHHHH (12) GVSWAMLVAR (12) GVSWAMLVAR (12) GSI ICLVFS (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSSLTMWVIF (12) GVSWAMLVAR (12) SGVVTWFIY (13) GVGFNLMVVY (13) GVGFNLMVVY (14) GYLCELIIVF (14) GYLCELIIVF (14) GYLCELIIVF (14) GYLCELIIVF (14) GYLCELIIVF (15) GYVKTAFFH (15) SYVKTAFFH (15) SYVKTAFFH (17) FEFTILMAA (8) FILDLIGHY (8) FILDLIGHY (17) TEEQRKIES (7) ICEQRKIES (7) ICEQRKIES (12) SYSLFIMAVS (12) SYS	5) SULV(CM T(PAP/OAS1 HHH HHHHH TCQL(5) ASTLV ICQL(5) ASTLV ICQL(5) ASTLV TCQR(46) LELLL TCQR(46) LELLL ILU(34) LCRLL (ILU(35) LCRQV ILU(35) LCRQV ILU(35) LCRQV ILU(35) LCRQV ILU(35) LCRQV WER(8) TAQGF WER(8) TAQGF ILLS(7) LCETM WMR(8) LNVAY SIGT(6) AGEN ZOTQ(11) LGLCF ILDS(7) LCETM WMN(8) LNVAY SIGT(6) AGEN ZOTQ(11) LGCF HHH HHHHH ILQN(8) KYPYL ILEA(9) RYACL IQN(8) KYPYL ILAQN(8) KYPYL ILQN(8) KYP	HHH HHHHHHHH HKFF(57) VSTRNVWVEEK NRFF(57) VSTRNVWVEEK NRFF(57) SSTKVUILQEY HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) TSTRVMVEEK HKFF(57) SSGLYRIRGEM LDFL(65) PLKRDFURHLE QRFA(95) ASRVHIQQEN HQFF(67) SSGLYRIRGEM LETV(42) LDNLAREVHLCR QTVL(53) THSWQRLAQEA RTVL(53) THSWQRLAQEA THA THA TH(14) RQTLNLTTNA THEN TLH(14) RQTLNLTTNA THH(14) RQTLNLTTNA THH(14) RQTLNLTSLA THH TLH(14) RQTLNLTSLA THH TLH(14) RQTLNLTSLA THH TLH(14) RQTLNLTSLA THM TLH(14) RQTLNLTSLA TM TLH(14) RQTLNLTSLA TLH(14) RQTLNLTSLA TM TLH(14) RQTLNLTSLA TM TLH(14) RQTLNC TLH(14) RQTLNC TLH(14) RQTLNC TLH(14) RQTLNC TLH(1	> HHHHHHH QGLATD(165) GGAPDLT(210) GGAPDLT(210) GGAPDLT(210) GGAPDLT(210) KARDTAL(4) KARDTAL(4
Inactive Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 2b4V_A Tb TUTase 2hf_A Tb TUTase 3hj1_A Tb TUTase 3ouy_A Af CCA 4x4n_A Af CCA 4x4n_A Af CCA 4x4n_A Af CCA 4x4n_A Af CCA 4x62_A Sc CAS 41g8_A Hs CAS 41g8_A HS CAS 4k8v_C Mm CGAS 4k8v_C Mm CGAS 4c68_A Hs CGAS 4u0n_A Vc CGAS 4at7_B Mm NF90 4at7_B Mm NF90 4at7_B Mm NF91 4at7_B Mm NF91 4at7_B Mm MiD49 Q961P4 Hs FAM46B Q5VWP2 Hs FAM46B Q5VWP2 Hs FAM46B Q5VWP2 Hs FAM46B Q9NVV4 HS TUT1 Q6PIY7 Hs TUT2 Q8NDFE Hs TUT3 Q5TAX3 HS TUT4 Q5XG87 Hs TUT5 Q9H6ES Ha TUT5	(997) LNI DQVGI NTT EEEEEEEEEEEEEEE PVTKI-C-F(2) IEI PIKK-V-E(2) ISI PIKF-V-E(4) IHI PVKR-S-H(4) FQO PVVK-G(3) VDP PVVR-K-G(3) VDP PVVR-K-G(3) VDP PVVR-K-G(3) VDP PVVR-K-G(3) VDP PVVR-K-G(3) VDP PVVR-V-C(2) VEV PVVR-V-V(2) VEV PVVR-V-V(2) VEV PVVR-V-V(2) VEV PVVR-V-U(8) VEP -ALSF-V-L(8) VEP PAVT-L-I(5) ISV PAVT-L-I(5) ISV PAVTL-L-I(3) ISV PAVT-K-N(5) LST PIRT-S-S(2) ATV AAIVT-K-N(5) LST PIRT-S-S(6) KNV PIKE-I-S(6) KNV DCWSL-I-S(6) KNV PIVKF-ER(3) VEP PIKL-T-D(4) VKV PIVKF-ER(3) LEG PIVKF-CHR(3) LEG PIVKE-CHR(3) LEG PIVKE-CHR(3) LEG PIVKE-CHR(4) IHG	SECENDIC (*) TEDUAREHE SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEEE SEEGE SEGE SEGE SEGE SEGE SEGE VUPC (13) DENPEHHKWLEG VUPC (13) DENPEHHKWLEG VUPA (37) - FELORDELKE SILA (55) FSHIEKE SILA (55) FSHIEKE SILA (55) FSHIEKE SILA (55) FSHIEKE SILA (53) LASIRHARWEDE SILT (23) LAAIRHARWEDE SILT (23) LAAIRHARWEDE SILT (23) LAAIRHARWESE HHHHHKHKE VILT (23) LAAIRHARWESE HILT (23) LAAIRHARWESE HILT (23) LAAIRHARWESE HILT (23) LASIRHARWESE HILT (23) LASIRHARWESE	RQNLESFIRQDE(H HHHHHHHHHHHHH HL (5) NFRLTIRAIKLMAKI EU (5) VFRLTIRAIKLMAKI EU (5) VFRLTIRAIKLMAKI EU (5) VFRLTIRAIKLMAKI EU (5) FRLTIRAIKLMAKI IY (5) RVRALVFSVRCMARI HL (5) TFRLTIRAVKLMAKI Y (5) RVRALVFSVRCMARI HL (5) TFRLTIRAVKLMAKI ST (5) FCRWLSMSIKRMSK AY (2) GKEDLKLMKYLLE CI (8) CRKECLKLMKYLLE CI (8) CRKECLKMKYLLE CI (8) CRKECLKMKYLE CI (3) KLSLFCSSKNG (HHHHHHHH (12) GVSWAMLVAR (12) GSI ICLVFS (12) GSI ICLVFS (13) NSITMVIF (12) GVSWAMLVAR (12) GSI ICLVFS (13) NSITMVIF (12) SGVVTWFIY (13) SGFNLMVY (13) SGFNLMVY (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (14) GYLCELLIVF (15) SHVKTAFFH (15) SILMAATVN (8) SILMAATVN (8) SILMAATVN (8) SILMAATVN (8) SILMAATVN (17) ISCUMATIN (17) ISCUMATIN	5) SULV (CMT(PAP/OAS1 HHH HHHHH PCQL(5) ASTLV (CQL(5) SAVIL CQL(5) SAVIL PCQR(46) LEILL PCQR(46) LEILL PCQR(46) LEILL (LU(34) LGRLL (LU(35) LGRCQV /LQE(48) VFCQ (GSF WEQ(6) TAQGF WEQ(6) TAQGF WEQ(6) TAQGF WEQ(6) TAQGF WEQ(6) TAQGF WEQ(6) TAQGF WEQ(6) TAQGF (GSF WEQ(6) TAQGF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (GSF (G	HHHH HHHHHHHH HKFF (57) VSTRWWWEER NRFF (57) VSTRWWWEER NRFF (57) SSTKWILQEFV LFF (58)NIRDIKKAPA KEFF (24) QSQLQKFVDLAR HKFF (57) TSTRTWWEEK NGFF (64) PAKFQLWKQEL LDFL (65) PLKRDFLRCHLE QRFA (95) ASRVRHLQQEN LGFT (67) SSGLYRIRGEM LETV (42) LDNLARFVHLQR TVU (53) THSWQRLAQEAR RTVL (53) THSWQRLAQEAR RTVL (53) THSWQRLAQEAR DNCV (29) KRSKEFLSKVE DNCV (29) KRSKEFLSKVE TNCV EXT (29) CPREMDIMSKE KITA (29) CPREMDIMSKE KI	> HHHHHH QGLATD(165) RGVQITN(210) GAFDLIT(26) GA
Inactive Active	Q5VYS8 Hs TUT7 1q78_A Bt PAP 2o1p_A Sc PAP 3nyb_A Sc PAP 3nyb_A Sc PAP 2b4v_A Tb TUTase 2ikf_A Tb TUTase 2ikf_A Tb TUTase 4e7x_A Sp TUTase 4e7x_A Sp TUTase 4e7x_A Sp TUTase 4e7x_A Af CCA 1px5_A Sc OAS 4ig8_A Hs OAS 4ig8_A Hs OAS 4ig8_A Hs OAS 4ig8_A Hs OAS 4ig8_A Hs CAS 4kw_C Mm CGAS 4kw_C Mm CGAS 4kw_C Am CGAS 4kw_A Hs CGAS 4u0_A Vc CGAS 4u0_A Vc CGAS 4u0_A Vc CGAS 4u1_A Sc Utp22 4mst_A Hs MID51 4woy_A Mm MiD49 Q96IP4 Hs FAM46A Q96A09 Hs FAM46B Q5VWP2 Hs TUT1 Q6PIY7 HS TUT2 Q8NDFB Hs TUT3 Q5TAX3 HS TUT4 Q5XG87 HS TUT5 Q9H6E5 HS TUT5 Q9H6E5 HS TUT5 Q54K5 HS TU	(997) LNI LDQVGI PUTKL-C-F(2) IEI PUTKL-C-F(2) IEI PITKT-K-F(2) IEI PITKF-V-E(4) IHI PUVR-S-H(4) FQO PVVRV-K-G(3) VDD PVVRV-K-G(3) VDD PVVR-V-V(2) VEV PVTR-L-T0(8) VED PVVR-L-V-(8) VED PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(5) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(3) ISV PAVTL-L-I(4) THI TGREI-S-C(1) FIL PITRL-S-C(1) FIL PITRL-S-C(6) KNVE DRWSL-I-S(6) KNVE PIVRE-SH(4) VED PITKL-T-D(4) VKVE PIVKF-CRR(3) LEG PITKL-T-D(4) VKVE PIVKF-CRR(3) LEG PITKL-T-D(4) VKVE PIVKF-CRR(3) LEG PITKL-T-D(4) VKVE PIVKF-CRR(1) LEG PIVKF-CRR(1) LEG PIVK	SECENDIC (*) TEDUAREHIP SEEEE HHHHHHHHHH SIGE	RQNLESFIRQDE H HHHHHHHHHHHHHHHH L(5)NFRLTIRAIKLMAKI EU(5)GURELVLIVGPLH IY(5)RVRALVFSVRCMARJ IY(5)STRLTIRAIKLMAKI IY(5)RVRALVFSVRCMARJ IY(5)STRLTIRAIKLMAKI IY(5)STRLTIRAIKLMAKI IY(5)STRLTIRAIKLMAKI IY(5)STRLTIRAIKLMAKI SY(5)SILMAKAGGKA AY(5)PCRWLSMSIRMSKQ AY(5)PCRWLSMSIRMSKQ SY(5)SILMAKAGGKA AY(5)PCRWLSMSIRMSKQ SY(5)SILMAKAGGKA AY(5)PCRWLSMSIRMSKQ SY(5)SILMAKAGKA AY(5)PCRWLSMKHKGEKG SY(5)SILMAKAGKA AY(15)CKLKMYLLEQ SY(18)CKECLKLMKYLLEQ SY(18)CKECLKLMKYLLEQ SY(18)CKECLKLMKYLLEQ SY(18)CKECLKLMKYLLEQ SY(18)CKECLKLMKYLLEQ SY(18)CKECLKLMKYLLEQ SY(18)CKECLKLMKYLLEQ SY(18)CKECLKLMKYLLEQ SYSSVIJIRURALGEN AY(18)CKECLKLMKYLLEQ SYSVVILEKARALD SYSVVILEKARALD SYSVVILEKARALD SYSVVILEKARKALD AY(10)	3) KLSLFCSSKNG(HHHHHHHH (12) GVSWAMLVAR (12) GVVTWHIY (12) STVTTMALM (12) STVLCELLIVF (11) STLALTTM (12) STLMATTM (12) STLMATTM (13) STLMATTM (13) SSLTNUTH (12) SVLTNUTH (12) SVLTNVLH (12) SVLTMAVE (12) SVLTMATFH (2) SU	PAP/OAS1 HHH HHHH PQQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (5) ASTLV [QQL (6) LGQL PLQ (46) LGQL [IL (43) LGSLL (IL (43) LGSLL (IL (43) LGSLL (IL (43) LGSLL (IL (43) LGSLT (IL (43) LGSLT (IL (43) LGSLT (IL (43) LGSCT (IL (11) LGCCF WER (8) TAQGF WER (8) TAQGF WER (8) TAQGF WER (8) TAQGF (IL (11) LGCCF [UDS (7) LGETM [LDS (7) LG	HHH HHHHHHHH HKFF (57) VSTRWVMVEER NRFF (57) ESTKKV ILQEP NRFF (57) ESTKKV ILQEP NRFF (57) TSTRTVMVEER KEFF (24) QSQLQKPVDLAR HKFF (57) TSTRTVMVEER LDFL (55) PLKRDFLRRHLD DRFA (95) ASRVRHLQEEN LDFL (65) PLKRDFLRRHD DRFA (95) ASRVRHLQEEN LETV (42) LDNLARFVHLQR LETV (42) LDNLARFVHLQR RTVL (53) PKGWRQLAQEAR TVL (53) PKGWRQLAQEAR DNCV (29) KRSKEFLSKLE DNCV (29) LRSKEFLSKLE DNCV (29) LRSKEFLSKLE LSF (43) DNLE LSF (53) - GAMQVKQVE DACH (57) SRVARLOVCRE LSF (> HHHHHHH QGLATTD(165) GAPDLT(26) GAPDLT(26) GAPDLT(26) GAPDLT(26) GAPDLT(26) GAAQCME(30) KARDTAL(4) ICLEMEAP(122) ICLEMEAP(228) ICLEMEAP(288) ICLEME

Figure 2. Multiple sequence alignment of human FAM46 proteins, human non-canonical poly(A) polymerases (TUT1-7) and all representative structures possessing both the NTase and PAP/OAS1 SBD domains. Only conserved regions of the domains are shown. Sequences are labelled with PDB code or UniProt ID. The numbers of excluded residues are specified in parentheses. Residue conservation is denoted with the following scheme: uncharged, highlighted in yellow; polar, highlighted in grey; invariant active site residues involved in catalysis, highlighted in black; critical substrate binding residues, highlighted in red. Locations of observed and predicted secondary structure elements are marked above the corresponding alignment blocks. Abbreviations: PAP, poly(A) polymerase; TUTase, terminal uridylyltransferase; CCA, CCA-adding enzyme; OAS, oligoadenylate synthetase; cGAS, cyclic GMP-AMP synthase; NF45 and NF90, nuclear factors NF45 and NF90; Utp22, U3 small nucleolar RNA-associated protein 22; MiD51 and MiD49, mitochondrial dynamics proteins MiD51 and MiD49; Ss. *S. scrofa*; Tb, *T. brucei*; Af, *A. fulgidus*; Hs, *H. sapiens*; Mm, *M. musculus*; Sp, *S. pombe*; Sc, *S. scrofae*; Vc, *V. cholerae*; Bt, *B. taurus*. Sequence-to-structure alignment for FAM46 proteins can be assigned higher confidence in the NTase domain.



Figure 3. Comparison of 3D model of human FAM46C and available structures of non-canonical poly(A) polymerase Trf4p (pdbl3nyb) and mitochondrial dynamics protein MiD51 (pdbl4oaf). Regions in MiD51 responsible for protein–protein interactions and their potential counterpart in FAM46C are coloured pink. The region between the fourth and fifth core α -helices of PAP/OAS1 SBD in FAM46C and Trf4p (critical for nucleobase binding) is shown in green. The region not modelled in FAM46C (70 amino acids) is denoted by red dots. The conserved active site carboxylates are shown in blue.



Figure 4. Comparison of the active sites of FAM46C, poly(A) polymerase (Pap1, pdbl1fa0), poly(U) polymerase (Cid1, pdbl4fhp), CCA-adding enzyme (pdbl4x4r), OAS (OAS1, pdbl4rwo) and cyclic GMP-AMP synthase (Mb21d1, pdbl4k97). Only NTase and PAP/OAS1 SBD domains are shown. The region not modelled in FAM46C (70 amino acids) is denoted by red dots. Conserved amino acids critical for catalysis and substrate binding are shown in blue.

the incoming NTP like Tyr212 in poly(U) polymerase Cid1 or makes van der Waals contacts with the ribose-base moiety of NTP similar to Val234 in poly(A) polymerase Pap1 (72). Finally, NTP β - and γ -phosphates most likely interact with the conserved Arg268 in addition to Ser74 from the hG[GS] motif.

Cellular localization and tissue specificity

According to various servers predicting subcellular localization, the FAM46 proteins seem to be localized in both the cytoplasm and nucleus. In addition, three human paralogs (FAM46B, FAM46C and FAM46D) harbour potential leucine-rich NES, located at the end of the C-terminal PAP/OAS1 SBD domain. As a consequence, it is likely that proteins belonging to the FAM46 family shuttle between the nucleus and cytoplasm.

We also analysed the gene expression data available in the BioGPS database (28) and found that each of the human FAM46 paralogs has a different tissue/cell expression pattern (Supplementary Table S1). Different expression patterns probably indicate various biological processes in which FAM46 proteins participate. According to the BioGPS database, FAM46A, FAM46B and FAM46C are potentially expressed in 81, 18 and 66 tissues/cells, respectively, while FAM46D can be found only in sperm (Supplementary Table S1).

To get more hints at the biological function of FAM46. we compared all human NTase superfamily members according to GO molecular function and biological process of their protein interactors identified in the BioGRID database (47) (Supplementary Figure S5). We found that FAM46 binding partners share a common GO functions and processes mostly with interactors of human NTase fold proteins from the following four groups (described in Supplementary Table S2): interleukin enhancer-binding factors, non-canonical poly(A) polymerases (TUT), poly(A) polymerases (PAP) and zinc finger RNA-binding proteins (ZFR). Similarly to FAM46 family members, proteins belonging to these groups also possess additional PAP/OAS1 SBD domain. In addition, almost all of them retain the same biological function-DNA/RNA binding, including poly(A) RNA binding and participate in the same process-transcription.

We also analysed all 61 FAM46 protein interactors identified in the BioGRID database and the literature in more details (Supplementary Table S3). We found that each FAM46 paralog has a different set of interacting protein partners; therefore, it is likely that each paralog participates in a different biological process in the cell. Most of the interacting partners play important roles in development, including cellular proliferation and cell differentiation. We noticed that many FAM46 interactors share a common molecular functions (e.g. nucleic acids binding, including binding of the mRNA poly(A) tail) or biological processes (e.g. protein modification, transcription). Specifically, 26 FAM46 interacting partners may directly bind RNA and/or DNA. This group includes: transcription (co)factors (RHOXF2, TBX4, NR2F2, SOX5, NRF1, TRIP6), transcription activators (Znf322, Cxxc5), RNA stabilization factors (ELAVL1, BCCIP, HDLBP, Pabpc1, Pabpc4), proteins involved in transcription (POLR1A, POLR1E, POLR2J), proteins participating in mRNA translation (EIF4G3, Pabpc1, Pabpc4), proteins taking part in DNA repair (POLE2, Rad23b, WRAP53) and other proteins, which play various roles such as DNA helicase DDX11, putative RNA exonuclease (44M2.3), mitochondrial translation optimization factor 1 (MTO1), SUMO-conjugating enzyme UBC9 (UBE2I), 14-3-3 protein zelta/delta (Ywhaz) involved in signal transduction (78) and ATXN1, which may participate in RNA export (79). Similarly to human FAM46A (80), nine protein interactors (ELAVL1, BCCIP, EIF4G3, MTO1, Ywhaz, TRIP6, UBE2I, Pabpc1, Pabpc4) participate or may participate in RNA poly(A) tail binding. Another seven FAM46 interacting partners (EGLN2, DAZAP2, KEAP1, DCAF6, KLHDC2, ZNHIT6, MVP) were shown or suggested to cooperate with or regulate proteins, which are able to bind directly to nucleic acids. Specifically, the Egl nine homolog 2 (EGLN2) targets the transcriptional complex HIF- α subunit for proteasomal degradation (81). KEAP1 targets transcription factor Nrf2 for ubiquitination and degradation (82). DCAF6 enhances the transcriptional activity of nuclear receptors NR3C1 and AR (83). The physical interaction between KLHDC2 and a bZIP transcription factor (LZIP) leads to the repression of the LZIP-

dependent transcription (84). DAZAP2 regulates stress or germ granules-ribonucleoprotein complexes (85.86). MVP is a part of an evolutionary highly conserved ribonucleoprotein particles (vaults) (87,88), while ZNHIT6 may be involved in snoRNP biogenesis (89). Another functional group of FAM46 interactors contains proteins, which take part in protein modification (mostly in protein degradation). This group embraces proteases (serine protease HTRA1, caspase-like protease ESPL1), Kunitz-type protease inhibitor 1 (SPINT1), Kelch-like ECH-associated protein 1 (KEAP1) participating in Nrf2 ubiquitination and degradation (82), EGLN2 targeting HIF- α for proteasomal degradation (81), E3 ubiquitin-protein ligases (RNF14, PARK2), CUL4-associated factor 6 (DCAF6) functioning as substrate-recruiting module for CUL4-DDB1 E3 ubiquitin-protein ligase complex (90), SUMO-conjugating enzyme UBC9 (UBE2I), ubiquitin carboxyl-terminal hydrolase 4 (USP4), a regulatory subunit 6A of the 26S proteasome (Psmc3) and BAG6, which is crucial in ubiquitinmediated protein degradation of defective or misallocated polypeptides (91). Among all FAM46 interactors, we were able to identify two kinases: Polo-like kinase 4 (PLK4) and non-receptor tyrosine-protein kinase SYK (SYK). PLK regulates cell cycle progression, mitosis and cytokinesis (92), while SYK mediates signal transduction and differentiation, particularly in B-cell development (93,94). Another interacting partner, the FYVE domain-containing protein 9 (ZFYVE9) identified as a SMAD2/3-binding protein, may also regulate the proliferation of hepatic cells during zebrafish embryogenesis (95). The last identified functional group contains proteins responsible for an intra and extracellular cell transport, including DYNC1H1-a heavy chain of cytoplasmic dynein 1 (96,97), RIN3-a small GT-Pase, which participates in intracellular membrane trafficking (98) and AP2B1, which plays a pivotal role in many vesicle trafficking pathways within the cell (99).

Mutations in cancers

Recent studies have identified numerous somatic mutations in various cancer patients leading to single point mutations in the human FAM46 proteins (Supplementary Table S4). For instance, the human FAM46C gene was reported as a causal driver of multiple myeloma (6). In addition, a single FAM46C mutation (Y247N) was identified in hem6 mice with hypochromic anemia, which affects terminal spermiogenesis and terminal stages of erythroid differentiation (100). This study showed that male *hem6* mice produce sperm with defects detectable by phase contrast microscopy and fluorescence microscopy. To analyze the role of these mutations, we mapped them onto a 3D model of FAM46C (Figure 5) and found that they can be divided into two groups. The first group includes mutations that are located in a highly conserved area of the active site and its close vicinity, and probably may decrease/increase FAM46 catalytic activity and/or affect substrate binding (e.g. change the preference for the type of incorporated NTP). This group embraces the mouse *hem6* mutation and the majority of mutations found in multiple myeloma patients as well as several mutations from other cancers. The average JSD (52) score for all mutated amino acid positions in multi-



Figure 5. Missense mutations in FAM46 family members found in cancer patients and *hem6* mouse. The positions of the corresponding single point mutations, mapped onto a 3D model of human FAM46C, are shown as spheres. The spheres are coloured according to JSD score, which refers to the amino acid conservation in FAM46 family.

ple myeloma (reported in Supplementary Table S4) and for the mutated residue 247 in *hem6* FAM46C is 0.56 and 0.74, respectively (higher JSD scores correspond to higher sequence conservation). In benchmarks JSD approach, which considers also estimated conservation of sequentially neighbouring sites, performed better than traditional measures (e.g. Shannon entropy or Sum-of-pairs measure) in identifying functionally important residues (52). In comparison, the average JSD score for five FAM46 active site residues: glutamic/aspartic acids, glycine and serine is 0.65. Mutations belonging to the second group are located mostly on the protein surface (usually with JSD scores below 0.4), in evolutionary low conserved regions. Those mutations may affect protein–protein interactions or alternatively might not play any crucial role in the reported cancers.

In addition, we selected all mutations of highly conserved residues with a JSD score higher than 0.65 (the average JSD for five FAM46 active site residues from conserved NTase motifs). It allowed us to identify mutations found in a number of malignancies (highlighted in orange in Supplementary Table S4), which probably have the largest impact on protein activity and may be connected with diagnosed cancers. Consequently, we suggest that, in addition to multiple myeloma. FAM46 genes may be also involved in pathogenesis of various other cancer subtypes including liver hepatocellular carcinoma, bladder urothelial carcinoma, head and neck squamous cell carcinoma, uterine corpus endometrial carcinoma, kidney renal papillary cell carcinoma, lung adenocarcinoma, ductal adenocarcinoma, colorectal adenocarcinoma, primary plasma cell leukemia and skin cutaneous melanoma.

Potential function

The results of our sequence and structure analyses suggest that the FAM46 proteins are active NTases, which have both the NTase fold and PAP/OAS1 SBD domains. Active NTases possessing the PAP/OAS1 SBD are known to participate in tRNA maturation (CCA-adding enzymes), RNA degradation (TUTases, poly(A) polymerase in TRAMP complex), mRNA maturation (poly(A) polymerases) and in a defense response to viruses and bacteria (2'-5'-oligoadenylate synthases and cyclic GMP-AMP synthases). Although it was shown that both FAM46A and FAM46C are induced by interferon I and II (14) and that FAM46C is one of the interferon-stimulated genes (ISGs)

which modify viral (YFV and VEEV) replication during infection, it is unlikely that FAM46 proteins are antiviral enzymes like OASes. Unlike replication inhibiting ISGs (such as OASL, Mab-21 and C6orf150), FAM46C slightly enhances the replication of certain viruses (14). FAM46 family members do not possess also an $H(X_5)CC(X_6)C$ motif (conserved among vertebrate cGAS members) located between the NTase and PAP/OAS1 SBD domains. This motif, which resembles most closely HCCC-type zinc-ribbons found in TAZ domains, is required for efficient cytosolic DNA recognition (101). Finally, we investigated the possibility that FAM46 proteins may be novel non-canonical poly(A) polymerases participating in RNA 3' end modification like TU-Tases or poly(A) polymerases GLD-2 and GLD-3. This hypothesis is consistent with M. Tian studies (100), where it was shown that mutated FAM46C may modulate the poly(A) tails of specific transcripts during erythroid differentiation. The author identified a single FAM46C mutation (Y247N) in *hem6* mice and showed that it might cause an accelerated, progressive shrinkage of the poly(A) tail in four transcripts (alpha-globin, Alas2, Hbb-bl and Ftll) and probably does not have any effect on poly(A) tails in two transcripts (Fth1 and beta-actin). Additionally, a Y247N mutation led to an increase of expression levels of 152 transcripts, resulted in a decrease of expression levels of 29 transcripts, and did not have any effect on 29 erythroid transcripts (100). It should be noted, however, that M. Tian used an indirect approach to analyze the poly(A) tail lengths in the six aforementioned transcripts. His strategy of indirect poly(A) tail length assay assumed that the poly(A) tails do not possess any nucleotides other than adenosines; therefore, he was not able to identify the real length of the modified poly(A) tails if they also contain other nucleotides. Thus, it is likely that he observed shortening of mRNA poly(A) tails for specific transcripts if FAM46C is responsible for the addition of adenosines to the RNA 3' end. The FAM46C mutation (Y247N) might have weakened the processivity of FAM46C resulting in poly(A) tails shrinkage. On the other hand, FAM46C may be a non-canonical poly(A) polymerase which adds cytidines or uridines to the RNA 3' end. In this scenario, FAM46C may participate in transcript degradation by modifying the poly(A) tails. This hypothesis is consistent with the fact that up to 152 transcripts increased expression levels in a *hem6* mutant. In this case, the observed shortening of the poly(A) tails in four

Nucleic Acids Research, 2016 11

transcripts may be a side effect of cell deregulation. In both scenarios. FAM46 proteins may play a very important role in mRNA stability as active non-canonical poly(A) polymerases rather than some other factors, which prevent early mRNA degradation by disrupting interactions between ribonuclease docking complex and RNA as suggested by M. Tian (100). Our functional assignment is also in line with the facts that mouse FAM46C may bind directly or through a complex to RNA CU-rich motifs (100) and FAM46A may bind to poly(A) tails (80). According to Chapman *et al.*. the expression of FAM46C is highly correlated with the expression of ribosomal proteins and initiation and elongation factors involved in protein translation (6). They proposed that FAM46C is functionally related in some way to the regulation of translation (e.g. as a mRNA stability factor), however, they did not assign any exact function to this protein. Recent studies revealed that the poly(A) tail length impacts gene expression in some processes such as inflammation, learning and memory (102), and there is a clear correlation between the poly(A) tail length and translational efficiency in early development stages in zebrafish and African clawed frogs (103). Therefore, it is possible that the correlation observed by Chapman et al. is the effect of length change of the poly(A) tails.

Both the M. Tian and Chapman et al. studies are consistent with the results of our analysis of FAM46 interactors and interacting partners of all remaining human NTase fold proteins. We found that FAM46 binding partners share a common GO functions and processes mostly with interactors of those active NTase fold superfamily members which belong to non-canonical poly(A) polymerases and poly(A) polymerases. We showed that over half of the 61 identified FAM46 interactors participate in DNA and/or RNA binding, including nine proteins which can bind mRNA poly(A) tails. Many of FAM46 interacting partners are involved in transcription or translation, like transcription (co)factors (RHOXF2, TBX4, NR2F2, SOX5, NRF1, TRIP6), transcription activators (Znf322, Cxxc5), RNA stabilization factors (ELAVL1, BCCIP, HDLBP, Pabpc1, Pabpc4), proteins involved in transcription (POLR1A, POLR1E, POLR2J), proteins participating in mRNA translation (EIF4G3, Pabpc1, Pabpc4), and proteins taking part in transcription regulation (EGLN2, KEAP1, DCAF6, KLHDC2) and mitochondrial translation optimization (MTO1). Finally, some FAM46 protein interactors regulate or are a part of ribonucleoprotein complexes.

Our domain architecture analysis revealed that proteins belonging to FAM46 family possess only two domains: NTase and PAP/OAS1 SBD, with single exceptions of some additional domains present in a few proteins. Importantly, we were not able to detect any additional conserved domains such as ferredoxin-like, which plays a critical role in processivity of canonical poly(A) polymerases or TU-Tases (Supplementary Figure S3). The ferredoxin-like domain provides additional interactions with RNA and may enhance its binding, allowing the NTase enzyme to add up to several hundred nucleotides. Therefore, FAM46 proteins acting as non-canonical poly(A) polymerases probably can add only a few nucleotides to the RNA 3' end.

FAM46 family members seem to be localized both in the cytoplasm and nucleus, like two other human noncanonical poly(A) polymerases, PAPD4 and PAPD5 (104). Considering the physiological functions of FAM46 interactors, we can speculate about the biological processes, in which FAM46 proteins may participate. FAM46A probably cooperates with a subunit RPB11-a of DNA-directed RNA polymerase II, eukaryotic translation initiation factor 4 gamma (eIF4G), high-density lipoprotein-binding protein (HDLBP, Vigilin), while FAM46C may bind to polyadenylate-binding proteins (Pabpc1, Pabpc4) and (together with FAM46A) to ELAV-like protein 1. As a consequence, proteins belonging to FAM46 family can be involved in mRNA (de)stabilization either in the nucleus or cytoplasm. DNA-directed RNA polymerase II transcribes all protein-coding genes and synthesizes many functional non-coding RNAs. The eIF4G3 subunit is a scaffold protein in eIF4F complex, which participates in the recruitment of eukaryotic mRNAs to the ribosome (105). Pabpc1 and Pabpc4 belong to cytoplasmic poly(A) binding proteins (PABPC), which bind specifically to the poly(A) tail of mRNA and are required for poly(A) shortening, ribosome recruitment and translation initiation (106). Another protein interactor, Xenopus Vigilin, can selectively protect in vitro vitellogenin mRNA from cleavage by endonuclease PMR-1 (107), while ELAVL1 is described in the literature usually as a stabilization factor, which prevents the degradation of mRNAs possessing short tails (108–110). FAM46 proteins can be also involved in a ribosome biogenesis (like POLR1A, POLR1E interactors (111,112)) or they can (de)stabilize a nuclear pool of extra-ribosomal RPL23 and the pre-60S trans-acting factor eIF6 (like BC-CIP interactor (113)). By interacting with telomerase Cajal body protein 1 (WRAP53), FAM46 family members may change 3' ends of small Cajal body RNAs, which are involved in modifying splicing RNAs (114). Together with the Box C/D snoRNA protein 1 (ZNHIT6), they may also participate in snoRNP biogenesis, which is essential for the processing and modification of rRNA (89). Finally, FAM46 proteins (together with DAZAP2 and major vault proteins (MVP)) may modify RNAs which build ribonucleoproteins complexes like stress granules (85) and vaults composed of MVP, vault poly(adenosine diphosphate-ribose) polymerases (VPARP), telomerase-associated proteins (TEP1) and small untranslated RNAs (vRNAs) (87,88).

The FAM46 family members seem to be highly regulated proteins. The process, in which these new non-canonical poly(A) polymerases participate, is probably determined by their tissue-specific expression and gene organization. As reported in the BioGPS database, tissue expression levels are different for each human FAM46 paralog. Moreover, the human FAM46 proteins are likely to be regulated by phosphorylation. Each human paralog has many phosphorylation patterns detectable with ELM predictor (27) with high probability scores (data not shown). For instance, FAM46A, FAM46B and FAM46C have two potential phosphoserine sites (a LIG_PLK pattern) recognized by the Polo-like kinase, which is a known human FAM46C interacting partner.

CONCLUSION

A comprehensive analysis of various biological information available in literature and databases combined with numerous sequence and structure analyses (including a state-ofthe-art distant homology detection, fold recognition and 3D modelling) allowed us to propose that FAM46 members function as cytoplasmic and/or nuclear non-canonical poly(A) polymerases. Four human FAM46 paralogs thus complement the group of already known non-canonical poly(A) polymerases in humans embracing seven proteins: RBM21 (U6 TUTase, Star-PAP, TUT6), hGLD2 (PAPD4, TUT2), hmtPAP (PAPD1, TUT1), POLS (TUT5), PAPD5 (TUT3), ZCCHC6 (TUT7) and ZCCHC11 (TUT4). ZC-CHC6 and ZCCHC11 mono-uridylate the 3' end of specific miRNAs involved in cell differentiation and Homeobox (Hox) gene control (115). The hmtPAP produces poly(A) tails in mitochondria (116). The RBM21 catalyzes the uridylation of U6 snRNA involved in pre-mRNA splicing (117). The hGLD2 generates poly(A) tails of selected cytoplasmic mRNAs (118). The PAPD5 participates in the polyadenylation-mediated degradation of aberrant prerRNA and in replication-dependent histone mRNA degradation (119). Unfortunately, we are not able to predict the exact type of RNA that can be modified by FAM46 proteins. However, taking into account all the identified FAM46 interacting partners, we can speculate that FAM46 proteins could modify the 3' end of mRNAs, small Cajal body RNAs and vRNAs. In addition, they may also participate in snoRNP and ribosome biogenesis, and (de)stabilize a nuclear pool of extra-ribosomal RPL23 and the pre-60S trans-acting factor eIF6.

The FAM46 family members as well as all the known non-canonical poly(A) polymerases share the two following domains: a PAP/OAS1 SBD with an inserted NTase domain right after the first core α -helix. In this work, we showed that proteins with such domain architecture, in addition to highly conserved NTase domain patterns ([DE]h[DE]h, h[DE]h and hG[GS]), possess also three additional, conserved amino acids critical for NTP binding. These residues embrace serine or threenine in the last α helix of the NTase domain, and lysine/arginine and a hydrophobic amino acid located in the second and third PAP/OAS1 SBD core α -helix, respectively. Although the FAM46 proteins retain serine or cysteine in the last α -helix of the NTase domain, it is possible that the conserved insertion between the last core β -strand and α -helix in FAM46 NTase domain may substitute the role of the conserved Ser/Thr at least for some family members, enabling them to catalyze the modification of selected RNA 3' ends.

We also performed a systematic search for missense mutations in human FAM46 genes, found in cancer patients. Collected mutation data from various databases and literature, combined with sequence/structure analyses suggest that, in addition to multiple myeloma, FAM46 genes may be also involved in the development of other major malignancies including lung, colorectal, hepatocellular, head and neck, urothelial, endometrial and renal papillary carcinomas and melanoma. We identified several single point mutations of highly conserved FAM46 amino acids that may affect the enzyme catalytic activity, processivity and substrate binding (e.g. by changing the preference for the type of incorporated NTP). Consequently, these mutations can lead to deregulation of specific RNAs as an oncogenic mechanism in multiple myeloma and other cancers. This is consistent with previous studies which showed a correlation between RNA deregulation (e.g. mRNA (120), microRNA (121,122), long non-coding RNA (123), small non-coding RNA (124)) and various diseases including cancers.

Summarizing, this work provides functional and structural annotation for novel and highly important enzymes involved in RNA metabolism in eukaryotes and thus may guide functional studies of these previously uncharacterized proteins. Further experimental investigations should address the predicted activity and clarify potential substrates to provide more insight into the detailed biological roles of these newly detected non-canonical poly(A) polymerases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Wayne Dawson for proofreading the manuscript.

FUNDING

Foundation for Polish Science [TEAM/2010-6 to K.G.]; Polish National Science Centre [2011/02/A/NZ2/00014 and 2014/15/B/NZ1/03357 to K.G.]; National Centre for Research and Development [INNOTECH-K2/HI2/19/184217/NCBR/13 to L.R.]; European Commission [FP7-KBBE-2011-289646 to L.R.]. A.M. and K.S. were the recipients of the fellowship from the Ministry of Science and Higher Education. Funding for open access charge: Polish National Science Centre [2014/15/B/NZ1/03357].

Conflict of interest statement. None declared.

REFERENCES

- Kuchta,K., Knizewski,L., Wyrwicz,L.S., Rychlewski,L. and Ginalski,K. (2009) Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res.*, 37, 7701–7714.
- Barragan, I., Borrego, S., Abd El-Aziz, M.M., El-Ashry, M.F., Abu-Safieh, L., Bhattacharya, S.S. and Antinolo, G. (2008) Genetic analysis of FAM46A in Spanish families with autosomal recessive retinitis pigmentosa: characterisation of novel VNTRs. *Ann. Hum. Genet.*, 72, 26–34.
- Etokebe,G.E., Bulat-Kardum,L., Munthe,L.A., Balen,S. and Dembic,Z. (2014) Association of variable number of tandem repeats in the coding region of the FAM46A gene, FAM46A rs11040 SNP and BAG6 rs3117582 SNP with susceptibility to tuberculosis. *PLoS One*, 9, e91385.
- Benjachat, T., Tongyoo, P., Tantivitayakul, P., Somparn, P., Hirankarn, N., Prom-On, S., Pisitkun, P., Leelahavanichkul, A., Avihingsanon, Y. and Townamchai, N. (2015) Biomarkers for refractory Lupus nephritis: a microarray study of kidney tissue. *Int.* J. Mol. Sci., 16, 14276–14290.
- Boyd, K. D., Ross, F.M., Walker, B.A., Wardell, C.P., Tapper, W.J., Chiecchio, L., Dagrada, G., Konn, Z.J., Gregory, W.M., Jackson, G.H. et al. (2011) Mapping of chromosome 1p deletions in myeloma

identifies FAM46C at 1p12 and CDKN2C at 1p32.3 as being genes in regions associated with adverse survival. *Clin. Cancer Res.*, **17**, 7776–7784.

- Chapman,M.A., Lawrence,M.S., Keats,J.J., Cibulskis,K., Sougnez,C., Schinzel,A.C., Harview,C.L., Brunet,J.P., Ahmann,G.J., Adli,M. *et al.* (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, **471**, 467–472.
- Kortum,K.M., Langer,C., Monge,J., Bruins,L., Zhu,Y.X., Shi,C.X., Jedlowski,P., Egan,J.B., Ojha,J., Bullinger,L. *et al.* (2015) Longitudinal analysis of 25 sequential sample-pairs using a custom multiple myeloma mutation sequencing panel (M(3)P). *Ann. Hematol.*, 94, 1205–1211.
- 8. Hamilton,S.M., Spencer,C.M., Harrison,W.R., Yuva-Paylor,L.A., Graham,D.F., Daza,R.A., Hevner,R.F., Overbeek,P.A. and Paylor,R. (2011) Multiple autism-like behaviors in a novel transgenic mouse model. *Behav. Brain Res.*, **218**, 29–41.
- Kuehl, W.M. and Bergsagel, P.L. (2012) Molecular pathogenesis of multiple myeloma and its premalignant precursor. *J. Clin. Invest.*, 122, 3456–3463.
- Barbieri, M., Manzoni, M., Fabris, S., Ciceri, G., Todoerti, K., Simeon, V., Musto, P., Cortelezzi, A., Baldini, L., Neri, A. *et al.* (2015) Compendium of FAM46C gene mutations in plasma cell dyscrasias. *Br. J. Haematol.*, doi:10.1111/bjh.13793.
- Bettoni, F., Filho, F.C., Grosso, D.M., Galante, P.A., Parmigiani, R.B., Geraldo, M.V., Henrique-Silva, F., Oba-Shinjo, S.M., Marie, S.K., Soares, F.A. *et al.* (2009) Identification of FAM46D as a novel cancer/testis antigen using EST data and serological analysis. *Genomics*, 94, 153–160.
- Colland, F., Jacq, X., Trouplin, V., Mougin, C., Groizeleau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P. and Gauthier, J.M. (2004) Functional proteomics mapping of a human signaling pathway. *Genome Res.*, 14, 1324–1332.
- Etokebe,G.E., Kuchler,A.M., Haraldsen,G., Landin,M., Osmundsen,H. and Dembic,Z. (2009)
 Family-with-sequence-similarity-46, member A (Fam46a) gene is expressed in developing tooth buds. *Arch. Oral Biol.*, 54, 1002–1007.
- Schoggins, J.W., Wilson, S.J., Panis, M., Murphy, M.Y., Jones, C.T., Bieniasz, P. and Rice, C.M. (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, 472, 481–485.
- Campbell,C.L., Torres-Perez,F., Acuna-Retamar,M. and Schountz,T. (2015) Transcriptome markers of viral persistence in naturally-infected andes virus (bunyaviridae) seropositive long-tailed pygmy rice rats. *PLoS One*, **10**, e0122935.
- Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402.
- Fu,L., Niu,B., Zhu,Z., Wu,S. and Li,W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28, 3150–3152.
- Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
- Capella-Gutierrez, S., Silla-Martinez, J.M. and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25, 1972–1973.
- Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B. *et al.* (2013) The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.*, **41**, D64–D69.
- Pierleoni, A., Martelli, P.L., Fariselli, P. and Casadio, R. (2006) BaCelLo: a balanced subcellular localization predictor. *Bioinformatics*, 22, e408–e416.
- Yu,C.S., Chen,Y.C., Lu,C.H. and Hwang,J.K. (2006) Prediction of protein subcellular localization. *Proteins*, 64, 643–651.
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*, 35, W585–W587.
- Chou,K.C. and Shen,H.B. (2010) A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS One*, 5, e9931.
- Hoglund, A., Donnes, P., Blum, T., Adolph, H.W. and Kohlbacher, O. (2006) MultiLoc: prediction of protein subcellular localization using

N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics*, **22**, 1158–1165.

- la Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K. and Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.*, 17, 527–536.
- Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Kruger, D., Grebnev, G., Kuban, M. et al. (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.*, 42, D259–D266.
- Wu,C., Macleod,I. and Su,A.I. (2013) BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res.*, 41, D561–D565.
- McCall,M.N., Uppal,K., Jaffee,H.A., Zilliox,M.J. and Irizarry,R.A. (2011) The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. *Nucleic Acids Res.*, **39**, D1011–D1015.
- Ginalski, K., von Grotthuss, M., Grishin, N.V. and Rychlewski, L. (2004) Detecting distant homology with Meta-BASIC. *Nucleic Acids Res.*, 32, W576–W581.
- Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. U. S. A.*, 95, 5857–5864.
- Finn,R.D., Tate,J., Mistry,J., Coggill,P.C., Sammut,S.J., Hotz,H.R., Ceric,G., Forslund,K., Eddy,S.R., Sonnhammer,E.L. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res.*, 36, D281–D288.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A. and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.*, 28, 33–36.
- 34. Koonin,E.V., Fedorova,N.D., Jackson,J.D., Jacobs,A.R., Krylov,D.M., Makarova,K.S., Mazumder,R., Mekhedov,S.L., Nikolskaya,A.N., Rao,B.S. *et al.* (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.*, 5, R7.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, 28, 235–242.
- Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol., 247, 536–540.
- Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol., 292, 195–202.
- Holm,L. and Sander,C. (1996) Mapping the protein universe. Science, 273, 595–603.
- Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22, 2577–2637.
- Ginalski, K., Elofsson, A., Fischer, D. and Rychlewski, L. (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*, 19, 1015–1018.
- Ginalski, K. and Rychlewski, L. (2003) Protein structure prediction of CASP5 comparative modeling and fold recognition targets using consensus alignment approach and 3D assessment. *Proteins*, 53(Suppl. 6), 410–417.
- Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol., 234, 779–815.
- Stagno, J., Aphasizheva, I., Rosengarth, A., Luecke, H. and Aphasizhev, R. (2007) UTP-bound and Apo structures of a minimal RNA uridylyltransferase. J. Mol. Biol., 366, 882–899.
- 44. Wang, Q., Canutescu, A.A. and Dunbrack, R.L. Jr (2008) SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nat. Protoc.*, **3**, 1832–1847.
- Wiederstein, M. and Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, 35, W407–W410.
- UniProt Consortium (2015) UniProt: a hub for protein information. Nucleic Acids Res., 43, D204–D212.
- Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, 34, D535–D539.
- Gene Ontology Consortium. (2015) Gene Ontology Consortium: going forward. Nucleic Acids Res., 43, D1049–D1056.

- 49. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal., 6, pl1.
- Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabe, R.R., Bhan, M.K., Calvo, F., Eerola, I., Gerhard, D.S. *et al.* (2010) International network of cancer genome projects. *Nature*, 464, 993–998.
- Gonzalez-Perez,A., Perez-Llamas,C., Deu-Pons,J., Tamborero,D., Schroeder,M.P., Jene-Sanz,A., Santos,A. and Lopez-Bigas,N. (2013) IntoGen-mutations identifies cancer drivers across tumor types. *Nat. Methods*, **10**, 1081–1082.
- Capra, J.A. and Singh, M. (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics*, 23, 1875–1882.
- 53. Lin,J. (1991) Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory*, **37**, 145–151.
- 54. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol., 59, 307–321.
- Anisimova, M. and Gascuel, O. (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.*, 55, 539–552.
- Letunic, I. and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.*, 39, W475–W478.
- Brunet, F.G., Roest Crollius, H., Paris, M., Aury, J.M., Gibert, P., Jaillon, O., Laudet, V. and Robinson-Rechavi, M. (2006) Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol. Biol. Evol.*, 23, 1808–1816.
- Peterson,K.J. and Butterfield,N.J. (2005) Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record. *Proc. Natl Acad. Sci. U.S.A.*, **102**, 9547–9552.
- 59. King,N., Westbrook,M.J., Young,S.L., Kuo,A., Abedin,M., Chapman,J., Fairclough,S., Hellsten,U., Isogai,Y., Letunic,I. *et al.* (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature*, **451**, 783–788.
- Sun,G., Yang,Z., Ishwar,A. and Huang,J. (2010) Algal genes in the closest relatives of animals. *Mol. Biol. Evol.*, 27, 2879–2889.
- Matsuzaka, Y., Tounai, K., Denda, A., Tomizawa, M., Makino, S., Okamoto, K., Keicho, N., Oka, A., Kulski, J.K., Tamiya, G. et al. (2002) Identification of novel candidate genes in the diffuse panbronchiolitis critical region of the class I human MHC. *Immunogenetics*, 54, 301–309.
- Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.P., Marth, C., Hofer, T. and Rippe, K. (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.*, **19**, 1185–1192.
- Hartmann, R., Justesen, J., Sarkar, S.N., Sen, G.C. and Yee, V.C. (2003) Crystal structure of the 2'-specific and double-stranded RNA-activated interferon-induced antiviral protein 2'-5'-oligoadenylate synthetase. *Mol. Cell*, **12**, 1173–1185.
- Kuhn,C.D., Wilusz,J.E., Zheng,Y., Beal,P.A. and Joshua-Tor,L. (2015) On-enzyme refolding permits small RNA and tRNA surveillance by the CCA-adding enzyme. *Cell*, 160, 644–658.
- 65. Zhu, D., Wang, L., Shang, G., Liu, X., Zhu, J., Lu, D., Wang, L., Kan, B., Zhang, J.R. and Xiang, Y. (2014) Structural biochemistry of a Vibrio cholerae dinucleotide cyclase reveals cyclase activity regulation by folates. *Mol. Cell*, **55**, 931–937.
- 66. Wolkowicz, U.M. and Cook, A.G. (2012) NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic Acids Res.*, 40, 9356–9368.
- Steczkiewicz,K., Muszewska,A., Knizewski,L., Rychlewski,L. and Ginalski,K. (2012) Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.*, 40, 7016–7045.
- Majorek, K.A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K. and Bujnicki, J.M. (2014) The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.*, 42, 4160–4179.
- 69. Sakon, J., Liao, H.H., Kanikula, A.M., Benning, M.M., Rayment, I. and Holden, H.M. (1993) Molecular structure of kanamycin

nucleotidyltransferase determined to 3.0-A resolution. *Biochemistry*, **32**, 11977–11984.

- Morar, M., Bhullar, K., Hughes, D.W., Junop, M. and Wright, G.D. (2009) Structure and mechanism of the lincosamide antibiotic adenylyltransferase LinB. *Structure*, **17**, 1649–1659.
- Xu, Y., Zhang, R., Joachimiak, A., Carr, P.D., Huber, T., Vasudevan, S.G. and Ollis, D.L. (2004) Structure of the N-terminal domain of Escherichia coli glutamine synthetase adenylyltransferase. *Structure*, **12**, 861–869.
- Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L. and Bohm, A. (2000) Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science*, 289, 1346–1349.
- Lunde,B.M., Magler,I. and Meinhart,A. (2012) Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res.*, 40, 9815–9824.
- Hamill, S., Wolin, S.L. and Reinisch, K.M. (2010) Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc. Natl Acad. Sci. U. S. A.*, 107, 15045–15050.
- Losón,O.C., Liu,R., Rome,M.E., Meng,S., Kaiser,J.T., Shan,S.O. and Chan,D.C. (2014) The mitochondrial fission receptor MiD51 requires ADP as a cofactor. *Structure*, 22, 367–377.
- 76. Lohofener, J., Steinke, N., Kay-Fedorov, P., Baruch, P., Nikulin, A., Tishchenko, S., Manstein, D.J. and Fedorov, R. (2015) The activation mechanism of 2'-5'-oligoadenylate synthetase gives new insights into OAS/cGAS triggers of innate immunity. *Structure*, 23, 851–862.
- 77. Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G. et al. (2013) Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell*, **153**, 1094–1107.
- Fu,H., Subramanian,R.R. and Masters,S.C. (2000) 14-3-3 proteins: structure, function, and regulation. *Annu. Rev. Pharmacol. Toxicol.*, 40, 617–647.
- Irwin,S., Vandelft,M., Pinchev,D., Howell,J.L., Graczyk,J., Orr,H.T. and Truant,R. (2005) RNA association and nucleocytoplasmic shuttling by ataxin-1. J. Cell Sci., 118, 233–242.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149, 1393–1406.
- 81. Jaakkola, P., Mole, D.R., Tian, Y.M., Wilson, M.I., Gielbert, J., Gaskell, S.J., von Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J. *et al.* (2001) Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science*, **292**, 468–472.
- Zhang, D.D. and Hannink, M. (2003) Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. *Mol. Cell. Biol.*, 23, 8137–8151.
- 83. Tsai, T.C., Lee, Y.L., Hsiao, W.C., Tsao, Y.P. and Chen, S.L. (2005) NRIP, a novel nuclear receptor interaction protein, enhances the transcriptional activity of nuclear receptors. *J. Biol. Chem.*, **280**, 20000–20009.
- Zhou,H.J., Wong,C.M., Chen,J.H., Qiang,B.Q., Yuan,J.G. and Jin,D.Y. (2001) Inhibition of LZIP-mediated transcription through direct interaction with a novel host cell factor-like protein. *J. Biol. Chem.*, 276, 28933–28938.
- Kim, J.E., Ryu, I., Kim, W.J., Song, O.K., Ryu, J., Kwon, M.Y., Kim, J.H. and Jang, S.K. (2008) Proline-rich transcript in brain protein induces stress granule formation. *Mol. Cell. Biol.*, 28, 803–813.
- Forbes, M.M., Rothhamel, S., Jenny, A. and Marlow, F.L. (2015) Maternal dazap2 regulates germ granules by counteracting dynein in zebrafish primordial germ cells. *Cell Rep.*, **12**, 49–57.
- Kedersha, N.L. and Rome, L.H. (1986) Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. J. Cell Biol., 103, 699–709.
- Kedersha, N.L., Heuser, J.E., Chugani, D.C. and Rome, L.H. (1991) Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry. *J. Cell Biol.*, **112**, 225–235.
- McKeegan,K.S., Debieux,C.M., Boulon,S., Bertrand,E. and Watkins,N.J. (2007) A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol. Cell. Biol.*, 27, 6782–6793.

- Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T. and Zheng, N. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature*, 443, 590–593.
- Rodrigo-Brenni, M.C., Gutierrez, E. and Hegde, R.S. (2014) Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol. Cell*, 55, 227–237.
- Arquint, C., Gabryjonczyk, A.M., Imseng, S., Bohm, R., Sauer, E., Hiller, S., Nigg, E.A. and Maier, T. (2015) STIL binding to Polo-box 3 of PLK4 regulates centriole duplication. *Elife*, 4, e07888.
- Cheng,A.M., Rowley,B., Pao,W., Hayday,A., Bolen,J.B. and Pawson,T. (1995) Syk tyrosine kinase required for mouse viability and B-cell development. *Nature*, **378**, 303–306.
- Carnevale, J., Ross, L., Puissant, A., Banerji, V., Stone, R.M., DeAngelo, D.J., Ross, K.N. and Stegmaier, K. (2013) SYK regulates mTOR signaling in AML. *Leukemia*, 27, 2118–2128.
- Liu, N., Li, Z., Pei, D. and Shu, X. (2013) Zfyve9a regulates the proliferation of hepatic cells during zebrafish embryogenesis. *Int. J. Dev. Biol.*, 57, 773–778.
- Fiorillo, C., Moro, F., Yi, J., Weil, S., Brisca, G., Astrea, G., Severino, M., Romano, A., Battini, R., Rossi, A. *et al.* (2014) Novel dynein DYNC1H1 neck and motor domain mutations link distal spinal muscular atrophy and abnormal cortical development. *Hum. Mutat.*, 35, 298–302.
- Vallee, R.B., McKenney, R.J. and Ori-McKenney, K.M. (2012) Multiple modes of cytoplasmic dynein regulation. *Nat. Cell Biol.*, 14, 224–230.
- 98. Hutagalung, A.H. and Novick, P.J. (2011) Role of Rab GTPases in membrane traffic and cell physiology. *Physiol. Rev.*, **91**, 119–149.
- Collins, B.M., McCoy, A.J., Kent, H.M., Evans, P.R. and Owen, D.J. (2002) Molecular architecture and functional model of the endocytic AP2 complex. *Cell*, **109**, 523–535.
- 100. Tian, M. (2010) The molecular cloning and characterization of Fam46c RNA stability factor. PhD dissertation, Harvard University.
- 101. Kranzusch, P.J., Lee, A.S., Berger, J.M. and Doudna, J.A. (2013) Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep.*, 3, 1362–1368.
- Weill,L., Belloc,E., Bava,F.A. and Mendez,R. (2012) Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat. Struct. Mol. Biol.*, 19, 577–585.
- 103. Subtelny,A.O., Eichhorn,S.W., Chen,G.R., Sive,H. and Bartel,D.P. (2014) Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, **508**, 66–71.
- Schmidt, M.J. and Norbury, C.J. (2010) Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. *Wiley Interdiscip. Rev. RNA*, 1, 142–151.
- 105. Villa, N., Do, A., Hershey, J.W. and Fraser, C.S. (2013) Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. *J. Biol. Chem.*, 288, 32932–32940.
- 106. Kuhn,U. and Wahle,E. (2004) Structure and function of poly(A) binding proteins. *Biochim. Biophys. Acta*, **1678**, 67–84.
- 107. Cunningham,K.S., Dodson,R.E., Nagel,M.A., Shapiro,D.J. and Schoenberg,D.R. (2000) Vigilin binding selectively inhibits cleavage of the vitellogenin mRNA 3'-untranslated region by the mRNA endonuclease polysomal ribonuclease 1. *Proc. Natl Acad. Sci. U. S. A.*, **97**, 12498–12502.

- 108. Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M. Jr, Tuschl, T., Ohler, U. and Keene, J.D. (2011) Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell*, **43**, 327–339.
- 109. Fan,X.C. and Šteitz,J.A. (1998) Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *EMBO J.*, **17**, 3448–3460.
- 110. Peng,S.S., Chen,C.Y., Xu,N. and Shyu,A.B. (1998) RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J.*, **17**, 3461–3470.
- Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D. and Cramer, P. (2013) RNA polymerase I structure and transcription regulation. *Nature*, **502**, 650–655.
- 112. Laferte, A., Favry, E., Sentenac, A., Riva, M., Carles, C. and Chedin, S. (2006) The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes Dev.*, 20, 2030–2040.
- 113. Wyler, E., Wandrey, F., Badertscher, L., Montellese, C., Alper, D. and Kutay, U. (2014) The beta-isoform of the BRCA2 and CDKN1A(p21)-interacting protein (BCCIP) stabilizes nuclear RPL23/uL14. FEBS Lett., 588, 3685–3691.
- 114. Venteicher, A.S., Abreu, E.B., Meng, Z., McCann, K.E., Terns, R.M., Veenstra, T.D., Terns, M.P. and Artandi, S.E. (2009) A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science*, **323**, 644–648.
- 115. Thornton, J.E., Du, P., Jing, L., Sjekloca, L., Lin, S., Grossi, E., Sliz, P., Zon, L.I. and Gregory, R.I. (2014) Selective microRNA uridylation by Zeche6 (TUT7) and Zeche11 (TUT4). *Nucleic Acids Res.*, 42, 11777–11791.
- Chang, J.H. and Tong, L. (2012) Mitochondrial poly(A) polymerase and polyadenylation. *Biochim. Biophys. Acta*, 1819, 992–997.
- 117. Trippe, R., Guschina, E., Hossbach, M., Urlaub, H., Luhrmann, R. and Benecke, B.J. (2006) Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA*, **12**, 1494–1504.
- 118. Kwak, J.E., Wang, L., Ballantyne, S., Kimble, J. and Wickens, M. (2004) Mammalian GLD-2 homologs are poly(A) polymerases. *Proc. Natl Acad. Sci. U.S.A.*, **101**, 4407–4412.
- 119. Mullen, T.E. and Marzluff, W.F. (2008) Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev.*, 22, 50–65.
- Lokody, I. (2014) RNA dynamics: destabilizing mRNAs promotes metastasis. *Nat. Rev. Cancer*, 14, 578.
- 121. Palmero, E.I., de Campos, S.G., Campos, M., de Souza, N.C., Guerreiro, I.D., Carvalho, A.L. and Marques, M.M. (2011) Mechanisms and role of microRNA deregulation in cancer onset and progression. *Genet. Mol. Biol.*, **34**, 363–370.
- 122. Jansson, M.D. and Lund, A.H. (2012) MicroRNA and cancer. *Mol. Oncol.*, **6**, 590–610.
- 123. Gao, W., Chan, J.Y. and Wong, T.S. (2014) Long non-coding RNA deregulation in tongue squamous cell carcinoma. *Biomed. Res. Int.*, 2014, 405860.
- 124. Ravo, M., Cordella, A., Rinaldi, A., Bruno, G., Alexandrova, E., Saggese, P., Nassa, G., Giurato, G., Tarallo, R., Marchese, G. *et al.* (2015) Small non-coding RNA deregulation in endometrial carcinogenesis. *Oncotarget*, 6, 4677–4691.