Additional File 1 for:

The *C. elegans* 3'UTRome V2: an updated genomic resource to study 3'UTR biology

Steber HS [1, 2], Gallante C[3], O'Brien S[2], Chiu P.-L[4], Mangone M*[1, 2].

[1]Molecular and Cellular Biology Graduate Program, School of Life Sciences 427 East Tyler Mall Tempe, AZ 85287 4501.

[2]Virginia G. Piper Center for Personalized Diagnostics, The Biodesign Institute at Arizona State University, 1001 S McAllister Ave, Tempe, AZ, USA

[3]Barrett, The Honors College, Arizona State University, 751 E Lemon Mall, Tempe, AZ 85281

[4]Center for Applied Structural Discovery, The Biodesign Institute at Arizona State University, 1001 S McAllister Ave, Tempe, AZ, USA

- To whom correspondence should be addressed. Tel: +1(480) 965-7957; Fax: +1(480) 965-3051; Email: mangone@asu.edu

Present Address: Marco Mangone, Arizona State University, Biodesign Institute Building A, 1001 S McAllister Ave, Tempe, AZ 85281 USA

This PDF includes:

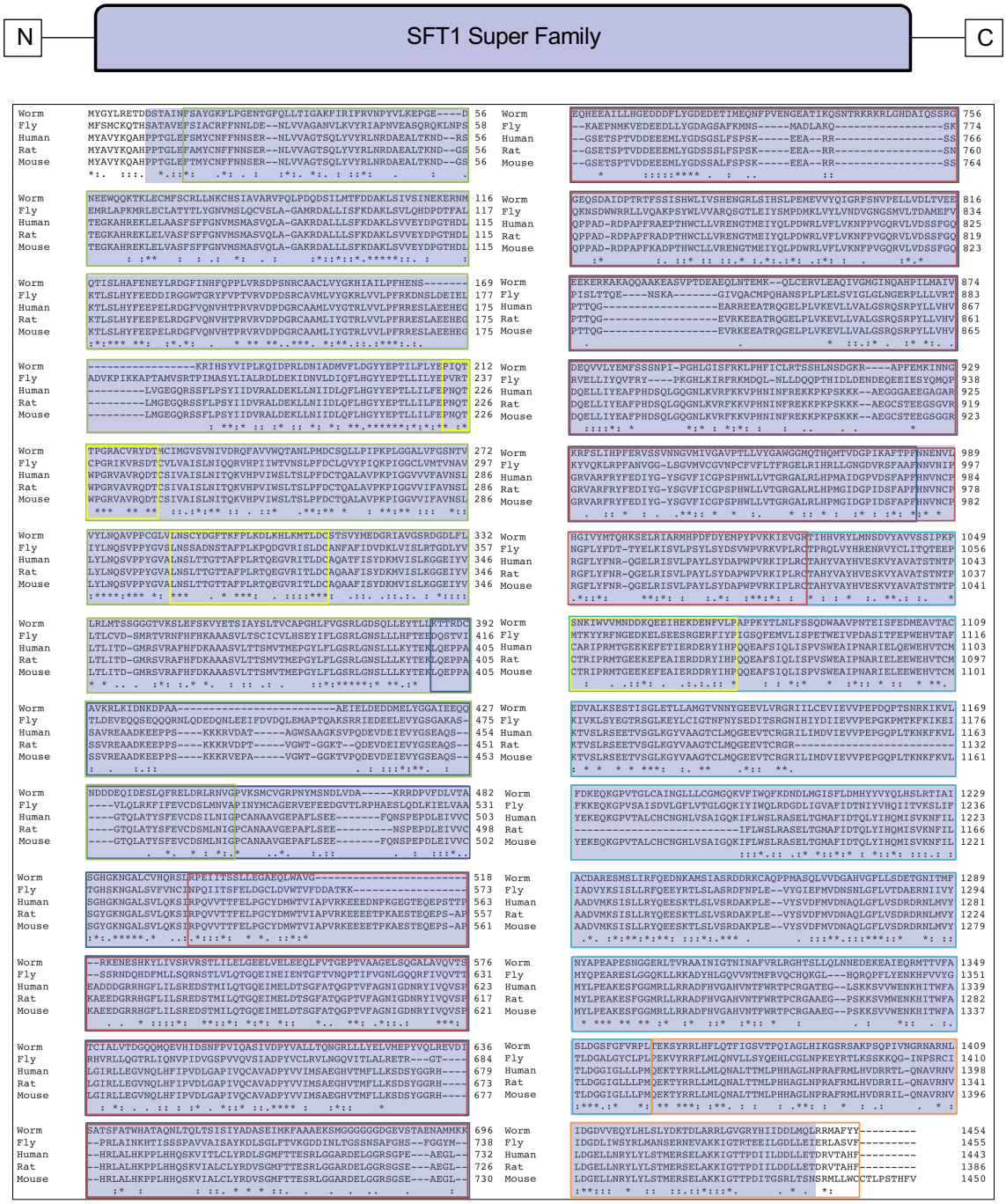
Figures S1-S9

Table of Contents

Additional File 1

CPSF Complex

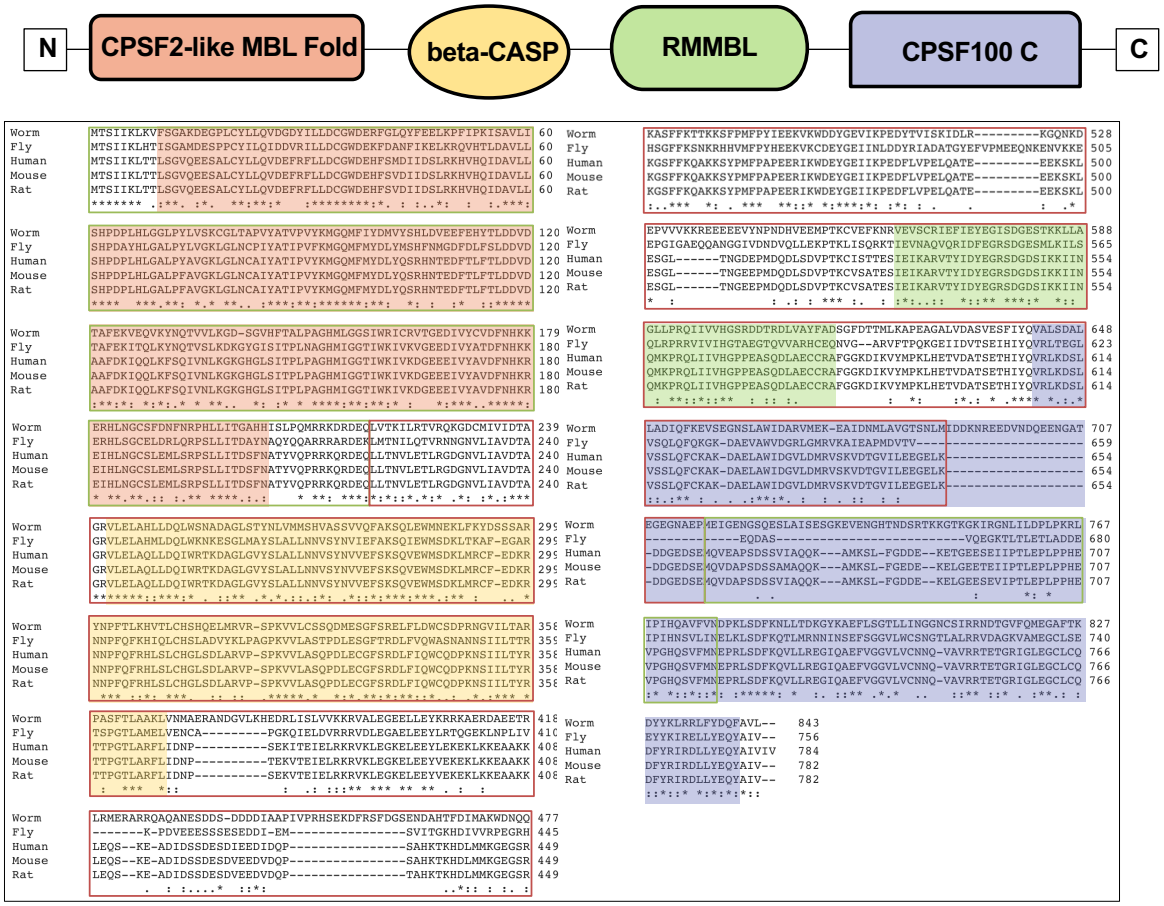
***cpsf-1* (CPSF160)**



Supplemental Fig. S1: Protein alignment of members of the CPC complex in five organisms. Amino acid sequence alignment for the members of the CPC from five organisms produced with Clustal Omega Multiple Sequence Alignment. Conserved domains produced with the Batch Conserved Domain Search on NCBI are represented by the highlighted regions of each figure. Known domains determined from previously published literature are outlined.

CPSF Complex

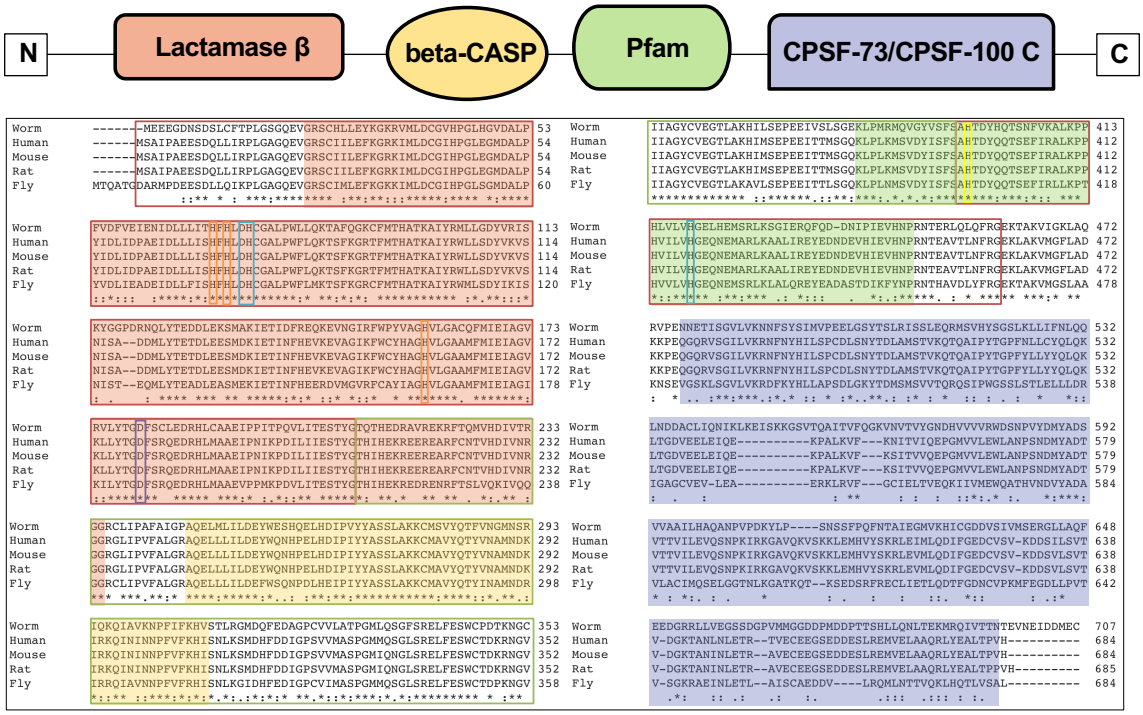
cpsf-2 (CPSF-100)



Metallo-beta-lactamase domain (Mandel et al., 2006)

Beta-CASP domain (Mandel et al., 2006)

cpsf-3 (CPSF-73)



Metallo-beta-lactamase domain (Mandel et al., 2006)

Beta-CASP domain (Mandel et al., 2006)

Bridging ligand to Zinc atoms (Mandel et al., 2006)

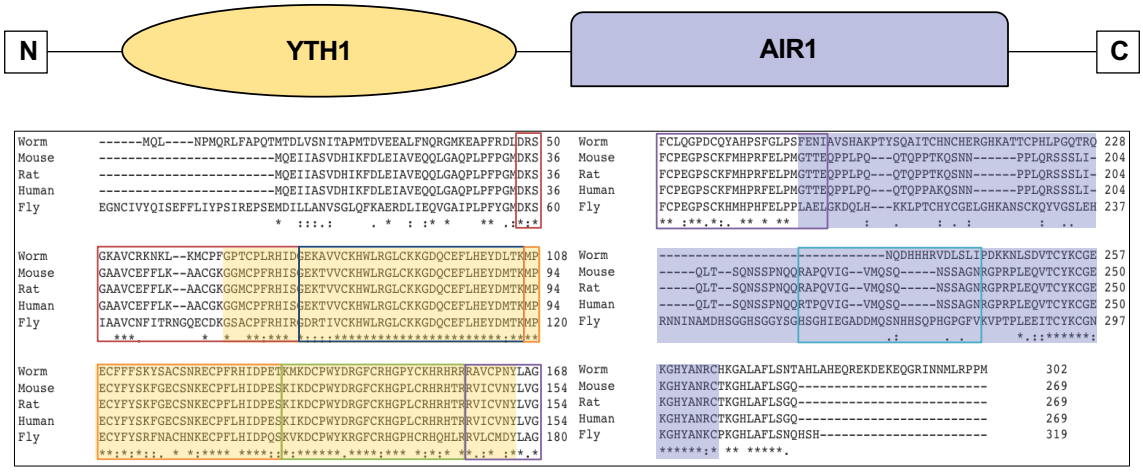
Ligands to Zn1 (Mandel et al., 2006)

Ligands to Zn2 (Mandel et al., 2006)

General acid for catalysis (Mandel et al., 2006)

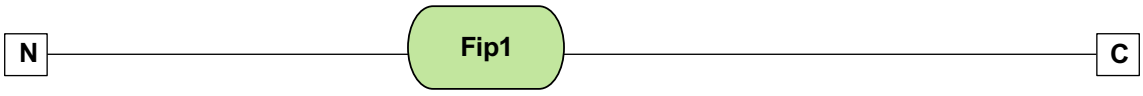
CPSF Complex

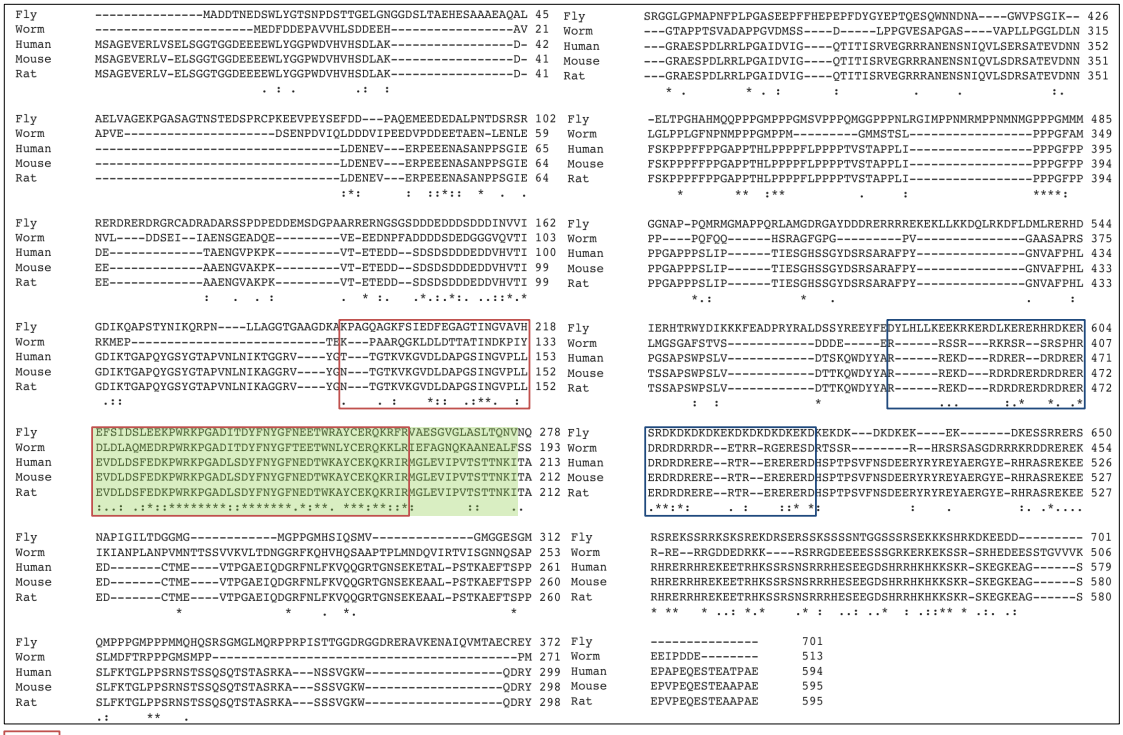
cpsf-4 (CPSF-30)



Zinc Finger 1 (Clerici et al., 2017)
Zinc Finger 2, contacts A1 and A2 of PAS (Clerici et al., 2017, Sjun et al., 2018)
Zinc Finger 3, contacts A4 and A5 of PAS (Clerici et al., 2017, Sjun et al., 2018)
Contact Wdr-33 and CPSF-160
Zinc Finger 4 (Clerici et al., 2017)
Zinc Finger 5 (Clerici et al., 2017)
Contact Fip-1
Zinc Knuckle (Clerici et al., 2017)

fipp-1 (FIP-1)

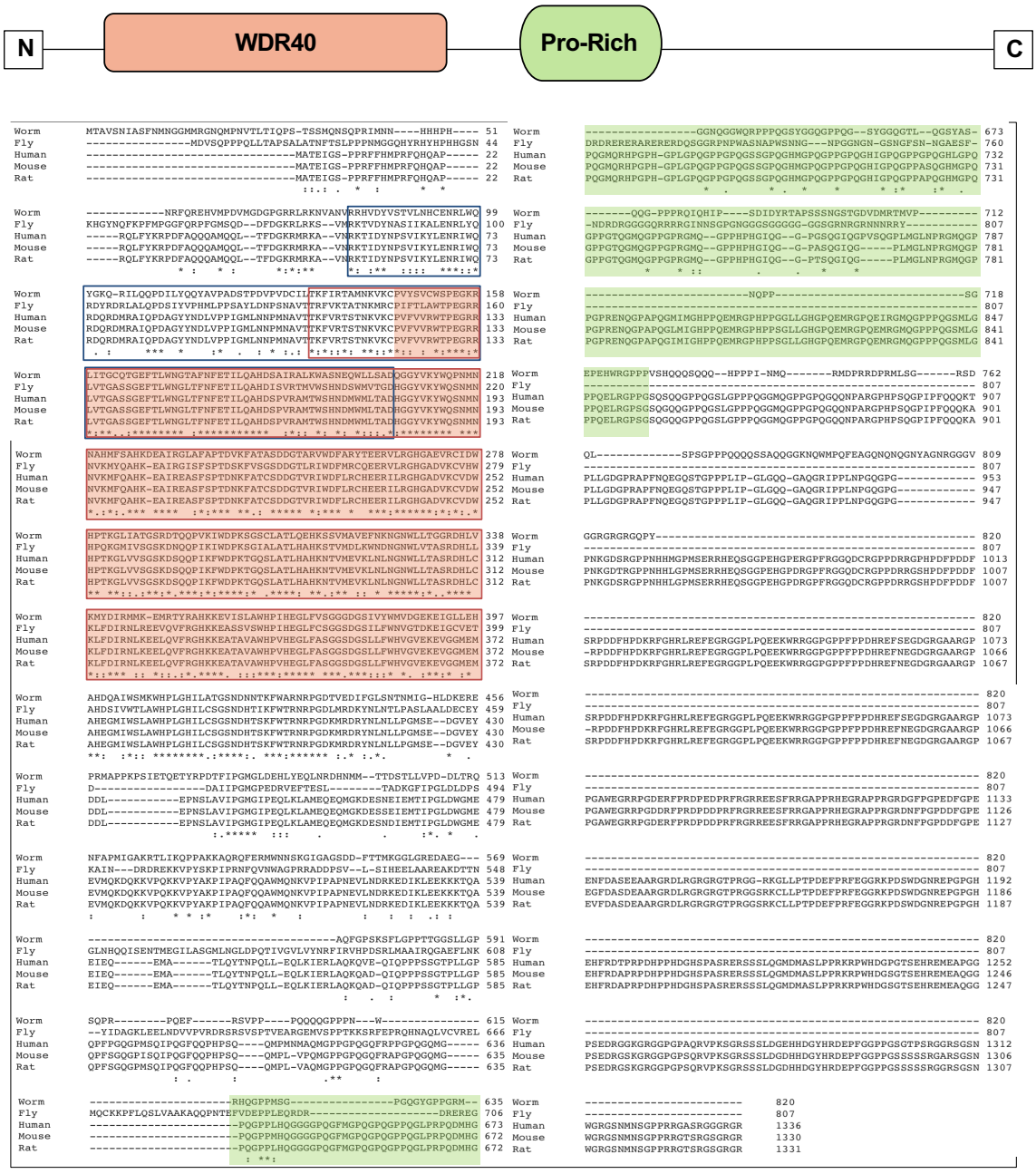




Conserved Domain contacts CPSF-30 (Clerici et al., 2017)
RE/D region interacts with CFIm68/59 (Zhu et al., 2018)

CPSF Complex

pfs-2 (WDR33)

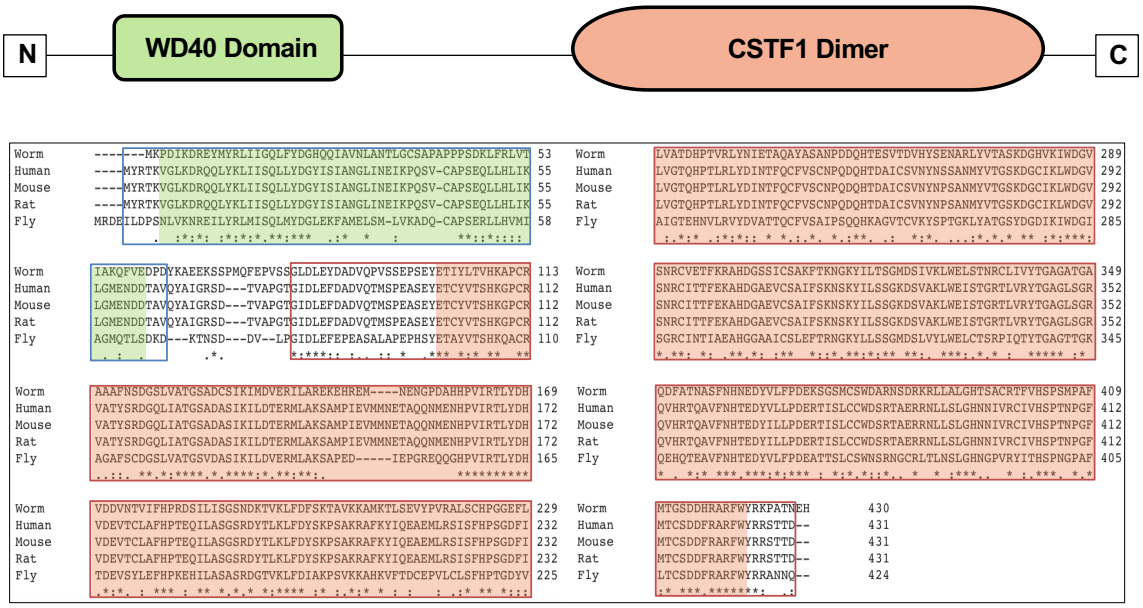


N-terminal domain, contacts CPSF-160 and CPSF-30 (Clerici et al., 2017)

WD40 repeats, contact U3 and A6 of PAS (Clerici et al., 2017)

CstF Complex

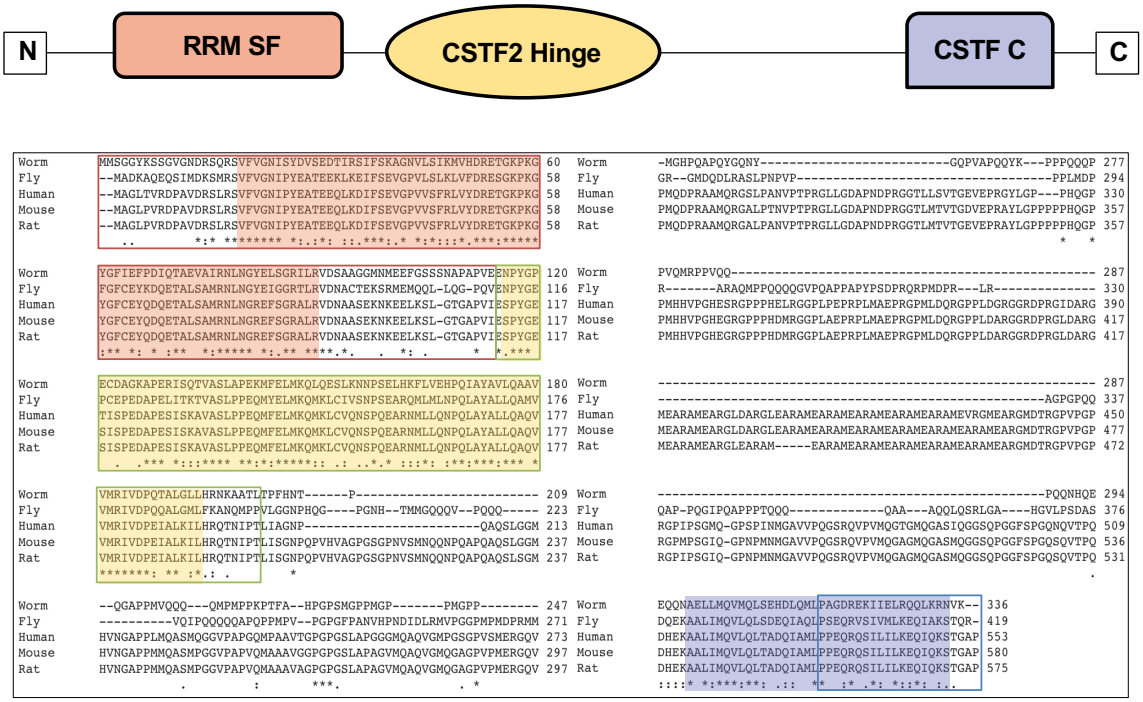
cpf-1 (CstF50)



WD40 Repeats (Yang et al., 2018)

N-terminal domain (Yang et al., 2018)

cpf-2 (CstF64)



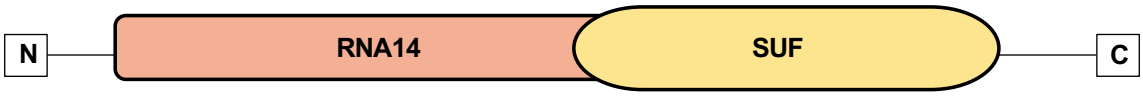
RRM domain (Yang et al., 2018)

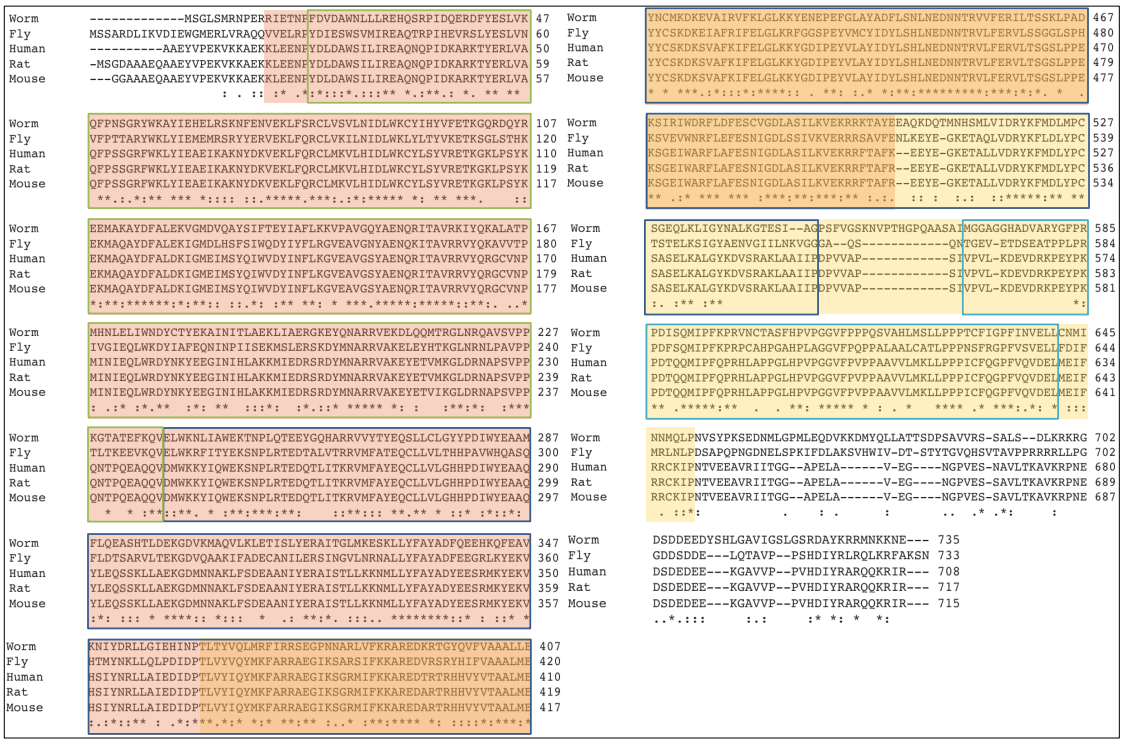
Hinge domain (Yang et al., 2018)

C-terminal domain (Yang et al., 2018)

CstF Complex

***suf-1* (CstF77)**





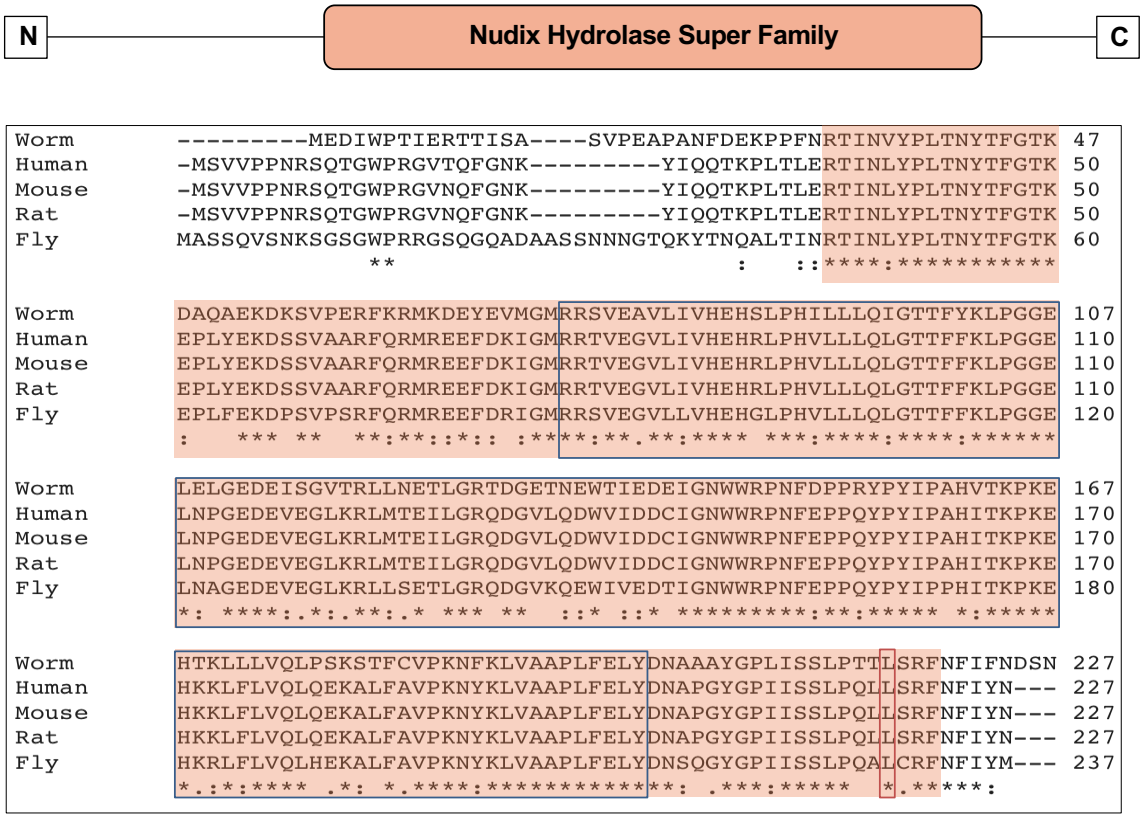
Proline-rich segment that facilitates interactions with CstF-64 and CstF-55 (Bai et al., 2007)

HAT-N Domain (Bai et al., 2007)

HAT-C Domain that interacts with CPSF-160 (Bai et al., 2007)

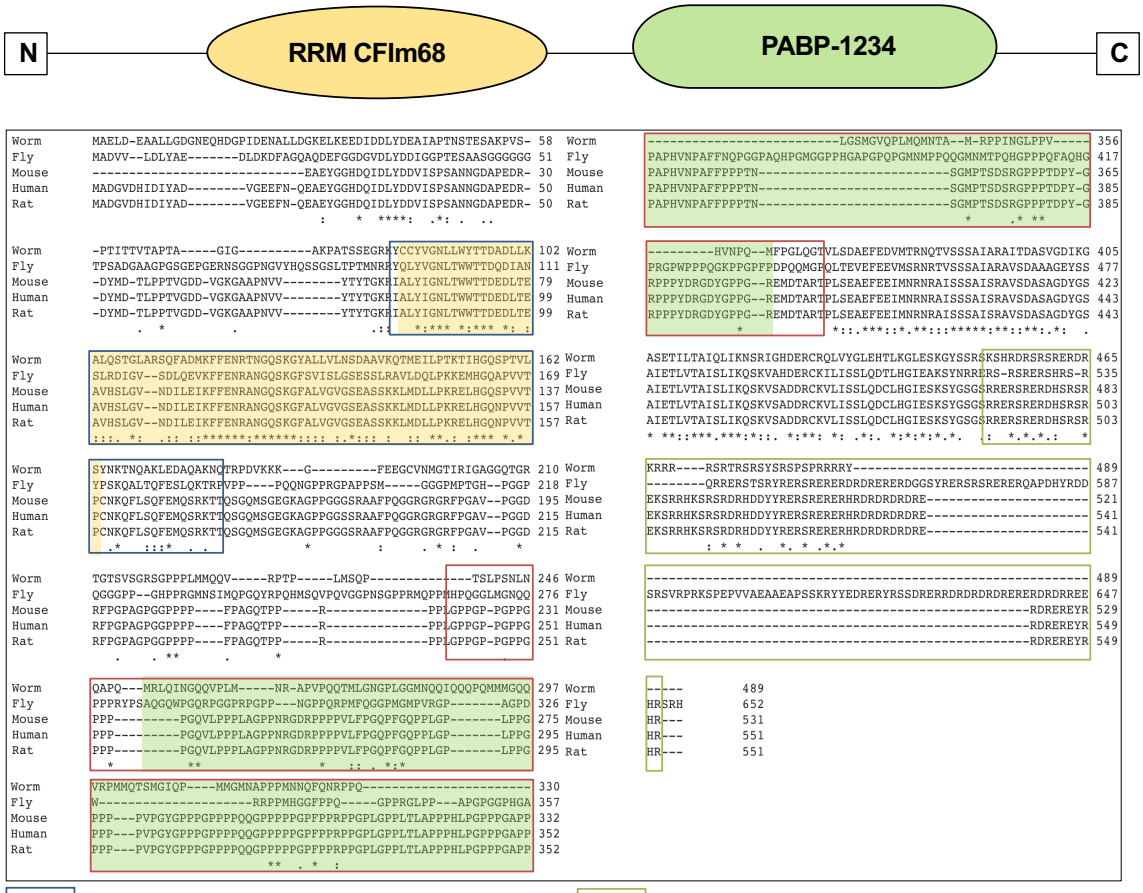
CFIm Complex

***cfim-1* (CFim25)**



Nudix domain (Yang et al., 2011) CFIm68 and CFIm59 tethering site (Zhu et al.,2017)

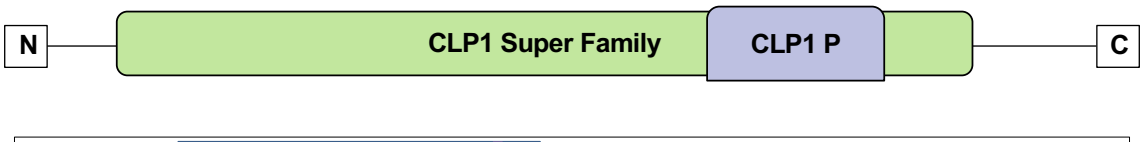
***cfim-2* (CFim68)**

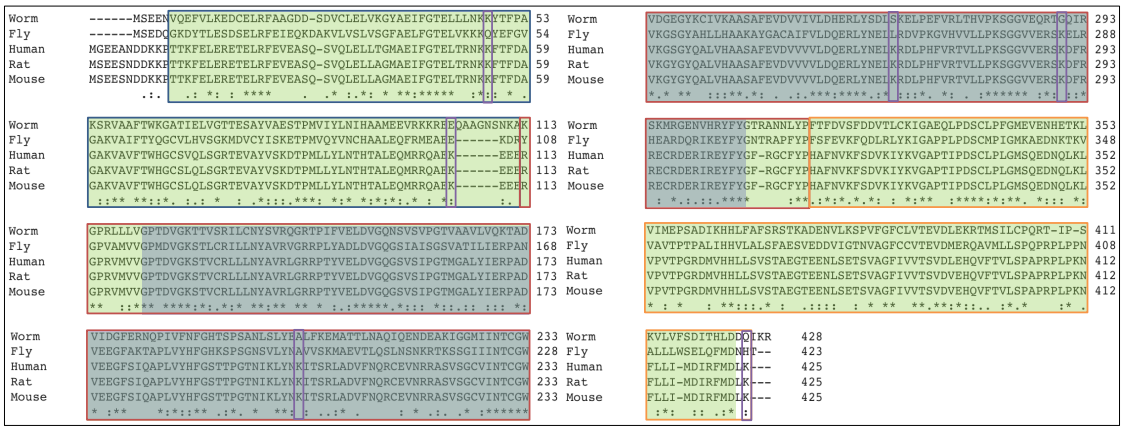


RRM domain (Yang et al., 2011, Martin et al., 2010) RS/RD/RE region (Yang et al., 2011)

Pro/Gly-rich region (Yang et al., 2011)

***clpf-1* (CLP1)**





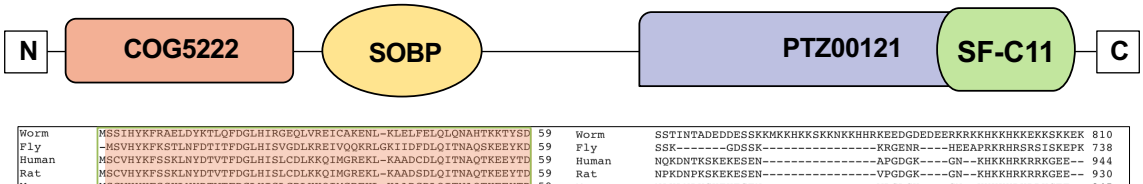
N-terminal domain (Dikfidan et al., 2014)

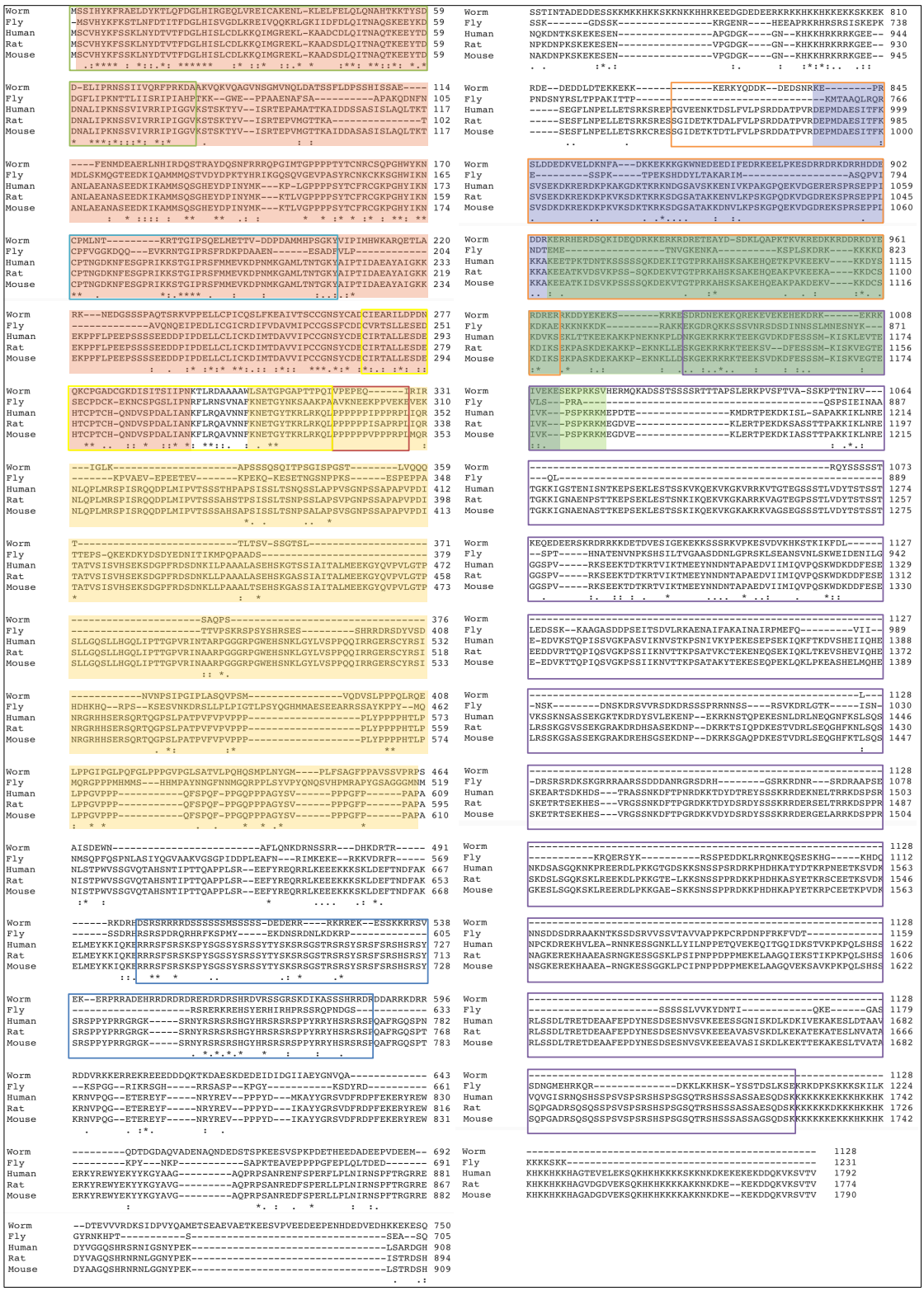
Polynucleotide kinase domain (Dikfidan et al., 2014)

C-terminal domain (Dikfidan et al., 2014)

Residues that crosslink to Pcf-11 (Schäfer et al., 2018)

rbpl-1 (RBBP6)





Domain With No Name (DWNN), ubiquitin-like (Pugh et al., 2006)

p53-binding domain (Pugh et al., 2006)

RING finger domain (Chibi et al., 2008, Lee et al., 2014)

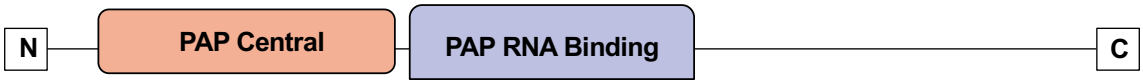
Proline-rich domain (Pugh et al., 2006)

SR domain (Pugh et al., 2006)

Rb-binding domain (Pugh et al., 2006)

Zinc knuckle (Lee et al., 2014)

***pap-1* (PAP1N)**



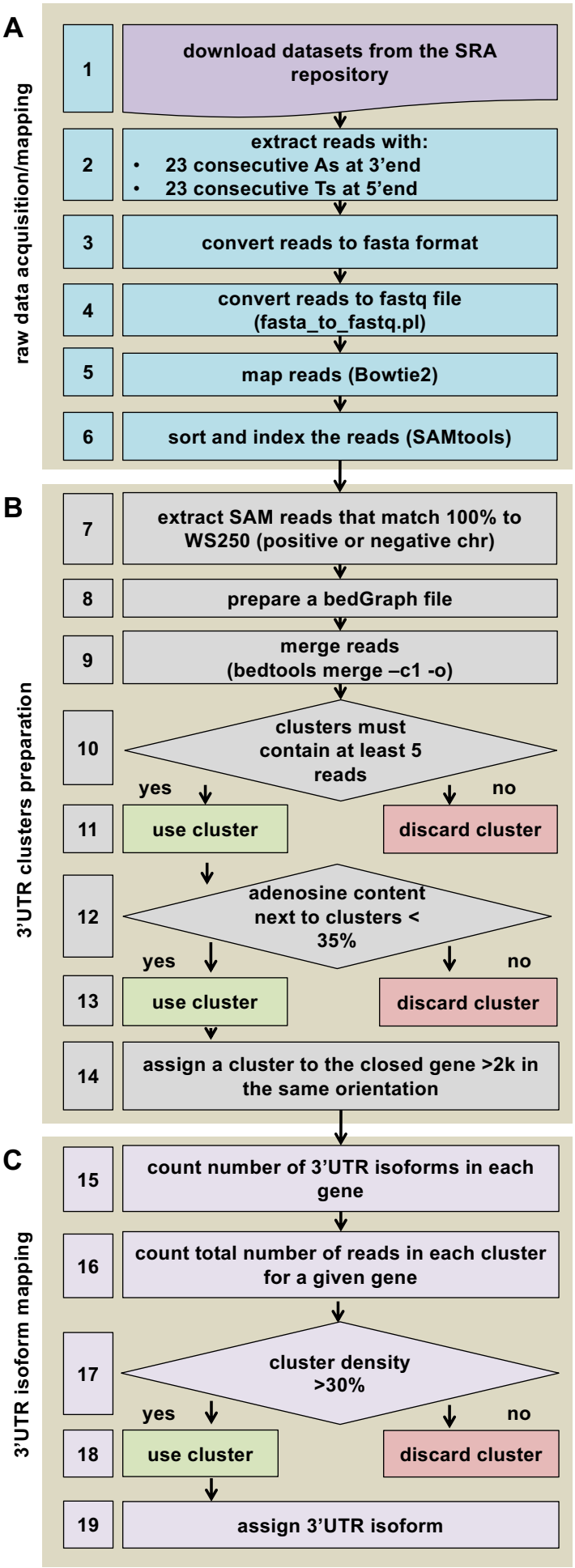

```
Worm     ----------------MSATEKDKTPLLGVSQPISLAHPDSKDIAQTTLLIETLKKFGSY  44   Worm     AEHA----KTLDLTNEIQRFKTNVELQASNVKGIGPNCQVQIDMFYVKRNSLIQVISAAD 499
Fly      MWNSEPTHRQHHQHNGNSTSGGPPAKQLGMTSAISLAEPRPEDLQRTDELRGSLEPYNVF  60   Fly      ER---SENLNVDLTESIQNFTEHVMMHGVNIKMLKEG--MTIDARHVKRKQLSLYLDSDF 530
Human    --------MPFPVTTQGSQQTQPPQKHYGITSPISLAAPKETDCVLTQKLIETLKPFGVF  52   Human    KKTENSENLSVDLTYDIQSFTDTVYRQAINSKMFEVD--MKIAAMHVKRKQLHQLLPNHV 501
Mouse    --------MPFPVTTQGSQQTQPPQRHYGITSPISLAAPKETDCLLTQKLIETLKPFGVF  52   Mouse    KKTENSENLSVDLTYDIQSFTDTVYRQAINSKMFELD--MKIAAMHVKRKQLHQLLPSHV 501
Rat      --------MPFPVTTQGSQQTQPPQRHYGITSPISLAAPKETDCLLTQKLIETLKPFGVF  52   Rat      KKTENSENLSVDLTYDIQSFTDTVYRQAINSKMFELD--MKIAAMHVKRKQLHQLLPSHV 501
                         *           *::. **** *   *   *  *  :*: :. :          .       .:*** .** *.  *  :. * * :  .  : *   :***:.*   :

Worm     EPKEETEQRMEVLRNLNRLVKEWVKNVTAM-KIPNGEGVNAGGKLFTFGSYRLGVHSSGA 103   Worm     LRRGRRWKKVVPIATNTSVSSSTPRSVVRTTSTSSVPTTPTGLAAPKTPLSASVSATNEP 559
Fly      ESQDELNHRMEILAKLNTLVKQWVKEISVSKNMPESAAEKLGGKIYTFGSYRLGVHHKGA 120   Fly      LKRERKSMESHNNFNNTLLAN-RKRLSTELAQSQ-DPLPP-GQQ-------PSSGN---- 576
Human    EEEEELQRRILILGKLNNLVKEWIREISESKNLPQSVIENVGGKIFTFGSYRLGVHTKGA 112   Human    LQKKKKHSTEGVKLT--ALNDSSLDLSMDSDNSMSVPSPT-SAT-KTSPLNSSGSS---- 553
Mouse    EEEEELQRRILILGKLNNLVKEWIREISESKNLPQSVIENVGGKIFTFGSYRLGVHTKGA 112   Mouse    LQKRKKHSTEGVKLT--ALNDSSLDLSMDSDNSMSVPSPT-SAM-KTSPLNSSGSS---- 553
Rat      EEEEELQRRILILGKLNNLVKEWIREISESKNLPQSVIENVGGKIFTFGSYRLGVHTKGA 112   Rat      LQKRKKHSTEGVKLA--ALNDSSLDLSMDSDNSMSVPSPT-SAM-KTSPLNSSGSS---- 553
         * ::* ::*: :* :** ***:*:::::   ::*:.   : ***::********** .**          *:: ::            : .          .:   *    .          * .

Worm     DIDTLAVVPRHIDRSDFFTSFKEMLNNDPNVTELHGVEEAFVPVMKLKYSGVELDILFAR 163   Worm     DSTTNGTPLSRKR-SMDEES---------STTVTSQISDESVPKKKTRDDTLEENRVSMV 609
Fly      DIDALCVAPRNIERTDYFQSFFEVLKKQPEVTECRSVEEAFVPVIKMNFDGIEIDLLFAR 180   Fly      RGRDSGAKIQRLSDSLTEENSNASSDMGAGTPTTPTTAQLSAPSFKSSGKNGSE------ 630
Human    DIDALCVAPRHVDRSDFFTSFYDKLKLQEEVKDLRAVEEAFVPVIKLCFDGIEIDILFAR 172   Human    QGRNSPAPAV-----------------TAASVTNIQATEVSVPQVNSSESSG-------- 588
Mouse    DIDALCVAPRHVDRSDFFTSFYDKLKLQEEVKDLRAVEEAFVPVIKLCFDGIEIDILFAR 172   Mouse    QGRNSPAPAV-----------------TAASVTSIQASEVSVPQANSSESPG-------- 588
Rat      DIDALCVAPRHVDRSDFFTSFYDKLKLQEEVKDLRAVEEAFVPVIKLCFDGIEIDILFAR 172   Rat      QGRNSPAPAV-----------------TAASVTSIQASEVSVPQANSSESPG-------- 588
         ***:*.*.**:::*:*:* ** : *: : :*.: :.********:*: :.*:*:*:**** .  . :                    .: ..   :: *.*. ::  .

Worm     LALKEVPDTQELSDDNLLRNLDQESVRSLNGCRVAEQLLKLVPRQKEFCVTLRAIKLWAK 223   Worm     VEVSNVVVEQRTKVVQEIVDLQADNGLNTSNGLEASEQKMEVPQSV-------------- 655
Fly      LSLKEIPDDFDLRDDNLLRNLDHRSVRSLNGCRVTDEILALVPNIENFRLALRTIKLWAK 240   Fly      -----------IDVVEQEPTQPHNNGNASSN-----TTTTEVACSXQPPTLHPHPPHQQP 674
Human    LALQTIPEDLDLRDDSLLKNLDIRCIRSLNGCRVTDEILHLVPNIDNFRLTLRAIKLWAK 232   Human    ----------------------------GTSS---------ESIPQTATQP-------AISP 606
Mouse    LALQTIPEDLDLRDDSLLKNLDIRCIRSLNGCRVTDEILHLVPNIDNFRLTLRAIKLWAK 232   Mouse    ----------------------------GPSS---------ESIPQTATQP-------AISP 606
Rat      LALQTIPEDLDLRDDSLLKNLDIRCIRSLNGCRVTDEILHLVPNIDNFRLTLRAIKLWAK 232   Rat      ----------------------------GPSS---------ESIPQTATQP-------AISP 606
         *:*: :*:  :* **.**:*** ..:********:::::* ***. .:* ::**:******                                        :*          .:  :

Worm     NHGIYSNSMGFFGGITWAILVARACQLYPNASPSRLVHRMFFIFSTWTWPHPVVLNEMNN 283   Worm     ---------------------------------------------------------- 655
Fly      KHGIYSNSLGYFGGVTWAMLVARTCQLYPNAAAATLVHKFFLVFSRWKWPNPVLLKHPDN 300   Fly      PPAPPQPPAHHHH---HQKFRKNNNAAAAAASNSIFNQRSILHAASSLSAALSITGHKRK 731
Human    RHNIYSNILGFLGGVSWAMLVARTCQLYPNAIASTLVHKFFLVFSKWEWPNPVLLKQPEE 292   Human    PPKPTVSRVVSSTRLVNPPPRSSGNAATSG--NA----------AT--KIPTPIVGVKRT 652
Mouse    RHNIYSNILGFLGGVSWAMLVARTCQLYPNAIASTLVHKFFLVFSKWEWPNPVLLKQPEE 292   Mouse    PPKPTVSRVVSSTRLVNPSPRPSGN--------T----------AT--KVPNPIVGVKRT 646
Rat      RHNIYSNILGFLGGVSWAMLVARTCQLYPNAIASTLVHKFFLVFSKWEWPNPVLLKQPEE 292   Rat      PPKPTVSRVVSSTRLVNPSPRPSGN--------A----------AT--KVPNPIVGVKRT 646
         .*.**** :*::**::**:****:*******  : ***::*::** * **:**:*:. ::

Worm     DRNDIPTLCELVWDPRRKNTDRFHVMPIITPAFPEQNSTHNVTRSTATVIKNEICEALEI 343   Worm     ---------------------------------------------------------- 655
Fly      V-----NLRFQVWDPRVNASDRYHLMPIITPAYPQQNSTFNVSESTKKVILTEFNRGMNI 355   Fly      QQTATSTTMISSSNIAHSNYTQHSTHNNNGGGGGSHYNKTYYIDDDDEDNNQPLLTPNNQ 791
Human    C-----NLNLPVWDPRVNPSDRYHLMPIITPAYPQQNSTYNVSVSTRMVMVEEFKQGLAI 347   Human    SSPHK-------------EESPKKTKT----------EEEQLDTET---------STTQ 679
Mouse    C-----NLNLPVWDPRVNPSDRYHLMPIITPAYPQQNSTYNVSVSTRMVMVEEFKQGLAI 347   Mouse    SSPNK-------------EESPKKTKT----------EEEQVDLET---------SAVQ 673
Rat      C-----NLNLPVWDPRVNPSDRYHLMPIITPAYPQQNSTYNVSVSTRMVMVEEFKQGLAI 347   Rat      SSPHK-------------EESPKKTKT----------EEEQVGMET---------SAAP 673
               .*   ***** : :**:*:*******:*:****.**: **  *:  *: ..: *

Worm     CRDISEGKSKWTALFEEVNFFSRYKHFIALIMAAPNEEEELNYGGFLESRIRLLVQSLER 403   Worm     ----------------------------------------------      655
Fly      TDEIMLGRIPWERLFEAPSFFYRYRHFIVLLVNSQTADDHLEWCGLVESKVRLLIGNLER 415   Fly      TQQQQKAVRITA---TSASPATSVPVAASAAPPAPTISL--------      827
Human    TDEILLSKAEWSKLFEAPNFFQKYKHYIVLLASAPTEKQRLEWVGLVESKIRILVGSLEK 407   Human    SETIQTAASLLASQKTSSTDLSDIP--ALPANPIPVIKNSIKLRLNR      724
Mouse    TDEILLSKAEWSKLFEAPNFFQKYKHYIVLLASAPTEKQRLEWVGLVESKIRILVGSLEK 407   Mouse    SETVPASASLLASQKTSSTDLSDIP--ALPANPIPVIKNSIKLRLNR      718
Rat      TDEILLSKAEWSKLFEAPNFFQKYKHYIVLLASAPTEKQRLEWVGLVESKIRILVGSLEK 407   Rat      SEAAPASASLVASQKTSSTDLSDIP--ALPANPIPVIKNSIKLRLNR      718
           :*  .:  *  *** .** :*:*:*.*:  : . .:.*:: *::**::*:*: .**:

Worm     NQDIIIAHNDPNKHKPSPNAKF--------------------DVNPENKRVTVWFIGLEF 443
Fly      NPHIALAHVNPKCFEFKKGQSANNSQNNSGNEDDLKQSQGNQSAVTSAPFCSMWFIGLEF 475
Human    NEFITLAHVNPQSFPAPKENP------------------------DKEEFRTMWVIGLVF 443
Mouse    NEFITLAHVNPQSFPAPKESP------------------------DREEFRTMWVIGLVF 443
Rat      NEFITLAHVNPQSFPAPKESP------------------------DREEFRTMWVIGLVF 443
         *  * :** :*: .                                  ::*.*** *
```

***symk-1* (Symplekin)**

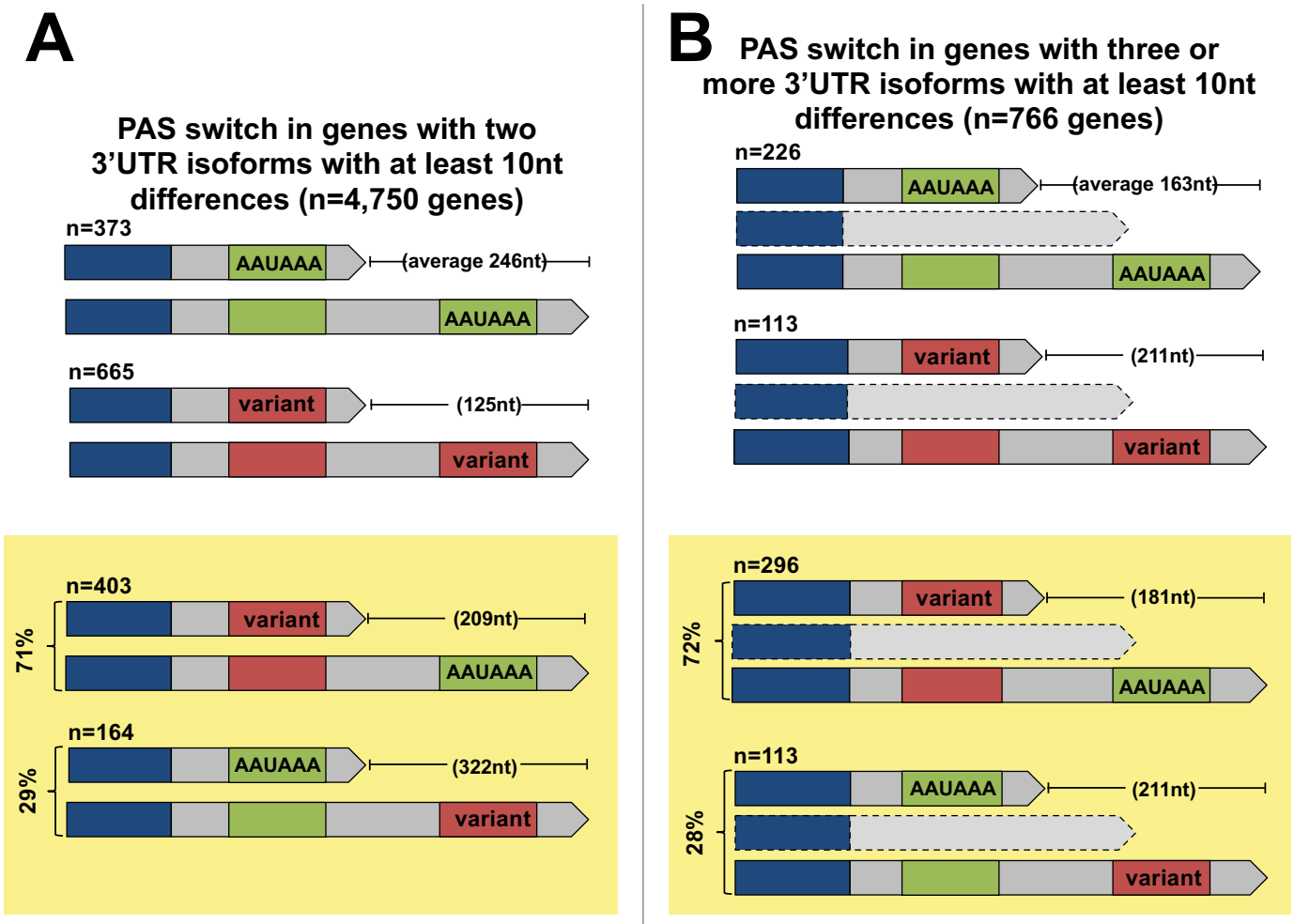

```
Worm      MDYIQG------------------LNEENETASERIGEALKEARDAETIEKKLLSLSTAM  42  Worm      -FYKVLLETPLLTPNAIERLKQVCLAKE-NEHGMAMLRELIMTRNRQRPQLLQFLFGLFF 730
Fly       MDSIIGRSQFVS-ETANLFTDEK-----TATARAKVVDWCNELVIASP-STKCELLAKVQ  53  Fly       LIRRVYLEAPILPEVSIGHLVQLSLDDEFSQHGLELIKDLAVLRPPRKNRFVRVLLNFSV 718
Human     MASGSGDSVTRRSVASQFFTQEEGPGIDGMTTSERVVDLLNQAALITN-DSKITVLKQVQ  59  Human     -FTKVVLEAPLITESALEVVRKYCEDESRTYLGMSTLRDLIFKRPSRQFQYLHVLLDLSS 722
Mouse     MASSSGDSVTRRSVASQFFTQEEGPSIDGMTTSERVVDLLNQAALITN-DSKITVLKQVQ  59  Mouse     -FTKVVLEAPLITESALEVIRKYCEDESRAYLGMSTLGDLIFKRPSRQFQYLHVLLDLSS 722
Rat       MASGSGDSVTRRSVASQFFTQEEGPGIDGMTTSERVVDLLNQAALITN-DSKITVLKQVQ  59  Rat       -FTKVVLEAPLITESALEVIRKYCEDESRAYLGMSTLGDLIFKRPSRQFQYLHVLLDLSS 722
          *    *                        *:  :: :  ::       ..*   *  .                 : :* **:*::   ::  : : .  ..    *:  : :* . *  :: : ::.*:.:

Worm      HLLIDPSLSISILDNFLTEMLEFAELNDSRILCLLVDFLLKASAKDFTLCNKTVERYSFY 102  Worm      MERPELRSSCLEVVKELCYLPF-IRSSLSDQARMQIHDCLQESPPMYMRSS-----EDSD 784
Fly       ETVL--GSCAELAEEFLESVLSLAHDSNMEVRKQVVAFVEQVCKVKVELLPHVINVVSML 111  Fly       HERLDLRDLAQAHLVSLYHVHKILPARIDEFALEWLKFIEQESPPAAVFSQDFGRPTEEP 778
Human     ELII--NKDPTLLDNFLDEIIAFQADKSIEVRKFVIGFIEEACKRDIELLLKLIANLNML 117  Human     HEKDKVRSQALLFIKRMYEKE-QLREYVEKFALNYLQLLVHPNPPSVLFGADKDTE-VAA 780
Mouse     ELII--NKDPTLLDNFLDEIIAFQADKSIEVRKFVIGFIEEACKRDIELLLKLIANLNML 117  Mouse     HEKDRVRSQALLFIKRMYEKE-QLREYVEKFALNYLQLLVHPNPPSVLFGADKDTE-VAA 780
Rat       ELII--NKDPTLLDNFLDEIIAFQADKSTEVRKFVIGFIEEACKRDIELLLKLIANLNML 117  Rat       HEKDRVRSQALLFIKRMYEKE-QLREYVEKFALNYLQLLVHPNPPSVLFGADKDTE-VAA 780
          . ::  .    : ::** .:: :   .. .:   :: *: :..  .. *  : :   .:                *:  :*. .   :  :      :   :.. *   ::   : .**  : .

Worm      LIPNKSIKRYESVIKRVVVASTNLYPIVLEFA--IMDKNDNAESCWDAFNLLKNRICMLV 160  Worm      QWTDEMYKNSLAVYSTLMPSDPL-LLIPLASVYAQSTNVFKRVVLRSLEPVFRQLSQ--E 841
Fly       LRD-----NSAQVIKRVIQACGSIYKNGLQYLCSLMEPGDSAEQAWNILSLIKAQILDMI 166  Fly       DWREDTTKVCFGLAFTLLPYKPEVYLQQICQVFVSTSAELKRTILRSLDIPIKKMGVESP 838
Human     LRD-----ENVNVVKAILTMTQLYKVALQWMVKSRVISELQEACWDMVSAMAGDIILLL 172  Human     PWTEETVKQCLYLYLALLPQNHK-LIHELAAVYTEAIADIKRTVLRVIEQPIRGMGMNSP 839
Mouse     LRD-----ENVNVVKAILTMTQLYKVALQWMVKSRVISDLQEACWDMVSSMAGEIILLL 172  Mouse     PWTEETVKQCLYLYLALLPQNHK-LIHELAAVYTEAIADIKRTVLRVIEQPIRGMGMNSP 839
Rat       LRD-----ENVNVVKAILTMTQLYKVALQWMVKSRVISDLQEACWDMVSSMAAEIILLL 172  Rat       PWTEETVKQCLYLYLALLPQNHK-LIHELAAVYTEAIADIKRTVLRVIEQPIRGMGMNSP 839
          *       .  .*:*:.: :  .:*   *::        .:  * .*: .. :   *  ::                * ::  * .: :  :*:* .    :  :. *:..:   :**.:** ::  :: :.

Worm      SDDHEGVRTVTVKFLEALILCQSPKPRELATGSNISWAREANTRFNRISLSDVPRSHRFL 220  Worm      MVISLIEDCPYGAETLVARLVVLLTERIT-PSTDLIQKLKILHDERKMDIRALLPIIGGL 900
Fly       DNENDGIRTNAIKFLEGVVVLQSFADED-----------SLKRDGDFSLADVPDHCTLF 214  Fly       TLLQLIEDCPKGMETLVIRIIYILTERVPSPHEELVRRVRDLYQNKVKDVRVMIPVLSGL 898
Human     DSDNDGIRTHAIKFVEGLIVTLSPRMADSE----I-----PRRQEHDISLDRIPRDHPYI 223  Human     ELLLLVENCPKGAETLVTRCLHSLTDKVP-PSPELVKRVRDLYHKRLPDVRFLIPVLNGL 898
Mouse     DSDNDGIRTHAIKFVEGLIVTLSPRMADSE----V-----PRRQEHDISLDRIPRDHPYI 223  Mouse     ELLLLVENCPKGAETLVTRCLHSLTDKVP-PSPELVKRVRDLYHKRLPDVRFLIPVLNGL 898
Rat       DSDNDGIRTHAIKFVEGLIVTLSPRMADSE----V-----PRRQEHDISLDRIPRDHPYI 223  Rat       ELLLLVENCPKGAETLVTRCLHSLTDKVP-PSPELVKRVRDLYHKRLPDVRFLIPVLNGL 898
          ..:::*:** ::**:*.:::  *    :                :   :**  :*     :                 :: *:*:** * **** * :  **:::  *  :*::::: *:.::  *:* ::*::.**

Worm      SYHKTQLEAEENFSALLKQTTVAEATSQNLITVIESLCMITRCRPQWENALPRVFDVIKA 280  Worm      EREEVVRLIPTFIFRAEYQKSVNVLFRKLYTVRDPQ--TGNLVFDPIEVIKEYHKIEPKN 958
Fly       RREKLQEEGNNILDILLQFHGTTHISSVNLIACTSSLCTIAKMRPIFMGAV---VEAFKQ 271  Fly       TRSELISVLPKLIKLNPA--VVKEVFNRLLGIGAEFAHQ-TMAMTPTDILVALHTIDTSV 955
Human     QYNVLWEEGKAALEQLLKFMVHPAISSINLTTALGSLANIARQRPMFMSEV---IQAYET 280  Human     EKKEVIQALPKLIKLNPI--VVKEVFNRLLGTQHGEGNSALSPLNPGELLIALHNIDSVK 956
Mouse     QYNVLWEEGKAAVEQLLKFMVHPAISSINLTTALGSLANIARQRPMFMSEV---IQAYET 280  Mouse     EKKEVIQALPKLIKLNPI--VVKEVFNRLLGTQHGEGNSALSPLNPGELLIALHNIDSVK 956
Rat       QYNVLWEEGKAALEQLLKFMVHPAISSINLTTALGSLANIARQRPMFMSEV---IQAYET 280  Rat       EKKEVIQALPKLIKLNPI--VVKEVFNRLLGTQHGEGNSALSPLNPGELLIALHNIDSVK 956
            .    *.:  .. **:       :* ** :   **. *:: ** : . :   .:. :                 :.*::  :*.:*        *: :*.:*                : * :::   *.*:

Worm      LHSNVPPMLSKGQVKFLRKSFKYNLLRFLKLPASVPLQQKITTMLTNYLGASPREVQQSI 340  Worm      DNEAELLVNNLEFLFDPALLKPDTASQAIEAVFKWENVPFLFLHSLYTLFHKFKTFESFV 1018
Fly       LNANLPPTLTDSQVSSVRKSLKMQLQTLLKNRGAFEFASTIRGMLVD-LGSSTNEIQKLI 330  Fly       CDIKAIVKATSLCLAERDLYTQEVLMAVLQQLVEVTPLPTLMMRTTIQSLTLYPRLANFV 1015
Human     LHANLPPTLAKSQVSSVRKNLKLHLLSVLKHPASLEFQAQITTLLVD-LGTPQAEIARNM 339  Human     CDMKSIIKATNLCFAERNVYTSEVLAVVMQQLMEQSPLPMLLMRTVIQSLTMYPRLGGFV 1016
Mouse     LHANLPPTLAKSQVSSVRKNLKLHLLSVLKHPASLEFQAQITTLLVD-LGTPQAEIARNM 339  Mouse     CDMKSIIKATNLCFAERNVYTSEVLAVVMQQLMEQSPLPMLLMRTVIQSLTMYPRLGGFV 1016
Rat       LHANLPPTLAKSQVSSVRKNLKLHLLSVLKHPASLEFQAQITTLLVD-LGTPQAEIARNM 339  Rat       CDMKSIIKATNLCFAERNVYTSEVLAVVMQQLMEQSPLPMLLMRTVIQSLTMYPRLGGFV 1016
          *:;*:** *:..**. :**.:* :*  .**  .:. :   *  :*.: **:   *: : :                 :   ::  .   : :  : . :.   .:: :.:   :* *::::    :  :  : . .**

Worm      PPELIQKIAPPRP--PQHPAEPVAKRPKIQNQIFEDDDDDDDEAGPSTSTV-NAKDARTE 397  Worm      ANLFYKVTEKKMYQQSDRWKQAFFKCIKELKTKAYPAVITFLS------FEEYEELKEVL 1072
Fly       PKMDKQEMARRQKRILENAAQSLAKRARLACE-QQDQQQREMELDTEE-----LERQKQK 384  Fly       MNLLQRLIIKQVWRQKVIWE-GFLKTVQRLKPQSMPILLHLPPAQLVDALQQCPDLRPAL 1074
Human     PSSKDTRKRPRDD-----SDSTL-KKMKLEPNLGEDDEDKDLEPGPSGTSKASAQISGQS 393  Human     MNILSRLIMKQVWKYPKVWE-GFIKCCQRTKPQSFQVILQLPPQQLGAVFDKCPELREPL 1075
Mouse     PSSKDSRKRPRDD-----TDSTL-KKMKLEPNLGEDDEDKDLEPGPSGTSKASAQISGQS 393  Mouse     MNILARLIMKQVWKYPKVWE-GFIKCCQRTKPQSFQVILQLPPQQLGAVFDKCPELREPL 1075
Rat       PSSKDSRKRPRDD-----TESTL-KKMKLEPNLGEDDEDKDLEPGPSGTSKASAQISGQS 393  Rat       MNILARLIMKQVWKYPKVWE-GFIKCCQRTKPQSFQVILQLPPQQLGAVFDKCPELREPL 1075
          *     .              . : *: ::  :  :*::: : * . .       :    .                 *:: ::  *::::    *: .*:*  :. * ::   :: :        :::  :*:  *

Worm      AIDMTAKFI-MECLNHETVMNLVKISLYTLPSEMPAAFASSYTPIANAGTEPNRQELSEL 456  Worm      G-------------DGIVAEFKIIYSTLATQQQK-----N--MDEKIKEELHDKERENR- 1111
Fly       STRVNEKFLAEHFRNPETVVTLVLEFLPSLPTEVPQKFLQEYTPIREMSIQQQVTNISRF 444  Fly       SEYAESMQDEPMNGSGITQQVLDIISGKSVDVFVTDESGGYISAEHIKKEAPDPSEISVI 1134
Human     DTDITAEFL-QPLLTPDNVANLVLISMVYLPEAMPASFQAIYTPVESAGTEAQIKHLARL 452  Human     LAHVRSFTPH--QQAHIPNSIMTILEASGKQ-----------EP-EAKEAPAGPLEEDD- 1120
Mouse     DTDITAEFL-QPLLTPDNVANLVLISMVYLPETMPASFQAIYTPVESAGTEAQIKHLARL 452  Mouse     LAHVRSFTPH--QQAHIPNSIMTILEATGKQ-----------EP-EVKEAPSGPLEEDD- 1120
Rat       DTDITAEFL-QPLLTPDNVANLVLISMVYLPETMPASFQAIYTPVESAGTEAQIKHLARL 452  Rat       LAHVRSFTPH--QQAHIPNSIMTILEASGKQ-----------EP-EVKEAPSGSLEEDD- 1120
             :. :*:     . :.* .**   :  **  :*  *   ***: . . : :  .::.:                      *  ..  * .  . :                . *:   .  . .

Worm      MAVQMTNKEIGPGYEWLQQQRKKEYEARNKARSEGMAIAQ----------------TP 498  Worm      ----------------------------ERDKRLRREEKKEKEEREKE-RTRESGKERS 1140
Fly       FGEQLSEKRLGPGAATFSRE--PPMRVKKVQAIESTLTAME--VDEDA---------VQK 491  Fly       STVPVLTSLVPLPVPPPIGSDLNQPLPPGED---------------------------- 1165
Human     MATQMTAAGLGPGVEQTKQCKEEPKEEKVV-KTESVLIKRRLSAQGQAISVVGSLSSMSP 511  Human     -LEPLTLA----------P--APAPRPPQDLIGLRLAQEKALKRQLEEEQKLKPGGVGAP 1167
Mouse     MATQMTAAGLGPGVEQTKQCKEEPKEEKVV-KPESVLIKRRLSVQGQAISVVGSQSTMSP 511  Mouse     -LEPLALALAPAPAPAPAP--APAPRPPQDLIGLRLAQEKALKRQLEEEQKQKPTGIGAP 1177
Rat       MATQMTAAGLGPGVEQTKQCKEEPKEEKVV-KPESVLIKRRLSVQGQAISVVGSQSTMSP 511  Rat       -LEPLALALAPALA----P--APAPRPPQDLIGLRLAQEKALKRQLEEEQKQKPTGVGAP 1173
          :. *::   :***    .:      . :     *.

Worm      IHEPN-MSNRVPAQIVKQSLQE-INTLPVI---QKAKKAFNLVEEAVVFDDKEAAEMFEL 553  Worm      SRR-------------------------------------------------------- 1143
Fly       LSEEEFQRKEEATKKLRETMERAKGEQTVIEKMKERAKTLKLQEITKPLPRNLKEKFLTD 551  Fly       -------------------------------------------------------- 1165
Human     LEEEAPQAKRRPEPII-------PVTQPRLAGAGGRKKIFRLSDVLKPLTDAQVEAMKLG 564  Human     SS---SSPSPSPSARPGPPPSEEAMDFREEGPECETPGIFISMDDDSGLTEAALLDSSLE 1224
Mouse     LEEEVPQAKRRPEPII-------PVTQPRLAGAGGRKKIFRLSDVLKPLTDAQVEAMKLG 564  Mouse     AACVSSTPSVPAAARAGPTPAEEVMEYREEGPECETPAIFISMDDDSGLAETTLLDSSLE 1237
Rat       LEEEVPQAKRRPEPII-------PVTQPRLAGAGGRKKIFRLSDVLKPLTDAQVEAMKLG 564  Rat       TSSVSSTPLVGPAARAGPTPAEEVMEYREEGPECETPAIFISMDDDSGLAETTLLDSSLE 1233
          : *     :.     :                :     * :.* :    :       :

Worm      AYESVLQAERRVVAGGARLMYQKLVVRLTRFWEDCTPFEEKLIEFVLADHKRNDLALL 613  Worm      ----------------------------------------------------  1143
Fly       AVRRILNSERQCIKGGVSSKRRKLVTVIAATFPD---NVRYGIMEFILEDIKQRIDLAFS 608  Fly       ----------------------------------------------------  1165
Human     AVKRILRAEKAVACSGAAQVRIKILASLVTQFNS---GLKAEVLSFILEDVRARLDLAFA 621  Human     GPLPKETAAGGLTLKEERSPQTLAPVGEDAMKTPSPAAEDAREPEAKGNS   1274
Mouse     AVKRILRAEKAVACSGAAQVRIKILASLVTQFDS---GFKAEVLSFILEDVRARLDLAFA 621  Mouse     GPLPKEAAAVGSSSKDERSPQNLSHAVEEALKTSSPE---TREPESKGNS   1284
Rat       AVKRILRAEKAVACSGAAQVRIKILASLVTQFDS---GFKAEVLSFILEDVRARLDLAFA 621  Rat       GPLPKEAAAVGPSSKDERSPQNLSHAVEEALKTSSPE---AREPESKGNS   1280
          * . :*.:*:    .*.     *::. :.: * .     ..  ::.*:* * : * ***:

Worm      WLCELYAQYQGYSNCALFMKEMIAGQEGLTQAQRLDRYDQAMCKMLDAMLERNMEKEAL- 672
Fly       WLFEEYSLLQGFTRHTYVK----------TENRPDHAYNELLNKLIFGIGERCDHKDKII 658
Human     WLYQEYNAYLAAGA-----------------SGSLDKYEDCLIRLLSGLQEKPDQKDGI- 663
Mouse     WLYQEYNAYLAAGT-----------------SGTLDKYEDCLICLLSGLQEKPDQKDGI- 663
Rat       WLYQEYNAYLAAGP-----------------SGTLDKYEDCLICLLSGLQEKPDQKDGI- 663
          ** : *    .                       . *:: :  :: .: *:  .*: :
```

Name	Experiment #	Eggs/Hatched	Lethality (%)
cpsf-1 *(CPSF160)*	1	163/11	**93.7**
	2	238/5	**97.9**
	3	149/1	**99.3**
cpsf-2 *(CPSF100)*	1	68/37	**64.8**
	2	178/28	**86.4**
	3	221/25	**89.8**
cpsf-4 *(CPSF30)*	1	323/5	**98.5**
	2	699/17	**97.6**
	3	204/3	**98.6**
cpf-1 *(CstF-50)*	1	251/76	**76.8**
	2	200/64	**75.8**
	3	185/25	**88.1**
cpf-2 *(CstF-64)*	1	446/54	**89.2**
	2	138/14	**90.8**
	3	61/3	**95.3**
cfim-1 *(CFIm25)*	1	249/23	**91.5**
	2	196/20	**90.7**
	3	154/2	**98.7**
cfim-2 *(CFIm68)*	1	137/1	**99.3**
	2	282/27	**91.3**
	3	260/32	**89.0**
lrp-2 *(CFIm59)*	1	311/116	**72.8**
	2	281/97	**74.3**
	3	222/93	**70.5**
symk-1 *(Symplekin)*	1	62/4	**93.9**
	2	19/4	**82.6**
	3	114/5	**95.8**
rbpl-1 *(RBBP6)*	1	293/0	**100**
	2	344/0	**100**
	3	116/0	**100**
pcf-11 *(CPF11)*	1	141/5	**96.6**
	2	105/0	**100**
	3	172/5	**97.2**
clpf-1 *(CLP1)*	1	286/23	**92.6**
	2	289/18	**94.1**
	3	208/17	**92.4**
pkc-3 *(negative control)*	1	5/0	**100**
	2	8/0	**100**
	3	18/1	**94.7**
	4	51/1	**98.1**
	5	3/0	**100**
	6	22/0	**100**
	7	53/0	**100**
	8	6/0	**100**
	9	13/0	**100**
	10	19/0	**100**
	11	5/0	**100**
	12	4/0	**100**

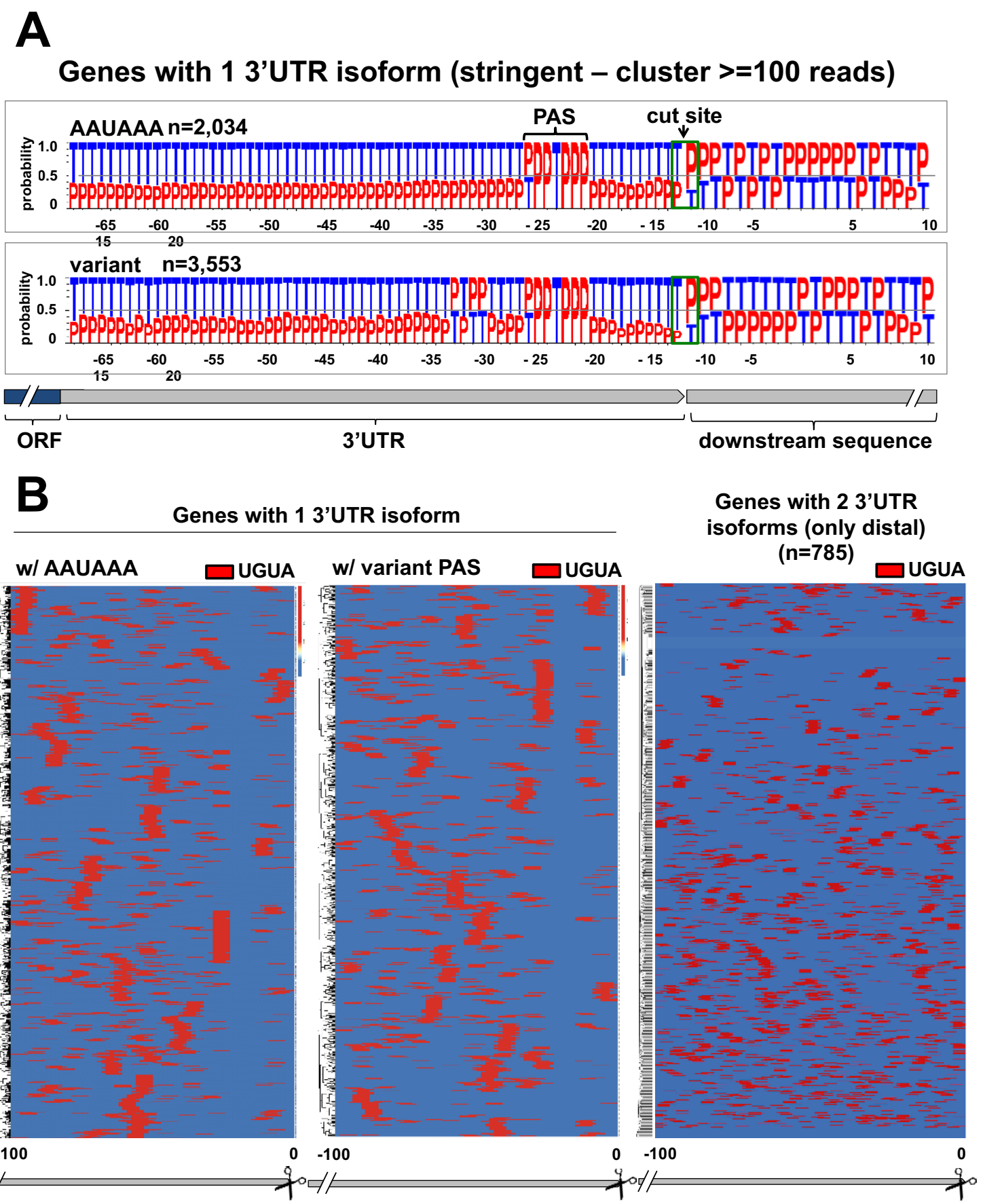
Supplemental Fig. S2: Results of the RNAi experiments of the *C. elegans* CPC: Twelve genes for the members of the *C. elegans* CPC were knocked-down using RNAi. Clones/rows are color-coded as from Figure 1. The human orthologs of each gene are shown in parenthesis in the first column. For each RNAi experiment we use 15 worms, and the number of eggs unhatched vs hatched at the end of the experiment were counted. The percent lethality was consistently high across all tested clones. *pkc-3* RNAi was used as a negative RNAi control, since it is known to induce strong embryonic lethality.



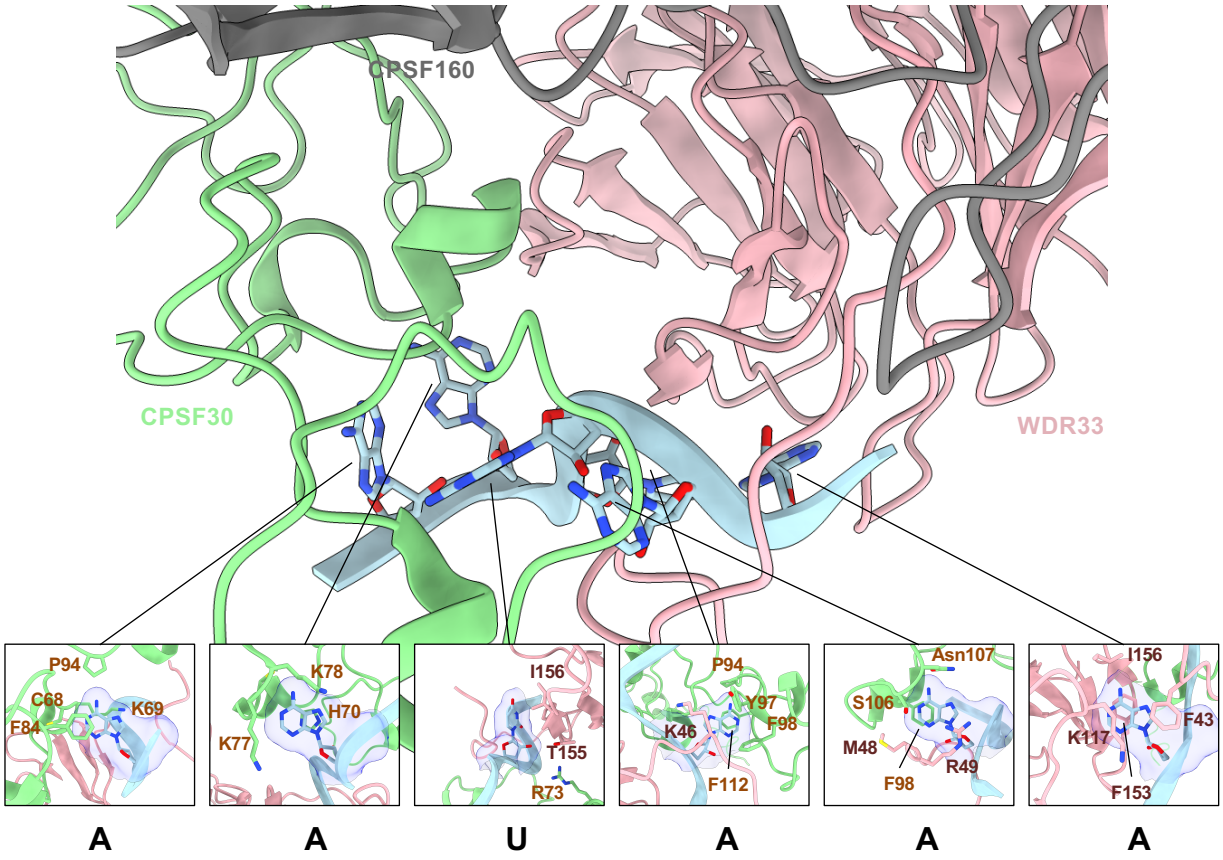
Supplemental Fig. S3: Bioinformatic Pipeline used in this study: The pipeline uses raw transcriptome datasets downloaded from the public repository SRA trace archive to extract and map 3'UTR end clusters to the closest protein-coding genes in the correct orientation.
The pipeline is divided in three large steps: A) Acquisition/Mapping, B) 3'UTR cluster preparation and C) 3'UTR isoforms mapping.
In the acquisition/mapping step, we used custom made Perl scripts to extract reads with 23 consecutive As at the 3' end or 23 consecutive Ts at the 5' end, and then mapped these filtered reads to the WS250 version of the *C. elegans* genome (Bowtie2).
We then sorted and indexed the reads for visualization purposes.
In the 3'UTR cluster preparation step we extracted SAM reads with 100% match to the WS250, and used them to prepare a new bedGraph file (BEDTools).
We then merged the reads and discarded the clusters with less than 5 reads.
Restrictive parameters for cluster identification and 3'UTR end mapping included the discard of clusters with an Adenosine content of <35% downstream of its end.
Clusters were assigned to a mapped 3'UTR end and attached to the closest gene with 2,000nt in the same orientation.
At the completion of these steps we performed the 3'UTR isoform mapping step, which consists of the counting and assignment the total number of 3'UTR isoforms to a given gene.
We discarded clusters with a density of less than 30% of the total number of reads.



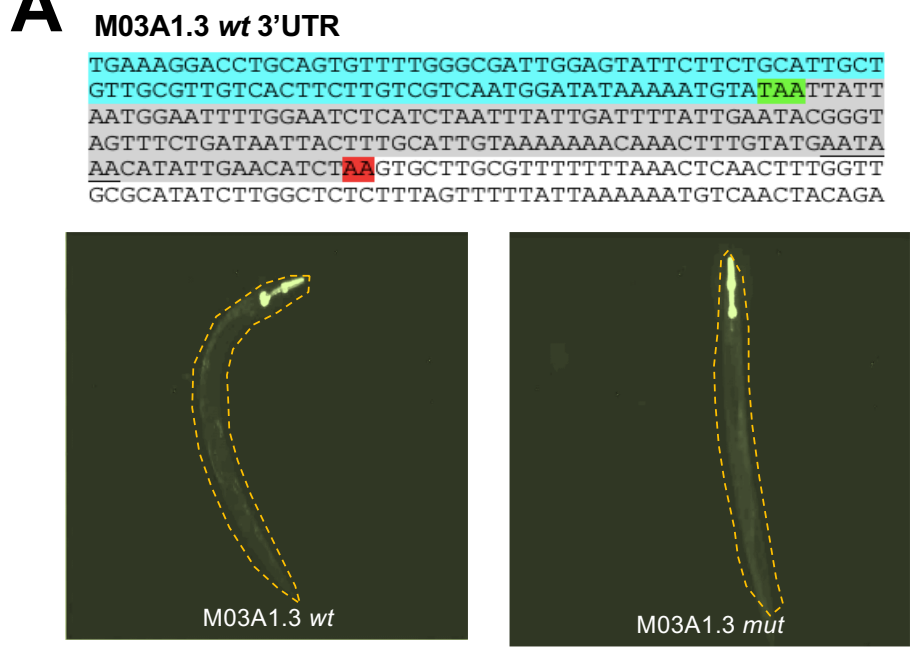
Supplemental Fig. S4: PAS site usage in genes with multiple 3'UTR isoforms. A) In genes with only two 3'UTR isoforms with a difference of at least 10nt between isoforms, 373 pairs of isoforms had canonical PAS elements in both isoforms with an average of 246nt difference between isoforms while 665 pairs had variant PAS elements in both isoforms with an average of 125nt difference between them. In isoform pairs where the type of PAS element switches, 71% have a shorter isoform with a variant PAS element and a longer isoform with a canonical PAS element with an average of 209nt between them while the remaining 29% have a canonical PAS element on the shorter isoform and a variant PAS element on the longer isoform with an average of 322nt between them. B) In genes with three or more 3'UTR isoforms, genes where the longest and the shortest isoform both have canonical PAS elements have an average of 163nt between them while genes where the longest and the shortest isoforms both have variant PAS elements have an average of 211nt between them. 72% of genes switch from a variant PAS elements in the short isoform to a canonical PAS element in the long isoform, with an average of 181nt between them. 28% of genes have canonical PAS elements in the short isoforms and variant PAS elements in the long 3'UTR isoform, with an average of 211nt between the two.



Supplemental Fig. S5: Detection of the 'UGUA' element in *C. elegans* 3'UTRs. A) Logo plot of genomic region within the cut site in genes with only one 3'UTR isoform and with canonical or variant PAS element. There are not immediate elements visible. (T) = Pyrimidine, (P) = Purine. B) Identification of the 'UGUA" motif (red) within 100nt upstream of the cleavage site in genes with one or two 3'UTR isoforms.



Supplemental Fig. S6: Nucleotide binding site of the human CPSF160-WDR33-CPSF30 complex. Ribbon representation of the cryo-EM structure of human CPSF160-WDR33-CPSF30 complex (PDB code: 6DNF) (Sun et al., 2018). The nucleotides of the bound RNA fragment do not show a specific interaction with either CPSF30 or WDR33. The interactions are mostly established by π-π ring stacking. Color gray shows the CPSF160, pink for WDR33, and light green for CPSF30. Sticks represent the RNA molecules bound with CPSF30 and WDR33. Surfaces in the inlets are for individual nucleotides.



Supplemental Fig. S7: *In vivo* cleavage assay for M03A1.3

A) M03A1.3 genomic region cloned downstream of the GFP reporter. Blue: terminal portion of the M03A1.3 ORF. Green: STOP codon. Gray: 3'UTR. Red: mutated terminal Adenosine nucleotides. The transgenic worms expressing the pmyo-3::GFP::M03A1.3_3'UTR *wt* and mutant cassette are shown below.

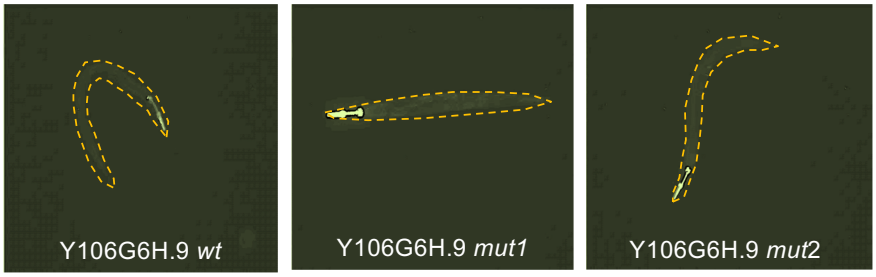
B) At the completion of the experiment, we recovered the total RNA and performed RT-PCR experiments using a forward primer annealing within the GFP ORF and a reverse polydT primer with two anchors containing Invitrogen Gateway adapters. The resultant amplicons were then subcloned in gateway vectors and sequenced to detect the cleavage site. An example of resultant trace files is shown.

C) Examples of 10 clones identified in this study for M03A1.3. The removal of the terminal genomic Adenosine nucleotide induces a cleavage site 3nt upstream of the canonical cleavage site in three clones (arrows), which also contain a terminal Adenosine nucleotide. The PAS element is boxed in blue color.

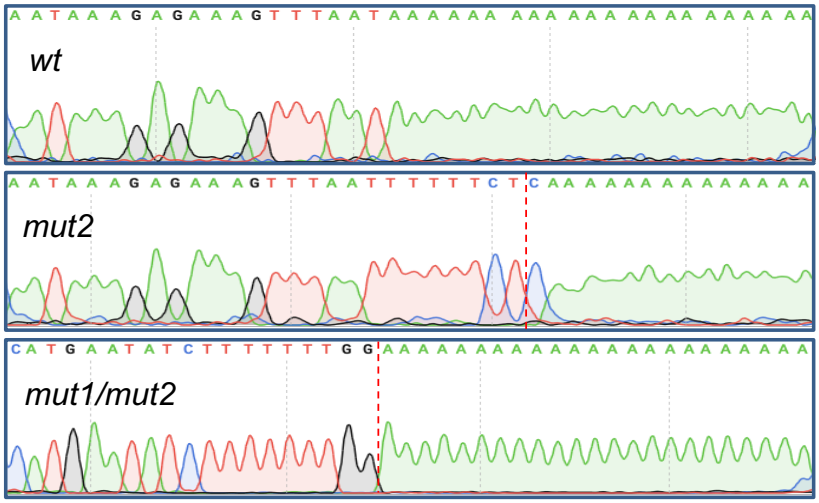
A

Y106G6H.9 *wt* 3'UTR

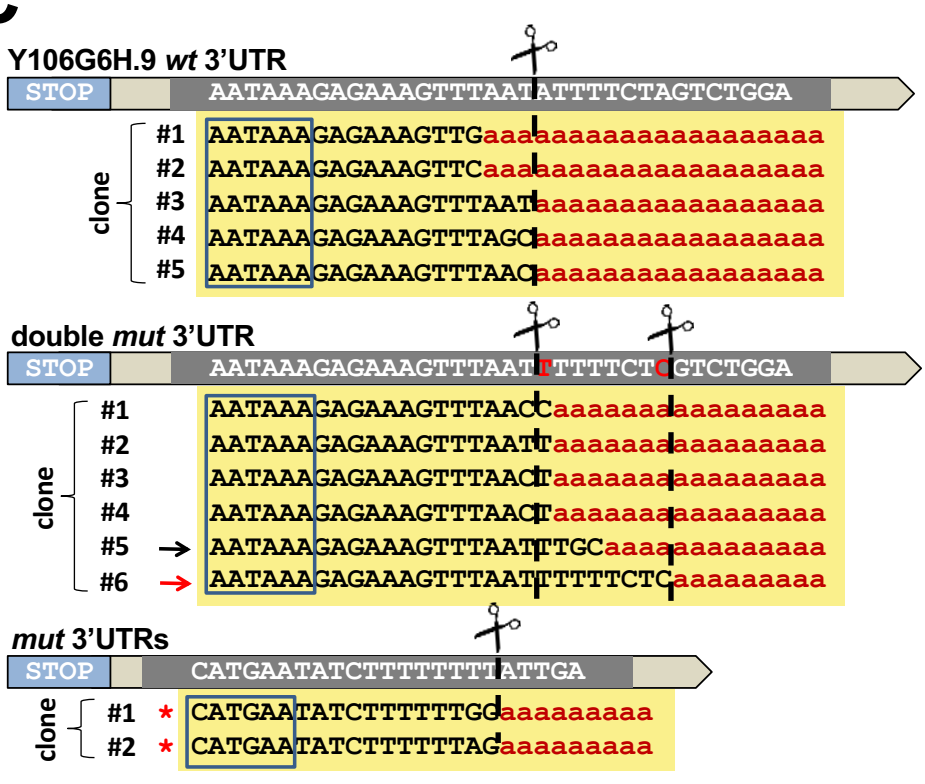
AGAGCCACGTGCACCTTCTATAAACATCCAAAAAAACTAAATATATATTT
TTTTGAAATGCAAACAACACTCCGCAGTTTTGTTTGGAAAACGAATTGGT
CTACTTCTTCATAAAACATATGCGGTTCAATTGATACTTTTATTTCCATT
GGAATTAAATTTAATGAATTGCTTCTTTAAATATTATTTCTATGCATCTG
TTCTTCCTTTTGATTCTTCCATGAATATCTTTTTTTTATTGATCCTACAG
GATCGTACAGGATCTTGTCACACTAAAGATATCTACATATTTAATAATGT
TCACCTTTGTTTTCTATTCTTCATGCCAATAAAGAGAAAGTTTAATATTT
TCTAGTCTGGAATTTTTATTTTTAAAAAGCTGTCAACTGACAAATTATTG
TCCACGACTTCGTCTGTTATTTTTAGTGAACTAAATGTTAGATCGACAGT



B



C

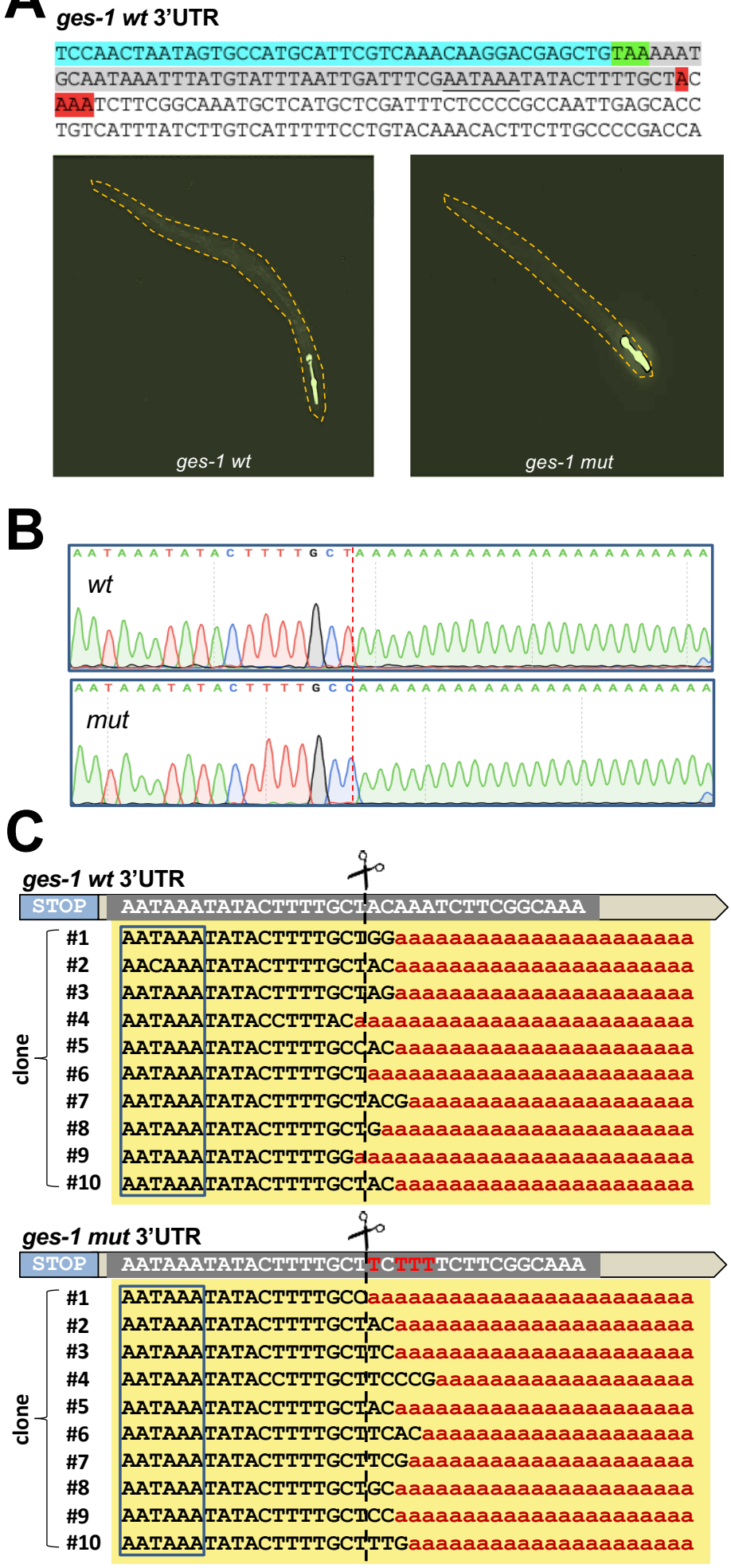


Supplemental Fig. S8: *In vivo* cleavage assay for Y106G6H.9

A) Y106G6H.9 genomic region cloned downstream of the GFP reporter. Blue: terminal portion of the Y106G6H.9 ORF. Green: STOP codon. Gray: 3'UTR. Red: mutated terminal Adenosine nucleotides. Red Asterisk: position of the cryptic cleavage site (see below). The transgenic worms expressing the pmyo-3::GFP:: Y106G6H.9_3'UTR *wt* and mutant cassette are shown below.

B) At the completion of the experiment we recovered the total RNA and performed RT-PCR experiments using a forward primer annealing within the GFP ORF and a reverse polydT primer with two anchors containing Invitrogen Gateway adapters. The resultant amplicons were then subcloned in gateway vectors and sequenced to detect the cleavage site. An example of resultant trace files is shown.

C) Examples of several clones identified in this study for Y106G6H.9. In the *wt* we were able to detect two classes of cleavage sites, both ending within 4nt of each other with a terminal Adenosine nucleotide. In the double mutant, the removal of the terminal genomic adenosine induces a cleavage skip in two clones (arrows). In one case (red arrow), the cleavage occurs 20nt downstream of the PAS element. Two of the mutant clones also shown an occurrence of a new cryptic cleavage site 100nt upstream of the natural site (red asterisks), which also contain a terminal Adenosine nucleotide at their 3'end. The PAS element is boxed in blue color.



Supplemental Fig. S9: *In vivo* cleavage assay for *ges-1*

A) *ges-1* genomic region cloned downstream of the GFP reporter. Blue: terminal portion of the *ges-1* ORF. Green: STOP codon. Gray: 3'UTR. Red: mutated terminal Adenosine nucleotides. The transgenic worms expressing the pmyo-3::GFP::ges-1_3'UTR *wt* and mutant cassette are shown below.

B) At the completion of the experiment we recovered the total RNA and performed RT-PCR experiments using a forward primer annealing within the GFP ORF and a reverse polydT primer with two anchors containing Initrogen Gateway adapters. The resultant amplicons were then subcloned in gateway vectors and sequenced to detect the cleavage site. An example of resultant trace files is shown.

C) Examples of 10 clones identified in this study for *ges-1*. The removal of the terminal genomic Adenosine nucleotide does not alter the cleavage site but makes it more variable. The PAS element is boxed in blue color.