Interaction Profiling Identifies the Human Nuclear EXosome Targeting (NEXT) Complex

Michal Lubas^{1,2,3,#}, Marianne S. Christensen^{1,#}, Maiken S. Kristiansen^{1,#}, Michal

Domanski¹, Lasse G. Falkenby⁴, Søren Lykke-Andersen¹, Jens S. Andersen⁴,

Andrzej Dziembowski^{2,3,*} and Torben Heick Jensen^{1,*}

¹Centre for mRNP Biogenesis and Metabolism, Department of Molecular Biology, Aarhus University, C. F. Møllers Allé 3, Bldg. 1130, DK-8000 Aarhus C, Denmark, ²Institute of Biochemistry and Biophysics, Polish Academy of Sciences, ul. Pawinskiego 5a, 02-106 Warsaw, Poland, ³Department of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, ul. Pawinskiego 5a, 02-106 Warsaw, Poland, ⁴Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Campusvej 55, DK-5230 Odense, Denmark

Equal contribution

*Corresponding authors: Torben Heick Jensen (thj@mb.au.dk), Andrzej Dziembowski (andrzejd@ibb.waw.pl)

Running title: human nuclear exosome activators

The RNA exosome is a conserved degradation machinery, which obtains full activity only when associated with co-factors. The most prominent activator of the yeast nuclear exosome is the RNA helicase Mtr4p, acting in the context of the Trf4p/Air2p/Mtr4p polyadenylation (TRAMP) complex. The existence of similar activator(s) in humans remains elusive. By establishing an interaction network of the human nuclear exosome, we identify the trimeric <u>Nuclear EX</u>osome <u>Targeting (NEXT)</u> complex, containing hMTR4, the Zn-knuckle protein ZCCHC8 and the putative RNA binding protein RBM7. ZCCHC8 and RBM7 are excluded from nucleoli and consistently, NEXT is specifically required for the exosomal degradation of promoter-upstream transcripts (PROMPTs). We also detect putative homologues TRAMP subunits hTRF4-2 (Trf4p) and ZCCHC7 (Air2p) in hRRP6 and hMTR4 precipitates. However, at least ZCCHC7 function is restricted to nucleoli. Our results suggest that human nuclear exosome degradation pathways comprise modules of spatially organized co-factors that diverge from the yeast model.

Introduction

Cells rely on processing and turnover pathways to control their steady-state RNA levels. In addition, the efficient elimination of malformed molecules and transcriptional by-products is constantly required, demanding a subset of robust nucleases. The major eukaryotic 3'-5' exonucleolytic activity is supplied by the multi-subunit RNA exosome complex. It resides in both cytoplasmic and nuclear compartments and participates in a wealth of reactions, including the processing of rRNA, snoRNA and snRNA, the turnover of mRNA and the surveillance of most cellular RNA species (for recent reviews see (Houseley and Tollervey, 2009; Lykke-Andersen et al., 2009)).

First discovered in the yeast *Saccharomyces cerevisiae*, the exosome consists of a catalytically inactive nine-subunit core, that obtains its ribonucleolytic activity from associated subunits. These include the nuclear/cytoplasmic and processive 3'-5' exo/endo-nuclease Dis3p/Rrp44p as well as the nuclear-specific and distributive 3'-5' exonuclease Rrp6p (Allmang et al., 1999; Mitchell et al., 1997). Although active *in vitro*, the exosome needs appropriate activators and adapters for full function *in vivo*. While activators are integral to exosome activity, adapters may rather direct the exosome to its many substrates and therefore be particular to discrete processing/degradation pathways or even recognize specific RNA features. However, as activators and adapters often co-appear in larger complexes, a clear biochemical distinction between the two is difficult to make. Central for exosome activation is an associated RNA helicase activity. In the yeast cytoplasm, this task is performed by the Ski2p DEVH ATPase in the context of the trimeric SKI complex (Anderson and Parker, 1998; Brown et al., 2000), while the DExH/D box RNA

helicase Mtr4p is an essential activator of the yeast nuclear exosome (Allmang et al., 1999; de la Cruz et al., 1998). Several Mtr4p activities are carried out in the context of the TRAMP complex, co-composed of one of the two non-canonical poly(A) polymerases, Trf4p or Trf5p, and one of the two Zn-knuckle proteins, Air1p or Air2p (LaCava et al., 2005; San Paolo et al., 2009; Wyers et al., 2005). TRAMP is suggested to target RNA, presumably through Air1p/Air2p, and trigger its processing/degradation via the 3'end-addition of an unstructured oligo(A) tail, enabling substrate unwinding by Mtr4p in preparation for exosomal nucleolysis. Interestingly, although Trf4p and Trf5p are close homologs and partially redundant (Houseley and Tollervey, 2006), Trf5p activity reportedly occurs in the yeast nucleolus, while Trf4p substrates are often non-nucleolar (Dez et al., 2007; Rougemaille et al., 2007; Wery et al., 2009; Wyers et al., 2005). Thus, exosome coactivator composition is not only different between the yeast nucleus and cytoplasm, but also inside the nucleus between the nucleolus and the remainder of the nucleoplasm. The mechanistic rationale for such division is not clear. Regardless, as a reflection of its broad nuclear localization, tRNA, sn/snoRNA, pre-rRNAs, aberrant mRNA and a class of non-coding cryptic unstable transcripts (CUTs) are all TRAMPdependent exosome substrates (Allmang et al., 1999; Milligan et al., 2005; Reis and Campbell, 2007; van Hoof et al., 2000). In addition to TRAMP, the nuclear exosome co-operates with additional RNA binding proteins, which are currently best described as exosome/RNA adapters; (i) Rrp47p binds structured RNA substrates, associates with Rrp6p and is needed for Rrp6p-dependent exosome activities (Mitchell et al., 2003; Stead et al., 2007), (ii) Mpp6p binds poly(U) RNA, is genetically linked to Rrp47p/Rrp6p and is important for many nuclear exosome activities (Milligan et al., 2008), and (iii) the Nrd1p, Nab3p and Sen1p proteins associate with RNA

polymerase II and nascent RNA to recruit the nuclear exosome and TRAMP to process or completely degrade short transcripts, snoRNAs and CUTs, respectively (Vasiljeva and Buratowski, 2006). Moreover, the Nrd1p/Nab3p/Sen1p complex targets some RNA polymerase III transcripts, such as hypomodified, unedited or unspliced tRNAs (Wlotzka et al., 2011).

The human RNA exosome also exists in cytoplasmic and nuclear forms, that share the highly conserved nine-subunit core. However, unlike in yeast, where Dis3p is the only processive exonuclease component, two Dis3p homologs have been found to associate with the human exosome core: hDIS3 present mainly in the nucleus and hDIS3L restricted to the cytoplasm (Tomecki et al., 2010). While nuclear hDIS3 appears largely excluded from nucleoli, the third exosome-associated nuclease, hRRP6 (PM/ScI100), accumulates in nucleoli, although it can also be detected in the non-nucleolar part of the nucleoplasm and to a minor extent in the cytoplasm (Lejeune et al., 2003; Staals et al., 2010; Tomecki et al., 2010). This differential composition and localization of exosome variants is also reflected by the substrate preferences of the complex *in vivo* (Lykke-Andersen et al., 2011).

Much less is known about human exosome activators/adapters. Functional homologs of Rrp47p (C1D), Mpp6p (hMPP6) and Mtr4p (hMTR4, also called SKIV2L2) have been identified and suggested to help recruit the nuclear exosome to its substrates (Milligan et al., 2008; Schilders and Pruijn, 2008; Schilders et al., 2007; Tomecki et al., 2010). However, mechanistic detail is still lacking. Moreover, clear experimental evidence of TRAMP-like complexes outside of *S. cerevisiae* and *S. pombe* has not been obtained, despite the fact that at the sequence level this complex appears well conserved in e.g. *T. brucei* (Cristodero and Clayton, 2007; Etheridge et al., 2009) and *D. melanogaster* (Nakamura et al., 2008). Also in human

cells, candidate TRAMP components, in addition to hMTR4, are present; the closest homologs of Trf4/5p are hTRF4-1 (POLS) and hTRF4-2 (PAPD5). The latter was recently found to contribute to the 3'end adenylation of RNA polymerase I transcripts targeted by the exosome (Shcherbik et al., 2010). However, this is the only direct example of adenylation potentially being part of an RNA decay mechanism in *Metazoa*. Finally, the Zn-knuckle protein ZCCHC7 has been suggested to be a candidate Air1p/Air2p sequence homolog (Houseley and Tollervey, 2008).

Here we use co-immunoprecipitation (co-IP) analyses of human nuclear exosome complexes coupled to high resolution mass spectrometry (MS) to discover a stable trimer, which we coin the <u>Nuclear EXosome Targeting</u> (NEXT) complex. NEXT contains hMTR4, the putative RNA binding protein RBM7 and the Zn-knuckle protein ZCCHC8. RBM7 and ZCCHC8 are excluded from nucleoli and their depletion leads to the selective stabilization of promoter upstream transcripts (PROMPTs). We also detect hTRF4-2 and ZCCHC7 in hRRP6/hMTR4 precipitates, and consistent with their nucleolar localization, these factors impact the 3'adenylation of rRNA degradation products. Our results identify new human exosome co-factors, demonstrate their differential nuclear partitioning and show a surprising divergence from the *S. cerevisiae* system.

Results

hMTR4 association with the nuclear exosome is stabilized by hRRP6

We previously reported that three nucleolytic subunits, hRRP6, hDIS3 and hDIS3L, associate with the human core exosome (Tomecki et al., 2010). Of these, hRRP6 and hDIS3 mainly localize to the nucleus. Moreover, exosome binding under the employed purification scheme was more robust for hRRP6. Therefore, in order to investigate the protein-protein interaction network of the human nuclear exosome, we first used hRRP6 as the bait in co-IP experiments. To this end, a Flp-In T-Rex HEK293 cell line stably expressing a tetracycline (tet) inducible and C-terminally FLAG-tagged version of hRRP6 was constructed. Specific hRRP6 interaction partners were determined using stable isotope labeling in cell culture (SILAC) methodology (Ong et al., 2003), followed by hRRP6-FLAG co-IP and high-resolution MS analysis. Induction conditions leading to only modest hRRP6 overexpression and the presence of the bait protein in the FLAG IP eluate were verified by western blotting analysis (Fig. S1A). To minimize RNA-dependent interactions, hRRP6-, and all other subsequent, purifications were conducted in the presence of RNase A.

The SILAC/MS approach classifies interactors by specificity (the SILAC ratio between peptide intensities of the bait-induced vs. un-induced samples) and protein abundance estimated from the sum of peptide signal intensities of a given protein normalized to its molecular mass (Fig. 1A). As expected, hRRP6 co-purified all nine core subunits of the exosome as well as C1D, hMPP6 and hMTR4 (Fig. 1A, 1B; Table S1). Also, in agreement with previous results, a less abundant - but highly specific - association of hDIS3 was detected. Interestingly, besides hMTR4, two other putative homologs of yeast TRAMP components, hTRF4-2 and ZCCHC7, were

identified, along with nucleolar WD40-proteins, with high SILAC ratios (Fig. 1A; Table S1) (Gallenberger et al., 2010; McMahon et al., 2010). This implies the existence of a human counterpart of yeast TRAMP (see below).

hMTR4 has previously been reported to interact with the exosome only by indirect studies (Schilders et al., 2007; Tomecki et al., 2010). Our MS data demonstrated hMTR4 co-purification with hRRP6-FLAG to near stoichiometric levels (Fig. 1A: Table S1), a salt-sensitive affinity, which was also evident from commassie-stained protein gels of the hRRP6-FLAG eluate (Fig. 1B). To investigate whether hRRP6 help facilitate hMTR4 interaction with the nuclear exosome, we conducted core component (hRRP41-FLAG) co-IP analysis from HEK293 cells in which hRRP6 had been depleted by RNAi (Fig. S1B), and compared it to a co-IP from control siRNAtreated cells. Upon hRRP6 siRNA administration, levels of the protein in the eluate, as determined by MS, dropped to \approx 30% of that of exosome core components (Fig. 1C). This drop was paralleled by a decrease in hMTR4 (Fig. 1C). In addition, levels of other exosome co-factors decreased (MPP6), or completely disappeared (C1D), concomitantly. Thus, hRRP6 appears to stabilize the association of hMTR4, as well as C1D and MPP6, with the exosome core (see Discussion). We note, however, that under physiological conditions, hMTR4 is likely able to interact/activate the exosome core hRRP6-independently. This is because hMTR4, but not hRRP6, is central for proper 5.8S rRNA 3'end processing ((Schilders et al., 2007); and see below).

Identification of the hMTR4-containing NEXT complex

Given its strong interaction with the nuclear exosome and its predicted central position as an exosome activator, we next characterized hMTR4 interaction partners by the SILAC/MS approach. To this end, physiological levels of hMTR4-FLAG was

stably expressed, and the presence of the bait protein in the FLAG IP was evaluated (Fig. S1C). In a SILAC experimental approach, the hMTR4-FLAG bait efficiently copurified the entire nuclear exosome core as well as hRRP6, C1D and MPP6 (Fig. 2A; Table S2). Interestingly, the hMTR4-FLAG co-IP disclosed a set of new and highly specific interaction partners harboring putative RNA recognition motifs, while hTRF4-2 and ZCCHC7 only became detectable upon increasing the scale of the experiment in a label-free setting (Fig. S1D; Table S3, see Materials and Methods). Some of these new interaction partners were previously localized to the nucleus: ZCCHC8, RBM7, ARS2 and CBP80 (Gruber et al., 2009; Gustafson et al., 2005; Nielsen et al., 2009; Wilson et al., 2008), whereas others are uncharacterized proteins, seemingly differentially localized between the non-nucleolar: ZC3H18 and C14orf102, and the nucleolar: ZFC3H1 and ZCCHC7, part of the nucleus (www.uniprot.org, www.lamondlab.com).

To uncover the more stably associated factors, we first measured protein IP levels after extracts from induced and un-induced samples were mixed (mixed extract 'ME') prior to the IP reaction, and compared it to IP levels obtained using the normal procedure of mixing FLAG-affinity beads (mixed beads 'MB') before co-elution (Fig. 2B). Since extract mixing allows for the exchange of labeled and non-labeled proteins, this experimental strategy assays the stability of interactions. Indeed, while hMTR4 association with nuclear exosome components was efficiently exchanged by the ME procedure, the RBM7, ZFC3H1 and ZCCHC8 proteins all purified with similar efficiencies whether the ME or the MB approach was applied (Fig. 2C). As hMTR4 co-purified ZCCHC8 with very high specificity and protein abundance (Fig. 2A), we also assayed complex formation by a reverse co-IP experiment, employing HEK293 cells stably expressing ZCCHC8-FLAG at physiological levels as the bait (Fig. S1E).

Gratifyingly, in addition to the entire nuclear exosome, MPP6 and C1D, ZCCHC8 also co-purified hMTR4 and RBM7; both with a very high efficiency and specificity (Fig. 2D and S1F; Tables S4 and S5). Moreover, analysis revealed that while exosome components could be 'chased' by mixing extracts, the association of ZCCHC8 with hMTR4 and RBM7 remained unaffected (Fig. 2E). Consistently, conducting hMTR4-FLAG or ZCCHC8-FLAG purifications at stringent conditions of 500mM NaCI resulted in the specific enrichment of hMTR4/ZCCHC8/RBM7 trimeric complexes (Fig. S2A and S2B; Table S6 and S7). Finally, we performed purification of a stably integrated RBM7-EGFP fusion construct, which revealed a near stoichiometric interaction with hMTR4 and ZCCHC8 (Fig. 2F). Collectively, these analyses demonstrate that hMTR4, ZCCHC8 and RBM7 form a stable trimeric complex, which we have named the Nuclear EXosome Targeting (NEXT) complex.

Sub-nuclear partitioning of exosome co-factors

To complement the proteomic analyses, we next examined the subcellular localization of NEXT components, and also included hTRF4-2 and ZCCHC7 in this analysis. Both N- and C-terminal EGFP fusions were generated and visualized by confocal microscopy in transiently transfected HeLa- and in stably transfected HEK293-cells, respectively. Generally, the positioning of the EGFP-tag did not affect localization of factors. Noticeably, ZCCHC8 and RBM7 both localized strictly to the non-nucleolar part of HeLa and HEK293 nuclei (Fig. 3A, 3B and S3A). Moreover, a revisit of the distribution of hMTR4 confirmed its previously reported nuclear localization with nucleolar accumulation, resembling the distributions of hRRP6, MPP6 and C1D (Schilders et al., 2007; Tomecki et al., 2010). Taken together with the co-IP analyses, we therefore suggest that the non-nucleolar pool of hMTR4

associates with ZCCHC8 and RBM7 to form the NEXT complex, whereas nucleolar hMTR4 is, among other factors, complexed with hRRP6, MPP6 and C1D.

Turning to putative TRAMP homologs, a remarkably strict nucleolar localization of ZCCHC7 was observed (Fig. 3A, 3B, and S3A). This result was confirmed using anti-ZCCHC7 antibodies for immuno-localization analysis of HeLa cells, a signal which disappeared upon depletion of the ZCCHC7 protein by RNAi (Fig. S3B). The nucleolar localization of ZCCHC7 is consistent with its association with factors previously reported to be nucleolar (see below). Finally, examination of hTRF4-2-EGFP expressing cells demonstrated its nuclear localization with nucleolar accumulation, fairly similar to hMTR4 (Fig. 3A, 3B, and S3A). This finding is consistent with localization data obtained for the closest Trf4/5p homolog in fission yeast, Cid14, and for DmTrf4 in *D. melanogaster* (Nakamura et al., 2008; Win et al., 2006). It is also compatible with the recent finding that hTRF4-2 is involved in the adenylation of rRNA degradation products (Shcherbik et al., 2010), but contrasts the unexpected cytoplasmic localization previously reported (Mullen and Marzluff, 2008).

To further characterize these putative nuclear exosome co-factors, we determined their sedimentation profiles in a 5%-40% glycerol gradient, separating whole cell extracts of HEK293 cells. Individual fractions of the gradient were subsequently analyzed by western blotting using available antibodies. The human exosome core has previously been reported to sediment into two differently sized pools (Mitchell et al., 1997; van Dijk et al., 2007), which we also found by probing fractions with hRRP40 and hRRP45 antibodies (Fig. 3C). Interestingly, hDIS3 was virtually excluded from the high molecular weight portion of the gradient, where instead hRRP6 accumulated. In contrast, hDIS3 was found in the light fractions, where hRRP6 was relatively less present. This rough separation of nuclear exosome sub-

populations parallels well the nucleolar exclusion and accumulation of hDIS3 and hRRP6, respectively (Tomecki et al., 2010). A similar partitioning of ZCCHC8 and ZCCHC7 was observed, with the former primarily present in the hDIS3 containing fractions, and the latter accumulating in the hRRP6 containing fractions (Fig. 3C). Again, this correlates well with the sub-nuclear localization of these factors. We were unable to probe the gradient fractions for RBM7 and hTRF4-2 as antibodies towards these factors were not available.

Interaction profiles of human putative TRAMP homologs

hMTR4, hTRF4-1, hTRF4-2 and ZCCHC7 display a domain composition similar to their putative yeast counterparts (Fig. 4A; for full alignments see Fig. S4A-C). Given their specific presence in the hRRP6-FLAG- and hMTR4-FLAG-precipitates (Fig. 1A and S1D), we focused on ZCCHC7 and hTRF4-2 for further label-free interaction studies. As in earlier purifications, moderate induction levels were used (Fig. S4D, lower left, ZCCHC7 was tested - for hTRF4-2 no antibodies were available) and the presence of bait proteins in the FLAG IPs was verified. The strongest indication of a human TRAMP complex comes from the ZCCHC7-FLAG bait purification, where hMTR4 and hTRF4-2 were specifically present in high amounts together with exosome components (Fig. 4B; Table S8). Moreover, in agreement with its localization, the ZCCHC7-FLAG IP revealed WDR36, WDR3, PWP2 and TBL3 proteins that are involved in rRNA biogenesis (Gallenberger et al., 2010). These proteins were also present in the hTRF4-2-FLAG eluate, as was hRRP6, hMTR4 and ZCCHC7 (Fig. 4C and 4D; Table S9). In addition, hTRF4-2 also interacted with a subset of splicing factors. This dual interaction profile with assigned nucleolar and non-nucleolar proteins likely reflects the sub-nuclear localization of hTRF4-2 (Fig. 3).

Although the ZCCHC7 and hTRF4-2 interaction analysis indicated the presence of a human TRAMP complex, we were not able to obtain enough material to verify its existence by SDS-PAGE and protein staining. Instead, we conducted western blotting analysis and confirmed the co-IP/MS data that ZCCHC7 and hTRF4-2 interact with each other and with hRRP6 in a salt-sensitive manner (Fig. 4D).

Substrate preferences of human exosome co-factors reflect their sub-nuclear localizations (NEXT components)

Provided that the proteins characterized in this paper are true co-factors of the human exosome, their removal should affect levels of its nuclear substrates. Hence, to examine the *in vivo* activity of NEXT components, hTRF4-2 and ZCCHC7, we depleted these proteins in HeLa cells using RNAi (Fig. 5A). In the case of RBM7 and hTRF4-2 where antibodies were not available, we tested the siRNA-directed depletion efficiency on cell lines stably expressing RBM7-EGFP (Fig. 5A, bottom left) or hTRF4-2-FLAG (Fig. 5A, bottom right). All siRNAs exhibited robust knock down efficiencies. In addition, and as previously reported for the hRrp40 knock down, which co-depletes hRrp6 (Kammler et al., 2008; Tomecki et al., 2010), depletion of key proteins in some cases decreased levels of others; i.e. hMTR4 knock down co-depleted ZCCHC8 and hRRP6 (Fig. 5A, lane 4) and administration of siRNA against hTRF4-2 caused decreased levels of ZCCHC7 (Fig. 5A, lane 3). This possibly reflects the tight individual interactions of these factors.

As the NEXT complex is confined to the non-nucleolar part of the nucleus, we turned to PROMPTs (Preker et al., 2008). These transcripts are degraded by either of the two nuclear nucleolytic exosome subunits hRRP6 or hDIS3, the latter of which is excluded from nucleoli (Tomecki et al., 2010). We therefore surmised that

exosome-directed PROMPT-removal cannot be nucleolar. In agreement with previous studies (Preker et al., 2008; Tomecki et al., 2010), selected PROMPTs were markedly stabilized in hRRP40 singly- as well as hDIS3/hRRP6 double-depleted cells, whereas individual depletion of hDIS3 or hRRP6 had smaller effects (Fig. 5B). Interestingly, depletion of NEXT complex components also resulted in robust PROMPT stabilization, whereas hTRF4-2- and ZCCHC7-depletions had no effect. Increased PROMPT levels were most pronounced in hMTR4-depleted samples, probably reflecting the pleiotropic effects of this knock down. We conclude that the NEXT complex cooperates with the nuclear exosome to remove PROMPTs from the non-nucleolar parts of human cell nuclei.

Substrate preferences of human exosome co-factors reflect their sub-nuclear localizations (ZCCHC7 and hTRF4-2)

Due to the nucleolar localizations of ZCCHC7 and hTRF4-2, we next turned to substrates assumed to be degraded in this sub-compartment. hTRF4-2 is involved in the adenylation of nascent 47S rRNA degradation intermediates (Fig. 6A), appearing as a result of treatment with a low concentration of Actinomycin D (Shcherbik et al., 2010). Presumably the drug triggers rRNA degradation by inhibiting RNA polymerase I activity, and since this is the hallmark of nucleolar assembly, we inferred that hTRF4-2 action may occur herein. To visualize adenylation of these pre-rRNA degradation intermediates, we adopted the assay of Shcherbik et al. and subjected total RNA from Actinomycin D treated cells to dT-primed semi-quantitative PCR followed by Southern blotting analysis using a radio-labeled 5'ETS specific probe (Fig. 6B). In accordance with published data, depletion of hRRP40 or hRRP6 resulted in increased levels of PCR products corresponding to adenylated 5'ETS

fragments, as did hMTR4-depletion (Fig. 6C). hDIS3 knockdown had no effect, implying that 5'ETS adenylation is indeed occurring in the nucleolus. Importantly, depletion of both hTRF4-2 and ZCCHC7 resulted in decreased levels of 5'ETS product as compared to the EGFP control (Fig. 6C, compare lane 1 with lanes 6 and 9), whereas ZCCHC8 and RBM7 depletions had no effect. Even more pronounced, co-depletion of hTRF4-2 or ZCCHC7 in the hRRP40- or hMTR4-depleted backgrounds, yielded a substantial down-regulation of the otherwise increased 5'ETS signals (Fig. 6C, compare lane 2 with lanes 7 and 10, or lane 5 with lanes 8 and 11). We conclude that ZCCHC7 and hTRF4-2 are co-factors of the human nucleolar exosome and that although ZCCHC7 has no adenylation activity of its own, it possibly aids hTRF4-2 in this function. Moreover, hMTR4 appears dispensable for adenylation of 5'ETS fragments, suggesting that a putative hTRF4-2/ZCCHC7 dimer may suffice to bind and adenylate target RNA, whereas hMTR4 facilitates recruitment of the exosome for transcript degradation.

Finally, we tested the effects of factor depletions on nuclear processing of the thoroughly studied exosome substrate 5.8S rRNA. As previously observed (Tomecki et al., 2010), hRRP40 and hRRP6/hDIS3 depletions caused the appearance of slower migrating 5.8S rRNA species when assayed by northern blotting analysis using a probe specific for the 3'extended region (Fig. 6D). Interestingly, hMTR4 depletion also yielded a robust 3'end processing defect, while neither ZCCHC8, RBM7, hTRF4-2 nor ZCCHC7 knock down showed a noticeable phenotype. It therefore seems that 5.8S rRNA processing requires the nuclear exosome complexed with hMTR4, whereas the NEXT complex and the putative TRAMP homologous factors are dispensable. Moreover, the partial phenotype seen upon

hDIS3 depletion argues that 5.8S rRNA 3'end processing in human cells occurs - at least in part - outside nucleoli.

Discussion

A vast amount of work on the *S. cerevisiae* exosome has convincingly demonstrated the need for co-factors to efficiently deal with its many cellular RNA substrates. Here, we uncover an interaction network of the human nuclear exosome, revealing a set of new partners, harboring an unanticipated complexity of composition and compartmentalization. Our data position hMTR4 as a central component in the coupling between the ribonucleolytic activity of the exosome and its other co-factors (summarized in Fig. 7). In sum, we propose that hRRP6 helps stabilize the tethering of hMTR4 to the exosome core, and that hMTR4 provides the link to compartment-specific activators/adaptors. Thus, similar to the ribonucleolytic activities of the human exosome, its nuclear co-factor complexes are also differentially distributed (Staals et al., 2010; Tomecki et al., 2010).

In S. cerevisiae, Mtr4p is indispensable for all known nuclear exosome activities, including 5.8S rRNA processing, which does not require other TRAMP complex components (Allmang et al., 1999; Bird et al., 2005; LaCava et al., 2005; Torchet et al., 2002; van Hoof et al., 2000). Our data suggest a parallel situation in human cells; i.e. depletion of hMTR4 causes, in contrast to other examined exosomal co-factors, an accumulation of 5.8S rRNA precursor species akin to that resulting from depletion of the hRRP40 core exosome subunit (Fig. 6C). Moreover, hMTR4 also exerts 'exosome core-like activity' on the analyzed PROMPT and 5'ETS substrates (Fig. 5B and 6B). Taken together with our data, that hRRP6 co-purifies hMTR4 with the exosome core at near stoichiometric levels (Fig. 1A), it seems appropriate to consider hMTR4 an integral activator of the human nuclear exosome. In further support of this notion, the sub-nuclear distributions of hMTR4, hRRP6, and the core

component hRRP45 (PM/ScI-75) are strikingly similar (Schilders et al., 2007). We note that Mtr4p also co-purifies with Rrp6p in yeast, thus, making a similar scenario possible in this organism (Peng et al., 2003). It is unlikely, however, that *in vivo* hRRP6 is the only factor involved in the tethering of hMTR4 to the exosome core as in both yeast and human cells, hRRP6 depletion yields relatively mild phenotypes on some substrates, e.g. 5.8S rRNA, as compared to hMTR4- or exosome core component-depletion (Fig. 6D, (Allmang et al., 1999; de la Cruz et al., 1998; LaCava et al., 2005; Schilders et al., 2007)). Thus, hMTR4 may also be involved in preparing substrates for the exosome, independent of its direct association.

The NEXT complex

S. cerevisiae Mtr4p is in excess of other exosome components and co-factors (Ghaemmaghami et al., 2003). Our co-IP experiments suggest a similar situation in human cells; i.e. hMTR4 is purified by hRRP6-FLAG in stoichiometric amounts with exosome components (Fig. 1A), but is in excess over same factors when hMTR4-FLAG is used as the bait (Fig. 2A). This may indeed enable hMTR4 to 'explore' the nucleus for additional co-factors, some of which are likely to serve as adaptors between the exosome and its substrates. Here, we describe the identification of the NEXT complex, consisting of hMTR4, ZCCHC8 and RBM7. This trimeric complex can be purified to near homogeneity at high salt conditions when RBM7-EGFP is used as the bait (Fig. 2F), and is very stable when challenged with exogenous extract in the ME/MB experimental scheme (Fig. 2C and 2E). NEXT factors also co-purify to near stoichiometry when physiological levels of ZCCHC8-FLAG are used as bait (Fig. 2D). However, consistent with the idea that hMTR4 is also engaged with other exosome co-factors, with the exosome itself and perhaps even with exosome-

unrelated proteins, RBM7 and ZCCHC8 are detected only as sub-stoichiometric partners when physiological levels of hMTR4-FLAG are used as bait (Fig. 2A).

We suggest that NEXT constitutes a non-nucleolar hMTR4-containing complex (Fig. 7), which targets, among other possible substrates, PROMPTs for rapid exosomal turnover. This is because nuclear ZCCHC8 and RBM7 are excluded from nucleoli (Fig. 3), and because their depletions result in PROMPT stabilization levels, that are reminiscent of those obtained when depleting hRRP40 or hRRP6/hDIS3 (Fig. 5B). Since RBM7 and ZCCHC8 harbor an RRM and a Zn-knuckle domain, respectively, it is possible that they provide substrate-targeting activity. Their tight assembly with hMTR4 in the NEXT complex could then facilitate the efficient funnelling of e.g. PROMPTs into the exosome. Such a model suggests the existence of a divergent pathway for nuclear ncRNA turnover in higher eukaryotes. As PROMPTs share many characteristics with S. cerevisiae CUTs, the participation of putative TRAMP homologs in PROMPT decay might have been expected. However, ZCCHC7 depletion has no discernable effect on PROMPT levels (Fig. 5B). Moreover, although hTRF4-2 can adenylate PROMPT 3'ends (Preker et al., submitted), this activity is not required for RNA degradation per se (Fig. 5B). Rather, PROMPT 3'ends, accumulating in the absence of a functional exosome, can provide substrates for hTRF4-2, consistent with its presence also in the non-nucleolar part of the nucleus. This is in contrast to S. cerevisiae, where both Trf4p and Air2p are directly required for the decay of CUTs (Wyers et al., 2005). Another difference between systems is the apparent lack of a human Nrd1p/Nab3p/Sen1p complex, which couples the transcription termination of CUTs to their rapid turnover (Lykke-Andersen and Jensen, 2006). Based on sequences, there are no strong candidate homologs of Nrd1p and Nab3p in the human genome. It thus remains an open question how

tightly the NEXT-directed decay of PROMPTs is coupled to the act of PROMPT transcription. As most human snoRNAs are cleaved out of larger precursor RNAs, in contrast to *S. cerevisiae* snoRNAs that are often transcribed from independent loci, it may be that the requirement of a Nrd1p/Nab3p/Sen1p-like complex to couple ncRNA transcription directly to exosomal processing/decay has been lost during evolution.

Interestingly, ZCCHC8 was previously found to be a target of the glycogen synthase kinase 3 (GSK-3) and to co-purify RBM7 and hMTR4, although no link to the RNA exosome was made (Gustafson et al., 2005). Moreover, RBM7 reportedly interacts with splicing factors (Guo et al., 2003). The relevance of these observations for the role of NEXT in RNA biology and exosomal targeting remains to be addressed.

A TRAMP complex in human cells?

Much effort has gone into the identification of TRAMP subunit homologs in higher eukaryotes, but apart from predictions, only little experimental support has been provided (Houseley and Tollervey, 2008). Here, we identify for the first time putative human TRAMP subunits hTRF4-2 (PAPD5) and ZCCHC7 (hAIR2) using hRRP6 (Fig. 1A) or hMTR4 (Fig. S1D) as purification baits. Validation by western blotting analysis verifies that these proteins indeed interact with the nuclear exosome and hMTR4, albeit less strongly than NEXT components (Fig. 4D). Consistently, hTRF4-2 is reportedly involved in the adenylation of rRNA degradation products (Shcherbik et al., 2010), an observation recapitulated in this paper and extended to also be valid for ZCCHC7 (Fig. 6C). As opposed to its assembly into the NEXT complex, we do not have firm proof that hMTR4 interacts directly with neither ZCCHC7 nor hTRF4-2. However, in favor of such a scenario, hMTR4 protein abundance was high in both

ZCCHC7-FLAG- and hTRF4-2-FLAG-IPs (Fig. 4B and 4C). Hence, we predict the existence of a nucleolar human 'TRAMP-like' complex (Fig. 7), which may even contain additional components like WDR3, WDR36, PWP2 and TBL3, scoring with similarly significant protein efficiencies in ZCCHC7-, hTRF4-2- and hMTR4-FLAG-precipitations (Fig. 4B, 4C and Fig. S1D). The exact composition of such complex(es) awaits further characterization.

Our reported nuclear localization of hTRF4-2 contrasts that reported by Mullen et al. (Mullen and Marzluff, 2008), who found the protein to be mainly cytoplasmic, a finding challenged by recent functional studies (Schmidt et al., 2010; Shcherbik et al., 2010). The hTRF4-2 clone used in this study exhibits a higher degree of conservation when compared to other vertebrate genomes (see Materials and Methods), and fusing it to EGFP, yielded the presented localization. Moreover, the amino acid composition of this isoform is consistent with the hTRF4-2 peptides obtained from both the ZCCHC7- and hMTR4-co-IPs, and we note that our hTRF4-2 IP also yields factors that are consistent with a nuclear function. Taken together, this strongly suggests that the present clone expresses physiologically relevant hTRF4-2. The functional relevance of hTRF4-2 outside nucleoli remains to be investigated, but the consistent presence of splicing factors in the hTRF4-2 co-IP may link the protein to pre-mRNA biology.

Additional targeting complexes of the nuclear human exosome?

Our data imply that additional nuclear exosome co-factors may exist. Most prominently present in the hMTR4- and ZCCHC8-FLAG-IPs are the Zn-knuckle proteins ZFC3H1, ZC3H18 and ARS2 (Fig. 2A, Fig. 2D, Fig. S1D, Fig. S1F and Fig. S2A). Interestingly, ZFC3H1 scores high for both specificity and abundance in the

hMTR4 IP even when conducted in 500mM NaCl (Fig. S2A). ZC3H18 is significantly present in the ZCCHC8 IP (Fig. 2D), indicating a non-nucleolar localization. Future work will focus on delineating the nature of possible RNA substrates for these proteins. Finally, both hMTR4- and ZCCHC8-FLAG baits consistently co-purify the ARS2 protein together with one or both of the cap-binding proteins CBP20 and CBP80 (Fig. 2A, Fig. 2D, Fig. S1D and Fig.S1F). Murine and *Drosophila* ARS2 was recently suggested to link CBP20/CBP80 to the RNAi system, possibly through the biogenesis of miRNA (Gruber et al., 2009; Nielsen et al., 2009; Sabin et al., 2009; Wilson et al., 2008). Our finding may therefore link the human exosome with miRNA metabolism.

The complexity and range of reported RNA exosome substrates have steadily increased since the first discovery of the *S. cerevisiae* complex (Mitchell et al., 1997). Most, if not all, nuclear RNA 3'ends have, at some physiological condition, a chance of encountering the exosome, posing the question how these substrates are targeted for processing or decay. Although the rules of such targeting as well as the mechanisms underlying exosome activation are not yet well understood, this work revealing a set of new co-factors should pave the way for such analysis in human cells.

Materials and Methods

Plasmids and cloning

Plasmids used in the study are listed in Table S10. Coding sequences were amplified by PCR from a HEK293 cDNA library using oligonucleotides listed in Table S11, and cloned into the BamHI/XhoI sites of pcDNA5/FRT/TO-FLAG (Invitrogen) using standard procedures . Tetracycline inducible cell lines stably expressing FLAG or EGFP fusions were generated according to the manufacturer using the Flp-In[™] T-REx[™] system (Invitrogen). Optimal expression conditions were tested for each construct using different concentrations (10-1000ng/mI) of tetracycline.

To amplify cDNA encoding hTRF4-2 HeLa- and HEK293-RNA was used. The longest cDNA variant was sub-cloned to obtain FLAG- and EGFP- protein fusions (as described above). In comparison to the previously localized isoform of hTRF4-2 (Q8NDF8) (Mullen and Marzluff G&D 2008), the one used here (submitted to UniProt) includes two insertions: (i) 12 amino acids in close proximity to the signal peptide (N-terminus) and (ii) 47 amino acids in the C-terminal region of the protein.

SILAC and label-free co-IPs

In SILAC experiments proteins were labelled with stable isotopes in cell culture. HEK293 Flp-In T-Rex cell lines harbouring FLAG-tagged versions of hRRP6, ZCCHC8 or hMTR4 were cultured for a minimum of five doublings in custom-made DMEM SILAC media, containing 15% dialyzed FBS (Gibco), 2mM L-glutamine, 100U/ml penicillin and 100ug/ml streptomycin. Subsequently, cells were grown for 24h in media containing Lys0 (${}^{12}C_{6}{}^{14}N_{2}$) and Arg0 (${}^{12}C_{6}{}^{14}N_{4}$), or Lys4: (${}^{2}H_{4}$) and Arg6 (${}^{13}C_{6}{}^{14}N_{4}$) (Sigma-Isotec), fresh media was added and the pool of cells growing on labelled amino acids was induced with tetracycline (hRRP6 - 30ng/ml, hMTR4 -30ng/ml, ZCCHC8 – 10ng/ml). 24h later, cells from one 145mm plate were collected and washed in 0,75ml of PBS (Gibco) and collected in RSB100 buffer (10mM Tris/HCI pH 7.4, 100mM NaCl, 2.5mM MgCl2), containing 0.5% Triton X100 and protease inhibitors (Roche). Cells were gently lysed by sonication: 3x10sek, 20W (Branson 250) and centrifuged (4000g, 15min) at 4°C. Supernatants were treated with RNase A (100ug/ml) for 15min on ice and centrifuged (16000g, 5min) at 4°C. 'Mixed extracts (ME)' and 'mixed beads (MB)' experiments were set-up as described in the results section and loaded onto agarose anti-FLAG (M2) beads (Sigma). After 3h of incubation, beads were washed 4 times with RSB100/0.5% Triton X100. For the MB approach, beads were pooled in one tube, and both MB and ME were washed 4 times with RSB100. For high-salt purifications all washing steps were performed in RSB500 (500mM NaCl). Proteins were eluted with FLAG peptide (0.5 mg/ml) in TBS (Sambrook and Russell, 2001). Additional elution was done by incubation with 2 x SDS loading buffer. Before MS analyses samples were analysed SDS-PAGE followed by silver staining (data not shown) and western blotting. 1/100 of the initial input and 1/10 of FLAG eluate, respectively, were loaded on gels.

Sample preparation, MS analyses and data treatment

Prior to MS analysis eluates were denatured with 8M Urea and digested in solution by endoproteinases: LysC in 6M Urea, 50 mM ammonium bicarbonate buffer, pH 7.8 (0.1AU/mL, Wako) for 3h at room temperature and trypsin (Promega) over night at 37°C. Released peptides were reduced in 10mM dithiotreitol (DTT) for 30min at RT and alkalized in 55mM iodoacetamide for 20min at room temperature. Finally, peptides extracted with 1% trifluoroacetic acid (TFA) and concentrated on homemade reverse-phase C18 columns (prewashed with methanol) were eluted with 80% acetonitrile/0.5% TFA, then dried and analyzed by LC-MS.

MS and data analyses were performed as previously described (Tomecki, et al. 2010). The LTQ-Orbitrap mass spectrometer was operated in a mode, which provides high-resolution of ion spectra (m/z 300–1500, resolution 60000 and ion accumulation to a target value of 5x10⁶ ions). MaxQuant software (version 1.0.13.13) was used to extract peptide ion and fragment ion m/z intensity signals from the mass spectra (Cox and Mann, 2008), which were submitted to the Mascot program (Matrix Science) for database searches (Human IPI sequence database). The same software was used for calculation of peptide isotope ratios as well as evaluation of the certainty of peptide identification using false discovery rate analysis.

Label free quantitation was done using the MaxQuant software. Peptide m/z signal intensity was found summing intensities over the entire elution of the peptide, and protein intensity as the sum of intensities for all peptides representing the protein. Proteins included in Tables S1-9 provide hits identified by a false discovery rate of 0.01 derived by decoy database searching, and only unique peptides were used for the quantitation. The previously reported non-specific binders also detected in a 'bead proteome' were removed from the datasets (Tomecki et al., 2010).

Purification of the NEXT complex

HEK293 RBM7-EGFP cells were induced by 10ng/ml of tetracycline. The standard co-IP protocol was scaled up to include ten 145mM plates and the entire procedure was performed in high salt conditions (RSB500). Elution was performed with 0.1M glycine pH 2.5, followed by neutralization with 1M Tris/HCl pH 8.0. 1/10 of the eluate was loaded on the gel.

Protein alignments and motif finding

Protein sequences (Table S12) were loaded into the CLC Main workbench 5.6.1 software (www.clcbio.com). Alignments were created using the create alignment tool with a gap open cost of 10 and a gap extension cost of 1. Protein motifs were found using the pfam domain search, using the 100 most common motifs and a significance cut-off at E=1.

Localization studies

Localization analyses were performed essentially as described (Tomecki et al., 2010). For localization of EGFP-tagged proteins in HeLa cells, the cells were seeded on 6-well Lab-Tek® Chamber Slides (Nunc), and 12h later transfected with appropriate constructs using a standard transfection protocol. Test media was supplemented with tetracycline to induce protein expression. 24h later localization experiments were performed directly on living cells, or indirectly by immunofluorescence analysis of formaldehyde-fixed cells. Nuclei were stained with Hoechst 33342 and nucleoli were visualized by both phase contrast light and antifibrillarin antibodies (Abcam 1:150). Cells were visualized by confocal microscopy -Olympus FluoView® FV10i – using a 60x water-immersion objective (NA 1,2).

siRNA-mediated knock-downs and RT-qPCR

Transfections were carried out using 20 nM siRNA (Table 13) for 2 days and repeated for another 2 days both times using Lipofectamin2000 (Invitrogen). Five µg of total RNA was treated with 1U of DNase I (Invitrogen), reverse transcribed using 0.5µg of a dT primer, which was anchored at the 3'end and carried an adaptor at the

5'end (APVN) and Superscript II in a volume of 20µl. Dilutions were subjected to qPCR analyses on a Stratagene Mx3005P using Platinum SYBR Green (Invitrogen) and 0.3µM of oligonucleotides at an annealing temperature of 59°C. Controls lacking Superscript II showed a negligible background. A complete list of RT-qPCR primers is found in Table S14.

Western blotting analysis

Cells were treated with siRNA directed against the protein of interest and equal amounts of cell lysate, or whole cell extract, were subjected to 10% SDS PAGE, and subsequently analyzed according to standard procedures (Sambrook and Russell, 2001) using the following primary antibodies: monoclonal mouse anti-FLAG M2 (Sigma-Aldrich), polyclonal mouse anti-ZCCHC8 and polyclonal rabbit anti-hMTR4 (Abcam), polyclonal mouse anti-RRP40 and anti-EGFP (Santa Cruz biotechnology), polyclonal rabbit anti-RRP40 (courtesy of Dr. Ger Pruijn), polyclonal rabbit anti-DIS3 (Abnova), polyclonal anti-ACTIN and polyclonal rabbit anti-RRP6 (Sigma-Aldrich) and polyclonal rabbit anti-ZCCHC7 (Atlas antibodies). Anti-mouse and anti-rabbit IgGs coniugated with HRP (Dako) were used as secondary antibodies .

Glycerol gradient

Whole cell extracts were prepared by washing cells with PBS followed by scraping into RSB100 buffer, containing 0.5% Triton X-100 and protease inhibitor (one Complete, mini, EDTA-free, Protease Inhibitor cocktail (Roche) tablet per 10ml of buffer). Cell slurries were sonicated on ice using a Branson Sonifier 250 at setting 1 (output: ~20W) for 3x10 s, and subsequently centrifuged at 4000g for 15min at 4°C. The supernatant was loaded onto a 5%-40% (v/v) glycerol gradient, prepared in

gradient buffer (RSB100 containing 0.5% Triton X-100). The gradient was centrifuged essentially as described (van Dijk et al., 2007), only using a SW41 (Beckman) rotor, and collecting 21 fractions. These were precipitated by TCA, re-suspended in SDS load buffer, separated by SDS-PAGE and transferred to a nitrocellulose membrane for western blotting analysis.

rRNA analysis

Five µg of total RNA from cells treated with siRNA was run on a 7% denaturing polyacrylamide gel and analyzed by standard northern blotting procedures. ITS2 extensions and mature 5.8S rRNA were visualised using an *in vitro* transcribed RNA probe kindly provided by Dr. Ger Pruijn (Schilders et al., 2007) and a 5'radiolabeled 5.8S rRNA specific DNA oligo: 5'-GTGTCGATGATCAATGTGTCCTGCAATTCA, respectively.

To analyze 5'ETS degradation products, HeLa cells were treated with siRNAs as described above. Before harvest, the medium was replaced with RPMI containing 20 ng/ml Actinomycin D for 60min. RNA was harvested and cDNA was prepared as described above, but using an unanchored adaptor-carrying dT primer (AP). The 5'ETS region was amplified by 19 cycles of PCR, using a 5'ETS-specific forward primer and the Abridged Universal Amplification Primer (AUAP), containing the sequence of the adaptor. PCR products were run on a 1% agarose gel and analyzed by standard southern blot procedures using an internal probe against 5'ETS. PCR using primers designed to amplify a part of GAPDH mRNA was used as control. For primers used in these experiments see Table S11.

Acknowledgements

We thank Stepanka Vanacova, John LaCava, Joanna Kufel and Rafal Tomecki for critical comments on the manuscript and members of the A.D., J.S.A. and T.H.J. laboratories for stimulating discussions. Kasia Kowalska, Ger Pruijn and Christian K. Damgaard are acknowledged for sharing reagents and for experimental advice. This work was supported by the Danish National Research Foundation, the Danish Cancer Society and the Lundbeck- and Novo Nordisk Foundations (to T.H.J.); the Foundation for Polish Science Team Programme co-financed by the EU European Regional Development Fund, the Operational Program Innovative Economy 2007-2013, the project support Agreement POIG.02.02.00-14-024/08-00 and an EMBO installation grant (to A.D.), and the European Commission's 7th Framework Programme (grant agreement HEALTH-F4-2008-201648/PROSPECTS (to J.S.A.). M.L. was supported by a Boehringer-Ingelheim PhD fellowship. The authors declare that they have no competing financial interests.

Figure legends

Fig. 1. hMTR4 interacts stoichiometrically with the exosome core

(A) hRRP6-FLAG SILAC co-IP result plotted by relative protein abundance (total peptide intensity divided by molecular weight (MW)) out the x-axis and log SILAC ratio (intensity of peptides originating from the hRRP6- *vs.* control-IP) up the y-axis. Note disruption of the x-axis to accommodate all detected proteins in the plot. Different types of interaction partners are indicated by color coding and relevant protein names are displayed. The entire data set is specified in Table S1. The asterisk (at hRRP6) indicates the FLAG-tagged bait protein.

(B) SDS-PAGE analysis of IP eluates from tet-induced (hRRP6-FLAG expressing) and un-induced HEK293 cells. Protein band ID's were achieved by MALDI-TOF MS. The migrations of IgG heavy- and light-chains as well as the HSPA1A protein contaminant are indicated.

(C) MS determination of relative levels of the indicated proteins in hRRP41-FLAG IP eluates from HEK293 cells depleted for hRRP6 *vs.* non-depleted cells. The calculation is done by MaxQuant (Cox and Mann, 2008), and based on the intensity of peptides from label-free experiments.

Fig. 2. Identification of the nuclear exosome targeting (NEXT) complex

(A) hMTR4-FLAG SILAC co-IP result plotted and labeled as in Fig. 1A. Highspecificity interactors with log SILAC ratio above 0.5 are indicated in orange. Note disruption of the x-axis to accommodate all detected proteins in the plot.

(B) Schematic outline of the two different SILAC purification strategies, 'mixed extracts' (ME) and 'mixed beads' (MB), to assay for dynamics of interactions. IP

experiments of differentially labeled samples are either carried out separately and mixing beads at the end (MB, left) or by mixing extracts and subjecting this material to common bead-binding and elution (ME, right). SILAC ratios that are constant between experimental strategies signify stable interactions.

(C) Interaction dynamics of most prominent hMTR4 binding partners. SILAC ratios are displayed for MB (black) and ME (red) experiments.

(D) ZCCHC8-FLAG co-IP result plotted and labeled as in Fig. 1A. The two groups of high-specificity (log SILAC ratio>0.5) and high-abundance (signal intensity/MW*10⁶>100) interactors are indicated in orange and blue, respectively. Note disruption of the x-axis to accommodate all detected proteins in the plot.

(E) Interaction dynamics of most prominent ZCCHC8 binding partners displayed as in(C).

(F) The NEXT complex: RBM7-EGFP fusion protein was purified from HEK293 cells using stringent conditions (500mM NaCl) and eluate was analyzed by SDS-PAGE. The MS-identification of commassie stained bands is indicated. Asterisks indicate contaminants.

Fig. 3. Differential nuclear partitioning of human exosome co-factors (A, B) Nuclear-localized hMTR4 and hTRF4-2 accumulate in nucleoli. ZCCHC7 strictly localizes to, and nuclear ZCCHC8 and RBM7 are excluded from, nucleoli. In (A), HeLa cells were transiently transfected with plasmids expressing the indicated proteins as C-terminal EGFP fusions. 12h post-transfection, protein expression was induced with tetracycline for 24h, cells were fixed and proteins were visualized by confocal microscopy. Fibrillarin staining served as a nucleolar marker and was overlayed with labeling of nuclei by Hoechst stain. In (B), HEK293 Flp-In T-REx cells

stably expressing N-terminally EGFP-tagged proteins were analyzed by live cell confocal microscopy. Protein expression was induced as above. Cells were visualized using phase contrast and overlaid with signal from EGFP fluorescence.

(C) ZCCHC8 and ZCCHC7 distribute in hDIS3- and hRRP6-containing low and high molecular weight glycerol gradient fractions, respectively. Western blotting analysis of 5%-40% glycerol gradient fractions of HEK293 cell extract, employing the indicated antibodies. Input material corresponding to 5% of the total cell extract was loaded in each of the two outer lanes of the gel.

Fig. 4. Interaction profiles of human putative TRAMP homologs

(A) Domain comparison of putative TRAMP subunits from *H. sapiens*, *D. melanogaster* and *S. pombe* with those of *S. cerevisiae* TRAMP components. Known domains are colored as indicated. For detailed sequence alignments see Fig. S4A-C. (B, C) ZCCHC7- and hTRF4-2-FLAG co-IP results plotted and labeled as in Fig. 1A. Only here, label-free IP's of ZCCHC7 and hTRF4-2 were conducted and peptide signal intensities were calculated by the label-free algorithm in MaxQuant software using normalization to the control (uninduced cell line) IP. Likely due to post-translational modifications, ZCCHC7 peptides were underrepresented in both MS spectra. Note disruption of the x-axes to accommodate all detected proteins in the plots. Full label-free datasets are labeled in grey.

(D) Verification of interactions by western blotting analysis. The hRRP6-, hMTR4-, ZCCHC8-, ZCCHC7- and hTRF4-2-FLAG eluates obtained after purification in the presence of 100mM, or 500mM, NaCl were probed with anti-hMTR4-, -hRRP6-, -ZCCHC8-, -ZCCHC7- and -hRRP40-antibodies as indicated. Asterisk denotes band corresponding to hRRP6 from previous hybridization.

Fig. 5. Substrate preference of NEXT reflects its sub-nuclear distribution

(A) Western blotting analysis of whole cell extracts showing protein depletion upon the indicated siRNA administrations. Top panels: HeLa cells were treated with the indicated siRNAs and control cells were treated with EGFP siRNA. Membranes were probed using specific antibodies as indicated to the right. Anti-Actin antibody was used as a loading control. HEK293 cells expressing EGFP-tagged RBM7 (bottom left) or FLAG-tagged hTRF4-2 (bottom right) were treated with the indicated siRNAs. In the EGFP-RBM7 experiment, TEL/AML siRNA was used as control. Protein depletion was assayed using anti-EGFP or anti-FLAG antibodies as indcated.

(B) PROMPTs are stabilized in cells depleted for exosome- and NEXT complexcomponents. Total RNA harvested from HeLa cells subjected to the indicated siRNA transfections were analyzed by dT-primed RT-qPCR using amplicons for the indicated PROMPTs; ID numbers from left to right: 40-9, -14, -16, -18, -38, -31, -13, -52, -33, -2b (Preker et al. 2008; Table S14). Data are displayed as mean values normalized to control (EGFP siRNA-treated cells). All data are normalized to GAPDH RNA as an internal control. Error bars represent standard deviations from biological repeats (n=3). Note disruption of y-axis to accommodate all data in the plot.

Fig. 6. Substrate preference of ZCCHC7 and hTRF4-2 reflects their sub-nuclear distributions

(A) Schematic representation of the 47S rRNA transcript. Boxes indicate the mature rRNAs 18S, 5.8S and 28S, which are flanked by the external spacers (ETSs) and separated by the internal spacers (ITSs).

(B) Schematic outline of the 5'ETS RNA adenylation assay. Adenylated 5'ETS RNAs arising from Actinomycin D-induced 47S rRNA degradation are reverse transcribed with the indicated dT-adaptor oligo followed by limited PCR using the indicated '5'ETS-1' and 'adaptor' primers. PCR products are subjected to Southern analysis using the 5'ETS-2 hybridization probe.

(C) Adenylation of 5'ETS degradation fragments is compromised upon ZCCHC7 or hTRF4-2 depletion. Southern blotting analysis of RT-PCR products derived from total RNA harvested from cells subjected to the indicated knock downs and using the reagents indicated in (B). A representative experiment from three repeats is shown. As an internal control, RT-PCR using GAPDH specific primers was performed. Probe signals were quantified, normalized to GAPDH levels and plotted relative to EGFP controls.

(D) hMTR4 and the core exosome are important for 5.8S rRNA 3'end prcessing. Total RNA harvested from HeLa cells subjected to the indicated siRNA-mediated knock downs was subjected to northern blotting analysis using a radiolabeled riboprobe targeting the ITS2 region (A). Mature 5.8S rRNA was visualized using a radiolabeled DNA oligo-probe. Probe signals arising from the 3'end extended species were quantified and plotted relative to EGFP controls.

Fig. 7. Sub-nuclear distribution of human nuclear exosome co-factors

Model overview of human nuclear exosome co-factors and their sub-nuclear localizations as derived from this study. The dually (nucleolar as well as non-nucleolar) localized hMTR4 (structure derived from its *S. cerevisiae* homolog (Weir et al., 2010) is centrally positioned and associates with the nuclear exosome (dashed arrows). In the non-nucleolar part of the nucleus, hMTR4 forms a stable trimeric

complex with ZCCHC8 and RBM7. This NEXT complex is excluded from the nucleoli, where hMTR4 instead cooperates with putative TRAMP homologous components ZCCHC7 and hTRF4-2. The latter is also present outside nucleoli.

Supplementary figure legends

Figure S1

(A) Western blotting analysis of hRRP6-FLAG SILAC mixed beads (MB) co-IP experiment. The analysis shows that the tagged protein was expressed at near endogenous level when induced with 30ng/ml of tetracycline. Input and eluate fractions were probed with α -RRP6 antibodies.

(B) Western blotting analysis of whole cell extracts showing hRRP6 depletion efficiency in tetracycline-induced cells, used for hRRP41 label-free co-IP. HEK293 cells were treated with siRNAs against hRRP6 and EGFP (control).

(C, E) Western blotting verification of hMTR4- and ZCCHC8-FLAG expression levels in cells induced with different concentration of tetracycline (top panel). The optimal expression conditions (closest to endogenous level, marked in red) were applied for subsequent experiments: SILAC mixed beads (MB) and mixed extracts (ME) co-IP (bottom panel; Fig. 2A, C), and label-free experiments (Fig. S1D, F ; S2A, B). Input and eluate fractions were probed with α -FLAG, and specific α -hMTR4 and α -ZCCHC8 antibodies, respectively.

(D, F) hMTR4- and ZCCHC8-FLAG label-free co-IPs performed in low stringent conditions (100mM NaCl). Results are plotted and labeled as in Fig. 2A. The entire dataset is specified in Table S3. Peptide signal intensities were calculated by the label-free algorithm in MaxQuant software, using normalization to the control (uninduced cell line) IP. Note disruption of the x-axes to accommodate all detected proteins in the plots. Names of high-specificity and high-abundance interactors were displayed. Full label-free datasets are labeled in grey.

Figure S2

(A, B) hMTR4 and ZCCHC8-FLAG label-free co-IPs performed in highly stringent (500mM) conditions. Results are plotted as in Fig. S1D with only NEXT components indicated. The calculations were performed as before (Fig. S1D, F).

Figure S3

(A) Intracellular localization of C-terminal EGFP-tagged components of the NEXT complex: ZCCHC8, RBM7 and hMTR; as well as hTRF4 and ZCCHC7. HEK293 cells stably expressing the fusion proteins were subjected to the standard immunofluoresce procedure followed by confocal microscopy. Hoechst stained nuclei and fibrillarin were overlayed as indicated.

(B) Nucleolar localization of ZCCHC7 protein in HeLa cells. The cells were subjected to α-ZCCHC7 immunostaining, which pattern was then overlaid with phase contrast light (left panel - nucleoli are pointed with arrows). Specificity of the ZCCHC7 signal was verified by treatment of HeLa cells with siRNA against ZCCHC7 or EGFP (control). Similar settings were used for visualizations.

Figure S4

(A-C) Multiple sequence alignments of yeast Air1p, Trf4p and Mtr4p with their closest homologs from human, fly and fission yeast, as determined by using Basic Local Alignment Search Tool (BLAST) on protein sequence. All alignments were done using the CLC main Main workbench 5.6.1 software, www.clcbio.com. For protein accession numbers used in this study see Table S2.

(D) Western blotting verification of ZCCHC7- and hTRF4-FLAG induction levels in label-free co-IPs. Cells were induced with the indicated concentration of tetracycline.

Conditions marked in red were chosen for MS analyses. Inputs and eluates were probed with α -FLAG and α -ZCCHC7 antibodies (there is no-available reagent to detect hTRF4-2).

References

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J *18*, 5399-5410.

Anderson, J.S., and Parker, R.P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J *17*, 1497-1506.

Bird, G., Fong, N., Gatlin, J.C., Farabaugh, S., and Bentley, D.L. (2005). Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing. Mol Cell *20*, 747-758.

Brown, J.T., Bai, X., and Johnson, A.W. (2000). The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. RNA *6*, 449-457.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cristodero, M., and Clayton, C.E. (2007). Trypanosome MTR4 is involved in rRNA processing. Nucleic Acids Res *35*, 7023-7030.

de la Cruz, J., Kressler, D., Tollervey, D., and Linder, P. (1998). Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. EMBO J *17*, 1128-1140.

Dez, C., Dlakic, M., and Tollervey, D. (2007). Roles of the HEAT repeat proteins Utp10 and Utp20 in 40S ribosome maturation. RNA *13*, 1516-1527.

Etheridge, R.D., Clemens, D.M., Gershon, P.D., and Aphasizhev, R. (2009). Identification and characterization of nuclear non-canonical poly(A) polymerases from Trypanosoma brucei. Mol Biochem Parasitol *164*, 66-73.

Gallenberger, M., Meinel, D.M., Kroeber, M., Wegner, M., Milkereit, P., Bosl, M.R., and Tamm, E.R. (2010). Lack of WDR36 leads to preimplantation embryonic lethality in mice and delays the formation of small subunit ribosomal RNA in human cells in vitro. Hum Mol Genet *20*, 422-435.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Gruber, J.J., Zatechka, D.S., Sabin, L.R., Yong, J., Lum, J.J., Kong, M., Zong, W.X., Zhang, Z., Lau, C.K., Rawlings, J., *et al.* (2009). Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. Cell *138*, 328-339.

Guo, T.B., Boros, L.G., Chan, K.C., Hikim, A.P., Hudson, A.P., Swerdloff, R.S., Mitchell, A.P., and Salameh, W.A. (2003). Spermatogenetic expression of RNAbinding motif protein 7, a protein that interacts with splicing factors. J Androl *24*, 204-214.

Gustafson, M.P., Welcker, M., Hwang, H.C., and Clurman, B.E. (2005). Zcchc8 is a glycogen synthase kinase-3 substrate that interacts with RNA-binding proteins. Biochem Biophys Res Commun *338*, 1359-1367.

Houseley, J., and Tollervey, D. (2006). Yeast Trf5p is a nuclear poly(A) polymerase. EMBO Rep 7, 205-211.

Houseley, J., and Tollervey, D. (2008). The nuclear RNA surveillance machinery: the link between ncRNAs and genome structure in budding yeast? Biochim Biophys Acta *1779*, 239-246.

Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. Cell *136*, 763-776.

Kammler, S., Lykke-Andersen, S., and Jensen, T.H. (2008). The RNA exosome component hRrp6 is a target for 5-fluorouracil in human cells. Mol Cancer Res *6*, 990-995.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *121*, 713-724.

Lejeune, F., Li, X., and Maquat, L.E. (2003). Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Mol Cell *12*, 675-687.

Lykke-Andersen, S., Brodersen, D.E., and Jensen, T.H. (2009). Origins and activities of the eukaryotic exosome. J Cell Sci *122*, 1487-1494.

Lykke-Andersen, S., and Jensen, T.H. (2006). CUT it out: silencing of noise in the transcriptome. Nat Struct Mol Biol *13*, 860-861.

Lykke-Andersen, S., Tomecki, R., Jensen, T.H., and Dziembowski, A. (2011). The eukaryotic RNA exosome: Same scaffold but variable catalytic subunits. RNA Biol 8.

McMahon, M., Ayllon, V., Panov, K.I., and O'Connor, R. (2010). Ribosomal 18 S RNA processing by the IGF-I-responsive WDR3 protein is integrated with p53 function in cancer cell proliferation. J Biol Chem *285*, 18309-18318.

Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A., and Tollervey, D. (2008). A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. Mol Cell Biol *28*, 5446-5457.

Milligan, L., Torchet, C., Allmang, C., Shipman, T., and Tollervey, D. (2005). A nuclear surveillance pathway for mRNAs with defective polyadenylation. Mol Cell Biol *25*, 9996-10004.

Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M., and Tollervey, D. (2003). Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. Mol Cell Biol *23*, 6982-6992.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. Cell *91*, 457-466.

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.

Nakamura, R., Takeuchi, R., Takata, K., Shimanouchi, K., Abe, Y., Kanai, Y., Ruike, T., Ihara, A., and Sakaguchi, K. (2008). TRF4 is involved in polyadenylation of snRNAs in Drosophila melanogaster. Mol Cell Biol *28*, 6620-6631.

Nielsen, A.F., Gloggnitzer, J., and Martinez, J. (2009). Ars2 and the Cap-Binding Complex Team up for Silencing. Cell *138*, 224-226.

Ong, S.E., Foster, L.J., and Mann, M. (2003). Mass spectrometric-based approaches in quantitative proteomics. Methods *29*, 124-130.

Peng, W.T., Robinson, M.D., Mnaimneh, S., Krogan, N.J., Cagney, G., Morris, Q., Davierwala, A.P., Grigull, J., Yang, X., Zhang, W., *et al.* (2003). A panoramic view of yeast noncoding RNA processing. Cell *113*, 919-933.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.

Reis, C.C., and Campbell, J.L. (2007). Contribution of Trf4/5 and the nuclear exosome to genome stability through regulation of histone mRNA levels in Saccharomyces cerevisiae. Genetics *175*, 993-1010.

Rougemaille, M., Gudipati, R.K., Olesen, J.R., Thomsen, R., Seraphin, B., Libri, D., and Jensen, T.H. (2007). Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. EMBO J *26*, 2317-2326.

Sabin, L.R., Zhou, R., Gruber, J.J., Lukinova, N., Bambina, S., Berman, A., Lau, C.K., Thompson, C.B., and Cherry, S. (2009). Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. Cell *138*, 340-351.

Sambrook, J., and Russell, D. (2001). Molecular cloning: a laboratory manual (2001) 3rd. Plainview (New York: Cold Spring Harbor Laboratories).

San Paolo, S., Vanacova, S., Schenk, L., Scherrer, T., Blank, D., Keller, W., and Gerber, A.P. (2009). Distinct roles of non-canonical poly(A) polymerases in RNA metabolism. PLoS Genet *5*, e1000555.

Schilders, G., and Pruijn, G.J. (2008). Biochemical studies of the mammalian exosome with intact cells. Methods Enzymol *448*, 211-226.

Schilders, G., van Dijk, E., and Pruijn, G.J. (2007). C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing. Nucleic Acids Res *35*, 2564-2572.

Schmidt, M.J., West, S., and Norbury, C.J. (2010). The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. RNA *17*, 39-44.

Shcherbik, N., Wang, M., Lapik, Y.R., Srivastava, L., and Pestov, D.G. (2010). Polyadenylation and degradation of incomplete RNA polymerase I transcripts in mammalian cells. EMBO Rep *11*, 106-111.

Staals, R.H., Bronkhorst, A.W., Schilders, G., Slomovic, S., Schuster, G., Heck, A.J., Raijmakers, R., and Pruijn, G.J. (2010). Dis3-like 1: a novel exoribonuclease associated with the human exosome. EMBO J *29*, 2358-2367.

Stead, J.A., Costello, J.L., Livingstone, M.J., and Mitchell, P. (2007). The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. Nucleic Acids Res *35*, 5556-5567.

Tomecki, R., Kristiansen, M.S., Lykke-Andersen, S., Chlebowski, A., Larsen, K.M., Szczesny, R.J., Drazkowska, K., Pastula, A., Andersen, J.S., Stepien, P.P., et al.

(2010). The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. EMBO J *29*, 2342-2357.

Torchet, C., Bousquet-Antonelli, C., Milligan, L., Thompson, E., Kufel, J., and Tollervey, D. (2002). Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs. Mol Cell *9*, 1285-1296.

van Dijk, E.L., Schilders, G., and Pruijn, G.J. (2007). Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways. RNA *13*, 1027-1035.

van Hoof, A., Lennertz, P., and Parker, R. (2000). Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. Mol Cell Biol *20*, 441-452.

Vasiljeva, L., and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. Mol Cell *21*, 239-248.

Weir, J.R., Bonneau, F., Hentschel, J., and Conti, E. (2010). Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. Proc Natl Acad Sci U S A *107*, 12139-12144.

Wery, M., Ruidant, S., Schillewaert, S., Lepore, N., and Lafontaine, D.L. (2009). The nuclear poly(A) polymerase and Exosome cofactor Trf5 is recruited cotranscriptionally to nucleolar surveillance. RNA *15*, 406-419.

Wilson, M.D., Wang, D., Wagner, R., Breyssens, H., Gertsenstein, M., Lobe, C., Lu, X., Nagy, A., Burke, R.D., Koop, B.F., *et al.* (2008). ARS2 is a conserved eukaryotic gene essential for early mammalian development. Mol Cell Biol *28*, 1503-1514.

Win, T.Z., Draper, S., Read, R.L., Pearce, J., Norbury, C.J., and Wang, S.W. (2006). Requirement of fission yeast Cid14 in polyadenylation of rRNAs. Mol Cell Biol *26*, 1710-1721.

Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., *et al.* (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725-737.

Figures 1-7 Click here to download Figure: Figures 1-7 final.pdf



Figure 2



Hoechst/Fibrillarin C-terminal EGFP

В



HeLa transiently transfected cells (fixed)



HEK293 stable cells (live imaging)

С



Glycerol gradient analysis

Figure 4



D



Figure 5





Figure 6







Supplemental Figures 1-4 Click here to download Supplemental Figure: Figures S1-S4 final.pdf

Figure S1













α-ZCCHC7

siRNA control

siRNA treated

HeLa cells (fixed)

HEK293 cells (fixed)

А

Sequence alignment of ZCCHC7 with its closest homologs

		10								10				140		180
5. cerevisiae AIR1p	M					STLLSEVE	5	IDTLP	YVKDTTPTGS		DSSSF			*********		
S. cerevisiae AIR2p	M					EK		NTAP	EVVDTAPTTP		P					
S. pombe AIR1	M					TYSNOKAE	SDDGSDIDDA	ALLOKINSLP	IDOSITNEVS	LEKHDFOGSD	DHD\$\$					
H. sapiens ZCCHC7	MM				FOGYETIE	AVEDDLYRDE	SSELSVDSE	VEFOLYSOIN	YAODLDDVIR	EEEHEEKNSG	NSESSSSKPN	QKKLIVLSDS	EVIQUADESE	VITLSDEDSI	YRCKOKNVRV	GAGENAHOLS 120
0. melanogaster SD06557n	MDATDYKMGL	SCITEPEGTS	VVASSTTPPM	HSPSCSVMSS	DEFIVORDIT	RULTECENAN	DEDLLVLTET	AIREAIGEEP	KOKEVVENAS	FEGEHOHAAD	TSSEVEEVPP	SRLEELKKNY	RVDDOOFRAL	DVYESESDLT	ESGLYSKAKS	KTTPTILRNY 160
		192		200		220		240		200	10	200		300		120
S. cerevisine AIR10		KLLAPSIE		DYDAN	PEELPTLRG.										OGRYFGITDY	DENGA IMEAE 73
E commission AIR20		KIVAPSIE		EVNEN	PNPI PAL PG	7.1.0.000.07		0.01991019991					10 - H 10 - 50 - 10 - 10 - 10 - 10 - 10 - 10 -		OGRYEGUED.	DOXDALKEAA 10
C nomine AID1		DISDETIE		DVEGS	EWLOV. S.										POPYEGS	DDSESIV
a pomperview	PRI ORNEL NO	VYCYPDIEVE	VALEBROUTE	EUMILENERS	EREPTIERO	DUVERNMILO	*					EURA	VARBULLNUV	POR NEWTERS	DOLINIELEON	DIELOIANNE ST
P. sapos zourur	DECCORECUE	ENUE DEVENT	KOLONOGOST	EVMITEYSSS	COOL COSCO	Develamento		VANIDALATA	HEADDHARKS	ATRIOKOKU			LAPPILLUCCE	CENSTIEGE	DODDDJIISUK	DIENGIARIA 22
D. melanogamer SU0055/p	USSSUSSSVE	EVVEPSVUNT	ARLAAASUST	NARSUSUPAN	DOGLODSDSD	DOVHATOVPO	TARUMAVERC	ARAINRISIR	HSSUUNSKAA	ALKINAPATA	AVBRADDEDA	SEDELPSARD	TAERLENUUK	GUGKEQVORG	DODDDATSTO	SUEUSHAEAE an
		T		T		+		1		7		7		7		
5. cerevisiae AIR1p	PK															····· CNNC IV
b. cerevisiae Aikzp	PK															
5. pombe AIR1																CHNC 92
H. sapiens 2CCHC7	TPGRWTQRYY	SA						********							*********	-NKNI I CRNC 24
D. melanogaster SD06557p	FRDAMTORLA	AVFERLDAKA	KKMEEDEEQV	NASGYISDEE	DSRDKSADTT	VRSDEDIDAS	NDLEETNPDK	NORDVTNEEQ	HVHVEHNLQE	LRDESLPKSG	DLIGWNEEMR	SFYNDSWNGE	NFSVHKVQRM	MTGHRHEWRI	NTADRCPVAR	PRSHAKCSNC 440
	* *	*	* *	* *	* * *	* 54	* *	* *	Carried States	1 125 125 125 125 1		500				1
5. cerevisiae AIR1p	SOROHLKRNC	PH	ICTYCOFMDD	HYSOHCPKAI	ICTNCNANGH	-YKSQC PH	KWKKVFCTLC	NSKRHSRERC	PSIWRSYLLK	TKDTNOGD	••F•••••	*********				168
S. cerevisiae AIR2p	SORGHLKKDC	PH	ICSYCGATOD	HYSRHCPKAI	QCSKCDEVGH	-YRSQC PH	KWKKVQCTLC	KSKKHSKERC	PSIWRAVILV	DDNEKAKP	KVL	*********				157
S. pombe AIR1	KGNGHISKDC	PH	LCTTCGAIDD	HISVRCPWTK	KCMNCGLLGH	- LAARCSEPR	KRGPRVCRTC	HTDTHTSSTC	PLIWRYYV	EKEHP	VR1					180
H. sapiens ZCCHC7	DKRGHLSKNC	PLPRKVRRCF	LCSRRG	HLLYSCP - AP	LCEYCPVPKM	-LDHSCLFRH	SWDK - QCDRC	HMLGHYTDAC	TEIWROYHLT	TKPGPPK - KP	KTP					
D. melanogaster SD06557p	FEMGHVRSKC	PRPRKPLVCF	ICGTMG	HAEPRCPNAL	-CFGCGSKQE	I YVQQCNKCS	FHSRLVCOLC	KMRGHGVDHC	POKWRRYHST	TRANTELDAR	VOYRSVOCSY	CAGRHSFENC	RORIGOFRST	NYTSQIVSHQ	KIYKNROGPI	GNLODLKSFF 63
		*		640	l.	700		721		740		790		180		800
5. cerevisiae AIR1p		DFOTV FCY														
S. cerevisiae AIR2p		DENTI-														and the second
		P P P P P P P P P P P P P P P P P P P														
S. northe AIR1		DUSEVRKICY														
S. pombe AIR1 H. saciens 200HC7		DVSEVRKICY														····· 18
S. pombe AIR1 H. sepiets 200HC7 D. melanopauler 5006557a	ALETPPHEKW	DVSEVRKFCY SRPSALAYCY NNPSMPKNCY	YSRFLVNANL	AKVQSVKRKS	TTAVSOIPEK	INTHENNEN	OVKAARGEVS	TGTAKKPLTK	OMPYVESIAE	NRTAEEPIGK	LNREDEKSKD	ILPLPEVPPE	VKOTODEKLV	EGOPNONOPD	EELPMKDDDQ	EEPSSSVORE 700
S. pombe AIR1 H. sapiens 200HC7 D. melanogaster SD06557p	ALETPPHEKW	DVSEVRK CY SRPSALAKCY NNPSMPKNCY	YSRFLVNANL	AKVQSVKRKS	TTAVSQIPEK	IVTHEVVPNA	QVKAARGEVS	TGTAKKPLTK	QMPVVESTAE	NRTAEEPIGK	LNREDEKSKD	ILPLPEVPPE NO	VKQTQDEKLV	EGGPNQNQPD	EELPMKDDDQ	EEPSSSVORE 797
S. ponthe AIR1 H. sepiens 200H07 D. melanogaster SD06567p	ALETPFHFKW	DVSEVRK CY SRPSALAKCY NNPSMPKNCY	YSRFLVNANL	AKVQSVKRKS * * ***	TTAVSQIPEK	IYTHEYVPNA	QVKAARGEVS	TGTAKKPLTK	QMPVVESTAE	NRTAEEPIGK	LNREDEKSKD	ILPLPEVPPE	VKQTQDEKLV	EGOPNONOPD	EELPMKDDDQ	EEPSSSVQRE 79
S. ponthe AIR1 H. sapiens 200H07 D. melanogaster SD06567p S. convisiae AIR1p S. convisiae AIR1p	AIETPFHFKW	DVSEVRK CY SRPSALAKCY NNPSMPKNCY	YSRFLVNANL	AKVQSVKRKS * * *** CGNAGHFG CQDKGHFG-	TTAVSQIPEK	IVTHEVVPNA	QVKAARGEVS	TGTAKKPLTK	QMPVVESIAE	NRTAEEPIGK *****	LNREDEKSKD RR-SSRVPNT KR-SSRVPNF	ILPLPEVPPE BG	VKQTQDEKLV	EGGPNQNGPD	EELPMKDDDG	EEPSSSVQRE 790 80 80 201
S. ponte AIR1 H. sepiens 200407 D. melanogaster 5006557p S. cenevisiae AIR1p S. cenevisiae AIR1p	AIETPFHFKW	DVSEVRK CY SRPSALACY NNPSMPKNCY	YSRFLVNANL	AKVQSVKRKS * * * * CGNAGHFG CGGKGHFG CGGKGHFG	TTAVSQIPEK		QVKAARGEVS	TGTAKKPLTK	QMPVVESIAE	NRTAEEPIGK ****	LNREDEKSKD RR-SSRVPNT KR-SSRVPNE	ILPLPEVPPE NP DG	VKQTQDEKLV	EGGPNQNQPD	EELPMKDDDG	EEPSSSVQRE 799
S. pontre AIR1 H. sepiens 2004027 D. melanogaster SD06557p S. cenevisiae AIR1p S. cenevisiae AIR1p S. pombe AIR1	AIETPPHPKW	DVSEVRK CY SRPSALAYCY NNPSMPKNCY 88	YSRFLVNANL	AKVQSVKRKS * * **** CGNAGHFG- CGCKGHFG- CASDEHFG CASDEHFG	TTAVSQIPEK		QVKAARGEVS		OMPVVESTAE	NRTAEEPIGK ************************************	LNREDEKSKD RR-SSRVPNT KR-SSRVPNE PS-RSNYP	ILPLPEVPPE 800 DG	VKOTQDEKLV	EGGPNQNOPD	EELPMKDDDQ	EEPSSSVQRE 799 900
S. pontie AR1 H. sapiens 200402 D. metanogeter 500657p S. cerevisiae AIR1p S. cerevisiae AIR1p S. pombe AIR1 H. sapiens 200407	A I ETPPHPKW	DVSEVRK CY SRPSALAKCY NNPSMPKNCY NPSMPKNCY	YSRFLVNANL	AKVQSVKRKS * * *** CGNAGHFG- CGSKGHFG- CASDEHFG- CASDEHFG-	TTAVSQIPEK		QVKAARGEVS	TGTAKKPLTK	QMPVVESIAE	NRTAEEPIGK ************************************	LNREDEKSKD RR - SSRVPNT KR - SSRVPNE PS - RSNYP REVYDPSPVS	ILPLPEVPPE 500 00 ES PF	VKQTQDEKLV	EGGPNONGPD	EELPMKDDDG	EEPSSSVQRE 79
5. pombe AR1 H. aspiens 200H07 D. melenogetor 5006557p S. convisiae AIR1p S. pomevisiae AIR1p S. pombe AIR1 H. aspiens 200H07 D. melanogaster 5006557p	AIETPPHFKW	DVSEVRK CY SRPSALAKCY NNPSMPKNCY TEPGLSRTPL	YSRFLVNANL	AKVQSVKRKS * * * *** CGNAGHFG CGCKOHFG CASDEHFG CASKGHYG YSFSEHFELP	TTAVSQIPEK	IVTHEYVPNA m	QVKAARGEVS	TGTAKKPLTK	GI SMCVNSSS	NRTAEEPIGK **** ODCAE ODCKE ODCTL HECPE TSLRTOETPE	LNREDEKSKD RR-SSRVPNT KR-SSRVPNE PS-RSNYP REVYDPSPVS PDSSEELPDV	ILPLPEVPPE 500 00 ES PF PSEGKIIMAR	VKQTQDEKLV	EGGPNQNOPD	EELPMKDDDQ	EEPSSSVQRE 790 800 201 201 201 201 201 201 201 201 201 2
S. pombe AR1 H. aspens 200Hc7 D. melanogaster 5D0657p S. correvisiae AIR1p S. pombe AIR1 H. aspens 200Hc7 D. melanogaster 5D0657p	AIETPPHPKW	DVSEVRK CY SRPSALAKCY NNPSMPKNCY NP TEPGLSRTPL	YSRFLVNANL	AKYQSYKRKS * * *** CGNAGHFG- CASDEHFG- CASDEHFG- CASKGHYG- YSFSEHFELP	TTAYSQIPEK	IVTHEVVPNA See	QVKAARGEVS	TGTAKKPLTK GRA DDFELSSNVQ Can 1	GI SMCVNSSS	NRTAEEPIGK	LNREDEKSKD RR-SSRVPNT KR-SSRVPNE PS-RSNYP REVYDPSPVS PDSSEELPDV	ILPLPEVPPE B00 DG DG ES PSEGKIIMAR CIM I	VKOTODEKLV	GRTFLAAAAK	EELPMKDDOG QCHVSVRMDF	EEP\$\$\$\$\U00e9 000 000 000 000 000 000 000
5. pombe AR1 H. aspens 200Hc7 D. melanogaster SD0657p S. cerevisiae AR1p S. cerevisiae AR1p S. pombe AR1 H. aspins 200Hc7 D. melanogaster SD0657p S. cerevisiae AR1p	AIETPPHPKW	DVSEVRK CY SRPSALACY NNPSMPKNCY NS SS SS SS SS SS SS SS SS SS SS SS SS	YSRFLVNANL N APHELDSDSN	AKVQSVKRXS * * *** CCNAGHFG - CCASDEHFG - CASDEHFG - CASDEHFELP 100 1	TTAVSQIPEK	TQLPLVRTMG	QVKAARGEVS	DDFELSSNVQ	GMPVVESIAE GISMCVNSSS	NRTAEEPIGK **** ODCAE ODCTL HECPE TSLRTOETPD	LNREDEKSKO RR-SSRVPNT KR-SSRVPNE PS-RSNYP- REVYDPSPVS PDSSEELPDV	ILPLPEVPPE sep DG DG ES PF PSEGKIIMAR cmm i	VKQTQDEKLV	EGGPNGNGPD HI GRTFLAAAAK	EELPMKDDDQ QCHVSVRMDF AFCGDNL	EEPSSSVQRE 795 80 201 201 201 201 201 201 201 20
5. pombe AIR1 H. sepiene ZCOHCF D. melanogaster SD005579 S. comvisiae AIR1p S. convrisiae AIR1p H. sepiene ZCCHCF D. melanogaster SD00579 S. correvisiae AIR1p S. correvisiae AIR1p	A I ETPPHPKW	DVSEVRK CY SRPSALA CY NNPSMPKNCY III TEPGLSRTPL	YSRFLVNANL N N APHELDSDSN	AKVQSVKRKS * * **** CGNAGHFG CGKRGHFG CASDEHFG CAGKGHYG TSFSEHFELP 100 1	TTAVŠQIPEK SSTTSDOQPN	I YTHEYYPNA min TQLPLVRTMO	QVKAARGEVS TLPDVIPLSK	TGTAKKPLTK 600 DDFELSSNVQ 1	GMPVVESTAE GISMCVNSSS	NRTAEEPIGK *** •••• ODCAE •••• ODCKE •••• ODCKE •	LNREDEKSKD RR - SSRVPNT KR - SSRVPNE PS - RSNYP REVYDPSPVS PDSSEELPDV	ILPLPEVPPE ND DG DG ES PF PSEOKIIMAR CM I	VKQTQDEKLV	EGOPNONOPD HE ORTFLAAAAK 50 S	EELPMKDDDQ OCHVSVRMDF AFCODNL AFTGSNL	EEP\$\$\$VORE 795 000 000 000 000 000 000 000 0
5. ponte Alti H. spenz 200407 D. melanogate S00657p S. convisiae AIR1p S. convisiae AIR1p S. spenz 200407 D. melanogaster S00657p S. convisiae AIR1p S. convisiae AIR1p S. ponte AIR1	AIETPPHPKW	DVSEVRK(CY SRPSALACY NNPSMPKNCY I	YSRFLVNANL N N APHELDSDSN	AKVOSVKRKS * **** CGNAOHFG - CGGKGHFG - CASCEHFG - CASCEHFG - YSFSEHFELP ton	TTAVSQIPEK SSTTSDOOPN	I YTHEYYPNA see TQLPLVRTMO	QVKAARGEVS TLPDVIPLSK	TGTAKKPLTK see DDFELSSNVG tm i	GISMCVN555	NRTAEEPIGK **** ····ODCAE ····ODCKE ····ODCKE ····ODCKE ····ODCKE ····ODCKE ····ODCKE ····ODCKE ····ODCKE ·····ODCKE ·····ODCKE ·····ODCKE ·····ODCKE ·····ODCKE ······ODCKE ······ODCKE ····································	LNREDEKSKD RR-SSRVPNT KR-SSRVPNE PS-RSNYP- REVYDPSPVS PDSSEELPDV	ILPLPEVPPE Buo DG DG ES PF PSEGKIIMAR Cam	VKOTODEKLV EESEYLFSAE	EGGPNGNGPD HI GRTFLAAAAK S S S T T	EELPMKDDDQ QCHVSVRMDF AFCQDNL AFCQDNL AFCQNL	EEP\$\$\$VQRE 79 100 100 100 100 100 100 100 10
5. ponte ARI H. sepiens 20CHC7 D. metanogater 5006557p S. convisiae AR9, S. ponte ARI H. sepiens 20CHC7 D. metanogaster 5006557p S. convisiae AR9, S. ponte ARI H. sepiens ACHC7	AIETPPHFKW	DVSEVRK CY SRPSALA CY NNPSMPKNCY TEPGLSRTPL Main J	YSRPLVNANL N N APHELDSDSN	AKYOSYKRKS * * *** CGNAGHFG CGGKGHFG CASDEHFG- CASDEHFG- CASDEHFG- TSFSEHFELP too 1	TTAVSQIPEK SSTTSDOQPN	IVTHEYVPNA min TQLPLVRTMG	QVKAARGEVS TLPDVIPLSK	TGTAKKPLTK see DDFELSSNVG tak	GISMCVNSSS	NRTAEEPIGK	LNREDEKSKD RR-SSRVPNT KR-SSRVPNP PS-RSNVP- PS-SSVP PDSSEELPDV	ILPLPEVPPE B0 DG DG ES PF PSEGKI MAR Cmm I	VKGTGDEKLV EESEYLPSAE	EGGPNQNOPD HE ORTFLAAAAX S S S S S S S S S S S S S S S S S	EELPMKDDDG GCHVSVRMDF AFCGDNL AFTGSNL AFCENC CYYDDKY	EEPSSVQRE 79 800 201 201 201 201 201 201 201 201 201 2
5. ponde ARI H. sapens 20CHC7 D. melanogater 5006057p S. convisiae AR31p S. convisiae AR31p S. ponde AR1 H. sapiens 20CHC7 D. melanogater 500657p S. ponde AR1 H. sapiens 20CHC7 D. melanogater 5006557p	AIETPPHFKW	DVSEVRK CY SRPSALA CY NNPSMPKNCY TEPGLSRTPL Manuary VKLLRRSOEV	YSRPLVNANL N APHELDSDSN KRKSIEFQSQ	AKYOSYKRKS * * *********************************	TTAVSQIPEK SSTTSDOQPN RFLRDGINSL	TQLPLVRTMG	QVKAARGEVS TLPDVIPLSK	TGTAKKPLTK Gen DDFELSSNVG CM I NTKNGFKMAE 100	GISMCVNSSS	NRTAEEPIGK	LNREDEKSKD RR-SSRVPNT KR-SSRVPNP PS-RSNVP- PS-SSVPN PDSSEELPDV THLDOLLVLL	ILPLPEVPPE B0 DG DG ES PF PSEGKIIMAR Cmm I RRLIDOSSPD 120	EESEYLFSAE	EGGPNQNOPD HE ORTFLAAAAX S S S LEEHWRMIFS I	EELPMKDDDG QCHVSVRMDF AFCGDNL AFTGSNL AFTGSNL CYYDDKY SYPHPNYDSL	EEPSSSVGE 79 m0 201 202 202 202 202 202 202 20
5. porde AR1 H. sepies 200407 D. melanopater 5006557 S. cervisiae AR51 S. cervisiae AR19 S. porte AR1 H. sepies 200457 D. melanopater 500657 S. portesiae AR19 S. cervisiae AR19 D. melanopater 5006657 D. melanopater 500666	AIETPPHPKW	VSEVARK CY SRPSALACY NNPSMPKACY TEPGLSRTPL	YSRFLVNANL N N APHELDSDSN KRKSIEFOSQ	AKVQSVKRKS * * *********************************	TTAVSQIPEK SSTTSDQQPN 	TQLPLVRTMG	QVKAARGEVS TLPDVIPLSK	TGTAKKPLTK BD DDFELSSNVQ Cas I NTKNGFKMAE SPASOBORE	OMPVVESIAE GISMCVNSSS KKRROLNMIL	NRTAEEPIGK	LNREDEKSKD RR-SSRVPNI KR-SSRVPNI PS-RSNYP- REVYDPSPVS PDSSELPDV THLDOLLVLL	ILPLPEVPPE BU DG ES PF PSEGKIIMAR RRLIDOFSPD STR WWKEVYOSTR	VKOTODEKLV EESEYLFSAE EHATPOLRNE	EGOPNQNOPD	EELPMKDDDO OCHVSVRMDF AFCGDNL AFTGSNL AFTGSNL CYYDDKY- SYPAPNYDSL	EEP\$\$\$VQET 75 80 20 20 21 20 20 20 20 20 20 20 20 20 20 20 20 20
S. portek ART H. saylers 200467 D. metanogaster 5006557 S. comvisien ARD S. portek ARD	AIETPPHPKW EITTNPDNPP GLKKHOEELQ	TEPGLSRTPL VKLLRRSDEV	YSRFLVNANL N APHELDSDSN KRKSIEFOSQ	AKVQSVKRKS CGNAGHFG- CGSRAHFG- CASCB	TTAVSQIPEK SSTTSDOQPN RFLRDGINSL	TQLPLVRTMG	QVKAARGEVS TLPDVIPLSK VVRIKELEGM LKDNK	TGTAKKPLTK	QMPVVESIAE GISMCVNSSS KXRRQLNMIL NEHEFNLLDY	NRTAEEPIGK	LNREDEKSKD RR-SSRVPNE PS-RSNVP- REVYDPS-VS PDSSEELPDV THLDOLLVLL PGSRTYRDKM	ILPLPEVPPE BO DG ES PSEGKIIMAR can RRLIDOFSPD San KWKKKVOSTR	VKQTQDEKLV EESEYLFSAE EHATPQLRNE NKNSSNN-	EGOPNONOPD	EELPMKDDDQ GCHVSVRMDF AFCGDNL AFTGSNL AFCEANC CYYDDKY SYPAPNYDSL 	EEPSSSVGE 77 20 21 21 21 21 21 21 21 21 21 21
 5. porte ARI; H. signitis - ZCHC2 D. melanigenets 500657p S. convision ARI1p G. convision ARI2p S. porte ARI1; H. sagens 2CCHC2; D. melanigenets 500657p S. convision ARI2p S. porte ARI1; H. sagens 2CHC2; D. melanigenets 500657p G. melanigenets 500657p S. convision ARI2p S. conversion ARI2p S. conversion ARI1p S. conversion ARI2p 	A I ETPPHPKW	TEPOLSRTPL VKLLRRSOEV	YSRFLVNANL N APHELDSDSN KRKSIEFOSQ	AKVQSVKRKS * **** CGNAOHFG CGAKOHFG CASDEHFG CASDEHFG CASDEHFELP **** ******************************	TTAVSQIPEK SSTTSDQQPN RFLRDQINSL 	TQLPLVRTMG	QVKAARGEVS TLPDVIPLSK VVRIKELEGM 	TGTAKKPLTK	GISMCVNSSS KXRRQLNMIL NEHEFNLLDY	NRTAEEPIGK	LNREDEKSKD RR-SSRVPNE PS-RSNVP- REVYDPSVS PDSSEELPDV THLDOLLVLL PGSRTYRDKM PLPRPSHKRH	ILPLPEVPPE DG CS ES PF SCALIMAR RRLIDDFSPD RRLIDDFSPD KWKGKVOSTR SQNDHSHSGR	VKOTODEKLV EESEYLFSAE EHATPOLRNE NKNSSNN	EGOPNONOPD III ORTFLAAAAK 50 5 5 5 7 1 IEEHWRMIFS PPYOKS PPYOKS	CCHVSVRMDF AFCGDNL AFTGSNL AFCEANC CYVDDKY SYPHPNYDSL NNRKK VVIQFTI	EEPSSSVGRE 77 80 80 80 80 80 80 80 80 80 80
S porte ARI H signs a COHC D indersgener 5000057 S comvise AND S porte ARI D signs a COHC D indersgener 500057 S comvise AND S porte ARI D signs 20057 D melangener 500057 D melangener 500	AIETPPHPKW EITTNPDNPP GLKKHQEELQ	VSEVAR CY SRPSALACY NNPSMPKNCY NNPSMPKNCY TEPGLSRTPL WKLLRRSGEV	YSRFLVNANL N APHELDSDSN KRKSIEFQSQ	AKVQSVKRKS CGNAGNFG- CGNAGNFG- CASGEHFG- CASGEHFG- CASGEHFG- CASGEHFG- T	TTAVŠQIPEK SSTTSDQOPN RFLRDQINSL 	TQLPLVRTMG TQLPLVRTMG TQLPLVRTMG ELKOEYVRI	QVKAARGEVS TLPDVIPLSK VVRIKELEGM LKDNK MRRS SG	TGTAKKPLTK MARKEPLTK DDFELSSNVO TO TO TO TO TO TO TO TO TO T	GISMCVNSSS GISMCVNSSS KKRRQLNWIL NEHEFNLLDY TSRKEF	NRTAEEPIGK	LNREDEKSKD RR-SSRVPNT PS-RSNYP- REVYDPS-VS PDSSEELPDV THLDOLLVLL POSRTYRDKM PLPRPSKL	I LPLPEVPPE DG DG ES PSEGKI IMAR RRL I DOFSPD tal RRKKVOSTR SONDH'SKSTR SONDH'SKSTR	VKQTQDEKLV EESEYLFSAE EHATPOLRNE NKNSSNN- NKHRASNFHP NRNOKSSKOR	EGOPNONOPD GRTFLAAAAK 1 M S S S S S S S S S S S S S	EELPMKDDDQ OCHVSVRMDF AFCGONL AFTGSNL AFCGNC CYVDDKY SYPHPNYDSL NVIGPTI NVIGPTI	EEPSSSVORE 75 80 10 10 10 10 10 10 10 10 10 1
5 porte AW H sayes 2CHCF 0 metrogener SCORS7 5 cenvise AH2 5 porte AW 5 cenvise AH2 5 cenv	AIETPPHPKW EITTNPDNPP GLKKHOEELQ	VVSEVAR CY SRPSALACY NNPSMPKKCY TEPOLSRTPL VKLLRRSDEV C	YSRPLVNANL N APHELOSDSN KRKSIEFOSG	AKVQSVKRKS CGNAGHFG CQQKQHFG CQQKQHFG CASCEHFG CASCEHFG VSFSEHFELP KPPKRIDVLI	TTAVSQIPEK SSTTSDQQPN RFLRDGINSL AT 		QVKAARGEVS TLPOVIPLSK VVRIKELEGM 	DDFELSSNVQ DDFELSSNVQ Ca NTKNGFKMAE REASQROHD DE PEPSKLPYIK	GISMCVNSSS KKRRQLNMIL NEHEFNLLDY NE	NRTAEEPIGK	LNREDEKSKD RR-SSRVPNE PS-SSRVPNE PS-SSRVPNE PDSSEELPDV THLDOLLVLL PGSRTYRDKM PLPRPSHKRH	ILPLPEVPPE Base DG CS PSEOKIIMAR RRLIDOPSPD SONDHSHSGR RWPOENKETO POENKETO	VKQTQDEKLV EESEYLPSAE EHATPQLRNE NKNSSNN- NKHRASNPHP NRNQKSPHP	EGOPNONOPD GRTFLAAAAK 10 S S T IEEEWRMIFS RYESG PPYOKS PPYOKS PPYOKS	EELPMKDDDQ QCHVSVRMDF AFCGONL AFTGSNL- AFTGSNL- CYVDDKY- SYPAPANTOSL NNRKK NNRKK - NNRKK NNRKK NNRKK NNRKK	EEPSSSVQRE 79 mp 10 10 10 10 10 10 10 10 10 10
S porte ARI H signer acceler D melanganter 500657p S comvise ARIDS S comvise ARIDS S porte ARI H signer 2COHCT D melanganter 500657p S porte ARIT D separat 200457 D melanganter 500657p S comvise ARIDS S comvise ARITS S comvise ARITS S comvise ARITS S comvise ARITS S separat 200457 D melanganter 500657p	AIETPPHPKW EITTNPDNPP GLKKHGEELG	VVSEVAR CY SRPSALACOV TEPGLSRTPL	YSRPLVNANL N APHELDSDSN KRKSIEFQSQ	AKVQSVKRKS * * *** CGNAQHFG CGNAQHFG CASDEHFG CASDEHFG KPPKRIDVLI *** PPSPTPLPPY	TTAVSQIPEK SSTTSDGOPN RFLRDGINSL AT SV P. RGDMAPPPPP	TQLPLVRTMG TQLPLVRTMG NTNLONAKSH KFKOHYFNG ELKOEYFNG GEREKRLKQK KMSKKKORGA	QVKAARGEVS TLPDVIPLSK VVRIKELEOM 	TOTAKKPLTK BDDFELSSNVO DDFELSSNVO III NTKNGFKMAE III NTKNGFKMAE IIII DDF PEPSKLPYK QOQOQOGLQ	GMPVVESIAE GISMCVNSSS KKRROLNMIL NEHEFNLLDY NEHEFNLLDY TRRK	NRTAEEPIOK	LNREDEKSKD RR-SSRVPNT KR-SSRVPNE PS-RSNY- PDSSEELPDV THLDOLLVLL PGSRTYRDKM PLPRPSHKRH	ILPLPEVPPE BO DG DG PSEGKIIMAR PSEGKIIMAR RRLIDDFSPD STAN KWKOKVOSTSR SONDOHSTSR SONDOHSTSR SONDOHSTSR RECORMETO	VKQTQDEKLV EESEYLFSAE EHATPOLRNE NKNSSNF NRNGVSSKQB KEMKIKIRNW	EGGPNONGPD With the second se	CCHVSVRMOF AFCGONL AFTGSNL AFTGSNL SYPHPNYDSL NNRKK NVIGPTI NSLA NVIGPTI 	EEPSSSVORE 750 800 101 101 101 101 101 101 10
5 porte AW H sayes 2CHC? C melanogenetr 500657 5 cenviae AH2 5 porte AW 5 porte	A LETPFHFKW	DVSEVER CY SRPSALACY NPSPPNCY TEPOLSRTPL VXLLRRSOEV VLLGLOOKKT	YSRFLVNANL N N APHELDSOSN KRKSIEFOSQ GSGPQSPSKQ	AXVOSVKRKS ************************************	TTAVSGIPEK SSTTSDOOPN RFLRDGINSL SV F RODMAPPPPP	TOLPLVHTMG TOLPLVHTMG 10 NTNLONAKSH KFKOHTFNO ELKOEYYRH OGREKKRCAK	QVKAAROEVS TLPDVIPLSK VVRIKELEOM 	TOTAKKPLTK B DDFELSSNVG US NTKNGFKMAE REASCROHD DE NDASSNEFFE PEPSKLPYIK GGOGOGORLO IIII IIIII IIIIIIIIIIIIIIIIIIIIIII	OMPVVESIAE GISMCVNSSS KXRRQLIMIL NEHEFNLLOY NE	NRTAREPIOK	LNREDEKSKO RR-SSRVPHE KR-SSRVPHE PS-RSNVP- PS-RSNVP- PDSSEELPOV THLDOLLVLL POSRTVRDKM PLPRPSNKRH 	ILPLPEVPPE DD	VKOTODEKLV EESEYLFSAE EHATPOLRHE NKNSSNN- NKHRASSIFHP NKNRASSIFHP KEMKIKIRHW HOQODINGYQ	EGGPNONOPD BE GRTFLAAAAK 10 10 10 10 10 10 10 10 10 10 10 10 10	EELPMKDDDG QCHVSVRMDF AFCGDNL AFTGSN AFTGSN SILA SILA SILA SILA SHPMPIYDSL ORHEY MPYNNLGQRS	EEPSSSVQRE 79 30 30 30 30 30 30 30 30 30 30
 S. porte ARI; H. sugina scalar S. comvise ARI; S. convise ARI; 	AIETPFHPRW EITTNFDNPP GLKKHOEELO NRLPSLHIDP VTKNKKY	DYSEVER CY SRPSALACE. NPSSPPAC. TEPOLSRTPL VKLLRRSGEV VLLGLOOKT VLLGLOOKT	YSRFLVNANL N APHELDSDSN KRKSLEFQSQ QSQPQSPSXQ RSSQNNRTND	ARVOSVKRKS	TTAVSGIPEK SSTTSDOOPN RFLROGINSL 	TOLPLURING TOLPLURING NTNLONAKSH KEKORTPNO OFREKRIKOK KINSKKORA	QVKAARGEVS TLPDVIPLSK VVRIKELEOM KONK MMRIS ISVUKKIGU LDGOTONOGR S	TOTAKKPLTK	GISMCVNSSS GISMCVNSSS KKRRGLNMIL NEMEPNLOV NEM-OTARKFF AANENPHOL GGHONGOGRL Y 300	NRTAEEPIOK	LNREDEKSKD RR-SSRVPNE SPS SRVP- PS - SRVP- PD SSRVPNE PD SSEELPOV THLDOLLVLL POSRTVRDKM RASWKSN LOODHOOGRV	ILPLPEUPPE BU DG DG SES PF SEGKI I MAR RRL I DOPSPD USU KWKGKVOSTR SCHONSINGO SCHONSINGO RRCO GOGRMOOOOD	VKGTGDEKLV EESEYLFSAE EHATPOLRNE NKNSSNN NRNGSSKOR KKHSKINW HOGODNIGYG	EGGPMONOPO B GRTFLAAAAK S S IEEHMINA -RYESG PYINKG EKHRKA GNLRRPOIDG	EELPMK000G OCHVSVRMOF AFCGONL	EEPSSSVQR 72 an KOYGYVLVI 75 KOYGYVLVI 75 CO CO CO CO CO CO CO CO CO CO
 S. porte AWI H. sayara ZCCHC? Intelerogender S000507 C. merkina AH2 S. porte AWI H. sayara ZCCHC? Intelerogender S000507 C. cervisia AH2 S. porte AWI H. sayara ZCCHC? Intelerogender S000507 C. cervisia AH2 S. porte AWI S. pore	ALETPFHFKW EITTNFONPP GLKKHOEELO NRLPSLHIDP VTKNKRV	DUSEVER CU SRPALLACU REPORTANCE TEPOLSETPL WILLERSOEV VILLERSOEV VILLERSOEV CONFILE	XSRFLVNANL N N APHELOSOSN KRKSIEFOSQ OSOPOSPSKG RSSONNATNO	ARVOSVARAS * * ********************************	TTAVSGIPEK SSTTSDOGPN RFLEDGINSL 	TOLPLVRTMG TOLPLVRTMG TOLPLVRTMG TOLPLVRTMG KFKOMTROG ELKOEYYRH GEREKRLKOK KUSKKKORGA GESSIK	QVKAARGEVS TLPDVIPLSK VVRIKELEGM LKDKK	DDFELSSNVQ COFELSSNVQ Car REASCROHD DE NDASNKEFFF GOGGOGORIG GRN0R- YSSTLCIKE	GISMCVNSSS GISMCVNSSS KKEROLINII NEKEFNLLDY NE	NRTAREPICK	LNREDEKSKO RR-SSRVPHE PS-RSNVP- PS-RSNVP- PS-RSNVP- PS-RSNVP- THLDOLLVLL PGSRTVBDKKRH PLPRPSHKRH RASWKSN LOGOHOGGRV	ILPLPEVPPE BDD- DD- PF PFSGRIIMAR PSGRIIMAR RR.100PSPD IMP SQRDHSHSGR SQRDHSHSGR RRPOEIKETQ GQQGMMQQQQ QQQMMQQQQQ	VKOTODEKLV EESEYLPSAE EHATPOLRNE NKNRSSNR NKNRSSKOR KGMIKKIRMW HOGOONNOTO	EGGPHONOPO GRTFLAAAAX 10 10 10 10 10 10 10 10 10 10	EELPHKODOG OCHVSVBMOF AFCODNL AFCOSNL AFCENC CYVDDK SYHENYDSL NNRPK NNIOPTI SIA ORREV MPYINLOGRS	EEPSSSVQRE 92 90 10 10 10 10 10 10 10 10 10 1
 S. porte ARI 	ALETPPHPKW ELTTNPONPP GLKKHOEELO NRLPSLHIOP VTKNKRV- ISKNSRV-	DUSEVER CO SPARALAZIO SPARALAZIO SPARALAZIO SPARA TEPOLSETPL SPARA VILLOLOGIENT VILLOLOGIENT VILLOLOGIENT SPARA SP	XSRPLVNARL N N APHELOSOSN KRKSIEFOSQ QSQPQSPSKQ RSSONNRTND SISENWYGSR R.KHSP9ES	ARVOSVKRKS * * * * * * * * * * * * * * * * * * *	TTAVSGIPPK SSTTSDOGPN RFLRDGINSL SV P CPPKOPKINKR NSISNSWYR NSISNSWYR	TOLPLURING TOLPLURING INTILORAKSH KFKOHFNOL ELKOEYTRH GERENRLKOK KKSKKORO GRSSNN. NYMSY-O. KFYKTSK.	QVKAARGEVS TLPDVIPLSK VVRIKELEOM MARIS IKVLKKIGVL LDGOTONOGR	DDFELSSNOD DDFELSSNOD REASORHD OC- NDASKEFFE QCOGOOGRIC QRNGR- YISOTLOKR NKKRK	GISMCVNSSS GISMCVNSSS KKRRGLMMIL NEMEPNLLOY NEME-OPT GGHONOGORL Y 300 R 344 W 313	NRTAEEPIG DCA DCA DCA DCA DCA DCA DCA DCA	LINREDEKSKO RR-SSRVPHE FS-SSRVPHE PS-RSIVP- REVYDSPESPS PSSELPPV THLDOLLVLL POSRTYNDKM PLPRPSIKKR RASWKSN LOGOHOGORY	LLPLPEVPPE BD DO DO DO DO DO DO DO DO DO D	VKQTQDEKLV EESEYLFSAE EHATPOLRHE NKNSSNN NKHRSSIFHP NRNOSSKOR KKHSKINN HQQQDNHQYQ	EGGPNONGPO BE GRTFLAAAAK S S IEEHMANIS IEEHMANIS IEEHMANIS IIEEHMA	EELPHKODDG OCHVSVRMOF AFCGONL AFCGNL AFCSNL AFCSNL CYDDK SLA SLA SLA SLA SLA SLA SLA SLA SLA	EEPSSSVQRE 79 mi KOYGVVLVI V KOYGVVLVI V S LINSYORLDOK 111 Tall KSPFSAONVK 303 RGETLSLININ 28 RGETLSLININ 28 RGETLSL
 S. porte AWI H. sayes a ZCHCF Intelengender S00057p C. merkina AH2p S. porte AWI S. cervisia AH2p S. porte AWI H. sagne ZCHCF D. melanogader S00057p S. cervisia AH2p S. porte AWI S. cervisia AH2p <li< td=""><td>ALETPFHFKW EITTNFONPP GLKKHOEELO NRLPSLHIDP ISKNSRY SRL FFROFKT</td><td>DVSEVER CV SRPALLACY RPSWFANCY TEPOLSRTPL WKLLRBSOEV VKLLBBSOEV VKLLBRSOEV - OTHPLOP - OTHPLOP - OTHPLOP - OTHPLOP</td><td>XSRFLVNANL N N APHELOSOSN KRKSIEFOSQ GSOPOSPSKG RSSONNATNO SISEMYOSR R KHSPPE</td><td>ARVOSVARAS</td><td>TTAVSOIPEK SSTTSDOOPN RFLROGINSL </td><td>TOLPLVRTMG TOLPLVRTMG TOLPLVRTMG TOLPLVRTMG KFRGMTFNG ELKOEYTRH GEREKRLKOK KUSKKKOK GRSSNK KFTKTSK- GKKEREKVE</td><td>QVKAARGEVS TLPDVIPLSK VVRIKELEGM LKDKK</td><td>DDFELSSNVQ DDFELSSNVQ Introoperation REASOROHD DE DASNKEFFE GOGOGORIG GRN0-N-R- YSSTLOKR NRK-RK REASOROHD</td><td>GIFVVESIAE GISMCVNSSS KKEROLINII NEKEFNLLDY NE</td><td>NRTAREPICK</td><td>LNREDEKSKO RR-SSRVPHE PS-RSNVP- PS-RSNVP- PS-RSNVP- REVYDP3PVS PD55EELPDV THLDOLLVLL PG5RTYBDKK PLPRP5HKRH RASMKSN LOGOHOGGRV</td><td>ILPLPEVPPE BDD</td><td>VKOTODEKLV EESEYLPSAE EHATPOLRNE NKNRSSNN NKNRSSSKOR KEMKNKRIMW HOGOONNOTO</td><td>EGGPHONGPO GRTFLAAAAX 10 10 10 10 10 10 10 10 10 10</td><td>EELPHKODOG OCHVSVBMOF AFCODIL- AFCONL- AFCONL- AFCENC- CYVDDK- SYHENYDSL - NNRRX- NVIOPTI- NSIO- ORREV- MPYINLOGRS</td><td>EEPSSSVORE 92 90 10 10 10 10 10 10 10 10 10 1</td></li<>	ALETPFHFKW EITTNFONPP GLKKHOEELO NRLPSLHIDP ISKNSRY SRL FFROFKT	DVSEVER CV SRPALLACY RPSWFANCY TEPOLSRTPL WKLLRBSOEV VKLLBBSOEV VKLLBRSOEV - OTHPLOP - OTHPLOP - OTHPLOP - OTHPLOP	XSRFLVNANL N N APHELOSOSN KRKSIEFOSQ GSOPOSPSKG RSSONNATNO SISEMYOSR R KHSPPE	ARVOSVARAS	TTAVSOIPEK SSTTSDOOPN RFLROGINSL 	TOLPLVRTMG TOLPLVRTMG TOLPLVRTMG TOLPLVRTMG KFRGMTFNG ELKOEYTRH GEREKRLKOK KUSKKKOK GRSSNK KFTKTSK- GKKEREKVE	QVKAARGEVS TLPDVIPLSK VVRIKELEGM LKDKK	DDFELSSNVQ DDFELSSNVQ Introoperation REASOROHD DE DASNKEFFE GOGOGORIG GRN0-N-R- YSSTLOKR NRK-RK REASOROHD	GIFVVESIAE GISMCVNSSS KKEROLINII NEKEFNLLDY NE	NRTAREPICK	LNREDEKSKO RR-SSRVPHE PS-RSNVP- PS-RSNVP- PS-RSNVP- REVYDP3PVS PD55EELPDV THLDOLLVLL PG5RTYBDKK PLPRP5HKRH RASMKSN LOGOHOGGRV	ILPLPEVPPE BDD	VKOTODEKLV EESEYLPSAE EHATPOLRNE NKNRSSNN NKNRSSSKOR KEMKNKRIMW HOGOONNOTO	EGGPHONGPO GRTFLAAAAX 10 10 10 10 10 10 10 10 10 10	EELPHKODOG OCHVSVBMOF AFCODIL- AFCONL- AFCONL- AFCENC- CYVDDK- SYHENYDSL - NNRRX- NVIOPTI- NSIO- ORREV- MPYINLOGRS	EEPSSSVORE 92 90 10 10 10 10 10 10 10 10 10 1

В

Sequence alignment of hTRF4-2 with its closest homologs

		10		40						100		120		140		146	
H. saciens TRF4-1	M																8
H. sapiens TRF4-2	M									YRSOER	LLGS - HALPA	E	QRDF	LPLETTN		NNN :	ii.
D. melanogaster TRF4-1	MOPEVOWEOK	EQEOPSICTW	LKIWETNAGE	MOALLNOOLO	QATTINGSTS	SSSSSSNSON	NNNNNNN 11	NNTITNTTNN	TONNSSAKPY	LSRPYSSLNR	VLNF-RADSL	EIL0000000	OLNOTTORNS	TNINTTSOOS	TSSSADSTTN	RONNSPANSS	158
D. melanogaster TRF4-2	M																6
S. pombe CID14	MOKKSVSFNR	NNYKKRKNER	TEPLPARIEK	NOKPSKEKSK	RKEKOKNSDA	YDEMLLNNNF	TLLDGEEPMV	EIGSKKSRND	NDSEGIRDKG	OVEISNKNDP	TIOFOKADPL	EPLEKPDLPE	EATKROE	PTILLOI	PKREGRKTNP	VHDKAVENNS	154
S. cerevisiae TRF4p	MGAKSVTASS	SKKIKNRHNG	KVKKSKKIKK	VRKP	KSISLNDENE	VEILPSRNEQ	ETNKLPKDHV	TADG1				VLEHKSDODE	GEDVYDGHED	NPTDIPSTTE	ESKTPSLAVH	GDEKDLANND -	121
S. cerevisiae TRF5p	MTRLKAKYSP	TKG	KRKEDKHTKR	MRKSSFTRTO	KMLEVFNDNR	SHF NKY	ESLAIDVEDD	DTFGN				VLMENDKSDV	DIPVIE	EVTSSED	EQRAESSKRN	NSLED NO	105
5.57 (200) TO (200)		180		200		228		240		240		380		500	Constant and the second second	830	635
H. sapiets TRF4-1						*********							SPCP	EEAANRREVV	KRIETVVKDL	WPTADVOIPG :	15
H. sapiens TRF4-2	NHHOPGAWAR	RAGSSASSPP	SASSSPHPSA	AVPAADPADS	ASGSSN		KRK	RONKASGGRA	AGGGRADGGG	VYYSGTPWKR	RNY NOGVY	GLHEEIS	DEVEYMSPRP	EEEKNRMEVV	NRIESVIKEL	WPSADVOIPO	165
D. melanogaster TRF4-1	STNGPGAGTG	TSTGAGGTGT	NSPATTASST	AATTTGPATS	MSDTSNNPPQ	STTTPASETN	SITYNPSRKK	RPENKAGGAH	YYMNNHMEMI	AKYKGEPWRK	PDYPYGEGVI	GLHEELE	HEYQYVLPTP	CEHA RNEVV	KRIEAVVHSI	WPGAVVEIFG !	116
D. melanogaster TRF4-2				-CDEANPA						KPWQL	POVVYGNGIP	ALCLLHOEIE	QFYNYIRSTP	TEFCLRAGAV	RRIEDVVLSI	WPSASVDLFG	13
5. pombe CID14	DFIKFOWNSD	EDEDSVSNDK	\$KN		-NESLKKSSK	NEIPGFMROR	GREFHEANEK	SDSNRKRKRG	AVELDSOSC-	PWHR	QYKVERE	VSRIFHODIL	HEIDYITPTP	EEHAVRKTLV	SRINGAVLOK	WPDVSLYVFG :	286
S. corevisiae TRF4p	DFISLSASSE	DEQA EQE	EEREK	ELEIKKEKOK	EILNTD	Y	*********			PWIL	-NHDHSKOKE	ISOWLTFEIK	DEVATISES	EEIE RNOTI	STIREAVKOL	WPDADLHVFG :	224
S. cerevisive TRF5p	DELAFSOSSE	DETEGIKEDD	DERSSELLTD	EHEVSKLTSO	OSLNTESTCN	VEY				PWIR	-NHCHSKORR	IADWLTSEIK	DEVHYLSPSK	NEIKCRNRTI	DKLRRAVKEL	WSDADLHVFG !	221
		348		300		380		400		41		440		460		480	_
H. sapiens TRF4-1	SESTGLYLPT	SDIDLVVFGK	WERPP	LOLLEGALRK	HNVAEPCSIK	VLDKATVPII	KLTDQETEVK	VDISEN-MET	GVRAAEFIKN	YMKKYSLLPY	LILVEROFEL	ORDLNEVF -T	GOISSYSLIL	MAISFLOLHP	RIDA - RRAD -	ENLOMLLVEF	186
H. sapiens TRF4-2	SEKTOLYLPT	SDIDLVVFGK	-WE NLP	LWTLEEALRK	HKVADEDSVK	VLOKATVPII	KLTDSFTEVK	VDISEN-VON	GVRAADLIKD	FTKKYPVLPY	LVLVLKOFLL	GROLNEVF-T	GOIGSYSLEL	MAVSFLOLHP	REDA - CIPN -	INYGULLIEF	116
D. melanogaster TRF4-1	SFRTGLFLPT	SDIDLYVLOL	-WEKLP	LRTLEFELVS	RGIAEACTVR	VLOKASVPII	KLTDRETOVK	VDISEN-MOS	GVOSAELIKK	FRRDYPVLEK	LVLVLKOFLL	LRDLNEVF-T	GOISSYSLIL	MCISFLOMHP	R-GI-YHDT-	ANLOVLLLEF	466
D. melanogaster TRF4-2	SERTGLNLPD	SDIDLVVYYK	FWNPRL	LHELONELVS	OGVTOPDTVT	VLDKASVPVV	KETDLISBIR	FOVTENSVAS	GVQAADL I KD	FURHFPELPK	LVMVLKOFLS	LHGENEVYNS	GOVSSYALTL	MULSELOOHA	RSNR - RLSEH	KLALLLIOF !	228
S. pombe CID14	SPETKLYLPT	SOLOLVIISP	EHHYRGTKKD	MEVLAHHLKK	LKLASE VO	VITTANVPLL	KEVDPLTKVH	VOISPNO-PG	GLETCLVVNG	FMKKYPALRP	LVITIKHELN	MRALNEVF -L	GOLSSYATVC	LVVSFLOLHP	RUSTOSMREE	DNFGVLLLEF	442
S. cerevisiae TRF4p	SYSTOLYLPG	SDIDCVVTSE	LGG-KESRNN	LYSLASHLKK	KNLATE VE	VVAKARVPII	KEVEPHSGIH	IDVSFER-TN	GIEAAKLIRE	WUDDTPOLRE	LVLIVKOFLH	ARRLNNVH-T	GGLGGFSIIC	LVESELHMHP	RIITNEIDPK	NLGVLLIEF !	179
S. corevisiae TRF5p	SFATDLYLPG	SDIDCVVNSR	NRD-KEDRNY	IVELARHLKN	KGLAIR ME	VIVKTRVPII	KFIEPOSOLH	IDVSPER - TN	GLEAAKLIRE	WLRDSPGLRE	LVLIIKOFLH	SRRLNNVH-T	GOLGOFTVIC	LVYSFLNMHP	RIKSNDIDVL	DNLOVLLIDE	170
		500		124	1	040		560		540	and the second se	400		620	100000000000000000000000000000000000000	640	
H. saciens TRF4-1	FELYGRNENY	LETGIRIKEG	GAYLAKEEIM	KANTSGYRPS	MUCIEDPLUP	GNDVGRSSYG	AMOVKOVEDY	AVIVLSH	AVSPLARSY	PNRDAE ST	LORIIKYTOR	VIDYREWIKE	KW05K	AHPS	POMOSRIKIK	ERIATCNG	127
H saciens TRF4-2	FELYGRHENY	LKTOIRIKDO	OSYVAKDEVO.	KNMLDQYRPS	MLYIEDPLOP	QNDVGRSSYQ.	AMOVKOAFDY	AYVVLSH	AVSPLAKYY	PNNETE SI	LORIIRVIDE	VATYROWISK	QWQL		P	SCNG (444
D melanonasiar TRF4.1	FELYGRRENY	MELGISIKNO	GRYMPKDELO	ROMVOGHERS	LICIEDRITE	ONDIGESSYG	VEOVODAEKC	AVEVLAL	AVEPINILIA	IDPRVN.	LORITHITOD	VIDYREWIRE	NEEHL VYVDR		P	TAYATANGAR	670
D melanotaster TRF4-2	LOYYGRKEDE	FKYGISVLGO	GGCVEKARLR	STLOENNWOS	VICIEDRATE	TNDIGRSSYG	VLOVMOGEGA	AFVELSK	+1.40 50	SSKIVG- PI	LANIVEVPOS	LINYRAWYHY	NFOHLLTPEL	PCADSL VOPS	P		100
S pombe CID14	LELYCKOFYY	DAVGLAVHNG	GEVESKKKM-	- GWL KPNOPY	LISIODEVDE	ONDVSKSSRG	LLRYKATEAN	GEDLLTSKLY	ALASRIEREG	VNRVKDEPSI	LETILSYDEG	VROHREHMLK	CYKN	NPV	PL		58.9
S. cerevisiae TRE4n	FELYGENEGY	DOVALGSSDG	YPVYFPKSTW	SALOPIKNEE	SLAIODPODE	SNNISPOSEN	IRDIKKAFAG	AFOLLTNRCF	FLHSATEKOR	LGK	LONVIKYRGK	ARDEKD					450
5. cerevisiae TRF5p	FELYGENEGY	DOVALSISDO	YASYIPKSCW	RTLEPSRSKP	SLAIODPODP	NNNISRGSEN	MKDIKKAFAG	AFELLVNKCW	ELNSATEKOR	VGK51	LONVIKYRGO	KRDFND					487
	Contraction of the second	560		680	and the second second second	700		729	1	741	1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.	784		780		800	1
H appints TRF4-1				!	mont	- marine the			NREPESPYOO	RLTL		SLSSPOLL	SSGSSASSVS	SLSGSDVDSD	TPPCTTPSVY	OFSLOAPAPL :	103
H. saciena TRF4-2													···· \$55564T	OSSSSDVDSD	ATPCKTPK		470
D. melanopaster TRF4-1	KYVIMPSGAV	VOOLYHHPVO	VETAHOHSHA	HSHSHGHAHP	GAHLCOPYVT	GTTVSAVTTT	TTMAYVTYGY	SAGGVODOOD	OQNATAHTHS	0001		ONOSOSRH	RRGSTSSGDD	SEDSKOGDVV	ETTSSAGEVY	DIALSTPAGE	161
D melanonaster TRF4-2													TOSTSPEAS	ASASED			174
S. nombe CID14														ALASIDVDK			LAS.
S. completion TREAD		F	BOL VINKALL	ENENYHKKRS	RUIHDEDEAF	DTVTSTATAT	TTODDYFLTN	PPAKKAKIE.	FKPESE	PAsses		KRNSG	FTYLTYS	SEDDOF			176
S. comvision TRESp.		F	ROLVONKALL	ENERYHKRRT	RIVOFOLFIN	OTED	PVEELYKLDE	PAKKKOKAKK	DEREGELEES	AIPSAPPDEG	VSRSKLKRKV	KKTDOGSLLH	ONNESIDDEM		0		404
a carena an ap		420	APETGARATI	840	Riffeberin	100	FILLINGE	100	CAREGETARD	000	- and - const	800	anne ar oben	940		960	1014
H anniene TREA 1	MAGI PTAL PM	P. SOKPOPTT	SOTI INTTIN	OTRETIRENT	LAVABURCED	AGVE			TASI KAVNUM	SEPAIDSASP		KOW	NONEL SHERE	HANTOGOOVS	SVASAAVDED	VONDOWN	617
H satisfiers TRE4-2	OLLCP.	P. 5	TON	P	Vason	VELE				OSTOTTATEN	STNKS	OH	GSARLERSS	KOF-OGTTOT	SHOSLMTNKO	HOOKSNN	10.1
D melanopaster TREA1	AN. MEMON	BUHAUGHBAS	NEWSONONON		RELANIANDE	MOREL EDOOD		NOFLOBODVO	TARNBOODO	ROODHYNORN	NUNSEAV. YU	OOYYVBBBBB		NYHOOHHUSH	SHOWHENDOO	нинооонино	10.7
D. melanopaster TRE4.2		FINATORFAS	namaununun	0433331033	FEIAMIAAGE	MOPECEOGGG	duanac 1 ada	NOT THE OUT O	Toankodaba	JUGANTAGAA			Garakanaaa	FRSGGRA	TIGEGROODE	PONID	100
E nombe CID14					18100	VOLOVVEDES	DEDETDAAKO						NO	HENEOKOALT	STOFOSSENS	DAN	140
E complete TRE40						TOLOTTEDES	DUDEIDAAAD	DEFRICE						nenronone i	DOVNEY		
5. cerevisiae TREEs														DOKORDT	RECODENER	ETH	102
a ceremene (19/sp		963		1.00		1.00	Sector Sector	1.04	1								-
M services TOTA 1	OVNETOW	PRESENTATED		Thrower of	and more second	T. Street Street	Harransee	1.	and the second	0.642							
H services TREA 3	OV	HAVYDYNYDD	ADIEDIC							B 573							
D malandonalas Tref4-2	ORROWL BY	CNRYOKELCO	SPILSAGE	NECENCENCE	REEGENNES	BRIDGTI				PDER 1001							
D melanogaster (RF4-1	WAP GHER VOG	UNRTURSE 00		nessnusriss	a a a a a a a a a a a a a a a a a a a	FFLRUILVNS		arss145555	**************	NUER 1001							
E combé CIDM			DEVIENCE	CODEF.				EVELIEBTTA	SPEESDANNI	DYDE and							
a, pompe CID14				SEUSE				ETFILEUITA	auceannini	TI 684							
a cerevisiae (10-4p								TUDA		0041 540							
p. cerevisiae THP5p								TVDA	WIRRUTWLSR.	0 MAL 642							

С

Sequence alignment of hMTR4 (SKIV2L2) with its closest homologs.

		20		40		60		80		100		120	e
H. sapiens hMTR4	MADAFGDELF	SVF	EG		DSTTAAGT	KKOKEKOKGK	WKGPPGSADK	A GKRFD -	GKLQSESTNN	GKNKR	DV		66
D. melanogaster L(2)35DF	M-DIEELF	DCFDEVTPSV	PPPALNQEEK		PSGSSASK	KGNKRQADGT	TASKEATSSK	G AEGDDK	GDVTEEPTKR	LKQEASTVAV	00		84
S. cerevisiae MTR4p	M DSTDLF	DVFE-EAPVE	LP		-TDSN	-GEKNADTNV	GDTPDHTODK	KHGLEEEKEE	HE ENNSEN	KKIKSNKSKT	ED		71
S. pombe MTR4	MEGGELDDAF	GVFEGKVPKS	LKEESKNSON	SONSOKIKRT	LTDKNASNQE	QGTKKLESSV	GEGESATKRA	K IENLKON	ODLIPNNDVN	G-IHINNSAV	ADTKHKPKIG	DIAADDISNE	117
		140		190		160		200		220		240	+
H. sapiens hMTR4	DFEGTDE	PIFGKKPRIE	ESITEDLALA	DLMPRVKVQS	VETV	EGCTHEVALP	AEEDYLPLKP	RVG KAAKE	YPEILDAFOR	EALOCYDNNO	SVLVSAHTSA	GKTVCAEYAI	175
D. melanogaster L(2)35DF	DNNADDE	PT RTLDID	DSAT LE	ALRTRIVTHL	LDAP	KSCTHEVAAH	PDGEYIPLKP	FSG VPAKE	YPEVLOPFOR	QAILCIDNSQ	SVLVSAHTSA	GKTVVAEYAI	187
S. cerevisiae MTR4p	KNKKVVV	P VLADSFE	QEASREVDAS	KGLTNSETLQ	VEODGKVRLS	HOVRHOVALP	PNYDYTPIAE	HKR - VNEART	YPETLOPFOD	TAISCIDRGE	SVLVSAHTSA	GKTVVAEYAI	185
S. pombe MTR4	VSIKNEGDTI	PEATVADSFE	QEASLOVAGK	VGMTEAKSST	EEV VELR	HOVRHOVSIP	PNYDYVPISK	HKSPIPPART	YPETLOPFOA	VSIACIERGE	SVLVSAHTSA	GKTVVAEYAV	234
	CALIFI CANADO OF DEDAMA	260		250		300		320		340		360	-
H. sapiens hMTR4	ALALREKORV	IFTSPIKALS	NOKYREMYEE	FODVGLMTGD	VTINPTASCL	VMTTEILRSM	LYRGSEVMRE	VAWVIFDEIH	YMRDSERGVV	WEETIILLPD	NVHYVFLSAT	IPNAROFAEW	295
D. melanogastor L(2)35DF	AKSLAAKORV	IYTTPIKALS	NOKFREFTDE	FKDVGLVTGD	VTINPSASCL	IMTTEILRNM	LYRGSEIMRE	VGWVVFDEIH	YMRDKERGVV	WEETLILLPD	NVRYVFLSAT	IPNAROFAEW	307
S. cerevisiae MTR4p	AQSLKNKORV	IYTSPIKALS	NOKYRELLAE	FGDVGLMTGD	ITINPDAGCL	VMTTEILRSM	LYRGSEVMRE	VAWVIFDEVH	YMRDKERGVV	WEETIILLPD	KVRYVFLSAT	IPNAMEFAEW	305
S. pombe MTR4	AGSLRDKORV	IYTSPIKALS	NOKYRELLAE	FGDVGLMTGD	VTINPDATCL	VMTTEILRSM	LYRGSEVMRE	VAWVIFDEIH	YMRDKERGVV	WEETIILLPD	KSHFVFLSAT	IPNAMOFAEW	354
		380		#30 1		+420		840 1		462 1		480	1
H. sapiens hMTR4	ICHLHKOPCH	VIYTDYRPTP	LOHYIFPAGG	DGLHLVVDEN	GDFREDNFNT	AMQVLRD	AG	DLAKGDOKGR	KGGTKGPS	NVFKIVKM	IMERNFORVI	IFSFSKKDCE	400
D. melanogaster L(2)35DF	VCHLHKOPCH	VVYTDYRPTP	LOHYIFPAGG	DGIHLIVDEK	GQFKEDNFTT	AMAVLAN	AG	EAGKGDOKGR	HGGIKGTNAG	QTNIFKIVKM	IMERNFAPVI	IFSFSKKDCE	416
S. cerevisiae MTR4p	ICKIHSOPCH	IVYTNFRPTP	LOHYLFPAHG	DGIYLVVDEK	STFREENFOK	AMASISNOIG	DDPNSTDSRG	KKG-QTYK	GGSAKGDAKG	DIYKIVKM	IWKKKYNPVI	VFSFSKRDCE	420
S. pombe MTR4	TTKIHROPCH	VVYTDERPTP	LOHYLFPSGS	DGIHLVVDEK	SNFREENFOR	AMSALMEROG	DDPAAMATKG	NAKKG - KTGK	GG . VKGPS	DIYKIVKM	IMVKNYNPVI	VFSFSKRECE	468
		The second se		1.00000000			BULKET LEU	FOR OLIVIE			ABWERGHERE		
H. sapiens hMTH4	ATALOMIKLD	FNTDEEKKMV	EEVFSNATDC	LSDEDRKLPQ	VEHVLPLLKR	GIGIHHGGLL	PILKETTETL	FSEGLIKALF	ATETFAMGIN	MPARTVLFTN	ARKFDGKDFR	WISSGEYIGM	520
D. meranogaster L(2)35DF	FLAL KMEKLD	FNIPDEKKLY	TRIENNAMOV	LIEEDRRLPO	VENVLPLLRR	GIGIHHGGLL	PILKETTETL	FOEGLIKALF	ATETERICIN	MPARIVLFIA	VERIDGARER	WISSGETTOM	530
S. Oerevisiae MTR4p	ALALOWSKLD	PRODUCEREAL	TTITUTUT	LAEKORELPO	INHILPELKK	GIGINHSGLL	PILKEVIEIL	FOEGELKVLF	ATETFOIGLE	MPARTYVETS	VRKWDGGGGFR	WUSGGETTOM	240
o, pombe MTR4	ALALUMBRED	620	TTTPNNAVNG	LOEKUKELFU 640	TENTLFLERK	GIGINHSULL (60)	PILKEVIEIL	FUEGLERVEF	ALEIPSIGEN	TOD TOD	VRAFDGRIFR	720	200
H sapiens hMTR4	SCRAGRRGMD	DEGLATIMAN	EKMSPTICKO	LIKGSADPIN	SAEHI TYNMY	INLIBUEEIN	PEYMI EKSEY	OFOHYRALPG	VVERVENSEE	OVNKIVIPNE	ESVVIVVKIR	OOLAKI GKEL	640
D. melacogaster 1/2/35DE	AGRAGERGID	DEGIVILMID	EKVSPAVGRD	LYOCKADPIN	SAFHI TYNMY	LNLLRVEEIN	PEYMLERSEY	OFONOAAL PG	LHDOVEOKTL	FUNKLTIKDE	HNIASYHHIR	SOLDOHGKOF	856
S. cerevisiae MTR4o	SGRAGRRGLD	DRGIVIMMID	EKMEPOVAKG	MVKGOADBLD	SAFHLGYNMI	LNLMRVEGIS	PEFMLEHSEE	OFONVISVEV	MEKKLAELKK	DEDGLEVEDE	ENVKEYYELE	QALKGYREDV	660
S. pombe MTR4	SGRAGRRGLD	DRGIVILMID	EKMOPPVAKS	MLKGEADRLD	SAFHLSYNMI	LNLLRVEGIS	PEFMLERCFF	OFONSLEVPK	LEAKLEESOO	HYDSETILDE	RPLEEYHTLK	TOLERYRTDY	708
0407.00000000000	and the second second second second	740		760		760		800		820		640	1
H. sapiens hMTR4	EEYIHKPKYC	LPFLOPGRLV	KVK - NEGDDF	GWGVVVNFSK	KSNVKPNSGE	LDP LYVVE	VLLRCSKESL	KNSAT	EAAKPAK	PDEKGEMQVV	PVLVHLLSAI	SSVRLYIPKD	749
D. melanogaster L(2)35DF	ROWITKPOYL	LPFLOPGRLV	KVA-AGSOEY	DWGIVLNEKK	ODOSRKNPLK	AEP SVTID	VLLHVSEAAA	KTGDT	E PCK	PNERGCMEVV	PVAHTLITOI	SSIRVYFPND	762
S. cerevisiae MTR4p	ROVVTHPANA	LSFLOPGRLV	EISVNGKDNY	GWGAVVDFAK	RINKRNPSAV	YTDHESYIVN	VVVNTMYIDS	PVNLLKPFNP	AFPEGIRPAE	EGEKSICAVI	PITLOSIKSI	GNLRLYMPKD	780
S. pombe MTR4	RTVVNHPNFC	LSFLOGGRLV	RVKV - GNEDF	DWGVVVNVSK	RPLPKGOSNE	YLPOESYIVH	TLV MVAS	DTGPLRIRSG	HLPEVHPPAA	E-DKGKFEVV	PFLLSSLDGI	AHIRVELPND	823
		860		880		900		820		940		6960	1 Carlos
H. sapiens hMTR4	LRPVDNRQSV	LKSIGEVOKR	FPDGIPLLDP	I DDMG I QDQG	LKKVIQKVEA	FEHRMYSHPL	HNDPNLETVY	TLCEKKAQIA	IDIKSAKREL	KKARTVLOMD	ELKCRKRVLR	RLGFATSSDV	869
D. melanogaster L(2)35DF	LRSADNRRAV	LKTIGEAKKR	FPLGPPVLNP	IDDMNIKDRE	FRDIVNTISO	FEKRLEEHPL	HKSPELERIH	RRYODKVTLO	KOLODLKAEL	KAARSLLOMD	ELKHRKRVLR	RMGYCKPGDV	882
S. cerevisiae MTR4p	IRASGOKETV	GKSLREVNRR	FPDGIPVLDP	VKNMKIEDGD	FLKLMKKIDV	LNTKLSSNPL	TNSMRLDELY	GKYSKKHDLN	EDMKOLKRKI	SESQAVIQLD	DLRRRKRVLR	REGECTPNDI	900
S. pombe MTR4	LKSQGQKLTV	GKALSEVKRR	FPEGITLLDP	VENMNIKEPT	FIKLMKKVNI	LESRLLSNPL	HNFSELEEKY	AEYLRKLALL	EEVKDLKKKL	SKARSIMOLD	ELNSRKRVLR	RLGFTTSDDV	943
		960		1.000		1.020		1.04	1	1.060		1.060	8
H. sapiens hMTR4	IEMKGRVACE	ISSADELLLT	EMMFNGLFND	LSAEQATALL	SCEVEQENSS	- EMPKLTEOL	AGPLROMOEC	AKRIAKVSAE	AKLEIDEETY	LSSFKPHLMD	VVYTWATGAT	FAHICKMTDV	888
D. melanogaster L(2)35DF	IEFKGRVACE	LSSADELLMT	EMIFNGVFND	LTAPOAVALL	SCEVCDEKSS	- ESVKSATEL	SGPLRSMODL	ARRIAKVSTE	CKLDLDADTY	VOKEKPELMO	VVLAWCKGSS	FLAVCKMTDI	1001
S. cerevisiae MTR4p	IELKGRVACE	ISSGDELLLT	ELIFNGNFNE	LKPEQAAALL	SCFAFGERCK	-EAPRLKPEL	AEPLKAMREI	AAKIAKIMKD	SKIEVVEKDY	VESFRHELME	VVYEWCRGAT	FTQICKMTDV	1019
S. pombe MTR4	IEVKGRVACE	ISSGDGLLLT	ELIFNGMEND	LTPEOCAALL	SCLVFQEKSE	VENORMKEEL	AGPLKILGEM	ARRIAKVSKE	SKQELNEEEY	VNSFKPSLME	VVYAWAHGAS	FAQICKMTDV	1063
M anniana hARTON		BI FELLBONG	OAAKA LONTE	TENERECIT	KINDOUVEAA	AL VI. 1013							
D. malappagater I (200505	FEGRILIRCHA	RIFELLROMC	OARKTICHTE	LENKEREGIE	LINDIVEAA	51 VI 1055							
S convision MTR4n	VEGSI I PMEK	PLEELINGMC	DVANTIONES	LENAFSEGIR	LINPOLVSAG	SL TL 1000							
o. cerevisiae in much	I LOOL I RMPK	RECEVICEV	DI NUL CHITA	LALAMEAVLA	LINNUTVOAG	0212 10/3							

D

ZCCHC7-FLAG IP



Supplemental Tables 1-14 Click here to download Supplemental Movies and Spreadsheets: Tables S1-S14 final.pdf