2354–2369 Nucleic Acids Research, 2013, Vol. 41, No. 4 Published online 28 December 2012
doi:10.1093/nar/gks1347

Alternative translation initiation augments the human mitochondrial proteome

Lawrence Kazak*, Aurelio Reyes*, Anna L. Duncan, Joanna Rorbach, Stuart R. Wood, Gloria Brea-Calvo, Payam A. Gammage, Alan J. Robinson, Michal Minczuk and Ian J. Holt*

MRC-Mitochondrial Biology Unit, Wellcome Trust-MRC Building, Cambridge CB2 0XY, UK

Received September 3, 2012; Revised December 2, 2012; Accepted December 3, 2012

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

ABSTRACT

Alternative translation initiation (ATI) is a mechanism of producing multiple proteins from a single transcript, which in some cases regulates trafficking of proteins to different cellular compartments, including mitochondria. Application of a genome-wide computational screen predicts a cryptic mitochondrial targeting signal for 126 proteins in mouse and man that is revealed when an AUG codon located downstream from the canonical initiator methionine codon is used as a translation start site, which we term downstream ATI (dATI). Experimental evidence in support of dATI is provided by immunoblotting of endogenous truncated proteins enriched in mitochondrial cell fractions or of co-localization with mitochondria using immunocytochemistry. More detailed cellular localization studies establish mitochondrial targeting of a member of the cytosolic poly(A) binding protein family, PABPC5, and of the RNA/DNA helicase PIF1α. The mitochondrial isoform of PABPC5 co-immunoprecipitates with the mitochondrial poly(A) polymerase, and is markedly reduced in abundance when mitochondrial DNA and RNA are depleted, suggesting it plays a role in RNA metabolism in the organelle. Like PABPC5 and PIF1α, most of the candidates identified by the screen are not currently annotated as mitochondrial proteins, and so dATI expands the human mitochondrial proteome.

INTRODUCTION

Products of nuclear genes dominate the mitochondrial proteome. They are synthesized by cytosolic ribosomes and imported into mitochondria via specific pathways according to their final destination in the organelle (1). The most extensively used system for importing matrix-destined mitochondrial proteins depends on a positively charged amphipathic α helix, located at the amino (N-) terminus of the protein, which functions as a mitochondrial targeting signal (MTS). Cytosolic proteins chaperone mitochondrial precursors to an import complex located on the outer surface of the mitochondrion, termed the translocase of the outer membrane (TOM) complex. The MTS can interact with import receptors and direct proteins across both the outer and inner mitochondrial membranes. Matrix-destined proteins depend additionally on the translocase of the inner membrane (TIM) complex, specifically TIM23, to direct them to the innermost compartment of the organelle. The insertion of proteins into the TIM23 channel requires a membrane potential across the inner mitochondrial membrane, and a further driving force is provided by the presequence translocase-associated motor complex. Upon entry to the matrix, many proteins have the MTS removed by the mitochondrial processing peptidase, and chaperones facilitate the proper folding of the mature protein into its active conformation (1,2). Although many genes encode dedicated mitochondrial proteins, an increasing number are recognised to specify multiple protein isoforms that are found in more than one cellular compartment. Protein variants that are targeted to different cellular compartments can be synthesized from a single gene, or transcript, via the use of alternative splice sites, transcription start sites or translation initiation sites (3). Alternative translation initiation (ATI), first discovered in viruses (4,5), and subsequently in eukaryotes (6), is a mechanism by which more than one initiation codon within a single mRNA results in the translation of proteins with distinct N-termini (3,7). ATI diversifies the proteome and may alter a protein's function or cellular location.

The use of an MTS lends itself to ATI, as essentially the same mature protein can be made for two compartments from one gene. RNase H1 is typical of this class of genes (8). Other documented examples of ATI-dependent

*To whom correspondence should be addressed. Tel: +44 1223 252840; Fax: +44 1223 252845; Email: holt@mrc-mbu.cam.ac.uk
Correspondence may also be addressed to Lawrence Kazak. Tel: +44 1223 252840; Fax: +44 1223 252845; Email: lawrence.kazak@gmail.com
Correspondence may also be addressed to Aurelio Reyes. Tel: +44 1223 252840; Fax: +44 1223 252845; Email: art@mrc-mbu.cam.ac.uk

© The Author(s) 2012. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

dual targeting include iron–sulfur cluster assembly enzyme (NFS1) and insulin-degrading enzyme (9,10). Translation initiation from the second, or a subsequent, AUG codon, which we term downstream ATI (dATI), is a less obvious method of achieving mitochondrial targeting, as the mature mitochondrial protein necessarily lacks a portion of the N-terminus that is present when initiation occurs from the first AUG codon. The thyroid hormone receptor, c-Erb A α1, was thought to be a rare case of mitochondrial targeting via dATI (11).

Our studies of nucleic acid-transacting proteins in mitochondria led us to the finding that dATI yields a mitochondrial isoform of flap endonuclease 1, FEN1 (manuscript in preparation). Taken together with the previous instance of dATI-mediated mitochondrial targeting of the thyroid hormone receptor, the possibility arose that this might be a commonplace mechanism of mitochondrial targeting. Therefore, using a computational approach, an inventory comprising 126 genes encoding candidate dATI-dependent mitochondrial isoforms was assembled. Experimental validation of a subset of the putative genes from the list provided empirical evidence of mitochondrial localization, indicating that mitochondrial targeting via dATI is much more widespread than recognized hitherto.

MATERIALS AND METHODS

Cell culture and transfections

Human 143B osteosarcoma (HOS) cells were maintained in DMEM supplemented with 0.1% penicillin/streptomycin and 10% FBS. Flp-In™ T-REx™ 293 (HEK293T) cells (Invitrogen) were cultured in DMEM, 0.1% penicillin/streptomycin, 10% tetracycline-free FBS, 15 μg/ml Blasticidin (InvivoGen) and 100 μg/ml Zeocin (Invivogen). HEK293T cells were co-transfected, using Lipofectamine 2000 (Invitrogen), with 1350 ng pOG44 plasmid (invitrogen) and 150 ng of cDNA ligated into pcDNA5/FRT/TO (Invitrogen). Stable transfectants were selected with 100 μg/ml Hygromycin B (Invivogen).

Construct design and site-directed mutagenesis

Pif1, Pabpc5 and Pop1 cDNAs were ligated as KpnI-XhoI fragments into the pcDNA5/FRT/TO MCS. Mutant constructs were made using the QuikChange site-directed mutagenesis kit (Stratagene).

Confocal microscopy

HOS cells were grown on glass coverslips and transiently transfected using Lipofectamine 2000 with 1.5 μg of Pif1, Pabpc5 and Pop1 cDNAs. Twenty-four hours after transfection, HOS cells were incubated with anti-HA (1:200; Roche Diagnostics) or anti-FLAG (1:400; Sigma) antibodies, followed by treatment with appropriate fluorescently labelled secondary antibodies. Coverslips were mounted with 1,4-diazabicyclo[2.2.2]octane (Sigma), containing 4',6-diamidino-2-phenylindole, dihydrochloride (DAPI). Images were acquired with an LSM 510 META confocal microscope (Carl Zeiss, Jena, Germany).

Immunoblotting

Proteins were resolved by 4–12% NuPAGE Bis–Tris SDS-PAGE (Invitrogen) and transferred to nitrocellulose membrane (Whatman), which were incubated at 4°C overnight with primary antibodies. The concentrations of primary antibodies from Abcam were as follows: GAPDH (1:20 000), HSP60 (1:60 000), SF2 (1:3000), BMS1L (1:200), B23 (1:1000), PABPC1 (1:1000), FIBRILLARIN (1:2000), ARAP1 (1:500), MTSS1L (1:1000), FBXL12 (1:2000), LEPRE1 (1:1000), PABPC5 (1:500), NOX3 (1:1000), TNIK (1:2000), MBRL (1:1000), FOXH1, (1:1000), PRPSAP2 (1:1000), CLASP2 (1:2000), GCN1L1 (1:2000), Histone H2A (1:1000), mtPAP (1:1000), MRPS18 (1:1000), COXIV (1:1000), Aconitase II (1:1000), CYT C (1:5000) and NRF1 (1:500). The concentrations of primary antibodies from Santa Cruz were as follows: PIF1 (1:200), POLG1 (1:500), PCNA (1:400), LRPPRC (1:2000), Utrophin (1:200), TOM20 (1:4000) and SSBP1 (1:500). The concentrations of primary antibodies from Sigma were as follows: FLAG (1:1000) and Tubulin (1:2000). Other antibodies used were as follows: NDUFB8 (1:1000; Invitrogen), HA (1:2000; Roche), TWINKLE (1:200; A. Suomalainen), TFAM (1:80 000; R. Wiesner) and TRIM32 (1:1000; D. Blake). Secondary antibodies were anti-rabbit or anti-mouse HRP (Promega). Membranes were visualized with enhanced chemiluminescence (ECL) plus immunoblotting detection system (GE Healthcare).

Mitochondrial import

[^{35}S]-methionine-labelled proteins were generated with the TNT® Quick Coupled Transcription/Translation System (Promega) according to the manufacturer's instructions. Labelled proteins were incubated with rat liver mitochondria prepared by differential centrifugation and import was assessed via trypsin protection and FCCP-dependent inhibition (12). Import reactions were carried out at 37°C for 1 h and then subjected to SDS-PAGE. Gels were dried, exposed to storage phosphor screens (GE Healthcare), visualized on the Typhoon 9410 Variable Mode Imager (Amersham Biosciences) and quantified using ImageQuant 5.2.

Nuclear, cytosolic and mitochondrial isolation from cultured cells

HEK293T cells were homogenized in hypotonic buffer (20 mM HEPES–NaOH [pH 7.8], 5 mM KCl, 1.5 mM $MgCl_2$, 2 mM DTT, 1 mg/ml BSA, 1 mM PMSF, protease inhibitor cocktail [Roche]). Low speed centrifugation of HEK293T cells resulted in a pellet that was used to isolate intact nuclei as previously described (13). Cytosolic extracts were obtained from post-mitochondrial supernatants, and mitochondria were prepared as described previously (14).

Submitochondrial fractionation

Freshly isolated mitochondria were resuspended in 1 ml of 20 mM potassium phosphate buffer, pH 7.4, with 150 mM KCl, and sonicated 3×10 s at 70 W (Soniprep 150).

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

Samples were centrifuged at 100 000*g* for 1 h at 4°C. The supernatants were retained as the matrix fractions, whereas pellets comprised mitochondrial membrane proteins.

Immunoprecipitation

Mitochondria were isolated from PABPC5^{33}.HA-, PABPC5^{33}.F-, or PDE12.F-overexpressing HEK293T cells and treated with proteinase K (PK) (0.02 mg/5 mg mitochondria) for 30 min on ice. PK was inactivated with PMSF and then mitochondria were lysed in lysis buffer (50 mM Tris–HCl, [pH 7.4]; 150 mM NaCl, 1 mM EDTA, 1% triton X-100). Debris was removed by centrifugation at 8000*g* max, and supernatant was incubated with EZview Red Anti-HA affinity gel (Sigma) or anti-FLAG M2 affinity gel (Sigma) for 1 h at 4°C. Beads were washed three times in wash buffer (50 mM Tris–HCl, pH 7.4; 150 mM NaCl). In the case of HA immunoprecipitation, beads were boiled, centrifuged and supernatant was used for immunoblotting. In the case of FLAG co-immunoprecipitation experiments, proteins were eluted from beads using 3X FLAG peptide (Sigma).

Iodixanol density gradients

Purified mitochondria were treated with 100 μg/ml trypsin for 30 min at room temperature, washed, lysed with 0.4% dodecyl maltoside (DDM) and loaded on to a 20–42.5% iodixanol density gradient (Sigma) and ultracentrifuged at 100 000*g* for 14 h at 4°C.

Mitochondrial DNA analysis

Mitochondrial DNA was extracted and analysed via Southern blotting as previously described (14).

In silico analysis for generating the dATI inventory

Protein sequences from the human and mouse genomes were downloaded from the ensembl genome browser, release 61 (15). Human–mouse orthology data were obtained from the ensembl genome browser via BioMart (www.ensembl.org/biomart.martview). For all protein sequences, each methionine occurring in the first quarter of the sequence was queried for possible dATI, with mitochondrial targeting predicted using Mitoprot (16), TargetP (17) and iPSORT (18). Genes were considered mitochondrial when they encoded at least one protein, which, when encoded from an alternative translation site, satisfied two of the following three criteria: (i) Mitoprot score >0.95; (ii) TargetP score >0.95 and (iii) positive iPSORT result. Retaining only those genes considered mitochondrial for both their human and mouse orthologs further filtered the inventory. Finally, the genes whose protein products were predicted to localize to mitochondria if their translation would start at M1 were removed from the inventory. Genes whose protein products that started at M1 were considered mitochondrial when they, satisfied two of the following three criteria: (i) Mitoprot score >0.5; (ii) TargetP score >0.5 and (iii) positive iPSORT result.

RESULTS

A genome-wide screen identifies 126 candidate dATI-dependent mitochondrial proteins

As a test of the idea that many proteins achieve mitochondrial targeting via downstream alternative translation initiation (dATI), every in-frame AUG codon within the first quarter of all annotated human and mouse protein-coding genes of the ensembl database was assessed as a potential translation initiation site. The choice to restrict the analysis to the first quarter of the longest annotated protein isoform was made to reduce the probability of identifying gene products that were truncated at the N-terminus to such an extent that functional domains within the open reading frame (ORF) were lost. Putative mitochondrial targeting was evaluated based on predictions from Mitoprot (16), TargetP (17) and iPSORT (18). Two of three conditions were required: a Mitoprot score >0.95, a TargetP score >0.95 and a positive iPSORT result. These criteria were established with the aim of minimizing false positives; however, this risked failing to identify genuine mitochondrial proteins that depend on dATI for targeting to the organelle.

When the computational analysis was applied, 886 entries (473 and 413, human and mouse genes, respectively) scored above the threshold set (Figure 1A). Human and mouse data sets were combined and only orthologous genes were retained, yielding 168 candidates. Forty-two of these were excluded (Supplementary Table S1) because the full-length protein has a plausible N-terminal MTS (an MTS starting at M1), and approximately half of these are found in the Mitocarta database (19). This suggests that these proteins can achieve mitochondrial targeting without the need for dATI. Thus, the final catalogue (Supplementary Table S2) comprised 126 genes that potentially depend on dATI to generate a mitochondrially targeted protein.

Antibody- and immunocytochemistry-based identification of dATI candidates

Of the 126 genes identified by the *in silico* screen (Supplementary Table S2), five are recognized mitochondrial proteins in the Mitocarta database (19), and 17 others can be found in the MitoMiner database (20). Thus, the computational screen identified >100 proteins that have had no prior evidence of mitochondrial localization. This might be taken to imply that the screen generated many false positives. However it is noteworthy that the established dATI-dependent mitochondrial protein, c-Erb A α1, is missing from the Mitocarta and Mitominer databases. As an initial test of the validity of the dATI screen, the cellular distribution of 26 proteins with no prior localization to mitochondria was analysed. Fifteen proteins predicted by the *in silico* screen to have a cryptic MTS, and 11 negative controls, were studied in enriched nuclear, cytosolic and mitochondrial preparations, by immunoblotting or by immunocytochemistry. The proteins were selected without bias, while aiming to reflect a range of biological processes. For immunoblotting analyses, the prediction, *a priori*, was that the

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

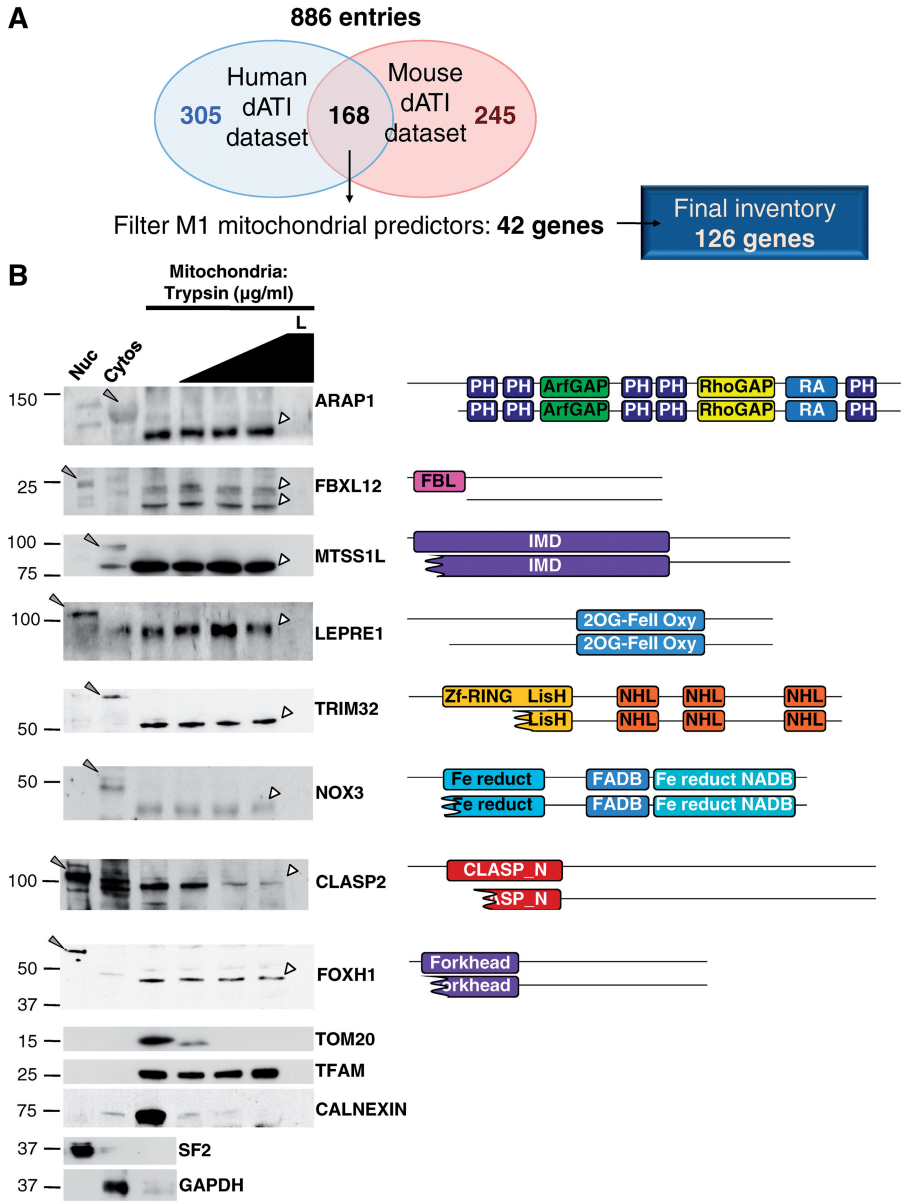

Figure 1. Output of a computational screen for genes encoding dATI-dependent mitochondrial isoforms and identification of eight truncated mitochondrial isoforms via immunoblotting. (**A**) Venn diagram of genome-wide *in silico* identification of putative genes targeted to mitochondria via dATI. 473 human and 413 mouse genes are predicted to encode a cryptic MTS, of which 168 are conserved between the two species. After filtering out 42 genes predicted to produce a mitochondrial protein based on translation starting at the first methionine, M1, 126 genes remained (Supplementary Table S2). (**B**) Immunoblot analysis of candidate gene products [ARAP1, FBXL12, MTSS1L, LEPRE1, TRIM32, NOX3, CLASP2 and FOXH1 from nuclear (Nuc), cytosolic (Cytos) and mitochondrial fractions of HEK293T cells]. Mitochondrial fractions were treated with 0, 10, 50 or 100 μg/ml trypsin. L, mitochondria lysed with 1% Triton X-100 and treated with 100 μg/ml trypsin. Gray arrowhead, full-length protein; white arrowhead, putative dATI isoform. Full-length and predicted dATI protein products of candidate genes are schematically depicted to the right of each immunoblot. Domains are as follows: PH, Pleckstrin homology; ArfGAP, Arf GTPase-activating protein; RhoGAP, Rho GTPase-activating protein; RA, Ras association; FBL, F-box-like; IMD, IRSp53/MIM homology domain; 2OG-Fell Oxy, 2OG-Fe(II) oxygenase superfamily; Zf-RING LisH: RING-type zinc-finger, LisH dimerization motif; NHL: NHL repeat; (NCL-1, HT2A and Lin-41); Ferric reduct: Ferric reductase like transmembrane component; FADB: FAD-binding domain; Fe reduct NADB: Ferric reductase NAD-binding domain; CLASP_N: N-terminal region of CLIP-associated proteins. TOM20 was used to show efficiency of trypsin treatment. TFAM, CALNEXIN, Splicing factor 2 (SF2) and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) were used as mitochondrial, endoplasmic reticulum, nuclear and cytosolic markers, respectively.

antibody would cross-react with a protein that was (i) enriched in mitochondrial extracts, (ii) shorter than the annotated full-length protein, while being maximally of a size predicted by dATI and (iii) resistant to trypsin degradation prior to mitochondrial lysis, like other internal mitochondrial proteins. For immunocytochemistry analysis, two versions of the cDNA were cloned, the full-length ORF, and a truncated form starting at the internal AUG predicted to mark the start of the cryptic MTS. In the immunoblotting experiments, 10 of 15 proteins tested fulfilled the criteria, suggesting they have a mitochondrial isoform. The 10 positive proteins were

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

ARAP1, FBXL12, MTSS1L, LEPRE1, TRIM32, NOX3, CLASP2 and FOXH1 (Figure 1B), as well as PABPC5 and PIF1α (see below). In the case of FBXL12, the MTS may not always be cleaved after import, as the antibody to this protein detected two bands specific to the mitochondrial cell fraction (Figure 1B).

There was no putative mitochondrial isoform in the cases of TNIK, PRPSAP2, MBRL or NRF1 (Supplementary Figure S1A). POP1 was another false positive, based on immunocytochemistry of the full-length and putative dATI isoform, as both long and short forms of the protein were targeted exclusively to nucleoli (Supplementary Figure S2). None of the negative controls tested yielded a short, trypsin-resistant, mitochondrial isoform based on immunoblotting (Supplementary Figure S1B). Therefore, our screen achieved a sensitivity and specificity of 77 and 72% respectively, with an estimated false discovery rate (FDR) of 33%. These results suggest that ~85 of the 126 identified genes will prove to encode dATI-dependent mitochondrial proteins, and thus are strong candidates for further validation. Nevertheless, multiple methods will be required to demonstrate that each predicted dATI-dependent gene product is a *bona fide* mitochondrial protein. On the other hand, a dATI variant that is predicted to be mitochondrial, but not detected in the organelle at first instance, should not be discounted, as dATI might be tightly regulated in some cases, and so the mitochondrial isoform might only be apparent in specific cell types, environmental conditions or stages of development.

The predicted polyA-binding protein PABPC5 has a dATI-dependent mitochondrial isoform

Poly(A) tails are attached to the 3′-end of almost all eukaryotic messenger RNAs, including those in mitochondria (21,22), and polyA-binding proteins (PABPs) bind to and modulate polyA tail length, with implications for mRNA stability and translation (22). Hence, mitochondria are expected to contain one or more PABPs, yet none had been identified hitherto. Therefore, the appearance of PABPC5 in the list of candidate proteins (Supplementary Table S2) was of particular interest, as it could be the long sought mitochondrial PABP (23).

Initiation from an internal AUG codon of Pabpc5 occurs *in vitro*, and human mitochondria contain a truncated isoform of PABPC5 concordant with dATI

The computational analysis predicted that translation initiation from methionine 33 of PABPC5 would generate a mitochondrial isoform of the protein. To test this prediction, Pabpc5 cDNAs were introduced into a coupled transcription/translation (TnT) system. Full-length Pabpc5 cDNA ($PABPC5^{1}$) produced two polypeptides (Figure 2A; lane 2) corresponding to translation products initiating at M1 and M33, based on comparisons with an N-terminally truncated template ($PABPC5^{33}$) (Figure 2A; lane 1), whereas a full-length mutant form of the protein, where M33 was replaced by isoleucine ($PABPC5^{M33I}$) (Figure 2A; lane 3), yielded a single polypeptide. Thus, the downstream AUG at position 33 is a functional start site.

An antibody to PABPC5 detected two proteins in the cytosol of HEK293T cells, the lower of which was also detected in trypsin-treated mitochondria (Figure 2B). Sub-mitochondrial fractionation localized the protein to the matrix compartment of the organelle (Supplementary Figure S3A), and *in vitro* synthesized $PABPC5^{33}$ resolved at the same position as the endogenous mitochondrial isoform of PABPC5 on SDS-PAGE gels (Figure 2C). In addition, full-length PABPC5 tagged at the carboxyl-terminus ($PABPC5^{1}$.HA) was targeted principally to the cytosol based on immunocytochemistry (Figure 2D), whereas deletion of the first 32 residues ($PABPC5^{33}$.HA) resulted in targeting of the protein to mitochondria, as predicted by the *in silico* screen (Figure 2E). Next, mitochondria were purified from transgenic HEK293T cells expressing FLAG-tagged PABPC5, starting at M33 ($PABPC5^{33}$.F), or full-length PABPC5 with a methionine-to-isoleucine mutation at the predicted internal start site ($PABPC5^{M33I}$.F). The full-length mutant protein was readily detectable in whole cell extracts, and as expected for a cytosolic protein, the little associated with isolated mitochondria was completely degraded with the addition of trypsin (Figure 2F). In contrast, tagged $PABPC5^{33}$ co-purifying with mitochondria survived trypsin treatment (Figure 2F). Therefore, translation initiation from methionine 33 of Pabpc5 yields a protein that is targeted to mitochondria in living cells.

Mitochondrial PABPC5 co-fractionates with mtDNA to a similar extent as DNA polymerase γ and mitochondrial single-stranded DNA-binding protein

To determine whether PABPC5 in mitochondria associated with nucleic acid, mitochondrial lysates were subfractionated through an iodixanol gradient. Southern blotting was used to detect mtDNA, and the co-fractionation of PABPC5 and known mtDNA replication proteins was assessed by immunoblotting. The majority of the signal from proteins known to be involved in mtDNA metabolism, such as TWINKLE, TFAM, SSBP1 and POLG1, resolved in the same fractions as mtDNA, whereas the chaperone protein HSP60 was mainly in the mtDNA-free fractions (Figure 3A). PABPC5 showed a distribution similar to SSBP1 and POLG1 (Figure 3A), suggesting that it is associated with mtDNA or RNA, as mitochondrial ribosomes and mRNAs co-fractionate with the DNA on iodixanol gradients (24).

Depletion of mtDNA destabilizes mitochondrial PABPC5

Many proteins associated with mitochondrial nucleic acid are rapidly turned over in its absence (25). To test whether mitochondrial PABPC5 was dependent on nucleic acid for its stability, HEK293T cells were largely depleted of their mitochondrial DNA and RNA by treatment with ethidium bromide (26,27). In these conditions, mitochondrial PABPC5 was markedly reduced in abundance, as were several known mtDNA and RNA-binding proteins, such as SLIRP, LRPPRC, DHX30, PTCD3, POLRMT

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

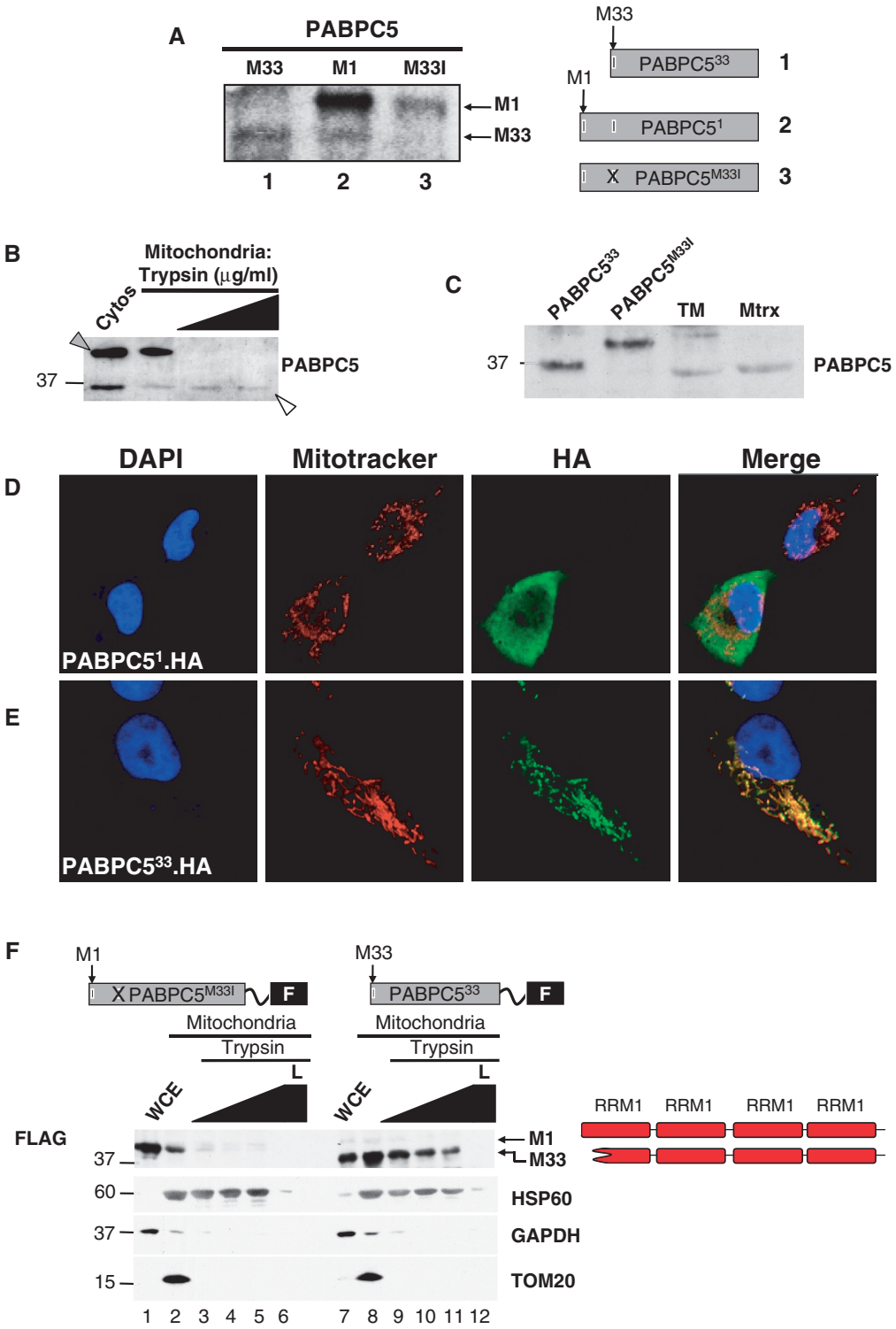

Figure 2. PABPC5 has a mitochondrial isoform consistent with dATI from M33. (**A**) SDS-PAGE of [^{35}S]-methionine-labelled PABPC5 polypeptide variants generated *in vitro*. M33 and M1 denote the methionines where translation starts, while M33I indicates a methionine to isoleucine point mutation at residue 33 of PABPC5. A consensus Kozak sequence (gccacc) was placed upstream of the first AUG (methionine) codon in each construct. Schematics of the constructs used as templates are indicated to the right of the representative gel image. 'X' in the cDNA schematic (lane 3) indicates mutation of the methionine residue. (**B**) Immunoblot analysis of PABPC5 cytosolic (Cytos), and mitochondrial fractions from HEK293T cells. Mitochondrial fractions were treated with 0, 10 or 50 μg/ml trypsin. Gray arrowhead, full-length protein; white arrowhead, putative dATI isoform. (**C**) Immunoblot analysis using anti-PABPC5 to PABPC5^{33} and PABPC5^{M33I} TnT products, total mitochondrial lysate (TM) and a mitochondrial matrix fraction (Mtrx). Confocal analysis of transiently transfected HOS cells with C-terminal HA-tagged cDNAs encoding (**D**) full-length human PABPC5 (PABPC5^{1}.HA) or (**E**) the dATI isoform with the first methionine at residue 33 (PABPC5^{33}.HA). Recombinant proteins were labelled with anti-HA antibody (green), while nuclei (blue) and mitochondria (red) were visualized by staining cells with DAPI and Mitotracker, respectively. (**F**) Mitochondria from PABPC5^{M33I}.F- and PABPC5^{33}.F-expressing cells (10 ng/ml doxycycline, 24 h) were purified

(continued)

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

and TFAM, whereas the mitochondrial protein Aconitase II, which does not interact with mtDNA, was unaffected by this treatment (Figure 3B). These results suggest that PABPC5 interacts with nucleic acid in mitochondria.

PABPC5 co-immunoprecipitates the mitochondrial poly(A) polymerase, but not mitochondrial ribosomes

Co-immunoprecipitation experiments, followed by immunoblotting, were used to determine potential protein–protein interactions, using the predicted dATI isoform of PABPC5, fused to a carboxy-terminal haemagglutinin (HA) or FLAG tag, as bait. Tagged PABPC5^{33} co-immunoprecipitated the mitochondrial poly(A) polymerase (mtPAP), and to a lesser extent LRPPRC, whereas a FLAG-tagged version of the mitochondrial deadenylase, PDE12 (28), did not interact with either of these proteins (Figure 3C and Supplementary Figure S3B). There was no enrichment of the mitochondrial RNA-binding protein SLIRP, nor was there any significant co-immunoprecipitation of the mitochondrial ribosomal protein, MRPS18, or the abundant respiratory chain subunit cytochrome c oxidase (COXIV) (Figure 3C and Supplementary Figure S3B). The specific interaction of PABPC5^{33} with mtPAP and LRPPRC, but not MRPS18, suggests it associates with mRNAs that are independent of mitochondrial ribosomes.

The RNA/DNA helicase PIF1α has a dATI-dependent mitochondrial isoform

Another dATI candidate from the *in silico* screen was the RNA/DNA helicase PIF1α. In yeasts, the PIF1 gene product is essential for mitochondrial DNA maintenance (29), and the mitochondrial isoforms of budding and fission yeast PIF1 are known to be generated via ATI, not dATI. Translation initiation from the first AUG codon of the ORF yields the mitochondrial isoform, whereas translation from a downstream in-frame AUG codon generates yeast nuclear PIF1 (30,31). In contrast, none of the annotated PIF1 transcripts in a variety of vertebrates possesses an in-frame AUG that could append an MTS to the protein (Supplementary Table S3). Nor do these PIF1 orthologs contain a proximal upstream ORF (uORF) adjacent to the main coding sequence (CDS), of the type associated with vertebrate RNase H1 genes (8), based on the annotated 5′ UTR sequences of all available PIF1 sequences in the NCBI and ensembl databases (Supplementary Table S3). It has been suggested that human PIF1α is a nuclear protein (32,33), and that a mitochondrial form of PIF1 (PIF1β) is generated via alternatively splicing (33). However, PIF1 was identified in our computational screen as a candidate dATI protein (Supplementary Table S2), and sequence alignments of PIF1 of several mammals revealed a potential MTS that would be revealed by dATI, from human M54 and mouse M67 (Figure 4A and B).

To screen for the predicted PIF1α product of dATI, cDNAs were translated *in vitro*. Full-length Pif1α cDNA produced two polypeptides (Figure 4C; lane 2), the shorter of which was not detected when the predicted dATI start site was modified by substituting methionine 54 with isoleucine (PIF1α^{M54I}) (Figure 4C; lane 3), and a PIF1α cDNA lacking the first 53 codons of the ORF (PIF1α^{54}) yielded a single polypeptide that was the same size as the short form of PIF1α^{1} (Figure 4C; lane 1). These results indicate that dATI occurs from the AUG encoding M54 of Pif1α, *in vitro*. Because evidence already existed for nuclear localization of PIF1α (32,33) and the full length protein is not detectable in mitochondria, we tested the ability of PIF1α^{54} to be imported into isolated rat liver mitochondria, using PIF1α^{1} and TFAM as negative and positive controls, respectively. When the PIF1α translation products were incubated with isolated mitochondria, the MTS of the predicted dATI isoform, PIF1α^{54}, was processed and imported in a membrane potential-dependent manner (Figure 4D, bottom panel), whereas there was no detectable mitochondrial import of full-length PIF1α (PIF1α^{1}) (Figure 4D middle panel). Although the efficiency of import of PIF1α^{54} into isolated mitochondria was lower than that of TFAM, this is also true of other well-established mitochondrial proteins, including components of the cytochrome *c* oxidase holoenzyme (12). The proposed mature form, PIF1α^{m}, was only seen when PIF1α^{54} translation products were incubated with mitochondria, and it alone was resistant to trypsin (Figure 4D, lanes 2 and 3, bottom panel), strongly suggesting import and processing of PIF1α^{54} by mitochondria. The size of PIF1α^{m} corresponds to an endogenous protein detected by a PIF1 antibody (Figure 3A) whose expression was reduced by a siRNA targeting PIF1 (Figure 4E) (34).

The predicted dATI forms of human and mouse PIF1 are targeted to mitochondria in cultured cells, and mutation of an internal AUG codon ablates mitochondrial localization of PIF1α

PIF1α was shown to display nuclear localization based on an N-terminal FLAG-tagged (FLAG.PIF1α) form of the protein (32,33). However, dATI bypasses N-terminal tags; therefore, cDNAs specifying full-length Pif1α tagged with HA, either at the N-terminus (HA.PIF1α^{1}) or the C-terminus (PIF1α^{1}.HA), were transiently expressed in

Figure 2. Continued
and subjected to trypsin protection assays, followed by immunoblotting. Schematic representations of the transgenes are depicted above the immunoblots. M1, indicates translation initiation at the annotated start methionine of PABPC5 according to ensembl genome browser, release 61; M33, indicates the band corresponding to the cDNA product starting from the dATI residue of PABPC5. Whole cell extract (WCE, lanes 1 and 7); black slope, indicates increasing trypsin concentrations (μg/ml) of 0 (lanes 2 and 8); 10 (lanes 3 and 9); 50 (lanes 4 and 10); and 100 μg/ml (lanes 5 and 11). L, mitochondria lysed with 1% Triton X-100 and treated with 100 μg/ml trypsin, lanes 6 and 12. PABPC5.F transgenes were detected with anti-FLAG antibody. Heat shock protein 60 (HSP60), mitochondrial marker; GAPDH, cytosolic marker. TOM20 was used to show efficiency of trypsin treatment. Full-length and predicted dATI protein products of candidate genes are schematically depicted to the right. RRM1, RNA recognition motif.

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

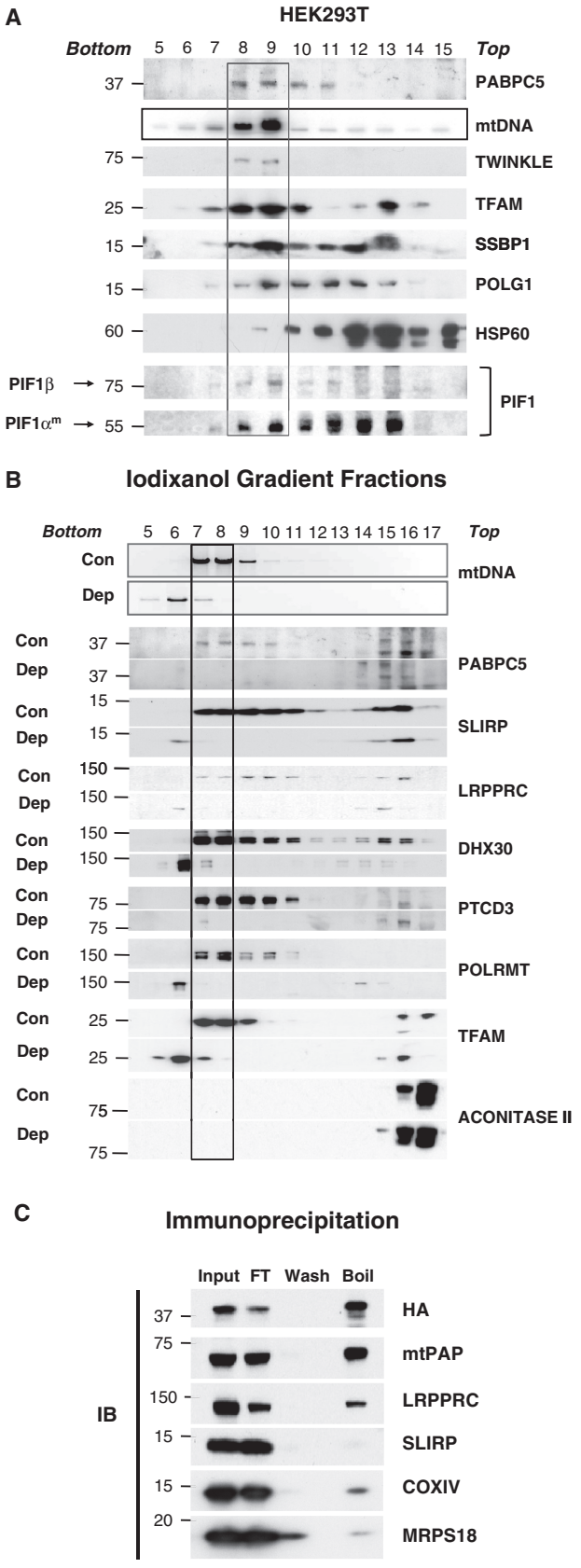

Figure 3. PABPC5 co-precipitates mitochondrial RNA-binding proteins and is dependent on mitochondrial nucleic acids for its stability. (**A**) Mitochondria were isolated from HEK293T cells, trypsin-treated (50 μg/ml), lysed and centrifuged through an iodixanol gradient. Fractions were collected from the bottom of the tube and then subjected either to DNA extraction, followed by Southern blotting, or to SDS-PAGE, followed by immunoblotting. Numbers above the immunoblots indicate iodixanol fractions, in order of collection. (**B**) After HEK293T cells were either grown under normal conditions (Con) or treated with ethidium bromide (100 ng/ml) for 72 h (Dep), mitochondrial lysates were fractionated and analysed as per panel A. (**C**) Mitochondria were isolated from cells overexpressing PABPC5^{33}.HA after 72 h of doxycycline (30 ng/ml) treatment, followed by immunoprecipitation with anti-HA agarose beads. Immunoblotting was used to determine the extent of co-immunoprecipitation of selected proteins using 2.5% of input, flow-through (FT) and wash, and 5% of the eluate.

human cells. HA.PIF1α^{1}, like FLAG.PIF1α, was located exclusively in the nucleus (Figure 5A), whereas the products of the C-terminally tagged, full-length PIF1α cDNA (PIF1α^{1}.HA) co-localized chiefly with mitochondria, with a small amount of signal detected in nuclei (Figure 5B). Enforced translation from M54 (PIF1α^{54}.HA) resulted in exclusive mitochondrial targeting of the protein (Figure 5C), whereas replacing M54 with isoleucine (I54) yielded a protein that did not co-localize with mitochondria (Figure 5D). These results show that a functional MTS is revealed when human PIF1α starts at M54, and that this residue is required to generate a mitochondrial isoform of PIF1α. Analysis of mouse PIF1 constructs provided further support for the generation of a mitochondrial-targeted form of PIF1 via dATI, as mPIF1^{67}.HA (the mouse equivalent of PIF1α^{54}.HA) co-localized with mitochondria (Figure 5E), whereas there was no detectable PIF1 in mitochondria after mutating the AUG codon predicted to yield translation initiation of the mitochondrial isoform of mouse PIF1 (mPIF1^{M67I}.HA) (Figure 5F).

N-terminally truncated human PIF1α is imported and processed by mitochondria in cultured cells

Subcellular fractions were prepared from human cells expressing PIF1α^{1}.HA, PIF1α^{54}.HA or PIF1α^{M54I}.HA to determine the location of the recombinant proteins. Tagged full-length PIF1α (M1) was the only signal detected in purified nuclei (Figure 6A, lane 1), and although some of this isoform was also associated with mitochondrial fractions (Figure 6A, lane 2), it was degraded by trypsin treatment (Figure 6B, lanes 3–5). The mitochondrial extracts additionally contained two shorter PIF1α products, one corresponding to the pre-protein (M54) and the other to a shorter protein (m) (Figure 6A, lane 2). We infer that the shorter of the two is the mature form of mitochondrial PIF1α, both because it is the more abundant and it is similar in size (the small difference being attributable to the tag) to an endogenous mitochondrial protein identified by the PIF1 antibody (Figures 3A and 4E). Consistent with this interpretation, cells expressing PIF1α^{54}.HA produced only the two presumed mitochondrial-specific species, M54 and m, both of which were resistant to degradation by trypsin (Figure 6C). Thus, full-length PIF1α is outside mitochondria, whereas M54 and m are located inside the organelle. Both these mitochondrial isoforms depend on

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

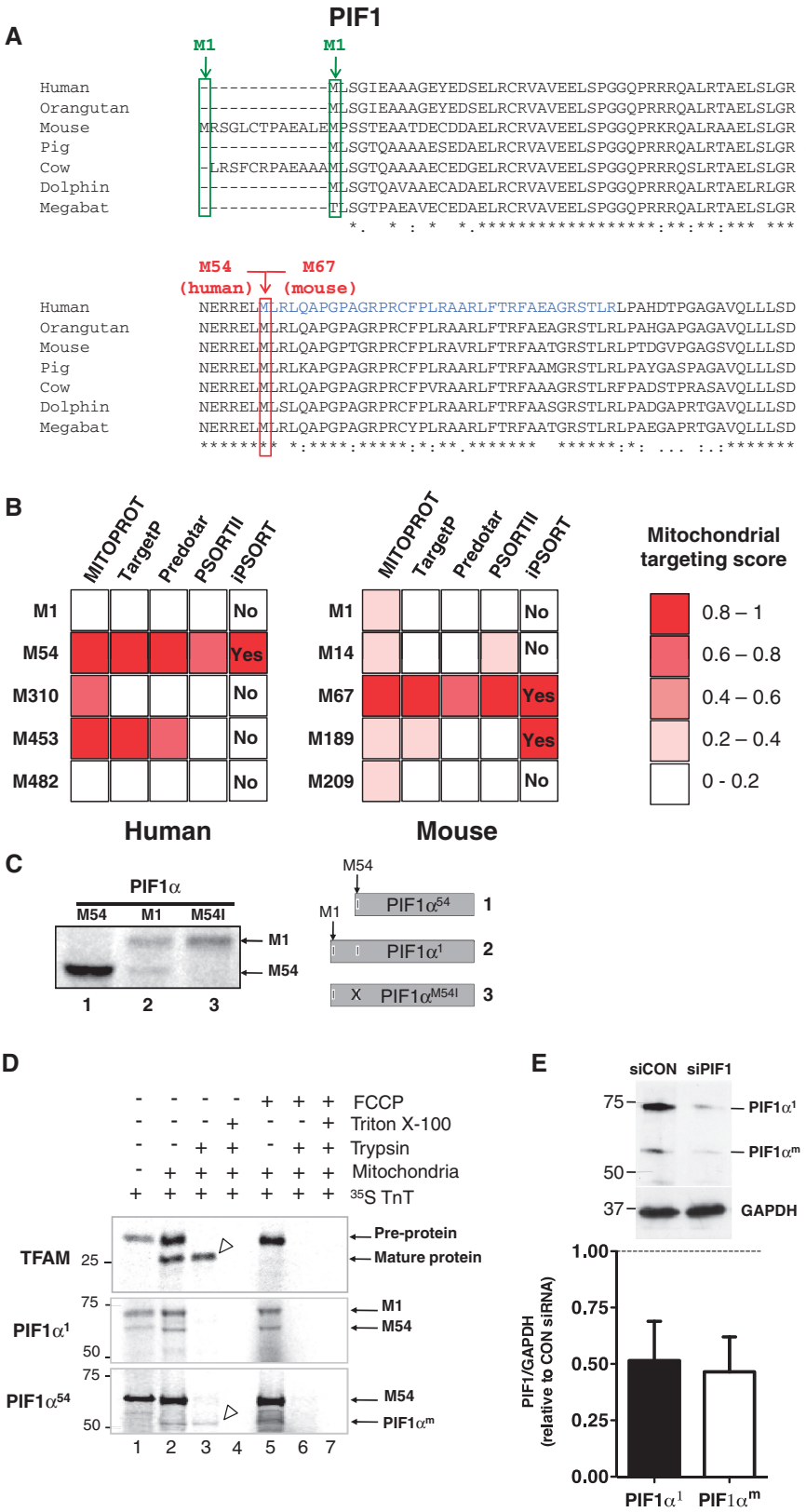

Figure 4. Mammalian PIF1 has a predicted dATI-dependent mitochondrial isoform, some of which can be imported into isolated mitochondria. (**A**) N-terminal amino acid sequence alignment of PIF1 between human, orangutan, mouse, pig, cow, dolphin and megabat. M1 and M54 above the alignment correspond to methionines 1 and 54, respectively of human PIF1α. Green font and box, canonical start sites in mouse and human; red font and box, putative methionine residue where dATI begins; blue font, Mitoprot-predicted cleavable MTS. (**B**) *In silico* mitochondrial targeting prediction scores from: Mitoprot, TargetP, Predotar (35), PSORTII (36) and iPSORT of human and mouse PIF1. 'M' followed by a number, on the left side of the heat map, indicate the methionine residue numbers; red-coloured boxes in the heat map indicate strong prediction for

(continued)

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

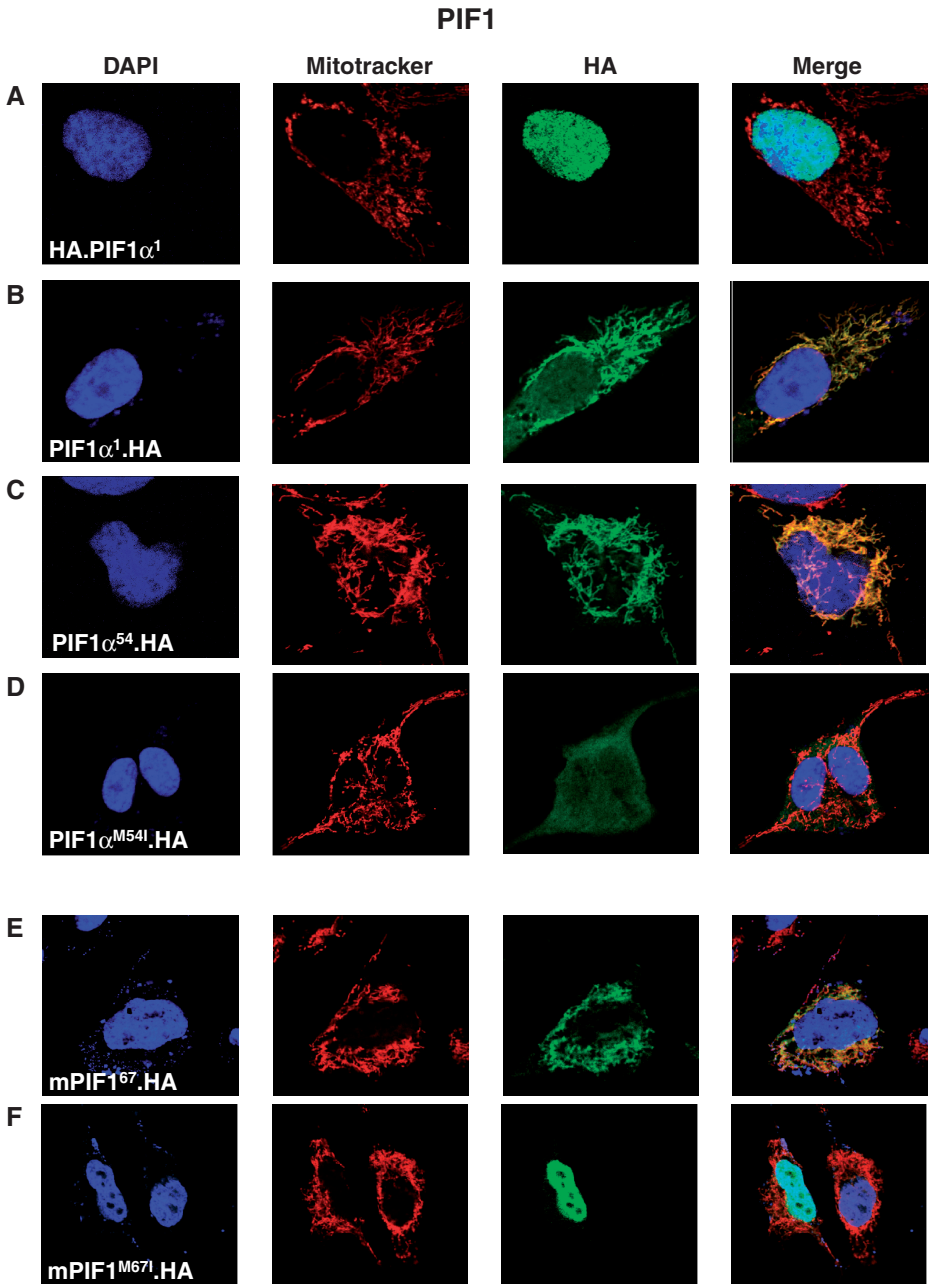

Figure 5. Immunocytochemistry of ectopically expressed human and mouse PIF1 suggests mitochondrial targeting via dATI. HOS cells were transiently transfected with human Pif1α full-length cDNA containing a HA tag on either the (**A**) N-terminus (HA.PIF1α^{1}) or (**B**) C-terminus (PIF1α^{1}.HA). The remaining Pif1α constructs were HA-tagged on the C-terminus, consisting of Pif1α.HA cDNAs that were (**C**) forced to start at M54 (PIF1α^{54}.HA) or (**D**) with a methionine to isoleucine substitution at residue 54 (PIF1α^{M54I}.HA). Mouse PIF1 (mPIF1) constructs were C-terminally tagged and included (**E**) a cDNA forced to start translation at M67 (mPIF1^{67}.HA) or (**F**) a cDNA encoding a methionine-to-isoleucine mutation at residue 67 (mPIF1^{M67I}.HA). Recombinant proteins were labelled with anti-HA antibody (green), nuclei were stained blue with DAPI, and mitochondria were stained red with Mitotracker.

Figure 4. Continued
mitochondrial targeting. (**C**) SDS-PAGE of [^{35}S]-methionine-labelled human PIF1α polypeptide variants generated *in vitro*. M54 and M1 refer to the methionines where translation starts, while M54I indicates a methionine to isoleucine point mutation at residue 54 of PIF1α. A consensus Kozak sequence (gccacc) was placed upstream of the first AUG (methionine) codon in each construct. Schematics of the constructs used as templates are indicated to the right of the representative gel image. 'X' in the cDNA schematic (lane 3) indicates mutation of the methionine residue. (**D**) [^{35}S]-methionine-labelled TFAM (positive control), PIF1α^{1} and PIF1α^{54} were incubated with isolated rat liver mitochondria. 1 μM FCCP (lanes 5–7) was used to dissipate membrane potential. White arrowheads indicate imported polypeptide. Import efficiency was determined relative to import of TFAM. Start methionines are indicated to the right of the gel images. (**E**) HEK293T cells were transfected with 200 pmol of scrambled siRNA (siCON) or siRNA targeted towards PIF1 (siPIF1). Forty-eight hours later, whole cell extracts were used for immunoblotting with anti-PIF1. GAPDH was used as a loading control. The chart accompanying the immunoblots shows the extent of knockdown relative to GAPDH protein (n = 3 independent experiments).

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

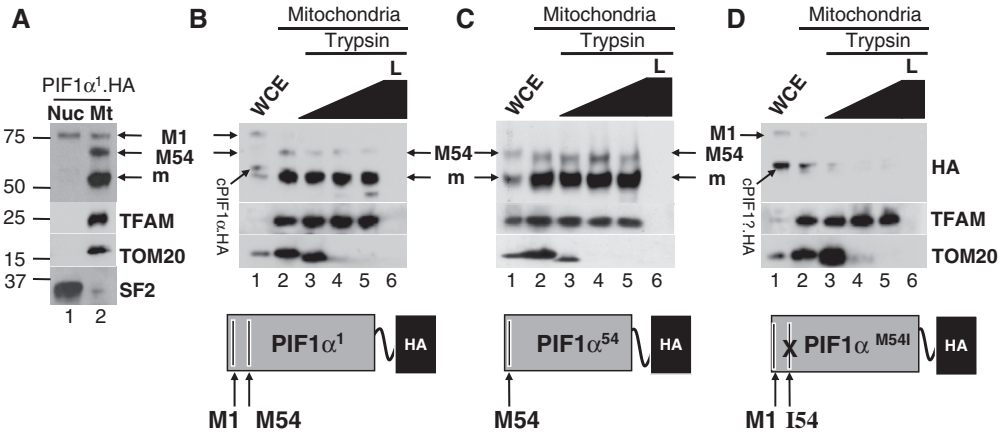

Figure 6. Subcellular fractionation confirms mitochondrial targeting and processing of the dATI forms of PIF1α^{54} in cells. (**A**) Purified nuclei (Nuc) and mitochondria (Mt) were isolated from HEK293T cells stably expressing PIF1α^{1}.HA followed by immunoblot analyses. Mitochondria from transgenic HEK293T cells expressing (**B**) PIF1α^{1}.HA, (**C**) PIF1α^{54}.HA, or (**D**) PIF1α^{M54I}.HA were purified and subjected to trypsin protection assays, followed by immunoblotting. Whole cell extracts (WCE) were fractionated alongside mitochondrial lysates to show all products of the cDNAs. Transgenes were expressed by the addition of 10 ng/ml doxycycline for 24 h. Schematic representations of the transgenes are depicted below the immunoblots. M1, indicates translation initiation at the annotated start methionine of PIF1α according to ensembl genome browser, release 61; M54, the band corresponding to the dATI residue of PIF1α; m, the putative mature product of PIF1α.HA after mitochondrial import and removal of the MTS. (B–D) WCE (lane 1); black slope, indicates increasing trypsin concentrations of 0, 10, 50 and 100 μg/ml (lanes 2–5, respectively). L, mitochondria lysed with 1% Triton X-100 and treated with 100 μg/ml trypsin (lane 6). PIF1α.HA transgenes were detected with anti-HA antibody. TFAM and SF2 are mitochondrial and nuclear markers, respectively. TOM20 was used to show efficiency of trypsin treatment. (B and D) cPIF1α.HA, processed full-length PIF1α that resides in the cytoplasm, not in mitochondria.

initiation at M54, as neither was detected in cells expressing the PIF1α^{M54I}.HA variant, and all derivatives of PIF1α^{M54I}.HA were degraded when mitochondria were exposed to trypsin (Figure 6D).

A species (cPIF1α) migrating between M54 and m was detected in cells expressing PIF1α^{1}.HA (Figure 6B, lane 1), and this was the major product in PIF1α^{M54I}.HA cells (Figure 6D, lanes 1 and 2). In the latter cells, the recombinant protein was dispersed throughout the cytosol (Figure 5D), and cPIF1α did not survive trypsin treatment of intact mitochondria (Figure 6B and D). Thus, cPIF1α probably represents a proteolytic cleavage product that is formed after export of tagged PIF1α from the nucleus, as occurs to the native protein during the course of the cell cycle (32). In contrast, murine PIF1^{M67I}.HA was concentrated in the nucleus (Figure 5F), and so it may not be recognized by the human nuclear export machinery. Similarly, the HA N-terminal tag may interfere with nuclear export of PIF1α^{1} (Figure 5A). In summary, dATI from the AUG corresponding to M54 of human Pif1α generates a pre-protein, PIF1α^{54}, which is cleaved after mitochondrial import, yielding a mature mitochondrial isoform, PIF1α^{m}, and mutation of the predicted dATI sites ablates mitochondrial isoforms of human and mouse PIF1.

Endogenous N-terminally truncated PIF1α is present in mitochondria and co-fractionates with mtDNA

The mitochondrial lysates fractionated by iodixanol gradient sedimentation and probed for PABPC5 were also used to evaluate endogenous forms of PIF1. The most abundant form of PIF1 detected by immunoblotting corresponded to PIF1α^{m} (the presumed mature dATI isoform of PIF1α), 30% of which co-fractionated with mtDNA (Figure 3A). The largest species detected by the PIF1 antibody was a polypeptide of ~75 kDa, which is the predicted size of PIF1β (33), and it was concentrated in the same fractions as the mtDNA. Thus, there appear to be two forms of PIF1 in human mitochondria, PIF1α^{m} and PIF1β, both of which may interact with mtDNA.

DISCUSSION

The mammalian mitochondrial proteome is estimated at 1500 proteins. If accurate, then some 400 mitochondrial proteins remain to be identified (19). The dATI-generated mitochondrial proteins predicted from our analysis could account for a substantial fraction of the 'missing' mitochondrial proteome, as >100 gene products in our list have not previously been annotated as mitochondrial proteins. Ten of 15 candidates tested from the 126-gene inventory appear to have a form of the protein in mitochondria corresponding to the predicted dATI isoform. Taking account of the estimated false discovery rate of 33%, then ~85 of the identified genes may prove to yield proteins targeted to mitochondria in a dATI-dependent manner. However, in view of the low specificities of mitochondrial prediction programs, it is inevitable that many genes will be missing from the inventory. The thyroid hormone receptor c-Erb A α1 (11) is a known example. Thus, the number of proteins targeted to mitochondria via dATI could be in the hundreds. Although the dATI screen could be refined further to include the likes of c-Erb A α1, simply lowering the threshold for acceptance will inevitably produce a marked increase in false positives. This might be offset in other

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

ways, such as a greater demand for conservation among diverse species, or the incorporation of additional algorithms that predict mitochondrial targeting (37). Conservation is a critical filter because randomly generated peptides can create an MTS (38). As noted above, evidence of mitochondrial involvement will be an important guide in many cases and this is likely to be facilitated by the explosion in RNA expression data providing comprehensive details of co-expression in a variety of cell and tissue types. Detailed mining of existing and new mitochondrial proteome studies will also doubtless reap reward, as it is likely that many mitochondrial proteins have been mistakenly dismissed as contaminants, β-actin being a case in point (14). Conversely, it is possible that some genes on the list have genuine mitochondrial isoforms that do not depend on a canonical N-terminal MTS, or are generated via alternative transcripts, rather than by dATI. Ultimately any *in silico* prediction of putative internal start codons will require experimental verification by multiple methods for each and every candidate, as per PIF1 and PABPC5.

Although the known or inferred functions of the 126 candidate mitochondrial dATI proteins are highly varied, they can be grouped into a number of categories (Table 1). Thirty-one of the candidates have links to RNA or DNA metabolism, and these can be further sub-divided into transcriptional regulation, post-transcriptional modification and DNA replication or repair. The earlier identification of nuclear transcription factors in mitochondria led others to suggest that the regulation of mitochondrial gene expression might share certain aspects with nuclear gene expression (39), and 17 candidate dATI-generated mitochondrial proteins are recognized transcriptional regulators (Table 1). Actomyosin has recently been found inside mammalian mitochondria and implicated in mtDNA maintenance (14), and 12 candidates have links to actin binding and organization, two of which were substantiated by immunoblotting (Figure 1B). Furthermore, the list includes two members of the WNK (WNK lysine-deficient protein kinase) serine/threonine kinase subfamily (WNK1 and WNK2), which act on Rho GTPases and control actin dynamics. These two proteins physically interact with one another (40), and WNK1 enhances the activity of the annotated mitochondrial protein, SGK1 (serum- and glucocorticoid-induced kinase 1) (41). Thus, dATI may be a mechanism of protein trafficking that maintains the balance between mitochondrial and cytoplasmic actin.

dATI-dependent mitochondrial isoforms confound conventional genetic analysis, as gene knockdown and ablation will affect all the protein isoforms, and so the relative contributions of the mitochondrial and non-mitochondrial proteins cannot be judged. Gene replacement of one particular isoform could circumvent this problem, as was achieved for the alternatively translated DNA ligase III, which identified the mitochondrial, not the nuclear, enzyme as the essential variant for cell viability (42).

Based on the immunoblotting results and the domain structure of some of the dATI candidates (Figure 1B), loss of a portion of the N-terminus will affect their function. For instance, the dATI isoform of FBXL12 is predicted to lack the F-box-like domain that gives it its name (Figure 1B), and half of a RNA recognition motif is missing from the mitochondrial isoform of PABPC5 (Figure 2F).

The dATI-mediated mitochondrial isoform of PABPC5

Although several mitochondrial proteins have been shown previously to bind RNA and poly(A) sequences in mitochondria (43–45), none binds poly(A) tails preferentially. Of the four human PABPC genes, three (*PABPC1*, *PABPC3* and *PABPC4*) are established polyA-binding proteins (46–48); however, none has a predicted N-terminal MTS based on the annotated first AUG codon, or downstream AUG codon, and only PABPC5 predicts a dATI-dependent mitochondrial isoform. Several lines of evidence provide support for the proposed dATI-mediated mitochondrial isoform of PABPC5. A cDNA of the complete ORF yields two proteins, one of which is of the size predicted by dATI, and when expressed in cells the shorter form is imported into mitochondria. Endogenous PABPC5 is present in mitochondria ($PABPC5^{m}$) based on immunoblotting, and it co-fractionates with mitochondrial nucleic acids. $PABPC5^{m}$ is unstable when mitochondrial DNA and RNA are depleted, and it interacts with mitochondrial RNA-binding proteins, implying PABPC5 in mitochondria is bound to RNA. Hence, $PABPC5^{m}$ is proposed to play a role in post-transcriptional mitochondrial RNA metabolism, and based on its high homology to other PABP family members, it is a highly credible candidate for the long sought mitochondrial poly(A)-binding protein. All four PABPCs contain four non-identical RNA recognition motifs (RRMs). Typically, these motifs are linked to a C-terminal domain through a proline-rich (P-rich) region (22). However, the C-terminus of PABPC1 can be deleted without affecting poly(A) binding *in vitro* (49), or compromising translation in *Xenopus* (50) or viability in yeast (51). Only RRM motifs 1 and 2 need be retained to preserve the protein's ability to bind to poly(A) tails (52). Moreover, whereas RRMs 1 and 2 are highly selective for polyadenylated RNA, RRMs 3 and 4 are less discriminative, and can also bind AU-rich RNA (48), and this may be an important feature of the mitochondrial variant of PABPC5, as some mitochondrial transcripts have been shown to be polyuridinylated, poly(U) (53,54). Therefore, although full-length PABPC5 lacks the P-rich linker region and the C-terminal domain of the other PABPCs (55), and the dATI form lacks RRM1 (Figure 2F), both cytosolic and mitochondrial isoforms of PABPC5 are still likely to have the capacity to bind to polyadenylated RNA. Nevertheless, the structural differences between PABPC5 and the other family members might be indicative of a distinct property or role, and the function of the shorter mitochondrial isoform may differ from its cytosolic counterpart.

The dATI mediated mitochondrial isoform of PIF1α

In the case of PIF1α, dATI exposes the MTS and simultaneously removes the nuclear localization signal (NLS).

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

Table 1. Functional categories of candidate dATI genes

Proposed function of dATI-dependent proteins					
Transcriptional regulation	Post-transcriptional modification	DNA Replication/Repair	Actin cytoskeleton	Kinases	Translation
ASH2L	**C19orf6**	C9orf102	ANK2	MINK1	RPL4
C11orf30	DHX15	TEP1	ANK3	MAP4K4	FTSJ3
PHTF1	ERI3	**PIF1**	**ARAP1**	MLKL	DUS4L
SMARCB1	**PABPC5**	ZNF335	FARP2	PRKCE	EIF2S1
TRRAP	**POP1**	ZRANB3	IQSEC1	**TNIK**	RPS15
ZMYM3	RBM25		IQSEC2	WNK1	BMS1
ZMYM4	SRRM2		MTSS1	WNK2	
ELK4	NFKBIL1		**MTSS1L**	WNK4	
FOXH1	YTHDC2		MYO10	*TBRG4*	
GATA4			MYO15A	CDK12	
MLF1			STARD13	DAPK1	
NR1I2			ARHGAP32	DGKD	
NRF1				DSTYK	
PAX4					
MYST3					
NR6A1					
SRFBP1					

Collagen	Microtubules	Beta oxidation	Receptors	GTP binding	Calcium sensing
COL16A1	**CLASP2**	*CPT1B*	CHRM1	C9orf86	MCTP1
COL27A1	SFI1	*CPT1C*	CHRM5	RASAL3	IQCF5
LEPRE1	ASPM	DECR2	GPR162		

Channels	Trafficking	Ubiquitination	Cell–cell communication	Proteases	Transporters
ACCN1	NUP155	**FBXL12**	DSP	MMP25	CNNM2
SCN11A	YIPF3	RNF111	DTX4	MMP28	ABCC6
TRPM2	DOPEY1	**TRIM32**	ICA1	PSMD1	ABCC4
CACNA1B	CLINT1	ZYG11B	PKP4	RHBDL3	
CACNA1F	CCDC157	FBXO38	SRCIN1	TMPRSS11F	
HCN1		COMMD10	C1QTNF3	CTSF	
TMC6				AGBL1	
CYBB				USP49	

Apoptosis	Phospho-lipase	TCA cycle	ROS	Nucleotide synthesis
SPATA17	PNPLA6	*IDH1*	**NOX3**	**PRPSAP2**
			NOX4	

Cysteine modification	Retinoic acid metabolism	Phospholipid biosynthesis	Unknown function
HHAT	CYP26A1	*PISD*	FAM160B1
		PLCB4	C14orf118

All 126 genes from Supplementary Table S2 were grouped according to known function.
Bold font, genes tested in this study, that have a product in mitochondria based on experiment, unless coloured gray; blue font, genes found in human Mitocarta; red font, genes found in Mitominer.

Prior research on PIF1α assigned it a nuclear location, based chiefly on N-terminal tagging of the recombinant protein (32,33). However, N-terminal tags mask mitochondrial targeting signals that are present at the N-terminus (24), and a protein derived from an internal MTS will be undetectable because dATI bypasses N-terminal tags. Therefore, the previous approach was not capable of revealing a mitochondrial isoform of PIF1α. Accordingly, C-terminal tags are most appropriate for localization studies of potential mitochondrial proteins.

Based on its mobility on denaturing gels, the mass of the dATI-dependent mitochondrial PIF1α ($PIF1\alpha^{m}$) is estimated at 55 ± 2.5 kDa, placing the cleavage site at residue 165 ± 30. Despite $PIF1\alpha^{54}$ being processed to an even shorter form upon mitochondrial import, to generate $PIF1\alpha^{m}$, the mature form is not expected to differ substantially from the full-length protein in terms of its core activities, because the first defined functional SFI helicase motif begins at residue 224 (56), which $PIF1\alpha^{m}$ appears to retain. Furthermore, a recombinant form of human PIF1α (PIF1ΔN) lacking the first 166 residues

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

has the same ATPase and helicase activities as the full-length protein (57). Therefore, the activities of PIF1α^{m} are expected to be similar to its full-length nuclear counterpart, although contextual differences between nuclei and mitochondria might mean that the function of PIF1α^{m} is distinct from the long form of PIF1α, *in vivo*.

Human PIF1β—the alternatively spliced mitochondrial isoform

The detection, with an antibody to PIF1, of a protein of a mass of ~75 kDa co-fractionating with mtDNA (Figure 3A), lends support to the proposal that alternative splicing gives rise to a dedicated transcript encoding a human mitochondrial PIF1 isoform, PIF1β (33). PIF1β is annotated as a manually verified transcript in the ensembl database. Assuming it is correctly assigned, PIF1β starts at M1 and so depends on signals located at the C-terminus (which PIF1α lacks) to achieve mitochondrial targeting (33). However, a genetic approach that targets PIF1β, while sparing PIF1α will be needed to clarify its physiological importance, as PIF1β does not appear to be conserved even among primates (Supplementary Figure S4). Orangutan lacks the consensus splice acceptor site, which in humans gives rise to Pif1β mRNA, and in mouse there are no annotated alternative mouse PIF1 isoforms. Although there is a putative AG splice acceptor site in the mouse Pif1 gene, the resultant four-nucleotide insertion directly downstream would create a variant PIF1 unlike human PIF1β (Supplementary Figure S4). Moreover, this hypothetical mouse PIF1 mRNA does not contain any appreciable homology (at the nucleotide level) to the portion of the human PIF1β protein that is required for mitochondrial targeting (33). Therefore, there is no evidence of a PIF1β mouse variant. Nevertheless, there is considerable sequence variation among mitochondrial targeting signals and so it remains possible that the mouse PIF1 sequence contains a carboxy-terminal MTS.

The extent of dATI-dependent mitochondrial targeting

An increasing number of nuclear DNA transacting proteins are also found in mitochondria (33,58–60), but the task of defining the organelle-specific forms is far from complete. Information regarding the subcellular localization of these proteins is essential for determining their precise roles within the cell. The identification of dATI-dependent mitochondrial isoforms of PABPC5 and PIF1α suggests that this trafficking mechanism will prove to be a significant contributor to the dual targeting of proteins. Other candidates well worthy of further investigation for a role in mitochondrial nucleic acid metabolism include the putative RNA exonuclease ERI3 and the RNA helicase YTHDC2 (Table 1 and Supplementary Table S2).

The *in silico* screen and cell biology data of this report strongly support the idea of dATI playing an important role in targeting proteins to mitochondria, potentially accounting for up to half of the unassigned mitochondrial proteins, or ~10% of the total mitochondrial proteome. This raises the question of how expression of this class of genes is regulated and the nature of the mitochondrial–nuclear and mitochondrial–cytoplasmic communication pathways involved. The use of dATI permits antagonistic regulation, as translation of one isoform inherently opposes translation of the other isoform (61); therefore, deregulated ATI could result in category of human disease.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3 and Supplementary Figures 1–4.

ACKNOWLEDGEMENTS

We thank Professors A. Suomalainen, R. Wiesner and D. Blake for generous gifts of the TWINKLE, TFAM and TRIM32 antibodies, respectively.

FUNDING

Cambridge University Commonwealth Trust, Fellowship (to L.K.); UK Medical Research Council. Funding for open access charge: UK Medical Research Council.

Conflict of interest statement. None declared.

REFERENCES

1. Chacinska,A., Koehler,C.M., Milenkovic,D., Lithgow,T. and Pfanner,N. (2009) Importing mitochondrial proteins: machineries and mechanisms. *Cell*, **138**, 628–644.
2. Wiedemann,N., Frazier,A.E. and Pfanner,N. (2004) The protein import machinery of mitochondria. *J. Biol. Chem.*, **279**, 14473–14476.
3. Bazykin,G.A. and Kochetov,A.V. (2011) Alternative translation start sites are conserved in eukaryotic genomes. *Nucleic Acids Res.*, **39**, 567–577.
4. Prats,A.C., De Billy,G., Wang,P. and Darlix,J.L. (1989) CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. *J. Mol. Biol.*, **205**, 363–372.
5. Curran,J. and Kolakofsky,D. (1988) Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. *EMBO J.*, **7**, 245–51.
6. Strubin,M., Long,E.O. and Mach,B. (1986) Two forms of the Ia antigen-associated invariant chain result from alternative initiations at two in-phase AUGs. *Cell*, **47**, 619–625.
7. Ivanov,I.P., Firth,A.E., Michel,A.M., Atkins,J.F. and Baranov,P.V. (2011) Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.*, **39**, 4220–4234.
8. Suzuki,Y., Holmes,J.B., Cerritelli,S.M., Sakhuja,K., Minczuk,M., Holt,I.J. and Crouch,R.J. (2010) An upstream open reading frame and the context of the two AUG codons affect the abundance of mitochondrial and nuclear RNase H1. *Mol. Cell. Biol.*, **30**, 5123–5134.
9. Leissring,M.A., Farris,W., Wu,X., Christodoulou,D.C., Haigis,M.C., Guarente,L. and Selkoe,D.J. (2004) Alternative translation initiation generates a novel isoform of insulin-degrading enzyme targeted to mitochondria. *Biochem. J.*, **383**, 439–446.
10. Land,T. and Rouault,T.A. (1998) Targeting of a human iron-sulfur cluster assembly enzyme, nifs, to different subcellular compartments is regulated through alternative AUG utilization. *Mol. Cell*, **2**, 807–815.

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

11. Wrutniak,C., Cassar-Malek,I., Marchal,S., Rascle,A., Heusser,S., Keller,J.M., Fléchon,J., Dauça,M., Samarut,J. and Ghysdael,J. (1995) A 43-kDa protein related to c-Erb A alpha 1 is located in the mitochondrial matrix of rat liver. *J. Biol. Chem.*, **270**, 16347–16354.
12. Petruzzella,V., Tiranti,V., Fernandez,P., Ianna,P., Carrozzo,R. and Zeviani,M. (1998) Identification and characterization of human cDNAs specific to BCS1, PET112, SCO1, COX15, and COX11, five genes involved in the formation and function of the mitochondrial respiratory chain. *Genomics*, **54**, 494–504.
13. Cooper,H.M. and Spelbrink,J.N. (2008) The human SIRT3 protein deacetylase is exclusively mitochondrial. *Biochem. J.*, **411**, 279–285.
14. Reyes,A., He,J., Mao,C.C., Bailey,L.J., Di Re,M., Sembongi,H., Kazak,L., Dzionek,K., Holmes,J.B., Cluett,T.J. *et al.* (2011) Actin and myosin contribute to mammalian mitochondrial DNA maintenance. *Nucleic Acids Res.*, **39**, 5098–5108.
15. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2012) Ensembl 2012. *Nucleic Acids Res.*, **40**, D84–D90.
16. Claros,M.G. and Vincens,P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.*, **241**, 779–786.
17. Emanuelsson,O., Brunak,S., von Heijne,G. and Nielsen,H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.*, **2**, 953–971.
18. Bannai,H., Tamada,Y., Maruyama,O., Nakai,K. and Miyano,S. (2002) Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics*, **18**, 298–305.
19. Pagliarini,D.J., Calvo,S.E., Chang,B., Sheth,S.A., Vafai,S.B., Ong,S.-E., Walford,G.A., Sugiana,C., Boneh,A., Chen,W.K. *et al.* (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell*, **134**, 112–123.
20. Smith,A.C. and Robinson,A.J. (2009) MitoMiner, an integrated database for the storage and analysis of mitochondrial proteomics data. *Mol. Cell. Proteomics*, **8**, 1324–1337.
21. Rorbach,J. and Minczuk,M. (2012) The post-transcriptional life of mammalian mitochondrial RNA. *Biochem. J.*, **444**, 357–373.
22. Kühn,U. and Wahle,E. (2004) Structure and function of poly(A) binding proteins. *Biochim. Biophys. Acta.*, **1678**, 67–84.
23. Wydro,M., Bobrowicz,A., Temperley,R.J., Lightowlers,R.N. and Chrzanowska-Lightowlers,Z.M. (2010) Targeting of the cytosolic poly(A) binding protein PABPC1 to mitochondria causes mitochondrial translation inhibition. *Nucleic Acids Res.*, **38**, 3732–3742.
24. He,J., Cooper,H.M., Reyes,A., Di Re,M., Sembongi,H., Litwin,T.R., Gao,J., Neuman,K.C., Fearnley,I.M., Spinazzola,A. *et al.* (2012) Mitochondrial nucleoid interacting proteins support mitochondrial protein synthesis. *Nucleic Acids Res.*, **40**, 6109–6121.
25. Poulton,J., Morten,K., Freeman-Emmerson,C., Potter,C., Sewry,C., Dubowitz,V., Kidd,H., Stephenson,J., Whitehouse,W. and Hansen,F.J. (1994) Deficiency of the human mitochondrial transcription factor h-mtTFA in infantile mitochondrial myopathy is associated with mtDNA depletion. *Hum. Mol. Genet.*, **3**, 1763–1769.
26. Maniura-Weber,K., Goffart,S., Garstka,H.L., Montoya,J. and Wiesner,R.J. (2004) Transient overexpression of mitochondrial transcription factor A (TFAM) is sufficient to stimulate mitochondrial DNA transcription, but not sufficient to increase mtDNA copy number in cultured cells. *Nucleic Acids Res.*, **32**, 6015–6027.
27. Wiseman,A. and Attardi,G. (1978) Reversible tenfod reduction in mitochondria DNA content of human cells treated with ethidium bromide. *Mol. Gen. Genet.*, **167**, 51–63.
28. Rorbach,J., Nicholls,T.J.J. and Minczuk,M. (2011) PDE12 removes mitochondrial RNA poly(A) tails and controls translation in human mitochondria. *Nucleic Acids Res.*, **39**, 7750–7763.
29. Foury,F. and Kolodynski,J. (1983) pif mutation blocks recombination between mitochondrial rho+ and rho- genomes having tandemly arrayed repeat units in Saccharomyces cerevisiae. *Proc. Natl Acad. Sci. USA*, **80**, 5345–5349.
30. Pinter,S.F., Aubert,S.D. and Zakian,V.A. (2008) The Schizosaccharomyces pombe Pfh1p DNA helicase is essential for the maintenance of nuclear and mitochondrial DNA. *Mol. Cell. Biol.*, **28**, 6594–6608.
31. Zhou,J., Monson,E.K., Teng,S.C., Schulz,V.P. and Zakian,V.A. (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science*, **289**, 771–774.
32. Mateyak,M.K. and Zakian,V.A. (2006) Human PIF helicase is cell cycle regulated and associates with telomerase. *Cell Cycle*, **5**, 2796–2804.
33. Futami,K., Shimamoto,A. and Furuichi,Y. (2007) Mitochondrial and nuclear localization of human Pif1 helicase. *Biol. Pharm. Bull.*, **30**, 1685–1692.
34. Gagou,M.E., Ganesh,A., Thompson,R., Phear,G., Sanders,C. and Meuth,M. (2011) Suppression of apoptosis by PIF1 helicase in human tumor cells. *Cancer Res.*, **71**, 4998–5008.
35. Small,I., Peeters,N., Legeai,F. and Lurin,C. (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics*, **4**, 1581–1590.
36. Nakai,K. and Horton,P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.*, **24**, 34–36.
37. Calvo,S., Jain,M., Xie,X., Sheth,S.A., Chang,B., Goldberger,O.A., Spinazzola,A., Zeviani,M., Carr,S.A. and Mootha,V.K. (2006) Systematic identification of human mitochondrial disease genes through integrative genomics. *Nat. Genet.*, **38**, 576–582.
38. Lemire,B.D., Fankhauser,C., Baker,A. and Schatz,G. (1989) The mitochondrial targeting function of randomly generated peptide sequences correlates with predicted helical amphiphilicity. *J. Biol. Chem.*, **264**, 20206–20215.
39. Leigh-Brown,S., Enriquez,J.A. and Odom,D.T. (2010) Nuclear transcription factors in mammalian mitochondria. *Genome Biol.*, **11**, 215.
40. Lenertz,L.Y., Lee,B.-H., Min,X., Xu,B., Wedin,K., Earnest,S., Goldsmith,E.J. and Cobb,M.H. (2005) Properties of WNK1 and implications for other family members. *J. Biol. Chem.*, **280**, 26653–26658.
41. Chen,W., Chen,Y., Xu,B., Juang,Y.-C., Stippec,S., Zhao,Y. and Cobb,M.H. (2009) Regulation of a third conserved phosphorylation site in SGK1. *J. Biol. Chem.*, **284**, 3453–3460.
42. Simsek,D., Furda,A., Gao,Y., Artus,J., Brunet,E., Hadjantonakis,A.K., Van Houten,B., Shuman,S., McKinnon,P.J. and Jasin,M. (2011) Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. *Nature*, **471**, 245–248.
43. Nakagawa,J., Waldner,H., Meyer-Monard,S., Hofsteenge,J., Jenö,P. and Moroni,C. (1995) AUH, a gene encoding an AU-specific RNA binding protein with intrinsic enoyl-CoA hydratase activity. *Proc. Natl Acad. Sci. USA*, **92**, 2051–2055.
44. Ponamarev,M.V., She,Y.-M., Zhang,L. and Robinson,B.H. Proteomics of bovine mitochondrial RNA-binding proteins: HES1/KNP-I is a new mitochondrial resident protein. *J. Proteome Res.*, **4**, 43–52.
45. Preiss,T., Chrzanowska-Lightowlers,Z.M. and Lightowlers,R.N. (1997) Glutamate dehydrogenase: an organelle-specific mRNA-binding protein. *Trends Biochem. Sci.*, **22**, 290.
46. Hosoda,N., Lejeune,F. and Maquat,L.E. (2006) Evidence that poly(A) binding protein C1 binds nuclear pre-mRNA poly(A) tails. *Mol. Cell. Biol.*, **26**, 3085–3097.
47. Kimura,M., Ishida,K., Kashiwabara,S. and Baba,T. (2009) Characterization of two cytoplasmic poly(A)-binding proteins, PABPC1 and PABPC2, in mouse spermatogenic cells. *Biol. Reprod.*, **80**, 545–554.
48. Sladic,R.T., Lagnado,C.A., Bagley,C.J. and Goodall,G.J. (2004) Human PABP binds AU-rich RNA via RNA-binding domains 3 and 4. *Eur. J. Biochem.*, **271**, 450–457.
49. Nietfeld,W., Mentzel,H. and Pieler,T. (1990) The Xenopus laevis poly(A) binding protein is composed of multiple functionally independent RNA binding domains. *EMBO J.*, **9**, 3699–3705.
50. Gray,N.K., Coller,J.M., Dickson,K.S. and Wickens,M. (2000) Multiple portions of poly(A)-binding protein stimulate translation *in vivo*. *EMBO J.*, **19**, 4723–4733.

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020

51. Sachs,A.B., Davis,R.W. and Kornberg,R.D. (1987) A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol. Cell. Biol.*, **7**, 3268–3276.
52. Kühn,U. and Pieler,T. (1996) Xenopus poly(A) binding protein: functional domains in RNA binding and protein-protein interaction. *J. Mol. Biol.*, **256**, 20–30.
53. Brown,T.A., Tkachuk,A.N. and Clayton,D.A. (2008) Native R-loops persist throughout the mouse mitochondrial DNA genome. *J. Biol. Chem.*, **283**, 36743–36751.
54. Szczesny,R.J., Borowski,L.S., Brzezniak,L.K., Dmochowska,A., Gewartowski,K., Bartnik,E. and Stepien,P.P. (2010) Human mitochondrial RNA turnover caught in flagranti: involvement of hSuv3p helicase in RNA surveillance. *Nucleic Acids Res.*, **38**, 279–298.
55. Blanco,P., Sargent,C.A., Boucher,C.A., Howell,G., Ross,M. and Affara,N.A. (2001) A novel poly(A)-binding protein gene (PABPC5) maps to an X-specific subinterval in the Xq21.3/Yp11.2 homology block of the human sex chromosomes. *Genomics*, **74**, 1–11.
56. Bochman,M.L., Sabouri,N. and Zakian,V.A. (2010) Unwinding the functions of the Pif1 family helicases. *DNA Repair*, **9**, 237–249.
57. Huang,Y., Zhang,D.-H. and Zhou,J.-Q. (2006) Characterization of ATPase activity of recombinant human Pif1. *Acta. Biochim. Biophys. Sin.*, **38**, 335–341.
58. Kalifa,L., Beutner,G., Phadnis,N., Sheu,S. and Sia,E. (2009) Evidence for a role of FEN1 in maintaining mitochondrial DNA integrity. *DNA Repair*, **8**, 1242–1249.
59. Lakshmipathy,U. and Campbell,C. (1999) The human DNA ligase III gene encodes nuclear and mitochondrial proteins. *Mol. Cell. Biol.*, **19**, 3869–3876.
60. Zheng,L., Zhou,M., Guo,Z., Lu,H., Qian,L., Dai,H., Qiu,J., Yakubovskaya,E., Bogenhagen,D.F., Demple,B. *et al.* (2008) Human DNA2 is a mitochondrial nuclease/helicase for efficient processing of DNA replication and repair intermediates. *Mol. Cell*, **32**, 325–336.
61. Mueller,J.C., Andreoli,C., Prokisch,H. and Meitinger,T. (2004) Mechanisms for multiple intracellular localization of human mitochondrial proteins. *Mitochondrion*, **3**, 315–325.

Downloaded from https://academic.oup.com/nar/article-abstract/41/4/2354/2414824 by University of Glasgow user on 10 March 2020