

ACCEPTED MANUSCRIPT



Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals

Fabian Schueren, Thomas Lingner, Rosemol George, Julia Hofhuis, Corinna Dickel, Jutta Gärtner, Sven Thoms

DOI: <http://dx.doi.org/10.7554/eLife.03640>

Cite as: eLife 2014;10.7554/eLife.03640

Received: 9 June 2014

Accepted: 22 September 2014

Published: 23 September 2014

This PDF is the version of the article that was accepted for publication after peer review. Fully formatted HTML, PDF, and XML versions will be made available after technical processing, editing, and proofing.

This article is distributed under the terms of the [Creative Commons Attribution License](#) permitting unrestricted use and redistribution provided that the original author and source are credited.

Stay current on the latest in life science and biomedical research from eLife.

[Sign up for alerts](#) at elife.elifesciences.org

Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals

Fabian Schueren¹†, Thomas Lingner²†, Rosemol George¹†, Julia Hofhuis¹,
Corinna Dickel¹, Jutta Gärtner^{1*}, Sven Thoms^{1*}

¹Department of Pediatrics and Adolescent Medicine, University Medical Center,
Georg-August-University Göttingen, Germany

²Department of Bioinformatics, Institute for Microbiology and Genetics,
Georg-August-University Göttingen, Germany

† Co-first author

* Correspondence: gaertnj@med.uni-goettingen.de (J.G.); sven.thoms@med.uni-goettingen.de
(S.T.)

Competing interest statement: The authors declare no competing interest.

ABSTRACT

Translational readthrough gives rise to low abundance proteins with C-terminal extensions beyond the stop codon. To identify functional translational readthrough, we estimated the readthrough propensity (RTP) of all stop codon contexts of the human genome by a new regression model in silico, identified a nucleotide consensus motif for high RTP by using this model, and analyzed all readthrough extensions in silico with a new predictor for peroxisomal targeting signal type 1 (PTS1). Lactate dehydrogenase B (LDHB) showed the highest combined RTP and PTS1 probability. Experimentally we show that at least 1.6% of the total cellular LDHB getting targeted to the peroxisome by a conserved hidden PTS1. The readthrough-extended lactate dehydrogenase subunit LDHBx can also co-import LDHA, the other LDH subunit into peroxisomes. Peroxisomal LDH is conserved in mammals and likely contributes to redox equivalent regeneration in peroxisomes.

33 INTRODUCTION

34 Translation of genetic information encoded in mRNAs into proteins is carried out by ribosomes.
35 When a stop codon enters the ribosomal A site release factors bind the stop codon, hydrolyze
36 the peptidyl-tRNA bond and trigger the release of the polypeptide from the ribosome. If instead
37 of release factor 1 (eRF1) a near-cognate aminoacyl-tRNA pairs with the stop codon in the
38 ribosomal A site, the stop signal is suppressed. Such decoding of a stop codon as a sense
39 codon is known as translational readthrough. As a consequence, translation ensues to the next
40 stop codon resulting in the synthesis of C-terminally extended proteins (Baranov et al., 2002;
41 Firth and Brierley, 2012; Namy et al., 2004). Mutant tRNAs are classical stop codon
42 suppressors, but also in normal physiology termination occurs with less than 100% efficiency.

43 A number of cis-elements on the mRNA, typically 3' of the stop together with trans-acting factors
44 are known to influence stop codon readthrough (Firth et al., 2011). A case of translational
45 readthrough dependent on RNA cis-elements has recently been found and termed programmed
46 translational readthrough (PTR) (Eswarappa et al., 2014). But it is also known that the stop
47 codon itself and the nucleotides before and after the stop codon affect readthrough. The three
48 stop codons differ in their tendency to be suppressed. In human UAA is least and UGA is most
49 likely to allow readthrough (Baranov et al., 2002; Beier and Grimm, 2001). Studies also show
50 that the nucleotide immediately downstream of the stop codon is biased and can strongly
51 influence readthrough (McCaughan et al., 1995). We here define translational readthrough that
52 is entirely dependent on the stop codon and the nucleotides in its immediate vicinity as basal
53 translational readthrough (BTR). Thus BTR is independent of cis-acting elements and also
54 differs from pharmacologically induced readthrough. Induction of readthrough, most prominently
55 by aminoglycoside antibiotics is an attractive strategy in the treatment of the large number of
56 genetic disorders caused by premature stop codons (Bidou et al., 2012; Keeling et al., 2014).

57 In viruses, readthrough optimizes the coding capacity of compact genomes (Firth and Brierley,
58 2012). In the yeast *S. cerevisiae*, the eukaryotic release factor eRF3 can form prion-like
59 polymers which introduces a level of epigenetic regulation not found in other eukaryotes (Tuite
60 and Cox, 2003). In fungi, translational readthrough extends cytosolic glycolytic enzymes by a
61 cryptic peroxisomal targeting signal (Freitag et al., 2012). In *Drosophila*, readthrough is known to
62 affect between 200 and 300 proteins (Dunn et al., 2013; Jungreis et al., 2011), and in mammals
63 readthrough has been described for more than 50 individual transcripts (Chittum et al., 1998;
64 Dunn et al., 2013; Eswarappa et al., 2014; Geller and Rich, 1980; Loughran et al., 2014;
65 Yamaguchi et al., 2012). Ribosome profiling and phylogenetic approaches provide powerful

66 methods for the systematic identification of readthrough in mammals (Dunn et al., 2013;
67 Eswarappa et al., 2014; Jungreis et al., 2011; Loughran et al., 2014).

68 We wanted to uncover a physiological role of translational readthrough in humans by identifying
69 C-terminal extensions with targeting signals that would create a functional difference between
70 the normal and the readthrough-extended form. To achieve this aim, we concentrated on
71 proteins deriving from BTR. Based on experimental data, we assigned regression coefficients to
72 all possible nucleotides in the stop codon context and, using those regression coefficients,
73 estimated the readthrough propensity (RTP) of all stop codons in the human genome or
74 transcriptome. We were able to formally derive a new nucleotide consensus for high RTP from
75 the regression coefficients of our model. Then we screened all predicted C-terminal extensions
76 for peroxisomal targeting signals because peroxisomes import most of their matrix proteins
77 through a short targeting signal (PTS1) at the very C-terminus (Smith and Aitchison, 2013). We
78 here show that LDHB combines a very high translational readthrough with a hidden, yet
79 functional and evolutionarily conserved PTS1. This peroxisomal isoform of LDH, containing the
80 readthrough-extended LDHBx subunit, is likely to be involved in the regeneration of redox
81 equivalents for peroxisomal β -oxidation.

82

83

84 **RESULTS**

85 **Genome-wide *in silico* analysis of basal translational readthrough**

86 To develop a computational assessment of the RTP of all human stop codon contexts (SCCs)
87 that would allow the identification of genes with high BTR, we focused on SCCs comprising 15
88 nucleotides including and surrounding the stop codon (nucleotides -6 to +9, stop codon at
89 positions 1 to 3). To be able to calculate a linear regression between the SCCs and their
90 experimental BTR values, we formalized SCCs using a binary vector that represented the stop
91 context in a multi-dimensional vector space (Figure 1A and Figure 1 – figure supplement 1). The
92 three stop codons were condensed into one position, so that the binary vector required 51
93 dimensions, for the four possible nucleotides in the six positions before and after the stop codon,
94 and for the three stop codons ($12 \times 4 + 3$). The vector was combined with experimentally
95 accessible BTR frequencies. For the first approximation model (LIN) we used 66 sequences
96 derived from human nonsense mutations (Floquet et al., 2012). The nucleotide sequences of
97 these stop contexts show no bias with respect to RTP, because the contexts and the stop
98 codons evolved independently, therefore the context nucleotides are random in relation to the
99 stop codon. We calculated a linear regression model for these SCCs and used only the
100 experimental BTR values that had been measured in the absence of aminoglycosides. The

101 model assigns regression coefficients to all possible nucleotides in the stop context (Figure 1 –
102 figure supplement 1).

103 For a first round of whole-genome RTP prediction, we extracted the SCCs for each transcript
104 from the Ensembl database and calculated RTP by adding up the regression coefficients of all
105 relevant positions. An outline of this algorithm is shown in Figure 1A and in more detail in Figure
106 1 – figure supplement 1. A sortable list of LIN RTP values for all human transcripts is contained
107 in Dataset 1 (Schueren et al., 2014).

108 To expand the data basis of the RTP algorithm and to obtain evidence that the algorithm indeed
109 predicts BTR values, we selected candidate transcripts with high, intermediate, and low RTP
110 and tested them by a dual reporter assay (Figure 1B and Table 1). For experimental analysis,
111 SCCs spanning 10 nucleotides upstream and downstream of the stop codon were expressed
112 with a 5'/N-terminal yellow fluorescent protein (Venus) and a 3'/C-terminal humanized Renilla
113 luciferase (hRluc) tag. Stop suppression leads to the expression of hRluc, Venus served as an
114 internal expression control. Readthrough is expressed as luciferase activity per Venus
115 fluorescence. This approach excludes introns and exon junction complexes and, due to the
116 relatively short stretch of variable nucleotides between the reporters, also does not allow for
117 extensive RNA structures that could modulate readthrough. In consequence, this form of the
118 dual reporter assay focusses on the assessment of BTR without influence by specific cis-
119 elements. The additional candidates tested showed BTR between 0.10 (± 0.006) and 2.91
120 (± 0.15) % relative to the 100% readthrough control expressing the Venus-hRluc fusion protein
121 without intervening stop codon region (Figure 1C and Table 1). Induction of readthrough by the
122 aminoglycoside antibiotic geneticin (G418) was between 3.25 (± 0.41) and 40.38 (± 5.33) -fold
123 (Figure 1C). The Geneticin could only increase the luciferase-per-Venus signal when a stop
124 codon separated Venus and luciferase, indicating that our dual reporter assay faithfully reports
125 readthrough. The finding that experimental readthrough could be increased by treatment with
126 aminoglycosides also excludes alternative mechanisms such as RNA editing or splicing that
127 could be considered to explain the relative increase of the luciferase over the Venus signal. The
128 highest levels of induction can only be reached when basal readthrough is low, and, vice versa,
129 a high BTR somewhat limits the maximum induction factor (Figure 1C), suggesting that maximal
130 BTR readthrough is limited to levels below 15%.

131 Next we added our candidate sequences and their experimentally determined readthrough
132 levels to obtain an iterative and extended RTP model (LINiter). Again, we applied this model to
133 all human transcripts (Dataset 1 (Schueren et al., 2014), model parameters in Table 2). We
134 measured the correlation of RTP and experimental BTR by leave-one-out cross-validation during
135 computation of the regression coefficients. For the LINiter model, we obtained a weak but
136 significant Pearson correlation coefficient of 0.34 ($p = 0.002$) (Figure 1 – figure supplement 2).

137 To understand the origin of the apparently strong non-linear contribution to RTP, we analyzed
138 the regression coefficients of the LINiter model. Nucleotide positions associated with coefficients
139 of large absolute value contribute most to the RTP. The relative contribution of nucleotides within
140 the stop codon context to the readthrough prediction is shown in Figure 2A.

141

142 **A consensus for high basal translational readthrough in humans**

143 The sequence-logo representation of regression factors in Figure 2A indicates that the three or
144 four nucleotides following the stop codon contribute to readthrough. The quantitative manner in
145 which we derived LINiter values allowed us to rationally derive a nucleotide motif permitting high
146 readthrough in humans. We identified the nucleotide positions with the strongest influence on
147 BTR in humans by feature selection, that is by successively eliminating those positions that
148 contribute least to the prediction (Figure 2B). One by one the nucleotide positions with the
149 smallest sum of squared regression coefficients were removed from the model. We find that two
150 reduced models improve the prediction. Models with either five or three relevant context
151 positions in addition to the stop codon correspond to the local and global residual error
152 minimum, respectively. LINfs5 comprises nucleotide positions -6, the stop codon, and positions
153 +4 to +7, and LINfs3 comprises only the stop codon and positions +4 to +6, that is the codon
154 following the stop (Figure 2B). The results of this analysis indicate that in humans the stop codon
155 and the three nucleotides immediately downstream of the stop codon have the largest influence
156 on BTR (LINfs3). The corresponding consensus is UGA CUA (stop codon underlined). Possibly
157 also the nucleotides at positions +7 (the fourth position after the stop) and -6 contribute to BTR.
158 The RTP-BTR correlation associated with LINfs3 was 0.41 ($p = 0.0001$) (Figure 2 – supplement
159 1). To test if the LINfs3 consensus confers indeed high BTR, we analyzed four additional
160 candidate SSCs. Three high-RTP were derived from AQP4, SYTL2, and CACNA2D4, and
161 DHX38 was used as a control with a low RTP. AQP4, SYTL2, and CACNA2D4 conform with the
162 LINfs3 consensus, whereas DHX38 does not. AQP4, SYTL2, and CACNA2D4 showed 2.29
163 (± 0.09), 0.99 (± 0.06), and 0.61 (± 0.02) % readthrough in HeLa cells, whereas for DHX38
164 readthrough was only 0.27 (± 0.04) % (Figure 2C), confirming that LINfs3 SCC indeed allows a
165 very high rate of stop suppression. Next we wanted to test if these conclusions obtained in HeLa
166 can be extended to other cell types. We therefore performed dual reporter experiments using the
167 HT1080 fibrosarcoma cell line, the human embryonic kidney cell line (HEK), and the U373 cell
168 line. In all these experiments the relative distribution of BTR values remained the same, with
169 AQP4 showing the highest and DHX38 the lowest BTR (Figure 2C). The finding that readthrough
170 is lower in CACNA2D4 than in AQP4 and SYTL2 can also be taken as evidence for a
171 contribution of the SCC position +7 (fourth after the stop). Taken together, these experiments

172 show that BTR is indeed a property of the respective SCC, and that readthrough may be
173 differently regulated in different tissues.

174 The linear approximation underlying the LINiter and the LINfs3 model led to the identification of
175 the UGA CUA (LINfs3) consensus conferring high BTR. A partially overlapping set of genes with
176 this consensus was recently tested (Loughran et al., 2014). An overview of all experimentally
177 confirmed cases of translational readthrough in Figure 2 – supplement 2 shows that ribosome
178 profiling, phylogenetic approaches, and RTP screening are complementary approaches. For
179 example, only one of the 42 readthrough genes found by ribosome profiling in foreskin
180 fibroblasts (Dunn et al., 2013) contains the UGA CUA consensus. The largely varying levels and
181 sequence requirements for efficient stop codon suppression suggest that multiple molecular
182 mechanisms can be the cause of readthrough in mammals.

183

184 **Identification of peroxisomal targeting signals in readthrough extensions**

185 The genome-wide *in silico* analysis of RTP provides the basis for the identification of
186 physiological functions of a readthrough protein. We have therefore screened the extensions for
187 possible elements that could confer functional differences between the normal and the extended
188 form of the protein. We screened the extensions for possible transmembrane domains (Krogh et
189 al., 2001), for prenylation sites (Zhang and Casey, 1996), for endoplasmic retention signals
190 (Stornaiuolo et al., 2003; Zerangue et al., 2001), and for glycosylation sites (Schwarz and Aebi,
191 2011; Zielinska et al., 2010).

192 To identify genes with a high BTR and a readthrough extension conferring a biological function,
193 we decided to focus on the detection of proteins carrying a hidden peroxisomal targeting signal
194 type 1 (PTS1) in the extension. This targeting mechanism had been shown to divert a small
195 fraction of cytosolic glycolytic proteins to peroxisomes in fungi (Freitag et al., 2012). PTS1 cover
196 more than 90% of the targeting motifs of peroxisomal matrix proteins. The alternative PTS2 is
197 found in only very few matrix proteins, and has even been lost in some organisms (Lanyon-Hogg
198 et al., 2010). PTS1 is localized at the very C-terminus of a substrate protein. However, the
199 quintessential PTS1, Ser-Lys-Leu (SKL), is neither necessary nor sufficient to support matrix
200 protein import into peroxisomes. Variations exist, and amino acids upstream of the terminal
201 tripeptide also contribute to targeting (Brocard and Hartig, 2006). Moreover, PTS1 do not confer
202 a binary decision (to import or not to import), but they are likely to determine an equilibrium
203 between cytosolic and peroxisomal localization. This is best exemplified by the peroxisomal
204 marker protein catalase, a considerable amount of which is not imported into peroxisomes due
205 to an inherently weak PTS1 which is associated with a low affinity to the cytosolic PTS1-receptor
206 PEX5 (Maynard et al., 2004). We took advantage of these scalable properties of PTS1 and
207 adapted to human PTS1 a prediction algorithm that we had previously developed for plants

208 (Lingner et al., 2011). This machine learning-based method has been shown to accurately
209 predict proteins with canonical and non-canonical PTS1 peptides and provides evidence for
210 peroxisome targeting in terms of a posterior probability (Lingner et al., 2011).

211 To program the human PTS1 prediction algorithm we conducted orthology searches on 24
212 known human PTS1 sequences in metazoa using BLAST against protein and EST databases.
213 The resulting dataset and several thousand metazoan sequences without peroxisomal
214 association were used as positive and negative examples in a discriminative machine learning
215 setup. Here, the sequences were represented by binary vectors encoding the presence or
216 absence of up to 15 C-terminal amino acids. Models were trained and validated using
217 regularized least squares classifiers (RLSC) and 5-fold cross-validation. A more detailed
218 description of the human PTS1 scoring can be found in the Materials and Methods section. We
219 calculated PTS1 posterior probabilities of all predicted C-terminal readthrough extensions
220 derived from the human transcriptome (Dataset 1 (Schueren et al., 2014)).

221

222 **LDHB is extended by translational readthrough**

223 Based on the assumption that a protein is more likely to target to peroxisomes by a cryptic PTS1
224 when the RTP and the extension's PTS1 scores are high, we used the product of RTP LINiter
225 scores and PTS1 posterior probabilities as a predictor of functional peroxisomal targeting by a
226 hidden PTS1 in the extension (Dataset 1 (Schueren et al., 2014)). To avoid negative product
227 scores, we scaled RTP between 0 and 1 before multiplication (now designated RTP⁺).

228 We identified LDHB, one of the two human lactate dehydrogenase (LDH) subunits, at the top
229 rank (position 1 of 42069 entries) of our sorted list of combined RTP⁺ and PTS1 scores (Dataset
230 1 (Schueren et al., 2014)). The distribution of RTP⁺ * PTS1 product scores over all human
231 transcripts indicates that other candidates must have considerably lower RTPs and/or targeting
232 efficiencies, because the score drops by 50% over the first 40 of 42069 transcripts (Figure 3A).

233 To experimentally confirm high BTR we expressed the human LDHB SCC in the Venus/hRluc
234 dual reporter assay. Readthrough was 1.55 (±0.09) % and mutation of the stop codon and/or the
235 consecutive nucleotide strongly suppressed readthrough (Figures 3B and Figure 3 – figure
236 supplement 1). Treatment with geneticin increased readthrough to 4.38 (±0.42) % (compare with
237 induction factors in Figure 3C).

238 To establish that the full-length protein is extended by stop suppression, LDHB including the
239 extension (designated LDHBx for “extended”) and mutants were expressed with N-terminal HA-
240 and C-terminal myc-tags and analyzed by Western blotting. Full-length LDHB showed
241 aminoglycoside-inducible readthrough, and the loss of readthrough upon exchange of the stop
242 codon or the nucleotide following the stop confirms the special function of the LDHB stop codon
243 context in stimulating translational readthrough (Figure 3D).

244

245 **Peroxisomal localization of LDHB depends on translational readthrough**

246 The identification of LDHB as virtually the only human protein with a high combined readthrough
247 and peroxisomal targeting probability is surprising, because a peroxisomal readthrough-
248 extended LDHBx entails at least one new LDH isoform. On the other hand, LDH activity and
249 isoforms inside peroxisomes were known for more than four decades (Baumgart et al., 1996;
250 Gronemeyer et al., 2013; McClelland et al., 2003; McGroarty et al., 1974; Osmundsen, 1982;
251 Völkl and Fahimi, 1985). In the apparent absence of known targeting signals, however, it has
252 been inexplicable how the protein can enter the peroxisome. Therefore we analyzed if the
253 extended human LDHBx protein and the predicted PTS1 therein leads to peroxisomal
254 localization. We expressed LDHBx as a fusion protein with an N-terminal enhanced yellow
255 fluorescent protein (YFP) and co-labelled cells by immunofluorescence with the peroxisomal
256 marker PEX14, a peroxisomal membrane protein. YFP-LDHB showed the expected cytosolic
257 localization (Figures 4A and 5A). We hypothesized that a large excess of cytosolic YFP-LDHB
258 masks the peroxisomal localization. To remove cytosolic YFP-LDHB, we permeabilized cells by
259 digitonin before fixation and washed out the cytosol using phosphate buffered saline (PBS). In
260 agreement with peroxisomal targeting through the cryptic PTS1, LDHBx is found localized in
261 peroxisomes after removal of the cytosol (Figures 4B and 5B). In control experiments, we show
262 complete removal of cytosolically expressed YFP by cytosol wash-out (Figure 5 - figure
263 supplement 1) and peroxisomal localization of a YFP variant fused to PTS1 of the peroxisomal
264 matrix protein ACOX3 (Figure 5 - figure supplement 2). To confirm that LDHB targeting to
265 peroxisomes is dependent on the putative PTS1 in the readthrough extension, we changed the
266 SRL terminus (PTS1 probability 94.3%) to SSI (0.002%) and to SR (Δ L, 0.00001%). These
267 mutations blocked YFP-LDHBx targeting to the peroxisome (Figures 4C and 5C-F). Remarkably,
268 exchange of the leaky UGA stop with the tighter UAA reduced peroxisomal localization of YFP-
269 LDHB (Figures 6A,B). Our results show that the high-RTP stop codon contexts as well as the
270 PTS1 in the extension after the stop codon are needed for peroxisome targeting. The extension
271 must be accessible to ribosomal translation and contain a functional PTS1. It is known that
272 PTS1-dependent targeting guides proteins into peroxisomes, and not only to the membrane. The
273 dependence of LDH targeting on the hidden PTS1 and on the nature of the stop codon thus
274 confirm that the protein is indeed inside the peroxisome. As expected, replacing the stop codon
275 by tryptophan-encoding UGG renders LDHBx entirely dependent on the PTS1 (Figures 6C,D).

276 To obtain more direct evidence for the readthrough-dependent low abundance targeting of
277 human LDHB to peroxisomes we analyzed untransfected wild-type cells by immunofluorescence
278 with anti-LDHB and anti-PEX14 antibodies. LDHB appears distributed in the cytosol (Figure 4D).
279 After cytosol depletion, however, the remaining LDHB signal is mainly peroxisomal (Figure 4E).

280 A small portion of LDHB may localize to other cellular locations protected against cytosol
281 removal. We confirmed these results in human skin fibroblasts, COS-7 cells (monkey kidney
282 fibroblast line), the human glioblastoma cell line U118, and freshly prepared rat cardiomyocytes
283 (Figure 7). Our data are in agreement with readthrough-dependent targeting of about 1.6% of
284 the LDHB to peroxisomes mediated by the cryptic PTS1 in the extension. Remarkably, treatment
285 of untransfected wild-type HeLa cells with geneticin increased LDHBx levels in the peroxisome
286 (induction factor 1.89, n=28, t-test $p < 0.0001$) suggesting elevated peroxisomal LDHBx levels as
287 a general pharmacological consequence of aminoglycoside treatment.

288 Next we wanted to test if there is evidence for differential regulation of translational readthrough
289 of LDHB in different cell types. We expressed LDHB and mutant dual reporter constructs in
290 COS-7 cells, U118 cells, and in HEK cells. Readthrough of LDHB ranged between 1.55 (± 0.09)
291 % in HEK and HeLa and 1.88 (± 0.14) % in COS-7. Surprisingly, in U118 cells LDHB readthrough
292 is increased to 5.09 (± 1.03) % (Figure 8). Geneticin induction factors ranged between 1.32
293 (± 0.09) and 2.82 (± 0.27) (Figure 8). LDHB stop suppression is thus not restricted to special
294 tissues, and may be differently regulated in different cell types.

295 Analysis of animal LDHB orthologs in vertebrates shows that the PTS1 in the extension is
296 exclusively and strictly conserved in mammals supporting the notion of a functional extension in
297 these proteins and an evolutionarily conserved targeting of LDHBx to peroxisomes in mammals
298 (Figure 9).

299

300 **Piggy-back co-import of LDHA with LDHB**

301 LDHB together with lactate dehydrogenase A (LDHA) can make up five tetrameric LDH
302 isoforms, two of which are homotetramers and three are heterotetramers (Boyer et al., 1963;
303 Markert, 1963) and peroxisomes have the unusual ability to import folded and even oligomeric
304 proteins (Lanyon-Hogg et al., 2010; McNew and Goodman, 1996). We therefore wanted to test if
305 peroxisomal LDHBx piggy-backs LDHA into peroxisomes. For this purpose we adapted a two-
306 hybrid assay previously used to analyze co-import of subunits of the dimeric peroxisomal
307 hydrolase Lpx1 in a heterologous system (Thoms et al., 2011). When LDHA was expressed as a
308 fusion protein with N-terminal YFP without co-expression of any form of LDHB the protein
309 localized to the cytosol as expected (Figure 10A). However when we co-expressed YFP-LDHA
310 with CFP-LDHBx[TGG], i.e. cyan fluorescent protein (CFP) fused to the readthrough form of
311 LDHB, we found YFP-LDHA in peroxisomes (Figure 10B). This experiment shows that the
312 readthrough form of LDHB, LDHBx can interact with LDHA, and that LDHBx is capable of
313 carrying LDHA into the peroxisome. To show that co-import of LDHA is dependent on the hidden
314 targeting signal in LDHBx, we mutated the targeting signal to SSI, or we deleted the terminal
315 leucine. Either LDHBx PTS1 mutation blocked co-import of LDHA (Figure 10 – figure

316 supplement 1). The peroxisome is thus accessible to all four new LDH isoforms containing
317 LDHBx. To support our data on LDHBx-LDHA co-import we drew a structural model of LDH-H,
318 the heart all-B isoform of LDH (Figure 10 – figure supplement 2). The C-terminal amino acid
319 leucine is extended by three amino acids not resolved in the structure, and, in LDHBx, by
320 additional seven amino acids. The model shows that this extension is protruding from the
321 tetramer and is located distal to the protomer-interaction site, confirming that oligomerization is
322 not hampered by the extension. The protruding LDHBx extension carrying the PTS1 is also
323 accessible on the tetramer surface for PEX5 binding and import into the peroxisome.

324

325 **DISCUSSION**

326 The study of translational readthrough goes back to the origins of molecular biology, but only
327 recently mammalian genes undergoing readthrough have come into focus and are being
328 identified by systemic approaches (Dunn et al., 2013; Eswarappa et al., 2014; Jungreis et al.,
329 2011; Loughran et al., 2014). Translational readthrough can be controlled by cis-acting
330 elements, RNA-structures on the transcript that, often mediated by trans-factors, influence the
331 termination process (Eswarappa et al., 2014; Firth et al., 2011). This mechanism has been
332 termed programmed translational readthrough (PTR) (Eswarappa et al., 2014). It is known,
333 however, that also the stop codon together with the preceding and immediately following
334 nucleotides (stop codon context, SCC) influence translational readthrough. We have termed this
335 process basal translational readthrough (BTR) to distinguish it, on the one hand, from PTR in
336 general, but also from pharmacologically induced readthrough. In this study we derive a motif
337 conferring high BTR from a linear regression model of SCCs and show that LDHBx undergoes
338 BTR, which in turn affects the intracellular distribution of LDH.

339

340 **A new lactate dehydrogenase subunit**

341 LDH is not only an enzyme with several isoforms, but it has been instrumental in devising the
342 enzyme isoform concept *per se*. The identification of the classical Muscle and Heart subunits
343 LDH-M (LDHA) and LDH-H (LDHB), in the late 1950s was followed by the identification of a
344 testes-specific LDHA-variant, LDHC (Boyer et al., 1963; Goldberg et al., 2009). Now we find that
345 readthrough-extended LDHBx is encoded by the well-known *LDHB* gene by translational stop
346 suppression and can give rise to new isoforms. Peroxisomal LDH is a novel isoform of LDH
347 containing at least one readthrough-extended LDHBx subunit. LDHB readthrough and
348 readthrough-dependent peroxisomal localization exist in various human cell types, suggesting
349 that the LDHBx subunit is expressed and localized to peroxisomes in all tissues that express

350 LDHB. LDHBx exemplifies a new mechanism of posttranscriptional diversification of the
351 genome's coding potential in mammals.

352 The 1.6% LDHBx stop codon readthrough that we find in our experiments correspond to 1.5 to
353 2% LDH activity found in association with peroxisomes (Baumgart et al., 1996; McGroarty et al.,
354 1974; Osmundsen, 1982), suggesting that cellular suppression of the stop codon is the only
355 pathway for LDHB into peroxisomes. Assuming that peroxisomes fill approximately 1-2% of the
356 cell volume, translational readthrough ensures almost equal concentration of LDH in cytosol and
357 in peroxisomes.

358

359 **A role for peroxisomal lactate dehydrogenase**

360 Fatty acid β -oxidation reactions are the hallmark of peroxisomes in most cell types and
361 organisms. In mammalian peroxisomes β -oxidation is involved in the degradation of very-long
362 chain fatty acids (VLCFA) and biogenetic reactions such as the synthesis of bile acids (Lodhi
363 and Semenkovich, 2014). Therefore patients with peroxisomal disorders accumulate VLCFA and
364 bile acid intermediates (Braverman et al., 2013). During fatty acid oxidation and other
365 peroxisomal processes nicotinamide adenine dinucleotide (NAD^+) is reduced to NADH.
366 However, the pathway of NAD^+ regeneration inside peroxisomes is not clear (Kunze and Hartig,
367 2013). For an efficient β -oxidation to occur it is necessary that a redox shuttle system exists for
368 NAD^+ regeneration, because peroxisomes are impermeable to NAD/NADH (Visser et al., 2007).
369 The identification of LDH inside the peroxisome suggested the existence of a lactate/pyruvate
370 shuttle involved in the regeneration of redox equivalents (Baumgart et al., 1996; Gladden, 2004;
371 McClelland et al., 2003). In the absence of a peroxisomal targeting signal, however, peroxisomal
372 LDH was not universally accepted.

373 Lactate/pyruvate shuttling could either occur directly through the peroxisomal membrane (Visser
374 et al., 2007) or make use of monocarboxylate shuttles in the peroxisomal membrane
375 (McClelland et al., 2003). Generally, functional LDHBx targeting to peroxisomes highlights the
376 role of intracellular lactate shuttle mechanisms (Brooks, 2009). In liver peroxisomes pyruvate
377 production is catalyzed by alanine-glyoxylate aminotransferase, an important enzyme in
378 glyoxylate detoxification. Glyoxylate, however, is itself a substrate of LDH (Salido et al., 2012).
379 Therefore peroxisomal LDH may also be involved in peroxisomal glyoxylate metabolism.

380 Peroxisomal LDH is not the first glycolytic enzyme found in peroxisomes. Trypanosomes have
381 sequestered the full set of glycolytic enzymes in specialized peroxisomes called glycosomes
382 (Gualdrón-López et al., 2012). And recently, in fungi, part of the glycolytic pathway upstream of
383 pyruvate including glyceraldehyde-3-phosphate dehydrogenase and 3-phosphoglycerate kinase,
384 was shown to be localized to peroxisomes by alternative splicing and/or translational

385 readthrough (Freitag et al., 2012). It is compelling that fungi as well as mammals use stop codon
386 suppression to localize a small fraction of glycolytic enzymes to peroxisomes. We hypothesize
387 that both, translational readthrough as well as PTS1 are easy to evolve, and so can divert a low
388 and steady amount of these enzymes to peroxisomes.

389 A small fraction of cytosolic LDHB is imported into peroxisomes. This fraction is likely to be
390 constant with respect to the overall LDHB expression levels in given tissue. We speculate that
391 the peroxisomal LDHB shunt helps coordinating redox processes between the cytosol and the
392 peroxisome. Importantly, our study reveals a new pharmacological effect of readthrough-
393 inducing drug such as the commonly prescribed aminoglycosides, as they will increase LDHB
394 readthrough and peroxisome import of LDHBx.

395 It is not known at the moment, whether translational readthrough is regulated in humans. The
396 very high readthrough of approximately 5% in a glioblastoma cell line suggests that readthrough
397 is differentially regulated in different tissues. Future experiments will show if the increased LDHB
398 readthrough we find in this cell line are a cancer-associated dysregulation linked to the Warburg
399 effect (Hsu and Sabatini, 2008), or if it just matches a higher abundance of peroxisomes in these
400 cells to ensure an equal concentration of LDH in cytosol and peroxisomes in these cells as
401 suggested above. It is also possible that glial cells generally have a higher demand of
402 peroxisomal LDH that could be involved in neuronal/glial lactate metabolism.

403

404 **A rational approach to translational readthrough**

405 The first mammalian readthrough proteins were identified by serendipity (Chittum et al., 1998;
406 Geller and Rich, 1980; Yamaguchi et al., 2012). Recently, two powerful and complementary
407 methods have been employed in the genome-wide identification of readthrough-extended
408 proteins. Ribosome profiling can identify translating ribosomes in 3'UTRs and thereby discover
409 readthrough and other recoding events outside of known coding regions (Dunn et al., 2013;
410 Ingolia et al., 2011). Phylogenetic approaches like those implemented in Phylo-CSF (Lin et al.,
411 2011) evaluate the coding potential of sequences before and after the stop codon to derive
412 prediction of readthrough and are particularly powerful when genome sequences from closely
413 related species are available (Jungreis et al., 2011; Loughran et al., 2014). Ribosome profiling,
414 however, depends on gene expression, and can identify readthrough events only when the cell
415 type in question is actually analyzed. Ribosome profiling may also fail to identify short
416 readthrough extensions. Phylogenetic approaches, on the other hand, may miss readthrough
417 when it is not conserved in the given dataset or when sufficiently dense datasets are not
418 available, or when the extensions are too short to provide a basis for phylogenetic comparison.

419 Our approach to systems-level identification of translational readthrough is based on the
420 formalization of SCCs and a linear regression model with experimental readthrough values. The

421 majority of the input sequences has been derived from patient nonsense-mutations. In
422 consequence, these sequences are neither biased by preselection by any pre-determined RTP
423 or experimental readthrough levels, nor by the stop codon contexts, because the contexts did
424 not evolve together with the respective stop codons. The algorithm we develop in this paper is
425 limited to the plus/minus six nucleotide positions before and after the stop codon. This approach
426 excludes the identification of extended RNA secondary structures involved in PTR and other
427 recoding events (Baranov et al., 2002; Eswarappa et al., 2014; Firth and Brierley, 2012;
428 Loughran et al., 2014). The identification of the LINfs3 consensus and the human genes
429 associated with this consensus justifies this approach. The LINfs3 motif, derived by feature
430 selection, encompasses the stop codon and the first codon after the stop: UGA CUA. Our
431 analysis suggests that also position +7 and -6 might further contribute to readthrough. We have
432 tested five of the 144 candidates in the genome with the UGA CUA motif and confirmed their
433 high BTR. Highest BTR appears to correlate with a G in position +7 (UGA CUA G) within the
434 LINfs5 consensus. This motif is found 30 times in the human genome and has recently been
435 shown to support high translational readthrough (Loughran et al., 2014). The motifs for high BTR
436 are distinct from the consensus UGA CAR YYA (R=A/G, Y= C/U) found in some viruses and
437 yeast (Harrell et al., 2002; Namy et al., 2001) but resembles the alphavirus-like high readthrough
438 stop context (Li and Rice, 1993). Interestingly, the same stop suppression context in the *LAMA3*
439 gene has been shown to alleviate disease severity of an otherwise fatal nonsense mutation in a
440 patient with junctional epidermolysis bullosa, the major and most devastating form of
441 epidermolysis bullosa (Pacho et al., 2011).

442 The existence of the consensus motif UGA CUA is the origin of the non-linear contribution to
443 RTP in our models. This is supported by the finding that correlation of BTR and RTP for LINfs3
444 is higher than for the LINiter model so that the reduced number of parameters in LINfs3 provides
445 a better model fit. This finding implies that with the currently small dataset, compact linear
446 models should be preferred over non-linear models with many parameters. The identification of
447 the few relevant nucleotide positions will help to create datasets with fully specified BTR for a
448 wide range of SCCs and cell types. A larger training set of sequences with verified readthrough
449 rates will allow the development of non-linear approximation models.

450 LDHBx show an unusually high readthrough of 1.6%, and its stop context UGA CUAG (stop
451 codon underlined) matches the LINfs3 consensus. The 18 nucleotide extension in LDHBx is
452 unlikely to contain extensive secondary structure that would suggest a combined effect of BTR
453 and PTR. The identification of LDHBx and the recently discovered readthrough-form of vascular
454 endothelial growth factor A, VEGF-Ax (Eswarappa et al., 2014) thus mark two extreme and
455 separable cases of physiological stop suppression: LDHBx appears independent of cis-factors
456 beyond the SCC and marks a prototypical example of BTR. In contrast, the readthrough of
457 VEGF-Ax is relatively independent of its SCC but instead requires a more distantly located cis-

458 element (Eswarappa et al., 2014). The distinction between PTR and BTR, however, is not
459 exclusive. A thorough analysis of readthrough in OPRK1 and OPRL1 indicates that readthrough
460 levels of more than 30% can be obtained by a combination of cis-elements and UGA CUA-
461 based BTR (Loughran et al., 2014).

462 The era of systematic analysis of translational readthrough in humans is only beginning. We
463 expect that a combination of *in silico* modelling and screening, ribosome profiling, phylogenetic
464 methods, and mass spectrometry will help identifying the “extensome”, the complete set of
465 readthrough-extended proteins in mammals.

466

467

468 **MATERIALS AND METHODS**

469 **Readthrough propensity (RTP) calculation algorithm**

470 To predict the readthrough propensity (RTP) of gene transcripts, we developed a linear
471 regression model based on the stop codon contexts and their experimentally determined basal
472 readthrough values. The stop codon context comprises the stop codon itself (positions +1 to +3)
473 and the nucleotide sequences surrounding the stop codon (-6 to +9). For the first-pass model
474 (LIN) we re-analyzed 66 stop codon contexts with known experimental basal readthrough values
475 (Floquet et al., 2012). The stop codons evolved independently of their contexts (Table 3).
476 Nucleotide sequences were represented by indicator vector coding. Here, 12*4 binary vector
477 entries are used to indicate the presence [1] or absence [0] of a nucleotide (A, C, G, or U) at a
478 particular position (-6 to -1, +4 to +9) surrounding the stop codon. Three further entries are
479 reserved to indicate the type of stop codon (UAA, UAG, or UGA, positions +1, +2, +3). The
480 resulting feature vectors of all sequences were normalized to euclidean unit length.

481 For the estimation of the regression model coefficients we performed a regularized least-squares
482 (“ridge”) regression (Hoerl and Kennard, 1970). Let \mathbf{X} be the $n \times d$ matrix of n sequence feature
483 vectors with dimensionality d and \mathbf{y} be the (n -dimensional) vector of readthrough values
484 associated with the sequences. Then the weight vector $\mathbf{w}=(\mathbf{X}^T\mathbf{X}+k*\mathbf{I})^{-1}*\mathbf{X}^T\mathbf{y}$ represents the
485 solution of the linear least-squares problem and $y=\mathbf{w}^T\mathbf{x}$ corresponds to the readthrough
486 propensity value y for a sequence feature vector \mathbf{x} . To evaluate the influence of the
487 regularization parameter k , we performed a leave-one-out cross-validation (loo-cv) with $k=\{10^j|j=-$
488 $3,-2.7,\dots,0,\dots,3\}$ for all model types. The minimum loo-cv error in terms of the sum of squared
489 deviations of predictions from known readthrough values was $4.75*10^{-7}$ for $k=10^{0.3}$ (ca. 1.995).

490 For genome-wide prediction of readthrough propensities for human transcripts we downloaded
491 all 215,621 coding sequences from the Ensembl BioMart (Flicek et al., 2012) using the Homo
492 sapiens Genes v74 section (GRCh37.p13) plus 300 nucleotides downstream the CDS end

493 (ensembl.org, November 2013). Transcripts corresponding to identical protein products, short
494 sequences (<15aa protein-coding) and incomplete (e.g. missing or mislocated stop codon) or
495 insufficiently sequenced (i.e. undetermined nucleotides) DNA were removed. Sequences with
496 identical 3'/C-termini (nucleotide positions -45 to +303) were aggregated to one representative
497 sequence, resulting in 42,069 unique transcripts. ORF extensions were identified by detection of
498 an in-frame stop codon within 300 nucleotides downstream the annotated stop codon.

499

500 **Iterative model refinement and feature selection**

501 To obtain a more comprehensive model for RTP prediction, we included 15 sequences and their
502 corresponding experimentally determined readthrough values from this study in the prediction
503 model (Dataset 1 (Schueren et al., 2014)). The regression coefficients for the iterative model
504 considering all twelve stop context positions (LINiter) were computed as described in the
505 previous section. The minimum regression error was 6.24×10^{-6} at $k=10^{0.3}$. A sequence logo
506 representation of the regression coefficients for this model is displayed in Figure 2A. The
507 sequence logo was created using the enoLOGOS web server (Workman et al., 2005).

508 Furthermore, we performed an evaluation of reduced model sizes by stepwise elimination of
509 context positions carrying no or little information for RTP prediction (feature selection). Starting
510 from the complete mode (LIN), we removed the position corresponding to the minimum sum of
511 squared regression coefficients. Regression error and coefficients were then calculated for the
512 remaining positions (including the stop codon) as described above. This procedure was
513 repeated until only the stop codon position was left. Figure 2B shows the development of the
514 regression error for subsequently reduced model sizes by eliminated positions. Here, a first local
515 minimum can be identified for the model LINfs5 with five positions remaining (-6, stop, +4 to +7)
516 and the global minimum corresponds to the model LINfs3 with three positions besides the stop
517 codon (stop, +4 to +6).

518

519 **PTS1 prediction algorithm**

520 To identify cryptic peroxisomal localization signals in readthrough extensions, we adapted a
521 peroxisomal targeting signal type 1 (PTS1) detection algorithm that was previously developed for
522 plant proteins (Lingner et al., 2011). For this purpose, we used 24 known human PTS1 proteins
523 (ACOT4, ACOX1, ACOX2, ACOX3, AGXT, AMACR, BAAT, CRAT, DAO, EHHADH, GNPAT,
524 HAO1, HAO2, HSD17B4, IDE, MLYCD, PRDX5, ACOT8, CROT, PECL, ECH1, LONP2, PECR,
525 PIPOX) and performed orthology searches to metazoan protein and EST sequences using a
526 bidirectional best BLAST hit strategy. Starting from each human protein sequence, we identified
527 significant BLAST hits ($e\text{-value} < 10^{-10}$) to metazoan sequences within the “nr” and “dbEST”
528 database. Then, the best hit of each organism was searched against the human proteome and

529 sequences not re-identifying the starting sequence were removed. Afterwards, the starting
530 sequences and putative orthologs were pooled and sequences with uncommon PTS1
531 tripeptides, i.e. tripeptides which occurred less than 3 times, were removed from the set. The
532 resulting set of sequences was used as positive examples for training of machine learning
533 models as previously published (Lingner et al., 2011). Briefly, a regularized least-square classifi-
534 cation algorithm was trained using indicator vector representations of up to 15 C-terminal amino
535 acids of positive and negative example sequences. A set of negative example sequences was
536 created by extracting all metazoan sequences without peroxisomal association from the Swiss-
537 Prot section of UniProt (<http://www.uniprot.org>) in November 2011. The best model (15 C-
538 terminal amino acids) was determined by 5-fold cross-validation and yielded a prediction
539 accuracy of 0.996 and 0.863 in terms of the area under receiver operating characteristic (ROC)
540 curve (auROC) and the area under the precision/recall curve (auPRC), respectively. When stop
541 was considered in the PTS1 prediction, the stop codon was scored as an undefined amino acid
542 ('X') without contribution to the PTS1 posterior probability.

543

544 **Multiple alignment analysis**

545 The multiple alignment of genomic sequences for the LDHB stop codon context (position -36 to
546 +48) was downloaded from the Ensembl database (www.ensembl.org) in November 2013. The
547 '21 amniota vertebrates' alignment was used and split into mammalian and non-mammalian
548 species. Sequences without residues in the extension region were deleted and the non-
549 mammalian alignment was augmented by LDHB sequences from the NCBI nucleotide database
550 (<http://www.ncbi.nlm.nih.gov/nucleotide>) in November 2013. In total, the alignments comprise 13
551 mammals and 9 non-mammalian vertebrates: *Homo sapiens* (human), *Mus musculus* (mouse),
552 *Rattus norvegicus* (rat), *Oryctolagus cuniculus* (rabbit), *Pan troglodytes* (chimpanzee), *Gorilla*
553 *gorilla*, *Pongo abelii* (orangutan), *Macaca mulatta* (rhesus macaque), *Felis catus* (cat), *Canis*
554 *familiaris* (dog), *Equus caballus* (horse), *Bos taurus* (cow), *Ovis aries* (sheep); *Xenopus*
555 *tropicalis* (western clawed frog), *Anolis carolinensis* (anole lizard), *Ficedula albicollis* (flycatcher),
556 *Taeniopygia guttata* (zebra finch), *Gallus gallus* (chicken), *Meleagris gallopavo* (turkey), *Alligator*
557 *mississippiensis*, *Salmo salar* (salmon), *Danio rerio* (zebrafish).

558 The genomic sequences were translated to amino acid sequences using the 'EMBOSS Transeq'
559 web server (http://www.ebi.ac.uk/Tools/st/emboss_transeq/). Species trees were obtained from
560 the Interactive Tree Of Life (iTOL) web site (<http://itol.embl.de/>) and visualized with the Phylip
561 package (Felsenstein, 1989). The JalView software (Waterhouse et al., 2009) was used to
562 visualize the alignments and to compute alignment quality and consensus. Here, the quality
563 score of an alignment column is inversely proportional to the average cost of all pairs of

564 mutations in terms of BLOSUM 62 substitution scores and the consensus reflects the fraction of
565 the most frequent residue for each column of the alignment.

566

567 **DNA Cloning**

568 Plasmids used in this study are listed in the table in Supplementary file 1. Oligonucleotides used
569 in this study are listed in the table in Supplementary file 2.

570 The dual reporter vector pDRV1 (PST1360) encoding an N-terminal Venus-tag and a C-terminal
571 hRluc tag was derived from pEXP-Venus-hRluc (a gift from Ania Muntau and Sören Gersting) by
572 introducing a short multicloning site (MCS) containing BstEII, ClaI, BspEI, and BsiWI restriction
573 sites. pDRV1 was created by ligating pre-annealed oligonucleotides OST963 and OST964 into
574 the XhoI site of pEXP Venus-hRluc. Dual reporter constructs PST1384-1385, 1387, 1393-1396,
575 1418-1426, 1430, 1435,1437, 1493, 1494, 1497, 1504 and PST 1444 were derived from pDRV1
576 by insertion of pre-annealed oligonucleotides OST1081-1084, 1086, 1087, 1117-1124, 1144-
577 1145, 1148-1157, 1160-1165, 1158-1159, 1190-1191, 1198-1199, 1229-1230, JH59-60, JH61-
578 62, JH 67-68, and JH81-82 into BspEI and BstEII sites, as listed in Supplementary file 2.

579 For cloning of pEYFP-LDHBx (PST1388) the LDHB open reading frame including the stop codon
580 and the 18 nucleotide 3' extension was PCR-amplified from pOTB7-LDHB using primers
581 OST1053 and 1054 and inserted into EcoRI and XbaI sites of pEYFP-C1.

582 The stop codon variants pEYFP-LDHBx[TGG] (PST1389), pECFP-LDHBx[TGG] (PST1440),
583 pEYFP-LDHBx[TAA] (PST1410), pEYFP-LDHBx[TAAT] (PST1411), and pEYFP-LDHBx[TGAT]
584 (PST1409) were created by amplifying LDHBx using primer OST1053 with reverse primers
585 OST1055, 1127, 1128, and 1129, respectively. Similarly, the PTS1 mutation variants pEYFP-
586 LDHBx[ΔL] (PST1407), pECFP-LDHBx[TGG, ΔL] (PST1512) (deletion of the last amino acid in
587 the cryptic PTS1 SRL) and pEYFP-LDHBx[SSI] (PST1408), pECFP-LDHBx[TGG, SSI]
588 (PST1513) (substitution of the PTS1 SRL by SSI) were created using forward primer OST1053
589 and reverse primers OST1125, 1263, 1126, and 1264, respectively. LDHA was amplified from
590 human cDNA using primers OST1130 and 1131 and cloned into EcoRI and XbaI sites of
591 pEYFP-C1 to yield pEYFP-LDHA (PST1434).

592 For cloning of pEXP Venus-PTS1 (PST1209), primers OST801 and 802 (encoding the PTS1 of
593 ACOX3) were annealed and inserted into pENTR-TOPO-D. Then the PTS1 tag was transferred
594 to pEXP-N-Venus using LR clonase II (Invitrogen).

595 Full-length dual reporter constructs pcDNA3.1-HA-LDHBx-myc and variants were cloned by
596 amplifying LDHB and stop codon variants from PST1388 (LDHB wt), PST1389 (LDHB [TGG]),
597 PST1409 (LDHB [TGAT]), PST1410 (LDHB [TAA]), PST1411 (LDHB [TAAT]), using primers

598 OST1202 and 1203 and cloning into NheI and BamHI restriction sites of pcDNA3.1/myc-His(-)A.
599 All plasmids were confirmed by DNA sequencing.

600

601 **Cell culture and transfection**

602 HeLa cells and human skin fibroblasts were maintained in low glucose Dulbecco's minimal
603 essential medium (DMEM), HEK cells, U118, and COS-7 cells in high glucose DMEM. Culture
604 media were supplemented with 1% (w/v) glutamine, 5-10% (v/v) heat inactivated fetal calf serum
605 (FCS), 100 units/ml penicillin and 100 µg/ml streptomycin. For U118 cells, 1% non-essential
606 amino acids and 1% pyruvate were added to the media.

607 Cells were transfected using Effectene transfection reagent (Qiagen) as described by the
608 manufacturer. Plasmids were diluted in Buffer EC and Enhancer and incubated for 5 min at room
609 temperature. Effectene was added and incubated for 10 min at room temperature. Prewarmed
610 medium was added to the HeLa cells and to the transfection mixture which was then added to
611 cells and incubated at 37°C in a humidified 5% CO₂ incubator for 24 h. Six hrs after transfection,
612 transfection reagent was removed, and, where indicated, geneticin (G418) was added at a
613 concentration of 100 µg/ml.

614

615 **Dual reporter assays and readthrough calculation**

616 Cells were washed with PBS and lysed by Renilla Luciferase Assay Lysis Buffer (Promega)
617 according to manufacturer's manual. Cells were spun down (14 krpm, 2 min, 4°C) and
618 supernatants were stored at -80°C. For Venus fluorescence measurement cell lysates was
619 diluted 1:25 in PBS and analyzed at 485 nm excitation, 530 nm emission (sensitivity: 130) using
620 a Synergy Mx plate reader (Biotek). PBS was used as a blank control for fluorescence
621 measurements.

622 Undiluted lysates (20 µl) were used to measure hRluc luminescence by the Renilla Luciferase
623 Assay System (Promega) and the Synergy Mx plate reader (Biotek). An automated injector was
624 used to add 100 µl Renilla Luciferase Assay Reagent. Luminescence was read 2 sec after
625 injection and integrated over 10 sec (sensitivity 150). Renilla Luciferase Assay Reagent was
626 used as a blank control for hRluc luminescence measurements. Each construct was analyzed in
627 3 to 7 biological replicates and each biological sample was measured in triplets.

628 To obtain readthrough rates the ratio of hRluc/Venus fluorescence was calculated, and the
629 readthrough of pDRVL was set to 100%. Ratio (y) and standard deviation of fluorescence (x₁)
630 and luminescence (x₂) signal for each replicate were calculated using uncertainty propagation
631 ($\sigma_y = [\sigma_{x1} * (dy/dx_1)^2 + \sigma_{x2} * (dy/dx_2)^2]^{0.5}$). Let $w_i = 1/\sigma_i^2$ be the weight of a readthrough value from

632 replicate i with σ_i being the error of the ratios. Then the weighted mean \mathbf{x}_m of the replicates and
633 its error σ_{x_m} was calculated according to $\mathbf{x}_m = (\sum_i(x_i w_i) / \sum_i w_i)$ and $\sigma_{x_m} = (\sum_i w_i)^{-0.5}$.

634

635 **Immunofluorescence, microscopy, and quantification**

636 Transfected LDHB and LDHA fusion constructs were detected in HeLa cells by combined
637 direct/immunofluorescence experiments. Endogenous LDHB was analyzed in HeLa, U118, and
638 COS-7 cells, and in primary rat cardiomyocytes by immunofluorescence. Approximately 1×10^5
639 cells were seeded on cover slips or on laminin-coated (Sigma) glass slides for HEK cells and
640 cardiomyocytes and transfected as indicated. For removal of cytosol, cells were treated with
641 0.02% (w/v) digitonin (Invitrogen) for 5 min at room temperature. Cells were fixed with 10% (w/v)
642 formaldehyde for 20 min, and permeabilized with 0.5% Triton X-100 for 5 min. After blocking for
643 20 min at 37 °C with 10 % BSA antigens were labeled with primary antibodies at 37 °C for 1 hr.
644 Antibody dilutions were 1:200 for anti-PEX14 rabbit polyclonal antibodies (ProteinTech) and
645 1:500 for anti-LDHB mouse monoclonal (Abnova). Secondary antibody labeling (1:200) was
646 done for 1 hr with antibodies labelled with Cy3 and/or Alexa647 (Jackson Immuno Research),
647 and/or Alexa488 (MoBiTech). Cover slips were mounted with Mowiol containing 0.01 mg/ml 4',6-
648 diamidino-2-phenylindole (DAPI). DAPI was omitted in cases when cells had been transfected
649 with CFP-expressing plasmids.

650 Fluorescence microscopy was done using a 100x oil objective (1.3 NA) with a Zeiss Imager M1
651 fluorescence wide field scope equipped with the Zeiss AxioCam HRm Camera and Zeiss
652 Axiovision 4.8 acquisition software. z-Stacks with 30 images and 0.25 μm spacing were
653 recorded and subjected to deconvolution. Where necessary, linear contrast enhancements were
654 applied (Axiovision).

655 To quantify induction of endogenous LDHB by geneticin, fluorescence images from samples
656 prepared with anti-LDHB and anti-PEX14 antibodies were recorded under identical conditions
657 and subjected to deconvolution. The LDHB/PEX14 intensities were measured applying the same
658 threshold ratios to all channel pairs (ImageJ). Induction is expressed as the ratio of
659 LDHB/PEX14 ratios with and without geneticin treatment, respectively.

660

661 **Western blot analysis**

662 Cells were lysed in RIPA lysis buffer (20 mM Tris-HCl, pH 7.4, 150 mM sodium chloride, 2 mM
663 EDTA, 1% NP40, 1 mM DTT, 0.1 mM PMSF, Complete protease inhibitors (Roche)) 24 h after
664 transfection. Proteins were separated by SDS-PAGE on a 12% gel, transferred to a
665 nitrocellulose membrane and probed with primary and secondary antibodies. The following
666 antibodies were used: anti-HA rabbit polyclonal (Abcam), anti-myc mouse monoclonal (Cell

667 signaling), anti-luciferase mouse monoclonal (Millipore), anti-GFP mouse monoclonal (Living
668 colors), anti-Actin mouse monoclonal (Sigma). HRP-conjugated goat anti-rabbit IgG and donkey
669 anti-mouse IgG (Jackson Immuno Research) were used as secondary antibodies. 1:1000
670 dilutions of primary antibody and 1:5000 for secondary antibody were used. Reactive bands
671 were revealed with Lumi-light and Lumi-light plus Western blotting substrate (Roche). Images
672 were scanned using Luminescent image analyzer LAS 4000.

673

674 DATA AVAILABILITY

675 **Dataset 1.** Spreadsheet containing predicted readthrough extensions, RTP scores (LIN. LINiter.
676 LINfs5, LINfs3), PTS1 scores, predictions of ER retentions signals, glycosylation motifs,
677 transmembrane domains, and transmembrane topology, and the LINiter⁺ x PTS1 product scores
678 for all human transcript termini. Publicly available at the Dryad Digital Repository with the DOI
679 [10.5061/dryad.j2n18](https://doi.org/10.5061/dryad.j2n18) (Schueren et al., 2014).

680

681 ACKNOWLEDGEMENTS

682 We thank Heiner Klingenberg for help with orthology searches of human PTS1 proteins, Ania
683 Muntau and Sören Gersting for plasmids, and Kristina Gamper and Viacheslav Nikolaev for the
684 rat cardiomyocytes. We are grateful to Ellen Krämer and Tanja Wilke for technical assistance,
685 and to Cindy Krause, Peter Meinicke, Olaf Jahn, Johannes Freitag, and Michael Bölker for
686 discussions. We thank Blanche Schwappach, Heinz Neumann, Heike Krebber, and Maya
687 Schuldiner for comments on the manuscript.

688

689 REFERENCES

- 690 Baranov PV, Gesteland RF, Atkins JF. 2002. Recoding: translational bifurcations in gene
691 expression. *Gene* **286**:187–201.
- 692 Baumgart E, Fahimi HD, Stich A, Völkl A. 1996. L-lactate dehydrogenase A4- and A3B isoforms
693 are bona fide peroxisomal enzymes in rat liver. Evidence for involvement in
694 intraperoxisomal NADH reoxidation. *J. Biol. Chem.* **271**:3846–3855.
- 695 Beier H, Grimm M. 2001. Misreading of termination codons in eukaryotes by natural nonsense
696 suppressor tRNAs. *Nucleic Acids Res.* **29**:4767–4782.
- 697 Bidou L, Allamand V, Rousset J-P, Namy O. 2012. Sense from nonsense: therapies for
698 premature stop codon diseases. *Trends Mol. Med.* **18**:679–688.
- 699 Boyer SH, Fainer DC, Watson-Williams EJ. 1963. Lactate dehydrogenase variant from human
700 blood: evidence for molecular subunits. *Science* **141**:642–643.

- 701 Braverman NE, D'Agostino MD, Maclean GE. 2013. Peroxisome biogenesis disorders:
702 Biological, clinical and pathophysiological perspectives. *Dev. Disabil. Res. Rev.* **17**:187–
703 196.
- 704 Brocard C, Hartig A. 2006. Peroxisome targeting signal 1: is it really a simple tripeptide?
705 *Biochim. Biophys. Acta* **1763**:1565–1573.
- 706 Brooks GA. 2009. Cell-cell and intracellular lactate shuttles. *J. Physiol.* **587**:5591–5600.
- 707 Chittum HS, Lane WS, Carlson BA, Roller PP, Lung FD, Lee BJ, Hatfield DL. 1998. Rabbit beta-
708 globin is extended beyond its UGA stop codon by multiple suppressions and translational
709 reading gaps. *Biochemistry (Mosc.)* **37**:10866–10870.
- 710 Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS. 2013. Ribosome profiling reveals
711 pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife*
712 **2**:e01179.
- 713 Eswarappa SM, Potdar AA, Koch WJ, Fan Y, Vasu K, Lindner D, Willard B, Graham LM,
714 DiCorleto PE, Fox PL. 2014. Programmed Translational Readthrough Generates
715 Antiangiogenic VEGF-Ax. *Cell* **157**:1605–1618.
- 716 Felsenstein J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**:164–
717 166.
- 718 Firth AE, Brierley I. 2012. Non-canonical translation in RNA viruses. *J. Gen. Virol.* **93**:1385–
719 1409.
- 720 Firth AE, Wills NM, Gesteland RF, Atkins JF. 2011. Stimulation of stop codon readthrough:
721 frequent presence of an extended 3' RNA structural element. *Nucleic Acids Res.*
722 **39**:6679–6691.
- 723 Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates
724 G, Fairley S, Fitzgerald S, Gil L, Garcia-Giron C, Gordon L, Hourlier T, Hunt S,
725 Juettemann T, Kahari AK, Keenan S, Komorowska M, Kulesha E, Longden I, Maurel T,
726 McLaren WM, Muffato M, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat
727 HS, Ritchie GRS, Ruffier M, Schuster M, Sheppard D, Sobral D, Taylor K, Thormann A,
728 Trevanion S, White S, Wilder SP, Aken BL, Birney E, Cunningham F, Dunham I, Harrow
729 J, Herrero J, Hubbard TJP, Johnson N, Kinsella R, Parker A, Spudich G, Yates A,
730 Zadissa A, Searle SMJ. 2012. Ensembl 2013. *Nucleic Acids Res.* **41**:D48–D55.
- 731 Floquet C, Hatin I, Rousset J-P, Bidou L. 2012. Statistical Analysis of Readthrough Levels for
732 Nonsense Mutations in Mammalian Cells Reveals a Major Determinant of Response to
733 Gentamicin. Ed. Kevin M. Flanigan. *PLoS Genet.* **8**:e1002608.
- 734 Freitag J, Ast J, Bölker M. 2012. Cryptic peroxisomal targeting via alternative splicing and stop
735 codon read-through in fungi. *Nature* **485**:522–525.
- 736 Geller AI, Rich A. 1980. A UGA termination suppression tRNATrp active in rabbit reticulocytes.
737 *Nature* **283**:41–46.
- 738 Gladden LB. 2004. Lactate metabolism: a new paradigm for the third millennium. *J. Physiol.*
739 **558**:5–30.
- 740 Goldberg E, Eddy EM, Duan C, Odet F. 2009. LDHC: The Ultimate Testis-Specific Gene. *J.*
741 *Androl.* **31**:86–94.
- 742 Gronemeyer T, Wiese S, Ofman R, Bunse C, Pawlas M, Hayen H, Eisenacher M, Stephan C,
743 Meyer HE, Waterham HR, Erdmann R, Wanders RJ, Warscheid B. 2013. The Proteome
744 of Human Liver Peroxisomes: Identification of Five New Peroxisomal Constituents by a
745 Label-Free Quantitative Proteomics Survey. Ed. Anand S. Mehta. *PLoS ONE* **8**:e57395.
- 746 Gualdrón-López M, Brennand A, Hannaert V, Quiñones W, Cáceres AJ, Bringaud F,
747 Concepción JL, Michels PAM. 2012. When, how and why glycolysis became

- 748 compartmentalised in the Kinetoplastea. A new look at an ancient organelle. *Int. J.*
749 *Parasitol.* **42**:1–20.
- 750 Harrell L, Melcher U, Atkins JF. 2002. Predominance of six different hexanucleotide recoding
751 signals 3' of read-through stop codons. *Nucleic Acids Res.* **30**:2011–2017.
- 752 Hoerl AE, Kennard RW. 1970. Ridge Regression: Biased Estimation for Nonorthogonal
753 Problems. *Technometrics* **12**:55–67.
- 754 Hsu PP, Sabatini DM. 2008. Cancer Cell Metabolism: Warburg and Beyond. *Cell* **134**:703–707.
- 755 Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells
756 reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**:789–802.
- 757 Jungreis I, Lin MF, Spokony R, Chan CS, Negre N, Victorsen A, White KP, Kellis M. 2011.
758 Evidence of abundant stop codon readthrough in Drosophila and other metazoa.
759 *Genome Res.* **21**:2096–2113.
- 760 Keeling KM, Xue X, Gunn G, Bedwell DM. 2014. Therapeutics Based on Stop Codon
761 Readthrough. *Annu. Rev. Genomics Hum. Genet.*
- 762 Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein
763 topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*
764 **305**:567–580.
- 765 Kunze M, Hartig A. 2013. Permeability of the peroxisomal membrane: lessons from the
766 glyoxylate cycle. *Front. Physiol.* **4**:204.
- 767 Lanyon-Hogg T, Warriner SL, Baker A. 2010. Getting a camel through the eye of a needle: the
768 import of folded proteins by peroxisomes. *Biol. Cell Auspices Eur. Cell Biol. Organ.*
769 **102**:245–263.
- 770 Li G, Rice CM. 1993. The signal for translational readthrough of a UGA codon in Sindbis virus
771 RNA involves a single cytidine residue immediately downstream of the termination
772 codon. *J. Virol.* **67**:5062–5067.
- 773 Lin MF, Jungreis I, Kellis M. 2011. PhyloCSF: a comparative genomics method to distinguish
774 protein coding and non-coding regions. *Bioinformatics* **27**:i275–i282.
- 775 Lingner T, Kataya AR, Antonicelli GE, Benichou A, Nilssen K, Chen X-Y, Siemsen T,
776 Morgenstern B, Meinicke P, Reumann S. 2011. Identification of Novel Plant Peroxisomal
777 Targeting Signals by a Combination of Machine Learning Methods and in Vivo
778 Subcellular Targeting Analyses. *Plant Cell* **23**:1556–1572.
- 779 Lodhi IJ, Semenkovich CF. 2014. Peroxisomes: A Nexus for Lipid Metabolism and Cellular
780 Signaling. *Cell Metab.* **19**:380–392.
- 781 Loughran G, Chou M-Y, Ivanov IP, Jungreis I, Kellis M, Kiran AM, Baranov PV, Atkins JF. 2014.
782 Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids*
783 *Res.*
- 784 Markert CL. 1963. Lactate Dehydrogenase Isozymes: Dissociation and Recombination of
785 Subunits. *Science* **140**:1329–1330.
- 786 Maynard EL, Gatto GJ, Berg JM. 2004. Pex5p binding affinities for canonical and noncanonical
787 PTS1 peptides. *Proteins Struct. Funct. Bioinforma.* **55**:856–861.
- 788 McCaughan KK, Brown CM, Dalphin ME, Berry MJ, Tate WP. 1995. Translational termination
789 efficiency in mammals is influenced by the base following the stop codon. *Proc. Natl.*
790 *Acad. Sci. U. S. A.* **92**:5431–5435.
- 791 McClelland GB, Khanna S, González GF, Eric Butz C, Brooks GA. 2003. Peroxisomal
792 membrane monocarboxylate transporters: evidence for a redox shuttle system?
793 *Biochem. Biophys. Res. Commun.* **304**:130–135.

794 McGroarty E, Hsieh B, Wied DM, Gee R, Tolbert NE. 1974. Alpha hydroxy acid oxidation by
795 peroxisomes. *Arch. Biochem. Biophys.* **161**:194–210.

796 McNew JA, Goodman JM. 1996. The targeting and assembly of peroxisomal proteins: some old
797 rules do not apply. *Trends Biochem. Sci.* **21**:54–58.

798 Namy O, Hatin I, Rousset JP. 2001. Impact of the six nucleotides downstream of the stop codon
799 on translation termination. *EMBO Rep.* **2**:787–793.

800 Namy O, Rousset J-P, Naphine S, Brierley I. 2004. Reprogrammed genetic decoding in cellular
801 gene expression. *Mol. Cell* **13**:157–168.

802 Osmundsen H. 1982. Factors which can influence beta-oxidation by peroxisomes isolated from
803 livers of clofibrate treated rats. Some properties of peroxisomal fractions isolated in a
804 self-generated Percoll gradient by vertical rotor centrifugation. *Int. J. Biochem.* **14**:905–
805 914.

806 Pacho F, Zambruno G, Calabresi V, Kiritsi D, Schneider H. 2011. Efficiency of translation
807 termination in humans is highly dependent upon nucleotides in the neighbourhood of a
808 (premature) termination codon. *J. Med. Genet.* **48**:640–644.

809 Read JA, Winter VJ, Eszes CM, Sessions RB, Brady RL. 2001. Structural basis for altered
810 activity of M- and H-isozyme forms of human lactate dehydrogenase. *Proteins* **43**:175–
811 185.

812 Salido E, Pey AL, Rodriguez R, Lorenzo V. 2012. Primary hyperoxalurias: disorders of
813 glyoxylate detoxification. *Biochim. Biophys. Acta* **1822**:1453–1464.

814 Schueren F, Lingner T, George R, Hofhuis J, Dickel C, Gärtner J, Thoms S. 2014. Data from:
815 Peroxisomal lactate dehydrogenase is generated by basal translational readthrough in
816 mammals. *Dryad Digital Repository*. doi:10.5061/dryad.j2n18

817 Schwarz F, Aebi M. 2011. Mechanisms and principles of N-linked protein glycosylation. *Curr.*
818 *Opin. Struct. Biol.* **21**:576–582.

819 Smith JJ, Aitchison JD. 2013. Peroxisomes take shape. *Nat. Rev. Mol. Cell Biol.* **14**:803–817.

820 Stornaiuolo M, Lotti LV, Borgese N, Torrisi M-R, Mottola G, Martire G, Bonatti S. 2003. KDEL
821 and KKXX retrieval signals appended to the same reporter protein determine different
822 trafficking between endoplasmic reticulum, intermediate compartment, and Golgi
823 complex. *Mol. Biol. Cell* **14**:889–902.

824 Thoms S, Hofhuis J, Thöing C, Gärtner J, Niemann HH. 2011. The unusual extended C-terminal
825 helix of the peroxisomal α/β -hydrolase Lpx1 is involved in dimer contacts but dispensable
826 for dimerization. *J. Struct. Biol.* **175**:362–371.

827 Tuite MF, Cox BS. 2003. Propagation of yeast prions. *Nat. Rev. Mol. Cell Biol.* **4**:878–890.

828 Visser WF, van Roermund CWT, Ijlst L, Waterham HR, Wanders RJA. 2007. Metabolite
829 transport across the peroxisomal membrane. *Biochem. J.* **401**:365–375.

830 Völkl A, Fahimi HD. 1985. Isolation and characterization of peroxisomes from the liver of normal
831 untreated rats. *Eur. J. Biochem.* **149**:257–265.

832 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2--a
833 multiple sequence alignment editor and analysis workbench. *Bioinforma. Oxf. Engl.*
834 **25**:1189–1191.

835 Workman CT, Yin Y, Corcoran DL, Ideker T, Stormo GD, Benos PV. 2005. enoLOGOS: a
836 versatile web tool for energy normalized sequence logos. *Nucleic Acids Res.* **33**:W389–
837 392.

838 Yamaguchi Y, Hayashi A, Campagnoni CW, Kimura A, Inuzuka T, Baba H. 2012. L-MPZ, a
839 Novel Isoform of Myelin P0, Is Produced by Stop Codon Readthrough. *J. Biol. Chem.*
840 **287**:17765–17776.

- 841 Zerangue N, Malan MJ, Fried SR, Dazin PF, Jan YN, Jan LY, Schwappach B. 2001. Analysis of
842 endoplasmic reticulum trafficking signals by combinatorial screening in mammalian cells.
843 *Proc. Natl. Acad. Sci. U. S. A.* **98**:2431–2436.
- 844 Zhang FL, Casey PJ. 1996. Protein prenylation: molecular mechanisms and functional
845 consequences. *Annu. Rev. Biochem.* **65**:241–269.
- 846 Zielinska DF, Gnad F, Wiśniewski JR, Mann M. 2010. Precision Mapping of an In Vivo N-
847 Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **141**:897–907.

848 **Figure titles and legends**

849

850 **Figure 1.** Genome-wide *in silico* analysis of basal translational readthrough (BTR) in humans.
851 **(A)** Schematic representation of the readthrough propensity (RTP) predictor algorithm. Binary
852 vector representations of stop codon contexts and their experimental readthrough values are
853 used to determine the coefficients of a linear regression model. For prediction of RTP for a given
854 stop codon context the position-specific regression coefficients are added up. **(B)** Readthrough
855 propensity (RTP) for selected human transcripts. LIN denotes first-pass RTP calculations,
856 LINiter iterative improvement of RTP scoring, and LINfs3 and LINfs5 the reduced models. RTP
857 of all human transcripts can be found in Dataset 1 (Schueren et al., 2014). **(C)** Experimental
858 readthrough by dual reporter assay in HeLa cells. Readthrough is expressed as luciferase per
859 Venus signal. The red line marks the background readthrough level obtained from a construct
860 containing two contiguous UAA stop codons separating the Venus and the hRluc. The
861 aminoglycoside geneticin (100 µg/ml) induces translational readthrough. SCC, stop codon
862 context. hRluc, humanized renilla luciferase. Error bars, SD.

863 The following figure supplements are available for figure 1:

864 **Figure supplement 1.** Schematic representation of the readthrough propensity (RTP) prediction
865 algorithm.

866 **Figure supplement 2.** Correlation of RTP and BTR in the LINiter model.

867

868 **Figure 2.** Characterization of basal translational readthrough (BTR): consensus and candidates.
869 (A) Sequence logo plot of regression coefficients of stop codon contexts in the LINiter model.
870 Character size corresponds to regression coefficients. The model treats stop codons as one
871 nucleotide position. Filled/upside-down letters correspond to positive/negative coefficients,
872 respectively. (B) Consensus motif for high RTP derived from the predictive model. The stop
873 codon together with the nucleotide triplet following the stop codon provides the best predictor for
874 RTP. The consensus was derived by feature selection: Starting from LINiter, positions with the
875 least contribution to prediction were successively eliminated as indicated by the gray arrow.
876 Nucleotide positions on the x-axis mark the removed positions upon transition to a reduced
877 model. LINfs3 (UGA CUA, stop codon underlined) represents the global minimum of regression
878 error (filled circle). The model LINfs5, corresponding to a local minimum additionally
879 encompasses positions +7 and -6, indicating that these positions could also contribute to high
880 BTR. (C) BTR determination of candidates from the genome wide *in silico* screen. Dual reporter
881 assays with Venus and humanized Renilla luciferase containing SCCs from AQP4 (UGA CUA
882 G), SYTL2 (UGA CUA G), CACNA2D4 (UGA CUA T) and DHX38 (UGA CUU G). AQP4, SYTL2,
883 and CACNA2D4 reveal high BTR in all tissues tested. HT1080, human fibrosarcoma cell line;
884 HEK, human embryonic kidney cells, U373 glioblastoma cell line. Error bars, SD.

885 The following figure supplements are available for figure 2:

886 **Figure supplement 1.** Correlation of RTP and BTR in the LINfs3 model.

887 **Figure supplement 2.** Venn diagram comparing previously identified readthrough proteins with
888 the high BTR candidates tested in this study.

889

890 **Figure 3.** LDHB is extended by translational readthrough. (A) Genomic distribution of RTP⁺ *
891 PTS1 product scores. Product scores are 0 for rank 5015 to 42069. Green cross: 50% of max.
892 score. LDHB has the highest product score, exceeding rank 2 by 24%. RTP⁺ denotes positively
893 scaled LINiter values. (B-C) Venus/hRluc dual reporter assay with LDHB wild-type and mutant
894 stop codon contexts. Error bars, SD. (B) Wild-type LDHB stop context shows high BTR.
895 Mutational analysis of the LINfs3 consensus of LDHB. Replacement of the stop codon and
896 mutations in positions +4 to +6 reduce readthrough. (C) LDHB readthrough induction by the
897 aminoglycoside geneticin. (D) Full-length LDHB is extended by readthrough. Western-blot of
898 dual tag assay with LDHBx with N-terminal HA- and C-terminal myc-tag. Molecular mass marker
899 in kDa.

900 The following figure supplement is available for figure 3:

901 **Figure supplement 1.** The LDHB stop context favors readthrough (Western blot of dual reporter
902 constructs).

903

904 **Figure 4.** LDHBx targets to the peroxisome by translational readthrough and a hidden
905 peroxisomal targeting signal type1 (PTS1) in the 3' extension. **(A-C)** Direct fluorescence
906 microscopy of transfected HeLa cells. Immunofluorescence with the peroxisome marker anti-
907 PEX14 (red). **(A)** YFP-LDHB (green) mainly localizes to the cytosol. The strong fluorescence
908 signal in the cytosol prevents detection of LDHB in other cellular compartment. **(B)** Upon plasma
909 membrane permeabilization and removal of cytosol (-CYT) a small fraction of LDHB remains co-
910 localized with the peroxisome marker. **(C)** Peroxisomal targeting of LDHB is dependent on the
911 cryptic PTS1 Ser-Arg-Leu (SRL) in the extension. Deletion of the L in SRL blocks import into
912 peroxisomes. **(D-E)** Endogenous LDHB is localized to peroxisomes in untransfected wild-type
913 cells. Immunofluorescence with anti-LDHB (green) and anti-PEX14 (red) antibodies. **(D)**
914 Endogenous LDHB is cytosolic. **(E)** Removal of cytosol (-CYT) reveals co-localization with
915 PEX14. Bar 10 μ m.

916
917 **Figure 5.** Peroxisome targeting of LDHBx is dependent on a hidden peroxisomal targeting signal
918 in the readthrough extension. Combined direct fluorescence and immunofluorescence in HeLa
919 cells. **(A)** YFP-LDHBx expression: LDHBx is mainly cytosolic. **(B)** LDHBx targets to the
920 peroxisome. Cells were permeabilized with digitonin, and cytosol was removed by washing with
921 PBS. **(C-F)** Mutation of the cryptic PTS1 in the extension blocks peroxisomal targeting of
922 LDHBx. **(C, D)** Deletion of the amino acid L of the SRL in the PTS1 readthrough extension gives
923 a wild-type cytosolic localization of LDHB and blocks import into the peroxisome completely. **(E,**
924 **F)** Similarly, the SRL-to-SSI substitution does not interfere with cytosolic expression of the LDHB
925 but completely blocks peroxisomal localization of LDHBx[SSI]. Bar 10 μ m.

926 The following figure supplements are available for figure 5:

927 **Figure supplement 1.** Complete removal of cytosol after permeabilization with digitonin.

928 **Figure supplement 2.** Positive control: Peroxisomal localization of a YFP variant fused to PTS1
929 of the peroxisomal matrix protein ACOX3.

930
931 **Figure 6.** Peroxisome targeting of LDHBx is dependent on the stop codon. Combined direct
932 fluorescence and immunofluorescence in HeLa cells. **(A, B)** Exchange of UGA stop to the tighter
933 stop UAA (YFP-LDHBx[TAA]) reduces peroxisomal localization of LDHB. **(C, D)** When UGA is
934 replaced by tryptophan-coding UGG (LDHBx[TGG]) a larger proportion of LDHB is targeted to
935 the peroxisome, and peroxisome localization becomes obvious without removal of the cytosol.
936 **(B, D)** Cytosol was removed after cell permeabilization with digitonin. Bar 10 μ m.

937

938 **Figure 7.** Endogenous LDHB is localized to peroxisomes in wild-type cells. Immunofluorescence
939 in wild-type cultured (**A-F**) or freshly prepared (**G, H**) cells with antibodies recognizing LDHB
940 (secondary antibody Alexa488-coupled) and the peroxisome marker PEX14 (secondary antibody
941 Cy3-coupled).(**A, B**) COS-7 cells, (**C, D**) human skin fibroblasts (**E, F**) U118 glioblastome, and
942 (**G, H**) primary rat cardiomyocytes. (**B, D, F, H**) Cytosol was removed after permeabilization with
943 digitonin (-CYT). Bar 10 μm .

944

945 **Figure 8.** Evidence for regulation of readthrough. LDHB stop codon readthrough in various
946 mammalian cell types. COS-7, HEK, and U118 cells were transfected with LDHB and mutant
947 dual reporter constructs and analyzed by Venus fluorescence and luciferase assays.
948 Readthrough is expressed as hRLuc/Venus signal. 100 $\mu\text{g/ml}$ geneticin induces readthrough.

949

950 **Figure 9.** LDHBx extensions including hidden PTS1 are strictly conserved in mammals.
951 Alignments of LDHBx termini from mammals and non-mammalian vertebrates. PTS1 extension
952 boxed. The conserved readthrough PTS1 extension is found exclusively in mammals and marks
953 the mammalian- non-mammalian border in vertebrates.

954

955 **Figure 10.** Piggy-back co-import of LDHA by LDHBx into peroxisomes. Direct fluorescence of
956 YFP-labelled LDHA (green) in the absence or presence of CFP-labelled LDHBx[TGG] (red)
957 combined with immunofluorescence with a peroxisome marker (blue). (**A**) YFP-LDHA
958 localization is mainly in the cytosol when expressed in the absence of LDHBx. (**B**) LDHA is
959 imported into peroxisomes when co-expressed with LDHBx[TGG]. Cytosol was removed after
960 permeabilization with digitonin. Bar 10 μm .

961 The following figure supplements are available for figure 10:

962 **Figure supplement 1.** Mutation of the cryptic PTS1 serine-arginine-leucine (SRL) in LDHBx by
963 deletion of L or substitution of RL by serine-isoleucine (SI) blocks co-import of LDHA into
964 peroxisomes.

965 **Figure supplement 2.** Positions of the C-termini in the tetrameric LDH structure.

Tables

Gene symbol	Stop codon context	Readthrough (%) (SD)
<i>ZNF-574</i>	GATCAGTGGC TGA CTCTGCCCGA	0.31 (0.020)
<i>LDHB</i>	AAAAGACCTG TGA CTAGTGAGCT	1.55 (0.087)
<i>PPP1R3F</i>	ATTCTCCCAA TAA AGCTTTACAG	0.18 (0.009)
<i>LDHB [TGAT]</i>	AAAAGACCTG TGA TTAGTGAGCT	0.17 (0.009)
<i>LDHB [TAA]</i>	AAAAGACCTG TAA CTAGTGAGCT	0.20 (0.009)
<i>LDHB [TAAT]</i>	AAAAGACCTG TAA TTAGTGAGCT	0.17 (0.009)
<i>LENG1</i>	CCTTACTCAC TGA CTCCTGAGGG	0.26 (0.009)
<i>VASN</i>	GCCCTACATC TAA GCCAGAGAGA	0.12 (0.004)
<i>MDH1</i>	TTCCTCTGCC TGA CTAGACAATG	2.91 (0.147)
<i>PRDM10</i>	CACCAAACCA TGA CTTCCACCCT	0.13 (0.005)
<i>FBXL20</i>	CATCATCCTA TGA CAATGGAGGT	0.10 (0.006)
<i>THG1L</i>	AGCCAGGCTT TGA CGGAAGAGTC	0.15 (0.006)
<i>EDEM3</i>	GGATGAGCTA TGA CTTGCTAAAC	0.66 (0.027)
<i>EDN1</i>	AGCACATTGG TGA CAGACCTTCG	0.25 (0.008)
<i>UBQLN1</i>	CCAGCCATCA TAG CAGCATTTCT	0.13 (0.009)
<i>IRAK3</i>	CAAAAAAGAA TAA ATTCTACCAG	0.10 (0.007)
<i>SLC3A1</i>	TACCTCGTGT TAG GCACCTTTAT	0.18 (0.008)
<i>LEPRE1</i>	GGATGAGCTA TGA CAGCGTCCAG	0.27 (0.010)

967

968 **Table 1.** Experimental dual reporter readthrough data of stop codon contexts used for the LINiter
 969 model. Stop codon constructs expressing plus/minus 10 nucleotides analyzed in HeLa cells.

970

971

972

973

974

975

976

LinIter model (SSC position -6 to +9)

Base \ Pos	-6	-5	-4	-3	-2	-1	4
A	-0.00041	0.00130	-0.00028	-0.00073	-0.00071	0.00016	-0.00037
C	-0.00105	0.00164	0.00075	-0.00004	0.00133	0.00109	0.00375
G	0.00060	-0.00077	-0.00041	0.00193	-0.00048	0.00043	-0.00156
U/T	0.00200	-0.00103	0.00108	-0.00002	0.00100	-0.00054	-0.00067

977

Base \ Pos	5	6	7	8	9	STOP	
A	-0.00068	0.00276	-0.00020	0.00105	-0.00081	-0.00026	TAA
C	-0.00097	-0.00026	-0.00062	-0.00017	0.00148	-0.00103	TAG
G	-0.00008	-0.00059	0.00245	-0.00058	0.00014	0.00243	TGA
U/T	0.00287	-0.00076	-0.00049	0.00084	0.00032		

978

979

LinFS3 model (STOP and pos. +4 to +6)

Base	4	5	6	STOP	
A	0.00006	-0.00071	0.00306	0.00005	TAA
C	0.00351	-0.00056	0.00021	-0.00052	TAG
G	-0.00111	0.00010	-0.00093	0.00229	TGA
U/T	-0.00064	0.00299	-0.00053		

980

981 **Table 2.** Regression factors of the LINiter and the LINfs3 model. These model weights are “raw”,
 982 that is as obtained from the ridge regression procedure. For prediction of RTP the weights
 983 associated with nucleotides within the stop codon context and the corresponding stop codon
 984 have to be added up. For calculation of our RTP score we normalized the model weight vectors
 985 (i.e. the complete stack of weights) to euclidean unit sum which corresponds to a division of
 986 weights by 0.0088 (LINiter) and 0.0063 (LINfs3), respectively. Furthermore the sequence feature
 987 vectors were normalized to euclidean unit sum which corresponds to a division by the square
 988 root of the length (3.6 and 2, resp.) As a shortcut to this, the sum of raw scores can be divided
 989 by 0.0317 and 0.0126, respectively.

990

991

Nucleotide	A	C	G	U
Position				
-6	0.2892	0.2530	0.2651	0.1928
-5	0.3253	0.2651	0.1446	0.2651
-4	0.1566	0.2289	0.3494	0.2651
-3	0.2410	0.3373	0.2410	0.1807
-2	0.2651	0.1807	0.2048	0.3494
-1	0.2410	0.2530	0.2651	0.2410
4	0.2289	0.3133	0.3373	0.1205
5	0.2651	0.2530	0.1446	0.3373
6	0.2771	0.2169	0.2530	0.2530
7	0.2530	0.3133	0.2892	0.1446
8	0.3253	0.1687	0.2410	0.2651
9	0.1807	0.2771	0.2771	0.2651
Stop codons	UAA	UAG	UGA	
1 to 3	0.1928	0.3373	0.4699	

992

993 **Table 3.** Nucleotide frequencies in each position of stop codon context. The nucleotide and stop
 994 codon frequencies for positions -6 to -1 and 4 to 9 calculated for the 81 sequences used in the
 995 RTP predictor (LINiter model).

996

997 **Legends to Figure supplements**

998 **Figure 1 – figure supplement 1.** Schematic representation of the readthrough propensity (RTP)
999 prediction procedure. This scheme summarizes, how regression coefficients were extracted from
1000 experimental basal readthrough (BTR) data. (1) Stop codon contexts (SCC. position -6 to +9,
1001 stop codon at position 1 to 3) with known experimental BTR values are formalized as binary
1002 vectors in 51-dimensional vector space. (2) The binary vector reserves four entries for the four
1003 possible bases in each position ($4 * 12$) and three for the stop codon. (3) In combination with
1004 their corresponding experimental readthrough values (in %) they are used to determine the (4)
1005 coefficients of a linear regression model. (5) For RTP calculation for a given SCC the position-
1006 specific regression coefficients are added up. The values used in the example are from the
1007 LINiter model. The algorithm is used to calculate RTP of stop codon contexts of 42,000 unique 3'
1008 transcript termini listed in Dataset 1 (Schueren et al., 2014).

1009 **Figure 1 – figure supplement 2.** Correlation of RTP and BTR in the LINiter model. Scatter plot
1010 indicating the correlation between RTP and experimental BTR. RTP were obtained by leave-
1011 one-out cross-validation. Pearson correlation coefficient 0.34 ($p = 0.002$).

1012
1013 **Figure 2 – figure supplement 1.** Correlation of RTP and BTR in the LINfs3 model. Scatter plot
1014 indicating the correlation between RTP and experimental BTR. RTP were obtained by leave-
1015 one-out cross-validation. Pearson correlation coefficient 0.41 ($p = 0.0001$).

1016 **Figure 2 – figure supplement 2.** Translational readthrough in humans. Venn diagram indicating
1017 experimentally verified human genes and SCCs associated with above-average translational
1018 readthrough. Genes were identified by ribosome profiling (Dunn et al., 2013), by phylogenetic
1019 approaches (Jungreis et al., 2011; Loughran et al., 2014), and by *in silico* profiling (this study).
1020 Gene products marked in boldface (black) correspond to sequences carrying the consensus
1021 motif UGA CUA (G) identified in this study and by Loughran et al., 2014. The human genome
1022 contains 144 (30) transcripts with the high-RTP motifs UGA CUA (G). Different experimental
1023 strategies lead to the identification of genes with high physiological readthrough rates, but the
1024 molecular mechanisms underlying readthrough are likely to vary.

1025
1026 **Figure 3 – figure supplement 1.** The LDHB stop context favors readthrough (Western blot).
1027 HeLa cells transfected with Venus/hRluc dual reporter constructs were analyzed by Western
1028 blot. Wild-type LDHB stop context (UGA CUA (stop underlined)) allows stop codon readthrough.
1029 Mutation of the stop codon (UAA CUA) and the downstream bases (UAA UUA) reduce
1030 readthrough. Geneticin treatment (100 μ g/ml) induces translational readthrough in all contexts.
1031 Molecular mass marker in kDa.

1032

1033
1034 **Figure 5 – figure supplement 1.** Permeabilization by digitonin allows complete removal of
1035 cytosol. Combined direct fluorescence with anti-PEX14 immunofluorescence. HeLa cells were
1036 transfected with the empty vector expressing YFP in the cytosol. **(A)** Cytosolic expression of
1037 YFP. **(B)** Complete removal of cytosolic after cell permeabilization and washing with PBS. Bar
1038 10 μm .

1039 **Figure 5 – figure supplement 2.** Cell permeabilization and removal of cytosol maintains
1040 peroxisomal integrity and co-localization of peroxisome marker (positive control) is shown.
1041 Combined direct fluorescence with anti-PEX14 immunofluorescence in HeLa cells. **(A, B)** Cells
1042 were transfected with a construct expressing the PTS1 of ACOX3 fused to the C-terminus of the
1043 YFP variant Venus. **(B)** Co-localization of PTS1 and PEX14 after removal of cytosol after cell
1044 permeabilization. Bar 10 μm .

1045
1046
1047
1048 **Figure 10 – figure supplement 1.** Mutation of the cryptic targeting signal SRL in LDHBx blocks
1049 co-import of LDHA into peroxisomes. PTS1 was mutated by deletion of the leucine (ΔL) or
1050 substitution of RL by SI (SSI) blocks co-import of LDHA into peroxisomes. **(A,B)** Co-expression
1051 of YFP-LDHA with CFP-LDHBx[TGG, ΔL]. **(C,D)** Co-expression of YFP-LDHA with CFP-
1052 LDHBx[TGG, SSI]. **(B,D)** Cytosol was removed after permeabilization with digitonin. Bar 10 μm .

1053 **Figure 10 – figure supplement 2.** Tetrameric lactate dehydrogenase (space fill model) from
1054 human heart (LDH-M, all B form). The individual subunits are shown in different colors, and the
1055 last resolved amino acid at the C-terminus (Leu in position 331) of all subunits is shown in
1056 yellow. This structural model shows that the termini are protruding from the compact tetramer.
1057 The readthrough LDH contains at least one subunit LDHBx that is extended by seven amino
1058 acids containing the PTS1. For import into the peroxisome at least one C-terminus has to bind to
1059 the soluble PTS1 receptor PEX5. The structure also shows that the added PTS1 is unlikely to
1060 block oligomerization of the protein, because the C-termini are far away from the interaction
1061 surface of the protomers. The PTS1 has to be protruding from the compact oligomer in an
1062 unstructured manner to be buried in the TPR domain pocket of PEX5. Therefore it is also
1063 unlikely that the PTS1 extensions generated by readthrough fold back onto the protein to induce
1064 a conformational change that would interfere with the subunit interaction. The structural model is
1065 derived from structure 1IOZ (Read et al. 2001) in the International Protein Database
1066 (www.pdb.org) and was rendered using Jmol: an open-source Java viewer for chemical
1067 structures in 3D (<http://www.jmol.org/>).

1068
1069

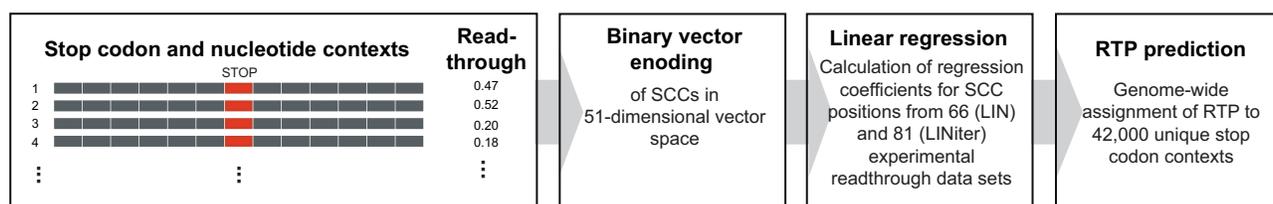
1070 **Titles to Supplementary files**

1071 **Supplementary file 1.** Plasmids used in this study.

1072 **Supplementary file 2.** Oligonucleotides used in this study.

1073

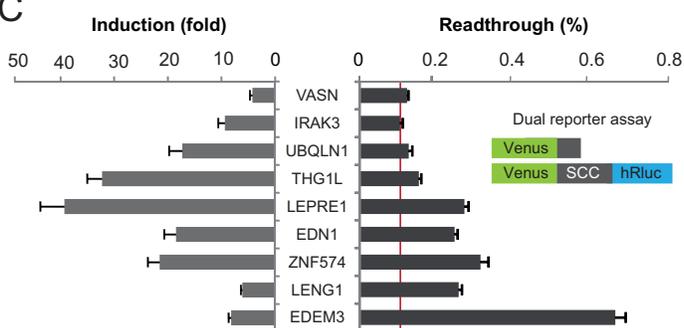
A

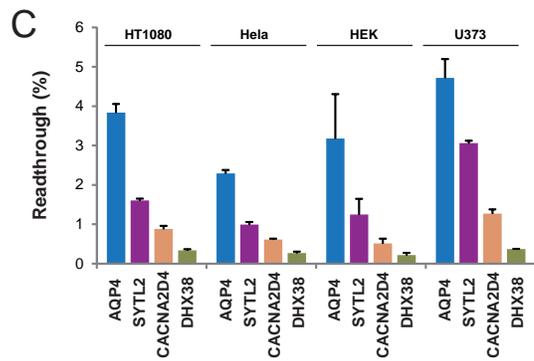
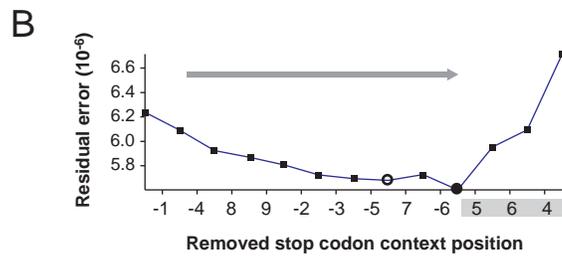
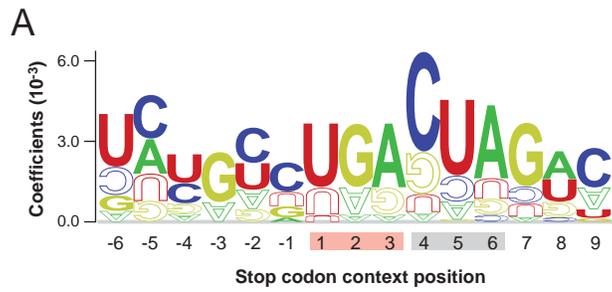


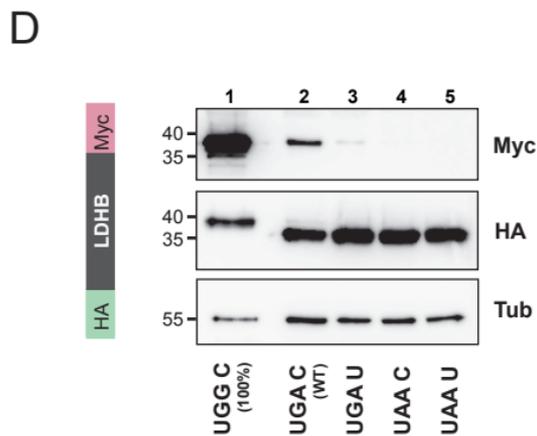
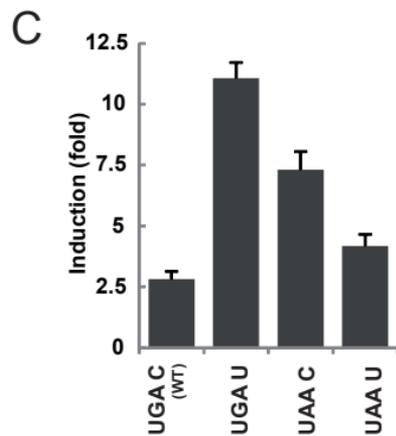
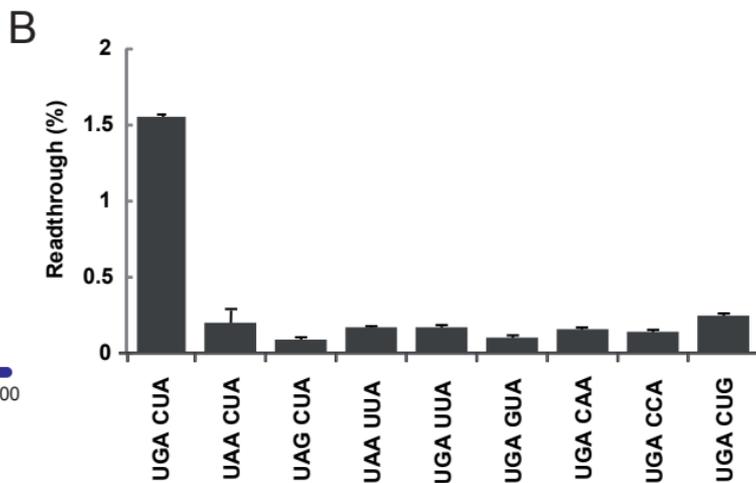
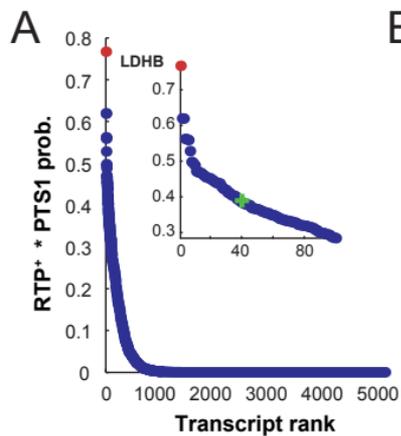
B

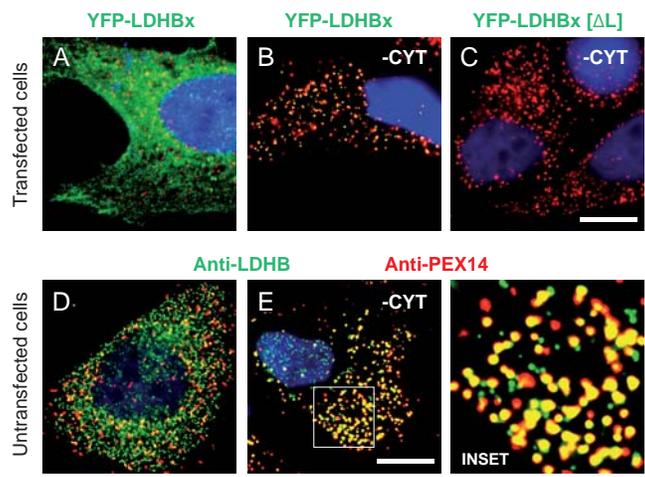
	Stop codon context (SCC)	RTP			
		LIN	iter	fs5	fs3
VASN	GCCCTACATC TAA GCCAGAGAGA	-0.10	0.02	0.00	-0.11
IRAK3	CAAAAAAGAA TAA ATTCTACCAG	0.14	0.09	0.05	0.20
UBQLN1	CCAGCCATCA TAG CAGCATTCT	0.02	0.13	0.05	0.11
THG1L	AGCCAGGCTT TGA CGGAAGAGTC	0.03	0.17	0.26	0.39
LEPRE1	GGATGAGCTA TGA CAGCGTCCAG	0.08	0.21	0.26	0.33
EDN1	AGCACATTGG TGA CAGACCTCG	0.17	0.23	0.24	0.33
ZNF574	GATCAGTGGC TGA CTCTGCCCGA	0.27	0.37	0.46	0.71
LENG1	CCTTACTCAC TGA CTCCTGAGGG	0.10	0.37	0.47	0.71
EDEM3	GGATGAGCTA TGA CTTGCTAAAC	0.15	0.43	0.63	0.65

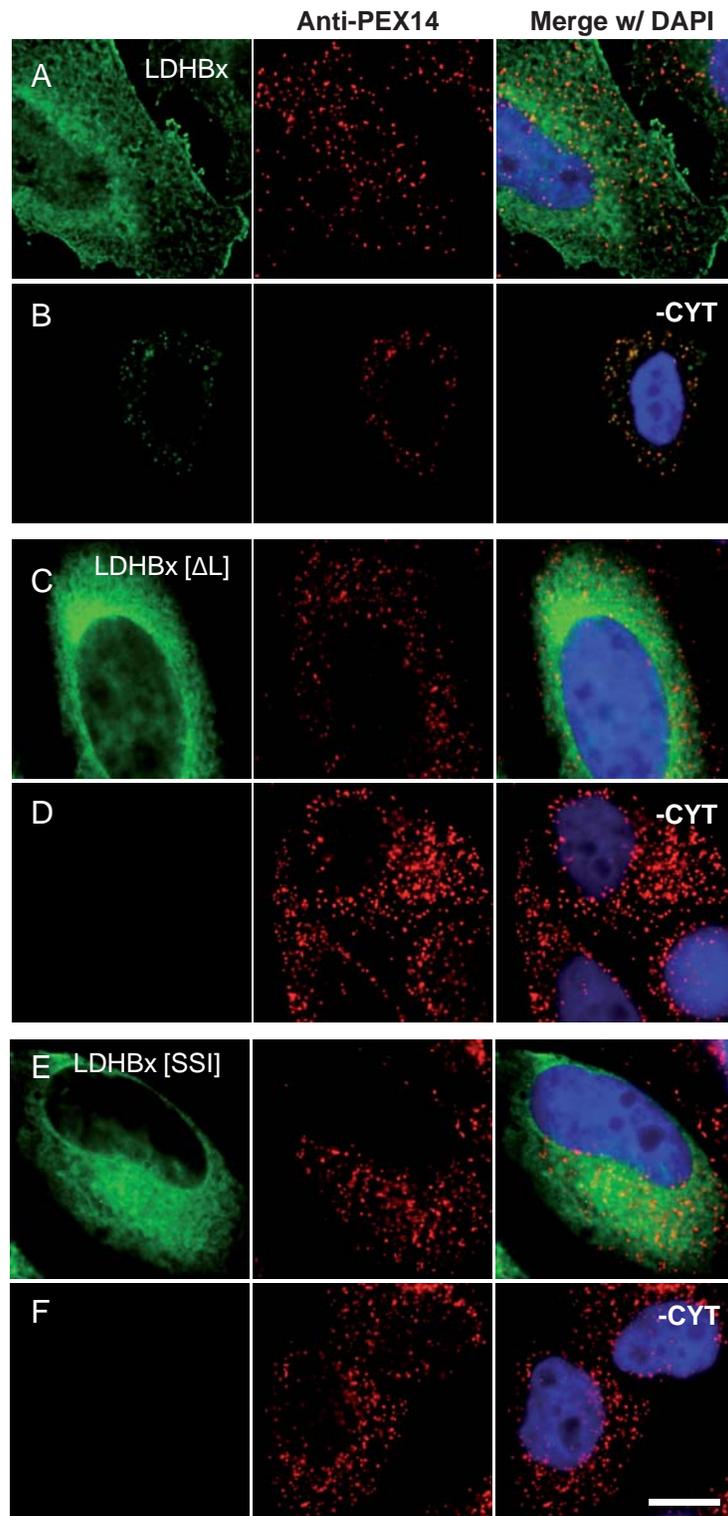
C

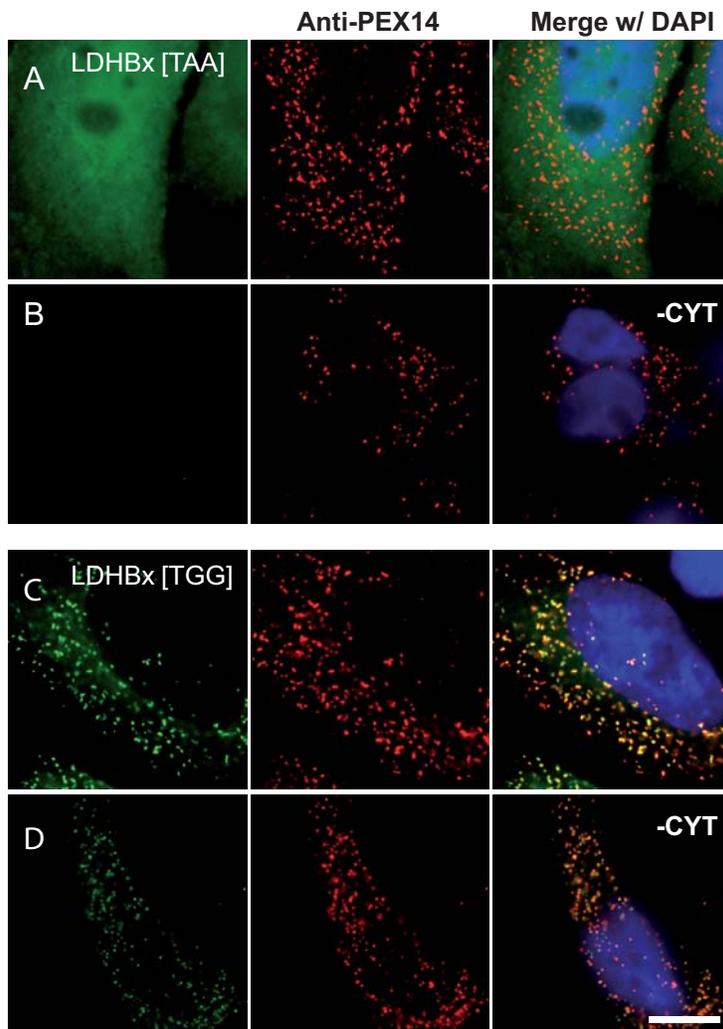


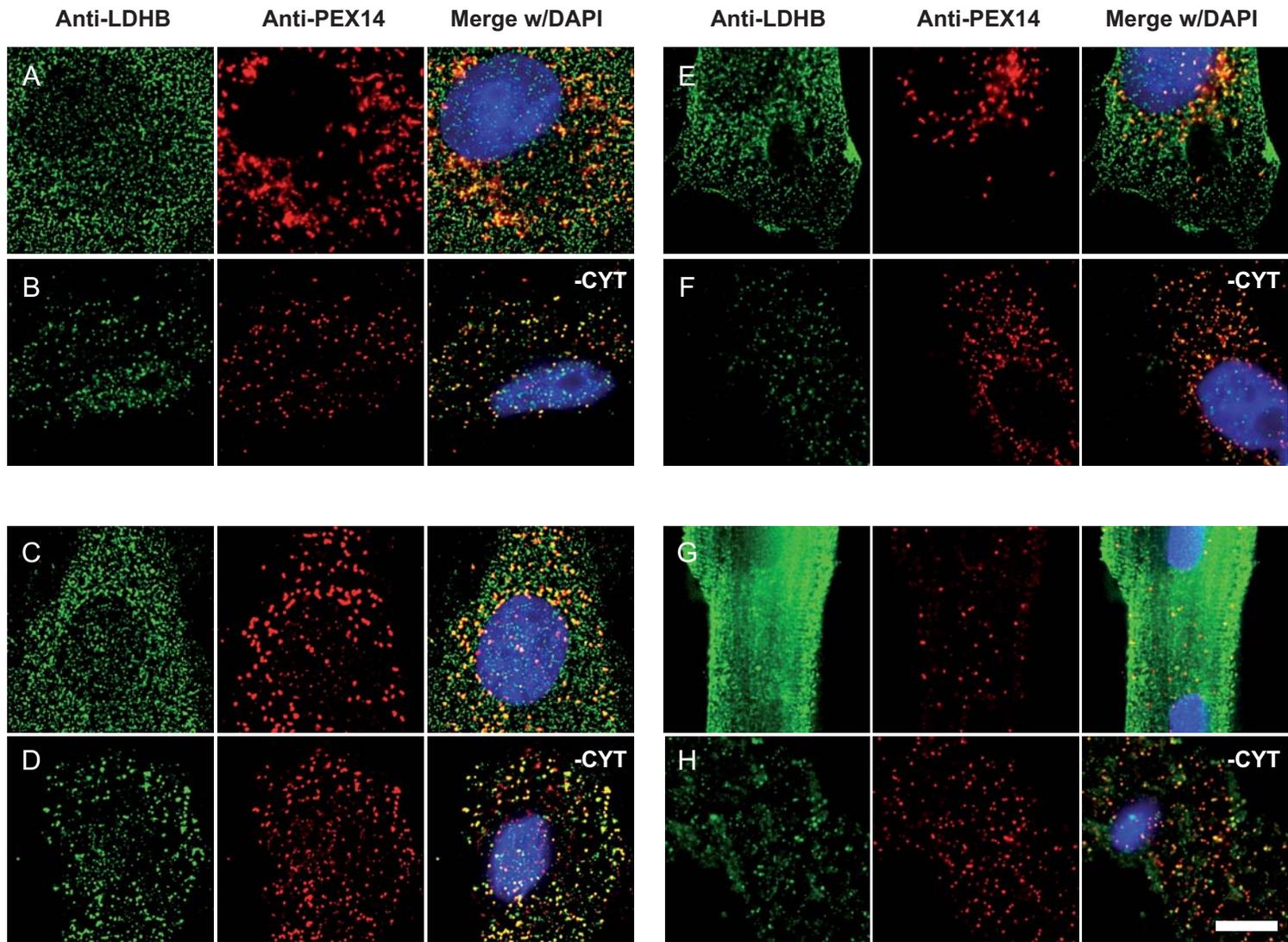




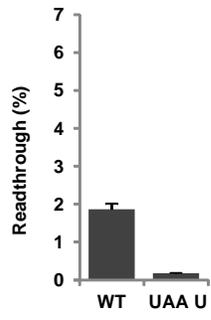




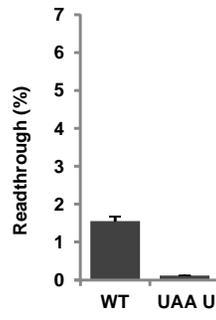




COS-7



HEK



U118

