



Published in final edited form as:

*DNA Res.* 2009 February ; 16(1): 45–58. doi:10.1093/dnares/dsn030.

## Database for mRNA half-life of 19,977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells

Lioudmila V. Sharova<sup>\*</sup>, Alexei A. Sharov<sup>\*</sup>, Timur Nedorezov, Yulan Piao, Nabeebi Shaik, and Minoru S.H. Ko

*Developmental Genomics and Aging Section, Laboratory of Genetics, National Institute on Aging, NIH, 251 Bayview Boulevard, Suite 100, Baltimore MD 21224, USA*

### Abstract

Degradation of mRNA is one of the key processes that control the steady state level of gene expression. However, the rate of mRNA decay for the majority of genes is not known. We successfully obtained the rate of mRNA decay for 19,977 non-redundant genes by microarray analysis of RNA samples obtained from mouse embryonic stem (ES) cells. Median estimated half-life was 7.1 hrs and only <100 genes, including *Prdm1*, *Myc*, *Gadd45g*, *Foxa2*, *Hes5*, and *Trib1*, showed half-life less than 1 hr. In general, mRNA species with short half-life were enriched among genes with regulatory functions (transcription factors), whereas mRNA species with long half-life were enriched among genes related to metabolism and structure (extracellular matrix, cytoskeleton). The stability of mRNAs correlated more significantly with the structural features of genes than the function of genes: mRNA stability showed the most significant positive correlation with the number of exon junctions per open reading frame (ORF) length, and negative correlation with the presence of PUF-binding motifs and AU-rich elements (AREs) in 3'-untranslated region (UTR) and CpG dinucleotides in the 5'-UTR. The mRNA decay rates presented in this report are the largest data set for mammals and the first for ES cells.

### Keywords

mRNA decay; mRNA degradation; microarray; transcript; exon junction; AU-rich elements (ARE); embryonic stem cells; Leukemia Inhibitory Factor (LIF); retinoic acids (RA); cell differentiation; mouse strain

## 1. Introduction

Post-transcriptional regulation of gene expression includes mRNA processing, splicing, editing, transport, stability, and translation. Degradation of mRNA is an important component of gene function which controls the steady state concentration of functional transcript levels in the cell<sup>1</sup>. For example, in human mRNA stability is estimated to control mRNA level of about 5–10% of all genes<sup>2</sup>. Quantification of this process is important for interpretation of gene-expression data obtained with microarrays as well as data from gene manipulation

Corresponding author: Minoru S. H. Ko, Developmental Genomics & Aging Section, Laboratory of Genetics, National Institute on Aging, IRP, NIH Biomedical Research Center, 251 Bayview Boulevard, Suite 100, Room 10C220, Baltimore, MD 21224, USA, Phone: 410-558-8359; Fax: 410-558-8331.

<sup>\*</sup>Equal contribution first authors.

experiments. Changes of gene expression may be delayed by several days if the corresponding transcript is stable. This may cause serious errors in reconstruction of gene regulatory networks.

Studies in several organisms revealed that majority of transcripts are stable<sup>3–5</sup>. It appeared that the specific half-life of each mRNA is precisely related to its physiological role<sup>4–8</sup>. For example, most housekeeping genes have long mRNA half-lives. In contrast, proteins that are required only for limited time in the cell, e.g., at a certain stage of the cell cycle, during development, growth or differentiation, or in response to external stimuli, often have mRNAs with short half-lives<sup>9</sup>. Transcriptionally inducible genes are disproportionately represented in the class of genes with rapid mRNA turnover<sup>5</sup>.

RNA stability can be altered in response to external stimuli such as hormones<sup>10</sup> or various types of stress. For example, response to UV-B exposure leads to stabilization of many short-lived mRNAs in mammalian cells<sup>11</sup>. Conversely, in yeast, mild heat shock promotes rapid degradation of mRNAs encoding the ribosomal proteins<sup>12</sup>, while glucose starvation, amino acid starvation, or sugar-induced osmotic stress causes stabilization of at least some yeast mRNAs<sup>13–15</sup>. Changes in RNA stability can also play an important role during embryonic development or cell differentiation. In plasmodium it has been shown that a major determinant of mRNA decay rate appears to be tightly linked to intra-erythrocytic development cycle, and to a lesser extent the functional category of the mRNAs themselves<sup>16</sup>. During PGC specification of *Oryzias latipes* (Japanese medaka) some specific RNA molecules are protected from degradation, thus establishing a specific mRNA expression pattern controlled by post-transcriptional mechanisms<sup>17</sup>. Expression of a wide variety of transcripts is controlled by changes in mRNA stability during neuronal development<sup>18, 19</sup>. During muscle differentiation mRNA half-lives for muscle specific genes - myogenin and myoD – have been shown to be the highest during differentiation, but declines when differentiation is completed<sup>20</sup>. Abnormal changes in RNA stability can be a cause of cell malfunction leading to cancer<sup>21, 22</sup> and other diseases like diabetic nephropathy<sup>22</sup>, muscular atrophy<sup>23</sup>, neurological disorders such as Alzheimer disease<sup>9</sup>.

Decay of mRNA is controlled by complex mechanisms that are not fully understood. This mechanism is integrated with other mRNA-related molecular processes including transcript elongation, splicing, polyadenylation, transport, and translation<sup>6, 9</sup>. RNA decay mechanisms include interaction between cis-acting elements and trans-acting factors. Major cis-factors affecting decay rates include AU-rich elements (ARE), iron-responsive elements (IRE), histone stem-loops, coding region determinants (CRD), Jun-kinase response elements (JRE), retained introns, and exon junctions<sup>24–26</sup>. Trans-factors that control mRNA decay include RNases which are often assembled into exosomes, RNA-binding proteins (poly-A binding, WD40 repeats, ARE-binding), siRNA, miRNA, piRNA. P-bodies in the cytoplasm can facilitate miRNA-mediated mRNA degradation induced by stress, and ribosomes control nonsense-mediated decay (NMD)<sup>27</sup>. Recent studies have shown that PUF-domain proteins can promote mRNA degradation of genes involved in development via Pop2-mediated deadenylation<sup>28</sup>. Binding motif of PUF protein PUM2 has been identified in mouse<sup>29</sup>.

Estimates of mRNA decay rates may depend on the method of measurement. Functional degradation, defined as decrease of the numbers of functional mRNA capable of producing correct protein, should be distinguished from physical degradation (or mass decay), which represents the decrease of mRNA numbers measured by PCR amplification or hybridization intensity<sup>30</sup>. Analysis of functional degradation is appropriate for individual genes, but it is not an option for high-throughput analysis targeted at the whole transcriptome. Thus, we will consider only physical degradation of mRNA with understanding that this measurement may not capture some aspects of mRNA function. The most widely used methods for high-throughput analysis of mRNA degradation is based on transcriptional inhibitors, for example

by actinomycin D, which inhibits transcription via binding to single-strand DNA at the transcription initiation complex and preventing elongation by RNA polymerase II<sup>31</sup>.

Multiple studies attempted to measure mRNA decay rates in human<sup>3, 5</sup>, but only one study has relatively large data of 5245 genes in hepatoma cells<sup>4</sup>. There are no data for mouse or any other mammals. Among eukaryotes, comparable data have been generated for plasmodium<sup>16</sup>, yeast<sup>7</sup>, and *Arabidopsis*<sup>32</sup>. To obtain baseline data for the mRNA half-life in mouse, we have selected mouse embryonic stem (ES) cells. ES cells are often used as experimental models for cell differentiation and for reconstruction of gene regulatory networks. Genes that control mRNA decay rates may also affect the differentiation of ES cells<sup>33</sup>. Despite increasing interest in the study of embryonic stem (ES) cells, there have been no attempts to measure mRNA decay rates in these cells. Here, we have described mRNA half-life of essentially all mouse genes in mouse ES cells.

## 2. Materials and Methods

### 2.1. Cell Culture

We used 2 mouse ESC lines: MC1 derived from mouse strain 129S6/SvEvTac, and MC2-B6 derived from strain C57BL/6J. These lines were purchased from The Transgenic Core Laboratory of the Johns Hopkins University School of Medicine (Baltimore, MD, USA). To remove contaminating feeder cells, ES cells were first cultured for 2 passages on gelatin-coated culture dish in complete ES medium (DMEM, 15% FBS; Leukemia Inhibitory Factor (LIF) (ESGRO) 1000 u/ml; 1mM sodium pyruvate; 0.1 mM NEAA, 2 mM glutamate, 0.1 mM beta-mercaptoethanol, and penicillin/streptomycin (50U/50 µg per ml). Cells were then seeded in gelatin coated 6-well plates at the density of  $1-2 \times 10^5$ /well ( $1-2 \times 10^4$ /cm<sup>2</sup>) and cultured for 3 days before the test for mRNA stability. Cells were incubated at 37 °C and 5% CO<sub>2</sub>. Medium was changed daily. Differentiated ES cells were obtained from MC1 cells by LIF withdrawal (complete medium without LIF, below referenced as LIF-) or by retinoic acid (RA) treatment (complete medium with 1µM RA, below referenced as RA) for 7 days before the test. To measure mRNA stability, transcription was blocked by adding actinomycin D (Sigma, cat # A-9415) to the medium at the concentration of 5 µg/ml. Cells were harvested at 0, 1, 2, 4, and 8 hr after addition of actinomycin D.

### 2.2. Microarray experiments

RNA was extracted using Phase lock gel™ (Eppendorf/Brinkman) columns according to the manufacturer's protocol after adding one ml/well of Trizol™ (Invitrogen) to the well. RNA was precipitated with isopropanol, washed with 70% ethanol, and dissolved in DEPC dH<sub>2</sub>O. RNA was hybridized to the NIA Mouse 44K Microarray v2.3 (whole genome 60-mer oligonucleotide probe; manufactured by Agilent Technologies, #014951)<sup>34</sup> in duplicate for each cell type and time point. Fluorescently labeled microarray targets were prepared from 2.5 µg aliquots of total RNA samples using a Low RNA Input Fluorescent Linear Amplification Kit (Agilent). A reference target (Cy5-CTP-labeled) was produced from Stratagene Universal Mouse Reference RNA (UMR), and all other targets were labeled with Cy3-CTP. Targets were purified using an RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol, and quantified on a NanoDrop scanning spectrophotometer (NanoDrop Technologies). All hybridizations were carried out by combining a Cy3-CTP-labeled experimental target and a Cy5-CTP-labeled UMR target. Microarrays were hybridized and washed according to Agilent protocol (G4140-90030; Agilent 60-mer oligonucleotide microarray processing protocol - SSC Wash, v1.0). Slides were scanned on an Agilent DNA Microarray Scanner, using standard settings, including automatic PMT adjustment.

### 2.3. Estimation of mRNA decay rates

Outliers in microarray data were detected using criterion of  $z > 7$  for log-transformed Cy3 and Cy5 data separately and removed (0.17% of the data). Then, Cy3 signals were normalized by Cy5 signals for the same probe in log scale, except for 4.34% of probes in which strong correlation between log Cy5 and Cy3 signals (slope of  $\log_{10}(\text{Cy5}/\text{Cy3}) > 0.25$ ) was likely a hybridization artifact (termed “chain effect”), because of low Cy5/Cy3 signal ratio ( $>3$  fold difference) and high variance ( $>0.05$ ) of  $\log_{10}(\text{Cy3})$  (Chain effect is described in the next section, 2.4). Degradation rate of mRNA was estimated using linear regression of log-transformed (base 10) signal intensity values  $y$  versus time  $t$ :

$$y = a - bt, \quad (\text{e1})$$

where  $t$  is time,  $b$  is the slope,  $a$  is intercept, and  $d = b/\ln(10)$  is instantaneous decay rate. The decrease of mRNA amount measured by microarray usually matched well with this model at all time points (Fig. 1A). However, for some mRNA species, the decay leveled off at some time point, which may be attributed to either heterogeneity in degradation rates (i.e., short-lived fraction degraded first) or non-specific hybridization to probes on the array (Fig. 1B). For these mRNA species we estimated decay rates after omitting the later time points to obtain the most accurate estimate of degradation rate. Time points were removed sequentially from the end of the time course till the lower confidence limit for  $b$  was maximized and the first time point ( $t = 0$ ) was within the confidence interval of the regression. Out of 32601 probes on the arrays, for which mRNA decay rates were estimated, 95.5% matched well to the exponent and all 5 time points were used for analysis; in 3.4, 0.9, and 0.2% cases decay rates were estimated using 4, 3, and 2 time points, respectively. The same number of time points was used for different cell types and culture conditions (MC2-B6-LIF+, MC1-LIF+, MC1-LIF-, and MC1-RA) to ensure proper comparison of mRNA degradation rates. Because the same amount of RNA was used for array hybridization, degradation of some mRNA species after block of transcription resulted in the increase of relative abundance of other stable mRNA species. Thus, additional correction was needed to account for global mRNA degradation. Yang et al.<sup>4</sup> used beta-actin for normalization assuming that it is very stable. However, *Actb* appeared not very stable in our experiments (half-life = 7.9 h). Thus, for global normalization we used 200 most stable non-redundant genes with average log intensity  $>2.5$  and for which decay rates were successfully measured for all 4 types of cells. Average mRNA degradation rate for these genes was estimated for each cell type using the equation (e1) (e.g.,  $d = -0.1012$  for undifferentiated MC1 cells) and then was subtracted from all estimated mRNA degradation rate for this cell type. Half-life of each mRNA species was estimated as  $H = \min [24, \ln(2)/d]$  for positive  $d$ , and  $H = 24$  h for negative  $d$ . We truncated half-life estimates by 24 h, because it is difficult to project mRNA half-life beyond this time point based on the experiment that lasted for 8 h. We also estimated the average decay rate of each gene in all examined cell types and the corresponding half-life. Statistical significance of the differential mRNA decay rates in various cell types was determined using False Discovery Rate (FDR  $< 0.05$ ), estimated from  $z$ -values:  $z = (d_1 - d_2) / \sqrt{SE_1^2 + SE_2^2}$ , where  $d_1$  and  $d_2$  are decay rates in cell types 1 and 2, respectively, and  $SE_1$  and  $SE_2$  are standard errors of these decay rates estimated from regression. FDR, which is necessary for multiple hypothesis testing, was estimated using the method of Benjamini and Hochberg<sup>35</sup>. In addition to FDR, we used the criterion of  $>2$  fold difference between half-lives to be significant.

### 2.4. Calibration of microarray signal intensities

Reliable estimation of mRNA degradation rates requires accurate scaling of expression values to ensure that the ratio of hybridization signal intensities is equal to the ratio of mRNA abundance uniformly across the entire range of values. We tested the scaling using a series of

dilutions (5x, 1x, 0.2x, and 0.04x) of the mRNA pool (obtained from newborn mice of C57BL/6J strain) labeled with Cy3 and hybridized together with the constant amount of the Universal Mouse Reference (UMR) sample labeled with Cy5. Regression of log-transformed Cy3 signals versus log RNA input was close to 1 for most probes with log signal >1.5 (Supplementary Figure S1), which indicates proper scaling. The results confirmed accurate calibration of intensity signals across the entire range of values.

Unexpectedly, Cy5 signals of some probes on the microarray also showed strong positive dependence on RNA input despite the fact that the amount of Cy5-labeled UMR was constant in all the arrays (Supplementary Figure S2). This effect was observed only when Cy3 signal was much stronger than Cy5 signal. The mechanism of this effect is unknown, but we hypothesize that it stems from the chain hybridization when free ends of Cy3-labeled mRNA bound to probes on the array can hybridize with Cy5-labeled mRNA either non-specifically or via rare antisense segments (termed “chain effect”). This effect may cause underestimation of gene expression difference between cell types if values are normalized by UMR signal. To avoid distortion, we did not use UMR for normalization for a small portion of probes where we expected chain effect.

Another challenge in interpretation of microarray measurements is that expression values often do not fall below some “background” level, which most likely results from non-specific hybridization. If a gene has low expression and is measured close to probe’s background level, it may show little change in expression value after actinomycin D treatment, which can be erroneously viewed as evidence of high mRNA stability. To avoid these errors, we estimated mRNA decay rates only for sets of data with expression values that were either high >2.5 in log<sub>10</sub> scale or substantially above the background <2 fold.

## 2.5. Microarray data analysis by Gene Ontology (GO) terms

Analysis of gene ontology (GO) terms in a selected list of genes was done using the hypergeometric distribution and  $FDR \leq 0.05$  criterion using NIA Mouse Gene Index (ver. mm8) software<sup>36</sup>. Only non-redundant genes with gene symbols were used for analysis. Because many GO categories are redundant (contain similar lists of genes), we adjusted the calculation of FDR in the following way. First, we identified all redundant pairs of GO terms that had a correlation of gene presence  $\geq 0.7$ . Second, in the list of GO terms ordered by increasing  $p$ -values,  $p_i$ , estimated from the hypergeometric distribution, GO term  $i$  was considered redundant, if it was redundant to at least one preceding term.

## 2.6. Real-time quantitative RT-PCR

Total RNAs were extracted from ES cells using Trizol™ (Invitrogen) and Phase lock gel™ (Eppendorf/Brinkman) columns according to the manufacturers’ protocols. RNAs were precipitated with isopropanol, washed with 70% ethanol, and dissolved in DEPC dH<sub>2</sub>O. Primers for quantitative reverse transcriptase PCR (qRT-PCR) were designed using the Vector NTI Advance 9.1 software (Invitrogen) and tested for SYBR Green chemistry using an established in-house protocol<sup>37</sup>. Reactions were run on the ABI 7500 Sequence Detection Systems using the default cycling program and data were processed using SDS 2.2 software (Applied Biosystems). Gene expression levels were normalized to a Cyclin D3 as an internal control and to total RNA amount.

# 3. Results and Discussion

## 3.1. Measuring mRNA decay rates in mouse ES cells

To quantify mRNA decay rates in mouse ES cells we measured changes in gene expression with whole-genome microarrays in a time-course (0, 1, 2, 4, and 8 h) after treating cells with

actinomycin D. Experiments were done for 2 ES cell lines - MC1 (129S6/SvEvTac) and MC2-B6 (C57BL/6J) cultured in the standard condition in the presence of LIF. To increase the number of genes that can be assayed, we also carried out the analysis of MC1 ES cells undergoing differentiation into two different culture conditions – in the absence of LIF and in the presence of RA for 7 days. Expression data for all genes are available in Supplementary Table S1. Proper calibration of signal intensities in the full range of gene expression values was confirmed using hybridization of different RNA amounts to microarrays (see details in Methods). Expression values for most genes were normalized using Universal Mouse Reference (UMR), except for 1824 probes (4.3%) where normalization was not possible because UMR signal was low and unstable (see details in the section 2.4). Decay rates were estimated using linear regression of log-transformed data (Fig. 1A, B) and normalized by average expression change of 200 most stable genes (see details in the section 2.4; Supplementary Table S1).

Because the earlier work did not fully address the data reproducibility issue, where analysis was limited to estimating statistical (technical) error and validation of expression measurements by PCR for a small number of genes, we felt that it was important to evaluate the reproducibility of the measurements. To this end, we used the following two methods: (1) comparing mRNA decay rates in two independent experiments with different ES cell lines (MC1 and MC2-B6) cultured in the standard conditions, and (2) comparing mRNA decay rates estimated using different oligonucleotide probes on the microarray for the main transcript of the same gene. The first method yielded very high consistency between estimates of mRNA degradation rate ( $r = 0.925$ ), indicating biological reproducibility of our database (Fig. 1C). The second method gave somewhat lower match ( $r = 0.744$ ), but the correlation was still within the range expected from the comparison of gene expression data obtained with different array platforms (Fig. 1D).

Discrepancy between mRNA half-lives estimated using different oligonucleotide probes for the same gene was likely related to the biological complexity of mRNA processing rather than to the error in measurement. For example, probes located in 3' untranslated region (UTR) often showed shorter mRNA half-lives than probes located in the open reading frame (ORF) (Supplementary Figure S3A). This may be caused by either alternative termination of transcription coupled with a differential degradation rate or active mRNA truncation at some early step of mRNA processing. Interestingly, some genes of histone H1 family had the opposite pattern of degradation: probes located in the short 3'UTR showed longer half-lives than probes located in the ORF (Supplementary Figure S3B). However, because of a high sequence similarity between histone genes and a partial cross-hybridization of probes, this effect may be an artifact.

To validate the microarray data, we carried out quantitative RT-PCR analysis and found that qRT-PCR results matched well to those obtained by DNA microarrays for 6 out of 7 selected genes (Supplementary Figure S4). An exception was *Sox17*, whose degradation rates measured by qRT-PCR were much lower than those measured by microarrays. The closer inspection of probe locations revealed that all microarray probes for *Sox17* were located in the 3'-UTR of the gene, whereas qRT-PCR primers were located in the ORF. This result was, therefore, consistent with differential degradation rates of *Sox21* measured with probes located in 3'UTR and ORF (Supplementary Figure S3A).

### 3.2. Effect of alternative transcription and splicing on estimation of RNA stability

It is conceivable that the influence of probe locations on the estimation of mRNA stability goes beyond the aforementioned simple difference between 3'-UTR and ORF. For example, oligonucleotide probes recognized different transcripts, which are produced by alternative transcription start sites (TSS), splicing, and polyA signals, may lead to the estimation of

different half-lives for the same gene. To determine the effect of alternative transcription and splicing on mRNA stability we compared decay rates identified with the best probe from the main transcript of each gene with decay rates estimated using other probes that matched the same gene. Probes from alternatively spliced transcripts often showed shorter mRNA half-lives compared to the best probe in the main transcript (Fig. 1E). This effect was strongest for probes located in the retained introns of alternative transcripts. Similar effect of alternative splicing was reported in *Arabidopsis*; however, it was not statistically significant because of the small number of alternatively spliced forms captured by the array<sup>32</sup>. Thus, we, for the first time, statistically confirm that alternative transcripts, especially those with retained introns, have shorter mRNA half-life compared to main transcripts.

This phenomenon can be simply explained by the possibility that alternative transcripts have a higher degradation rate due to specific sequences located in retained introns or alternatively spliced exons. Alternatively, it is also possible that the degradation of alternative transcripts appears to be accelerated by the splicing that may continue even after the blocking of transcription by actinomycin D. Because RNAs were extracted from whole cells in our experiments, it may contain heterogeneous nuclear RNA with some retained introns, which are able to hybridize with intron-matching probes in the array. If splicing continues for a while after the blocking of transcription by actinomycin D, the number of heterogeneous nuclear RNA would decrease. Further experiments are needed to distinguish between the two possible mechanisms. All subsequent analysis was done only for main transcripts of each gene. We would like to note that for 1391 genes, mRNA decay rates were not consistent between different microarray probes (> 2-fold difference). Therefore, mRNA half-lives for these genes estimated with the best oligonucleotide should be considered with caution (marked by an asterisk in Supplementary Table S2).

### 3.3. General trends in stability of mRNAs in ES cells

We were able to estimate mRNA half-lives of 19977 non-redundant genes (Supplementary Table S2). For 14815 genes, estimates are available for two ES cell lines and three culture conditions. We were not able to estimate mRNA half-lives of 7260 genes due to the low level of expression in ES cells and those of 541 genes due to the lack of acceptable probes in the main transcript. Most of the genes had long half-lives, which was in concordance with published data<sup>3-5</sup>. In our analysis, median estimated half-life for all genes was 7.1 hrs (Fig. 1F), which was less than reported for human hepatoma cells (10 h)<sup>4</sup>; however, this difference may be related to normalization method, as discussed below (see the section 3.9). In our analysis, only 54 genes (including *Prdm1*, *Myc*, *Gadd45g*, *Foxa2*, *Hes5*, and *Trib1*) showed a half-life shorter than 1 hr. Published data indicate that more genes showed RNA half-life shorter than 30 min<sup>3</sup>. Therefore, it is possible that the number of genes with very short mRNA life may be underestimated in our study, because we did not collect samples at the time points earlier than 1 hr.

### 3.4. Functional annotation of stable and unstable transcripts

We analyzed over representation of Gene Ontology (GO) terms in the 610 genes with half-life shorter than 2 hrs and in 4310 genes with half-life longer than 12 hrs to identify functional categories of genes with unstable and stable mRNA, respectively (Supplementary Tables S3 and S4). Using GO database and additional information on protein domains we manually assembled lists of genes with functions that were characteristic to stable and unstable mRNA species (Supplementary Table S2). The major functional categories of genes with unstable mRNA were “Regulation of transcription”, “Cell cycle”, “Apoptosis”, “Signal transduction”, and “Development”, whereas the major functional categories of genes with stable mRNA were “Metabolism”, “Protein biosynthesis”, “Extracellular matrix” and “Cytoskeleton” (Fig. 1G). Most of these results confirm findings published earlier<sup>4, 5, 32</sup>; however, overrepresentation

of “Extracellular matrix” and “Cytoskeleton” groups among stable mRNA species is reported here for the first time. Many transcription factors showed very short mRNA half-life, which may be important for the quick expression changes of target genes. Genes with half-life from 1 to 2 hr ( $N=557$ ) included early response and cell cycle regulation and also some genes important for pluripotency and early development: for example *Sox2*, *Klf4*, *Foxd3*, *Sox7*, *Nodal*, *Ctgf*, *Id1*, *Id2*, *Id3*, and Wnt signaling.

### 3.5. Strain-dependent stability of mRNA

In the previous study, we have shown that there are genes differentially expressed between ES cells derived from 129 mouse strain and ES cells derived from C57BL/6 mouse strain (strain-specific signature genes)<sup>38</sup>. We, therefore, asked if the observed differences were related to differences in mRNA half-life of corresponding genes. We found that a small number of genes ( $N = 92$ ) had statistically significant differences in mRNA decay rates between MC1 (129S6/SvEvTac) and MC2-B6 (C57BL/6J) ES cells (FDR <0.05 and more than 2 fold change of half-life) (Supplementary Table S5) and only 10 of these genes overlapped with strain-specific signature genes. Although differential gene expression of at least some strain-specific genes can be explained partially by differences in mRNA stability, we were not able to identify correlation between differential gene expression and mRNA stability between strains on a global scale.

### 3.6. Alteration of mRNA half-life in differentiating ES cells

Differentiation of ES cells in the absence of LIF or in the presence of RA caused changes of mRNA stability in a large set of genes ( $N = 3332$  in LIF<sup>-</sup> and  $N = 2257$  in RA) (Fig. 2A, B, Supplementary Tables S6, S7). These changes were much stronger than the strain difference in mRNA stability between MC1 and MC2-B6 ES cells (Fig. 1C). In general, decay rates of mRNA increased after LIF withdrawal: the median half-life declined from 7.07 h to 5.48 h; and out of 3332 genes with significant change of mRNA stability, 77.9% had increased decay rates compared to undifferentiated ES cells (Fig. 2A, C). In contrast, differentiation of ES cells in RA conditions caused a general increase of mRNA stability: the median half-life increased from 7.07 h to 8.60 h; and out of 2257 genes with significant change of mRNA stability, 87.1% had decreased decay rates compared to undifferentiated ES cells (Fig. 2B, C). This indicates a substantial difference between two protocols of ES cell differentiation. But despite these global differences, gene-specific changes of mRNA stability in these differentiating ES cells showed a good correlation: the change of decay rates of mRNA (in log scale) in RA conditions was positively correlated with the change of decay rates (in log scale) in LIF<sup>-</sup> conditions ( $r = 0.593$ ) (Supplementary Figure. S5A). This means that changes of mRNA decay rates can be viewed as a superposition of two effects: (1) gene-specific effect which is consistent between LIF<sup>-</sup> and RA conditions, and (2) non-specific effect associated with the culture medium which caused general decrease of stability in LIF<sup>-</sup> conditions and general increase of mRNA stability in RA conditions.

To select genes whose mRNA were either destabilized or stabilized in differentiated cells both in LIF<sup>-</sup> and RA conditions we used one standard deviation cut-off of log decay rates (Supplementary Figure S5A). Genes with increased rate of mRNA decay upon differentiation were enriched in GO terms “transcription regulation”, “nervous system development”, “apoptosis”, “ubiquitin cycle”, “ribosome” (mostly mitochondrial ribosomal proteins), “chromatin binding” and “methyltransferase activity” (Supplementary Table S8). Genes with decreased rate of mRNA decay upon differentiation were enriched in GO terms “extracellular space”, “extracellular matrix”, “calcium ion binding”, “actin cytoskeleton”, and “lysosome” (Supplementary Table S9). During the ES cell differentiation, unstable mRNA species became even less stable, whereas stable mRNA species became more stable (Supplementary Figure S5B).



It is believed that in general the differentiation of ES cells is associated with destabilization of pluripotency-related genes and stabilization of lineage-specific genes. Indeed, our data showed that mRNAs of some pluripotency-related genes (*Sall4*, *Eed*, *Esrrb*, *Notch4*, *Mras*, *Tbx3*, *Btbd14b*, and *Sap30*) were destabilized upon differentiation (Fig. 2D). Interestingly, the strongest increase in mRNA decay rate was observed for *Btbd14b* (*Nac1*), which is a known cofactor of NANOG and POU5F1 proteins<sup>39, 40</sup>. Similarly, mRNAs of some differentiation-related genes (e.g., *Bmp1*, *Acta1*, *Igfbp1*, *Igfbp4*, *Igfbp6*, *Frzb*, *Cited1*, *Flt1*, *Slit2*, *Krt8*, *Krt14*, *Cnn1*, *Mest*, *Wnt3*, and *Amot*) were stabilized. On the other hand, the change of mRNA stability for all genes had no correlation with the change of gene expression levels (Supplementary Figure S5C, D), indicating that mRNA stability has no global role in controlling gene expression changes during ES cells differentiation.

One way to control mRNA degradation is to modulate the expression levels of particular ribonucleases. In metazoans, it is known that the modulation of the expression of ribonucleases and/or associated factors play an important role in the control of mRNA stability during development<sup>6</sup>. It has also been shown that many ribonucleases and associated factors are differentially regulated during development. For example, in *Drosophila*, the expression of ribonucleases *pacman/XRN1*, *twister/SKI2*, and *tazman/DIS3/RRP44*, varies during development at both RNA and protein levels<sup>41–43</sup>. We, therefore, focused on genes that are involved in the mRNA degradation or binding and found that 103 genes changed their expression at >2-fold between undifferentiated and differentiated ES cells (Supplementary Table S10). The multiplicity of mRNA degradation pathways makes it difficult to relate expression of these genes with observed patterns of mRNA degradation. However, we noted that the expression levels of *Lsm3*, *Lsm5*, *Lsm6*, *Lsm7*, *Lsm12*, *Lsm1*, *Cnot4*, *Cnot10*, *Dcp2*, *Exosc1*, which are known to promote mRNA degradation, decreased in the differentiated ES cells, especially in the RA condition. This may explain increased stability of mRNA in cells differentiated in the RA condition. In contrast, degradation-promoting genes - *Zfp361l* and *Cnot6l* - showed increased expression in differentiated ES cells, which may have contributed to lower mRNA stability in the LIF- condition, but it does not explain increased mRNA stability in the RA condition.

### 3.7. Multivariate analysis of gene structural features that affect mRNA stability

Decay rates of mRNA vary substantially between genes (Fig. 1F); however, little is known about factors and mechanisms responsible for these differences. Two major structural factors that affect mRNA stability are known: (i) the presence of introns in genes (i.e., multi-exon genes), which makes mRNA more stable<sup>44</sup>; and (ii) the presence of AU-rich elements (AREs) in the 3'UTR of genes, which makes mRNA less stable<sup>45, 46</sup>. However, all the previous results have been obtained using single-factor analysis, which may easily lead to erroneous conclusions if factors are correlated. We, therefore, used multivariate regression analysis, which helps to distinguish direct causal effects from mere correlations. We incorporated as many structural characteristics of genes as possible, including the length of open reading frame (ORF), the length of 5'UTR, the length of 3'UTR, the number of exons, the occurrence of different kinds of ARE, the presence of CpG di-nucleotides, and the presence of PUF binding motifs in all parts of the transcript (Supplementary Table S11). AREs were detected by using three types of reported consensus sequences: AUUUA (ARE1), UUAUUUAWW (ARE2), WWWUAUUUAUWWW (ARE3)<sup>45, 47</sup>, and non-specific A/U 12-mers with no more than one mismatch (ARE4). We used UGUANAUA as PUF consensus motifs<sup>29, 48, 49</sup>. CpG elements were added to the list after we unexpectedly found that they can affect mRNA stability. We did not analyze the effects of miRNA on the mRNA stability, because it would require extensive data mining and integration with existing miRNA databases.

Multiple regression analysis showed that mRNA stability was correlated with positively with the number of exons and negatively with the ORF length, number of ARE2, ARE4 and PUF in 3'UTR, number of ARE4 and PUF in ORF, and number of CpG di-nucleotides in 5'UTR (Supplementary Table S11). The effect of ORF length was nearly equal to the effect of the number of exons with the opposite sign, which indicated that the number of exon junctions per unit ORF length may be a better predictor of mRNA stability than the number of exons per gene. This is also supported by the fact that most exon junctions are located in the ORF region of genes. Indeed, log-transformed rate of mRNA decay correlated stronger with log-transformed number of exon junctions per unit ORF length ( $r = -0.326$ ) than with log-transformed number of exons ( $r = -0.111$ ). Therefore, below we always used the ratio of exon junctions per 1 Kb of ORF length as a predictor of mRNA stability.

The final form of regression after removing non-significant terms for MC1 cells in the standard culture conditions (LIF+) was

$$y = -0.592 - 0.179 \cdot x_1 - 0.020 \cdot x_2 + 0.054 \cdot x_3 + 0.022 \cdot x_4 + 0.040 \cdot x_5 + 0.051 \cdot x_6 + 0.071 \cdot x_7 + 0.044 \cdot x_8 \quad (\text{e2})$$

where  $y$  is log-transformed rate of mRNA degradation,  $\log_{10}(b + 0.1)$ ;  $x_1$  is log-transformed number of exon junctions per 1 Kb ORF length;  $x_2$  is log-transformed ORF length;  $x_3$  is log-transformed number of ARE4 in 3'UTR;  $x_4$  is log-transformed number of ARE4 in ORF; and  $x_5$  is log-transformed number of ARE2 in 3'UTR;  $x_6$  is log-transformed number of CpG di-nucleotides in 5'UTR;  $x_7$  is log-transformed number of PUF motifs in 3'UTR; and  $x_8$  is log-transformed PUF motifs in ORF. Log-transformation was used because it increased the predictive power of the regression and made all regression coefficients dimensionless and comparable with each other;  $\log_{10}(n + 1)$  was used to handle zero counts as arguments. This regression accounted for 19.8% of variation in mRNA degradation rates, which was only slightly lower than the regression with all 19 factors (20.0%). The equation (e2) was not precise enough to predict mRNA degradation rates for individual genes, but it worked well to estimate the average degradation rate in groups of genes identified in microarray experiments.

According to the regression equation (e2), we conclude that the number of exon junctions per unit ORF length shows the strongest and the most consistent effect on mRNA stability between cells types. This conclusion is further supported by the alternative data presentation shown in Fig. 3A. The proportion of genes with  $\geq 5$  exon junctions per unit ORF length was higher in genes with longer mRNA half-life, whereas the proportion of genes associated with the destabilizing factors (CpG di-nucleotides in 5'UTR, and ARE4 and PUF in 3'UTR) was higher in genes with shorter mRNA half-life (Fig. 3A). Although the role of introns in stability of mRNA has been reported both in experimental and statistical studies<sup>44</sup>, here we present the first evidence that the number of exon junctions per unit ORF length is the most important known structural factor that increases mRNA stability. After intron is excised by spliceosome, the exon junction location is marked by binding of the exon junction complex (EJC), which consists of at least 10 proteins<sup>50</sup>. EJC remains bound to mRNA after transport to cytoplasm, can stimulate translation, and mediate nonsense-mediated decay (NMD) of mRNA during the pioneer round of translation<sup>50</sup>. It is indeed possible that some components of EJC can remain bound to mRNA after the first round of translation and decrease the rate of mRNA degradation<sup>44</sup>. But the mechanism of this process remains unknown.

According to the regression equation (e2), the second strongest factor affecting the mRNA stability was the number of PUF motifs in 3'UTR, which decreased the mRNA stability. The third strongest factor affecting the mRNA stability was the presence of AREs. Unexpectedly,

the sequence non-specific ARE4 in 3'UTR (to a smaller extent, in ORF) decreased mRNA half-life more strongly than the consensus-matching AREs. Among different consensus matching AREs, only ARE2 in 3'UTR showed a significant effect on mRNA stability. Several mechanisms are known to be involved in ARE-mediated mRNA degradation. For example, ARE-binding proteins can recruit PARN or exosomes<sup>51</sup>. It is also known that the stability of mRNA can be modified by immune factors via AREs in 3'UTR<sup>52</sup>.

The fourth strongest factor that affects mRNA stability was the number of CpG di-nucleotides in 5'UTR. Because this was not known before, this is indeed the first report about the effect of CpG di-nucleotides on the mRNA stability. We suspected that the occurrence of CpG di-nucleotides in 5'UTR may be correlated with the presence of CpG islands at transcription start site (TSS). We, therefore, estimated the number of CpG di-nucleotides near TSS (from -400 to +600 bp) and included it into the regression analysis. We found that the numbers of CpG di-nucleotides in 5'UTR appeared to have a 5-fold stronger effect on mRNA stability than their numbers in the CpG island adjacent to TSS (Supplementary Table S12), which indicates that CpG di-nucleotides need to be located within 5'UTR to affect mRNA decay. The mechanism of this effect is unknown, but considering a strong effect of splicing on mRNA stability we can hypothesize that it may be mediated by methyl CpG-binding proteins which regulate splicing<sup>53</sup>.

We have, thus far, shown that both functional categories of genes (Section 3.4) and the structural features of genes (this Section) affect the mRNA stability. It will be, thus, interesting to see which factors affect mRNA stability more strongly. To this end, we carried out the regression analysis after including both structural features and functional categories (Fig. 3B, Supplementary Table S13). Among functional groups of genes, the strongest destabilizing effect was shown in the category of "Regulation of transcription" followed by "Signal transduction" and "Cell cycle". In contrast, "Metabolism", "Extracellular matrix", and "Cytoskeleton" functional categories increased mRNA stability, which is consistent with our previous results (Fig. 1G). However, we found that all the effects of structural factors listed in the equation (e2) remained strong, statistically significant, and the sign of regression coefficients did not change after adding functional gene categories. This indicates that the stability of mRNAs is more significantly correlated with the structural features of genes than the function of genes.

Furthermore, the differential mRNA stability in various functional groups of genes can be partially explained by structural features of genes. For example, low stability of mRNA among genes in the functional category of "Regulation of transcription" and "Development" can be explained by the low number of exon junctions per ORF length and high number of ARE4 and PUF in 3'UTR and CpG in 5'UTR (Fig. 3C). Similarly, high stability of mRNA among genes in the category of "Metabolism" can be explained by the high number of exon junctions per ORF length and low number of ARE4 and PUF in 3'UTR and CpG in 5'UTR (Fig. 3C). However, several functional categories of genes (e.g., "Regulation of transcription" or "Extracellular matrix") had a significant effect on mRNA stability based on regression (Fig. 3B). This indicates that the effect of gene function cannot be fully explained by structural characteristics of transcripts included in the regression.

As mentioned above, differentiation of ES cells in both LIF- and RA conditions caused changes of mRNA stability in a large set of genes (Fig. 3A, B, Supplementary Tables S6, S7). To identify structural and functional factors that affected these global changes in mRNA stability we compared coefficients of multiple regression that related decay rates (log scale) with various factors in undifferentiated and differentiated ES cells in LIF- and RA conditions (Fig. 3B, Supplementary Table S13). Degradation-promoting effects of CpG di-nucleotides in 5'UTR and ARE in 3'UTR appeared much stronger in cells differentiated in the LIF- condition

compared to undifferentiated cells and cells differentiated in the RA condition. Cells differentiated in the RA condition had weaker degradation-promoting effects of ARE4 in 3' UTR than in undifferentiated cells. These effects may explain the increased mRNA decay rates in ES cells differentiated in LIF<sup>-</sup> and the decreased decay rates in ES cells differentiated in the RA condition. Protection of mRNA from degradation by exon junctions was stronger in both types of differentiated ES cells than in undifferentiated cells.

### 3.9. Comparison with human data

Interspecies comparison of mRNA decay rates has been done only among distant evolutionary lineages (plant, yeast, and human). The new mouse data set provides an opportunity to compare two mammalian species: mouse and human. Using Homologene database<sup>54</sup> we identified 4153 pairs of orthologous genes for which mRNA rates were estimated both in mouse (this study) and human (Supplementary Table S3). For compatibility with our data we converted estimates of mRNA decay rates in human from  $\log_2$  to  $\log_e$ . Because human data are normalized using beta-actin, whose mRNA is not highly stable (based on our data), a large set of human genes (11.6%) shows negative degradation rates<sup>4</sup>. We, therefore, re-normalized estimates of human mRNA decay rates using the method we used for mouse (by average decay rate of 200 genes with most stable mRNA). Median rate of mRNA decay for all orthologous genes in mouse ( $d = 0.098 \text{ h}^{-1}$ ) was lower than in human ( $d = 0.137 \text{ h}^{-1}$ ). Because these experiments differed not just in species (i.e., mouse vs. human), but also in cell types (ES cells vs. hepatoma), and were done in different laboratories using different microarray platforms, it was impossible to attribute this difference to any specific factor. However, correlation between mRNA decay rates in mouse and human was high ( $r = 0.610$ , Fig. 4), which was only slightly lower than correlation between estimates of mRNA decay rates obtained with different probes that matched the same mouse transcript ( $r = 0.744$ , see the section 3.1). This finding indicates that pathways of mRNA degradation are mostly conserved between mammalian species. Interestingly, mRNA decay rates in human hepatoma were more similar to decay rates in differentiated ES cells ( $r = 0.605$  for LIF<sup>-</sup> and  $r = 0.616$  for RA) than to decay rates in undifferentiated ES cells ( $r = 0.534$ , SE = 0.013,  $p = 0.0001$ ). This may be explained by the fact that hepatoma cells are differentiated, and thus, the profile of mRNA decay rate in hepatoma matched better with that in differentiated ES cells. However, this similarity needs to be confirmed with further experiments that use the same cell types in both species.

## 4. Conclusion

In this paper we present a whole-genome analysis of mRNA decay rates in mouse ES cells. Decay rates were reliably measured for 19,977 non-redundant genes from time course of microarray data in undifferentiated ES cells originated from different mouse strains and in differentiated ES cells obtained by LIF withdrawal (LIF<sup>-</sup>) or retinoic acid (RA) treatment. The novelty of this study can be summarized as follows: (1) This is the first study of mRNA decay rates in mouse; (2) This is the first study of mRNA decay rates in ES cells of any mammalian species; (3) We have analyzed the difference in mRNA decay rates between cells derived from two different mouse strains and between undifferentiated and differentiated cells; (4) The set of genes for which mRNA decay rates estimated is 3.8 times larger than in the most extensive earlier study for human; (5) Close match between data from different cell lines has confirmed high reproducibility of mRNA decay rate estimates; (6) The microarrays have been calibrated; (7) We have applied multivariate statistics for the analysis of factors affecting mRNA decay rates, which identifies direct effects and distinguishes them from mere correlations; (8) We have explored the effect of various structural features of transcripts and gene function on mRNA stability in both general and during ES cell differentiation. The major result is that the number of exon junctions per ORF is the strongest factor promoting mRNA stability, whereas decay is promoted by CpG di-nucleotides in 5'UTR region as well as by non-canonical ARE

(without specific consensus sequence) in 3'UTR and in ORF. We believe that the presented data on mRNA decay rates provides a valuable resource for future experiments in interpreting various gene manipulation and gene expression studies as well as the computational modeling and simulation on gene regulatory networks.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

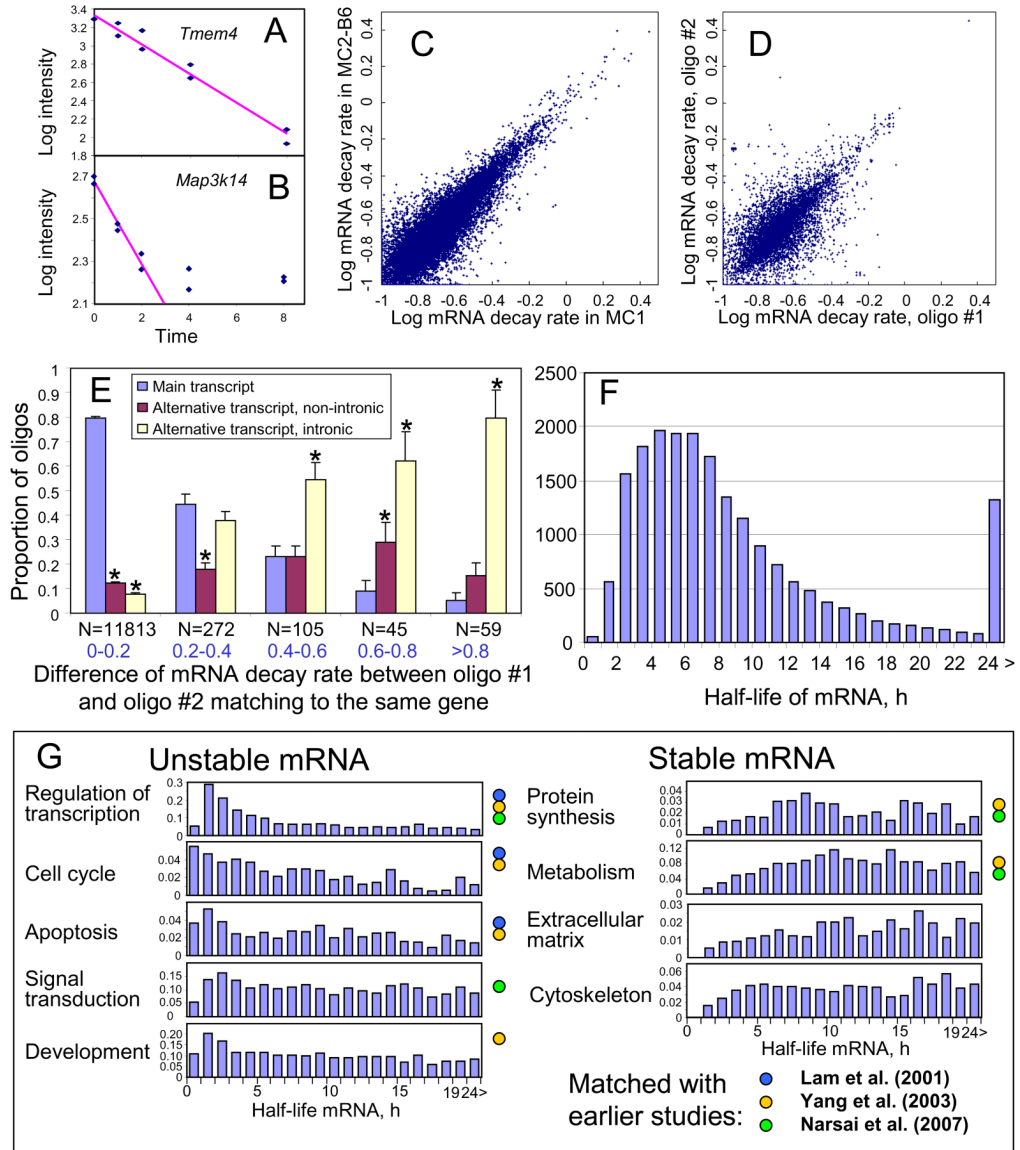
The work was supported by the Intramural Research Program of the National Institute on Aging, NIH (Z01 AG000656).

## References

1. Raghavan A, Bohjanen PR. Microarray-based analyses of mRNA decay in the regulation of mammalian gene expression. *Brief Funct Genomic Proteomic* 2004;3:112–124. [PubMed: 15355594]
2. Bolognani F, Perrone-Bizzozero NI. RNA-protein interactions and control of mRNA stability in neurons. *J Neurosci Res* 2008;86:481–489. [PubMed: 17853436]
3. Raghavan A, Ogilvie RL, Reilly C, et al. Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Res* 2002;30:5529–5538. [PubMed: 12490721]
4. Yang E, van Nimwegen E, Zavolan M, et al. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res* 2003;13:1863–1872. [PubMed: 12902380]
5. Lam LT, Pickeral OK, Peng AC, et al. Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biol* 2001;2:RESEARCH0041
6. Newbury SF. Control of mRNA stability in eukaryotes. *Biochem Soc Trans* 2006;34:30–34. [PubMed: 16246172]
7. Garcia-Martinez J, Aranda A, Perez-Ortin JE. Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell* 2004;15:303–313. [PubMed: 15260981]
8. Wang J, Chang YF, Hamilton JI, Wilkinson MF. Nonsense-associated altered splicing: a frame-dependent response distinct from nonsense-mediated decay. *Mol Cell* 2002;10:951–957. [PubMed: 12419238]
9. Hollams EM, Giles KM, Thomson AM, Leedman PJ. mRNA stability and the control of gene expression: implications for human disease. *Neurochem Res* 2002;27:957–980. [PubMed: 12462398]
10. Sahmi M, Nicola ES, Price CA. Hormonal regulation of cytochrome P450 aromatase mRNA stability in non-luteinizing bovine granulosa cells in vitro. *J Endocrinol* 2006;190:107–115. [PubMed: 16837615]
11. Gowrishankar G, Winzen R, Bollig F, et al. Inhibition of mRNA deadenylation and degradation by ultraviolet light. *Biol Chem* 2005;386:1287–1293. [PubMed: 16336123]
12. Herruer MH, Mager WH, Raue HA, Vreken P, Wilms E, Planta RJ. Mild temperature shock affects transcription of yeast ribosomal protein genes as well as the stability of their mRNAs. *Nucleic Acids Res* 1988;16:7917–7929. [PubMed: 3047675]
13. Greatrix BW, van Vuuren HJ. Expression of the HXT13, HXT15 and HXT17 genes in *Saccharomyces cerevisiae* and stabilization of the HXT1 gene transcript by sugar-induced osmotic stress. *Curr Genet* 2006;49:205–217. [PubMed: 16397765]
14. Benard L. Inhibition of 5' to 3' mRNA degradation under stress conditions in *Saccharomyces cerevisiae*: from GCN4 to MET16. *Rna* 2004;10:458–468. [PubMed: 14970391]
15. Jona G, Choder M, Gileadi O. Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast. *Biochim Biophys Acta* 2000;1491:37–48. [PubMed: 10760568]

16. Shock JL, Fischer KF, DeRisi JL. Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol* 2007;8:R134. [PubMed: 17612404]
17. Herpin A, Rohr S, Riedel D, Kluever N, Raz E, Scharl M. Specification of primordial germ cells in medaka (*Oryzias latipes*). *BMC Dev Biol* 2007;7:3. [PubMed: 17217535]
18. Deschenes-Furry J, Perrone-Bizzozero N, Jasmin BJ. The RNA-binding protein HuD: a regulator of neuronal differentiation, maintenance and plasticity. *Bioessays* 2006;28:822–833. [PubMed: 16927307]
19. Ratti A, Fallini C, Cova L, et al. A role for the ELAV RNA-binding proteins in neural stem cells: stabilization of Msi1 mRNA. *J Cell Sci* 2006;119:1442–1452. [PubMed: 16554442]
20. Figueroa C, Vojtek AB. Akt negatively regulates translation of the ternary complex factor Elk-1. *Oncogene* 2003;22:5554–5561. [PubMed: 12944902]
21. Champelovier P, Pautre V, Elatifi M, et al. Resistance to phorbol ester-induced differentiation in human myeloid leukemia cells: a hypothetic role for the mRNA stabilization process. *Leuk Res* 2006;30:1407–1416. [PubMed: 16757024]
22. Benjamin D, Moroni C. mRNA stability and cancer: an emerging link? *Expert Opin Biol Ther* 2007;7:1515–1529. [PubMed: 17916044]
23. Heier CR, Gogliotti RG, DiDonato CJ. SMN transcript stability: could modulation of messenger RNA degradation provide a novel therapy for spinal muscular atrophy? *J Child Neurol* 2007;22:1013–1018. [PubMed: 17761657]
24. Stoecklin G, Mayo T, Anderson P. ARE-mRNA degradation requires the 5′–3′ decay pathway. *EMBO Rep* 2006;7:72–77. [PubMed: 16299471]
25. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *Embo J* 2001;20:4987–4997. [PubMed: 11532962]
26. Le Hir H, Izaurralde E, Maquat LE, Moore MJ. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *Embo J* 2000;19:6860–6869. [PubMed: 11118221]
27. Yamasaki S, Stoecklin G, Kedersha N, Simarro M, Anderson P. T-cell intracellular antigen-1 (TIA-1)-induced translational silencing promotes the decay of selected mRNAs. *J Biol Chem* 2007;282:30070–30077. [PubMed: 17711853]
28. Goldstrohm AC, Hook BA, Seay DJ, Wickens M. PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol* 2006;13:533–539. [PubMed: 16715093]
29. White EK, Moore-Jarrett T, Ruley HE. PUM2, a novel murine puf protein, and its consensus RNA-binding site. *Rna* 2001;7:1855–1866. [PubMed: 11780640]
30. Kennell D. Processing endoribonucleases and mRNA degradation in bacteria. *J Bacteriol* 2002;184:4645–4657. [PubMed: 12169587]discussion 4665
31. Sobell HM. Actinomycin and DNA transcription. *Proc Natl Acad Sci U S A* 1985;82:5328–5331. [PubMed: 2410919]
32. Narsai R, Howell KA, Millar AH, O’Toole N, Small I, Whelan J. Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell* 2007;19:3418–3436. [PubMed: 18024567]
33. Wegmuller D, Raineri I, Gross B, Oakeley EJ, Moroni C. A cassette system to study embryonic stem cell differentiation by inducible RNA interference. *Stem Cells* 2007;25:1178–1185. [PubMed: 17218405]
34. Carter MG, Sharov AA, VanBuren V, et al. Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. *Genome Biol* 2005;6:R61. [PubMed: 15998450]
35. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of Royal Statistical Society, B* 1995;57:289–300.
36. Sharov AA, Dudekula DB, Ko MS. Genome-wide assembly and analysis of alternative transcripts in mouse. *Genome Res* 2005;15:748–754. [PubMed: 15867436]
37. Carter MG, Hamatani T, Sharov AA, et al. In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling. *Genome Res* 2003;13:1011–1021. [PubMed: 12727912]

38. Sharova LV, Sharov AA, Piao Y, et al. Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains. *Dev Biol* 2007;307:446–459. [PubMed: 17560561]
39. Wang J, Rao S, Chu J, et al. A protein interaction network for pluripotency of embryonic stem cells. *Nature* 2006;444:364–368. [PubMed: 17093407]
40. Wang J, Lévassieur DN, Orkin SH. Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* 2008;105:6326–6331. [PubMed: 18436640]
41. Cairrao F, Arraiano C, Newbury S. *Drosophila* gene tazman, an orthologue of the yeast exosome component Rrp44p/Dis3, is differentially expressed during development. *Dev Dyn* 2005;232:733–737. [PubMed: 15704111]
42. Seago JE, Chernukhin IV, Newbury SF. The *Drosophila* gene twister, an orthologue of the yeast helicase SKI2, is differentially expressed during development. *Mech Dev* 2001;106:137–141. [PubMed: 11472843]
43. Till DD, Linz B, Seago JE, et al. Identification and developmental expression of a 5′–3′ exoribonuclease from *Drosophila melanogaster*. *Mech Dev* 1998;79:51–55. [PubMed: 10349620]
44. Zhao C, Hamilton T. Introns regulate the rate of unstable mRNA decay. *J Biol Chem* 2007;282:20230–20237. [PubMed: 17478421]
45. Chen CY, Shyu AB. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci* 1995;20:465–470. [PubMed: 8578590]
46. Shaw G, Kamen R. A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 1986;46:659–667. [PubMed: 3488815]
47. Bakheet T, Frevel M, Williams BR, Greer W, Khabar KS. ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins. *Nucleic Acids Res* 2001;29:246–254. [PubMed: 11125104]
48. Xie X, Lu J, Kulbokas EJ, et al. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* 2005;434:338–345. [PubMed: 15735639]
49. Cora D, Di Cunto F, Caselle M, Provero P. Identification of candidate regulatory sequences in mammalian 3′ UTRs by statistical analysis of oligonucleotide distributions. *BMC Bioinformatics* 2007;8:174. [PubMed: 17524134]
50. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 2007;76:51–74. [PubMed: 17352659]
51. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* 2007;8:113–126. [PubMed: 17245413]
52. Tebo J, Der S, Frevel M, Khabar KS, Williams BR, Hamilton TA. Heterogeneity in control of mRNA stability by AU-rich elements. *J Biol Chem* 2003;278:12085–12093. [PubMed: 12556523]
53. Young JI, Hong EP, Castle JC, et al. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci U S A* 2005;102:17551–17558. [PubMed: 16251272]
54. Wheeler DL, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2008;36:D13–21. [PubMed: 18045790]

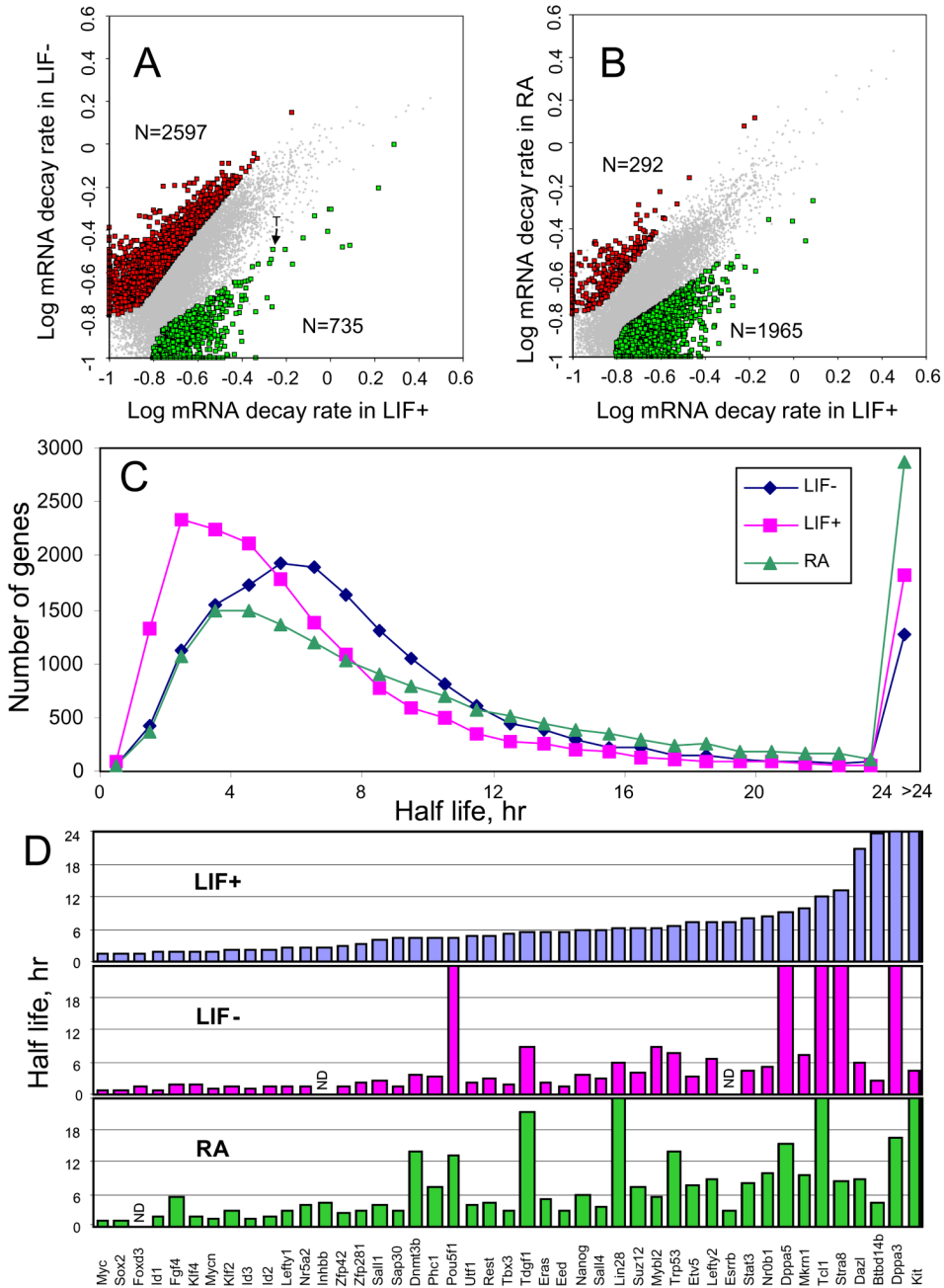


**Figure 1. Stability of mRNA in mouse ES cells**

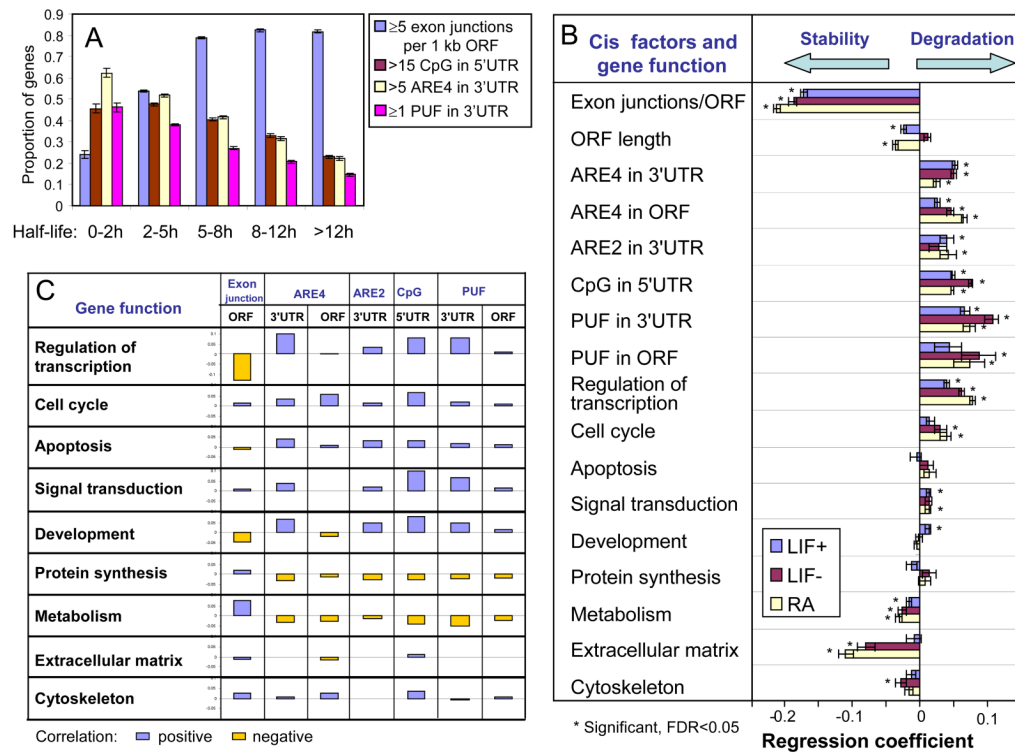
(A) (B) Regression of gene expression level (microarray signal intensity in log scale) versus time after suppressing transcription by actinomycin D ( $y = a + bx$ ) is used to estimate the rate of mRNA decay,  $d = b/\ln(10)$ . All 10 data points are used for an example (A) and only 6 data points are used for an example (B). (C) Comparison between mRNA decay rates estimated for MC1 ES cells and those estimated for MC2-B6 ES cells. Rates are log-transformed,  $\log_{10}(d + 0.1)$ . (D) Comparison between mRNA decay rates estimated using different probes (#1 and #2) on the microarray, which match to the main transcript of the same gene. Only the data for MC1 cells are shown. (E) Comparison of mRNA decay rates identified with the best probe (#1) from the main transcript of each gene with decay rates estimated using other probe (#2) that matched the same gene; pairs of probes were first classified into groups according to the difference in estimated mRNA decay rate (group size is shown below bars,  $N$ ), and then the proportion of probes #2 that matched different types of transcripts and/or introns was estimated and plotted for each group. (F) Frequency distribution of mRNA half-life among all genes. (G) Frequency distribution of mRNA half-life in different functional categories of



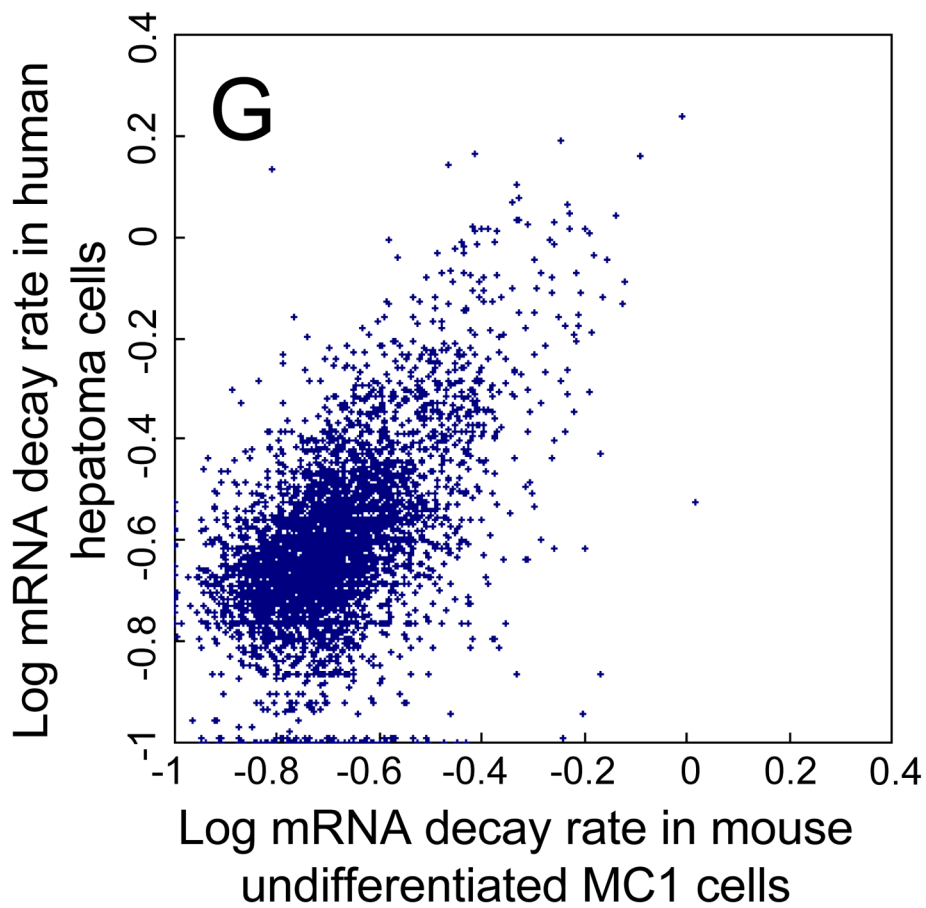
genes; color circles indicate that the functional group was previously identified as being over-represented among stable or unstable mRNA species.



**Figure 2. Change in mRNA decay rates in ES cells after differentiation**  
 (A) Change of mRNA decay rates after 7 days of differentiation in LIF- conditions; significant changes (FDR < 0.05 and >2 fold change of half-life) is shown by red and green dots; rates are log-transformed,  $\log_{10}(d+0.1)$ . (B) Change of mRNA decay rates after 7 days of differentiation in RA conditions. (C) Frequency distribution of mRNA half-life in undifferentiated MC1 cells (LIF+) compared to cells differentiated after LIF withdrawal (LIF-) and retinoic acid treatment (RA). (D) Half-life of mRNA of genes associated with pluripotent state of ES cells in undifferentiated cells (LIF+) and upon differentiation (LIF- and RA).



**Figure 3. Effect of structural and functional characteristics of genes on the rate of mRNA decay** (A) Proportion of genes with increased number of exon junctions, CpG di-nucleotides in 5' UTR, and ARE4 and PUF motifs in 3'UTR in groups of genes with different range of mRNA half-life. (B) Regression coefficients for the effect of structural elements (cis-factors) and gene function on the mRNA decay rate; abundance of structural elements and mRNA decay rates are log-transformed. (C) Correlation between functional characteristics of genes and abundance of structural elements (log-transformed).



**Figure 4. Comparison of mRNA decay rates for orthologous genes in mouse ES cells and in human hepatocytes (from the reference<sup>4</sup>)**

Rates are log-transformed,  $\log_{10}(d + 0.1)$ . Human mRNA decay rates were re-normalized using average decay rate of 200 most unstable mRNA species for proper comparison.