Published in final edited form as:

Mol Genet Metab. 2007 June ; 91(2): 148–156.

Pleiotropic Effects and Compensation Mechanisms Determine Tissue Specificity in Mitochondrial Myopathy and Sideroblastic Anemia (MLASA)

Yelena Bykhovskaya¹, Emebet Mengesha¹, and Nathan Fischel-Ghodsian¹

IMedical Genetics Institute, Ahmanson Department of Pediatrics, Steven Spielberg Pediatric Research Center, Cedars-Sinai Medical Center and David Geffen School of Medicine at UCLA, Los Angeles, California, USA

Abstract

The tissue specificity of mitochondrial diseases is poorly understood. Recently, tissue-specific quantitative differences of the components of the mitochondrial translation system have been found to correlate with disease presentation in fatal hepatopathy caused by mutations in mitochondrial translation factor EFG1. MLASA is an autosomal recessive inherited progressive oxidative phosphorylation disorder that affects muscle and erythroid cells. The disease is caused by the homozygous point mutation C656T (R116W) in the catalytic domain of the *pseudouridylate synthase 1 (PUS1)* gene, which leads to a complete lack of pseudouridylation at the expected sites in mitochondrial and cytoplasmic tRNAs. Despite the presence of these altered tRNAs, most tissues are unaffected, and even in muscle and erythroid cells the disease phenotype only slowly emerges over the course of years. In order to elucidate intracellular pathways through which the homozygous mutation leads to tissue-restricted phenotype, we performed microarray expression analysis of EBV-transformed lymphoblasts from MLASA patients, heterozygous parents, and controls using human Beadchip microarray with 47,296 transcripts. Genes coding for proteins involved in DNA transcription and its regulation, and metal binding proteins, demonstrated major differences in expression between patients and all other individuals with normal phenotype. Genes coding for ribosomal proteins differed significantly between individual with at least one copy of the mutated *PUS1* gene and controls. These findings indicate that the lack of tRNA pseudouridylation can be overcome by compensatory changes in levels of ribosomal proteins, and that the disease phenotype in affected tissues is likely due to pleiotropic effects of PUS1p on non-tRNA molecules involved in DNA transcription and iron metabolism. Similar combinations of mechanisms may play a role in the tissue specificity of other mitochondrial disorders.

Keywords

Oxidative phosphorylation disorder; pseudouridylation; expression; microarray; sideroblastic anemia; myopathy

Correspondence should be addressed to Dr. Nathan Fischel-Ghodsian, Department of Pediatrics, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Suite SSB378, Los Angeles, CA 90048; Telephone (310) 423-6392; Fax: (310) 423-0620; Email: nathan.fischel@cshs.org

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Mitochondrial myopathy and sideroblastic anemia (MLASA, MIM 600462) is a progressive oxidative phosphorylation (OXPHOS) disorder that affects muscle and erythroid cells [1]. Hallmark features of MLASA include progressive exercise intolerance during childhood, onset of sideroblastic anemia around adolescence, basal lactic acidemia, and mitochondrial myopathy. Individuals with MLASA frequently become dependent on blood transfusions. A molecular basis of the disorder was recently identified to be a point mutation (C656T) leading to the arginine to tryptophan change at position 116 in the catalytic domain of the *pseudouridylate synthase 1* gene (*PUS1*, MIM 608109) [2]. All affected individuals were found to be homozygous for the C656T (R116W) change, all the parents to be heterozygous, and all unaffected siblings to be either heterozygous carriers or negative for the C656T mutation. Biochemical studies revealed that individuals homozygous for the mutated allele of *PUS1* show absent or greatly reduced tRNA pseudouridylation at specific sites of specific cytoplasmic and mitochondrial tRNAs in EBV-transformed lymphoblasts [3]. Despite the presence of these altered tRNAs, most tissues are unaffected, and even in muscle and erythroid cells the disease phenotype only slowly emerges over the course of years.

The tissue specificity of OXPHOS disorders has been well documented but poorly understood [4,5]. Recently, advances in understanding some of the aspects of tissue specificity came from the investigation of patients with a fatal hepatopathy due to mutations in the mitochondrial translation elongation factor EFG1 [6]. The synthesis of OXPHOS complexes paralleled the levels of the mutant EFG1 protein (undetectable in liver) and was also dependent on the ratio of other translation elongation factors EFTu and EFTs (decreased from 1:1 to 1:4 in liver). This study provides the first documentation that differences among tissues in the organization of the mitochondrial translation system and its response to dysfunction by altered expression of components of this system, may be responsible for the phenomenon of tissue specificity.

We hypothesized that the mechanisms for tissue specificity in MLASA may be twofold. A compensatory response occurs in patients, and to a lesser extent in heterozygote carriers, which compensates for the absence of tRNA pseudouridylation at PUS1p controlled sites. In muscle and erythroid cells, however, PUS1p may have pleiotropic effects on non-tRNA molecules that have no compensatory response available. Such additional targets of PUS1p have recently been described [7,8], and for one of them a direct relationship to iron metabolism exists (8, and see Discussion). Separately, in the only other human disease linked to a pseudouridylation defect, dyskeratosis congenita, the mutated protein dyskerin (coded by *DKC1* gene, MIM 300126) also has multiple functions, including pseudouridylation of specific uridine residues of ribosomal RNA [9], and binding and modifying small nucleolar and telomerase RNA [10]. Similar to MLASA, dyskeratosis congenita is a tissue-restricted progressive disease appearing in early childhood and leading to bone marrow failure [11].

To evaluate the potential validity of our hypotheses regarding the compensatory and pleiotropic effects associated with the *PUS1* mutation in MLASA, we performed global RNA expression profiling of EBV-transformed lymphoblasts. These cells were the only tissue available from patients and their parents, and transcriptional profiling of lymphoblastoid cell lines have led to new insights into possible pathogenic pathways in such diverse diseases as familial combined hyperlipidemia (FCHL, 12), autism [13], and rheumatoid arthritis [14]. In the autosomal recessive disease Ataxia telangiectasia, significant differences were uncovered by comparing expression profiles of lymphoblasts of heterozygous carriers to normal control individuals [15].

We used Illumina's Sentrix Human-6 Expression Beadchips, each containing six arrays with 47,296 unique probes, to assess global gene expression in MLASA patients, their parents and

related and unrelated controls. Differential expression analysis revealed that genes coding for proteins involved in DNA transcription and its regulation, and metal binding proteins are the ones with major differences in expression between EBV-transformed lymphoblasts of MLASA patients ('Affected') and all other individuals with no phenotypic presentation ('Unaffected'). Differential expression analysis between individuals with at least one copy of abnormal *PUS1* gene (patients homozygous for the C656T change and their heterozygous parents, 'Mutation Carriers') and homozygous wild-type controls ('Controls' group) revealed highly significant differences between expression levels of transcripts coding for ribosomal proteins. This provides preliminary data to suggest that the lack of pseudouridylation of mitochondrial and cytoplasmic tRNAs can be overcome by compensatory changes in levels of ribosomal proteins, and that the disease phenotype in affected tissues may be caused by pleiotropic effects of PUS1p on other molecules involved in DNA transcription and iron metabolism. The pathways of MLASA pathogenesis suggested by the results of this study in EBV-transformed lymphoblasts will be investigated in erythroid, muscle, and other tissues from *PUS1*-knockout mice to further elucidate mechanisms of tissue-restriction in MLASA.

Materials and Methods

Patients, Cell lines, RNA Isolation

We analyzed samples from two affected brothers and their two affected female cousins homozygous for the MLASA – associated C656T mutation in the *PUS1* gene; four parents – heterozygous carriers of the C656T mutation, and an unaffected sister carrying wild-type genotype at the *PUS1* gene (Figure 1). In addition, two females and one male with normal hearing from an Arab-Israeli family with nonsyndromic deafness carrying wild-type *PUS1* sequence were used as controls.

Lymphoblastoid cell lines have previously been established from peripheral blood lymphocytes of study participants [1] and immortalized with Epstein-Barr virus [16]. Cells were grown in suspension in T flasks in RPMI-1640 medium (Invitrogen, Inc., Carlsbad, CA), containing 2 mmol/L L-glutamine, 100 ug/ml streptomycin, and 10% fetal calf serum. Total RNA was extracted using Trizol reagent (Invitrogen, Inc.) according to the manufacturer's protocol.

The study was approved by the Institutional Review Board and informed consent was obtained from the families.

cRNA amplification

cRNA amplification and labeling with biotin were performed using Illumina® TotalPrep RNA amplification kit, manufactured by Ambion, Inc (Austin, TX) according to the manufacturer's protocol using 100 ng of total RNA as input material. cRNA yields were quantified with NanoDrop® spectrophotometer (NanoDrop Technologies, Inc.). 1ug of labeled cRNAs were hybridized to the Beadchip arrays.

Expression microarray

Three Sentrix Human-6 (WG-6) Expression Beadchips, each containing six arrays, were purchased from Illumina, Inc. Each chip contains six identical sets of 47,296 unique probes, including 24,385 RefSeq annotated genes, known alternative splice regions, putative transcripts predicted by Gnomon (<http://web.ncbi.nlm.nih.gov/genome/guide/gnomon.html>), and from Unigene clusters (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>).

All the reagents and equipment used for hybridization of cRNA samples to the beadchips were purchased from Illumina, Inc. cRNA was hybridized to arrays for 16 hours at 55°C before

being washed and stained with streptavidin-Cy3 according to the manufacturer's protocol. Beadchips were scanned on the Illumina BeadArray Reader confocal scanner for 90 minutes according to the manufacturer's instructions.

Data analysis

Our complete data set had been deposited in the Gene Expression Omnibus (GEO) database (accession # GSE6374) at the GEO website <http://www.ncbi.nlm.nih.gov/geo/>.

Data visualization, clustering, and differential analysis were carried out with BeadStudio software package (Illumina Inc., San Diego, CA) and GeneSpring GX computer program, release 7.3. (Silicon Genetics, Inc., San Jose, CA)

Signal quality control

Gene signals were ranked relative to the distribution of signals of the negative controls and detection scores were calculated by the BeadStudio program.

Signal normalization

BeadStudio—With the Illumina Bead technology, a single hybridization of RNA from one cell line sample to an array produces approximately 30 intensity values for each bead type, i.e. transcript. These background-corrected values for a single bead type were summarized, averaged, and normalized using average method of normalization.

GeneSpring GX 7.3—Raw microarray data was output from BeadStudio without background subtraction and directly downloaded into the Genespring GX computer program. Signal data was filtered based on detection score of 0.99 or above in at least three of the samples. Default data transformation to set measurements below 0.01 to 0.01 was applied. Default settings were used for experiment normalization which included chip normalization to 50th percentile and gene normalization to the median.

Differential expression analysis

Grouping samples—Group of 'Affected' consisted of four affected children from two nuclear families homozygous for C656T mutation. RNA sample from patient Affected 4 was duplicated on the same chip, RNA from Affected 7 individual was duplicated on different chips. Group of 'Parents' included four parents carrying C656T mutation in the heterozygous form. 'Controls' (or wild-type) group included four independent samples, consisting of the unaffected child Control 8 duplicated on two chips, and three Arab-Israeli individuals. Group of 'Unaffected' included individuals from both Parents and Controls groups, i.e. wild-type homozygotes and heterozygotes for the C656T mutation. Group of 'Mutation Carriers' included individuals with at least one copy of C656T mutation – both groups of Parents and Affected.

BeadStudio—Pair-wise differential expression analysis was done comparing reference group to so-called condition group. The comparisons were done using T-test with the assumption of equal variance error model. Differential score (Diff score) was calculated as $-\log(p\text{-value}) \times 10$. Diff score of 13 corresponds to the p-value of 0.05, diff score of 20 to the p-value of 0.01, diff score of 20 to the p-value of 0.001. Positive Diff Score represents up-regulation, negative - down-regulation.

GeneSpring GX 7.3—2-sided Student T-test was used to compare gene expression levels between the two groups of samples. One-Way ANOVA testing was performed for

simultaneous comparison of three groups of samples. P-value cut off value of 0.05 and no multiple testing correction was applied. Equal variance was assumed in the analysis.

Quantitative RT-PCR

Gene-specific TaqMan® Gene Expression Assays were purchased from Applied Biosystems, Inc., Foster City, CA for confirmation of microarray results. We included assay for human *glyceraldehyde-3-phosphate dehydrogenase* gene (*GAPDH*, MIM 138400) as a control. All PCR reactions were performed using TaqMan universal PCR master mix (Applied Biosystems) according to the manufacturer's instructions. The 7900HT Sequence Detection System (Applied Biosystems) was used to collect real-time fluorescent signal. The threshold cycle (C_T), was recorded for each individual RNA sample. To ensure accurate quantification, reactions were performed in triplicates.

Results

Samples correlation

Global pair-wise expression comparisons performed using BeadStudio revealed an exceptionally high degree of similarity between duplicated samples. Linear correlation coefficient (r^2) of 0.9976 was calculated for sample Affected4 duplicated on the same chip; 0.9941 for sample Affected7 and 0.9971 for sample Control8 both duplicated on different chips. Correlation coefficient of 0.9879 was calculated between 'Controls' and 'Affected' groups, 0.9882 between 'Controls' and 'Parents', and 0.9948 between 'Affected' and 'Parents'.

Principal Component Variance Analysis

To try to unveil the most significant pattern(s) in our dataset we performed principal component variance analysis using GeneSpring GX 7.3. We used expression patterns of 12,983 transcripts detected at the greater than 99% detection confidence limit in at least three of twelve samples. The results were visualized using 3-D scatter plots based on the condition for the three groups of affected, unaffected, and parents. The results on Figure 2 revealed some degree of separation between both 'Parents', shown in red squares, and 'Affected' children, light blue squares, from the 'Controls', dark blue squares. No clear separation is observed between groups of 'Parents' and 'Affected'.

Clustering analysis and study of differential expressed transcripts in MLASA patients

We performed unsupervised clustering analysis using both BeadStudio and Genespring GX software packages. Initially we used expression profiles of all 47,296 transcripts, followed by analysis of 12,983 transcripts detected at the greater than 99% detection confidence limit in at least three of twelve samples. Different clustering algorithms failed to separate individual patients with MLASA from their parents, unaffected sibling, and controls.

We therefore attempted to identify 'MLASA-associated transcripts' whose expression intensities differed significantly between patients affected with MLASA ('Affected') and all others without clinical symptoms of MLASA ('Unaffected'). Statistical testing identified 897 genes to be differentially expressed with Diff Score equal to or over 13 (p-value ≤ 0.05 , Supplementary Table 1).

Of the 897 transcripts, 816 had associated Genebank accession numbers while others were either Unigene clusters, or Gnomon predictions. We performed functional grouping of these annotated 816 genes based on functional similarity using Functional Classification tool available at the public bioinformatics resource DAVID 2.1 - Database for Annotation, Visualization, and Integrated Discovery [17]. It is designed to identify smaller sets of genes significantly co-occurring in the tested gene list as well as in the extensive gene list compiled

and categorized by DAVID, using powerful Fisher Exact test as a co-occurrence scoring system incorporated into the EASE method [18]. A number of functional clusters were identified. The most prominent and statistically significant functional groups are summarized in Table 1. The largest functional group consists of 68 transcripts related to transcription, regulation of transcription, DNA-binding and nucleic acid metabolism. The second largest cluster are transcripts related to metal ion binding, and three more functional groups are all related to RNA and protein metabolism, including ribosomal proteins.

We performed similar analysis of differential expression using Genespring GX program. The results revealed smaller number of differentially expressed transcripts divided into similar major functional groups (data not shown).

Study of differential expressed transcripts in individuals with at least one copy of mutated *PUS1* gene

To help elucidate which potential compensatory mechanisms are involved in carriers of mutated *PUS1* allele and in unaffected tissues of affected individuals, we compared gene expression between homozygotes and heterozygotes combined ('Mutation Carriers'), and wild-type controls ('Controls') using Genespring GX program. We performed 2-sided Student T-test as described in Materials and Methods and identified 740 differentially expressed transcripts, of which 553 were associated with Genebank accession numbers (Supplementary Table 2). These annotated genes were subjected to the Functional Classification tool from DAVID. The most significant differences between the two groups were related to protein biosynthesis and ribosomal proteins (Table 1).

Analysis of differentially expressed genes between the three groups of Affected, Parents, and Controls

To identify changes in expression potentially related to the mutated allele dosage, we performed 1-way ANOVA testing of three groups, 'Affected', 'Parents', and 'Controls', and identified 520 transcripts as being differentially expressed (Supplementary Table 3). We analyzed functional classification of 398 genes with Genebank accession numbers with Functional Classification tool and identified several groups the most prominent of which were transcription and metal ion binding (Table 1).

Investigation of known functional KEGG and BioCarta pathways related to the differentially expressed genes

Lists of differentially expressed genes between 'Affected' and 'Unaffected', and between those with at least one mutated *PUS1* allele ('Mutation Carriers') and 'Controls', were compared to human KEGG (Kyoto Encyclopedia of Genes and Genome, Kyoto University and Human Genome Center, <http://www.genome.ad.jp/kegg/>), and BioCarta (BioCarta, Inc., <http://www.biocarta.com/genes/index.asp>) pathways using Functional annotation tool from DAVID 2.1 and Genespring GX. Only single overlapping hits were identified when genes differentially expressed between 'Affected' and 'Unaffected' were analyzed. Strikingly, twenty-six differentially expressed genes between 'Mutation Carriers' and 'Controls' overlapped with one hundred and twelve genes represented in human ribosome KEGG pathway, with a dramatically significant p-value of 2.28E-19. It is important to note that the pathway is composed of components of cytoplasmic ribosome only, therefore five mitochondrial ribosomal proteins differentially expressed between 'Mutation Carriers' and 'Controls' were not identified in the pathway. List of all ribosomal proteins, differentially expressed between 'Mutation Carriers' and 'Controls' is shown in Table 3.

When we performed the same analysis using a list of differentially expressed genes identified between the groups of 'Affected', 'Parents', and 'Controls' the number of overlapping genes

decreased to eleven with a similarity p-value of 2.35E-5, not including four mitochondrial ribosomal proteins.

Targeted expression analysis of selected transcripts coding for iron metabolism related proteins

We selected a list of twenty-two transcripts possibly affected by abnormal SRA pseudouridylation by PUS1p (see Discussion), and compared their expression levels in different groups of samples. Eleven transcripts were expressed at more than 99% confidence, and three were identified as having Diff Scores above 13, corresponding to the nominally significant p-value of 0.05 (Table 2). The most statistically significant differences were calculated for FTL (ferritin light chain) with 1.5 fold change between groups of 'Affected' and 'Controls' with Diff Score of 31.68 (p-value <0.001).

Real-time RT-PCR analysis

To confirm expression results derived from Illumina WG-6 whole genome expression chip, we performed confirmation studies using TaqMan® Gene Expression Assays (Applied Biosystems). We included four genes MRPS10 (mitochondrial ribosomal protein S10, Genebank NM_018141), RASSF6 (Ras association (RalGDS/AF-6) domain family 6, Genebank NM_177532), NOLA2 (Nucleolar Protein Family A, member 2, MIM 606470), FTL (Ferritin Light Chain, MIM134790), and GAPDH (glyceraldehyde-3-phosphate dehydrogenase, MIM 138400). Their intensity levels varying from 110 to 16807, fold changes between 'Affected' and 'Unaffected' groups ranging from 1.0 (no difference) to 2.2, and Diff scores ranging from 0.65 (not significant) to 44.72 (p-value<0.001). The absolute levels of expression for all five transcripts from real-time PCR and from the microarray data were very similar (data not shown). Fold changes between groups of individuals calculated from microarray and real-time PCR data varied no more than 0.3 for all the transcripts, and no more than 0.2 for the transcripts with highly significant results with Diff scores above 40 (p-value<0.001).

Discussion

Global expression analysis identifies global transcriptional changes, and is used as a tool to interrogate suspected genes as well as to provide important clues on biological pathways involved in the pathogenesis of a disease prior not necessarily known to be involved. We chose Illumina's platform with an average 30-fold sequence redundancy for high-throughput expression analysis [19,20]. An independent study undertaken by the FDA (Food and Drug Administration) to assess reproducibility of the microarray technology and improve quality of the data, titled MicroArray Quality Control (MAQC) project (<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/>), demonstrated Illumina's technology to be highly reproducible and highly correlated both with the microarray data produced with other platforms such as Affymetrix and with gene-specific qRT-PCR results.

Quality data was obtained from analysis of RNA samples extracted from EBV-transformed lymphoblastoid cell lines with Illumina's WG-6 whole genome expression beadchips. Correlation between samples duplicated on the same or different chips was extremely high. Correlation between different samples was also very high as expected for RNAs which came from the same tissue. Our verification experiments using real-time RT-PCR technique validated relative amounts of expression for all tested transcripts as well as fold changes above 1.5 between groups of samples for each tested transcript.

Unsupervised clustering analysis done independently using two different computer programs and using global gene expression did not show phenotype-related clustering among C656T

homozygotes, heterozygotes, and wild-type controls. This can be explained by the fact that in the analysis of twelve thousand transcripts, which greatly vary in the levels of expression signal, individual differences and/or systematic technical variations probably exceed the effect of the disease. To reduce the dimensionality of our large data set, we performed Principal Component Analysis (PCA), a decomposition technique that produces a set of expression patterns known as principal components. PCA is not a clustering technique; it is a tool to characterize the most abundant themes, or principal components, and then use them to create a plot of the samples. PCA analysis revealed higher degree of separation between individuals carrying at least one copy of mutated *PUS1* gene (heterozygous parents and their affected children) and the wild-type controls, than between parents and their affected children. This observation is consistent with our expectations that some compensatory mechanisms occur in carriers of the *PUS1* defect, although the closer genetic relationship between individuals in the groups of 'Affected' and 'Parents' can contribute to this result by an indeterminate extent. Analysis of up- and down-regulated transcripts in individuals with at least one copy of mutated *PUS1* gene identified a significant number of cytoplasmic and mitochondrial ribosomal proteins. While levels of some of them were higher, others were lower, suggesting that the ratio of the components of the ribosomes to be important to its proper function in the presence of the *PUS1* defect. These results are in good correlation with the compensatory mechanism described in an EFG1 hepatopathy (see Introduction). Different ratios, rather than absolute levels, of the components of the mitochondrial translation system have been found to either compensate, or fail to compensate, for the defect in EFG1 protein [6]. Since our study is limited to the single available tissue from the MLASA patients, we can only identify the combination of ribosomal factors that compensate for the defect. However, we hypothesize that unlike the EFG1 hepatopathy, in MLASA the compensation mechanism works in all cell types, and the lack of pseudouridylation of non-tRNA targets of PUS1p is responsible for the disease phenotype.

Comparison of global gene expression and targeted analysis of iron metabolism-related genes between patients with MLASA and all others, identified DNA transcription, including regulation and nucleic acid metabolism, and binding of metal ions (including iron metabolism) as major molecular pathways up- and down-regulated in patients with MLASA. These preliminary results obtained from EBV-transformed lymphocytes support the hypothesis of pleiotropic effects involved in MLASA pathogenesis. A cascade of events, which starts with abnormal post-transcriptional modification of multiple RNA substrates by mutated hPUS1p leads to abnormal nucleic acid metabolism as well as metal ion binding and iron metabolism which leads to a damaging effect on iron metabolism in erythroid cells, and cause defective oxidative phosphorylation and energy metabolism in muscle cells. Similar pleiotropic mechanisms may operate in the case of the mutated *dyskerin* gene, responsible for the only other known disease of pseudouridylation, dyskeratosis congenital. Abnormal pseudouridylation of telomerase RNA leads to compromised telomerase function and telomere maintenance abnormalities [10,21], and abnormal pseudouridylation of ribosomal RNA leads to an impaired translation of IRES (internal ribosome entry site)-containing mRNAs [22].

Similar in-depth biochemical studies of tRNA and non-tRNA substrates of PUS1p will be required to clarify the specific mechanisms involved. Pseudouridylation is an abundant post-transcriptional modification of ribosomal RNA (rRNA), transfer RNA (tRNA), and small nuclear RNA (UsnRNA) carried out by members of a distinct protein family of pseudouridine synthases [23-27]. Prokaryotic and eukaryotic pseudouridine synthases modify specific positions in their respective RNA substrates with some of them displaying multi-site and multi-substrate capabilities [28-30]. Mammalian pseudouridine synthases have been shown to modify multiple positions in cytoplasmic and mitochondrial tRNAs [3,31], as well as additional non-tRNA substrates: U2 snRNA [7] and SRA [8]. *SRA* (*Steroid Receptor Activator*) related pathway may play a particularly important role in the sideroblastic anemia phenotype of MLASA. Pus1p and SRA cooperatively enhance mRAR (retinoic acid receptor) gamma-

mediated transcription of a number of target genes [8], including erythropoietin gene (EPO) [32]. Erythropoietin regulates cellular iron homeostasis in erythroid cells through the direct effect on the interaction of RNA stem loop structures, called iron-responsive elements (IREs), with iron-regulatory protein (IRP)-1 and IRP-2 [33–35]. Increase or decrease in binding activity of IRP-1 and IRP-2 to IREs present in ferritins (ferH, ferL, and mitochondrial ferritin), delta-aminolevulinic acid synthase (ALAS) isoforms 1 and 2, and transferrin receptor (TfR) directly affects their translation and consecutively iron storage, heme biosynthesis, and iron uptake by the cell [36–42]. Therefore, all of these functions could be affected by the loss of pseudouridylation of SRA by mutated PUS1p in affected erythroid precursors in the bone marrow of MLASA patients. This complex mechanism may also include differential splicing as a way of generating coding and noncoding SRA isoforms [43].

We are currently generating *PUS1* knockout mice. Clinical, histological, and biochemical analyses of these mice, and high-throughput expression studies of affected and unaffected tissues, may help clarify mechanisms of pathogenesis and compensation in this tissue-restricted disorder.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge support from NIH/NIDDK grant RO1-DK74368. This project was supported in part by General Clinical Research Center grant MO1-RR00425. We thank Dr. Kent Taylor for technical support.

References

1. Casas KA, Fischel-Ghodsian N. Mitochondrial myopathy and sideroblastic anemia. *Am J Med Genet* 2004;A125:201–204.
2. Bykhovskaya Y, Casas KA, Mengesha E, Inbal A, Fischel-Ghodsian N. Missense mutation in pseudouridine synthase 1 (*PUS1*) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *Am J Hum Genet* 2004;74:1303–1308. [PubMed: 15108122]
3. Patton JR, Bykhovskaya Y, Mengesha E, Bertolotto C, Fischel-Ghodsian N. Mitochondrial myopathy and sideroblastic anemia (MLASA): missense mutation in the pseudouridine synthase 1 (*PUS1*) gene is associated with the loss of tRNA pseudouridylation. *J Biol Chem* 2005;280:19823–19828. [PubMed: 15772074]
4. DiMauro S, Schon EA. Mitochondrial respiratory-chain diseases. *N Engl J Med* 2003;348:2656–2668. [PubMed: 12826641]
5. DiMauro S. The many faces of mitochondrial diseases. *Mitochondrion* 2004;4:799–807. [PubMed: 16120434]
6. Antonicka H, Sasarman F, Kennaway NG, Shoubbridge EA. The molecular basis for tissue specificity of the oxidative phosphorylation deficiencies in patients with mutations in the mitochondrial translation factor EFG1. *Hum Mol Genet* 2006;15:1835–1846. [PubMed: 16632485]
7. Behm-Ansmant I, Massenet S, Immel F, Patton JR, Motorin Y, Branlant C. A previously unidentified activity of yeast and mouse RNA:pseudouridine synthases 1 (Pus1p) on tRNAs. *RNA* 2006;12:1583–1593. [PubMed: 16804160]
8. Zhao X, Patton JR, Davis SL, Florence B, Ames SJ, Spanjaard RA. Regulation of nuclear receptor activity by a pseudouridine synthase through postranscriptional modification of steroid receptor RNA activator. *Mol Cell* 2004;15:549–558. [PubMed: 15327771]
9. Mochizuki Y, He J, Kulkarni S, Bessler M, Mason PJ. Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *Proc Natl Acad Sci U S A* 2004;101:10756–10761. [PubMed: 15240872]
10. Mitchell JR, Wood E, Collins K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 1999;402:551–555. [PubMed: 10591218]

11. Dokal I. Dyskeratosis congenita in all its forms. *Br J Haematol* 2000;110:768–779. [PubMed: 11054058]
12. Morello F, de Bruin TW, Rotter JJ, Pratt RE, van der Kallen CJ, Hladik GA, Dzau VJ, Liew CC, Chen YD. Differential gene expression of blood-derived cell lines in familial combined hyperlipidemia. *Arterioscler Thromb Vasc Biol* 2004;24:2149–2154. [PubMed: 15388524]
13. Baron CA, Liu SY, Hicks C, Gregg JP. Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. *J Autism Dev Disord* 2006;36:973–982. [PubMed: 16845580]
14. Haas CS, Creighton CJ, Pi X, Maine I, Koch AE, Haines GK, Ling S, Chinnaiyan AM, Holoshitz J. Identification of genes modulated in rheumatoid arthritis using complementary DNA microarray analysis of lymphoblastoid B cell lines from disease-discordant monozygotic twins. *Arthritis Rheum* 2006;54:2047–2060. [PubMed: 16804865]
15. Watts JA, Morley M, Burdick JT, Fiori JL, Ewens WJ, Spielman RS, Cheung VG. Gene expression phenotype in heterozygous carriers of ataxia telangiectasia. *Am J Hum Genet* 2002;71:791–800. [PubMed: 12226795]
16. Pressman S, Rotter JJ. Epstein-Barr virus transformation of cryopreserved lymphocytes: prolonged experience with technique. *Am J Hum Genet* 1991;49:467. [PubMed: 1651051]
17. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003;4:P3. [PubMed: 12734009]
18. Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003;4(10):R70. [PubMed: 14519205]
19. Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, Zhao C, Che D, Dickinson T, Wickham E, Bierle J, Doucet D, Milewski M, Yang R, Siegmund C, Haas J, Zhou L, Oliphant A, Fan JB, Barnard S, Chee MS. Decoding randomly ordered DNA arrays. *Genome Res* 2004;14:870–877. [PubMed: 15078854]
20. Kuhn K, Baker SC, Chudin E, Lieu MH, Oeser S, Bennett H, Rigault P, Barker D, McDaniel TK, Chee MS. A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res* 2004;14:2347–2356. [PubMed: 15520296]
21. Ruggero D, Grisendi S, Piazza F, Rego E, Mari F, Rao PH, Cordon-Cardo C, Pandolfi PP. Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science* 2003;299:259–262. [PubMed: 12522253]
22. Yoon A, Peng G, Brandenburger Y, Zollo O, Xu W, Rego E, Ruggero D. Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science* 2006;312:902–906. [PubMed: 16690864]
23. Patton JR. Pseudouridine formation in small nuclear RNAs. *Biochimie* 1994;76:1129–1132. [PubMed: 7748947]
24. Koonin EV. Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res* 1996;24:2411–2415. [PubMed: 8710514]
25. Massenet, S.; Mougin, A.; Branlant, C. Posttranscriptional modifications in the U Small Nuclear RNAs. In: Grosjean, H.; Benne, R., editors. *The modification and editing of RNA*. ASM Press; Washington, DC: 1998. p. 201–228.
26. Charette M, Gray MW. Pseudouridine in RNA: What, where, how, and why. *IUBMB Life* 2000;49:341–351. [PubMed: 10902565]
27. Ofengand J. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett* 2002;514:17–25. [PubMed: 11904174]
28. Motorin Y, Keith G, Simon C, Foiret D, Simos G, Hurt E, Grosjean H. The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA* 1998;4:856–869. [PubMed: 9671058]
29. Massenet S, Motorin Y, Lafontaine DL, Hurt EC, Grosjean H, Branlant C. Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase Pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol Cell Biol* 1999;19:2142–2154. [PubMed: 10022901]

30. Behm-Ansmant I, Urban A, Ma X, Yu YT, Motorin Y, Branlant C. The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:{Psi}-synthase also acting on tRNAs. *RNA* 2003;9:1371–1382. [PubMed: 14561887]
31. Chen J, Patton JR. Mouse pseudouridine synthase 1: gene structure and alternative splicing of pre-mRNA. *Biochem J* 2000;352:465–473. [PubMed: 11085940]
32. Makita T, Hernandez-Hoyos G, Chen TH, Wu H, Rothenberg EV, Sucov HM. A developmental transition in definitive erythropoiesis: erythropoietin expression is sequentially regulated by retinoic acid receptors and HNF4. *Genes Dev* 2001;15:889–901. [PubMed: 11297512]
33. Weiss G, Houston T, Kastner S, Johrer K, Grunewald K, Brock JH. Regulation of cellular iron metabolism by erythropoietin: activation of iron-regulatory protein and upregulation of transferrin receptor expression in erythroid cells. *Blood* 1997;89:680–687. [PubMed: 9002972]
34. Busfield SJ, Tilbrook PA, Callus BA, Spadaccini A, Kuhn L, Klinken SP. Complex regulation of transferrin receptors during erythropoietin-induced differentiation of J2E erythroid cells—elevated transcription and mRNA stabilisation produce only a modest rise in protein content. *Eur J Biochem* 1997;249:77–84. [PubMed: 9363756]
35. Zoller H, Decristoforo C, Weiss G. Erythroid 5-aminolevulinate synthase, ferrochelatase and DMT1 expression in erythroid progenitors: differential pathways for erythropoietin and iron-dependent regulation. *Br J Haematol* 2002;118:619–626. [PubMed: 12139757]
36. Klausner RD, Rouault TA. A double life: cytosolic aconitase as a regulatory RNA binding protein. *Mol Biol Cell* 1993;4:1–5. [PubMed: 8443405]
37. Hentze MW, Kuhn LC. Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proc Natl Acad Sci U S A* 1996;93:8175–8182. [PubMed: 8710843]
38. Ponka P. Tissue-specific regulation of iron metabolism and heme synthesis: distinct control mechanisms in erythroid cells. *Blood* 1997;89:1–25. [PubMed: 8978272]
39. Hanson, ES.; Leibold, EA. *Molecular and Cellular Iron Transport*. Templeton, DM., editor. M. Dekker; New York: 2002. p. 207-235.
40. Sadlon TJ, Dell'Oso T, Surinya KH, May BK. Regulation of erythroid 5-aminolevulinate synthase expression during erythropoiesis. *Int J Biochem Cell Biol* 1999;31:1153–1167. [PubMed: 10582344]
41. Casey JL, Koeller DM, Ramin VC, Klausner RD, Harford JB. Iron regulation of transferrin receptor mRNA levels requires iron-responsive elements and a rapid turnover determinant in the 3' untranslated region of the mRNA. *EMBO J* 1989;8:3693–3699. [PubMed: 2583116]
42. Theil EC. Regulation of ferritin and transferrin receptor mRNAs. *J Biol Chem* 1990;265:4771–4774. [PubMed: 2156853]
43. Hube F, Guo J, Chooniedass-Kothari S, Cooper C, Hamedani MK, Dibrov AA, Blanchard AA, Wang X, Deng G, Myal Y, Leygue E. Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines. *DNA Cell Biol* 2006;25:418–428. [PubMed: 16848684]

Figure 1.
Pedigree structure of MLASA family.

Figure 2.
3-D Scatter Plot of Principal Component Analysis generated by the Genespring GX program. 'Affected' are indicated by light blue squares, 'Parents' by red squares, and 'Controls' by dark blue squares.

Table 1
Molecular function annotations of genes differentially expressed between different groups of patients samples

[[Group]]	[[Genes present]]	[[Function]]	[[P-value]]
[[1]]	[[68]]	Group of 'Affected' vs. group of 'Unaffected'	
		Transcription; nucleobase, nucleoside, nucleotide and nucleic acid metabolism; transcription regulation; DNA-binding	[[2.94E-44]]
		[[Metal ion binding]]	[[3.00E-28]]
[[2]]	[[28]]	Contain RNA recognition motif; mRNA splicing and processing	[[4.18E-16]]
[[3]]	[[9]]	[[Protein biosynthesis; ribosome]]	[[1.53E-13]]
[[4]]	[[10]]	[[Protein transport]]	[[4.32E-12]]
[[5]]	[[7]]	Group of 'Mutation Carriers' vs. group of 'Controls'	
[[1]]	[[37]]	[[Protein biosynthesis; ribosome]]	[[0.00]]
[[2]]	[[58]]	Transcription; nucleobase, nucleoside, nucleotide and nucleic acid metabolism; transcription regulation; DNA-binding	[[4.46E-40]]
[[3]]	[[6]]	[[mRNA processing]]	[[2.58E-08]]
[[4]]	[[15]]	Zinc ion binding; metal ion binding	[[1.92E-07]]
[[5]]	[[6]]	[[Cellular protein catabolism]]	[[1.96E-07]]
[[1]]	[[54]]	Group of 'Affected' vs. group of 'Parents' vs. group of 'Controls'	
		Transcription; nucleobase, nucleoside, nucleotide and nucleic acid metabolism; transcription regulation; DNA-binding	[[3.04E-39]]
[[2]]	[[14]]	[[Protein biosynthesis; ribosome]]	[[3.15E-20]]
[[3]]	[[7]]	Protein Serine/Threonine kinase activity	[[3.66E-10]]
[[4]]	[[9]]	Zinc ion binding; metal ion binding	[[2.66E-09]]
[[5]]	[[5]]	[[Cellular protein catabolism]]	[[3.23E-06]]

Table 2
Differential expression analysis of candidate genes related to iron metabolism

[[[GeneBank Accession Number]]]]	[[[Gene Description]]]]	[[['Controls'/'Affected']]	[[['Unaffected'/'Affected']]	[[Expression Diff Score]] [[['Controls'/'Mutation Carriers']]	[[['Parents'/'Affected']]
[[NM_000146.2]] [[XM_370714.2]]	[[Ferritin, light polypeptide]] Similar to Ferritin heavy chain (Ferritin H subunit)	31.68 <u>14.15</u>	14.81 [[4.35]]	[[12.33]] 13.75	[[4.4]] [[0.26]]
[[NM_002032.1]]	[[Ferritin, heavy polypeptide1]]	[[[-3.81]]]	[[[-3.62]]]	[[[-4.16]]]	[[[-0.11]]]
[[XM_375850.1]]	Similar to Ferritin heavy chain (Ferritin H subunit)	[[[-3.64]]]	[[[-4.41]]]	[[[-2.97]]]	[[[-1.27]]]
[[NM_002197.1]] [[XM_039114.4]]	[[Aconitase 1, soluble]] Iron-responsive Element binding protein 2	[[[-3.73]]] [[11.07]]]	[[[-3.66]]] [[10.94]]]	[[[-4.76]]] [[6.39]]]	[[[-0.05]]] [[9.92]]]
[[NM_004136.1]]	Iron-responsive element binding protein 2	[[8.53]]]	[[3.25]]]	15.23	[[[-0.13]]]
[[NM_000688.4]]	Aminolevulinate, delta-, synthase 1, transcript variant 1	[[1.05]]]	[[0.28]]]	[[0.83]]]	[[0.8]]]
[[NM_000966.3]]	Retinoic acid receptor, gamma	[[1.72]]]	[[0.81]]]	[[2.87]]]	[[0.28]]]
[[NM_000121.2]] [[NM_018471.1]]	[[Erythropoietin receptor]] Likely ortholog of mouse immediate early response, erythropoietin 4	[[3.46]]] [[7.91]]]	[[1.67]]] [[4.05]]]	[[5.94]]] [[8.18]]]	[[0.17]]] [[1.8]]]

Expression Diff Score represents statistical confidence that the gene's expression is different between the two groups compared as calculated by Beadstudio. Statistically significant Diff Scores are in bold and underlined. Negative scores represent up-regulation, positive scores – down-regulation.

Table 3
Cytoplasmic and mitochondrial ribosomal proteins differentially expressed between groups of ‘Controls’ and ‘Mutation Carriers’

[[Protein Name]]	[[GeneBank Accession Number]]	[[T-test P- value]]
[[S6]]	[[NM_001010]]	[[0.000414]]
[[L18a]]	[[NM_000980]]	[[0.00154]]
[[S10]]	[[NM_001014]]	[[0.00175]]
[[L19]]	[[NM_000981]]	[[0.00564]]
[[S28]]	[[NM_001031]]	[[0.00636]]
[[S14]]	[[NM_005617]]	[[0.00687]]
[[L35]]	[[NM_007209]]	[[0.00723]]
[[L11]]	[[NM_000975]]	[[0.0092]]
[[L35a]]	[[NM_000996]]	[[0.00931]]
[[L24]]	[[NM_000986]]	[[0.00981]]
[[L30]]	[[NM_000989]]	[[0.0104]]
[[L12]]	[[NM_000976]]	[[0.013]]
[[S2]]	[[NM_002952]]	[[0.0157]]
[[L31]]	[[NM_000993]]	[[0.0169]]
[[S21]]	[[NM_001024]]	[[0.0189]]
[[S7]]	[[NM_001011]]	[[0.019]]
[[S25]]	[[NM_001028]]	[[0.0195]]
[[mitochondrial S21]]	[[NM_031901]]	[[0.0225]]
[[mitochondrial L52]]	[[NM_181307]]	[[0.023]]
[[mitochondrial L43]]	[[NM_032112]]	[[0.0232]]
[[L5]]	[[NM_000969]]	[[0.0255]]
[[L3]]	[[NM_000967]]	[[0.0271]]
[[L29]]	[[NM_000992]]	[[0.0279]]
[[L18]]	[[NM_000979]]	[[0.0296]]
[[S30]]	[[NM_001997]]	[[0.0312]]
[[L4]]	[[NM_000968]]	[[0.0316]]
[[L37a]]	[[NM_000998]]	[[0.0352]]
[[S15a]]	[[NM_001019]]	[[0.0359]]
[[mitochondrial L55]]	[[NM_181465]]	[[0.0392]]
[[mitochondrial L47]]	[[NM_020409]]	[[0.0456]]
[[S19]]	[[NM_001022]]	[[0.0486]]

T-test performed by Genespring computer program.