The Modulatory Effects of the Polymorphisms in *GLA* 5'-Untranslated Region Upon Gene Expression Are Cell-Type Specific

Susana Ferreira • Carlos Reguenga •
João Paulo Oliveira

Received: 17 September 2014 / Revised: 01 December 2014 / Accepted: 13 February 2015 / Published online: 13 March 2015
© SSIEM and Springer-Verlag Berlin Heidelberg 2015

Abstract Lysosomal α -galactosidase A (α Gal) is the enzyme deficient in Fabry disease (FD). The 5'-untranslated region (5'UTR) of the α Gal gene (*GLA*) shows a remarkable degree of variation with three common single nucleotide polymorphisms at nucleotide positions c.-30G>A, c.-12G>A and c.-10C>T. We have recently identified in young Portuguese stroke patients a fourth polymorphism, at c.-44C>T, co-segregating in *cis* with the c.-12A allele. In vivo, the c.-30A allele is associated with higher enzyme activity in plasma, whereas c.-10T is associated with moderately decreased enzyme activity in leucocytes. Limited data suggest that c.-44T might be associated with increased plasma α Gal activity. We have used a luciferase reporter system to experimentally assess the relative modulatory effects on gene expression of the different *GLA* 5'UTR polymorphisms, as compared to the wild-type sequence, in four different human cell lines. Group-wise, the relative luciferase expression patterns of the various

GLA variant isoforms differed significantly in all four cell lines, as evaluated by non-parametric statistics, and were cell-type specific. Some of the post hoc pairwise statistical comparisons were also significant, but the observed effects of the *GLA* 5'UTR polymorphisms upon the luciferase transcriptional activity in vitro did not consistently replicate the in vivo observations.

These data suggest that the *GLA* 5'UTR polymorphisms are possible modulators of the α Gal expression. Further studies are needed to elucidate the biological and clinical implications of these observations, particularly to clarify the effect of these polymorphisms in individuals carrying *GLA* variants associated with high residual enzyme activity, with no or mild FD clinical phenotypes.

Introduction

Insufficient activity of lysosomal alpha-galactosidase A (α Gal; EC 3.2.1.22) impairs the catabolism of several neutral glycosphingolipids, particularly of trihexosylceramide (Kint 1970). Accumulation of these glycosphingolipids in the endothelium and smooth muscle cells of blood vessels, as well as in the heart, the kidney and the nervous system, is the pathologic hallmark of Fabry disease (FD, OMIM #301500) (Desnick et al. 2001). The severity of the clinical phenotype of FD is roughly related to residual enzyme activity (REA) of α Gal, as assayed in vitro: the lower the REA, the earlier is the age of onset and the more severe and multi-systemic are the clinical manifestations, while in patients with higher levels of REA the resultant phenotypes are more organ restricted (Germain 2010).

The primary transcript of the α Gal gene (*GLA*) contains an unusually polymorphic 5'-untranslated region (5'UTR) of 110 nucleotides, encoded by exon 1 (Nucleotide

Communicated by: Viktor Kozich

Competing interests: None declared

Electronic supplementary material: The online version of this chapter (doi:10.1007/8904_2015_424) contains supplementary material, which is available to authorized users.

S. Ferreira (✉) • J.P. Oliveira

Department of Genetics, Faculty of Medicine, University of Porto, Porto, Portugal
e-mail: susanadg@med.up.pt

C. Reguenga

Department of Experimental Biology, Faculty of Medicine, University of Porto, Porto, Portugal

C. Reguenga

Institute for Molecular and Cell Biology, Porto, Portugal

J.P. Oliveira

Genetics Clinic, São João Hospital Centre, Porto, Portugal

database, National Center for Biotechnology Information – NCBI, reference sequence: NML_000169.2; http://www.ncbi.nlm.nih.gov/nucore/NML_000169.2; National Library of Medicine, Bethesda, MD, USA). Three single nucleotide polymorphisms (SNPs) respectively identified by reference numbers rs2071225, rs3027585 and rs3027584 at the NCBI SNP database (dbSNP, <http://www.ncbi.nlm.nih.gov/snp/>), which result from cytosine-to-thymine (C>T) or adenine-to-guanine (G>A) transitions at cDNA nucleotide positions c.-10(C>T), c.-12(G>A) and c.-30(G>A), counting backwards from the translation initiation codon, are relatively common in several ethnically different populations (Davies et al. 1993; Saifudeen et al. 1995; Wu et al. 2011; Ferri et al. 2012), including the Portuguese (Oliveira et al. 2008a). The dbSNP lists three additional *GLA* 5'UTR SNPs, at positions c.-8(C>G), c.-18(T>C) and c.-105(A>G), respectively identified as rs371291716, rs545597063 and rs3027583, but these variants have never been reported in the Portuguese population. Another *GLA* 5'UTR variant at c.-44(C>T), not yet registered at the dbSNP, was identified in young Portuguese stroke patients (Baptista et al. 2010), co-segregating in cis with the c.-12A allele.

Compared to the wild-type (WT) allele, the c.-30A allele is associated with increased plasma α Gal activity (Saifudeen et al. 1995; Fitzmaurice et al. 1997), the c.-10T allele is associated with decreased activity of α Gal in leukocytes (Oliveira et al. 2008a, b) and the c.-44T allele might be associated with increased plasma α Gal activity (Baptista et al. 2010). Contrastingly, the SNP c.-12(G>A) seems to have no effect upon *GLA* gene expression (Oliveira et al. 2008a).

Because the *GLA* 5'UTR contains a binding site for methylated DNA-binding protein (MDBP)/regulatory factor X (RFX) transcription factor family (Zhang et al. 1990; Samac et al. 1998), as well as partially overlapping binding motifs for the nuclear factor kappa-B (NF κ B) and E-26 (Ets) transcriptional regulatory factors (Saifudeen et al. 1995; Fitzmaurice et al. 1997), it has been previously suggested that protein ligands to these sites might be involved in the regulation of *GLA* gene expression (Saifudeen et al. 1995; Fitzmaurice et al. 1997). Since the SNPs c.-10(C>T) and c.-30(G>A) involve nucleotides respectively located at each of those binding sites, they might act as polymorphic modulators of *GLA* gene expression, with possible clinical relevance particularly in males carrying *GLA* sequence variants associated with high REA.

To compare the relative impact upon in vitro gene expression of each of the *GLA* 5'UTR SNPs that have been identified in the Portuguese population, we have carried out luciferase reporter assays in several human cell lines of distinct embryological origins, representing some of the cell types that are critically involved in the pathogenesis of FD.

Materials and Methods

A detailed description of the laboratory methods is available online (Supplementary Information and Supplementary Tables 1 and 2).

Synopsis of the Laboratory Protocol

Five genomic DNA (gDNA) samples from adult males known to carry either the *GLA* 5'UTR WT sequence or one of the four SNPs identified in the Portuguese population were obtained from an anonymised biorepository. The five distinct isoforms were amplified by polymerase chain reaction (PCR) and the amplicons were inserted into the *HindIII* site of the pGL3-Control Vector (pGL3 Luciferase Reporter Vectors; Promega, Madison, WI, USA). The pGL3-Control Vector contains a cDNA encoding a modified firefly luciferase (luc+) under the control of the Simian virus 40 (SV40) promoter. The recombinant pGL3 vectors were transiently transfected into each of four human cell lines of different embryological origin: (1) human embryonic kidney, HEK-293 (Shaw et al. 2002); (2) human cervical carcinoma, HeLa (Macville et al. 1999); (3) human dermal microvascular endothelial cells, HDMEC (Richard et al. 1999); and (4) human T cell lymphoblast-like, Jurkat (Schneider et al. 1977). A plasmid (pCDNA3.3-LacZ; Invitrogen, Life Technologies, Carlsbad, CA, USA) containing the *Escherichia coli* (*E. coli*) β -galactosidase gene (pGAL) was used as an internal control for transfection efficiency. The relative luciferase activity (RLA) of each sample was calculated as the ratio of luciferase to β -galactosidase luminometric readings. A minimum of nine independent successful co-transfection experiments were analysed per cell type.

To allow comparisons between independent experiments in each cell type, a normalised RLA (nRLA) was calculated by dividing the RLA calculated for each sample by the average RLA of the WT construct vector samples of the corresponding experiment.

Statistical Analyses

The non-parametric Kruskal–Wallis one-way analysis of variance by ranks was used as the first-tier statistical testing for significant differences in the nRLA raw data of the various *GLA* 5'UTR vector constructs, in each cell type.

As this test does not identify where the observed differences occur or how many differences actually occur, further statistical analyses were based on parametric testing with one-way analysis of variance (ANOVA) and post hoc Dunnett's test. To this end, outlier raw RLA data points were first excluded by the fourth-spread method, as recommended by other investigators (Jacobs and Dinman

Table 1 Non-parametric statistical analysis by the Kruskal–Wallis test of the normalised relative luminometric activity (nRLA) of the five *GLA* 5'UTR plasmid vector constructs in four different cell lines

<i>GLA</i> 5'UTR plasmid vector constructs	Cell lines							
	HEK-293		HeLa		HDMEC		Jurkat	
	<i>N</i>	Mean rank (nRLA)	<i>N</i>	Mean rank (nRLA)	<i>N</i>	Mean rank (nRLA)	<i>N</i>	Mean rank (nRLA)
WT	13	20.54	16	38.03	15	45.23	15	45.17
c.-10T	13	46.62	16	48.28	15	39.77	15	27.07
c.-12A	13	30.92	16	28.69	15	26.73	15	27.83
c.-30A	13	25.96	12	10.00	15	48.60	9	23.94
c.-44T	13	40.96	16	60.38	15	29.67	15	46.57
Total (<i>N</i>)	65		76		75		69	
<i>Kruskal–Wallis test</i>								
Chi-square (<i>df</i> = 4)	16.652		41.993		11.501		15.831	
Asymptotic significance	0.002		0.000		0.021		0.003	

WT wild-type, *N* number of experiments validated per each cell line, *df* degrees of freedom

Table 2 One-way analysis of variance (ANOVA) of the natural logarithms of the normalised relative firefly luciferase activities expressed as a percentage of the wild-type (nRLA%), comparing the means of the five *GLA* 5'UTR plasmid vector constructs in four different cell lines

ANOVA results						
Cell lines		Sum of squares	<i>df</i>	Mean square	<i>F</i>	<i>p</i> -value
HEK-293	Between groups	2.184	4	0.546	3.615	0.010
	Within groups	9.062	60	0.151		
	Total	11.246	64			
HeLa	Between groups	10.423	4	2.606	22.796	0.000
	Within groups	8.116	71	0.114		
	Total	18.538	75			
HDMEC	Between groups	1.401	4	0.350	3.721	0.009
	Within groups	6.305	67	0.094		
	Total	7.706	71			
Jurkat	Between groups	1.005	4	0.251	4.278	0.004
	Within groups	3.641	62	0.059		
	Total	4.646	66			

df degrees of freedom

2004), the nRLA values were expressed as percentage of the WT nRLA (nRLA%), and the natural logarithms of the latter values were used for the statistical analyses.

A two-sided significance level of <0.05 was assumed for all the statistical tests except for post hoc Dunnett's test where one-sided *p*-values were used instead. All statistical analyses were performed with the SPSS software, Version 19.0.

Results

HDMEC cells with more than 12 subculture passages showed non-uniform and significantly slower doubling rates, with more erratic transfection efficiencies (data not shown). For that reason, only the RLA assays obtained in younger cell cultures were validated for the final analyses.

Fig. 1 Normalised luciferase/β-galactosidase luminometric ratios of the *GLA* 5'UTR sequence variant isoforms c.-10T, c.-12A, c.-30A and c.-44T, expressed as percent of wild-type (WT), in HEK-293, HeLa, HDMEC and Jurkat human cell lines. The average bars are shown with the upper limit of the standard error of the mean. As explained in

the text, the c.-44T allele was present in *cis* with the c.-12A allele. All pairwise comparisons between each of the variant alleles and the WT *GLA* 5'UTR isoform that reached statistical significance in the post hoc analyses are shown in the charts, with the corresponding one-sided Dunnett's test *p*-value

The distributions of the nRLA of the five *GLA* 5'UTR constructs, as evaluated by non-parametric statistics (Table 1), differed significantly in all cell types. By the fourth-spread method, one of the WT and one of the c.-44T allele data point values obtained in the Jurkat cells, as well as three of the c.-30A allele data point values obtained in the HDMEC cells, were classified as outliers and removed from the dataset for the subsequent parametric statistical analyses. ANOVA was performed on the normalised, log-transformed nRLA% data, with results similar to the non-parametric statistical approach (Table 2). Post hoc parametric statistical analyses showed that the SNPs c.-12 (G>A) and c.-30 (G>A) were associated with significantly lower nRLA%, respectively in HDMEC ($p = 0.002$) and HeLa ($p < 0.000$) cells, while the SNPs c.-10 (C>T) and c.-44 (C>T) were associated with significantly higher nRLA%, respectively in HEK-293 ($p = 0.032$) and HeLa ($p < 0.000$) cells. In the Jurkat cells, none of the *GLA* SNPs was associated with a statistically significant difference in nRLA%, but both the c.-10T ($p = 0.067$) and c.-12A ($p = 0.086$) alleles showed a trend to lower activity (Fig. 1).

Discussion

Luciferase-based genetic reporter assays are a standard in vitro approach to study DNA sequences and molecular processes that control gene expression, in various cellular contexts (Brogan et al. 2012). Herein we report the results of luciferase expression assays designed to investigate the relative efficiency of four different human *GLA* 5'UTR SNPs, in comparison to the WT sequence, as potential modulators of gene expression. One of the most common research applications of chimeric genetic reporter systems is in the analysis of *cis*-acting elements, like gene promoters; however, to the best of our knowledge, the functional consequences of the 5'UTR SNPs upon *GLA* gene expression have never been assayed in this manner. Overall, our results are consistent with the working hypothesis, based on observations in vivo that some of the minor 5'UTR SNPs alleles significantly affect αGal activity levels (Saifudeen et al. 1995; Fitzmaurice et al. 1997; Oliveira et al. 2008a, b), and provide indirect evidence that the human *GLA* 5'UTR indeed contains sequences that are involved in the regulation of gene expression. These data

also demonstrate, for the first time, that the 5'UTR-dependent modulation of *GLA* gene expression may vary among different cell types. The expression of alternative 5'UTRs represents an evolutionary gain of transcriptional and translational control pathways, allowing tissue-specific expression patterns and expanding the repertoire of expression from a single gene locus (Barrett et al. 2012). Although *GLA* is a housekeeping gene, α Gal activity levels vary greatly from organ to organ and in different cell types (Brady et al. 1967; von Scheidt et al. 1991). It is possible that either transcription factors (TF) or RNA-binding proteins (RBP) that bind to specific sequences in the *GLA* 5'UTR contribute to the regulation of *GLA* gene expression in a tissue-specific manner.

The three common SNPs of the human *GLA* 5'UTR were originally described in 1993 with a combined minor allele frequency of 10% in the British population (Davies et al. 1993). Because they were neither translated nor part of the mRNA Kozak consensus sequence for translation initiation, they were regarded as biologically neutral, but subsequent studies showed that the c.-30A allele was associated with increased α Gal activity in plasma (Saifudeen et al. 1995) and the c.-10T allele with decreased α Gal activity in leukocytes (Oliveira et al. 2008a, b), as compared to the corresponding WT alleles. Since the *GLA* 5'UTR c.-10 and c.-30 positions are within binding sites respectively for the MDBP and the NF κ B and Ets families of TF, the effect of those two nucleotide transitions upon α Gal expression in vivo might be mediated by modulation of transcriptional activity.

On electrophoretic mobility shift assays (EMSA), binding of nuclear extract proteins to synthetic oligonucleotides containing the *GLA* 5'UTR NF κ B/Ets binding site, either with a G or an A at the position corresponding to c.-30, was significantly less when adenine was present (Saifudeen et al. 1995), showing that the WT sequence has higher affinity to the putative NF κ B/Ets ligands. Furthermore, in vitro translation of mRNAs from cloned WT and c.-30A alleles resulted in similar levels of α Gal protein, indicating that the G>A transition does not enhance translation (Fitzmaurice et al. 1997), and studies performed on α Gal derived from the WT and the c.-30A alleles, partially purified from plasma and lymphoblasts, revealed that the high plasma activity was not due to altered post-translational processing (Fitzmaurice et al. 1997). Overall, these findings suggest that the *GLA* 5'UTR c.-30G>A transition results in enhanced transcription, presumably by interfering with the binding of negatively acting TF which normally decrease α Gal expression in various cells (Fitzmaurice et al. 1997).

Surprisingly, the c.-30A allele was not associated with higher protein expression in comparison to the WT allele, in any of the four cell types assayed in our luciferase reporter

studies. Since plasma α Gal most probably has a multiplicity of cellular sources (Fitzmaurice et al. 1997; Warnock 2005), a possible explanation for the inconsistency between the in vivo observations and the experimental in vitro data might be that none of the cell types used in our reporter studies is the right model to assay the 5'UTR-related modulation of *GLA* gene expression at the transcriptional level, at least for cells that most significantly contribute to the pool of circulating α Gal protein. Therefore, testing additional cell types, ideally representing highly vascularised tissues that are particularly rich in α Gal activity, like the spleen and the liver (Brady et al. 1967), might help to clarify these discrepancies.

Both the c.-12(A>G) and the c.-10(C>T) SNPs are located within the *GLA* 5'UTR MDBP consensus sequence, respectively, at its first and third nucleotide positions (Zhang et al. 1990), but only the c.-10(C>T) SNP seem to affect enzyme expression in vivo (Oliveira et al. 2008a). The amount of α Gal identified by western blot analyses of leukocyte protein extracts was significantly lower in carriers of the c.-10T allele as compared to carriers of the WT allele (Oliveira et al. 2008b). Experimental data (Samac et al. 1998) have shown that sequence changes that increase the affinity of the *GLA* 5'UTR MDBP binding site for its cognate ligands exert a strong repressive effect upon gene expression. Furthermore, a C>T transition in the third nucleotide position of a MDBP binding site in the human cytomegalovirus significantly increased ligand binding, resulting in 10-fold reduction of reporter gene expression (Schneider et al. 1977). However, to the best of our knowledge, no EMSA studies have ever been performed with the human *GLA* 5'UTR MDBP binding site, to assess the relative affinities of the c.-10(C > T) alleles. Although the human *GLA* 5'UTR c.-12(G>A) SNP does not seem to change α Gal expression in vivo, at least in plasma and leukocytes (Oliveira et al. 2008a), studies in the MDBP binding site of plasmid pBR322 have shown that the A>G transition in its first nucleotide position significantly increased the binding of MDBP extracted from the human placenta (Macville et al. 1999).

The results of our luciferase reporter assay studies are consistent with the hypothesis that the G>A transition at c.-12 has indeed a biological effect on gene expression. It is of note that in HDMEC and Jurkat cells the *GLA* c.-10T and c.-12A alleles are associated with relatively lower reporter gene expression as compared to the WT alleles, but in HEK-293 and HeLa cells the c.-10T allele is associated with increased reporter gene expression. The trend for a lower reporter gene expression associated with the c.-10T allele in the Jurkat cells is in agreement with the decreased leukocyte α Gal activity observed in vivo (Oliveira et al. 2008a). The lack of statistical significance of all the post hoc analyses of the relative luciferase reporter expression

data in the Jurkat cells, as well as of some of the comparisons made in other cell types, can be attributed to statistically underpowered datasets. Confirmation of the present findings should be made using larger datasets of validated reporter gene expression readings per cell type, possibly using a dual reporter assay for better control of experimental normalisation (Jacobs and Dinman 2004).

The relative luciferase expression profile of the four human *GLA* 5'UTR SNPs, as normalised to the WT sequence, was different for each of the transfected cell lines in our experimental protocol. The diverse embryological lineages of HEK-293 (Shaw et al. 2002), HeLa (Macville et al. 1999), HDMEC (Richard et al. 1999), and Jurkat (Schneider et al. 1977) cells are a logical explanation for this finding, as the *GLA* gene may be under different constitutive expression regulation on different cells.

To identify other potential TF binding sites in the *GLA* 5'UTR, we queried the Transcription Factor Database (TRANSFAC). According to TRANSFAC, the c.-12G and the c.-10C nucleotides are both part of a consensus binding motif for zinc finger C4-type domains of nuclear receptors (ZFC4-NR), particularly the estrogen receptor alpha (ER- α), and the G>A and C>T transitions at those positions result in the loss of ZFC4-NR binding affinity. Similarly, the G>A transition at the c.-30 nucleotide, which is part of a conserved consensus binding sequence for both the Ets and the adenovirus E2 promoter binding (E2F) families of TF, also suppresses the site ligand affinity. On the other hand, the C>T transition at the c.-44 nucleotide creates a novel ZFC4-NR binding site, particularly for hepatocyte nuclear factor-4-alpha (HNF-4 α). It is of note that, in the Portuguese population, the c.-44T allele has been found exclusively in *cis* with the c.-12A allele: therefore, the results of our luciferase reporter assays represent the in vitro combined effects of the two nucleotide transitions.

By predicting binding sites from position weight matrix in the database of RNA-binding specificities (RBPDB), the G>A transition at the c.-30 nucleotide results in the loss of binding affinity with the eukaryotic translation initiation factor 4B (EIF4B), but no changes are reported for the c.-10 (C>T) and the c.-12(G>A) SNPs. On the other hand a C>T transition at the c.-44 nucleotide originates a novel binding site to KH-type splicing regulatory protein. According to these results either the transcriptional machinery or post-transcriptional events, or both, might be influencing the gene expression.

In conclusion, as assessed in vitro by luciferase reporter assays, the various *GLA* 5'UTR SNPs modulate the gene expression in a manner that seems to be cell-type specific. The biological context and clinical implications of these observations are not yet clear, and further studies will be necessary to elucidate such questions. This may be of

particular relevance considering the recent data suggesting that subnormal α Gal activity levels above the range of enzyme insufficiency associated with FD phenotypes may be a risk factor in the pathogenesis of sporadic Parkinson disease (Wu et al. 2008, 2011) and stroke (Baptista et al. 2010). Although the 5'UTR-related modulation of *GLA* gene expression is not anticipated to be of much relevance for male patients carrying pathogenic mutations associated with very low or absent α Gal activity, elucidation of the specific cellular contexts and of the molecular mechanisms underlying the effect of *GLA* 5'UTR SNPs, particularly of the negatively acting, might be clinically relevant in individuals with several other *GLA* genotypes. First: in patients with later-onset, organ-limited FD phenotypes, because most of such cases are associated with *GLA* mutations leading to altered enzyme stability (Garman 2007), which makes the final amount of α Gal protein available in each tissue more dependent of the local gene expression profile. Second: in females coinheritance a negatively acting *GLA* 5'UTR SNP in *trans* with a pathogenic *GLA* mutation, the compound heterozygosity may significantly decrease the REA and aggravate the clinical phenotype since the α Gal activity would be subnormal in all cells, irrespective of which of the two X chromosomes is inactivated. Third: in individuals carrying *GLA* variants associated with small decreases in REA, which otherwise would not be the cause of FD clinical phenotypes – like the p.Arg118Cys (Ferreira et al. 2015) and the p.Asp313Tyr (Yasuda et al. 2003; Niemann et al. 2013) – the additive effect of a negatively acting *GLA* 5'UTR SNP in *cis* could decrease REA into the typical range of mutations associated with later-onset phenotypic variants of FD (Ferreira et al. 2015), thereby modifying the expected phenotype. This might also affect the probability of identifying individuals carrying such mutations in large case-finding studies of FD among high-risk patients, when the primary screening method is based on α Gal activity analysis. Finally, at the population level, the inadvertent inclusion of carriers of nonneutral *GLA* 5'UTR SNPs in cohorts of healthy individuals used to define the normal laboratory ranges of α Gal activity will affect the values of the upper limit in plasma, like the c.-30A allele, and of the lower limit in leukocytes, like the c.-10T allele.

Acknowledgements These studies have been partially supported by a donation from Sanofi-Genzyme (Portugal) for research in Fabry disease. However, Sanofi-Genzyme was not involved in any way either with the design or the development of this study, including the analyses of the experimental data, or with the decision to publish their results.

We thank Deolinda Lima, M.D., Ph.D., for providing the facilities of the Laboratory of Support to Research in Molecular Medicine (LAIMM), at the Department of Experimental Biology, of the Faculty of Medicine, University of Porto (FMUP), Portugal, where most of the experimental work was performed.

We thank Ana Moço, MSc Student, from the Biocomposites Group, Institute of Biomedical Engineering (INEB), Faculty of Engineering, University of Porto (FEUP), Portugal, for generously providing the HDMEC cells; José Pedro Castro, Ph.D. Student, from the Department of Experimental Biology, Centre for Medical Research, FMUP, for generously providing the Jurkat cells; and Ana Grangeia, Ph.D., from the Genetics Department, FMUP, for generously providing the HEK-293 cells.

Synopsis

The polymorphisms in the α -galactosidase gene (*GLA*) 5'-untranslated region have cell-type-specific modulatory effects upon gene expression.

Compliance with Ethics Guidelines

This study was conducted according to the applicable national and institutional regulatory and ethical standards. Anonymised male genomic DNA samples were used as the source of the various polymorphic *GLA* 5'-untranslated region nucleotide sequences assayed in the luciferase reporter system experiments reported herein. The original samples had been obtained and genotyped in prior research projects carried out in the laboratory of molecular genetics of the Department of Genetics, Faculty of Medicine, University of Porto, Portugal, which were approved by the institutional Health Ethics Board and subject to written informed consent. The preservation of the anonymised genomic DNA samples was authorised by the Portuguese Data Protection Authority.

The study protocol did not involve any studies performed with human or animal subjects.

This work is part of Susana Ferreira's Ph.D. thesis, supervised by João Paulo Oliveira and co-supervised by Carlos Reguenga.

All authors were involved in the conception and design of the study, had full access to all the data and were involved in their analysis and interpretation. Susana Ferreira performed the laboratory work and, as corresponding author, had the final responsibility for the decision to submit the manuscript for publication.

The authors state that none of the material contained in the submitted manuscript has been published previously. Preliminary results of this research project were presented as posters at the European Conference of Human Genetics, Paris, June 2013, and at the Fabry Expert Lounge, Rome, March 2014.

Conflict of Interest Declaration

Susana Ferreira has received unrestricted research grants and funding for research projects from Genzyme Corporation;

conference registration fees and travel grants from Genzyme Corporation and Shire Human Genetic Therapies.

Carlos Reguenga declares no conflicts of interest related to the subject matter of this manuscript.

João Paulo Oliveira is member of the European Advisory Board of the Fabry Registry, a global observational registry of patients with Fabry disease sponsored by Genzyme Corporation. He has received unrestricted research grants and funding for research projects from Genzyme Corporation; consulting honoraria and speaker's fees from Genzyme Corporation; conference registration fees and travel grants from Genzyme Corporation, Shire Human Genetic Therapies and Amicus Therapeutics.

References

- Baptista MV, Ferreira S, Pinho EMT et al (2010) Mutations of the *GLA* gene in young patients with stroke: the PORTYSTROKE study—screening genetic conditions in Portuguese young stroke patients. *Stroke* 41:431–436
- Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* 69:3613–3634
- Brady RO, Gal AE, Bradley RM, Martensson E (1967) The metabolism of ceramide trihexosides. I. Purification and properties of an enzyme that cleaves the terminal galactose molecule of galactosylgalactosylglucosylceramide. *J Biol Chem* 242: 1021–1026
- Brogan J, Li F, Li W, He Z, Huang Q, Li CY (2012) Imaging molecular pathways: reporter genes. *Rad Res* 177:508–513
- Davies JP, Winchester BG, Malcolm S (1993) Sequence variations in the first exon of alpha-galactosidase A. *J Med Genet* 30:658–663
- Desnick RJ, Ioannou YA, Eng CM (2001) α -Galactosidase A deficiency: Fabry disease In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds). *The metabolic and molecular basis of inherited disease*. McGraw-Hill, New York, pp. 3733–3774
- Ferreira S, Ortiz A, Germain DP et al (2015) The alpha-galactosidase A p.Arg118Cys variant does not cause a Fabry disease phenotype: Data from individual patients and family studies. *Mol Genet Metab* 114(2):248–258.
- Ferri L, Guido C, la Marca G et al (2012) Fabry disease: polymorphic haplotypes and a novel missense mutation in the *GLA* gene. *Clin Genet* 81:224–233
- Fitzmaurice TF, Desnick RJ, Bishop DF (1997) Human alpha-galactosidase A: high plasma activity expressed by the -30G→A allele. *J Inher Metab Dis* 20:643–657
- Garman SC (2007) Structure-function relationships in alpha-galactosidase A. *Acta Paediatr (Suppl)* 96:6–16
- Germain DP (2010) Fabry disease. *Orphanet J Rare Dis* 5:30
- Jacobs JL, Dinman JD (2004) Systematic analysis of bicistronic reporter assay data. *Nucleic Acids Res* 32:e160
- Kint JA (1970) The enzyme defect in Fabry's disease. *Nature* 227: 1173
- Macville M, Schrock E, Padilla-Nash H et al (1999) Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. *Cancer Res* 59:141–150
- Niemann M, Rolfs A, Giese A et al (2013) Lyso-Gb3 indicates that the alpha-galactosidase A mutation D313Y is not clinically relevant for fabry disease. *JIMD Rep* 7:99–102

- Oliveira JP, Ferreira S, Barcelo J et al (2008a) Effect of single-nucleotide polymorphisms of the 5' untranslated region of the human alpha-galactosidase gene on enzyme activity, and their frequencies in Portuguese caucasians. *J Inherit Metab Dis* 31(Suppl 2):S247–S253
- Oliveira JP, Ferreira S, Reguenga C, Carvalho F, Mansson JE (2008b) The g.1170C>T polymorphism of the 5' untranslated region of the human alpha-galactosidase gene is associated with decreased enzyme expression—evidence from a family study. *J Inherit Metab Dis* 31(Suppl 2):S405–S413
- Richard L, Velasco P, Detmar M (1999) Isolation and culture of microvascular endothelial cells. *Methods Mol Med* 18:261–269
- Saifudeen Z, Desnick RJ, Ehrlich M (1995) A mutation in the 5' untranslated region of the human alpha-galactosidase A gene in high-activity variants inhibits specific protein binding. *FEBS Lett* 371:181–184
- Samac S, Rice JC, Ehrlich M (1998) Analysis of methylation in the 5' region of the human alpha-galactosidase A gene containing a binding site for methylated DNA-binding protein/RFX1-4. *Biol Chem* 379:541–544
- Schneider U, Schwenk HU, Bornkamm G (1977) Characterization of EBV-genome negative “null” and “T” cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. *Int J Cancer* 19:621–626
- Shaw G, Morse S, Ararat M, Graham FL (2002) Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *Faseb j* 16:869–871
- von Scheidt W, Eng CM, Fitzmaurice TF et al (1991) An atypical variant of Fabry's disease with manifestations confined to the myocardium. *N Engl J Med* 324:395–399
- Warnock DG (2005) Fabry disease: diagnosis and management, with emphasis on the renal manifestations. *Curr Opin Nephrol Hypertens* 14:87–95
- Wu G, Yan B, Wang X et al (2008) Decreased activities of lysosomal acid alpha-D-galactosidase A in the leukocytes of sporadic Parkinson's disease. *J Neurol Sci* 271:168–173
- Wu G, Pang S, Feng X et al (2011) Genetic analysis of lysosomal alpha-galactosidase A gene in sporadic Parkinson's disease. *Neurosci Lett* 500:31–35
- Yasuda M, Shabbeer J, Benson SD, Maire I, Burnett RM, Desnick RJ (2003) Fabry disease: characterization of alpha-galactosidase A double mutations and the D313Y plasma enzyme pseudo-deficiency allele. *Hum Mutat* 22:486–492
- Zhang XY, Asiedu CK, Supakar PC, Khan R, Ehrlich KC, Ehrlich M (1990) Binding sites in mammalian genes and viral gene regulatory regions recognized by methylated DNA-binding protein. *Nucleic Acids Res* 18:6253–6260