A large-scale screening of the normalized mammalian mitochondrial gene expression profiles.

Anisimov, Sergey

Published in:
Genetical Research

DOI:
10.1017/S0016672305007718

2005

Genet. Res., Camb. (2005), **86**, *pp.* 127–138. *With 7 figures.*
doi:10.1017/S0016672305007718 Printed in the United Kingdom

A large-scale screening of the normalized mammalian mitochondrial gene expression profiles

SERGEY V. ANISIMOV*
Neuronal Survival Unit, Wallenberg Neuroscience Center, Lund University, 221 84 Lund, Sweden

(*Received 24 January 2005 and in revised form 7 July 2005*)

Summary

Mammalian mitochondrial genomes are organized in a conserved and extremely compact manner, encoding molecules that play a vital role in oxidative phosphorylation (OXPHOS) and carry out a number of other important biological functions. A large-scale screening of the normalized mitochondrial gene expression profiles generated from publicly available mammalian serial analysis of gene expression (SAGE) datasets (over 17·7 millions of tags) was performed in this study. Acquired SAGE libraries represent an extensive range of human, mouse, rat, bovine and swine cell and tissue samples (normal and pathological) in a variety of conditions. Using a straightforward *in silico* algorithm, variations in total mitochondrial gene expression, as well as in the expression of individual genes encoded by mitochondrial genomes are addressed, and common patterns in the species- and tissue-specific mitochondrial gene expression profiles are discussed.

1. Introduction

Mitochondrial function is essential to eukaryotic cells. Mitochondria carry out a variety of metabolic processes, their major functions being to supply the cell with energy generated by oxidative phosphorylation (OXPHOS) and to participate in protein synthesis (Fernandez-Silva *et al.*, 2003). Both parts of mammalian genomes: nuclear DNA (nDNA) and maternally inherited mitochondrial DNA (mtDNA) are involved in OXPHOS (Enriquez *et al.*, 1999). OXPHOS features, regulated in part by gene expression, vary significantly depending on cellular activities (Kagawa *et al.*, 1999). Considering the direct involvement of mitochondria-encoded genes in many basic cellular processes (such as heat production, protein folding, calcium homeostasis, apoptosis, ageing and carcinogenesis) and human diseases, directed studies of mitochondrial gene expression in tissues and cell types might expand our understanding of the biological relevance of mitochondrial transcription.

Serial Analysis of Gene Expression (SAGE) is a sequence-based technique to study messenger RNA (mRNA) transcripts in cell populations (Velculescu *et al.*, 1995). Two major principles underline SAGE: (i) short (10 bp) expressed sequenced tags (ESTs) are sufficient to identify individual gene products, and (ii) multiple tags (up to 100) can be concatenated (i.e. linked together as a chain or cascade) and identified by sequence analysis. SAGE results are reported in either absolute or relative numbers of tags, which permits direct comparisons between libraries. Advances in SAGE protocol and sequencing techniques, supplemented with the expansion of sequence information available in public databases, have further augmented the perception of SAGE as an important tool of modern molecular biology widely used in a number of applications.

In the current study, mitochondrial gene expression was evaluated in a wide array of normal and pathological samples using the complete set of currently available SAGE data for mammals, including *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Bos taurus* and *Sus scrofa*. The aim of this study was to perceive variations and commonalities in species-specific, tissue-specific and cell-state-specific profiles of mitochondrial gene expression in mammals. Taken together, over 17·7 million SAGE tags derived from SAGE libraries generated with commonly used *NlaIII* (Velculescu *et al.*, 1995) and *Sau3A I* (Virlon *et al.*,

* Corresponding author: Sergey V. Anisimov, Neuronal Survival Unit, Department of Experimental Medical Science, Wallenberg Neuroscience Center, Lund University, BMC A10 221 84 Lund, Sweden. Tel: +46 46 222 4369. Fax: +46 46 222 0531. e-mail: Sergey.Anisimov@med.lu.se

1999) anchoring enzymes were subjected to the analysis. Clustering analysis has allowed identification of some common patterns in the profiles of mitochondrial gene expression and the potential biological significance of variations in the total mitochondrial gene expression is discussed.

2. Methods

(i) *SAGE*

SAGE libraries subjected to the analysis were generally constructed as previously described (Velculescu *et al*., 1995). Briefly, cDNA is prepared from poly $(A)^+$ mRNA with biotinylated oligo $(dT)_{18}$ and restriction enzymes of type I (*NlaIII*/*Sla3A I*) and type II (*BsmFI*) are used to generate tags. After a concatemerization step, the DNA is cloned into pZeRO-1, PCR-amplified and sequenced. Sequencing results are typically analysed using the SAGE 2000 software (available at http://www.sagenet.org).

The complete set of tag abundance data for publicly available SAGE libraries was downloaded from the Gene Expression Omnibus (GEO) database (National Center of Biotechnology Information (NCBI); http://www.ncbi.nlm.nih.gov/geo/; Edgar *et al*., 2002). Duplicate dimers were extracted, and all linker-generated tags were excluded from the analysis. Furthermore, all publicly available SAGE libraries constructed by the author and co-workers (namely, human: GSM1515 and GSM1515, Potapova *et al*., 2002; mouse: GSM580, GSM1681, GSM1682, GSM1683, GSM1684, GSM7759, GSM11348 and GSM11349, Anisimov *et al*., 2002*a*, *b*, *c* and unpublished data) were subjected to a 'clean-up' procedure (Anisimov & Sharov, 2004), in which clones containing four or fewer tags were excluded. All SAGE libraries were annotated and sorted based on the number of tags sequenced. Non-informative $(A)_{10}$ sequences were extracted from SAGE libraries when detected, and tags per million (tpm) values were recalculated accordingly as the transcript's raw tag count divided by the number of reliable tags in the library and multiplied by 10^6.

(ii) *Mitochondrial genomes*

Complete mammalian mitochondrial genomes generally established as the 'reference' sequences were downloaded from the GenBank database (NCBI; http://w ww.ncbi.nlm.nih.gov/entrez/) as the following: *Homo sapiens*, accession number X93334 (a Caucasian from southern Sweden (Lund, Skåne); Arnason *et al*., 1996; Nilsson *et al*., 2003) and NC_001807 (an African (Yoruba ethnic group); Ingman *et al*., 2000); *Mus musculus*, AB042432 (Mizutani *et al*., 2001) and J01420 (Bibb *et al*., 1981); *Rattus norvegicus*, AY172581 (Rat Genome Sequencing Project Consortium, 2004) and NC_001665 (Gadaleta *et al*., 1989); *Bos taurus*, AY526085; and *Sus scrofa*, NC_000845 (Lin *et al*., 1999). Single nucleotide polymorphisms (SNPs) or genetic variations in mitochondrial sequences were addressed using the BLAST2 alignment tool (NCBI; http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html; Tatusova & Madden, 1999).

(iii) *SAGE tag annotation*

Complete mammalian mitochondrial genomes were searched for *NlaIII* and *Sau3A I* SAGE anchoring enzymes recognition sequences that were then annotated to the sequences for distinctive mitochondrial genes, ribosomal and transfer RNAs. These were then matched to individual SAGE libraries using the MS Access 2002 software package Query function. Individual queries were merged using MS Excel 2002 software, and calculations of relative expression of mitochondrial transcripts were performed using normalized tpm values.

(iv) *Clustering analysis*

Clustering analysis was performed using EPCLUST clustering, visualization and analysis software (http://ep.ebi.ac.uk/EP/EPCLUST/; credits to J. Vilo, P. Kemmeren & M. Kapushesky). Hierarchical clustering was performed via average linkage (average distance, UPGMA (unweighted pair group method with arithmetic mean)) clustering based on the linear correlation-based distance (Pearson, centred) method. K-mean clustering analysis was performed with initial cluster centres chosen by most distant (average) transcripts. The optimal number of clusters was determined empirically to produce the most balanced ratio of entries per cluster to the number of clusters.

To discover trends in the differences in mitochondrial gene expression in normal versus cancer and undifferentiated versus differentiated SAGE samples, average tag frequency values were calculated for the pools of SAGE libraries (e.g. prostatic cancer, 8 SAGE libraries). After ensuring that the average number of analysed tags per library was within the same order of magnitude in the groups being compared (e.g. normal prostatic tissue, 48 729 tags; prostatic cancer, 49 033 tags), average tag frequency values were compared using the following equation: $[N1-k(N1)^{1/2}]-[N1-k(N1)^{1/2}]>0$, where N1 represents the larger tag frequency and $k=3{\cdot}29$ for $P<0{\cdot}001$, $k=2{\cdot}58$ for $P<0{\cdot}01$, $k=1{\cdot}96$ for $P<0{\cdot}05$ (Madden *et al*., 1997).

3. Results

Tag frequency data for the complete set of publicly available mammalian SAGE libraries was downloaded

Table 1. *Summary of SAGE catalogs analyzed*

Species	GEO platform	Anchoring enzyme	Number of SAGE catalogs	Number of SAGE tags
Homo sapiens	GPL4	*NlaIII*	239	13 672 568
	GPL6	*Sau3A I*	10	457 493
Mus musculus	GPL11	*NlaIII*	43	2 029 235
	GPL275	*Sau3A I*	9	218 412
Rattus norvegicus	GPL23	*NlaIII*	10	718 162
Bos taurus	GPL223	*NlaIII*	10	586 001
Sus scrofa	GPL1270	*NlaIII*	2	84 616
Total			323	17 766 487

GEO, Gene Expression Omnibus database (NCBI).

from the GEO database, for a total of 282 human, 96 mouse, 16 rat, 11 bovine and 2 swine SAGE libraries (see Supplemental Table 1.; all Supplemental Tables can be found at *CUPwebsite*). Among these, 2 human (Potapova *et al*., 2002) and 8 mouse (Anisimov *et al*., 2002*a*, *b*, *c* and unpublished data) SAGE libraries constructed and sequenced following the original protocol (Velculescu *et al*., 1995), with a number of minor modifications in the protocol (Kenzelmann & Mühlemann, 1999; Anisimov *et al*., 2002*c*), were analysed after being subjected to tag extraction procedures (Anisimov & Sharov, 2004). Taken together, acquired SAGE libraries have represented a wide spectrum of cell and tissue samples (normal and pathological) in a variety of conditions. The following criteria were applied to these when selecting libraries for the analysis of mitochondrial gene expression, and SAGE libraries were selected only if they represented (i) genetically unmodified species or cell cultures, (ii) untreated samples and (iii) SAGE libraries with a total of at least 10 000 reliable tags and a complete dataset available. For example, samples GSM14917 (human) and GSM3677 (mouse), representing telomerase overexpression-immortalized primary fibroblast culture and the A1 × RAG-1$^{-/-}$ TCR transgenic mouse, respectively, were excluded since they did not satisfy criterion (i). Similarly, samples GSM3832 (murine bone marrow-derived dendritic cells, LPS-treated) and GSM718 (normal human mammary gland epithelium; 1423 tags) were excluded as not satisfying criteria (ii) and (iii), respectively.

Three hundred and twenty-three SAGE libraries generated from human samples and mammalian species widely employed in general biology and molecular genetics were selected for further analysis. Together, these libraries consist of over 17·7 million SAGE tags. For *Homo sapiens* and *Mus musculus*, the GEO database provided SAGE libraries constructed using both *NlaIII* and *Sau3A I* anchoring enzymes common in SAGE applications, while rat, bovine and swine samples were all constructed using *NlaIII* anchoring enzyme (Table 1).

Complete 'reference' sequences of mammalian mitochondrial genomes were downloaded from the GenBank database. In cases where more than one complete mitochondrial genome sequence was established as 'reference', all these were analysed. Sequences were compared pair-wise using the BLAST2 sequence alignment tool, and a number of SNPs or genetic variations within anchoring enzyme (*NlaIII* and *Sau3A I*) recognition sites and corresponding SAGE tags were identified in human and rat complete mitochondrial genome sequences (Figs 1, 2). Sequences were formatted and individual ribosomal RNAs, mitochondrial genes and transfer RNAs were identified in these, followed by the identification of anchoring enzyme (*NlaIII*, all species; *Sau3A I*, *Homo sapiens* and *Mus musculus*) recognition sites. SAGE tags 10 bp long located immediately 3′ to the detected anchoring enzyme recognition sites were then retrieved from these sequences. Taken together, 47 and 25 human (for *NlaIII* and *Sau3A I* anchoring enzymes, respectively), 55 and 35 mouse (*NlaIII* and *Sau3A I*), 59 rat, 44 bovine and 53 swine (all – *NlaIII*) SAGE tags associated with mitochondrial H-strands were identified (see Supplemental Table 2). A number of these (1 and 1 for human, *NlaIII* and *Sau3A I*, respectively and 9 for rat, *NlaIII*) were originated from genetic variations and/or SNPs within the tags sequences. Two SAGE tags matching NADH dehydrogenase 1 (*mt-Nd1*) and cytochrome *c* oxidase subunit I (*mt-Co1*) rat mitochondrial genes were represented by 4 variant tags each, while tags matching rat *mt-Co2, -Nd4 and -Nd5*, human *mt-Nd1* (*NlaIII*) and *mt-Nd6* (*Sau3A I*) mitochondrial genes were each represented by 2 variant tags (Fig. 1, 2B). Supplemental Fig. 1 (all Supplemental Figures can be found at *CUPwebsite*) shows the relative dimensions and sequence breakdown, as well as the distribution and relative positions of SAGE *NlaIII* anchoring enzyme restriction sites, in human, mouse, rat, bovine and swine mitochondrial genomes, illustrating the abundance of mitochondrial gene expression data that could be extracted from SAGE samples.

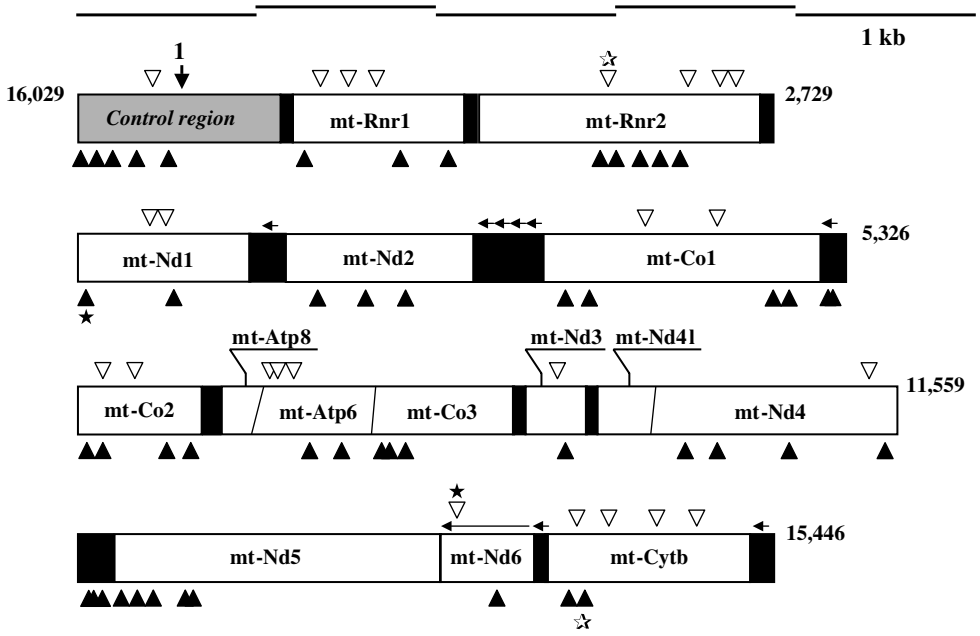

Fig. 1. Distribution of SAGE anchoring enzymes restriction sites in the human mitochondrial genome. Circular mitochondrial DNA is plotted in linear form. Numbers at the end of each segment correspond to the base number of the RNA encoded by the heavy strand (H strand); oblique strokes indicate overlaps in the coding sequences; the vertical arrow (↓) indicates the starting position; scale bars correspond to 1 kb. Two complete human mitochondrial DNA samples representing a Caucasian from the southern Sweden (Lund, Skåne; GenBank acc. no. X93334) and an African (Yoruba ethnical group; NC_001807) were used for analysis. Shaded areas represent tRNAs (black) and control region (D-loop; grey). Left-headed arrow (←) indicate areas encoded by the light strand (L strand). Filled triangles under the bars (▲) and empty triangles above the bars (▽) indicate *NlaIII* and *Sau3A I* restriction sites, respectively. Filled asterisks (★) indicate potential single nucleotide polymorphisms (SNPs) in SAGE tags for *NlaIII* (mt-Nd1, $^{G}/_{A}$CCAACCTCC) and *Sau3A I* (mt-Nd6, CTCCCGAAT$^{C}/_{G}$) anchoring enzymes. Empty asterisks (☆) indicate potential SNPs in the recognition sequences of *NlaIII* (mt-Cytb, CAT$^{G}/_{A}$) and *Sau3A I* (mt-Rnr2, GAT$^{C}/_{T}$) anchoring enzymes.

Supplemental Table 3 lists the size of individual mammalian mitochondrial genome sequences. Owing to the existing overlap in the coding sequences of *mt-Atp8*/*Atp6*, *mt-Atp6*/*Co3*, *mt-Nd4l*/*Nd4* (all species) and *mt-Nd5*/*Nd6* (all but *Homo sapiens*) genes, sequence lengths of these genes were summed.

Primary analysis of mitochondrial gene expression profiles has demonstrated great variability in total mitochondrial gene expression. In the assessed datasets it ranged from 0·36% to 25·44% (calculated to the total (mitochondrial + nuclear) gene expression), with an average value of 4·59% for human, and from 0·46% to 22·64%, with an average value of 4·76%, for mouse. From 249 human SAGE libraries tested, 6 had a fraction of mitochondrial gene expression to the total number of analysed tags below 1%, and in 24 SAGE libraries it was above 10% (Table 2). Similarly, of 52 murine SAGE libraries tested, a fraction of mitochondrial gene expression was below 1% in 6 and above 10% in 13 SAGE libraries (Table 3). At the same time, initially smaller number of SAGE libraries generated from rat (10), bovine (10) and swine (2) samples did not lead to identification of many samples with extreme values of total mitochondrial gene expression. Average values of total mitochondrial gene expression were 7·19% for *Rattus norvegicus*, 2·14% for *Bos taurus* and 17·88% for *Sus scrofa*. Three of the rat samples (GSM581, Extraocular muscle; GSM1679 and GSM12532, Hippocampus) demonstrated a fraction of mitochondrial gene expression >10% (15·72%, 13·39% and 13·25% to the total number of analysed tags, respectively), while in 2 bovine samples (GSM11027, Madin-Darby bovine kidney (MDBK) cells and GSM24604, Circulating γ/δ T cells) this fraction was <1% (0·97% and 0·42%, respectively).

Supplemental Table 2 shows a distribution of SAGE tag frequencies in mammalian mitochondrial genomes. Clearly, this distribution is reflective of the relative expression of individual mitochondrial transcripts in the sample. In many samples, a significant proportion of total mitochondrial expression was granted by only a few transcripts. In the most abundant array of human *NlaIII*-derived SAGE libraries, the *mt-Co2* most 3′ tag alone (CCCATCGTCC) accounted for 12·5% of the total mitochondrial gene expression (by the averaged data). Relative frequency values for individual tags have reached extreme ranges in certain libraries: in a human PC3 cell SAGE library (Potapova *et al*., 2002), for example, the same tag has accounted for 26·7% of total mitochondrial gene expression (and 0·37%

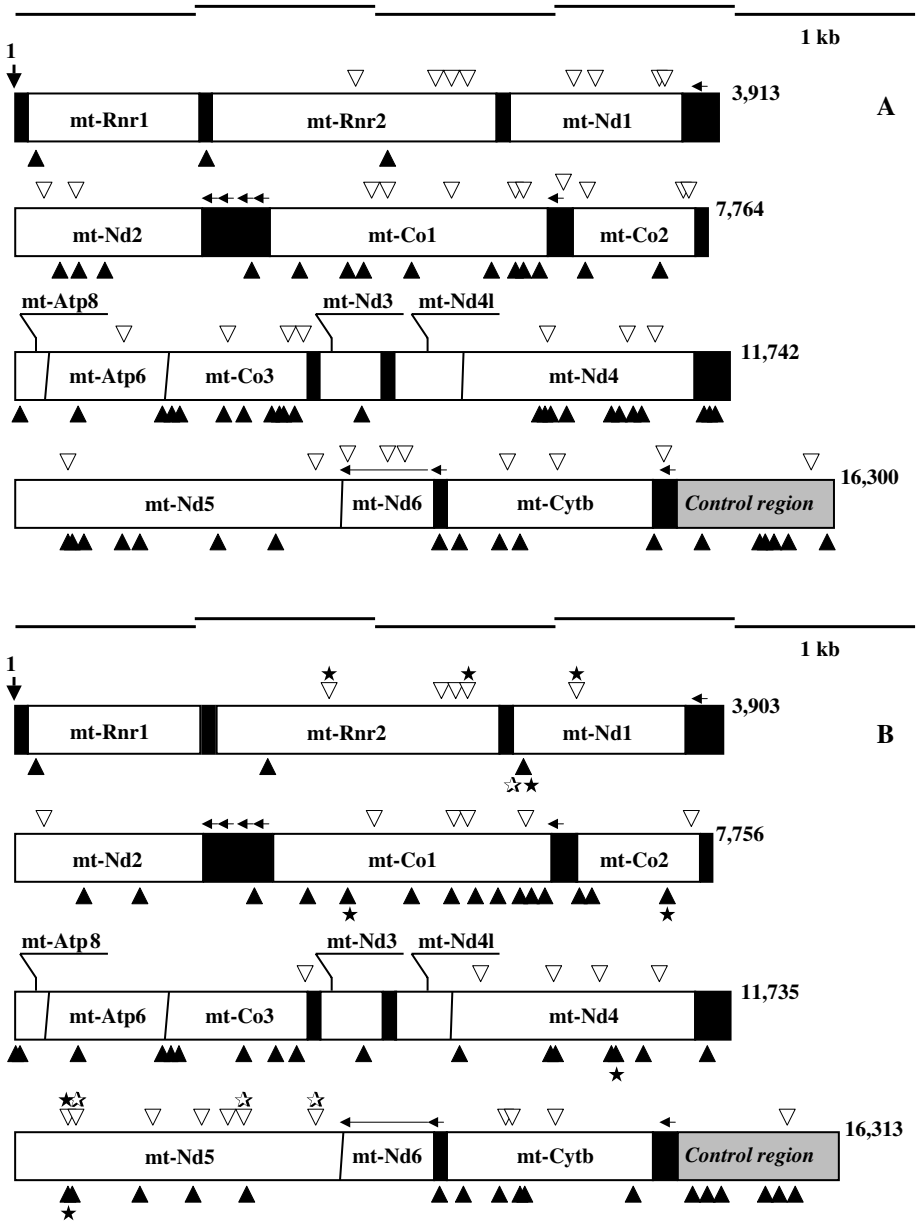

Fig. 2. Distribution of SAGE anchoring enzyme restriction sites in mouse (*A*) and rat (*B*) mitochondrial genome. Circular mitochondrial DNA is plotted in linear form. Numbers at the end of each segment correspond to the base number of the RNA encoded by the heavy strand (H strand); oblique strokes indicate overlaps in the coding sequences; the vertical arrow (↓) indicate the starting position; scale bars correspond to 1 kb. Two complete mouse (GenBank accession nos. AB042432 and J01420) and rat (AY172581 and NC_001665) mitochondrial DNA samples representing *Mus musculus* and *Rattus norvegicus* strains, respectively, were used for each analysis. Shaded areas represent tRNAs (black) and control region (D-loop; grey). Left-headed arrows (←) indicate areas encoded by the light strand (L strand). Filled triangles under the bars (▲) and empty triangles above the bars (▽) indicate *NlaIII* and *Sau3A I* restriction sites, respectively. Filled asterisks (★) indicate potential single nucleotide polymorphisms (SNPs) or genetic variations in SAGE tags for *NlaIII* (mt-Nd1, $G^{C}/_{G}C^{T}/_{C}$CTCACCC; mt-Co1, CTGG$^{A}/_{G}G^{C}/_{T}$ATC; mt-Co2, AGCC$^{G}/_{A}$TCCCT; mt-Nd4, GGAGC$^{T}/_{C}$ACAA; mt-Nd5, ATCAAT$^{C}/_{T}$ATA) and *Sau3A I* (mt-Rnr2, C$^{A}/_{-}$TAG$^{A}/_{-}$AATA and T$^{G}/_{A}$AGT$^{T}/_{C}/^{A}/_{-}/^{G}/^{A}/_{-}$; mt-Nd1, $^{A}/_{G}$GGATGAGC$^{A}/_{C}$; mt-Cytb, AT$^{T}/_{C}$CTACTAT) anchoring enzymes in rat mitochondrial sequences. Empty asterisks (☆) indicate potential SNPs in the recognition sequences of *NlaIII* (mt-Nd1, CAT$^{G}/_{A}$) and *Sau3A I* (mt-Nd5, GA$^{T}/_{C}$C, $^{C}/_{G}$ATC and GA$^{C}/_{T}{}^{T}/_{C}$) anchoring enzymes in rat mitochondrial sequences.

of total gene expression), while tags matching *tRNA-Ser* and *mt-Co1* have accounted for 22·8% and 14·3% of total mitochondrial gene expression, respectively. At the same time, many other SAGE libraries have different mitochondrial expression profiles. Predictably, similar diversity was observed in other mammalian species and in datasets generated using an alternative SAGE anchoring enzyme (*Sau3A I*).

Table 2. *Extreme values of total mitochondrial gene expression in human SAGE libraries*

N	SAGE library	AE	Sample	Mt[a]
			<1·00%	
1	GSM14751	*NlaIII*	Skin, melanoma	0·36%
2	GSM754	*NlaIII*	Prostate, cell line	0·42%
3	GSM41382	*NlaIII*	Cartilage chondrosarcoma grade 3	0·80%
4	GSM668	*NlaIII*	Kidney, embryonic cell line 293, uninduced cells	0·87%
5	GSM675	*NlaIII*	Ovary	0·97%
6	GSM32266	*NlaIII*	Microvascular endothelial cells exposed to sustained high shear stress	0·99%
			>10·00%	
1	GSM786	*NlaIII*	Frontal cortex	10·14%
2	GSM7800	*NlaIII*	Primary gastric cancer tissue, poorly differentiated (scirrhous type)	10·24%
3	GSM1499	*NlaIII*	Normal heart tissue	10·30%
4	GSM14780	*NlaIII*	Normal gastric epithelial tissue from the antrum	10·37%
5	GSM725	*NlaIII*	Mammary gland, adenocarcinoma	10·75%
6	GSM14742	*NlaIII*	Brain cortex, oligodendroglioma	10·76%
7	GSM2451	*NlaIII*	Brain, astrocytoma	11·08%
8	GSM14783	*NlaIII*	Brain anaplastic medulloblastoma	11·15%
9	GSM14796	*NlaIII*	Brain, substantia nigra	11·49%
10	GSM728	*NlaIII*	Normal colonic epithelium	11·75%
11	GSM695	*NlaIII*	Brain, normal, greater than 95% white matter	11·77%
12	GSM763	*NlaIII*	Brain, normal	12·11%
13	GSM755	*NlaIII*	Adenocarcinoma, colon	12·23%
14	GSM729	*NlaIII*	Normal colonic epithelium	13·53%
15	GSM10427	*Sau3A I*	Kidney distal convoluted tubule	13·59%
16	GSM676	*NlaIII*	Brain, normal, greater than 95% white matter	13·64%
17	GSM10425	*Sau3A I*	Kidney medullary thick ascending limb	13·73%
18	GSM784	*NlaIII*	Normal gastric epithelial tissues	13·79%
19	GSM819	*NlaIII*	Normal muscle (old)	16·01%
20	GSM824	*NlaIII*	Normal muscle (young)	16·66%
21	GSM14771	*NlaIII*	Placenta	17·95%
22	GSM10424	*Sau3A I*	Kidney proximal straight tubule	18·79%
23	GSM708	*NlaIII*	Normal kidney tissue	19·31%
24	GSM713	*NlaIII*	Brain, normal thalamus	25·44%

Indexes (GSM#) represent GEO database accession numbers for SAGE libraries.
AE, anchoring enzyme.
[a] Total mitochondrial gene expression (as percentage of the total number of analyzed tags).

To address the cellular basis for the evident dissimilarity in mitochondrial gene expression profiles in an integrated way, clustering analysis was performed using EPCLUST clustering, visualization and analysis software. Hierarchical clustering was performed using a correlation-measure-based distance (Pearson, centred)/average linkage (average distance) clustering method and hierarchical trees built for individual datasets (Fig. 3; Supplemental Figs 2, 3). The optimal number of K-mean clusters (*Homo sapiens*: *NlaIII*, 10, *Sau3A I*, 4; *Mus musculus*: *NlaIII*, 5, *Sau3A I*, 3; *Rattus norvegicus*: *NlaIII*, 3; *Bos taurus*: *NlaIII*, 3) was determined empirically and these have been built for individual datasets resulting in pattern recognition (Figs 3, 4 and Supplemental Fig. 3). Notable diversity was observed among possible transcription profiles (Fig. 4, Supplemental Fig. 3), with either relatively high expression characteristics for just a few individual transcripts (e.g. Supplemental Fig. 3B, Cluster 4; Supplemental Fig.3J), or with expression profiles being more proportional (e.g. Fig. 4, Cluster 5). It was anticipated that samples generated from the tissues/cell types with similar characteristics would fall into the same cluster, as demonstrated in Fig. 3 where a complete dataset for human *Sau3A I* SAGE libraries is comprised of only two major transcription profiles, clearly associated with either kidney tissue structures or U937 monoblast/early monocyte cell line (Fig. 3). However, the analysis of more complex datasets failed to provide solid support for that hypothesis, and the analysis of abundant SAGE datasets based on sample origin (tissue or sample type, etc.) has identified very few groups of these falling into the same cluster. For example, while in a human *NlaIII* SAGE dataset broken into 10 K-mean clusters 5 of 12 samples falling in Cluster 7 were generated from normal mammary gland tissue (Fig. 4), one more sample of the same origin was detected in Cluster 5 and 2 in Cluster 10. Similarly, 12 of 20 samples falling in Cluster 2 represented either carcinomas (e.g. GSM14753, GSM1730) or tumours generating metastatic carcinomas (GSM696) or

Table 3. *Extreme values of total mitochondrial gene expression in mouse SAGE libraries*

N	SAGE library	AE	Sample	Mt[a]
			< 1·00%	
1	GSM3833	*NlaIII*	Bone marrow-derived dendritic cells	0·46%
2	GSM3686	*NlaIII*	CD4 + CD25 + spleen cells	0·47%
3	GSM1684	*NlaIII*	P19 EC cells, differentiation Day 3 + 3·0	0·69%
4	GSM34768	*NlaIII*	Total testis	0·72%
5	GSM3828	*NlaIII*	Thymoma cell line	0·83%
6	GSM5436	*NlaIII*	Fetal testis somatic cells	0·95%
			> 10·00%	
1	GSM12541	*NlaIII*	Hippocampus: long attack latency (LAL) mice	10·33%
2	GSM17430	*NlaIII*	Cerebellum	10·37%
3	GSM12540	*NlaIII*	Hippocampus: short attack latency (SAL) mice	10·40%
4	GSM4871	*Sau3A I*	Whole brain (8- to 12-week-old)	10·73%
5	GSM4876	*Sau3A I*	Somatosensory cortex (8- to 12-week-old)	10·80%
6	GSM4875	*Sau3A I*	Nucleus accumbens (8- to 12-week-old)	14·35%
7	GSM9194	*NlaIII*	Whole kidney	14·43%
8	GSM4873	*Sau3A I*	Lateral striatum (8- to 12-week-old)	16·20%
9	GSM10430	*Sau3A I*	Adult kidney (8 to 10-weeks-old)	17·21%
10	GSM4874	*Sau3A I*	Medial striatum (8- to 12-week-old)	17·63%
11	GSM1681	*NlaIII*	Heart (young)	18·68%
12	GSM15116	*NlaIII*	Heart (old)	22·27%
13	GSM23547	*NlaIII*	Heart	22·64%

Indexes (GSM#) represent GEO database accession numbers for SAGE libraries.
AE, Anchoring enzyme.
[a] Total mitochondrial gene expression (as a percentage of the total number of analyzed tags).

preceding carcinoma development (GSM780). At the same time, many other carcinoma samples were detected in further clusters. Only one group of samples that fell entirely into a single cluster was identified within the course of analysis. Mitochondrial profiles of all 4 SAGE samples generated from human retina (*NlaIII* dataset; GSM571-574), including human retinal pigment epithelium (RPE), peripheral and central retina (Sharon *et al*., 2002), were found in Cluster 2 (Fig. 4). Notably, all these samples also fell into a single cluster when the dataset (which contains 239 entries) was analysed with the number of K-mean clusters increased up to 51.

4. Discussion

The mammalian mitochondrial genome is a circular molecule organized in a conserved and extremely compact manner. The mitochondrial DNA (mtDNA) encodes 2 ribosomal RNAs and 22 transfer RNAs that play an essential role in mitochondrial translation. A non-coding mitochondrial control region contains the main regulatory sequences for transcription and replication initiation, including a triple-stranded D-loop. Though the great majority of mitochondrial proteins are encoded by the nuclear genome, mtDNA encodes 13 proteins involved in oxidative phosphorylation (for review see Hiendleder & Wolf, 2003 and Fernandez-Silva *et al*., 2003). Individual mammalian mitochondrial-genome-encoded genes have demonstrated a number of other important biological functions, such as serving as histocompatibility antigens for the NADH dehydrogenase 1 (*mt-Nd1*) and cytochrome *c* oxidase subunit I (*mt-Co1*)-derived peptides (Loveland *et al*., 1990; Morse *et al*., 1996) or playing a role in protein folding for 12S and 16S ribosomal RNA (*mt-Rnr1* and *-Rnr2*; Sulijoadikusumo *et al*., 2001).

Mammalian mitochondrial genomes share a significant level of sequence homology and genes in these are organized in a virtually identical manner (Supplemental Fig. 1). Though there have been reports on the variations in mtDNA structure (triplex/duplex forms) in large versus small mammals (Annex & Williams, 1990), it is expected that basic mitochondrial gene expression patterns should be similar within mammalian species. Numerous SNPs or genetic variations were detected in human and rat reference mitochondrial sequences, and average allele distribution could be estimated from the pooled datasets. For example, the G-allele of human *mt-Nd6* gene CTCCCGAAT$^{C}/_{G}$ tag (*Sau3A I* anchoring enzyme) was not detected in either 11 SAGE libraries constructed using the pooled RNA from 9 individuals (Chabardes-Garonne *et al*., 2003) or human U937 monoblast/early monocyte cell line, while the A-allele of human *mt-Nd1* gene $^{G}/_{A}$CCAACCTCC tag (*NlaIII*) was detected in 7 SAGE libraries of 239 tested, with the average allele frequency being only 0·004 (Fig. 3A; Supplemental Table 2). Similarly, of

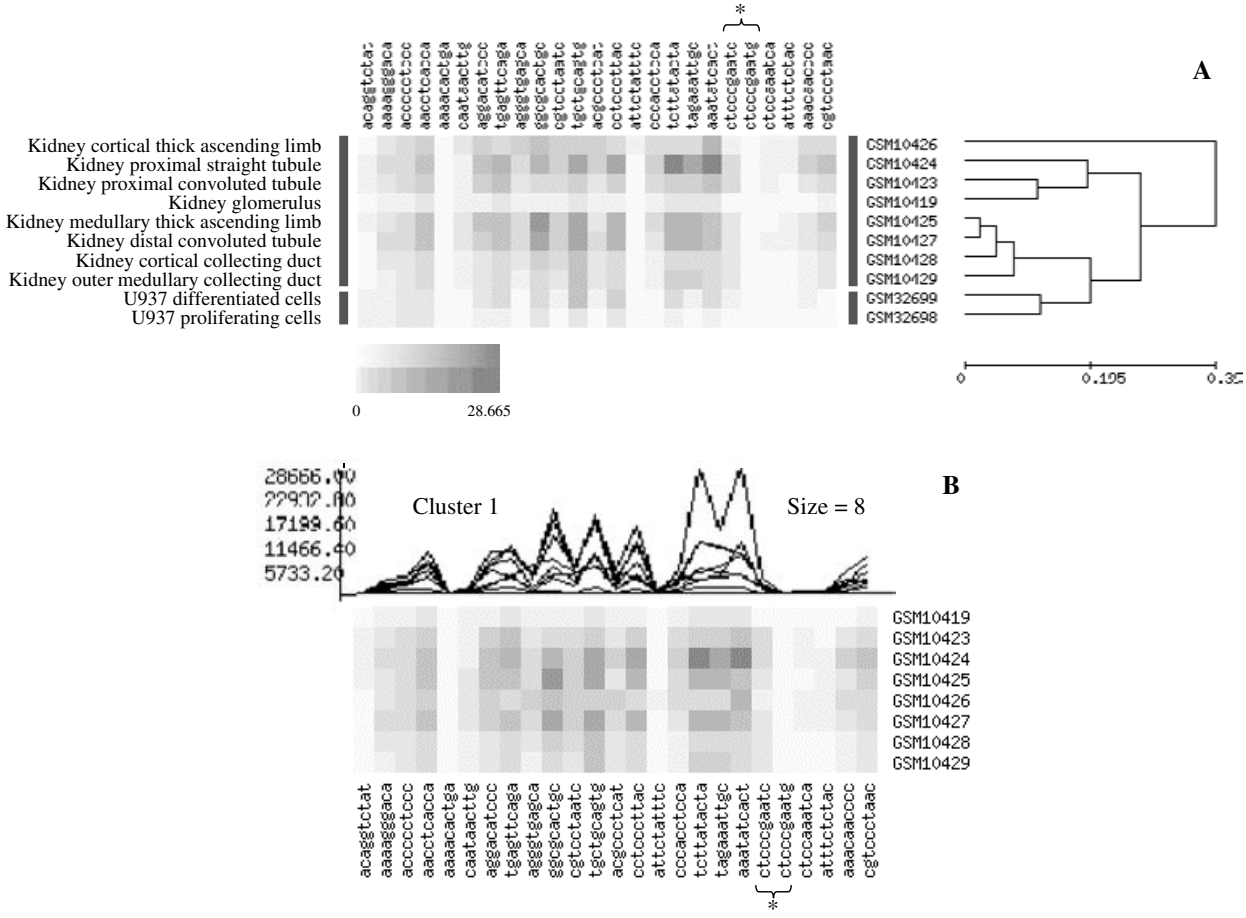

Fig. 3. Hierarchical cluster tree for human *Sau3A I*-derived SAGE dataset (10 libraries) (*A*) and an example of gene expression pattern recognized by K-mean clustering analysis (*B*). Vertical bars denote individual clusters. Harmonic shades of grey code (10 colors) is based on the normalized tpm values. Asterisk (*) denotes variant SAGE tags. Two major transcription profiles are identified in the dataset, being clearly associated with either kidney tissue structures (*B*: Cluster 1, Size = 8) or U937 monoblast/early monocyte cell line (Cluster 2, Size = 2).

4 potential allelic variants of rat *mt-Nd1* gene $G^{C}/_{G}C^{T}/_{C}$CTCACCC tag (*NlaIII*), only 3 were detected in 10 SAGE libraries, with the average allele frequency being 0·991, 0·005 and 0·004 for CT, GT and CC allelic variants, respectively (Supplemental Fig. 3E). In contrast, mitochondrial genome sequence identity was observed for mouse reference sequences, consistent with earlier observations and based on a common descent of all mouse inbred strains along the maternal line (Silver, 1995).

In the current study, a large-scale screening (323 SAGE libraries, over 17·7 million SAGE tags) of normalized mitochondrial gene expression profiles was performed. Normalization via recalculating tag frequency in individual SAGE libraries to tag per million (tpm) values was employed as a generally accepted normalization method widely used in the field. A great variability in the total mitochondrial gene expression was observed (ranging from 0·36% to 25·44% of the total gene expression in human and from 0·46% to 22·64% in mouse), consistent with the wide range of cell and tissue types screened. The results of this screening confirm that total mitochondrial gene expression varies considerably depending on a cell/tissue type. Since the majority of genomic-scale studies employ tissues rather than individual cells or pure cellular populations, the main type of cell present in the tissue determines the number of mitochondria per cell and tissue/organism metabolic activity will result in tissue-specific variations in the total mitochondrial gene expression and that of individual mitochondrial-genome-encoded genes. There are therefore considerations that discrete cell types are able to partially 'mask' each other's gene expression profiles within the tissue, due to the potential disparity of their own overlapped distinct gene expression profiles. It is thus important to evaluate these complex tissues and cell populations as an integral function of the total gene expression of the cells constituting the sample. Moreover, some SAGE tags could be associated with more than one transcript. As a search with the SAGEmap (http://www.ncbi.nlm.nih.gov/projects/SAGE) annotation tool demonstrates, this is true for many 10 bp SAGE

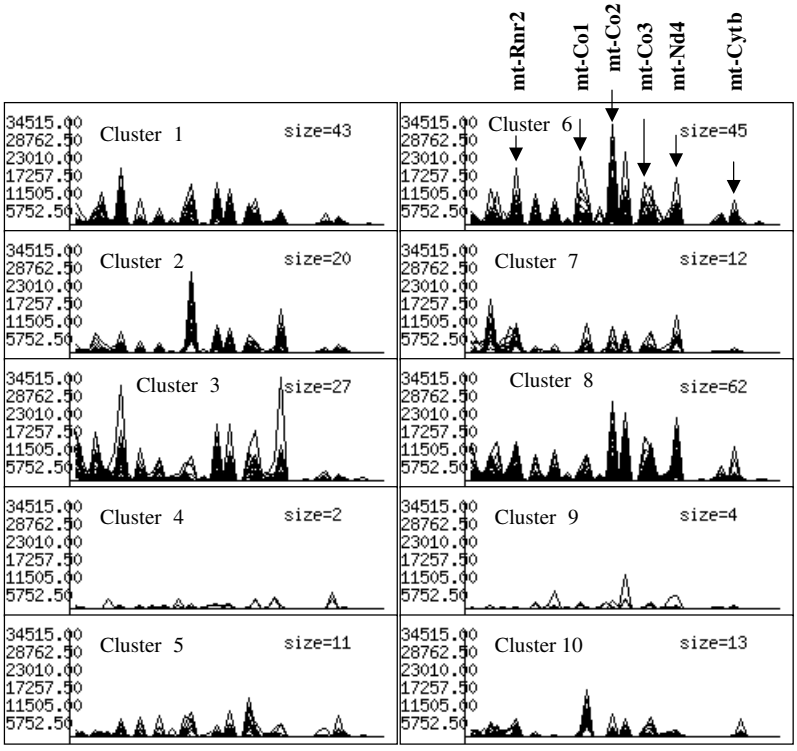

Fig. 4. Mitochondrial gene expression patterns as identified by K-mean clusters in human *NlaIII*-derived SAGE dataset (239 libraries). Individual SAGE tags (including variant tags) are spotted at abscissa in the order following mitochondrial H-strand (see Supplemental Table 2). Most 3′-located tags for *mt-Rnr2*, *-Co1*, *-2*, *-3*, *-Nd4* and *-Cytb* genes are denoted with arrows in the representative cluster. Normalized tpm values are spotted on the ordinate. Groups of numerous individual samples fall into the limited number of clusters with analogous expression profiles of individual mitochondrial transcripts. Note the diversity among possible transcription profiles with either relatively high expression characteristics for a few individual transcripts (e.g. Clusters 2 and 3), or with more proportional expression profiles (e.g. Clusters 5 and 9).

tags. For example, tag ATACTGACAT (most 3′ tag for murine *mt-Co3* gene (*NlaIII* dataset), highly abundant in many SAGE libraries) matches 543 different UniGene clusters, 15 of which are considered 'reliable'. Though this is a common concern in all studies utilizing conventional SAGE technology (i.e. not 'LongSAGE' or its analogues), it has be taken into account that since there is currently no way to distinguish between such multiple matches, a certain degree of bias could therefore be introduced into the analysis. Moreover, some mitochondrial-encoded genes do not bear restriction sites for anchoring enzymes routinely employed in SAGE (such as *NlaIII*). Examples include *mt-Atp8* and *mt-Nd4l* genes for *Homo sapiens*, and *mt-Nd1*, *mt-Nd4l* and *mt-Nd6* for *Mus musculus*. Supplementing SAGE data with an alternative anchoring enzyme (e.g. *Sau3A I* or *MmeI*) or alternative technique (such as hybridization-based microarray technology) should address this issue, generating the most comprehensive and accurate expression profiles.

Tables 2 and 3 list human and murine SAGE libraries with extreme values of total mitochondrial gene expression, the highest levels belonging mostly to brain, muscle (cardiac and skeletal) and kidney tissue, as well as to placenta and several tumour subtypes. High requirements for energy production and/or protein synthesis are known for all of these. Justifiably, the large amount of energy produced by OXPHOS in mitochondria is directly associated with continuous contraction of a muscle fibre, or transmembrane ion pumping involved in the transmission of a nerve signal or the modification of primary urine. Within all these tissues, regional variations in total mitochondrial gene expression could be associated with different spectra of cell subtypes or activities. For example, most of the central nervous system regions (including those studied by SAGE, such as brain cortex, thalamus and *substantia nigra*) differ in the relative presence of neuronal cell subtype fractions, and their firing rate.

Expression of individual SAGE tags varied to an even greater extent: while some tags (including variant ones) were either not detected in all SAGE libraries studied or had low frequency (see Supplemental Table 2), others demonstrated generally high frequency levels, with extreme abundance in certain cell types. Maximum normalized values detected were 103 686 tpm for mouse *NlaIII* (GCTGCCCTCC, *mt-Co1*, GSM15116, old heart; Anisimov *et al*., 2005)

and 59 623 for *Sla3A I* (TGCTTCAATA, *mt-Co2*, GSM4874, medial striatum) datasets, 38 589 for rat (GATGCCCCCC, *mt-Co1*, GSM581, extraocular muscle; Cheng & Porter, 2002) and 32 868 for human (CCCATCGTCC, *mt-Co2*, GSM708, normal kidney tissue) *NlaIII* datasets. Notably, all these represented the most 3′ tags of the corresponding genes; this is further illustrated by the examples presented in Fig. 4 and Supplemental Fig. 3. While 4 bp SAGE anchoring enzyme recognition sites are abundant in mammalian mitochondrial genomes (see Figs 1, 2 and Supplemental Fig. 1), and many mitochondrial transcripts bear more than one SAGE tag, most SAGE tags originate from the most 3′ anchoring enzyme cleavage sites. Still, a number of SAGE tags match more 5′ anchoring enzyme cleavage sites in mitochondrial genes (Supplemental Table 2). It is therefore appropriate to evaluate the expression of individual mitochondrial-encoded genes in a cumulative way, i.e. as a sum of all corresponding transcripts. Furthermore, a few mitochondrial genes, such as *mt-Atp6/8* and *mt-Nd4/4l* (Supplemental Fig. 1, Supplemental Table 3) have 3′-to-5′-end overlaps and are transcribed together, with separate proteins being produced from these by frameshifting during the subsequent translation. At the same time, shifts in the distribution of frequency values for multiple tags related to the individual genes encoded by the mitochondrial genome might be indicative of defects in mitochondrial transcription potentially associated with ageing and disease (Kowald & Kirkwood, 2000).

In order to identify mitochondrial gene expression profiles with patterns characteristic for individual tissues or cell types, an extensive clustering analysis (by K-means) was performed. Figs 3 and 4 and Supplemental Fig. 3B, D, F, H and J clearly demonstrate that groups of individual samples fall into the limited number of clusters with analogous expression profiles of individual mitochondrial transcripts, and it was anticipated that samples generated from the tissues/cell types with similar characteristics would fall into the same cluster. However, the analysis of complex datasets (such as human and murine *NlaIII* datasets) failed to provide solid support for that hypothesis. Although a few groups of samples of the same sample origin (tissue or sample type, etc.) demonstrated a trend towards similar mitochondrial expression profiles, SAGE libraries generated from human retina was the only identified group falling completely into a single cluster (human *NlaIII* dataset, Cluster 2; Fig. 4). It could therefore be suggested that the origin of the tissue and sample could determine mitochondrial gene expression profile on rare occasions.

Numerous reports validate the assumption that mitochondrial gene expression (i.e. the abundance of mtRNA transcripts) is a function of cell type and state, as well as of the metabolic/respiratory activity of the proband (Annex & Williams, 1990; Kagawa *et al*., 1999; Unami *et al*., 2004). Importantly, the replication and transcription of the mitochondrial genome depends on nuclear gene products, and post-transcriptional mechanisms play an important role in regulating mitochondrial expression (Ostronoff *et al*., 1996; Escriva *et al*., 1999; Wu *et al*., 2002). While the mtDNA/mitochondrion ratio is essentially constant in all cell types in mammalian species, the number of mitochondria per cell and the amount of mtDNA per cell seem to be closely regulated within a given cell type but differ widely between cell types (Robin & Wong, 1988). Total mitochondrial gene expression seems to be a direct function of ageing, and generally declines with age, as has been demonstrated in a number of studies (Bodyak *et al*., 2002; Lu *et al*., 2004), while the activity of mitochondrial enzymes is suggested as a biochemical marker of ageing (Navarro *et al*., 2004). Importantly, ageing might affect the expression of individual mitochondrial genes to various extents (Tollet-Egnell *et al*., 2001). Quantitatively accurate data from both publicly available ageing-related SAGE studies (Welle *et al*., 2000; Anisimov *et al*., 2005) provide a degree of support for this hypothesis. It has been demonstrated that ageing in human skeletal muscle is associated with lower expression levels for individual mitochondrial transcripts, including those for *mt-Nd1*, *mt-Co2*, *mt-Atp8*, *mt-Atp6*, *mt-Nd3* and *mt-Nd4* (Welle *et al*., 2000); moreover, preliminary SAGE data for mouse cardiac muscle suggest that ageing in this tissue might be associated with lower expression levels for a few mitochondrial transcripts, such as *mt-Atp8* and *mt-Nd5* (Anisimov *et al*., 2005).

There are indications that other important physiological and pathological processes, such as cell differentiation (Angenieux *et al*., 2001), immortalization (Kim *et al*., 2001) or tumorigenesis (Ojala *et al*., 2002; Haugen *et al*., 2003; Baris *et al*., 2004), are also associated with alterations in total mitochondrial gene expression. However, the applied algorithm failed to associate these processes with characteristic alterations in expression of mitochondrial transcripts. Instead, when analysing SAGE libraries representing undifferentiated versus differentiated or normal versus cancer cells diverse changes in the expression of individual mitochondrial genes were detected. Though total mitochondrial gene expression values have demonstrated a trend to increase in association with tumorigenesis, no regular patterns in these or in the expression profiles of individual mitochondrial genes have been revealed. Supplemental Table 4 shows an example of such an analysis performed for normal versus cancer cells generated from mammary gland (average data from 8 vs 21 SAGE libraries, respectively), pancreas (7 vs 5) and prostate (6 vs 8).

Statistical analysis of averaged mitochondrial gene expression data demonstrates a notable diversity in the observed alterations caused by cellular transformation. Observed alterations in the expression of a few mitochondrial transcripts (*mt-Rnr1*, *mt-Rnr2*, *mt-Co2*, *mt-Co3*) are in opposite directions in different tissues, illustrating the complexity of this issue. Similar analysis performed for undifferentiated versus differentiated cells also failed to reveal regular patterns in expression profiles of individual mitochondrial genes (data not shown).

In this study a large-scale screening of the normalized mitochondrial gene expression profiles generated from publicly available mammalian SAGE datasets was performed, and variations in total mitochondrial gene expression, as well as in the expression of individual genes encoded by mitochondrial genomes, were analysed. Though the analysis demonstrates that total mitochondrial gene expression is a function of the cell type/cell state of the sample, the approach failed to reveal a direct link between the former and expression profiles of individual mitochondrial genes. At the same time, the trends in the diversity of mitochondrial gene expression profiles are remarkable, and it could be suggested that on some occasions the origin of tissue and sample could determine the mitochondrial gene expression profile.

Accumulated data illustrate the utility of SAGE technology that, unlike microarray technology, can generate quantitatively accurate and normalized gene expression information. The Gene Expression Omnibus database (GEO, NCBI) has become a convenient and broad public tool for comparative gene expression studies. A moderate volume of important biological information can be extracted from this database using straightforward *in silico* algorithms and subjected to further, more directional studies. Being validated with conservative molecular biology methods, such information can serve as an important tool in discovering delicate mechanisms of vital cellular functions.

This study was supported by a grant from The United States Army Medical Research Acquisition Activity (USAAMRAA) Award No. W81XWH-04-1-0366, National Institute of Health Grant Number 1 R21 NS043717-01A1 and by The Swedish Parkinson Foundation. The author is grateful to all GEO database contributors and to Professor Patrik Brundin, Assistant Professor Jia-Yi Li, Dr Emma Lane and Ana Sofia Correia for their critical review of this manuscript.

References

Angenieux, C., Fricker, D., Strub, J. M., Luche, S., Bausinger, H., Cazenave, J. P., Van Dorsselaer, A., Hanau, D., de la Salle, H. & Rabilloud, T. (2001). Gene induction during differentiation of human monocytes into dendritic cells: an integrated study at the RNA and protein levels. *Functional & Integrative Genomics* **1**, 323–329.

Anisimov, S. V. & Sharov, A. A. (2004). Incidence of 'quasi-ditags' in catalogs generated by Serial Analysis of Gene Expression (SAGE). *BMC Bioinformatics* **5**, 152.

Anisimov, S. V., Tarasov, K. V., Lakatta, E. G. & Boheler K. R. (2005). SAGE analysis of age- and sex-associated changes in cardiac gene expression. In *Serial Analysis of Gene Expression: Current Technologies and Applications* (ed. S. M. Wang), pp. 341–364. Norwich, UK: Horizon Scientific Press.

Anisimov, S. V., Tarasov, K. V., Riordon, D., Wobus, A. M. & Boheler K. R. (2002*a*). SAGE identification of differentiation responsive genes in P19 embryonic cells induced to form cardiomyocytes in vitro. *Mechanisms of Development* **117**, 25–74.

Anisimov, S. V., Tarasov, K. V., Stern, M. D., Lakatta, E. G. & Boheler, K. R. (2002*b*). A quantitative and validated SAGE transcriptome reference for adult mouse heart. *Genomics* **80**, 213–222.

Anisimov, S. V., Tarasov, K. V., Tweedie, D., Stern, M. D., Wobus, A. M. & Boheler, K. R. (2002*c*). SAGE identification of gene transcripts with abundances unique to pluripotent mouse R1 embryonic stem cells. *Genomics* **79**, 169–176.

Annex, B. H. & Williams, R. S. (1990). Mitochondrial DNA structure and expression in specialized subtypes of mammalian striated muscle. *Molecular and Cellular Biology* **10**, 5671–5678.

Arnason, U., Xu, X. & Gullberg, A. (1996). Comparison between the complete mitochondrial DNA sequences of *Homo* and the common Chimpanzee based on non-chimeric sequences. *Journal of Molecular Evolution* **42**, 145–152.

Baris, O., Savagner, F., Nasser, V., Loriod, B., Granjeaud, S., Guyetant, S., Franc, B., Rodien, P., Rohmer, V., Bertucci, F., Birnbaum, D., Malthiery, Y., Reynier, P. & Houlgatte, R. (2004). Transcriptional profiling reveals coordinated up-regulation of oxidative metabolism genes in thyroid oncocytic tumors. *Journal of Clinical Endocrinology and Metabolism* **89**, 994–1005.

Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W. & Clayton, D. A. (1981). Sequence and gene organization of mouse mitochondrial DNA. *Cell* **26**, 167–180.

Bodyak, N., Kang, P. M., Hiromura, M., Sulijoadikusumo, I., Horikoshi, N., Khrapko, K. & Usheva, A. (2002). Gene expression profiling of the aging mouse cardiac myocytes. *Nucleic Acids Research* **30**, 3788–3794.

Chabardes-Garonne, D., Mejean, A., Aude, J. C., Cheval, L., Di Stefano, A., Gaillard, M. C., Imbert-Teboul, M., Wittner, M., Balian, C., Anthouard, V., Robert, C., Segurens, B., Wincker, P., Weissenbach, J., Doucet, A. & Elalouf, J. M. (2003). A panoramic view of gene expression in the human kidney. *Proceedings of the National Academy of Sciences of the USA* **100**, 13710–13715.

Cheng, G. & Porter, J. D. (2002). Transcriptional profile of rat extraocular muscle by serial analysis of gene expression. *Investigative Ophthalmology And Visual Science* **43**, 1048–1058.

Edgar, R., Domrachev, M. & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* **30**, 207–210.

Enriquez, J. A., Fernandez-Silva, P. & Montoya, J. (1999). Autonomous regulation in mammalian mitochondrial DNA transcription. *Biological Chemistry* **380**, 737–747.

Escriva, H., Rodriguez-Pena, A. & Vallejo, C. G. (1999). Expression of mitochondrial genes and of the

transcription factors involved in the biogenesis of mitochondria Tfam, NRF-1 and NRF-2, in rat liver, testis and brain. *Biochimie* **81**, 965–971.

Fernandez-Silva, P., Enriquez, J. A. & Montoya, J. (2003). Replication and transcription of mammalian mitochondrial DNA. *Experimental Physiology* **88**, 41–56.

Gadaleta, G., Pepe, G., De Candia, G., Quagliariello, C., Sbissa, E. & Saccone, C. (1989). The complete nucleotide sequence of the *Rattus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *Journal of Molecular Evolution* **28**, 497–516.

Haugen, D. R., Fluge, O., Reigstad, L. J., Varhaug, J. E. & Lillehaug, J. R. (2003). Increased expression of genes encoding mitochondrial proteins in papillary thyroid carcinomas. *Thyroid: Official Journal of the American Thyroid Association* **13**, 613–620.

Hiendleder, S. & Wolf, E. (2003). The mitochondrial genome in embryo technologies. *Reproduction in Domestic Animals* **38**, 290–304.

Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713.

Kagawa, Y., Cha, S. H., Hasegawa, K., Hamamoto, T. & Endo, H. (1999). Regulation of energy metabolism in human cells in aging and diabetes: FoF(1), mtDNA, UCP, and ROS. *Biochemical and Biophysical Research Communications* **266**, 662–676.

Kenzelmann, M. & Mühlemann, K. (1999). Substantially enhanced cloning efficiency of SAGE (serial analysis of gene expression) by adding a heating step to the original protocol. *Nucleic Acids Research* **27**, 917–918.

Kim, H., You, S., Kim, I. J., Farris, J., Foster, L. K. & Foster, D. N. (2001). Increased mitochondrial-encoded gene transcription in immortal DF-1 cells. *Experimental Cell Research* **265**, 339–347.

Kowald, A. & Kirkwood, T. B. (2000). Accumulation of defective mitochondria through delayed degradation of damaged organelles and its possible role in the ageing of post-mitotic and dividing cells. *Journal of Theoretical Biology* **202**, 145–160.

Lin, C. S., Sun, Y. L., Liu, C. Y., Yang, P. C., Chang, L. C., Cheng, I. C., Mao, S. J. & Huang, M. C. (1999). Complete nucleotide sequence of pig (*Sus scrofa*) mitochondrial genome and dating evolutionary divergence within Artiodactyla. *Gene* **236**, 107–114.

Loveland, B., Wang, C. R., Yonekawa, H., Hermel, E. & Lindahl, K. F. (1990). Maternally transmitted histocompatibility antigen of mice: a hydrophobic peptide of a mitochondrially encoded protein. *Cell* **60**, 971–980.

Lu, T., Pan, Y., Kao, S. Y., Li, C., Kohane, I., Chan, J. & Yankner, B. A. (2004). Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891.

Madden, S. L., Galella, E. A., Zhu, J., Bertelsen, A. H. & Beaudry, G. A. (1997). SAGE transcript profiles for p53-dependent growth regulation. *Oncogene* **15**, 1079–1085.

Mizutani, J., Chiba, T., Tanaka, M., Higuchi, K. & Mori, M. (2001). Unique mutations in mitochondrial DNA of senescence-accelerated mouse (SAM) strains. *Journal of Heredity* **92**, 352–355.

Morse, M. C., Bleau, G., Dabhi, V. M., Hetu, F., Drobetsky, E. A., Lindahl, K. F. & Perreault, C. (1996). The COI mitochondrial gene encodes a minor histocompatibility antigen presented by H2-M3. *Journal of Immunology* **156**, 3301–3307.

Navarro, A. (2004). Mitochondrial enzyme activities as biochemical markers of aging. *Molecular Aspects of Medicine* **25**, 37–48.

Nilsson, M. A., Gullberg, A., Spotorno, A. E., Arnason, U. & Janke, A. (2003). Radiation of extant marsupials after the K/T boundary: evidence from complete mitochondrial genomes. *Journal of Molecular Evolution* **57** (Suppl. 1), S3–12.

Ojala, P., Sundstrom, J., Gronroos, J. M., Virtanen, E., Talvinen, K. & Nevalainen, T. J. (2002). mRNA differential display of gene expression in colonic carcinoma. *Electrophoresis* **23**, 1667–1676.

Ostronoff, L. K., Izquierdo, J. M., Enriquez, J. A., Montoya, J. & Cuezva, J. M. (1996). Transient activation of mitochondrial translation regulates the expression of the mitochondrial genome during mammalian mitochondrial differentiation. *The Biochemical Journal* **316**, 183–191.

Potapova, O. U., Anisimov, S. V., Gorospe, M., Dougherty, R. H., Gaarde, W. A., Boheler, K. R. & Holbrook, N. J. (2002). Identification of targets of JNK2 signaling involved in regulation of human tumor cell growth. *Cancer Research* **62**, 3257–3263.

Rat Genome Sequencing Project Consortium (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521.

Robin, E. D. & Wong, R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology* **136**, 507–513.

Sharon, D., Blackshaw, S., Cepko, C. L. & Dryja, T. P. (2002). Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE). *Proceedings of the National Academy of Sciences of the USA* **99**, 315–320.

Silver, L. M. (1995). Origin of the classical inbred strains. In *Mouse Genetics: Concepts and Applications*, pp. 26–27. New York: Oxford University Press.

Sulijoadikusumo, I., Horikoshi, N. & Usheva, A. (2001). Another function for the mitochondrial ribosomal RNA: protein folding. *Biochemistry* **40**, 11559–11564.

Tatusova, T. A. & Maden, T. L. (1999). Blast 2 sequences: a new tool for comparing protein and nucleotide sequences. *FEMS Microbiology Letters* **174**, 247–250.

Tollet-Egnell, P., Flores-Morales, A., Stahlberg, N., Malek, R. L., Lee, N. & Norstedt, G. (2001). Gene expression profile of the aging process in rat liver: normalizing effects of growth hormone replacement. *Molecular Endocrinology* **15**, 308–318.

Unami, A., Shinohara, Y., Kajimoto, K. & Baba, Y. (2004). Comparison of gene expression profiles between white and brown adipose tissues of rat by microarray analysis. *Biochemical Pharmacology* **67**, 555–564.

Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* **270**, 484–487.

Virlon, B., Cheval, L., Buhler, J. M., Billon, E., Doucet, A. & Elalouf, J. M. (1999). Serial microanalysis of renal transcriptomes. *Proceedings of the National Academy of Sciences of the USA* **96**, 15286–15291.

Welle, S., Bhatt, K. & Thornton, C. A. (1999). Inventory of high-abundance mRNAs in skeletal muscle of normal men. *Genome Research* **9**, 506–513.

Welle, S., Bhatt, K. & Thornton, C. A. (2000). High-abundance mRNAs in human muscle: comparison between young and old. *Journal of Applied Physiology* **89**, 297–304.

Wu, H., Kanatous, S. B., Thurmond, F. A., Gallardo, T., Isotani, E., Bassel-Duby, R. & Williams, R. S. (2002). Regulation of mitochondrial biogenesis in skeletal muscle by CaMK. *Science* **296**, 349–352.