Identification of abnormal nuclear and mitochondrial genes in esophageal cancer cells

ZONGWEN LIU[1*], TING CHAI[2*], YAN ZHANG[3], ALAN CHU[1], BING LIANG[2], XING GUO[1], ZHENJIANG GUO[1], RUI SONG[1], GE HOU[1], JINJIN YUAN[1], YAOHE LIU[1] and YUHAO ZHANG[4]

Departments of [1]Radiotherapy and [2]Oncology, The Second Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450014; [3]Department of Pharmacology, School of Basic Medical of Zhengzhou University, Zhengzhou, Henan 450001; [4]Department of Obstetrics and Gynecology, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, P.R. China

Abstract. The present study aimed to detect the mutation characteristics of mitochondrial DNA (mtDNA) in Eca109 of Ec9706 cells, and to investigate their association with the nuclear genome (nDNA), thus providing a basis for gene targeting therapies for esophageal squamous cell carcinoma (ESCC). *In vitro*-cultured Ec9706 and Eca109 cells were analyzed the changes of single-nucleotide polymorphisms (SNPs), insertions/deletions (INDELs), copy number variation, and structure variation (SV) of their genome by high-throughput sequencing. The loci with SV on chromosome 1-12 of the two ESCC cell lines were ≥5% of the mtDNA, but SV on chromosome 13-22, X and Y was ≤3%; >40% of loci exhibited gain or loss; intergenic loci with INDEL changes and SNP features accounted for the majority of mutations. The affected genes encoded proteins including nDNA-encoding intra-mitochondrial-transporting proteins, ATP energy generation-associated proteins and mitochondrial electron respiratory chain proteins, and these proteins were all nucleus-encoded mitochondrial proteins. The transcription, duplication, and translation of the abnormally expressed mtDNA in Ec9706 and Eca109 cells were closely associated with disorders of nuclear DNA products.

Introduction

Eesophageal squamous cell carcinoma (ESCC) is one of the most common malignant diseases world wide,

Correspondence to: Professor Yan Zhang, Department of Pharmacology, School of Basic Medical of Zhengzhou University, 100 Kexue Avenue, Zhengzhou, Henan 450001, P.R. China
E-mail: cnyzdoc@163.com

*Contributed equally

particularly in China, where it is the fourth most common cause of cancer-associated mortality (1). Unlike cancers that have been extensively studied, such as breast and colon cancers, the outcome of ESCC remains unchanged during the last several decades, with a 5-year survival rate ranging from 15-25% (2). However, in general, understanding of the genomic abnormalities in this disease is limited to reports from small cohorts (3-5). Advances in next-generation sequencing (NGS) technology are facilitating the identification of novel disease-associated genes and promising to transform the routine clinical diagnosis of inherited disease. Since the initial reports in 2009 (6,7), NGS has aided the discovery of >50 novel disease genes in research settings. Thus, a there is a compelling requirement to extensively identify genomic abnormalities underlying ESCC by NGS, including single-nucleotide polymorphisms (SNPs), insertions/deletions (INDELs), structure variations (SVs) and other information variations (8,9), to elucidate its molecular basis, and to aid the development of effective targeted therapies.

In-depth studies of mitochondria identified mitochondrial DNA (mtDNA) at the basal membrane of mitochondria, which is a type of special and unique genetic material, as it is located outside the nuclei (10). Though mtDNA has independent genetic function, the majority of proteins (including the outer mitochondrial membrane and the matrix proteins) are still encoded by nuclear genome (nDNA), among which ~1,500 proteins have important roles in maintaining mitochondrial functions (11). The present study performed high-throughput sequencing on two ESCC cell lines to determine the common features of genomic variation in ESCC cells, identify ESCC-associated abnormally expressed nDNA and mtDNA, and determine their interactions that have important roles in the occurrence and development of ESCC, thus providing direction for ESCC basic research and an important experimental basis for the clinical treatment of ESCC gene targeting therapies.

Materials and methods

Cell recovery and culture. The Ec9706 cell line was provided by the National Key Laboratory of Molecular Oncology, Chinese Academy of Medical Sciences (Beijing, China) and

the Eca109 cell line was provided by the laboratory, School of Pharmacy, Zhengzhou University (Zhengzhou, China). The Ec9706 and Eca109 cells cryopreserved in liquid-nitrogen were quickly removed and placed in a water bath at 37˚C for 1-2 min thawing. The cell suspension was then diluted with an appropriate amount of RPMI-1640 medium containing 10% fetal bovine serum (both from Beijing Solarbio Science and Technology Co., Ltd., Beijing, China). Following centrifugation at 173 x g at room temperature for 5 min, the supernatant was removed, and medium was added; this process was repeated twice. The cells were diluted in an appropriate amount of culture medium containing 10% fetal bovine serum, the cells were seeded into culture flasks for the culture at 37˚C and 5% CO_2. The culture medium was replaced the next day, and subsequently the cells were routinely cultured.

DNA extraction. Genomic DNAs was extracted according to the instructions of the DNA extraction kit (Shanghai Genmed Pharmaceutical Technology Co., Ltd., Shanghai, China).

The concentration and purity of the genomic DNA extracted from each sample were determined three times using ultra-trace spectrophotometry (Multiskan MK3; Thermo Fisher Scientific, Inc., Waltham, MA, USA) with pure water as the control, and the average of the three measurements was used for the calculation. For pure genomic DNA, the ratio of absorbance at 260 nm/absorbance at 280 nm should be close to 1.8 (>1.9, indicating RNA contamination; <1.6, indicating protein or phenol pollution).

High-throughput sequencing. The high-throughput sequencing process was performed by Genewiz, Inc. (Beijing, China). Workflow of the experiment is presented in Fig. 1. Genomic DNA was tested and randomly broken into manageable fragments by ultrasound to facilitate the construction of an insert library. In human genome re-sequencing, paired-end libraries with a span size of 400-500 bp were usually adopted. The template DNA fragments of the constructed libraries were hybridized to the surface of flow cells and amplified to form clusters, and then sequenced with the Illumina HiSeq X sequencing system (Illumina, Inc., San Diego, CA, USA). Currently, read length of 150 bp paired-end sequencing strategy is used in high-throughput whole genome sequencing.

Detected SNPs and INDELs were annotated against a collection of comprehensive functional annotation databases, including 1000 Genomes Project, Cosmic, GWAS, PolyPhen-2, VISIFT, NHLBI, ClinVar, NCI60, YH genome, gene/protein structure, germline variations (dbSNP; https://www.ncbi.nlm.nih.gov/projects/SNP/), and functional elements (transcription binding sites, microRNA targets, conserved elements). Structural variations were detected and annotated by the chromosomal locations. Copy number variations (CNV) were also calculated based on the read depth and summarized in detail. WGC024118D and WGC024119D were the sequencing IDs of Ec9706 and Eca109 cells, respectively.

Data analysis. The sample passed sample quality control and yielded enough high quality sequencing data with an average 117.222G raw data after Illumina pass filtering (PF). The
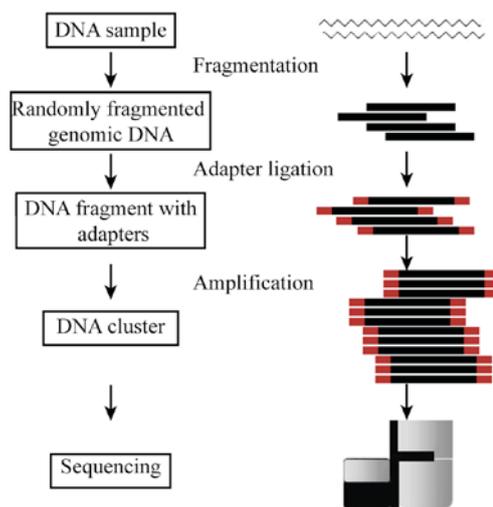

Figure 1. Workflow of the high-throughput sequencing.

passing filter was a default standard processing of Illumina HiSeq sequencers, to remove any reads that do not meet the overall quality as measured by the Illumina chastity filter. The chastity of a base call was calculated as the ratio of the brightest intensity divided by the sum of the brightest and second brightest intensities. Clusters passed filter if no more than one base call in the first 25 cycles had a chastity of <0.6.

A comprehensive annotation package, including 10 different databases, including 1000 Genomes Project (http://www.internationalgenome.org), dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/), Cosmic (http://cancer.sanger.ac.uk/cosmic), GWAS (https://www.ebi.ac.uk/gwas/), Poly-Phen-2 (http://genetics.bwh.harvard.edu/pph2/), VISIFT (http://sift.jcvi.org/), NCI60 (http://genome-www.stanford.edu/nci60/index.shtml), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), YHgenome (http://www.yhdatabase.com/), and National Heart, Lung, and Blood Institute (https://www.nhlbi.nih.gov), was applied to annotate the bioinformatics analysis of identified SNPs, INDELs, CNVs and SVs from the sequencing data.

Results

Summary of bioinformatics analysis. The PF sequencing reads were mapped to the human genome version 19 (hg19) using Burrows-Wheeler Aligner (12). On average, 98.99% of whole genome regions were sequenced and a 39.35-fold mappable-reads-coverage was achieved in this sample set (Table I).

SV detection and annotation. The categories of SV include insertions, duplications, deletions, inversions, recurring mobile elements and other rearrangements, typically defined as those covering 50 or more base pairs (Fig. 2). A total of 187 structural variants were identified, including seven inter-chromosomal translocations, three intra-chromosomal translocations, zero inversion, 177 large insertions or deletions. Although many SVs were detected, the majority of SVs may be common SVs present in normal samples, or false detections because of the limitations of NGS (Fig. 3).

Table I. Data coverage analysis.

Quality control statistics sample	WGC024118D	WGC024119D
Paired-end read length	150*2	150*2
Total effective data yield (Gb)	116.434	118.01
Total reads number (M)	776.23	786.73
Reads mapping rate	99.04%	98.70%
Properly paired mapping reads rate	96.97%	96.81%
No-mismatch mapping reads rate	61.04%	60.11%
Mismatch alignment bases rate	0.77%	0.74%
Mean coverage sequencing depth	39.1	39.6
Reference genome coverage	98.99%	98.99%
Reference genome coverage ≥4X	98.53%	98.53%
Reference genome coverage ≥10X	97.21%	97.28%
Reference genome coverage ≥20X	90.00%	90.49%
Polymerase chain reaction duplication rate	8.62%	8.64%

Table II. Summary of INDEL detection results.

Feature	WGC024118D	WGC024119D
High-confidence INDEL no.	740783	766563
Deletion	392984	408253
Insertion	347799	358310
Heterozygotes	386393	409200
Homozygotes	327829	330279
dbSNP	455370 (61.5%)	465321 (60.7%)
1000 Genomes Project	237781 (32.1%)	239886 (31.3%)
NA	17882	18327
3'UTR	6195	6354
5'UTR	810	865
3' UTR5, UTR3	6	3
Downstream	5402	5600
Exonic	615	645
Exonic, splicing	3	2
Intergenic	403796	419365
Intronic	270105	277821
ncRNA 3'UTR	127	128
ncRNA 5'UTR	23	26
ncRNA 5'UTR, ncRNA 3'UTR	1	1
ncRNA exonic	1241	1299
ncRNA intronic	29459	30801
ncRNA splicing	13	14
Splicing	91	95
Upstream	4851	5049
Upstream, downstream	163	168
Exon frameshift deletion	99	108
Exon frameshift insertion	87	89
Exon nonframeshift deletion	189	194
Exon nonframeshift insertion	148	158
Exon stopgain SNV	7	8
Exon stoploss SNV	1	2
Exon unknown	87	88

INDEL, insertion/deletion; UTR, untranslated region; SNV, single-nucleotide variant; NA, no annotation.

CNV and annotation. As presented in Fig. 4, >40% genes exhibited gain or loss CNVs, and the main form of CNV in WGC024118Ds was loss, and in WGC024119D it was gain. Almost half genes on chromosome 9 (Chr 9) and 11 in WGC024118D and WGC024119Ds had gain or loss, and nearly half genes on Chr 10 in these two samples had gain or loss.

Functional annotations of short INDEL. Loci with INDEL in these two ESCC cell lines predominantly occurred in the upstream, downstream, exon and intron as well as among genes, among which inter-gene INDEL accounted for the majority (Table II).

As demonstrated in Table III, the bases with short insert or deletion ranged between 1-20 nt, and the mitochondrial proteins mutated include mitochondrial electron respiratory chain-related proteins, such as **NADH:Ubiquinone oxidore**-ductase subunit S5 (NDUFS5), cytochrome P450 family 27 subfamily A member 1 (CYP27A1), fatty acid metabolism-associated proteins, such as 3-hydroxyisobutyryl-CoA hydrolase (HIBCH), acyl-CoA dehydrogenase family member 9 (ACAD9), ATP energy generation-associated proteins, such as ATP synthase mitochondrial F1 complex assembly factor 1 (ATPAF1), and these proteins are all nDNA-encoding intra-mitochondrial-transporting proteins.

Functional annotations of SNP. As presented in Table IV, the loci with SNP features in these two ESCC cell lines are located in the upstream, downstream, exon and intron, and within genes. The majority of loci with SNP features were located in genes and in exons.

Additionally, the mitochondrial protein genes with SNP features were mostly located in exons (Table IV), including nDNA-encoding intra-mitochondrial-transporting proteins, such as mitochondrial *trans*-2-enoyl-CoA reductase (MECR), fatty acid metabolism-associated proteins, such as 3-hydroxymethyl-3-methylglutaryl-CoA lyase (HMGCL), HIBCH, and MECR, ATP energy generation-associated proteins, such as ATPAF1, and succinate dehydrogenase complex. All these proteins are nDNA-encoded intra-mitochondrial-transporting proteins.
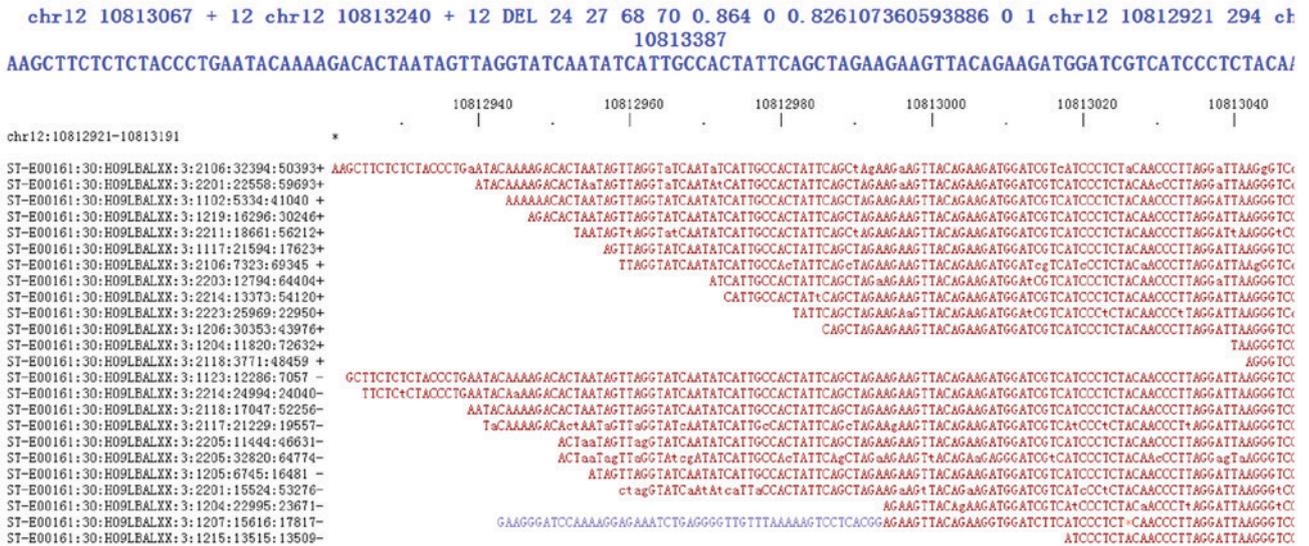
Figure 2. Detection results of partial single nucleotide polymorphisms. The '+' or '-' before the sequence indicates the sequence alignment orientation for each read; mismatches or inserts/deletions to the hg19 reference genome are brown; low-quality bases (<20 phred score) are in lower case and in gray if they match the reference genome.
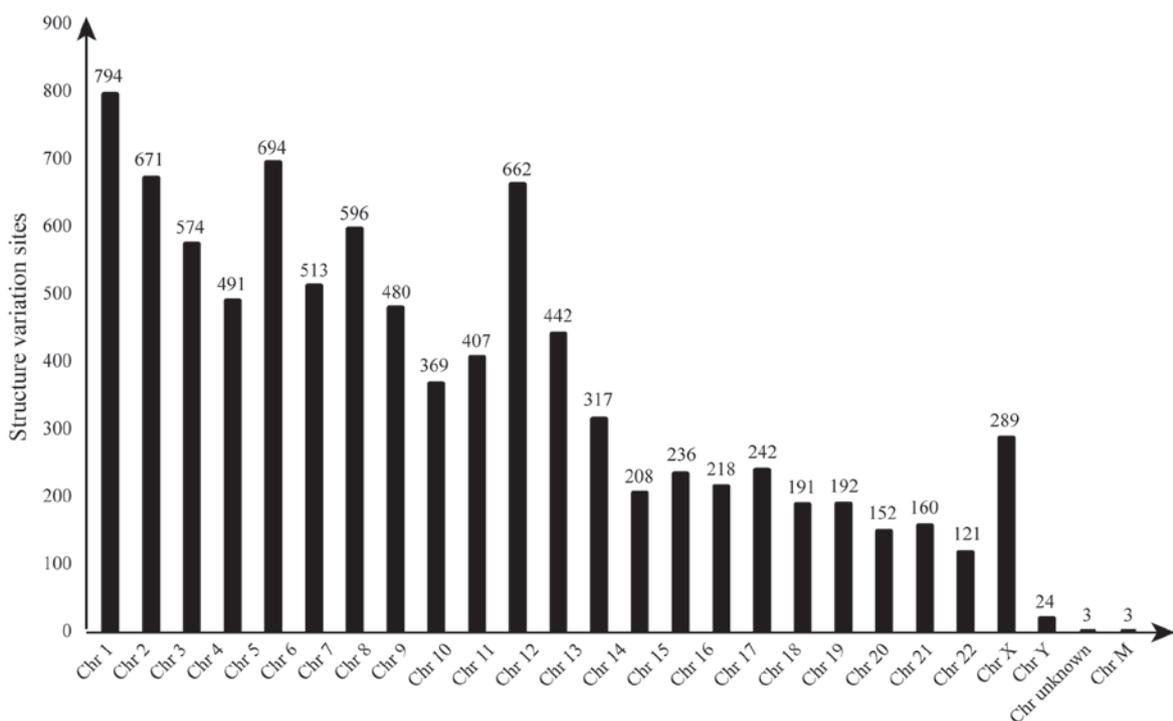

Figure 3. Structure variation loci within chromosomes. In nest-generation sequencing, the loci with structure variation on Chr 1-12 were ≥5%, and structure variation on Chr 13-22, X and Y were ≤3%. Chr, chromosome.

Discussion

Oncogenes and tumor suppressor genes have important roles in the occurrence and development of tumors, Changes to mtDNA in tumor tissue has received increasing research interest. It has been previously indicated that various solid tumors and hematological malignancies have different mtDNA mutations, deletions, and microsatellite instabilities at different sites, and these may be associated with the occurrence of a variety of tumors (13). In recent years, more researchers hypothesize that the interactions between mtDNA and nDNA are involved in the tumorigenic process (14). Compared with nDNA, mtDNA has higher mutation rate due to the poor proofreading ability of mtDNA-replicating DNA polymerase and the environment with high concentrations of reactive oxygen species (ROS). Furthermore, mtDNA has hologynic features, in which mutations accumulate continuously along the matriarchal line, and the accumulation of mutations increases the risk of tumorigenesis (9). In addition, mtDNA has more polymorphisms (15), appearing
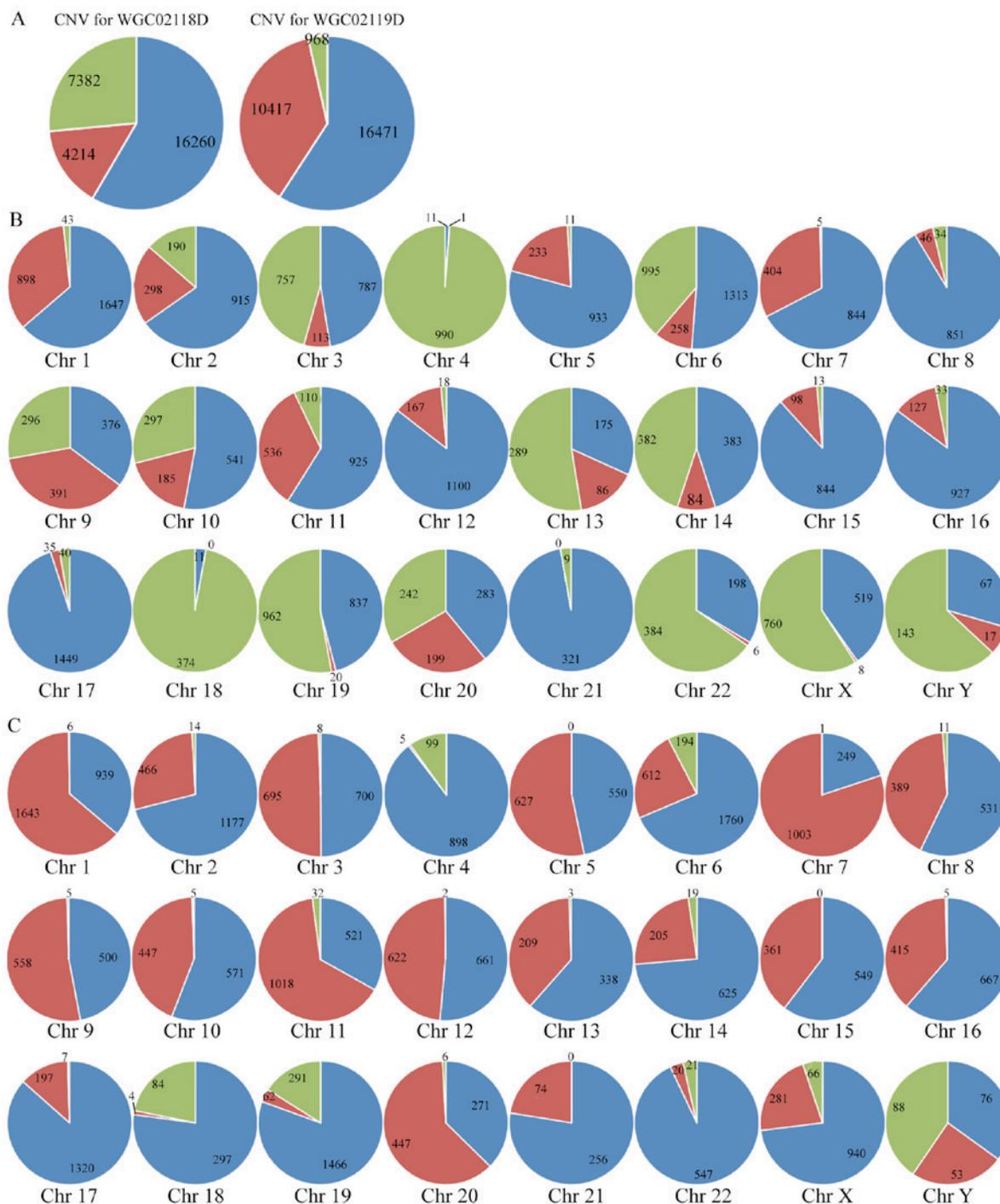
Figure 4. CNVs in the Ec9706 (WGC024118D) and Eca109 (WGC024119D) genome. (A) CNVs in the whole genome; the CNV change trend in the whole genome is that loci with gain or loss account for >40% of the entire sequence sites, the main form of CNV in WGC024118D is loss, and for WGC024119D is gain. Red indicates loss; green indicates gain; and blue indicates sequence matching the reference hg19. (B) CNVs in WGC024118D and (C) WGC024119D. Almost all of the loci on Chr 4, 17, and 18 in WGC024118D have gain or loss; more than half loci on Chr 3, 9, 11, 13, 14, 19 and 20 in WGC024118D, and Chr 1, 5, 7, 9, 11, 20 and Y in WGC024119D have gain or loss; almost half of loci on Chr 6, 10, 22, X and Y in WGC024118D and Chr 3, 8, 10, 12, 13, 15, and 16 in WGC024119D have gain or loss; few loci on Chr 8, 12, 15, 16, and 17 in WGC024118D and Chr 4 and 22 in WGC024119D have gain or loss, maintaining nearly normal sequence. Red indicates gain; green indicates loss; and blue indicates normal sequence. Chr, chromosome.

as significant sequence differences among different races and in different regions. The function of mtDNA is known to be involved in tumorigenesis in head and neck cancer (16), bladder cancer (10), breast cancer (17) and lung cancer (11). Hu *et al* (18) also identified gastric cancer and esophageal cancer-associated genes using whole-genome sequencing, including p53, Janus kinase 3, BRCA2, fibroblast growth

factor 2, F-box and WD repeat domain containing 7, mutS homolog 3, patched 1, neurofibromin 1, ErbB-2 receptor tyrosine kinase 2 and checkpoint kinase 2, and identified number of novel potential cancer-associated genes, including KISS1 receptor, anti-Mullerian hormone, motor neuron and pancreas homeobox 1, WNK lysine deficient protein kinase 2, THAP domain containing 12; furthermore, certain

Table III. Genetic screening of nuclear DNA with INDEL.

Chr	Start	End	Reference sequence	Sequence alteration	QUALITY	Alteration ratio (%)	Variation type	Gene location	Gene	Name (description)
Chr 1	39500839	39500839	C	-	841.73	100	Deletion	Downstream	NDUFS5	NADH:Ubiquinone oxidoreductase subunit S5, (nuclear gene encoding mitochondrial protein, transcript variant 2)
Chr 1	45804431	45804431	-	T	819.73	100	Insertion	Intronic	MUTYH	mutY DNA glycosylase, (nuclear gene encoding mitochondrial protein, transcript variant α3)
Chr 2	191091321	191091334	CAAAAAAAAAAAAA	-	1600.73	100	Deletion	Intronic	HIBCH	3-hydroxyisobutyryl-CoA hydrolase (nuclear gene encoding mitochondrial protein, transcript variant 2)
Chr 2	219652424	219652424	-	CCTCTTACCTG	3035.73	100	Insertion	Intronic	CYP27A1	Cytochrome P450 family 27 subfamily A member 1 (nuclear gene encoding mitochondrial protein)
Chr 3	179339343	179339343	-	GGTCTCGG	1811.73	100	Insertion	Intronic	NDUFB5	NADH:Ubiquinone oxidoreductase subunit B5 (nuclear gene encoding mitochondrial protein, transcript variant 1)
Chr 3	128614563	128614563	-	CTC	2388.73	100	Insertion	Intronic	ACAD9	acyl-CoA dehydrogenase family member 9 (nuclear gene encoding mitochondrial protein, transcript variant 1)
Chr 4	106312189	106312189	-	C	806.73	100	Insertion	Intronic	PPA2	Pyrophosphatase (inorganic) 2 (nuclear gene encoding mitochondrial protein, transcript variant 1)
Chr 4	89197868	89197875	GACTGTCC	-	1078.74	100	Deletion	Intronic	PPM1K	Protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K (nuclear gene encoding mitochondrial protein)
Chr 1	29538224	29538225	CT	-	1835.73	100	Deletion	Intronic	MECR	Mitochondrial *trans*-2-enoyl-CoA reductase (nuclear gene encoding mitochondrial protein)
Chr 1	47107042	47107042	-	G	1521.73	100	Insertion	Intronic	ATPAF1	ATP synthase mitochondrial F1 complex assembly factor 1 (nuclear gene encoding mitochondrial protein)

Table IV. Summary of SNP detection results.

Feature	WGC024118D	WGC024119D
High-confidence SNP no.	3851527	3861523
Heterozygotes	1938738	1944372
Homozygotes	1911032	1915229
dbSNP	3685859 (95.7%)	3694088 (95.7%)
1000 Genomes Project	3520593 (91.4%)	3520791 (91.2%)
3'UTR	25886	25946
5'UTR	5578	5570
5'UTR, 3'UTR	10	12
Downstream	23646	23744
Exonic	23196	23198
Exonic, splicing	6	7
Intergenic	2255882	2264134
Intronic	1327100	1327665
ncRNA 3'UTR	615	618
ncRNA 5'UTR	97	105
ncRNA exonic	9941	10103
ncRNA intronic	156029	156787
ncRNA splicing	52	55
Splicing	71	68
Upstream	22691	22772
Upstream, downstream	727	739
Exon nonsynonymous SNV	10720	10774
Exon stopgain SNV	98	92
Exon stoploss SNV	15	13
Exon synonymous SNV	11962	11915
Exon unknown	407	411

SNP, single nucleotide polymorphism; UTR, untranslated region; SNV, single-nucleotide variant.

chromosomes were reported to have mutations in >30% of tumor genes, including MACRO domain containing 2, fragile histidine triad (FHIT) and parkin RBR E3 ubiquitin protein ligase (18). However, the application of whole genome sequencing to detect mitochondrial changes in ESCC cell lines has rarely been reported previously. Thus, identifying nDNA and mtDNA mutation involved the occurrence and development of ESCC, and investigating their impact on the occurrence and development of ESCC is important for developing novel ESCC gene targeting therapies.

Detecting SVs can help to investigate the mechanisms of gene mutation in cancer, understand the biological differences and identify novel therapeutic targets. However, the complexities and mutation mechanisms of SVs throughout the whole ESCCs genome are still unclear. Cheng et al (19) reported chromosomal abnormalities in malignant metastatic ESCCs and the function of SVs-derived target genes in these abnormal chromosomes are diverse, indicating that the SVs map of the whole genome important for the prevention, diagnosis and treatment of ESCCs. The present study detected

many SVs; however, most of the SVs may be common SVs present in normal samples, or the false detections because of the limitation of NGS. The categories of structural variants include insertions, duplications, deletions, inversions, recurring mobile elements, and other rearrangements. A total of 187 SVs were identified in the current study, including seven inter-chromosomal translocations, three intra-chromosomal translocations and 177 large insertions or deletions. The loci with SVs on Chr 1-12 of the two ESCC cell lines were ≥5%, whereas those with SVs on Chr 13-22, X and Y were ≤3%. The SVs of mitochondrial genes detected in this study may provide important information for further studies into the nDNA and mtDNA mutations, and roles in the occurrence and development of ESCC.

DNA copy number alterations in tumor cells are key genetic events in the development and progression of human cancers (20). These alterations are typically the result of genomic events producing gains and losses of chromosomes or chromosomal regions. Losses and gains of DNA can contribute to alterations in the expression of tumor suppressor genes and oncogenes. Therefore, identifying DNA copy number alterations in tumor genomes may help to identify critical genes associated with cancer and, eventually, to improve therapeutic approaches. Errors during mitosis and meiosis can result in duplications or deletions of genes on a chromosomal level. These differences are termed CNV, which may profoundly affect health and lead to various disorders. In this study, CNVs were identified based on read depth. Miyawaki et al (21) investigated the important roles of MYC and FHIT gene CNV in selecting the optimal treatment strategy in strategy in resected ESCC patients. The detection results of CNV the current study demonstrated that >40% of genes exhibited gain or loss. Almost half of loci on Chr 9 and 11 in WGC024118D and WGC024119D exhibited gain or loss, and nearly half loci of on Chr 10 exhibited gain or loss, indicating that detecting CNV in ESCC is important for identifying the oncogenic gene gains of the genome and in genes involved in the development of tumors.

The results of SNPs and INDELs sequencing demonstrate the potential functional impact. Cao et al (22) used the whole exon sequencing and revealed lots of genetic heterogeneities within ESCC. Xu et al (23) previously investigated the polymorphisms of ESCC phosphatase and tensin homolog gene in the Chinese Han population, and the interactions among genes. Similarly, detecting short insertion and deletion in genes was applied to detect the gene mutations of ESCC (24,25). The detection results of SNPs and INDELs in the current study revealed potential adverse alterations that have potentially important functions and are associated with gene regulation, including transcription factor binding, microRNA target, conserved elements. The inter-gene loci occurred INDEL and had SNP features account for the majority, and these genes-encoded proteins include mitochondrial electron respiratory chain-associated proteins, such as CYP27A1, fatty acid metabolism-associated proteins, such as HIBCH and ACAD9, nDNA-encoding intra-mitochondrial-transporting proteins, such as MECR, and ATP energy generation-related proteins, such as ATPAF1. The sequencing results also demonstrate that the detection of SNPs and INDELs can be used as the basis for the development of gene therapy.

Table V. Screening of nuclear DNA genes with SNP features.

Chr	Start	End	Reference sequence	Sequence alteration	QUALITY	Alteration ratio (%)	Variation type	Gene location	Gene	Name (description)
Chr1	28564279	28564279	A	G	1000.77	100	SNP	Exonic	ATPIF1	ATPase inhibitory factor 1 (nuclear gene encoding mitochondrial protein)
Chr1	29542637	29542637	A	G	1442.77	100	SNP	Exonic	MECR	Mitochondrial *trans*-2-enoyl-CoA reductase (nuclear gene encoding mitochondrial protein)
Chr1	24134721	24134721	T	C	794.77	100	SNP	Exonic	HMGCL	3-Hydroxymethyl-3-methylglutaryl-CoA lyase (nuclear gene encoding mitochondrial protein)
Chr1	36927733	36927733	G	A	1452.77	100	SNP	Exonic	MRPS15	Mitochondrial ribosomal protein S15 (nuclear gene encoding mitochondrial protein)
Chr1	47133811	47133811	T	C	457.77	100	SNP	Exonic	ATPAF1	ATP synthase mitochondrial F1 complex assembly factor 1 (nuclear gene encoding mitochondrial protein)
Chr1	24134721	24134721	T	C	1025.77	100	SNP	Exonic	HMGCL	3-Hydroxymethyl-3-methylglutaryl-CoA lyase (nuclear gene encoding mitochondrial protein)
Chr1	29529699	29529699	T	C	982.77	100	SNP	Exonic	MECR	Mitochondrial *trans*-2-enoyl-CoA reductase (nuclear gene encoding mitochondrial protein)
Chr1	17380497	17380497	G	T	896.77	100	SNP	Exonic	SDHB	Homo sapiens succinate dehydrogenase complex iron sulfur subunit B (nuclear gene encoding mitochondrial protein)
Chr1	19201956	19201956	A	G	847.77	100	SNP	Exonic	ALDH4A1	Aldehyde dehydrogenase 4 family member A1 (nuclear gene encoding mitochondrial protein)

Under certain conditions, mtDNA can also have effects on nDNA due to mitochondrial defects, thus resulting in corresponding gene expression changes; this process is termed reverse regulation (26). This reverse regulation is associated with disturbances in nDNA expressions, mtDNA mutations, respiratory function and mitochondrial protein synthesis inhibition. This reverse regulation is also involved in the process of apoptosis, during which mitochondrial damages can cause the increasing of mitochondrial membrane permeability; such apoptosis-inducing factors as cytochrome c, apoptotic protease activating factor-1 and apoptosis inducing factor are then released from mitochondria, and can be directly transported to the nuclei, thus inducing the expressions of certain genes in the nuclei and triggering the apoptotic cascade. It is also suggested that the reverse regulation may also be associated with the ROS pathway or the ratio of ATP/ADP (26). In the present study, the identified nDNA-encoded mtDNA with SVs, CNV, SNPs and INDELs provides a good experimental basis and ideas for investigating the relationships between nDNA and mtDNA in the occurrence and development of ESCC.

Tumorigenesis not only depends on intranuclear genetic materials, but also is also closely associated with ectonuclear mtDNA. As the biological metabolism and energy conversion center, mitochondrial genome replication, gene expression, respiratory chain function, and mitochondrion-associated cell function are inextricably associated with nDNA. nDNA and mtDNA changes will cause corresponding changes of mitochondrial functions. The transcription, replication and translation of mtDNA need a variety of different nDNA products, and the disorders of nDNA products are also associated with mtDNA mutations and biosynthesis inhibition of mitochondrial proteins. These changes can lead to various diseases, and play vital roles particularly in the occurrence and development of tumors. Thus, nDNA-encoded mtDNA protein variants in Ec9706 and Eca109 detected in this study using high-throughput sequencing can provide very reliable theoretical basis and data support for the gene targeting therapies of ESCC.

In this study, the identified mtDNA with abnormal transcription, duplication and translation in Ec9706 and Eca109 cells by high-throughput sequencing were closely associated with the disorders of nDNA products. This study into these nDNA-associated mtDNA and the interaction between the two is important for basic research for ESCC, pointed out the direction and provided a very reliable theoretical basis for the gene targeting therapies of ESCC in clinical.

Acknowledgements

References

1. Zhao P, Dai M, Chen W and Li N: Cancer trends in China. Jpn J Clin Oncol 40: 281-285, 2010.
2. Ell C and Lorenz D: Diagnosis and treatment of oesophageal carcinoma: Changes in every respect. Viszeralmedizin 31: 314, 2015.
3. Shigaki H, Baba Y, Watanabe M, Murata A, Ishimoto T, Iwatsuki M, Iwagami S, Nosho K and Baba H: PIK3CA mutation is associated with a favorable prognosis among patients with curatively resected esophageal squamous cell carcinoma. Clin Cancer Res 19: 2451-2459, 2013.
4. Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, Lawrence MS, Kiezun A, Fernandes SM, Bahl S, et al: Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nat Commun 6: 8866, 2015.
5. Knoppers BM, Zawati MH and Sénécal K: Return of genetic testing results in the era of whole-genome sequencing. Nat Rev Genet 16: 553-559, 2015.
6. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, et al: Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci USA 106: 19096-19101, 2009.
7. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al: Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461: 272-276, 2009.
8. Frank K, Barta E, Bana NÁ, Nagy J, Horn P, Orosz L and Stéger V: Complete mitochondrial genome sequence of a Hungarian red deer (Cervus elaphus hippelaphus) from high-throughput sequencing data and its phylogenetic position within the family Cervidae. Acta Biol Hung 67: 133-147, 2016.
9. Brunelle JK, Shroff EH, Perlman H, Strasser A, Moraes CT, Flavell RA, Danial NN, Keith B, Thompson CB and Chandel NS: Loss of Mcl-1 protein and inhibition of electron transport chain together induce anoxic cell death. Mol Cell Biol 27: 1222-1235, 2007.
10. Yang Ai SS, Hsu K, Herbert C, Cheng Z, Hunt J, Lewis CR and Thomas PS: Mitochondrial DNA mutations in exhaled breath condensate of patients with lung cancer. Respir Med 107: 911-918, 2013.
11. Cheng M, Guo Z, Li H, Li Z, Li C and Geng C: Identification of sequence polymorphisms in the mitochondrial displacement loop as risk factors for sporadic and familial breast cancer. Tumour Biol 35: 4773-4777, 2014.
12. Li H and Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595, 2010.
13. Hiyama T, Tanaka S, Shima H, Kose K, Tuncel H, Ito M, Kitadai Y, Sumii M, Yoshihara M, Shimamoto F, et al: Somatic mutation in mitochondrial DNA and nuclear microsatellite instability in gastric cancer. Oncol Rep 10: 1837-1841, 2003.
14. Liang BC: Evidence for association of mitochondrial DNA sequence amplification and nuclear localization in human low-grade gliomas. Mutat Res 354: 27-33, 1996.
15. Soares VY, Silva JC, Silva KR, Pires e Cruz Mdo S, Santos MP, Ribolla PE, Alonso DP, Coelho LF, Costa DL and Costa CH: Identification of blood meal sources of Lutzomyia longipalpis using polymerase chain reaction-restriction fragment length polymorphism analysis of the cytochrome B gene. Mem Inst Oswaldo Cruz 109: 379-383, 2014.
16. Montebugnoli L, Leonardi E, Morandi L, Farnedi A, Gissi DB, Marchetti C, Tarsitano A, Balbi T, Gentile L, Cocchi R and Foschini MP: Genetic relationship between multiple squamous cell carcinomas arising in the oral cavity. Head Neck 36: 94-100, 2014.
17. Shakhssalim N, Houshmand M, Kamalidehghan B, Faraji A, Sarhangnejad R, Dadgar S, Mobaraki M, Rosli R and Sanati MH: The mitochondrial C16069T polymorphism, not mitochondrial D310 (D-loop) mononucleotide sequence variations, is associated with bladder cancer. Cancer Cell Int 13: 120, 2013.
18. Hu N, Kadota M, Liu H, Abnet CC, Su H, Wu H, Freedman ND, Yang HH, Wang C, Yan C, et al: Genomic landscape of somatic alterations in esophageal squamous cell carcinoma and gastric cancer. Cancer Res 76: 1714-1723, 2016.
19. Cheng C, Zhou Y, Li H, Xiong T, Li S, Bi Y, Kong P, Wang F, Cui H, Li Y, et al: Whole-genome sequencing reveals diverse models of structural variations in esophageal squamous cell carcinoma. Am J Hum Genet 98: 256-274, 2016.

20. Gough DJ, Corlett A, Schlessinger K, Wegrzyn J, Larner AC and Levy DE: Mitochondrial STAT3 supports RAS-dependent oncogenic transformation. Science 324: 1713-1716, 2009.
21. Miyawaki Y, Kawachi H, Ooi A, Eishi Y, Kawano T, Inazawa J and Imoto I: Genomic copy-number alterations of MYC and FHIT genes are associated with survival in esophageal squamous-cell carcinoma. Cancer Sci 103: 1558-1566, 2012.
22. Cao W, Wu W, Yan M, Tian F, Ma C, Zhang Q, Li X, Han P, Liu Z, Gu J and Biddle FG: Multiple region whole-exome sequencing reveals dramatically evolving intratumor genomic heterogeneity in esophageal squamous cell carcinoma. Oncogenesis 4: e175, 2015.
23. Xu X, Chen G, Wu L and Liu L: Association of genetic polymorphisms in PTEN and additional gene-gene interaction with risk of esophageal squamous cell carcinoma in Chinese Han population. Dis Esophagus 29: 944-949, 2016.
24. Doucet-O'Hare TT, Sharma R, Rodić N, Anders RA, Burns KH and Kazazian HH Jr: Somatically acquired LINE-1 insertions in normal esophagus undergo clonal expansion in esophageal squamous cell carcinoma. Hum Mutat 37: 942-954, 2016.
25. Abnet CC, Huppi K, Carrera A, Armistead D, McKenney K, Hu N, Tang ZZ, Taylor PR and Dawsey SM: Control region mutations and the 'common deletion' are frequent in the mitochondrial DNA of patients with esophageal squamous cell carcinoma. BMC Cancer 4: 30, 2004.
26. Rafael G and Carmen G: Animal mitochondrial biogenesis and function: A regulatory cross-talk between two genomes. Gene 263: 1-16, 2001.