Insights Breast Cancer through mitochondrial genomics

Raúl Cuauhtemoc Baptista-Rosas (✉ raul.baptista@academicos.udg.mx)

Universidad de Guadalajara - Centro Universitario de Tonalá <https://orcid.org/0000-0002-0273-4740>

Arieh Roldán Mercado-Sesma

Universidad de Guadalajara - Centro Universitario de Tonalá

Luis Enrique Hernández

Universidad de Guadalajara - Centro Universitario de Tonalá

Jorge Hermilo Vega Avalos

Universidad de Guadalajara - Centro Universitario de Tonalá

Jonathan Alejandro O'Connor Miranda

Universidad de Guadalajara - Centro Universitario de Tonalá

Alma Aurora Arreola Cruz

Universidad Autonoma de Baja California - Campus Ensenada

José Alfonso Cruz Ramos

Hospital Civil de Guadalajara

Luis Daniel Hernández Ortega

Universidad de Guadalajara - Centro Universitario de Tonalá

Alberto Daniel Rocha Muñoz

Universidad de Guadalajara - Centro Universitario de Tonalá

Felipe de Jesús Orozco Luna

Data Analysis of Supercomputing Center, University of Guadalajara, Mexico

Marco Andres Perez Rodriguez

Data Analysis and Supercomputing, University of Guadalajara, Mexico

Rocío Ortiz López

Instituto Tecnológico y de Estudios Superiores de Monterrey

Research article

Keywords: Breast cancer, mitochondria, polymorphism, mitochondrial genomics

Posted Date: September 24th, 2019

DOI: <https://doi.org/10.21203/rs.2.14726/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Breast cancer has a predominant incidence and prevalence in Caucasian women. Although alterations in the mitochondrial genome probably play an important role in carcinogenesis, the actual evidence is ambiguous and inconclusive. The purpose of the present work was to determine the frequency of polymorphisms associated with breast cancer exploring mitochondrial sequences of clinical cases diagnosed from different origins reusing information available in the public free access database GenBank.

Results

110 mtDNA sequences associated to breast cancer cases were identified, of which 72 sequences were complete mitochondrial chromosome and 38 partial sequences. Of these 110 identified sequences, 85 belong to patients with a confirmed diagnosis of breast cancer and 25 complete sequences were obtained from healthy mammary tissue of the same patients. In addition, we obtained 49 complete mtDNA sequences from Eskimo and Inuit individuals, a low prevalence breast cancer population, used as controls. From patients diagnosed with breast cancer, T16519C was found in 60% of the breast cancer sequences and less than 20% of controls sequences ($P = 0.427$). Two changes were found in D130 in all sequences analyzed, the first characterized by the insertion of base C in position 315 and, the second, a transition of cytosine by at least two cytosines and one thymine at position 309 always followed by another transition of thymine by cytosine in the position 310.

Conclusions

Both polymorphisms, T16519C and D130, are evolutionarily related to haplogroup H of Caucasian origin of Indo-European and Euro Asiatic origins. However, they were also found in all sequences of non-European origin with breast cancer. The changes in the mtDNA of patients with breast cancer are polymorphic and not associated with a few unique mutations.

Background

Breast cancer has ranked first for more than a decade as the leading cause of death from malignant neoplasms in women, mainly impacting age groups between the fourth and sixth decade of life and affecting all socioeconomic levels (1,2). Mortality rates for this neoplastic disease have shown a significant increase in the last five decades. Between 1955 and 1960, the rate was around two to four deaths per 100,000 women, rising steadily in adult women of all groups of age, but with a greater impact from 30 years of age (3).

The actual evidence indicates that breast cancer is a multifactorial disease, with modifiable and nonmodifiable risk factors. Obesity, postmenopausal state, smoking, and physical inactivity are the most important modifiable factors. Inherited genetic autosomal mutations have been associated in ~10% of total cases with breast cancer (4).

In Latin America, there are important regional differences regarding incidence and prevalence, with a higher frequency of breast cancer in areas with the highest socioeconomic incomes compared with indigenous areas where the socioeconomic level is lower. This evidence is compatible with the epidemiology of breast cancer worldwide, where it affects the developed countries with greater importance and has a lower incidence and prevalence rates in third world populations (5,6). More than a socioeconomic or environmental component, several authors have mentioned the relationship with the genetics of different human populations, with special attention to

the different mitochondrial haplogroups. However, there is a lack of evidence of mitochondrial haplogroups with sporadic (7) or familial hereditary breast cancer (8).

Some authors have suggested that mitochondrial DNA (mtDNA) alterations play an important role in carcinogenesis (9–17) involving the control region *D-loop* of the mtDNA, due to it containing essential sequences for transcription and replication. Previous studies have suggested that polymorphisms in these non-coding regions of the control area could play an important role in the pathogenesis of breast cancer (18–22). These changes in the sequence can be associated with a particular phenotype and serve as markers for the development of various malignancies (23). However, some studies suggest that different mitochondrial polymorphisms vary from one population to another. In addition, some polymorphisms are more related to the presence of breast cancer than others (24–26).

Many works have focused on polymorphisms associated with regions related to the synthesis of proteins and RNA for mitochondrial metabolism. Among the most cited in literature, the A10398G polymorphism, related with synthesis of the protein NADH-ubiquinone oxidoreductase 3 (ND3), has been identified as a biomarker in different populations as Polish (27), Indian (23,28), Chinese (29,30), and mainly in Afro-descendant groups (23,31–33). This polymorphism has also been associated with metabolic syndrome and mental disorders in populations of Asiatic origin (34). However, in a meta-analysis, it has been identified that there is no association when analyzing this polymorphism individually, without correlating with other mitochondrial polymorphisms in women affected by this malignancy (35). Although in Caucasian American human groups with European descent, the A10398G substitution confers an increase in the risk of breast cancer, other polymorphisms with more statistical significance have been identified, such as the T16519C located in the control region (36–41).

Other studies have focused on T16519C (19,31,32,34,36). In particular, on the mutation in the *D-loop* due to an increased breast cancer risk, either occurring singularly or in association with other mitochondrial protein-coding genes alterations such as A10398G, G13368A or C14766T (34,36). Moreover, the association of several variants resulted in a significant predictive breast cancer factor. Indeed, A10398G, together with some other mutations such as T4216C, G9055A, A12308G or T16519C, is considered to increase the risk of developing breast cancer in women (34, 38).

The frequency of two novel polymorphisms in the *D-loop*, one at position 16290Tins and the other at position 16293Adel, was statistically significant more prevalent in breast cancer patients than in control subjects. The study suggests that two novel polymorphisms in the *D-loop* may be candidate biomarkers for breast cancer diagnosis in Bangladeshi women (24).

Despite these promising results, and that most of these mtDNA polymorphisms have functional consequences, associations with specific polymorphisms and the risk of cancer have been subject to intense debate. Several studies involving the association of specific polymorphisms with cancer risk have been thoroughly analyzed due to an erroneous experimental design, interpretation and low-quality data (42). Many of these mtDNA variants may not be conclusive due to artifacts related to genotyping errors or inadequate experimental design (7,43,44). However, due to its potential usefulness as a diagnostic tool, the mitochondrial DNA study and its relation to cancer must remain an important focus of oncological biomarker research with an adequate study design, population stratification and independent replication from the results.

The purpose of the present work was to determine the frequency of polymorphisms associated with breast cancer exploring mitochondrial sequences of clinical cases diagnosed from different origins reusing information available in the public free access database GenBank.

Methods

The study protocol was approved by the research and ethical committee of the University Center of Tonalá of University of Guadalajara in Mexico. The search for complete and partial mtDNA sequences was performed in the *NCBI Genbank Nucleotide* database (National Center for Biotechnology Information) (<https://www.ncbi.nlm.nih.gov/nucleotide>). For the search of complete chromosomes, fragments smaller than 15,400 base pairs were not considered. Sequences of a small size were researched considering fragments no greater than 1,500 bp.

The search strategy was carried out using the keywords *Homo sapiens*, *mitochondrion*, and *Breast Cancer*, using Booleans and filters to select the results and excluding the term "*ancient human remains*" to facilitate the search in the database. As control sequences, we obtained samples from *Eskimo* and *Inuit*, a low prevalence population of breast cancer.

Once the sequences were identified in the database, the accompanying metadata were analyzed to validate the diagnosis of breast cancer and eliminate those that did not correspond or did not specify the presence of the neoplastic disease. From the selected sequences, the reference of the work appointment was obtained to evaluate the end of the sampling and its objectives in the experimental design of each particular work.

Once this was done, using the sequence identification number of *GenBank*, the complete sequences were obtained in FASTA format files. Subsequently, The FASTA files were used as inputs to feed the University of California Santa Cruz Genome Browser (<http://genome.ucsc.edu/>), (45) and aligned the sequences against the *revised Cambridge Reference Sequence* or (*rCRS*) with the latest available version (*GRCh38 / hg38 Assembly*). The FASTA files were aligned using *BLAT*, a *BLAST*-like alignment tool (46) and created custom sequence tracks in the browser with mtDNA obtained from breast cancer cases, allowing us to observe the variants in a general way and panoramic perspective in the mitochondrial genomic landscape.

The haplotyping and accounting of polymorphisms was carried out using the MITOMASTER application (<https://www.mitomap.org/foswiki/bin/view/MITOMASTER>), building a database in a *.xlsx* spreadsheet (*Microsoft Excel*) for tabulation.

The criteria for the classification of the different haplogroups can be found in the *Phylotree* database (<http://www.phylotree.org/>) (47). Once the haplotype was performed and the variants in the sequences were identified, a database was constructed to quantify the haplogroups, haplotypes and, main subclades of the population analyzed with breast cancer.

To evaluate the presence of phantom polymorphisms and discordance between the different haplogroups, the *Haplogrep* package (available in <http://haplogrep.uibk.ac.at/>) was used with the latest version of *PhyloTree 17*. Furthermore, using as a reference the sequence NC_012920 deposited in *GenBank* as *rCRS* (28) and with the tool *Haplogrep 2* (v.2.1.19) accessible in the same platform, we constructed a graphical phylogenetic tree to explore the population structure and identify the possible phantom polymorphisms not assigned to the resulting haplogroup. All variants with a <3 score were considered and which are found in at least two different sequences according to the criteria of Soares et al. (48)

Only the complete mtDNA sequences were analyzed and a phylogenetic tree was generated, graphically representing the presence of polymorphisms associated to the various haplogroups, global and local private mutations, regressive mutations, as well as the expected loss of polymorphisms for each haplogroup.

For the analysis of the structure of the populations, a single file with all the sequences in FASTA format was used as an input for the Multiple sequence alignment application with *CLUSTAL W* using the *MEGA (Molecular Evolutionary Genetics Analysis)* package (<https://www.megasoftware.net/>). Once aligned, mitochondrial single nucleotide polymorphism (mtSNP) profile and frequency of the package *mtDNAprofiler* (<http://mtprofiler.yonsei.ac.kr/>) was used. From these tools, changes in the sequence were identified as well as insertion, deletion, and heteroplasmy sites. Also, all the sequence alignments were reviewed to detect polymorphic heteroplasmic sites.

The number of mtSNPs as total polymorphic loci, the number of fixed differences, the polymorphic and monomorphic mutations among the populations, the shared polymorphisms and the average number of different nucleotides between populations were also estimated. In this way, the identification of the most common polymorphisms in the control region and in the coding region of mtDNA was facilitated.

To evaluate its correlation, the breast cancer group frequency was compared with select polymorphisms in breast cancer mtDNA sequence and controls using the χ^2 test using *SPSS* software (IBM Corporation, Armonk, NY) version 21.

Results

In the search in the *NCBI GenBank* database with the previously defined criteria, 110 mtDNA sequences were identified, of which 72 sequences were complete (16.5 kb) and 38 partial sequences (Table 1). Of the 3 partial sequences smaller than 0.5 kb, 12 sequences covered the HVR1 region and 26 sequences from the HVR2 region of the mitochondrial chromosome. When reviewing the information contained in the metadata of each sequence, the origin of each sequence was confirmed.

Of these 110 identified sequences, 85 belonged to patients with a confirmed diagnosis of breast cancer and 25 complete sequences were used as controls (Table 1). These control sequences were distributed in three categories: 6 distant normal tissue sequences and 7 para-cancerous normal tissue sequences obtained from Wang et al. 2007 (49); also 12 sequences obtained from cells were identified as normal under laser capture microdissection obtained from biopsies of patients diagnosed with invasive breast carcinoma reported in the work of Fendt et al. (50). It is important to note that 42 sequences (38%) were reported in Europe by Gasparre et al. (51) and Fendt et al. (50), and about half of those complete sequences obtained (50.6%) and 28.9% partial sequences were related with Caucasian or Euro Asiatic origin with the haplogroup H (Tables 2 and 3).

About the partial sequences, 14 sequences of the HVR1 region were published by Ghatak, et al. (52) and 24 remaining sequences, 12 from the HVR1 region and 12 from the HVR2 region, were associated with direct referrals without related publications provided by Darvishi, et al. (53) (Table 1).

The number of total polymorphic *loci* without considering potential phantom variants was analyzed and counted, an obtained difference significative associated to less polymorphic sites in cancer sequences compared with controls and Eskimo sequences (Supplemental material Breast cancer database). The total number of polymorphic *loci* without considering potential phantom variants was analyzed and near half of the variations related with polymorphic sites were in the control region with average 10.7 (with a minimum of 1 and maximum of 14 per sequence). The results of haplotyping in the population with breast cancer showed different clades and subclades distributed in a general way in 15 haplogroups (Table 2 and Supplemental material Figure S1). Seven possible phantom polymorphisms (54) shown in Table 3 were identified in the 47 complete mtDNA sequences and were excluded from the subsequent analyses.

The most frequent haplogroup in the sample studied with breast cancer was H with 23 individuals representing about half of the population, followed by haplogroup B in 7 individuals (13.5%) and haplogroups C, N and J, with 3 individuals each one, representing as a whole 11.4% of the population. Haplogroups A, B, and C, usually associated with Asiatic and Amerindian populations, were found in 12 individuals from the breast cancer group, which together represented almost a quarter of the population (23.1%) (Table 2 and Supplemental material Figure S1).

After haplotyping and navigation in *UCSC* genome browser exercise, we found 214 different polymorphisms shared sequences that we used as controls, 241 unique polymorphisms found only in breast cancer and only one polymorphism, A3480G was found exclusive in a control sample (Figure 1 and Supplemental material Figure S2).

The cumulative analysis shows that most of the variation in breasts cancer sequences are mainly unique polymorphisms and only 13 mutations are related in almost 50% of the sequences studied. In contrast, in the control group, we observed that in a half of the sequences there are up to 22 different polymorphisms, being proportionally less frequent if we compare it with the samples associated with breast cancer (Figures 1 and 2). Most of the variation was found in unique polymorphisms distributed along the mitochondrial genome (Supplemental material Figure S5), but the most frequent repetitions in different sequences are concentrated in the control region between positions 16024 and 576 of the mtDNA (Supplemental material Figure S4).

The most frequently shared polymorphism between breast cancer and controls were A16183C in 7 sequences, T16217C in 5 sequences, T16298C in 4 sequences, A249d, T401d, C16184A, T16092C, A16235G, A263G, A15326G, A1438G and, A4769G in 3 sequences (Table 4 and Figure 2). Moreover, 32 polymorphisms were present in two sequences and 60 sequences with a single polymorphism in each sequence. In the position C16147T, there were two different polymorphisms: a transition C/T in 3 sequences and C/A in one.

When we compare the relative frequencies of each sequence, we can observe that there is a significant proportion of exclusive polymorphisms in breast cancer which are practically absent in the control samples (Figure 3). When comparing the absolute frequencies of breast cancer polymorphisms, we could define the variants that were less frequent and which appear more than expected, taking as reference the proportions of the controls (Supplemental material Figure S5).

Even though the most frequent variants in breast cancer and less incidence in controls was G3010A, T16311C, T16189C, and T16519C, only T16519C present a statistical significance after Pearson χ^2 test (Table 6). T16519C, the most important polymorphism, only represented 2.6% of the total sequences shared analyzed (Figure 2). Moreover, it was found in nearly 60% of the breast cancer sequences and only in 20% of the controls (Figure 1 and Supplemental material Figure S3).

From the total of 72 mtDNA sequences obtained from patients diagnosed with breast cancer, 6 common polymorphisms were identified in the sequences: A263G, A750G, A1438G, A4769G, A8860G and, A15326G. These represented 98% of the analyzed sequences obtained of malignant tissue. However, they were also found in most of the control samples.

The polymorphism A10398G, usually associated with breast cancer in other studies, was found only in 17 (24%) of the mtDNA sequences analyzed, without a statistical significance (Table 5), and the frequency of 315.1C. and T1659C, the most recurrent polymorphisms found in our analysis, meeting in isolation or both in the same sequence, but without any statistical significance (Table 5 and Supplemental material Figure S5). With respect to heteroplasmic in mtDNA sequences shown in the summaries in the tables 8 and 9, our search showed in complete

sequences, 22 heteroplasmic changes distributed only in 17 sequences. On the other hand, partial sequences present only 7 heteroplasmic changes in only 4 sequences, all obtained by laser capture microdissection (LCM). These changes were not identified in the rest of the sequences with heteroplasmic somatic transitions described previously by Wang [49]. The polymorphism T2275C was found in the primary cancerous tissue identified as EF114285.1, but not in other normal tissues. The polymorphism A8601G related to primary cancerous sequence EF114276.1 and paracancerous normal sample EF114277.1, obtained from cells without dysmorphic or malignant changes under macro and microscopic analysis, but with polymorphic changes related with neoplastic genotype.

Other common findings were polymorphisms in the control region, clearly showing a characteristic pattern related to the presence of one polymorphism and the absence of other (Supplemental material Figure S5). We were able to identify that in 47 mitochondrial sequences analyzed, all had alterations in D310, characterized only by two situations: The first, manifested by an insertion of three bases, two cytosines and one thymine at position 309, always accompanied by a thymine transition by cytosine at position 310 (C309CCT-T310C); and the second option, characterized by an insertion of a cytosine in the next position 310.1C with HAPLOGREP and 315.1C with MITOMASTER). An interesting observation is that these changes were observed in almost half of the sequences of non-European origin, C309CCT-T310C in 23.4% (11 of 47 sequences) and 315.1C in 21.2% of the cases (10 of the 47 sequences) (Table 5).

The insertion of base C at position 315 (315.1C) was found in 36 (50%) of the cancer samples and only in 12 (48%) of the samples used as controls. When this polymorphism was absent, we identified that in the rest of the analyzed sequences 36 sequences (50%) present the T310C polymorphisms and in position 309. Even though the most frequent polymorphism in position 309 was C309CCT, our analysis found C309CCCT in three sequences (4%) and C309CCCCT in one sequence (1.4%). After the statistical analysis using the X^2 test, C309CCT-C310T and 315.1C were not significantly different, we also ruled out as a potential artifact associated with ghost mutations (Tables 5 and 6).

Discussion

Apparently, there is no relationship between the different haplogroups associated with protection against breast cancer. However, when we analyzed the sequence information, it is possible to observe polymorphisms commonly related to breast cancer. It is important to note that most of the samples have a European origin and these are similar to half of the population with breast cancer (44.9%) and to the sequence used as the *rCRS* reference genome in the database and genomic browsers.

Otherwise, of the 241 polymorphisms exclusive in neoplastic tissue, only 9 were repeated in more than 3 different sequences in our analysis: A16183C repeat in 11 different sequences, T16217C in 5, and T16298C in 4 sequences. Some of these changes in the mtDNA sequences are related with other malignant neoplastic diseases (49).

These variants in D310 were present in 98% of the analyzed sequences. Although most of these polymorphisms are commonly associated with specific haplogroups, it was identified that the 315.1C polymorphism, evolutionarily is associated with haplogroup H of Caucasian origin and infrequent in Indo-European haplogroups with a frequency lower than 1.5%. In our analysis in 99% of sequences of non-European origin with breast cancer.

It is an important focus that T16189C and T16311C are positioned in the HVR1 and T16519C is placed in a non-coding position in the control region. In our analysis we found a characteristic pattern related to breast cancer characterized by the presence of 315.1C and the absence of C309CCT and C310T; and in the same way, when

polymorphisms C309CCT and C310T are present, 315.1C was invariably absent. Although when analyzed individually, each of these changes has no statistical significance (tables 5 and 6), the 315.1C polymorphism was identified related in the literature mainly to haplogroup H of Caucasian origin and infrequent in Indo-European haplogroups with a frequency lower than 1.5%, in our analysis it was found in 99% of the non-European sequences with breast cancer.

In a previous study, we identified that the polymorphism 315.1C is not associated with breast cancer. However, it has as a reference two samples of a medium prevalence and very low for neoplastic disease (55). This change into the sequence is written 315.1C, where .1C means that a copy of the base C has been inserted at position 315 compared to the *rCRS*, so it is translated into these sequences and the *rCRS* is slightly (a base) more extensive than many other sequences in mitochondrial genomes.

The polymorphism 315.1C is considered to be one of the most recent mutations occurred in the last 60,000 years. This is the type of mutations, now well accepted and detailed in many of the phylogenetic trees, related to haplogroup H (56). Although this haplogroup is related to populations of ethnic origin, it has been found to have a high prevalence in Latin American mestizos and in other populations with different haplogroups (55–61).

The polymorphisms, 315.1C and C309CCT-T310C, which are found in the non-coding HVR2 fragment of the *D-Loop* at the mtDNA control region, are considered as an access point due to the high frequency of changes. Furthermore, the modifications of these have been given the go-ahead, may have implications with the appropriate transcription and regulation of mitochondrial genome expression. This Poly-C hotspot area is located in the mitochondrial genome between positions 303–315, defined as D310, and is considered very polymorphic and can be different between direct relatives by the mother line (48,62–64).

Due to the high prevalence in Caucasian groups, 315.1C usually is not considered during the bioinformatic analysis and is not used in the construction of phylogenetic trees. This is the reason why it is usually going unnoticed. However, there are reports in the literature where the 315.1C polymorphism is associated with various forms of cancer and other chronic-degenerative diseases (53,63). This poly-C tract in the mitochondrial *D-Loop* located commonly between 303 and 315 nucleotides has been identified as a frequent *hotspot* mutation region in human neoplasia, including breast cancer (56–58), suggesting that mtDNA instability in this site may be a common characteristic in this malignant disease.

In Mexico, there is a high prevalence and incidence of breast cancer, however, is lower than the United States of America or Western Europe, due to both populations have Caucasian ancestry. In the northern of Mexico there is a higher prevalence of breast cancer compared to central and southern of the country. This discrepancy could explain it because in the north there is European ancestry to difference to the south where is mainly indigenous and the principal variants are C309CCT-C310T, del306 and C310T-G306C, most commonly related to Amerindian populations with lower frequency of breast cancer (data not published by Baptista Rosas et al. 2019).

Is important to remark that in our analysis the presence of T16519C, was found in near 60% of the breast cancer sequences and less than 20% of controls sequences. Other studies have focused on this polymorphism, apparently, without relation with other polymorphisms previously described in the available literature (20,32–35,38). Moreover, the association of several variants resulted in a significant predictive breast cancer factor. Indeed, A10398G, usually associated with breast cancer in other studies, (28,29,31–34,52) was found only in 24% of the mtDNA sequences analyzed, without statistical significance and in association with other mutations such as T4216C, G9055A, A12308G or T16519C. It was found to increase the risk of developing breast cancer (36).

T16519C is associated in the *Phylotree* database between the branches to haplogroup L5 and Haplogroup L2 with other 11 mutations: T16311C, T16189C, C16187T, A15301G, C13506T, A13105G, T10810C, G10688A, C8655T, T825A, and G247A, back to its CRS value and usually found in Euro Asiatic mitogenomes. These ancient polymorphisms have arisen 140,000 years ago, and are considered part of the suggested *Mitochondrial Eve genome*, supporting the theory to first human migration outside of African continent (56).

Once again, the high number of mutations suggests that there was a significant bottleneck in human evolution at the time, perhaps around 120,000 years ago, which might have lasted for many thousands of years (56).

Finally, the molecular mechanisms underlying that increased risk of cancer, due to these specific mtDNA polymorphisms, are still unclear. The control region is important for the regulation of mitochondrial genome replication and expression. The polymorphisms in this region might affect mtDNA replication and lead to electron transport chain alteration, resulting in compensatory increases in glycolytic ATP production (66). However, one of the inevitable products of these alterations is an increased release of highly reactive oxygen species, which may lead to mitochondrial abnormalities. These abnormalities invoke a mitochondria-to-nucleus retrograde response and finally result in nuclear genome damage, which contributes to initial events related to carcinogenesis (67–69). The regulation of mitochondrial genome replication from the control region might also lead to the mtDNA damage (70–72) and, with a critical number of mitochondrial genome changes, to cellular apoptosis (73), which finally could induce cancer development.

The evidence so far suggests that these changes previously described in the literature and the findings in our analysis are probably the result of damage by oxidative stress at the level of the mitochondrial genome rather than being the origin of the changes associated with carcinogenesis.

Further analysis is required with the objective to evaluate more sequences and calculate the correlation with the risk in the development of malignancy.

This approach will give us a general perspective about the importance of consolidating the evidence in specialized repositories and that has not been crossed yet. Finally, it is important to include populations of different origins than Caucasian with the purpose of showing the genetic differences related to breast cancer, regardless of high and low prevalence. With this focus, our understanding of these malignant diseases will improve through the interpretation of the complete mitogenome.

Conclusions

Both polymorphisms, T16519C and D130, are evolutionarily related to haplogroup H of Caucasian origin of Indo-European and Euro Asiatic origins. However, they were also found in all sequences of non-European origin with breast cancer. The changes in the mtDNA of patients with breast cancer are polymorphic and not associated with a few unique mutations.

List Of Abbreviations

BMI: Body Mass Index; *ENSANUT*: National Survey of Health and Nutrition (Mexico); *IDF*: International Diabetes Federation. *NCDs*: Non-communicable Diseases; *OR*: odds ratio; *PNS*: National Health Survey (Brazil).

Declarations

Ethics approval and consent to participate

The study protocol was approved by the research and ethical committee of the University Center of Tonalá of University of Guadalajara in Mexico.

Consent for publication

Not applicable

Availability of data and materials

Data are available with the author.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding was obtained.

Author contributions

RCBR planned, design and coordinate the study, submitted the final draft as the corresponding author. LEH, JHVA, MAPR and FJOL participated in data collection and bioinformatic analysis. ARMS and JAOM performed the statistical analysis of the data and participated in the interpretation of results. AAAC, JACR, LDHO, ADRM, and ROL were involved in the discussion of results and style correction of the final draft. All authors helped to draft the manuscript and read and approved the final manuscript.

Acknowledgment

We appreciate the comments and recommendations of Robert Kuhn to improve the analysis using *UCSC* Genome Browser and Ian Capstick for manuscript style correction.

References

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics. *CA Cancer J Clin*, 2018;68;7–30.
2. Bray F, Ferlay J, Soerjomataram I. Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 2018;68;394–424.

3. Carioli G, Malvezzi M, Rodriguez T, Bertuccio P, Negri E, La Vecchia C. Trends and predictions to 2020 in breast cancer mortality: Americas and Australasia. *Breast*, 2018;37;163–169.
4. Rojas K, Stuckey A. Breast Cancer Epidemiology and Risk Factors. *Clin Obstet Gynecol*, 2016;59(4);651–672.
5. Rodríguez CSA, Capurso GM. Epidemiología del cáncer de mama. *Ginecol Obstet Mex*, 2006;74;585–593.
6. Sifuentes-Álvarez A, Castañeda Martínez LY, Lugo-Nevarés M, Reyes Romero MA. Factores de riesgo asociados con el cáncer de mama en mujeres del estado de Durango, México. *Ginecol Obstet Mex*, 2015;83:662–669.
7. Riley V, Mesut Erzurumluoglu A, Rodriguez S, Bonilla C. Mitochondrial DNA Haplogroups and Breast Cancer Risk Factors in the Avon Longitudinal Study of Parents and Children (ALSPAC). *Genes (Basel)*, 2018;9(8):395.
8. Gutiérrez Povedano C, Salgado J, Gil C, Robles M, Patiño-García A, García Foncillas J. Analysis of BRCA1 and mtDNA haplotypes and mtDNA polymorphism in familial breast cancer. *Mitochondrial DNA*, 2015;26(2);1940–1744.
9. Mambo E, Chatterjee A, Xing M, Tallini G, Haugen BR, Yeung SC, et al. Tumor-specific changes in mtDNA content in human cancer. *Int J Cancer*, 2005;116(6);920–924.
10. Birch-Machin MA. The role of mitochondria in ageing and carcinogenesis. *Clin Exp Dermatol*, 2006;31(4):548–552.
11. Brandon M, Baldi P, Wallace DC. Mitochondrial mutations in cancer. *Oncogene*, 2006;25(34);4647–4662.
12. Putignani L, Raffa S, Pescosolido R, Aimati L, Signore F, Torrisi MR, et al. Alteration of expression levels of the oxidative phosphorylation system (OXPHOS) in breast cancer cell mitochondria. *Breast Cancer Res Treat*, 2008;110(3):439–452.
13. Rohan TE, Won LJ, Wang T, Haines J, Kabat GC. Do Alterations in Mitochondrial DNA Play a Role in Breast Carcinogenesis? *J Oncol*, 2010;2010:604–304.
14. Yadava N, Schneider SS, Jerry DJ, Kim C. Impaired mitochondrial metabolism and mammary carcinogenesis. *J Mammary Gland Biol Neoplasia*, 2013;18(1):75–87.
15. Jezierska-Drutel A, Rosenzweig SA, Neumann CA. Role of Oxidative Stress and the Microenvironment in Breast Cancer Development and Progression. *Adv Cancer Res*, 2013;119:107–125.
16. Chandra D, Singh KK. Genetic insights into OXPHOS defect and its role in cancer. *Biochimica et Biophysica Acta - Bioenergetics*, 2011;1807(6):620–625.
17. Singh RK, Srivastava A, Kalaiarasan P, Manvati S, Chopra R, Bamezai RNK. mtDNA germ line variation mediated ROS generates retrograde signaling and induces pro-cancerous metabolic features. *Sci Rep*, 2014;4:6571.
18. Ye C, Shu XO, Pierce L, Wen W, Courtney R, Gao YT, et al. Mutations in the Mitochondrial DNA D-loop Region and Breast Cancer Risk. *Breast Cancer Res Treat*, 2010;119(2):431–436.
19. Duan-Jun Tan, Ren-Kui Bai, Lee-Jun C. Wong. Comprehensive Scanning of Somatic Mitochondrial DNA Mutations in Breast Cancer. *Cancer Research*, 2002;62:972–976.
20. Tommasi S, Fabia P, Weigl S, Bianco A, Pilato B, Russo L, et al. Mitochondrial DNA variants and risk of familial breast cancer: An exploratory study. *International Journal of Oncology*, 2014;44:1691–1698.
21. Parrella P, Xiao Y, Fliss M, Sanchez-Céspedes M, Mazzarelli P, Rinaldi M, et al. Detection of mitochondrial DNA mutations in primary breast cancer and fine-needle aspirates. *Cancer Res*, 2001;61(20):7623–7626.
22. Yu M, Shi Y, Zhang F, Zhou Y, Yang Y, Wei X, et al. Sequence variations of mitochondrial DNA D-loop region are highly frequent events in familial breast cancer *Journal of Biomedical Science*, 2008;15:535–543
23. Singh KK, Kulawiec M. Mitochondrial DNA polymorphism and risk of cancer. *Methods in Molecular Biology*, 2009;471:291–303.

24. Sultana GN, Rahman A, Shahinuzzaman AD, Begum RA, Hossain CF. Mitochondrial DNA mutations—candidate biomarkers for breast cancer diagnosis in Bangladesh. *Chin J Cancer*, 2012;31(9):449–454.
25. Li LH, Kang T, Chen L, Zhang W, Liao Y, Chen J, et al. Detection of mitochondrial DNA mutations by high-throughput sequencing in the blood of breast cancer patients. *Int J Mol Med*, 2014;33(1):77–82.
26. Grzybowska-Szatowska L, Slaska B, Rzymowska J, Brzozowska A, Floriańczyk B. Novel mitochondrial mutations in the ATP6 and ATP8 genes in patients with breast cancer. *Mol Med Rep*, 2014;10(4):1772–1778.
27. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res*, 2016;44(W1): W58–63.
28. Francis A, Pooja S, Rajender S, Govindaraj P, Rao TN, Surekha D, et al. A mitochondrial DNA variant 10398GNA in breast cancer among South Indians: An original study with meta-analysis. *Mitochondrion*, 2013;13:559–565.
29. Jiang H, Zhao H, Xu H, et al. Peripheral blood mitochondrial DNA content, A10398G polymorphism, and risk of breast cancer in a Han Chinese population. *Cancer Sci*, 2014;105(6):639–645.
30. Yu Y, Lv F, Lin H, Qian G, Jiang YS, Pang LX, et al. Mitochondrial ND3 G10398A mutation: a biomarker for breast cancer. *Genet Mol Res*, 2015;14(4):17426–17431.
31. Canter JA1, Kallianpur AR, Parl FF, Millikan RC. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Research*, 2005;66(3):1880–1881.
32. Mims MP, Hayes TG, Zheng S, Millikan RC. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Research*, 2006;66(3):1880–1881.
33. Setiawan VW1, Chu LH, John EM, Ding YC, Ingles SA, Bernstein L, et al. Mitochondrial DNA G10398A variant is not associated with breast cancer in African-American women. *Cancer Genetics and Cytogenetics*, 2008;181(1):16–19.
34. Kishida K, Tominaga M, Matsubara K, Taguchi M, Noguchi M, Tsunawake N, et al. An Association Analysis between Mitochondrial DNA A10398G Polymorphism and Temperament in Japanese Young Adults. *PLoS ONE*, 2009;4(11):e7763.
35. Czarnecka AM, Krawczyk T, Zdrozny M, Lubiński J, Arnold RS, Kukwa W, et al. Mitochondrial NADH-dehydrogenase subunit 3 (ND3) polymorphism (A10398G) and sporadic breast cancer in Poland. *Breast Cancer Res Treat*. 2010;121(2):511–518.
36. Covarrubias D, Bai RK, Wong LJ, Leal SM. Mitochondrial DNA variant interactions modify breast cancer risk. *J Hum Genet*, 2008;53(10):924–928.
37. Czarnecka AM, Krawczyk T, Plak K, Klemba A, Zdrozny M, Arnold RS, et al. Mitochondrial genotype and breast cancer predisposition. *Oncol Rep*, 2010;24(6):1521–1534.
38. Tipirisetti NR, Lakshmi RK, Govatati S, Govatati S, Vuree S, Singh L, et al. Mitochondrial genome variations in advanced stage breast cancer: a case-control study. *Mitochondrion*, 2013;13(4):372–378.
39. Verma M, Naviaux RK, Tanaka M, Kumar D, Franceschi C, Singh KK. Meeting report: mitochondrial DNA and cancer epidemiology. *Cancer Res*, 2007;67(2):437–439.
40. Bai RK1, Leal SM, Covarrubias D, Liu A, Wong LJ. Mitochondrial genetic background modifies breast cancer risk. *Cancer Research*, 2007;67(10):4687–4694.
41. Stefania Weigl, Angelo Paradiso, Stefania Tommasi. Mitochondria and Familial Predisposition to Breast Cancer. *Curr Genomics*, 2013;14(3):195–203.

42. Chatterjee A, Dasgupta S, Sidransky D. Mitochondrial subversion in cancer. *Cancer Prevention Research*, 2011; 4(5):638–654.
43. Salas A, García-Magariños M, Logan I, Bandelt HJ. The saga of the many studies wrongly associating mitochondrial DNA with breast cancer. *Cancer*, 2014;14:659.
44. Mao Q, Gao L, Liu Q, Zou Y, Yang X, Wang H, et al. The A10389G polymorphism of ND3 gene and breast cancer: A meta-analysis. *Biomed Rep*, 2013;1(2):259–264.
45. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*, 2002;12:996–1006.
46. Kent WJ. BLAT - the BLAST-like alignment tool. *Genome Res*, 2002;12:656–664.
47. Van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*, 2009;30(2):E386-E394.
48. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *Am J Hum Genet*, 2009;84(6):740–759.
49. Wang CY, Wang HW, Yao YG, Kong QP, Zhang YP. Somatic mutations of mitochondrial genome in early stage breast cancer. *Int J Cancer*, 2007;121(6):1253–1256.
50. Fendt L, Niederstätter H, Huber G, Zelger B, Dünser M, Seifarth C, et al. Accumulation of mutations over the entire mitochondrial genome of breast cancer cells obtained by tissue microdissection. *Breast Cancer Res Treat*, 2011;128(2):327–336.
51. Gasparre G, Porcelli AM, Bonora E, Pennisi LF, Toller M, Iommarini L, et al. Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. *Proc Natl Acad Sci USA*, 2007;104(21):9001–9006.
52. Ghatak S, Lallawmzuali D, Lalmawia, Sapkota R, Zothanupia, PautuJL, et al. Mitochondrial D-loop and cytochrome oxidase C subunit polymorphisms among the breast cancer patients of Mizoram, Northeast India. *Curr Genet*, 2014;60(3):201–212.
53. Darvishi K, Sharma S, Bhat AK, Rai E, Bamezai RN. Mitochondrial DNA G10398A polymorphism imparts maternal Haplogroup N a risk for breast and esophageal cancer. *Cancer Lett* 2007; 249(2):249–255.
54. Brandstätter A, Sängler T, Lutz-Bonengel S, Parson W, Béraud-Colomb E, Wen B, et al. Phantom mutation hotspots in human mitochondrial DNA. *Electrophoresis*, 2005;26(18):3414–3429.
55. Baptista Rosas RC, Hernández LE, Vega Avalos JH, Mercado Sesma AR, Arreola Cruz AA. Reevaluation of polymorphisms associated to a higher risk of breast cancer in Mexican population through available genetic information from public databases. *Genética médica y Genómica*, 2019;3:51–66.
56. Logan I. A Suggested Genome for “Mitochondrial Eve”. *J Genet General*, 2007;3(2):72–77.
57. Picornell A, Giménez. PJA, Castro.RMM. Mitochondrial DNA sequence variation in Jewish populations. *International Journal of Legal Medicine*, 2006;120(5):271–281.
58. Fridman C, Cardena MMSG, Kanto EA, Godinho MBC, Gonçalves FT. SNPs in mitochondrial DNA coding region used to discriminate common sequences in HV1–HV2–HV3 region. *Forensic Science International: Genetics*, 2011; Supl3: e75–e76.
59. Röck AW, Dür A, van Oven M, Parson W. Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic Science International: Genetics*, 2013;7:601–609.
60. Verma K, Sharma S, Sharma A, Dalal J, Bhardwaj T. Sequence polymorphism data of the hypervariable regions of mitochondrial DNA in the Yadav population of Haryana. *Data in Brief*, 2018;18:164–171.

61. Bhatti S, Abbas S, Aslamkhan M, Attimonelli M, Trinidad MS, Aydin HH, et al. Genetic perspective of uniparental mitochondrial DNA landscape on the Punjabi population, Pakistan. *Mitochondrial DNA Part A*, 2018;29(5):714–726.
62. Bandelt HJ, Quintana-Murci L, Salas A, Macaulay B. The Fingerprint of Phantom Mutations in Mitochondrial DNA Data. *Am J Hum Genet*, 2002;71(5):1150–1160.
63. Aral C, Kaya H, Ataizi-Çelikel A, Akkiprik M, Zönmez Ö, Güllüoglu BM, et al. A novel approach for rapid screening of mitochondrial D310 polymorphism. *BMC Cancer*, 2006;6:21.
64. Alhomidi MA, Vedicherla B, Movva S, Rao PK, Ahuja YR, Hasan Q. Mitochondrial D310 instability in Asian Indian breast cancer patients. *Tumor Biology*, 2013;34:2427–2432.
65. Sequeira A, Martin MV, Rollins B, Moon EA, Bunney WE, Macciardi F, et al. Mitochondrial Mutations and Polymorphisms in Psychiatric Disorders. *Front Genet*, 2012;3:103.
66. Warburg O. On the origin of cancer cells. *Science*, 1956;123:309–314.
67. Bandy B, Davison AJ. Mitochondrial mutations may increase oxidative stress: Implications for carcinogenesis and aging? *Free Radic Biol Med*, 1990;8:523–539.
68. Gille JJ, Joenje H. Cell culture models for oxidative stress: Superoxide and hydrogen peroxidative versus normobaric hyperoxia. *Mutat Res*, 1992;275:405–414.
69. Scarpulla RC. Nuclear control of respiratory gene expression in mammalian cells. *Cell Biochem*, 2006;97:673–683.
70. Lee HC, Li SH, Lin JC, Wu CC, Yeh DC, Wei YH. Somatic mutations in the D-loop and decrease in the copy number of mitochondrial DNA in human hepatocellular carcinoma. *Mutat Res*, 2004;547:71–78.
71. Xie H, Lev D, Gong Y, Wang S, Pollock RE, Wu X, et al. Reduced mitochondrial DNA copy number in peripheral blood leukocytes increases the risk of soft tissue sarcoma. *Carcinogenesis*, 2013;34:1039–1043.
72. Carew JS, Huang P. Mitochondrial defects in cancer. *Mol Cancer* 2002;1:9.
73. Wang X. The expanding role of mitochondria in apoptosis. *Genes Dev*, 2001;15:2922–2933.

Tables

Table 1. Origin and references of sequences associated with breast cancer and controls used for the analysis. For detailed information about characteristics, *GenBank ID* and references, consult *Supplemental Table S1*.

Type of sequence	Number of sequences	Origin of the samples	References
Complete (More than de 16.5 kb).	47	Breast cancer	Gasparre et al. 2007, Wang et al. 2007, and Fendt et al. 2011.
	25	Controls samples obtained of some patients with breast cancer	Wang et al. 2007 and Fendt et al. 2011
	49	North American and Siberian eskimo used as control because low prevalence of breast cancer	Derbeneva et al. 2002, Tamm et al. 2007, Volodko et al. 2008, Gilbert et al. 2008 and Dryomov et al. 2015.
Parcial (Less than 0.5 Kb	38	Delhi y Cachemira in India	Ghatak et al. 2014 and Darvishi et al. 2015.

Table 2. Comparative summary of macrohaplogroups distribution obtained from complete and partial mitochondrial sequences from patients diagnosed with breast cancer. More than half of complete sequences and more than a quarter of partial samples studied belongs to European origin. For detailed information about haplogroups, haplotypes and subclades assigned, consult *Supplemental Table S2*.

Complete mtDNA sequences (All mitochondrial chromosome)			Partial mtDNA sequences (Control region and hypervariable regions)	
Macrohaplogroup	Number of individual sequences	%	Number of individual sequences	%
A	2	2.4	-	-
B	3	3.5	3	7.9
C	1	1.2	1	2.6
D	4	4.7	3	7.9
E	1	1.2	1	2.6
F	1	1.2	-	-
H	43	50.6	11	29.0
J	4	4.7	-	-
K	1	1.2	-	-
M	2	2.4	1	
				0
16189d	2	0		
16422A	2	0		
397T	2	0		
37G	2	0		
16012d	2	0		
574C	2	2		
16353T	2	1		
16390R	2	0		
513.1CA	2	0		
284.1C	2	0		
285.1A	2	0		

Table 3. Comparative summary of the distribution of possible phantom polymorphisms in 47 complete mtDNA sequences with breast cancer diagnosis. To more detailed information like *Genbank* ID number and haplogroups assigned, consult *Supplemental Table S4*.

Position	Amount	Soares Score
3106d	47	0
3107C	47	0
310.1C	39	0
514d	9	0
515d	9	0
398d	8	0
248d	6	0
430A	4	0
49T	4	0
16011d	4	0
8279d	3	0
8272d	3	0
8271d	3	0
8274d	3	0
8273d	3	0
8276d	3	0
8275d	3	0
8278d	3	0
8277d	3	0
16184A	3	1
12562G	2	0
16032.1G	2	0
16258d	2	0
16189d	2	0
16422A	2	0
397T	2	0
37G	2	0
16012d	2	0
574C	2	2
16353T	2	1
16390R	2	0
513.1CA	2	0
284.1C	2	0
285.1A	2	0

Table 4. Frequent polymorphisms in different groups.

Polymorphism	Breast cancer	Control	Distant normal tissue	Para-cancerous normal tissue	LCM normal
		N=25	N=6	N=7	N=12
C309CCT	25 (N=72)	14	3	3	8
T310C	26 (N=72)	13	3	4	6
C315CC	46 (N=72)	12	3	3	6
A10398G	9 (N=47)	8	2	2	4
T16519C	28 (N=47)	5	1	1	3
A16183C	12 (N=59)	0	0	0	0
T16189C	16 (N=59)	2	1	1	0
T16311C	14 (N=59)	0	0	0	0

*Includes different polymorphism found in the same position: C309CCT, C309CCCT and C309CCCCT.

Table 5. Statistical analysis of polymorphisms in subjects with breast cancer and control.

Polymorphisms,	<i>Breast Cancer</i>	<i>Controls</i>	<i>p</i>
	<i>Present/Absent</i>	<i>Present/Absent</i>	
C309CCT	23/24	13/12	0.804
T310C	23/24	13/12	0.804
C315CC	24/23	12/13	0.804
A10398G	9/38	8/17	0.222
A16183C	4/43	0/25	0.133
T16189C	7/40	2/23	0.400
T16311C	7/40	5/20	0.580
T16519C	18/29	12/13	0.427

*Statistical significance $P = 0.05$. χ^2 test**Table 6.** Relation with polymorphisms and presence of breast cancer.

Polymorphism, frequency (%)	Breast cancer group (n=47)	Control group (n=25)	<i>p</i>
<i>No present</i>	13 (18.0)	7 (9.7)	0.236
<i>315.1C</i>	16 (22.2)	6 (8.3)	0.188
<i>T16519C</i>	10 (13.8)	6 (8.3)	0.677
<i>315.1C + T16519C</i>	8 (11.1)	6 (8.3)	0.856
<i>Any polymorphism</i>	34 (47.2)	18 (25.0)	0.310
<i>Any polymorphism vs No present*</i>	34 (47.2)	7 (9.7)	0.010*

*Statistical significance

Table 7. Summary of the heteroplasmic in mtDNA sequences by position in cases diagnosed with breast cancer. The sequence GU592023.1 was identified a heteroplasmic position in 16401, while in GU592042.1 and GU592046.1 was identified in position 16400. The significance of heteroplasmic polymorphism are code following IUPAC single letter codes: symbol *R* name *Purine* and remarks *A* or *G*; symbol *Y* name *Pyrimidine* and remarks *C* or *T*; symbol *S* name *Strong* and remarks *C* or *G*; and symbol *M* are used for *A* or *C*.

Complete sequences heteroplasmy

Position	Genbank ID sequence number	Heteroplasmic polymorphism
1632	GU592046.1	Y
2145	GU592041.1	R
2275	EF114285.1	Y
2998	GU592041.1	Y
5703	GU592036.1	R
7379	GU592035.1	R
7814	EF660946.1	R
7818	EF660934.1	R
8601	EF114276.1	R
9119	EF660937.1	R
9387	EF660950.1	R
12803	GU592045.1	S
13466	EF660947.1	R
15341	GU592040.1	Y
15623	GU592038.1	R
15843	EF660946.1	Y
16106	EF660946.1	M
16182	GU592037.1	M
16189	GU592040.1	M
16189	GU592046.1	M
16390	GU592042.1	R

Control sequence heteroplasmy

Position	Genbank ID sequence number	Heteroplasmic polymorphism
152	GU592032.1	Y
215	GU592030.1	R
8601	GU592047.1	R
5192	GU592031.1	R
5390	GU592031.1	R
15078	GU592032.1	R
15623	GU592023.1	R
16391	GU592023.1	R

Table 8. Summary of the heteroplasmic polymorphisms detected by sequence, not related with haplotyping in mtDNA sequences of cases diagnosed with breast cancer and controls. The meaning of heteroplasmic polymorphism are code following IUPAC single letter codes: symbol *R* name *Purine* and remarks *A* or *G*; symbol *Y* Name *Pyrimidine* and remarks *C* or *T*; symbol *S* name *Strong* and remarks *C* or *G*; and symbol *M* are used for *A* or *C*. LCM: laser capture microdissection.

Breast cancer mitochondrial heteroplasmy

<i>Genebank ID</i>	<i>Position in mtDNA sequence</i>	<i>Type of Heteroplasmy</i>	<i>Haplogroup assigned</i>
number sequence			
GU592046.1	1632 16189	Y M	M35b (M35b2)
GU592041.1	2145 2998	R Y	K1a (K1a1b1c)
EF660946.1	7814 15843 16106	R Y M	V2 (V2)
GU592040.1	15341 16189	Y M	H65a (H65a)
EF114285.1	2275	Y	B5a (B5a2a1a)
GU592036.1	5703	R	R0a (R0a1a3)
GU592035.1	7379	R	H1a (H1ay)
EF660934.1	7818	R	H1h (H1h1)
EF114276.1	8601	R	D4j (D4j3a)
EF660937.1	9119	R	N1a (N1a3a3)
EF660950.1	9387	R	U5a (U5a2c2)
GU592045.1	12803	S	HV11a (HV11a)
EF660947.1	13466	R	H7a (H7a1a)
GU592038.1	15623	R	H11a (H11a5)
GU592037.1	16182	M	H1g (H1g1)
GU592042.1	16390	R	H3q (H3q1)
GU592047.1	8601	R	J1c (J1c11a)

Control sequences

<i>Genebank ID</i> number sequence	Position in mtDNA sequence	Type of Heteroplasmy	Haplogroup assigned	Source
GU592031.1	5192	R	M35b (M35b2)	LCM normal cells control sample of
	5390	R		GU592046.1
GU592023.1	15623	R	H11a (H11a5)	LCM normal cells control sample of
	16391	R		GU592038.1
GU592032.1	152	Y	J1c (J1c11a)	LCM normal cells control sample of
	15078	R		GU592047.1
GU592030.1	215	R	HV11 (HV11a)	LCM normal cells control sample of
				GU592045.1

Figures

Figure 1

Comparative cumulative proportions of breast cancer and control polymorphisms.

Figure 2

Frequency of shared and exclusive polymorphism between breast cancer and controls.

Figure 3

Proportion of shared and unique polymorphism between breast cancer and control sequences.

Figure 4

D310 analysis Haplogroup H vs different H.

Figure 5

D310 analysis Haplogroup H vs different H vs Eskimo.

Figure 6

D310 analysis Haplogroup A.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [3.SupplementalmaterialFigS3Comparisonofbreastcancer.pdf](#)
- [TableS2.docx](#)
- [BaptistaRosas2019Breastcancerdatabase.xlsx](#)
- [1.SupplementalmaterialFigS1Phylogenetictree.pdf](#)
- [Breastcancerdatabase.xlsx](#)
- [6.SupmatFigS6Positions315.pdf](#)
- [5.SupplementalmaterialFigS5Proportional.pdf](#)
- [2.SupplementalmaterialFigS2Mitochondrialgenome.pdf](#)

- [4.SupplementalmaterialFigS4ParetochartofBreastcancer.pdf](#)
- [TableS1Originandreferencesofthemitochondrial.docx](#)
- [TableS3.docx](#)