Scientific Research Publishing

Open Journal of Genetics, 2022, 12, 43-56
https://www.scirp.org/journal/ojgen
ISSN Online: 2162-4461
ISSN Print: 2162-4453

S53G SNP in Cathepsin B: Can Be Used as a Molecular Biomarker to Breast Cancer?

Letícia Cristina Dalzy Castro[1,2], Danielle Queiroz Calcagno[1,3], Carlos Eduardo de Melo Amaral[4], Luiz Carlos Santana da Silva[1,2,3]

[1]Programa de Pós-graduação em Oncologia e Ciências Médicas, Universidade Federal do Pará, Belém, Brasil
[2]Laboratório de Erros Inatos do Metabolismo, Universidade Federal do Pará, Belém, Brasil
[3]Núcleo de Pesquisa em Oncologia, Universidade Federal do Pará, Belém, Brasil
[4]Fundação Centro de Hematologia e Hemoterapia do Pará, Belém, Brasil
Email: leticiadalzy7000@gmail.com

How to cite this paper: Castro, L.C.D., Calcagno, D.Q., de Melo Amaral, C.E. and da Silva, L.C.S. (2022) S53G SNP in Cathepsin B: Can Be Used as a Molecular Biomarker to Breast Cancer? *Open Journal of Genetics*, **12**, 43-56.
https://doi.org/10.4236/ojgen.2022.124005

Received: September 1, 2022
Accepted: December 4, 2022
Published: December 7, 2022

Abstract

Cathepsin B (CTS B) is a proteolytic enzyme that participates in several important biological processes. However, when it is altered can be involved in development of breast cancer, a disease with high incidence and mortality rate among women. Many studies have shown a correlation between high CTS B expression and breast tumor. Furthermore, it has been shown that the SNP rs1803250 (S53G) leads to structural and functional changes in protein that can make it pathogenic. The present study aimed to evaluate a possible association of SNP rs1803250 (S53G) with breast cancer. For this, real-time PCR was performed on a sample collected in the State of Pará, with 127 patients and 122 controls. The SNP frequency in this region was 0.12, according to a research project in progress that aims to identify Amerindians molecular alterations. This indicates that the SNP is found in region with a distribution close to worldwide frequency of 0.09. Our results showed that the SNP was in Hardy-Weinberg Equilibrium in collected sample, but C variant allelic frequency was 0.08 in both patients and control groups, which is extremely similar to global average. Moreover, homozygotes CC was not found in the sample and SNP genotypes frequency in patients and control groups was not significantly different. In addition, statistical analysis showed that the SNP did not have to correlate with tumor subtypes nor with tumor staging. Therefore, according to the analyzed sample, the SNP rs1803250 has no association with breast cancer and it cannot be considered a molecular biomarker for breast cancer.

Keywords

Cathepsin B, Polymorphism, Breast Cancer

1. Introduction

Breast cancer is the most incident and is the leading cause of death among cancers in women [1]. In Brazil, it is the second cancer type more incident in women with a rate of 43.74 per 100 hundred women in 2021. In the same year, this rate was 18.24 in state of Pará [2].

Since it is a very heterogeneous pathology, patients prognosis have many variables as the breast tumor subtypes based on HER2 and hormonal receptors expressions [3]. Thus, there are four categories: luminal-A, luminal B, HER2-positive and basal. Luminals are defined when the estrogen (ER+) and/or progesterone (PR+) receptors are present, and when there is high expression, they are called luminal-A and when they have low expression and a high proliferation rate, they are called luminal-B [4]. The prognosis for patients who have these types of tumors is generally good due to the existence of specific target drugs such as tamoxifen [5] [6].

The category that is classified as HER2-positive is characterized by high expression of these receptors in tumor cells compared to normal tissues. Like the previous ones, this one has a specific target therapy favoring a good prognosis for patients [4] [7]. The basal category is composed by triple negative tumors (TN), they do not have hormone receptor expression and HER2 also does not show increased expression. However, it is necessary to analyze the expression of a large set of genes through molecular tests to define this tumor type. So, there is a group of patients that is TN non-basal, and these are considered aggressive and have a worse prognosis due to absence of target-specific therapies since some treatments are only effective in TN basal type [4] [8] [9] [10].

In addition, there is a distinction that can be made between pure HER2+ and luminal-HER2+ individuals due to the fact that patients who have both hormone receptor and HER2 receptor expression may have a different prognosis than luminal-HER2 negative patients and HER2-pure patients. This occurs because the presence of hormone receptors and HER2 in the same patient can interfere with target treatments effectiveness [11] [12].

Several studies analyzed cathepsins expression in breast cancer, and many demonstrated increased expression of these enzymes compared to control group [13] [14]. These proteins are a diverse group of proteolytic enzymes classified based on their structure and catalytic site into: serines (A and G), cysteines (B, C, F, H, L, K, O, S, V, W and X) and aspartic (D and E) [15] [16]. These enzymes need an acidic pH to become active. In this way, the organelle where they predominate is lysosome and there, they do their main function, recycling of proteins through peptide bonds irreversible cleavage [16] [17] [18] [19].

Cathepsins are also present in other places such as nucleus, cytosol and even in extracellular environment, playing important roles, including, organism growth and development, immune response and collagen biochemical route in bones and cartilages [15] [20] [21]. However, cathepsins altered levels can harm body homeostasis and may contribute to development of pathologies, including breast

cancer [17] [18].

Dysregulated cathepsins can cause, for example, exacerbated instability of collagen molecules, which can lead to accelerated extracellular matrix degradation, so that malignant cells leak into the blood, causing possible metastases in distant tissues [17]. Thus, Cathepsin B (CTS B) may be used by cancer cells since malignant breast tissues were able to increase its expression, being even more significant in ER+ tumors, and this was associated with invasiveness [22] [23]. There are, already, studies about drugs to inhibit Cat B expression in breast cancer as Rhenium(I)-dyselenether that when present inhibit Cat B causing loss of cell viability of malignant MDA-MB231 cells *in vitro* [24].

These biochemical aspects may have a cause in molecular factor. Studies with cathepsins have shown an association of single nucleotide polymorphisms (SNPs) presence with dysregulation of these enzymes in breast cancer [25] [26]. Tointou *et al.*, 1994, observed that a SNP in pro-CTS D, with exchange of cytosine for thymine, causes pro-enzyme increased secretion and delay in its maturation, leading to participation in breast tumors development [27]. Regarding the rs10030044 SNP, which exchanges a thymine for a guanine 5' of the CTS O gene, it was shown that even being an intronic gene, it can positively regulate CTS O expression, in addition, this SNP were associated with a lower overall survival and lower disease-free survival in ER+ breast cancer patients who were using targeted tamoxifen therapy to prevent recurrence [25] [26].

CTS B SNP rs1803250 is located on chromosome 8p23.1, at position 11852665. This mutation is caused by the replacement of T nucleotide by C variant, resulting in exchange of the original serine amino acid residue by a glycine. This is a missense mutation, that cause changes in amino acid sequence and one study performed analyzes on silica and molecular dynamics simulations that suggested that SNP 1803250 could effectively lead to structural or functional changes in the translated enzyme, therefore its presence could be pathogenic [28] [29].

According to Genome Aggregation Databases (GenomAD), which aggregates information about genome and exome sequences, the worldwide frequency of C variant allelic of the SNP rs1803250 is 0.09 [28]. A research project in progress that aims to identify Amerindians molecular alterations, sparce information in international databases, sequenced the exome of 95 individuals with gastric cancer in regional population of this study, and the SNP rs1803250 had a frequency of 0.12, which indicates that the SNP is found in this region with a distribution close to global average (unpublished data).

Thus, the present study analyzed the SNP rs1803250, investigating presence or absence of a significant association in patients with breast cancer in relation to global average. In addition, the study proposes to analyze whether there is a correlation between this SNP and tumor staging and/or tumor subtypes, in order to determine whether SNP rs1803250 could be the reason for CTS B high expression in patients with breast cancer, and thus would be indicated to be part of a molecular biomarkers panel for the disease.

2. Materials e Methods

2.1. Samples Obtaining

Blood samples were collected from 127 breast cancer patients at the João de Barros Barreto University Hospital (HUJBB/UFPA) in Belém in the State of Pará. The patients were all women and were aged between 24 and 88 years. In addition, data on tumor subtypes and TMN staging were collected, and only patients with no previous history of other cancer types were accepted, according to medical records. As a control, 102 women were selected, who underwent several tests, including mammography, and absence of a malignant breast tumor was confirmed. All signed the free and informed acceptance term, giving the sample for this research. The tube used for collection contains EDTA as an anticoagulant and the samples were stored at −80˚C degrees, if necessary.

2.2. DNA Extraction and SNP Genotyping

DNA extraction was performed following phenol-chloroform extraction protocol, through steps of red blood cell lysis, leukocyte lysis and phenol-chloroform addition for DNA isolation, and isopropanol addition, precipitating it. SNP genotyping was performed using real-time PCR (qPCR). The SNP rs1803250 specific probe used has the following sequence:
GTACCACATAGCCTCTTCAAGTAGC
[T/C]CATGTCCAGTTGTAGAAGTTGTGC. This shows VIC® fluorescence for variant C allele and FAM™ for wild type T allele.

Reactions were performed with addition of sample with extracted DNA, nucleotide-free water and Master Mix, containing Taq Polymerase, nucleotides, magnesium chloride and a buffer solution in an eppendorf, with a final volume of 25 μL. Eppendorfs were inserted in Rotor-Gene-Q automatic thermocycler and the reaction occurred following this parameters: initial denaturation at 95˚C for 10 minutes, followed by 40 cycles for amplification with denaturation at 95˚C for 15 seconds and extension and annealing at 60˚C for 1 minute. Subsequently, graphs generated by the reaction were analyzed.

2.3. Statistical Analysis

The data obtained were analyzed using Hardy-Weinberg Equilibrium Test. The population was considered to be in equilibrium with $p > 0.05$. Genotypic and allelic frequencies correlation with clinicopathological data from patients with breast cancer was performed using Fisher's exact test and Chi-Square test. Values of $p < 0.05$ were considered statistically significant. Analyzes were performed using BioEstat Software version 5.3.

2.4. Ethics Committee

The study was approved by the Research Ethics Committee of the Oncology Research Center with number 3,847,758.

3. Results

3.1. Sample Data Analysis

The data obtained from patients medical records were analyzed and, for this, the ages were categorized every ten years, forming 5 groups: under 40 years old, 40 to 49 years old, 50 to 59 years old, 60 to 69 years old and over 70 years old. In this sense, most patients were aged between 40 and 49 years (31.49%), and the mean age was ±54 years. Table 1 details the samples characteristics from the patient group.

Among the 127 patients analyzed, data were obtained about tumor subtype of breast cancer from 86 individuals, which led to categorization of these women into 5 subgroups: luminal A, luminal B, pure HER2+, luminal-HER2+, and triple

Table 1. Clinical dates from patient group.

Ages	
<40 anos	14 (11.02%)
40 a 49	40 (31.49%)
50 a 59	33 (25.98%)
60 a 69	21 (16.53%)
>70	19 (14.96%)
Tumor subtypes	**86**
Luminal A	15 (17.44%)
Luminal B	18 (20.93%)
HER2+ Pure	20 (23.25%)
Triplo negativo	16 (18.60%)
Luminal-HER2	17 (19.76%)
Tumor size	**48**
T1	2 (4.16%)
T2	22 (45.83%)
T3	21 (43.75%)
T4	3 (6.25%)
Lymph nodes status	**48**
N0	19 (39.58%)
N1	18 (37.50%)
N2	10 (20.83%)
N3	1 (2.08%)

T1: Tumor size ≤ 20 mm; T2; Tumor size between 20 mm and 50 mm; T3: Tumor size > 50 mm; T4: Any tumor size with ulceration or skin nodules; N0: No metastases in near lymph nodes; N1: Metastases in near lymph nodes level I and II, movable; N2: Metastases in near lymph nodes level I and II, fixed, or metastases in mammary chain without lymph node involvement; N3: Near lymph node metastasis level III, or breast chain metastasis with lymph node involvement.

negative. Data were also collected about TMN staging of breast tumors from 48 individuals. This staging system classifies tumors by size (T), presence of spread to nearby lymph nodes (N) and presence or absence of distant metastases (M). The graduation is from T0 to T4, N0 to N3 and M0 to M1, where the higher the number, more advanced the tumor stage is [30]. About category M, the 48 patients did not suffer metastasis (M0) or could not be properly evaluated (MX) according to data from the medical record, so this category was not included in the analysis.

3.2. qPCR Fluorescence Standard of SNP rs1803250 Genotypes

Samples genotyping performed by qPCR detected homozygous wild TT and heterozygous TC genotypes. Both in patient population and control group, no homozygous CC genotype was found. **Figure 1(a)** shows the reaction graph, where fluorescence intensity is shown on X axis and number of cycles on Y axis. The observed standard is TT genotype, because the fluorescence signal was detected only from yellow channel, which is equivalent to FAN™ in the probe used, demonstrating presence only of the T allele. Otherwise, heterozygous TC standard (**Figure 1(b)**), where fluorescence was detected in both yellow channel and green channel, which is equivalent to VIC® in the probe used, demonstrating presence of both alleles.

Figure 2 shows a scatter plot of genotypes according to fluorescence signal detected. Patients are represented in green color and controls in blue. Individuals in which was detected only the T allele are concentrated in fourth quadrant,

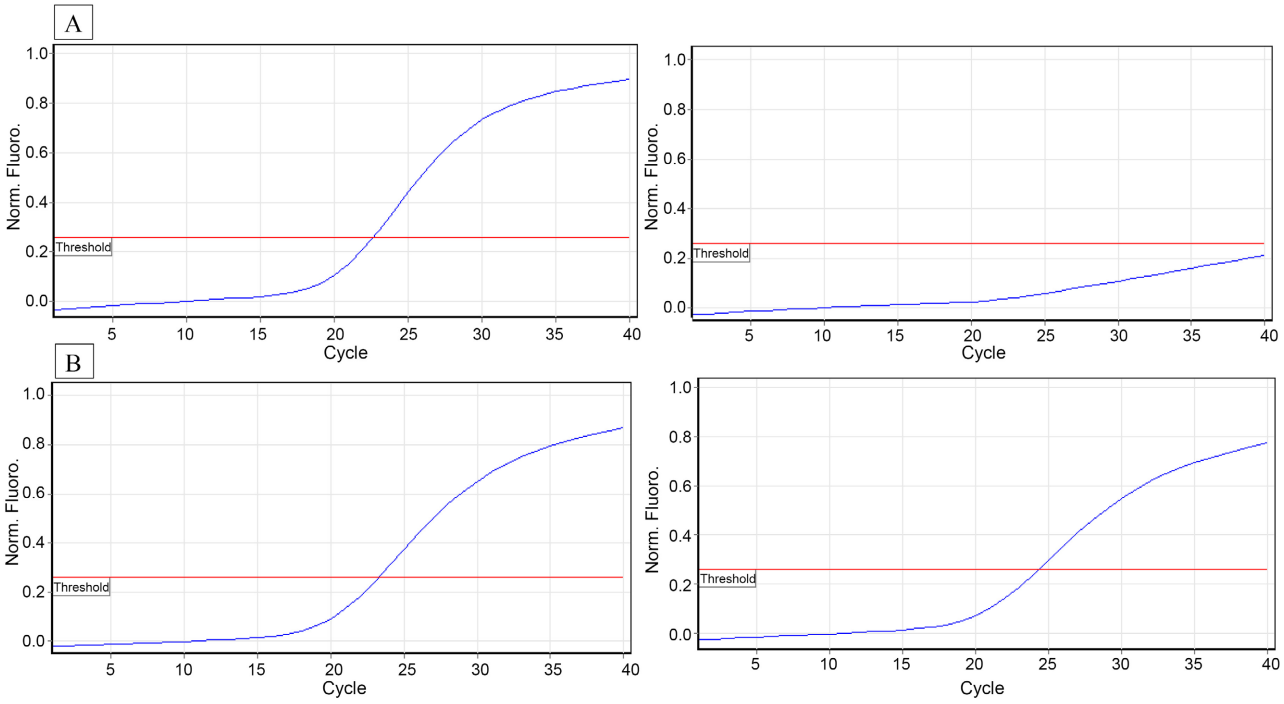

Figure 1. The two genotypes graphs of qPCR reaction of SNP rs1803250 in an individual. (A) TT homozygote standard; (B) TC heterozygote standard.

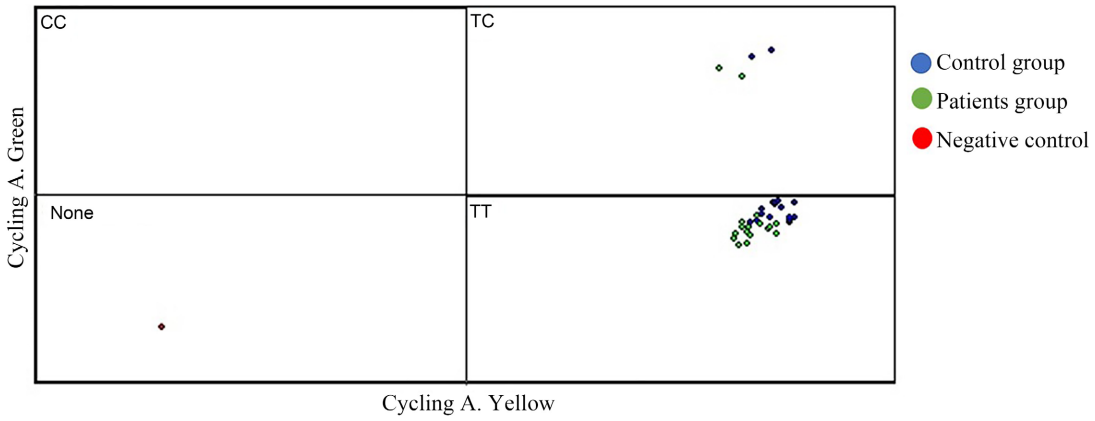

Figure 2. Scatter plot of genotypes according to samples fluorescence detected.

therefore, they are TT homozygotes. Those which detection was of both alleles are concentrated in first quadrant, with heterozygotes being TC. If there were any CC individuals in studied sample, they would be represented in second quadrant. The third quadrant represents samples without any fluorescence detection, where negative control used in reaction is represented in red.

3.3. Hardy-Weinberg's Equilibrium of SNP rs1803250

The Hardy-Weinberg Equilibrium Test demonstrated that the SNP rs1803250 was in equilibrium, with P value of 0.159. In addition, the minor allele frequency in study population was 0.08, which is in agreement with the worldwide allelic frequency according to GenomAD database (0.09) and with the frequency among Amerindians with gastric cancer in region population (0.12). **Table 2** presents the Hardy-Weinberg Equilibrium Test result as well as the global and regional frequencies.

3.4. Allelic and Genotype Frequency of SNP rs1803250

The sample frequency of C allele was 8.33% in control group and 8.66% in patient group. The T allele frequency was 91.67% in control group and 91.33% in patient group. The homozygous TT genotype was observed in 83.33% in control group and in 82.67% in patient group and heterozygous TC genotype was found in 16.66% in control group and 17.32% in patient group. Homozygous CC genotype was not observed in either group. There was no correlation between presence of SNP rs1803250 variant C allele and breast cancer ($P = 0.51$). **Table 3** presents the two groups allelic and genotypic frequencies.

3.5. SNP rs1803250 Association with Tumor Subtypes

The investigation of correlation between SNP 1803250 with breast cancer tumor subtypes was carried out using the Chi-Square Test with data obtained from 86 patients. The C allele frequency among luminal A individuals was 10.00%, among luminal B individuals 11.11%, among pure HER2+ patients 5.00%, among triple negatives 9.37% and among luminal-HER2+ patients was 5.88%. After statistical

Table 2. Hardy-Weinberg's equilibrium test of SNP rs1803250.

Most frequent allele	Less frequent allele	*P* value of HWB* test	MAF*	MAF* GenomAD	MAF* Amerindians
T	C	0.159	0.08	0.09	0.12

*MAF: Minor allele frequency; *HWB: Hardy-Weinberg Equilibrium.

Table 3. Allelic and genotypic frequencies of SNP rs1803250 in patient and control groups.

	TT	TC	CC	*P* value*	T	C	*P* value*
Control (102)	83.33% (85)	16.66% (17)	0.0% (0)	0.51	91.67% (187)	8.33% (17)	0.51
Patients (127)	82.67% (105)	17.32% (22)	0.0% (0)		91.33% (232)	8.66% (22)	

*Fisher's exact test.

analysis, it was observed that there was no significant difference between allelic frequencies between the 5 classifications, thus, the presence of SNP 1803250 presented a similar distribution in all groups ($P = 0.84$).

The frequency of TT genotype in luminal subtype A patients was 80.00%, in luminal subtype B 77.77%, in pure HER2+ patients 80.00%, in triple negatives 81.25%, and in luminal-HER2+ 88.23%. After statistical analysis, it was observed that there was no significant difference between the 5 categories ($P = 0.82$). Thus, statistical data demonstrate that there is no significant correlation between the presence of SNP rs1803250 and classifications of tumor subtypes in breast cancer. Table 4 presents data on allelic and genotypic frequencies of patients according to tumor subtypes.

3.6. SNP rs1803250 Association with Tumor Staging

Correlation between SNP 1803250 and tumor staging was also analyzed using Fisher's exact test in 48 patients, whose data could be obtained. Regarding primary tumor size, C allele frequency among individuals that fit into initial stages (T1 + T2) was 6.25%, an exactly equal frequency value obtained in individuals with larger tumors (T3 + T4), indicating that there was no correlation between SNP and tumor size ($P = 0.66$).

Concerning tumor spread to near lymph nodes, the C allele frequency among patients where spread is null or for few lymph nodes (N0 + N1) was 8.10%. In none of the individuals with disseminations for more lymph nodes (N2 + N3) was detected C allele presence, all being TT genotype. However, this difference was not significant ($P = 0.18$).

Thus, statistical data showed that there was no significant correlation between SNP rs1803250 presence and tumor size or dissemination to near lymph nodes. Table 5 presents data of allelic and genotypic frequencies of patients according to staging T and N.

4. Discussion

This study obtained samples from patients with breast cancer with a mean age of

Table 4. Genotypic and allelic association of SNP rs1803250 with subtypes of breast tumors.

	Genotypes		*P* value*	Alelos		*P* value*
Subtypes	TT	TC		T	C	
Luminal A	12 (80.00%)	3 (20%)		27 (90.00)	3 (10.00%)	
Luminal B	14 (77.77%)	4 (22.22%)		32 (88.88%)	4 (11.11%)	
HER2+ pure	18 (80.00%)	2 (20.00%)	0.82	38 (95.00%)	2 (5.00%)	0.84
TN	13 (81.25%)	3 (18.75%)		29 (90.62%)	3 (9.37%)	
Luminal-HER2+	15 (88.23%)	2 (11.76%)		32 (94.11%)	2 (5.88%)	

*Chi-Square Test.

Table 5. Genotypic and allelic association of rs1803250 SNP with tumor staging.

	Genotypes		*P* value*	Alleles		*P* value*
Staging	TT	TC		T	C	
T1 + T2	21 (87.50%)	3 (12.50%)	0.66	45 (93.75)	3 (6.25%)	0.66
T3 + T4	21 (87.50%)	3 (12.50%)		45 (93.75)	3 (6.25%)	
N0 e N1	31 (83.78%)	6 (16.21%)	0.18	68 (91.89%)	6 (8.10%)	0.18
N2 e N3	11 (100%)	0		22 (100%)	0 (0%)	

T1: Tumor size ≤ 20 mm; T2; Tumor size between 20 mm and 50 mm; T3: Tumor size > 50 mm; T4: Any tumor size with ulceration or skin nodules; N0: No metastases in near lymph nodes; N1: Metastases in near lymph nodes level I and II, movable; N2: Metastases in near lymph nodes level I and II, fixed, or metastases in mammary chain without lymph node involvement; N3: Near lymph node metastasis level III, or breast chain metastasis with lymph node involvement. *Fisher's exact test.

±54 years, with a greater representation of the group of individuals aged between 40 and 49 years (31.49%). Among the data obtained about breast tumor subtype, most patients were classified as HER2+ Pure (23.25%). In addition, most (45.83%) of the patients had tumor size classified as T2, but in the present study, initial and more advanced stages were grouped, wherein 50.00% of the patients were classified in T1 + T2 group and 50.00% in T3 + T4 group. As for lymph node dissemination, the most expressive group was patients whose dissemination did not occur (39.58%). Likewise, this attribute was grouped between individuals with more lymph node metastases and those without or with a few spreads for lymph node metastases, being that last group the most found (77.08%).

Breast cancer is the most incident and considered the leading cause of death among women cancers worldwide [1]. Several factors participate in prognosis determining of these patients. Subtypes based on hormone and HER2 receptors expression, tumor size and lymph node status are some of these factors and studies in some cathepsins have already revealed that SNPs may influence them [31].

Several SNPs like the SNP in pro-CTS D and the SNP at 5' of the CTS O gene can impact some of these factors, as well as participate in the development of

breast cancer [25] [26] [27]. Furthermore, the association between high cathepsin B expression and breast cancer is proven, including its relationship with tumor invasiveness [22] [23] [32] [33]. Therefore, the investigation about SNPs that may be linked to this protein dysregulation during the disease is interesting.

The present study chose to investigate SNP 1803250, due to its exonic location in Cat B, which is able to lead to changes in the mutated protein structure and alter its function [29]. In addition, the SNP 1803250 presents a frequency in individuals from the region where the study sample was collected very similar to the global distribution, meaning that changes in frequency rates among patients could be related to the disease development [28].

The initial statistical performed was Hardy-Weinberg Equilibrium test, which demonstrated that the sample was in equilibrium. Thus, it was possible to proceed with allelic and genotypic frequencies analysis to investigate the association of SNP rs1803250 with breast cancer. No studies have been done before about this SNP, but an association between other CTS B SNPs and cancers has been reported. This is the case of SNP rs13332, a switch from adenine to cytokine, which was significantly associated with risk of developing hepatocarcinoma in a study performed by genotyping 135 patients [34].

Furthermore, Stiblar-martincic & Hajdinjak (2009) investigated the SNP rs12338 in CTS B, which exchanges a cytokine for guanine and leads to an exchange of lysine for valine at position 26 of protein. Through genotyping of 168 patients, they observed that this SNP was associated with a 1.71 times higher risk of developing prostate cancer in patients homozygous for this variant compared to heterozygous and wild-type homozygotes [35]. However, the present study result was different, as there was no association of SNP rs1803250 with breast cancer patients, since the difference in allelic and genotypic frequencies between control and patient groups was not significant.

Analysis of association between the rs1803250 SNP and tumor subtypes was also performed. Knowing that CTS B is significantly more expressed in patients with ER+ breast cancer, the SNP could also be associated with luminal patients, since they are hormone-responsive [14].

In addition, the correlation of a SNP in Cat D with breast cancer that leads to the exchange of alanine for valine has been observed. This study was carried out by mass spectrometry in breast cell lines that showed high expression of HER2 receptor when the SNP was present [36]. In the present study, however, there was no significant association of the rs1803250 SNP and none of the breast tumors subtypes.

Considering that CTS B contributes to disease progression [14], association of SNP rs1803250 with tumor staging of breast cancer patients was also investigated in our study. Stiblar-Martincic & Hajdinjak (2009) demonstrated that the SNP rs12338 present in CTS B was associated with early stages of prostate cancer [35]. Chen *et al.* (2014) also reported that this same SNP was correlated with tumor size in hepatocellular carcinoma. In our study, association of SNP 1803250 was

analyzed, both with regard to tumor size and to dissemination to lymph nodes, but there was no significant association with any degree of tumor staging [34].

No studies have been done previously with SNP 1803250, neither in relation to breast cancer or any other cancer and only two studies on SNPs in CTS B investigating association with malignant tumors have been published [34] [35]. These demonstrated a relationship with the risk of developing prostate cancer and tumor size in hepatocellular carcinoma, indicating that these SNPs can be used as molecular biomarkers for these pathologies.

Since regional population of the study is similar to global population with regard to the mutant allele frequency and that sample size was equivalent to those studies with SNPs in CTS B in cancers, our findings are pioneering and relevant. There was a limitation only in obtaining data about staging factor, where only 48 patients entered in statistical analysis. Still, other studies with more patients can be carried out in order to confirm or not the results obtained.

Furthermore, the lack of association of the SNP rs1803250 with breast cancer apparently indicates that it is not responsible for CTS B high expression during the pathology. Considering the possibility of finding others SNPs in CTS B associated with breast cancer, including the already documented rs12338 and 13332, further investigation may contribute to elaboration of molecular biomarkers panels for this illness.

5. Conclusion

This study points to the fact that there is no correlation between SNP 1803250 and breast cancer development. Likewise, there is no correlation of the SNP with tumor subtypes, nor with tumor staging with regard to tumor size and lymph node dissemination. Despite this, further studies can be carried out with others SNPs in CTS B in relation to breast cancer, in order to investigate the molecular processes involved in high expression of the protein during the course of this disease and to search for biomarkers that can compose a molecular panel for breast cancer.

Acknowledgements

We thank "Núcleo de Oncologia Médica of Hospital Universitário João de Barros Barreto" (NPO-HUJBB) for providing the samples. We also thank "Fundação Centro de Hematologia e Hemoterapia do Pará" (HEMOPA) for help in carrying out the experiments. And we thank "Instituto Nacional de Genética Médica e Populacional"—INAGEMP (CNPq: 573993/2008-4) for providing information.

Author Contributions

Conception/Design: Castro, L.C.D.; Silva, L.C.S.

Data and sample collect: Castro, L.C.D.; Calcagno, D.Q.

DNA extraction and genotyping: Castro, L.C.D.; Amaral, C.E.M.

Data analysis and interpretation: Castro, L.C.D.; Amaral, C.E.M.; Calcagno,

D.Q; Silva, L.C.S.

Manuscript writing: Castro, L.C.D; Silva, L.C.S.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Global Cancer Observatory (2018) Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. https://www.who.int/health-topics/cancer

[2] Instituto Nacional do Câncer (2021) A situação do câncer de mama no Brasil: Síntese de dados dos sistemas de informação. https://www.inca.gov.br/publicacoes/livros/situacao-do-cancer-de-mama-no-brasil-sintese-de-dados-dos-sistemas-de-informacao

[3] Sørlie, T., *et al.* (2001) Gene Expression Patterns of Breast Carcinomas Distinguish Tumor Subclasses with Clinical Implications. *Annals of Oncology*, **30**, 1044-1050.

[4] Coates, A.S., *et al.* (2015) Tailoring Therapies-Improving the Management of Early Breast Cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Annals of Oncology*, **26**, 1533-1546. https://doi.org/10.1093/annonc/mdv259

[5] Yip Ch, R.A. (2014) Estrogen and Progesterone Receptors in Breast Cancer. *Future Oncology*, **10**, 2293-2301. https://doi.org/10.2217/fon.14.110

[6] Begam, A.J., Jubie, S. and Nanjan, M.J. (2017) Estrogen Receptor Agonists/Antagonists in Breast Cancer Therapy: A Critical Review. *Bioorganic Chemistry*, **71**, 257-274. https://doi.org/10.1016/j.bioorg.2017.02.011

[7] Krishnamurti, U. and Silverman, J.F. (2014) HER2 in Breast Cancer: A Review and Update. *Advances in Anatomic Pathology*, **21**, 100-107. https://doi.org/10.1097/PAP.0000000000000015

[8] Alshareeda, A.T., *et al.* (2018) The Effect of Human Placental Chorionic Villi Derived Mesenchymal Stem Cell on Triple-Negative Breast Cancer Hallmarks. *PLOS ONE*, **13**, e0207593. https://doi.org/10.1371/journal.pone.0207593

[9] Prat, A., Adamo, B., Cheang, M.C.U., Anders, C.K., Carey, L.A. and Perou, C.M. (2013) Molecular Characterization of Basal-Like and Non-Basal-Like Triple-Negative Breast Cancer. *Oncologist*, **18**, 123-133. https://doi.org/10.1634/theoncologist.2012-0397

[10] Kunc, M., Biernat, W. and Senkus-Konefka, E. (2018) Estrogen Receptor-Negative Progesterone Receptor-Positive Breast Cancer—"Nobody's Land" or Just an Artifact? *Cancer Treatment Reviews*, **67**, 78-87. https://doi.org/10.1016/j.ctrv.2018.05.005

[11] Kolářová, I., *et al.* (2019) Is There a Benefit of HER2-Positive Breast Cancer Subtype Determination in Clinical Practice? *Klinicka Onkologie*, **32**, 25-30. https://doi.org/10.14735/amko201925

[12] Pernas, S. and Tolaney, S.M. (2020) Targeting HER2 Heterogeneity in Early-Stage Breast Cancer. *Current Opinion in Oncology*, **32**, 545-554. https://doi.org/10.1097/CCO.0000000000000685

[13] Ruibal, A., Herranz, M. and Arias, J.I. (2012) Clinical and Biological Significance of Cathepsin D Levels in Breast Cancer Cytosol in Women over 70 Years. *Biomarkers*

in Cancer, **4**, 1-6. https://doi.org/10.4137/BIC.S9096

[14] Sun, T., Jiang, D., Zhang, L., Su, Q., Mao, W. and Jiang, C. (2015) Expression Profile of Cathepsins Indicates the Potential of Cathepsins B and D as Prognostic Factors in Breast Cancer Patients. *Oncology Letters*, **11**, 575-583. https://doi.org/10.3892/ol.2015.3960

[15] Patel, S., Homaei, A., El-Seedi, H.R. and Akhtar, N. (2018) Cathepsins: Proteases That Are Vital for Survival but Can Also Be Fatal. *Biomedicine & Pharmacotherapy*, **105**, 526-532. https://doi.org/10.1016/j.biopha.2018.05.148

[16] Reif, M.M., MacH, L. and Oostenbrink, C. (2012) Molecular Insight into Propeptide-Protein Interactions in Cathepsins L and O. *Biochemistry*, **51**, 8636-8653. https://doi.org/10.1021/bi300802a

[17] Reiser, J., Adair, B. and Reinheckel, T., (2010) Specialized Roles for Cysteine Cathepsins in Health and Disease. *Journal of Clinical Investigation*, **120**, 3421-3431. https://doi.org/10.1172/JCI42918

[18] Turk, V., *et al.* (2012) Cysteine Cathepsins: From Structure, Function and Regulation to New Frontiers. *Biochimica et Biophysica Acta—Proteins Proteomics*, **1824**, 68-88. https://doi.org/10.1016/j.bbapap.2011.10.002

[19] Tan, G.-J., Peng, Z.-K., Lu, J.-P. and Tang, F.-Q. (2013) Cathepsins Mediate Tumor Metastasis. *World Journal of Biological Chemistry*, **4**, 91-101. https://doi.org/10.4331/wjbc.v4.i4.91

[20] Fonović, M. and Turk, B. (2014) Cysteine Cathepsins and Extracellular Matrix Degradation. *Biochimica et Biophysica Acta—General Subjects*, **1840**, 2560-2570. https://doi.org/10.1016/j.bbagen.2014.03.017

[21] Chen, S., Dong, H., Yang, S. and Guo, H. (2017) Cathepsins in Digestive Cancers. *Oncotarget*, **8**, 41690-41700. https://doi.org/10.18632/oncotarget.16677

[22] Nouh, M.A., *et al.* (2011) Cathepsin B: A Potential Prognostic Marker for Inflammatory Breast Cancer. *Journal of Translational Medicine*, **9**, Article No. 1. https://doi.org/10.1186/1479-5876-9-1

[23] Ibrahim, S.A., *et al.* (2016) Hormonal-Receptor Positive Breast Cancer: IL-6 Augments Invasion and Lymph Node Metastasis via Stimulating Cathepsin B Expression. *Journal of Advanced Research*, **7**, 661-670. https://doi.org/10.1016/j.jare.2016.06.007

[24] Collery, P., Veena, V., Desmaële, D., Harikrishnan, A. and Lakshmi, B. (2021) Effects of Rhenium(I)-Diselenoether and of Its Diselenide Ligand on the Production of Cathepsins B and S by MDA-MB231 Breast Malignant Cells. *Anticancer Research*, **41**, 5997-6001. https://doi.org/10.21873/anticanres.15418

[25] Hato, Y., *et al.* (2016) Prognostic Impact of a Single-Nucleotide Polymorphism near the CTSO Gene in Hormone Receptor-Positive Breast Cancer Patients. *International Journal of Clinical Oncology*, **21**, 539-547. https://doi.org/10.1007/s10147-015-0913-5

[26] Cairns, J., Ingle, J.N., Wickerham, L.D., Weinshilboum, R., Liu, M. and Wang, L. (2017) SNPs near the Cysteine Proteinase Cathepsin O Gene (CTSO) Determine Tamoxifen Sensitivity in ERα-Positive Breast Cancer through Regulation of BRCA1. *PLOS Genetics*, **13**, e1007031. https://doi.org/10.1371/journal.pgen.1007031

[27] Touitou, I., Capony, F., Brouillet, J.P. and Rochefort, H. (1994) Missense Polymorphism (C/T224) in the Human Cathepsin D Pro-Fragment Determined by Polymerase Chain Reaction-Single Strand Conformational Polymorphism Analysis and Possible Consequences in Cancer Cells. *European Journal of Cancer*, **30**, 390-394. https://doi.org/10.1016/0959-8049(94)90261-5

[28] Genome Aggregation Databases (2021). http://www.ensembl.org/Homo_sapiens/Variation/Explore?db=core;r=8:11852165-11853165;v=rs1803250;vdb=variation;vf=165688495

[29] Singh, G., *et al.* (2019) Structural, Functional and Molecular Dynamics Analysis of Cathepsin B Gene SNPs Associated with Tropical Calcific Pancreatitis, a Rare Disease of Tropics. *PeerJ*, **7**, e7425. https://doi.org/10.1101/308932

[30] Instituto Nacional do Câncer (2021) Estadiamento. https://www.inca.gov.br/estadiamento

[31] He, Y., Liu, H., Chen, Q., Shao, Y. and Luo, S. (2019) Relationships between SNPs and Prognosis of Breast Cancer and Pathogenic Mechanism. *Molecular Genetics & Genomic Medicine*, **7**, e871. https://doi.org/10.1002/mgg3.871

[32] Bronger, H., *et al.* (2017) Induction of Cathepsin B by the CXCR3 Chemokines CXCL9 and CXCL10 in Human Breast Cancer Cells. *Oncology Letters*, **13**, 4224-4230. https://doi.org/10.3892/ol.2017.5994

[33] Wang, Z., *et al.* (2018) *In Situ* Proapoptotic Peptide-Generating Rapeseed Protein-Based Nanocomplexes Synergize Chemotherapy for Cathepsin-B Overexpressing Breast Cancer. *ACS Applied Materials & Interfaces*, **10**, 41056-41069. https://doi.org/10.1021/acsami.8b14001

[34] Chen, T.P., Yang, S.F., Lin, C.W., Lee, H.L., Tsai, C.M. and Weng, C.J. (2014) A4383C and C76G SNP in Cathepsin B Is Respectively Associated with the High Risk and Tumor Size of Hepatocarcinoma. *Tumor Biology*, **35**, 11193-11198. https://doi.org/10.1007/s13277-014-2004-z

[35] Štiblar-Martinčič, D. and Hajdinjak, T. (2009) Polymorphism L26V in the Cathepsin B Gene May Be Associated with a Risk of Prostate Cancer and Differentiation. *Journal of International Medical Research*, **37**, 1604-1610. https://doi.org/10.1177/147323000903700539

[36] Whelan, S.A., *et al.* (2012) Mass Spectrometry (LC-MS/MS) Identified Proteomic Biosignatures of Breast Cancer in Proximal Fluid. *Bone*, **11**, 5034-5045. https://doi.org/10.1021/pr300606e