SciencePG
Science Publishing Group

# Prevalence of Mutations of the MED12 and CYP17A1 Genes in Mammary Fibroadenomas in Senegalese Women

**Gueye Rokhaya[1], [\*], Tendeng Jacques Noël[2], Kénémé Bineta[1], Sembene Pape Mbacké[1], [3]**

[1]Department of Animal Biology, Faculty of Science and Technology, University Cheikh Anta Diop of Dakar, Dakar, Senegal

[2]Service of General Surgery, Regional Hospital Center of Saint-Louis, University Gaston Berger of Saint-louis, Saint-Louis, Senegal

[3]Biology of Animal Populations Sahelo-soudanian, Institute of Research and Development (IRD), Campus of Bel-air, Dakar, Senegal

**Email address:**
rokhaya25.gueye@ucad.edu.sn (G. Rokhaya), jacques-noel.tendeng@ugb.edu.sn (T. J. Noël), bineta.keneme@ucad.edu.sn (K. Bineta), mbacke.sembene@ucad.edu.sn (S. P. Mbacké)
[\*]Corresponding author

**Abstract:** Fibroadenoma is the most common benign breast tumor in women under 30 years. This study aimed to contribute to the knowledge of the genetic factors involved in the occurrence and progression of mammary fibroadenomas. *MED12* and *CYP17A1* were sequenced in fibroadenomas and blood in 43 Senegalese women. The Alamut-visual software, which includes the pathogenicity prediction software SIFT, Polyphen2 and MutationTaster, was used to search for mutations. DnaSP version 5.10.01, MEGA version 7.0.14 and Arlequin version 3.5.1.3 were used to determine phylogenetic parameters including indices of genetic variability and diversity and genetic differentiation parameters. A deletion in the poly-A tail of *MED12* was identified in our study population. An alteration of Methionine (M1) was observed on exon 1 of *CYP17A1*. Our results also show that most of the variants found on exon 2 of *MED12* and exon 1 of *CYP17A1* have the probability of causing the appearance of breast fibroadenomas according by the three pathogenicity prediction software. We found 23 new variants on the *MED12* gene and 109 new variants on the *CYP17A1* gene. The amino acid frequency distribution between blood and fibroadenomas shows a statistically significant difference in Glycine, Arginine and Valine for *MED12* and Cysteine, Phenylalanine, Histidine, Asparagine, Arginine, Tryptophan and Tyrosine for *CYP17A1*. In addition the selection test shows that codon 20 of exon 1 of *CYP17A1* which codes for Arginine (p.20Arg) is under positive selection in mammary fibroadenomas. Genetic differentiation parameters show a clear difference between blood and breast fibroadenomas. These results show for the first time the involvement of the *CYP17A1* gene in breast fibroadenomas and confirm the involvement of *MED12*. Codon 20 of exon 1 of *CYP17A1* being under positive selection could be used as a biomarker in breast fibroadenomas.

**Keywords:** Fibroadenoma, Benign, Mammary, Breast, *MED12*, *CYP17A1*, Arginine, Codon 20

## 1. Introduction

The breast can be the site of the development of different types of benign or malignant tumors, including breast fibroadenoma. It is the most common benign tumor pathology. Fibroadenoma is the most common benign breast tumor in women under 30 years. It represents 68% of all breast masses and 44% to 94% of breast biopsies in adolescents [1]. A recent pathological review indexed fibroadenomas as the most common lesion followed by cystosarcomaphylloides and fibrocystic disease [2]. In the United States, each year more than 1.5 million breast biopsies are evaluated [3], most of which are not malignant but have a number of pathological lesions that constitute benign breast disease. A study in Africa, specifically in Nigeria, showed that fibroadenoma accounted for 55.6% of all benign breast disorders, which is much higher than in western populations [4]. In Senegal, a study of benign tumors showed that fibroadenoma is the most common benign

tumors [5].

Although these are benign tumors, there is a large pathological difference between adolescent and adult, pregnant or breastfeeding women and menopausal, black and white skin women. Thus, age, physiological status, reproductive factors and body mass index (BMI) appear to influence the prevalence of fibroadenomas [2, 6, 7].

Despite efforts to study the etiological factors involved in the occurrence of mammary fibroadenoma, understanding of all mechanisms is far from complete. Nevertheless, genetic factors appear to be the most likely determinants. Since estrogens and their metabolites are both inducers and promoters of breast tumors growth [8], genes encoding enzymes involved in estrogen metabolism have been hypothesized to be involved in breast tumors pathogenesis. In this study we contribute to the knowledge of the genetic factors involved in the occurrence and progression of mammary fibroadenomas by investigating mutations of *MED12* and *CYP17A1* genes, assessing the variability of these genes and determining the degree of involvement of these two genes in the occurrence and/or progression of breast fibroadenomas.

# 2. Materials and Methods

## 2.1. Samples Collection

This study involves Senegalese women with breast fibroadenoma. These patients were recruited from the surgical department of the Juliot Curie Institute at Aristide Le Dantec Hospital. Each patient programmed for surgery was interviewed. A duly completed and signed informed consent is required for admission to the study, after which a structured questionnaire is administered. After approval, each patient gave a blood sample forming the control group named B and a breast fibroadenoma tissue sample named BF. Blood samples are placed in EDTA tubes and stored at -4°C. Tumor tissue samples were stored in 96% alcohol. These samples are sent to the BIOPASS (Biology of Sahelo-Sudanian Animal Populations) laboratory of the IRD (Research Institute for Development) for the various stages of genetic analysis. The ethical approval of this study was obtained after review in accordance with the rules laid down by the National Ethics Committee for Health Research (NECHR) of Senegal and in accordance with the procedures established by the University Cheikh Anta DIOP of Dakar (UCAD) for any research involving human participants.

## 2.2. DNA Extraction and Sequencing of MED12 and CYP17A1 Genes

Total DNA was extracted from patients tissues and blood using the DNase Blood and Tissue Kit (Qiagen). The Exon 2 and its flanking regions of *MED12* have been amplified using the forward 5'-GCCCTTTCACCTTGTTTTCCTT-3' and reverse 5'-TGTCCCTATAAGTCTTCCCAACC-3' primers. The exon 1 and the 5' UTR region of *CYP17A1* have been amplified using the forward 5'-CCACAAGGCAAGAGAGAGAGATAACA-3' and reverse 5'-AGGGTAAGCAGCAGCAAGAGAGAGC-3' primers. An electrophoretic migration on 1.5% agarose gel was performed to confirm amplification. Sequencing was performed from 30µl of the PCR product with the forward primer for the *MED12* gene and the reverse primer for the *CYP17A1* gene. Sequencing reactions were performed in a Thermal cycler MJ Research PTC-225 Peltier type with ABI PRISM BigDye TM Terminator Cycle kits. Each sample was sequenced using the primer corresponding. Fluorescent fragments were purified with the BigDye Xterminator purification protocol. The samples were suspended in distilled water and subjected to electrophoresis in ABI 3730xl sequencer (Applied Biosystems).

## 2.3. Molecular Analysis

For the search for mutations, chromatograms obtained with MutationSurveyor version 5.0.1 (®DNA Variant Analysis Software) were submitted to Alamut-Visual version 2.12.0 (©Interactive Biosoftware) [9]. For our study, this software provides us with the location of variants, their predicted protein change, their accession number if they have already been listed in the dbSNP variant database [10], their clinical significance through ClinVar [11] and the prediction of variant pathogenicity through SIFT prediction software [12], Polyphen2 [13] and MutationTaster [14]. Any variation not listed in the variant database is considered as new.

The sequences obtained were cleaned and corrected using BioEdit software version 7.1.9 [15]. Then these sequences are aligned using the ClustalW algorithm [16] to highlight similarities, thus showing the position of insertions, deletions or substitutions. Phylogenetic analysis were performed including the determination of variability index and genetic diversity and genetic differentiation parameters. Genetic variability parameters including the number of sites (N), the number of invariable and variable sites, the number of haplotype (h), the average number of nucleotide difference (k), the number of mutations (Eta), the nature of mutations (%transitions; %transversions) and the estimated Transition/Transversion bias (R), the amino acids frequencies and the codon selection test were obtained through the DnaSP software version 5.10.01 [17] and MEGA version 7.0.14 [18]. For the frequency distribution of amino acids, the RStudio software version 1.0.153 [19] was used, Shapiro Wilk's normality test was performed to see if the data follow a normal distribution. In the case of a normal distribution, the Student t-test is performed for the comparison of averages; otherwise the Wilcoxon test is used. A materiality threshold of 5% has been used. The estimated Transition/Transversion bias (R) was estimated under the kimura 2-parameter model. The codon selection test, obtained with MEGA 7.0.14 was determined for *MED12* exon 2 and *CYP17A1* exon 1. For each codon, estimates of the numbers of synonymous (dS) and nonsynonymous (dN) substitutions were made. These

estimates are produced using the joint Maximum Likelihood (ML) reconstructions of ancestral states under a Muse-Gaut model of codon substitution and Tamura-Nei model of nucleotide substitution. The test statistics dN - dS are used to detecting codons that have under positive selection. A positive value for the test statistic indicates an overabundance of nonsynonymous substitutions. In this case, the probability of rejecting the null hypothesis of neutral evolution (P-value) is calculated. Values of P less than 0.05 are considered significant at a level of 5%. To determine the degree of genetic differentiation between controls and fibroadenomas, the Nei genetic distance obtained with MEGA version 7.0.14 and the factor of genetic differentiation (Fst) obtained with Arlequin software version 3.5.1.3 [20] were extracted. Values of P less than 0.05 are considered significant at a 5% level.

# 3. Results and Discussion

## 3.1. Results

### 3.1.1. Variants of MED12

A total of 26 variants were found on the *MED12* gene, including 7 (26.92%) variants at intron 1, 17 (65.38%) at exon 2 and 2 (7.69%) at intron 2. Only 3 of these variants are known in the variant database (dbSNP), namely the 2 variants found on exon 2: c.131G>A and c.131G>T located at codon 44 and c.204+25G>T located at intron 2. The remaining 23 variants are considered as new variants. All variants at exon 2 have the probability of influencing disease according to the three pathogenicity prediction software programs. (Table 1)

Legend table 1:

*mutation already listed in the variants database (dbSNP)

Calculation of the frequency of mutations in %: (Number of sequences where the mutation is found / Total number of sequences (37)) x 100

*Table 1. Variants of exon 2 of MED12 and its flanking regions.*

| Region | Variant | Number of sequences | Predicted protein change | SIFT | Polyphen2 | Mutation Taster |
|---|---|---|---|---|---|---|
| Intron 1 | c.100-35G>A | 3 (8.11%) | | | | |
| Intron 1 | c.100-33A>G | 7 (18.92%) | | | | |
| Intron 1 | c.100-28del | 8 (21.62%) | | | | |
| Intron 1 | c.100-22A>C | 1 (2.70%) | | | | |
| Intron 1 | c.100-12T>G | 8 (21.62%) | | | | |
| Intron 1 | c.100-5C>A | 1 (2.70%) | | | | |
| Intron 1 | c.100-1G>C | 1 (2.70%) | | | | |
| Exon 2 | c.110C>G | 11 (29.73%) | p.T37R | Deleterious | Probably damaging | Causing disease |
| Exon 2 | c.113C>T | 7 (18.92%) | p.A38V | Tolerated | Probably damaging | Causing disease |
| Exon 2 | c.114C>G | 1 (2.70%) | p.A38A | | | |
| Exon 2 | c.115T>G | 1 (2.70%) | p.L39V | Tolerated | Benign | Causing disease |
| Exon 2 | c.120T>A | 4 (10.81%) | p.N40L | Tolerated | Possible damage | Causing disease |
| Exon 2 | c.128A>G | 6 (16.22%) | p.Q43R | Deleterious | Probably damaging | Causing disease |
| | c.131G>A* | 8 (21.62%) | p.G44D | Deleterious | Probably damaging | Causing disease |
| Exon 2 | c.131G>T* | 5 (13.51%) | p.G44V | Deleterious | Probably damaging | Causing disease |
| Exon 2 | c.143A>C | 1 (2.70%) | p.Q48P | Deleterious | Possible damage | Causing disease |
| | c.143A>G | 1 (2.70%) | p.Q48R | | Probably damaging | Causing disease |
| Exon 2 | c.147T>C | 1 (2.70%) | p.P49P | | | |
| Exon 2 | c.148G>T | 1 (2.70%) | p.A50S | Deleterious | Probably damaging | Causing disease |
| Exon 2 | c.160G>C | 1 (2.70%) | p.D54H | Deleterious | Probably damaging | Causing disease |
| Exon 2 | c.173G>C | 2 (5.41%) | p.S58T | Tolerated | Probably damaging | Causing disease |
| | c.173G>T | 1 (2.70%) | p.S58I | Deleterious | Possible damage | Causing disease |
| Exon 2 | c.180G<C | 1 (2.70%) | p.K60N | Tolerated | Benign | Causing disease |
| Exon 2 | c.188G>T | 1 (2.70%) | p.S63I | Deleterious | Benign | Causing disease |
| Intron 2 | c.204+25G>T* | 1 (2.70%) | | | | |
| Intron 2 | c.204+47G>T | 1 (2.70%) | | | | |

### 3.1.2. Variant of CYP17A1

A total of 131 variants have been found on this gene, including 26 (19.85%) in the 5' UTR region and 105 (80.15%) in exon 1, 22 of these variants have already been listed in the variants database (dbSNP), 8 of which are not yet classified (meaning not known). The remaining 109 variants are considered as new variants. 7 of the variants already listed have had clinical significance, 5 are classified as benign (c.-34T<A, c.-34T>C, c.138C>T, c.195G>A and c.195G>T) and

2 as pathogenic (c.51G>A and c.81C>A). On exon 1, 69 (65.71%) of the variants were predicted to influence pathogenicity. Mutations affecting the initiator codon p.Met1 are also found. (Table 2)

Legend table 2:

*mutation already listed in the variants database (dbSNP)

Calculation of the frequency of mutations in %: (Number of sequences where the mutation is found / Total number of sequences (26)) x 100

***Table 2.*** *Variants of exon 1 and 5'UTR region of CYP17A1.*

| Region | Variant | Number of sequences | Predicted protein change | SIFT | Polyphen2 | Mutation Taster |
|---|---|---|---|---|---|---|
| (5'UTR) | c.-81A>T | 1 (3.85%) | | | | |
| (5'UTR) | c.-76G>T | 2 (7.69%) | | | | |
| (5'UTR) | c.-60G>A | 22 (84.62%) | | | | |
| (5'UTR) | c.-58G>A* | 26 (100%) | | | | |
| (5'UTR) | c.-45T>A | 5 (19.23%) | | | | |
| | c.-45T>G | 1 (3.85%) | | | | |
| (5'UTR) | c.-43C>A | 22 (84.62%) | | | | |
| | c.-43C>T | 2 (7.69%) | | | | |
| (5'UTR) | c.-40C>A | 5 (19.23%) | | | | |
| | c.-40C>G | 17 (65.38%) | | | | |
| (5'UTR) | c.-34T>A* | 2 (7.69%) | | | | |
| | c.-34T>C* | 1 (3.85%) | | | | |
| (5'UTR) | c.-32C>A | 2 (7.69%) | | | | |
| | c.-32C>G | 4 (15.38%) | | | | |
| | c.-15C>T* | 2 (7.69%) | | | | |
| (5'UTR) | c.-14G>A* | 12 (46.15%) | | | | |
| | c.-14G>C | 14 (53.85%) | | | | |
| (5'UTR) | c.-12C>G | 1 (3.85%) | | | | |
| (5'UTR) | c.-10C>G | 1 (3.85%) | | | | |
| (5'UTR) | c.-7A>C | 19 (73.08%) | | | | |
| | c.-7A>G* | 3 (11.54%) | | | | |
| (5'UTR) | c.-5C>G | 1 (3.85%) | | | | |
| (5'UTR) | c.-4C>A | 20 (76.92%) | | | | |
| | c.-4C>G | 5 (19.23%) | | | | |
| (5'UTR) | c.-2C>A | 20 (76.92%) | | | | |
| | c.-2C>G | 5 (19.23%) | | | | |
| Exon 1 | c.1A>G | 6 (23.08%) | | | | |
| | c.1A>T | 1 (3.85%) | Altered M1 | | | |
| Exon 1 | c.3G>A* | 17 (65.38%) | | | | |
| | c.3G>C | 1 (3.85%) | Altered M1 | | | |
| Exon 1 | c.8A>C | 25 (96.15) | p.E3A | Tolerated | Benign | Polymorphism |
| | c.8A>T | 1 (3.85%) | p.E3V | Tolerated | Benign | Polymorphism |
| Exon 1 | c.14T>A | 16 (61.54%) | p.V5E | Deleterious | Possible damage | Polymorphism |
| | c.14T>G* | 1 (3.85%) | p.V5G | Deleterious | Possible damage | Polymorphism |
| Exon 1 | c.24G>A | 25 (96.15) | p.L8L | | | |
| Exon 1 | c.32C>T* | 5 (19.23%) | p.T11I | Tolerated | Benign | Polymorphism |
| Exon 1 | c.35T>A | 16 (61.54%) | p.L12Q | Tolerated | Probably damaging | Polymorphism |
| | c.35T>C | 3 (11.54%) | p.L12P | Tolerated | Probably damaging | Polymorphism |
| Exon 1 | c.37G>A | 19 (73.08%) | p.A13T | Tolerated | Benign | Polymorphism |
| | c.37G>C | 1 (3.85%) | p.A13P | Tolerated | Possible damage | Polymorphism |
| Exon 1 | c.41A>G | 4 (15.38%) | p.Y14C | Tolerated | Possible damage | Polymorphism |
| | c.41A>T | 3 (11.54%) | p.Y14F | Tolerated | Probably damaging | Polymorphism |
| Exon 1 | c.42T>A | 18 (69.23%) | p.Y14X | | | |
| | c.42T>C | 1 (3.85%) | p.Y14Y | | | |
| Exon 1 | c.51G>A* | 7 (26.92%) | p.W17X | | | |
| Exon 1 | c.52C>T | 16 (61.54%) | p.P18S | Tolerated | Benign | Causing disease |
| Exon 1 | c.55A>C | 1 (3.85%) | p.K19Q | Tolerated | Probably damaging | Polymorphism |
| | c.55A>G | 13 (50%) | p.K19E | Tolerated | Possible damage | Polymorphism |
| Exon 1 | c.56A>G | 13 (50%) | p.K19R | Tolerated | Benign | Polymorphism |
| | c.56A>T | 5 (19.23%) | p.K19M | Deleterious | Probably damaging | Polymorphism |
| Exon 1 | c.57G>A | 9 (34.62%) | p.K19K | Tolerated | Probably damaging | Polymorphism |
| | c.57G>C | 1 (3.85%) | p.K19N | | | |
| Exon 1 | c.58A>C | 2 (7.69%) | p.R20R | Tolerated | Benign | Polymorphism |
| | c.58A>G | 16 (61.54%) | p.R20G | | | |
| Exon 1 | c.59G>A | 10 (38.46%) | p.R20K | Tolerated | Benign | Polymorphism |
| | c.59G>T | 3 (11.54%) | p.R20I | Tolerated | Benign | Polymorphism |
| Exon 1 | c.61A>G* | 4 (15.38%) | p.R21G | Tolerated | Benign | Polymorphism |
| | c.61A>T | 6 (23.08%) | p.R21W | Tolerated | Probably damaging | Polymorphism |
| Exon 1 | c.62G>A* | 17 (65.38%) | p.R21K | Tolerated | Benign | Polymorphism |
| Exon 1 | c.64T>A | 4 (15.38%) | p.C22R | Tolerated | Benign | Polymorphism |
| | c.64T>C* | 21 (80.77%) | p.C22S | Tolerated | Benign | Polymorphism |
| Exon 1 | c.66C>G* | 2 (7.69%) | p.C22W | Deleterious | Possible damage | Polymorphism |
| Exon 1 | c.68C>T* | 25 (96.15) | p.P23L | Deleterious | Benign | Polymorphism |
| Exon 1 | c.72T>A | 4 (15.38%) | p.G24G | | | |
| | c.72T>G | 5 (19.23%) | p.G24G | | | |
| Exon 1 | c.77A>G | 3 (11.54%) | p.K26R | Tolerated | Probably damaging | Causing disease |

| Region | Variant | Number of sequences | Predicted protein change | SIFT | Polyphen2 | Mutation Taster |
|---|---|---|---|---|---|---|
| | c.77A>T | 14 (53.85) | p.K26M | Deleterious | Probably damaging | Causing disease |
| Exon 1 | c.80A>C* | 2 (7.69%) | p.Y27S | Tolerated | Benign | Polymorphism |
| | c.80A>T | 18 (69.23%) | p.Y27F | Tolerated | Benign | Polymorphism |
| Exon 1 | c.81C>A* | 5 (19.23%) | p.Y27X | | | |
| Exon 1 | c.96G>A | 2 (7.69%) | p.L32L | | | |
| | c.96G>T | 6 (23.08%) | p.L32L | | | |
| Exon 1 | c.100C>A | 25 (96.15%) | p.L34M | Deleterious | Probably damaging | Causing disease |
| | c.100C>T | 1 (3.85%) | p.L34L | | | |
| Exon 1 | c.108G>A* | 3 (11.54%) | p.L36L | | | |
| | c.108G>C | 23 (88.46%) | p.L36L | | | |
| Exon 1 | c.113G>A | 16 (61.54%) | p.G38D | Deleterious | Probably damaging | Causing disease |
| | c.113G>T | 1 (3.85%) | p.G38V | Deleterious | Probably damaging | Causing disease |
| Exon 1 | c.115A>C | 3 (11.54%) | p.S39R | Deleterious | Probably damaging | Causing disease |
| | c.115A>G | 8 (30.77%) | p.S39G | Deleterious | Probably damaging | Causing disease |
| Exon 1 | c.116G>A | 5 (19.23%) | p.S39N | Tolerated | Probably damaging | Causing disease |
| | c.116G>T | 1 (3.85%) | p.S39I | Deleterious | Probably damaging | Causing disease |
| Exon 1 | c.129C>A | 5 (19.23%) | p.L43L | | | |
| | c.129C>G | 18 (69.23%) | p.L43L | | | |
| Exon 1 | c.132C>A | 16 (61.54%) | p.P44P | | | |
| | c.132C>G | 4 (15.38%) | p.P44P | | | |
| Exon 1 | c.134G>A | 17 (65.38%) | p.R45K | Tolerated | Benign | Polymorphism |
| | c.134G>C | 1 (3.85%) | p.R45T | Tolerated | Possible damage | Polymorphism |
| Exon 1 | c.138C>A | 3 (11.54%) | p.H46Q | Tolerated | Possible damage | Polymorphism |
| | c.138C>T* | 1 (3.85%) | p.H46H | | | |
| Exon 1 | c.158T>A | 15 (57.69%) | p.F53Y | Tolerated | Probably damaging | Polymorphism |
| | c.158T>G | 4 (15.38%) | p.F53C | Tolerated | Probably damaging | Causing disease |
| Exon 1 | c.163A>G | 18 (69.23%) | p.K55E | Tolerated | Benign | Causing disease |
| | c.163A>T | 2 (7.69%) | p.K55X | | | |
| Exon 1 | c.165G>A | 3 (11.54%) | p.K55K | Tolerated | Possible damage | Causing disease |
| | c.165G>T | 3 (11.54%) | p.K55N | | | |
| Exon 1 | c.171G>A | 17 (65.38%) | p.Q57Q | Deleterious | Probably damaging | Causing disease |
| | c.171G>T | 1 (3.85%) | p.Q57H | | | |
| Exon 1 | c.178T>A | 17 (65.38%) | p.Y60N | Deleterious | Probably damaging | Causing disease |
| | c.178T>C | 5 (19.23%) | p.Y60H | Deleterious | Probably damaging | Causing disease |
| Exon 1 | c.180T>C* | 3 (11.54%) | p.Y60Y | | | |
| Exon 1 | c.185C>T | 4 (15.38%) | p.P62L | Tolerated | Probably damaging | Causing disease |
| Exon 1 | c.187A>G | 26 (100%) | p.I63V | Tolerated | Possible damage | Polymorphism |
| Exon 1 | c.188T>C | 25 (96.15%) | p.I63T | Deleterious | Possible damage | Polymorphism |
| | c.188T>G | 1 (3.85%) | p.I63S | | Probably damaging | Polymorphism |
| Exon 1 | c.189C>G | 3 (11.54%) | p.I63M | Deleterious | Probably damaging | Polymorphism |
| Exon 1 | c.190T>C | 4 (15.38%) | p.Y64H | Deleterious | Probably damaging | Causing disease |
| | c.190T>G | 19 (73.08%) | p.Y64D | Deleterious | Probably damaging | Causing disease |
| Exon 1 | c.191A>G | 1 (3.85%) | p.Y64C | Deleterious | Probably damaging | Causing disease |
| | c.191A>C | 25 (96.15%) | p.Y64S | | Probably damaging | Causing disease |
| Exon 1 | c.193T>A | 1 (3.85%) | p.S65T | Tolerated | Probably damaging | Causing disease |
| | c.193T>C | 16 (61.54%) | p.S65P | Tolerated | Probably damaging | Causing disease |
| Exon 1 | c.195G>A* | 25 (96.15%) | p.S65S | | | |
| | c.195G>T* | 1 (3.85%) | p.S65S | | | |
| Exon 1 | c.201T>A | 10 (38.46%) | p.R67R | | | |
| | c.201T>C | 1 (3.85%) | p.R67R | | | |
| Exon 1 | c.207C>A | 14 (53.85) | p.G69G | | | |
| | c.207C>G | 8 (30.77%) | p.G69G | | | |
| Exon 1 | c.209C>A | 3 (11.54%) | p.T70N | Tolerated | Possible damage | Polymorphism |
| | c.209C>G | 4 (15.38%) | p.T70S | Tolerated | Benign | Polymorphism |
| Exon 1 | c.211A>T | 7 (26.92%) | p.K71X | | | |
| Exon 1 | c.214A>C | 8 (30.77%) | p.T72P | Tolerated | Benign | Polymorphism |
| | c.214A>G | 2 (7.69%) | p.T72A | Tolerated | Benign | Polymorphism |
| Exon 1 | c.220G>A | 16 (61.54%) | p.V74M | Deleterious | Probably damaging | Causing disease |
| Exon 1 | c.224T>A | 8 (30.77%) | p.I75N | Tolerated | Probably damaging | Causing disease |
| | c.224T>G | 17 (65.38%) | p.I75S | Tolerated | Probably damaging | Causing disease |
| Exon 1 | c.226G>A | 20 (76.92%) | p.V76I | Tolerated | Benign | Polymorphism |
| | c.226G>C | 5 (19.23%) | p.V76L | Tolerated | Probably damaging | Polymorphism |
| Exon 1 | c.229G>A | 1 (3.85%) | p.G77S | Tolerated | Probably damaging | Causing disease |
| Exon 1 | c.231C>A | 3 (11.54%) | p.G77G | | | |
| | c.231C>T | 20 (76.92%) | p.G77G | | | |

### 3.1.3. Index de Variability and Genetic Diversity

The relative values of the variability and genetic diversity of *MED12* and *CYP17A1* in fibroadenomas are shown in Table 3. These parameters indicate that there is a slight variability in the controls, especially for *CYP17A1*. The nature of the mutations shows that the transversions are much higher than the transitions. The estimated Transitions/Transversions bias (R) confirms the superiority of transversions. The polymorphism analysis revealed a high value of haplotypic diversity (hd) and a low value of nucleotide diversity (Pi). The average number of nucleotide differences (k) is very high for *CYP17A1*.

**Table 3.** *Index of variability and genetic diversity.*

| Genes | MED12 | | CYP17A1 | |
|---|---|---|---|---|
| Parameter | B | BF | B | BF |
| Number of sequences | 11 | 26 | 10 | 24 |
| Number of sites, *N* | 226 | 226 | 609 | 609 |
| Invariable sites | 225 | 204 | 579 | 534 |
| Variable sites | 1 | 22 | 30 | 75 |
| Informative sites | 0 | 9 | 30 | 63 |
| Number of haplotypes, *h* | 2 | 23 | 2 | 23 |
| Total number of mutations, *Eta* | 1 | 24 | 30 | 119 |
| Transitions (%) | 27.1 | | 42.8 | |
| Transversions (%) | 72.9 | | 57.2 | |
| estimated Transition/Transversion bias *R* | 0.41 | | 0.67 | |
| Haplotypic diversity, *hd* | 0.988+-0.016 | | 0.996+-0.013 | |
| Nucleotide diversity, *Pi* | 0.0199+-0.0023 | | 0.0469+- 0.0034 | |
| Average number of nucleotide differences, *k* | 4.483 | | 28.551 | |

The amino acids frequency of the *MED12* and *CYP17A1* genes between controls and fibroadenomas is shown in Table 4. For *MED12*, Glycine, Arginine and Valine are the only amino acids with a statistically significant difference. Isoleucine is absent in controls and present in tumor tissue. For *CYP17A1*, Cysteine, Phenylalanine, Histidine, Asparagine, Arginine, Tryptophan and Tyrosine are the only amino acids with a statistically significant difference.

Legend table 4:

P-value < 0.05 → Statistically significant difference (* intensity of significance)

**Table 4.** *Amino acid frequencies.*

| Genes | MED12 | | | CYP17A1 | | |
|---|---|---|---|---|---|---|
| Amino acids | B | BF | P-value | B | BF | P-value |
| Alanine | 11.428 | 10.549 | 0.0635 | 7.382 | 6.592 | 0.05455 |
| Cysteine | 0 | 0 | - | 1.342 | 0.279 | 0.0000289*** |
| Aspartic acid | 5.714 | 6.483 | 0.09064 | 1.342 | 1.005 | 0.08921 |
| Glutamic acid | 5.714 | 5.714 | - | 2.013 | 2.067 | 0.8595 |
| Phenylalanine | 5.714 | 5.714 | - | 5.369 | 3.743 | 0.00008736*** |
| Glycine | 8.571 | 7.142 | 0.004342** | 8.724 | 8.603 | 0.4007 |
| Histidine | 2.857 | 2.967 | 0.5543 | 5.369 | 4.078 | 0.000118*** |
| Isoleucine | 0 | 0.219 | 0.3732 | 2.684 | 2.178 | 0.06755 |
| Lysine | 8.571 | 8.901 | 0.3702 | 6.711 | 7.877 | 0.1886 |
| Leucine | 5.714 | 5.604 | 0.5543 | 18.120 | 18.659 | 0.7152 |
| Methionine | 0 | 0 | - | 6.040 | 5.810 | 0.6355 |
| Asparagine | 14.285 | 13.956 | 0.3702 | 3.355 | 4.413 | 0.005228** |
| Proline | 5.714 | 5.824 | 0.5543 | 9.395 | 8.603 | 0.07305 |
| Glutamine | 5.454 | 4.945 | 0.2442 | 2.013 | 2.402 | 0.8613 |
| Arginine | 0.259 | 1.868 | 0.002144** | 2.013 | 3.743 | 0.01367* |
| Serine | 8.571 | 8.241 | 0.5174 | 5.369 | 6.312 | 0.3493 |
| Threonine | 2.857 | 1.868 | 0.05184 | 5.369 | 5.307 | 0.7147 |
| Valine | 8.571 | 10 | 0.007641** | 4.026 | 4.357 | 0.9385 |
| Tryptophan | 0 | 0 | - | 3.355 | 2.458 | 0.0002559*** |
| Tyrosine | 0 | 0 | - | 0 | 1.508 | 0.0001604*** |

The codon selection test showed a heterogeneous distribution of mutations in the *MED12* gene. However, for the *CYP17A1* gene, most of the alterations affect codon 20 (p.20Arg) which is under positive selection (p=0.0076). (Table 5)

Legend table 5:

P-value < 0.05 → *significant

**Table 5.** *Selection test CYP17A1.*

| Codon | Triplet | dS | dN | dN-dS | P-value |
|---|---|---|---|---|---|
| 1 | ATG | 5.4683 | 3.6968 | -1.7714 | 0.8355 |
| 2 | TGG | 0 | 0 | 0 | N/A |
| 3 | GAG | 0 | 0.9984 | 0.9984 | 0.4471 |

| Codon | Triplet | dS | dN | dN-dS | P-value |
|---|---|---|---|---|---|
| 4 | CTC | 0 | 0 | 0 | N/A |
| 5 | GTG | 0 | 2.4345 | 2.4345 | 0.1710 |
| 6 | GCT | 0 | 0 | 0 | N/A |
| 7 | CTC | 0 | 0 | 0 | N/A |
| 8 | TTG | 1.0671 | 0 | -1.0671 | 1 |
| 9 | CTG | 0 | 0 | 0 | N/A |
| 10 | CTT | 0 | 0 | 0 | N/A |
| 11 | ACC | 0 | 1.4930 | 1.4930 | 0.3004 |
| 12 | CTA | 0 | 2.7996 | 2.7996 | 0.0969 |
| 13 | GCT | 0 | 2.5 | 2.5 | 0.1316 |
| 14 | TTG | 0 | 0 | 0 | N/A |
| 15 | TTT | 0 | 0 | 0 | N/A |
| 16 | CCC | 0 | 3 | 3 | 0.0877 |
| 17 | AAG | 4.9956 | 6.5826 | 1.5869 | 0.4273 |
| 18 | AGA (p.20Arg) | 0 | 6.9605 | 6.9605 | 0.0076* |
| 19 | TGC | 2.0780 | 1.4817 | -0.596 | 0.8063 |
| 20 | CCT | 0 | 0.5 | 0.5 | 0.6666 |
| 21 | GGT | 6 | 0 | -6 | 1 |
| 22 | GCC | 0 | 0 | 0 | N/A |
| 23 | AAG | 0 | 2.1866 | 2.1866 | 0.3220 |
| 24 | CCC | 0 | 0 | 0 | N/A |
| 25 | AAG | 0 | 0 | 0 | N/A |
| 26 | AGC | 0 | 0 | 0 | N/A |
| 27 | CTC | 0 | 0 | 0 | N/A |
| 28 | CTG | 2.9932 | 0 | -2.9932 | 1 |
| 29 | TCC | 0 | 0 | 0 | N/A |
| 30 | CTG | 0 | 0.6790 | 0.6790 | 0.9657 |
| 31 | CCC | 0 | 0 | 0 | N/A |
| 32 | CTG | 2.8358 | 0 | -2.8358 | 1 |
| 33 | GTG | 0 | 0 | 0 | N/A |
| 34 | GGC | 0 | 2.1633 | 2.1633 | 0.2714 |
| 35 | AGC | 0 | 3.3626 | 3.3626 | 0.1564 |
| 36 | CTG | 0 | 0 | 0 | N/A |
| 37 | CCA | 0 | 0 | 0 | N/A |
| 38 | TTC | 0 | 0 | 0 | N/A |
| 39 | CTC | 3.7324 | 0 | -3.7324 | 1 |
| 40 | CCC | 6 | 0 | -6 | 1 |
| 41 | AGA | 0 | 2.8315 | 2.8315 | 0.1849 |
| 42 | CAC | 2.4266 | 0.7866 | -1.6400 | 0.9470 |
| 43 | GGC | 0 | 0 | 0 | N/A |
| 44 | CAT | 0 | 0 | 0 | N/A |
| 45 | ATG | 0 | 0 | 0 | N/A |
| 46 | CAT | 0 | 0 | 0 | N/A |
| 47 | AAC | 0 | 0 | 0 | N/A |
| 48 | AAC | 0 | 0 | 0 | N/A |
| 49 | TTC | 0 | 2.7146 | 2.7146 | 0.3795 |
| 50 | CTG | 0 | 0 | 0 | N/A |
| 51 | CAG | 6.1798 | 0.5068 | -5.6730 | 0.9990 |
| 52 | AAA | 0 | 0 | 0 | N/A |
| 53 | AAA | 0 | 0 | 0 | N/A |
| 54 | TAT | 6.1456 | 2.8420 | -3.3036 | 0.9221 |
| 55 | GGC | 0 | 0 | 0 | N/A |
| 56 | CCC | 0 | 2 | 2 | 0.1975 |
| 57 | ATC | 3.0137 | 1.4965 | -1.5172 | 0.9013 |
| 58 | TAT | 0 | 3.5 | 3.5 | 0.0616 |
| 59 | TCG | 2 | 4.5706 | 2.5706 | 0.2332 |
| 60 | GTT | 0 | 0 | 0 | N/A |
| 61 | CGT | 4.3027 | 0 | -4.3027 | 1 |
| 62 | ATG | 0 | 0 | 0 | N/A |
| 63 | GGC | 8 | 0 | -8 | 1 |
| 64 | ACC | 0 | 2.2939 | 2.2939 | 0.2024 |
| 65 | ACT | 0 | 2.5 | 2.5 | 0.1316 |
| 66 | ACA | 0 | 0 | 0 | N/A |
| 67 | GTG | 0 | 2.2755 | 2.2755 | 0.4608 |
| 68 | ATT | 0 | 2.6543 | 2.6543 | 0.4056 |
| 69 | GTC | 0 | 1.8382 | 1.8382 | 0.2767 |
| 70 | GGC | 4.0598 | 0.4974 | -3.5624 | 0.9961 |

Analysis of the intragroup genetic distance shows that there is almost no difference between controls for the *MED12* gene. For the *CYP17A1* gene there is a slight difference between the controls. Tumors tissues, on the other hand, show genetic differentiation. In addition, a statistically significant differentiation is noted between controls and tissues. (Table 6)

Legend table 6:

*significant → P-value < 0.05

***Table 6.** Genetic distance and Factor of genetic differentiation (Fst).*

| Genes | MED12 | | CYP17A1 | |
|---|---|---|---|---|
| genetic distance of Nei | | | | |
| Groups | intragroup | intergroup | intragroup | intergroup |
| B | 0.001 | 0.004 | 0.028 | 0.006 |
| BF | 0.020 | | 0.049 | |
| Factor of genetic differentiation (Fst) | | | | |
| Groups | intergroup | P-value | Intergroup | P-value |
| B | 0.13561 | 0.00099* | 0.15815 | 0.00000* |
| BF | | | | |

## 3.2. Discussion

The overall objective of this study is to contribute to research on the prevalence of mutations of the *MED12* and *CYP17A1* genes in mammary fibroadenomas in Senegalese women. The choice of genes to be studied is based on these two genes involved in estrogen synthesis. Thus, we looked for mutations and determined the degree of involvement of these two genes in the occurrence of fibroadenomas. The variability, genetic diversity and genetic differentiation of these genes were studied.

The sequencing of *MED12* which codes for a protein involved in transcription regulation showed 26 variants including 25 (96.15%) substitutions and 1 (3.85%) deletion. This deletion is located at the poly-A tail of intron 1: c.100-28del. This result is consistent with that of Kénémé *and al.* [21] who showed for the first time a deletion located at the poly-A tail of intron 1. This mutation could influence the alternative splicing mechanism of the intronic regions resulting in an aberrant MED12 protein. Thus, further studies will be needed to determine the exact function of this region in the synthesis of the MED12 protein. Our results also show that the variants found on *MED12* exon 2 are predicted to be involved in the pathogenicity of breast fibroadenoma. The variants found on intron 1 are the most frequent because they are found in several patients: c.100-33A>G (18.92%), c.100-28del (21.62%) and c.100-12T>G (21.62%). Variants c.131G>A (21.62%) and c. 131 G>T (13.51%) are already listed in the variant database. They have been found in uterine fibroids [22, 23]. This follows the assumption that fibroadenomas and uterine fibroids may share a common genetic etiology. Moreover, the mutations of codon 44, which is the most preserved codon of exon 2, are found in these two pathologies, showing the important role they play in their occurrence. A study by Bourbon [24], involving different species, showed that codon 44 is the most preserved codon of the *MED12* gene, which stipulates that this codon would play an important role in the normal functioning of the protein. Thus, the false-sense mutations observed on this codon 44 in particular may render the translated protein non-functional or result in a gain of function in tumor tissue, indicating the specific importance of this amino acid for *MED12* function. Moreover, according to the work of Turunen *and al* [25], the binding domain of cyclin C resides in the N-terminal region encoded by exons 1 and 2 of the *MED12* gene and codon 44 would play a role in this adherence.

The sequencing of *CYP17A1* involved in estrogen metabolism showed 131 variants all being substitutions. 7 variants have a clinical scope including 2 located in the 5' UTR region (c.-34T>A and c.-34T>C) which are benign (ClinVar) mutations. Indeed, the c.-34T>C mutation located in the 5' promoter region at 34 bp upstream of the translation initiation site and 27 bp downstream of the transcription initiation site has been found in several pathologies. This mutation creates a new CCACC box site and therefore an additional promoter. Carey *and al* [26] first identified this T27C SNP in the 5' UTR of *CYP17α* and hypothesized that the C allele could over-regulate gene expression by first increasing serum hormones including androstenedione and estradiol (E2). This polymorphism is common: the CC genotype is present in 11-19% of white North American women and 6-16% of African Americans [27]. Several studies have hypothesized that the C allele of CYP17α may be a marker for increased steroidogenesis [28, 29]. Sun *and al.* [30] concluded that rs743572 (-34T>C) may increase the risk of breast cancer in postmenopausal women. Kaur *and al.* [31] showed that polymorphism -34T>C is associated with polycystic ovary syndrome in North India. This syndrome affects women of childbearing age and causes menstrual disorders, infertility and overproduction of androgen by the ovaries. According to the three pathogenicity prediction software programs, almost all exon 1 mutations have an impact on the appearance of fibroadenoma except for certain variations that do not induce a change in amino acids. The variant c.51G>A (26.92%) which corresponds to a premature Tryptophan stop codon in codon 17 (p.W17X) has been identified in ClinVar and is considered pathogenic. This same variant was found in the study by Suzuki *and al.* [32] and was reported to induce the genetic disorder of cytochrome P450c17 resulting in a 17 alpha-

hydroxylase/17,20-lyase deficiency. 81C>A (19.23%) is also a pathogenic variant also found in the study of Müssig *and al.* [33] causing a deficiency in 17 alpha-hydroxylase / 17,20-lyase. The variants c.195G>A (25/6.15%) and c.195G>T (1/3.85%) are recognized as benign in ClinVar and are identified in congenital adrenal hyperplasia. The mutations c.-58G>A, c.8A>C, c.24G>A, c.68C>T, c.100C>A, c.187A>G, c.188T>C, c.191A>C and c.195G>A, are also found in over 90% of our population demonstrating the altered mechanism of the *CYP17A1* gene in mammary fibroadenomas. Indeed, *CYP17A1* possesses both 17-alpha-hydroxylase activities necessary for the production of glucocorticoids and 17,20-lyase, a key activity of steroidogenesis producing progestins, androgens and estrogens. Mutations in this gene are associated with an isolated steroid-17-alpha-hydroxylase deficiency and 17-alpha-hydroxylase/17,20-lyase deficiency. This deficiency in enzyme activity is thought to induce an inhibition of the catabolism of progesterone to 17hydroxy-progesterone and thus an overproduction of the latter in tumor tissues. Progesterone has been shown to play an important role in the developmental physiology of breast fibroadenomas.

Analysis of the variability of *MED12* and *CYP17A1* in mammary fibroadenomas shows that transversions are greater than transitions. In other words, the majority of mutations induce a conformational change in the 3D structure of the mutated protein. This is in accordance with the general characteristics of nuclear DNA mutations which seem to be quality mutations. Indeed, nuclear DNA is in a protected environment, surrounded by the nuclear envelope and subject to a repair mechanism. So the fact that there are many transversions stipulates the role of the *MED12* and *CYP17A1* mutations in the occurrence of fibroadenomas.

The polymorphism of *MED12* and *CYP17A1* in breast fibroadenoma in Senegalese women is characterized by high haplotypic diversity (Hd=0.988+-0.016; Hd=0.996+-0.013) and low nucleotide diversity (Pi=0.0199+-0.0023; Pi=0.0469+-0.0034). This suggests a rapid evolution of tumor pathology in patients due to an accumulation of different genetic mutations from one patient to another. This could be explained by the heterogeneity of the breast fibroadenoma. There are different types of fibroadenoma, ranging from simple to complex in which are found cysts, sclerosing adenosis, fibrosis and epithelial calcifications. Some may therefore evolve more rapidly than others, resulting in differences in size and histology. It would be interesting to include the type of fibroadenoma in our parameters in order to understand its involvement in the evolution of breast fibroadenoma.

The frequency distribution of amino acids for *MED12* shows a statistically significant difference in Glycine, Arginine and Valine between controls and tissues. And for the *CYP17A1* gene, Cysteine, Phenylalanine, Histidine, Asparagine, Arginine, Tryptophan and Tyrosine have a statistically significant difference between controls and benign tumors. Arginine is a conditionally essential amino acid, which means that the body can synthesize sufficient

amounts of Arginine to meet basal metabolic demands and in some cases it will have to be reported through the diet. Many studies have shown that Arginine is necessary for cell growth and can become limiting in rapidly growing states such as malignancy. Given its vital role in cell growth, proliferation and immune responses. Arginine has been investigated as a potential target for cancer treatments [34]. Glycine is a non-essential amino acid because it can be synthesized by the body. RNA, DNA, creatine, serine and heme are generated by several pathways using Glycine. It acts as a neurotransmitter in the central nervous system and plays many roles as an antioxidant, anti-inflammatory, cryoprotective and immunomodulator in peripheral and nervous tissues [35]. It has recently been shown that the absorption of Glycine and its catabolism are capable of promoting tumorigenesis and malignancy, suggesting that Glycine metabolism could in principle be a target for therapeutic intervention [36]. Valine is an essential amino acid and therefore must be provided by the diet. It has an elementary role in the neurotransmitters of the brain. Cysteine is a non-essential amino acid, its use is important to maintain homeostasis and cancer cell survival [37]. Phenylalanine is an essential amino acid that is involved in the nervous system and neurotransmitters. It has been shown that any increase in Tryptophan and Phenylalanine levels in cancerous tissue may be correlated with an increased risk of breast cancer [38]. Histidine is an essential amino acid which is involved in the construction of proteins as well as several metabolic functions of our body. Asparagine is a non-essential amino acid. It is involved in the stability and biosynthesis of proteins and enzymes. Krall *and al.* [39] showed that Asparagine is an important regulator of amino acid homeostasis in cancer cells, anabolic metabolism and proliferation. Tryptophan plays an important role in the proliferation of T cells, which are the main actors in immune rejection reactions that can lead to the elimination of tumor tissue. The absence or decrease in Tryptophan levels in tumor tissues may be a risk factor for tumor progression. Tyrosine, which is a non-essential amino acid, is present only in low doses on fibroadenoma. A particular observation is made on Methionine, which is considered to be the codon that initiates protein translation. It is absent for exon 2 of *MED12* and an alteration of it has been observed in exon 1 (codon 1) of *CYP17A1* when looking for mutations. This reflects the importance of these amino acids in mammary fibroadenoma. Therefore, a much more in-depth study on the role of these amino acids in mammary fibroadenoma would be necessary.

The codon selection test showed that codon 20 of *CYP17A1* gene, which codes for Arginine is under positive selection (p=0.0076), characterizing the superiority of non-synonymous mutations compared to synonymous mutations. In other words, all the mutations affecting codon 20 cause a modification of the amino acids and thus an aberrant function of the CYP17A1 protein. On codon 20, Arginine is mutated into Glycine corresponding to the variant c.58A>G found in 61.54% of our study population, Lysine c.59G>A (38.46%) and Isoleucine c.59G>T (11.54%). The mutations in codon

20 can change the function of the CYP17A1 protein. In addition, the enzyme CYP17A1 has 20-lyase activity which allows it to convert dihydroepiandosterone (DHEA) into estrogen, and alterations in this region could confer a gain in function to the enzyme and thus an overproduction of sexual steroid hormones. Although the direct role of estrogen on the incidence of mammary fibroadenoma has not been specified and none of the steroidal hormone receptors have been expressed in breast fibroadenoma [40, 41], estrogen dependence has been suggested for their growth. In fact, the estrogen receptor (ER-β) is the only hormone receptor expressed by the breast fibroadenoma stroma at both the protein and mRNA levels [42]. Cases of breast fibroadenoma in young patients with highly ER-β-positive stroma cells indicate a hormone-receptor mechanism involved in growth regulation [40]. In addition, epidemiological studies show that 90% of mammary pathologies could be caused by environmental pollution and that the main development factor is cumulative exposure to endogenous and exogenous sources of estrogens [43] as well as aryl hydrocarbon receptors (AhR) such as dioxins, polychlorinated biphenyls (PCBs) etc., which mimic estrogenic activities [44]. Compared to control cells, breast cells exposed to long-term estrogens had higher levels of AhR in vitro [45]. Recently studies by Bidgoli *and al.* [41] have shown that higher AhR levels in young premenopausal women with fibroadenoma deregulate the expression of other tumor proliferation genes and increase the risk of tumor growth.

The study of differentiation parameters shows a differentiation between controls and breast fibroadenomas. This differentiation could be explained by the fact that breast fibroadenomas are sensitive to hormones. When they occur in teenagers or pregnant patients, they can grow remarkably large, due to the rapid increase in hormonal stimulation.

## 4. Conclusion

Advances in tumor pathology research technologies have enabled researchers to detect several genetic polymorphisms (SNPs) that influence the development of certain tumors. This led to the idea of studying the genes involved in estrogen metabolism in malignant and benign tumors. And it is by following this logic that our study had the general objective of contributing to the knowledge of the genetic factors involved in the occurrence and even the progression of breast fibroadenomas. At the end of this study, our results show that in addition to the SNP -34T>C of *CYP17A1* and codon 44 of *MED12*, 23 new variants of the *MED12* gene and 109 new variants of the *CYP17A1* gene were found. Our results also open up prospects for a study on mammary fibroadenomas in order to better understand its evolution and prevalence. Mutations in the Arginine of codon 20 influence the pathogenicity of this disease. To further investigate the role of *CYP17A1* in mammary fibroadenomas, codon 20 would be a good target to establish biomarkers. It would also be interesting to study this codon for therapeutic perspectives. Further elucidation of the involvement of Arginine in mammary fibroadenoma tumorigenesis would be of interest. It is clear that further studies would be necessary to determine the factors influencing the development of mammary fibroadenomas. To this end, broadening the study population and including clinico-pathological parameters would be a good way to better understand their evolutionary process in order to find a treatment other than surgical excision.

## Acknowledgements

## References

[1] Lee M. and Soltanian H. T. (2015). Breast fibroadenomas in adolescents: current perspectives. *Dove Press: Adolescent Health Medecine and Therapeutics*. 6. 159-163. DOI: 10.2147/AHMT.S55833.

[2] Alagar S. R. and Purushothaman R. (2015). An analytical study on fibroadenoma of the breast: ESIC Medical College and hospital, Coimbatore, Tamilnadu, India. *CIBTech Journal of Surgery*. 4 (2). 40-45.

[3] Shaik A. N., Ruterbusch J. J., Abdulfatah E., Daaboul M. H. D. F., Pardeshi V., Visscher D. W., Bandyopadhway S., Ali-Fehmi R. and Cote M. L. (2018). Breast fibroadenomas are not associated with increased breast cancer risk in an African American contemporary cohort of women with benign breast disease. *Breast Cancer Research*. 20 (1). 91p. DOI: 10.1186/s13058-018-1027-6.

[4] Ihekwaba F. N. (1994). Benign breast disease in Nigerian Women: a study of 657 patients. Journal of the Royal College of Surgeons of Edinburgh. 39 (5). 280-283.

[5] Gueye S. M. K., Gueye M., Coulibaly M. T., Mahtouk D. et Moreau J. C. (2017). Tumeurs bénignes du sein à l'unité de sénologie du Centre Hospitalier Universitaire Aristide Le Dantec de Dakar (Sénégal). *The Pan African Medical Journal*. (27). 251. DOI: 10.11604/pamj.2017.27.251.12262.

[6] Oluwole S. F. and Freeman H. P. (1979). Analysis of Benign Breast Lesions in Blacks. *The American Journal of Surgery*. 137. 786-789.

[7] Klinger K., Bhimani C., Shames J. and Sevrukov A. (2019). Fibroadenoma: From Imaging Evaluation to Treatment. Journal American Osteopathic College Radiology. 8 (2).

[8] Hefler L. A., Tempfer C. B., Grimm C., Lebrecht A., Ulbrich E., Heinze G., Leodolter S., Schneeberger C., Manfred W. Mueller M. W., Muendlein A. and, Koelbl H. (2004). Estrogen-Metabolizing Gene Polymorphisms in the Assessment of Breast Carcinoma Risk and Fibroadenoma Risk in Caucasian Women. *Cancer* 101 (2): 264-269.

[9] Interaction biosoftware a SOPHiA GENETICS company. ALAMUT VISUAL. Available on: https://www.interactive-biosoftware.com/alamut-visual/

[10] Sherry S. T., Ward M. H., Kholodov M., Baker J., Phan L., Smigielski E. M. and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, *29* (1), 308–311. DOI: 10.1093/nar/29.1.308.

[11] Landrum M., Lee J. M., Riley G. R., Jang W., Rubinstein W. S., Church D. M. and Maglott D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*. 42 (Database issue). D980-D985.

[12] Sim N-L., Kumar P., Hu J., Henikoff S., Schneider G. and Ng P. C. (2012). SIFT Web Server: Predicting Effects of Amino Acid Substitutions on Proteins ». *Nucleic Acids Research*. 40 (W1). W452-457.

[13] Adzhubei I., Jordan D. M., and Sunyaev S. R. (2013). Predicting Functional Effect of Human Missense Mutations Using Polyphen-2. *Current Protocols in Human Genetics*. Chapter7. 7.20.1-7.20.41.

[14] Schwarz J. M., Cooper D. N., Schuelke M. and Seelow D. (2014). MutationTaster2: Mutation Prediction for the Deep-Sequencing Age. *Nature Methods*. 11 (4). 361-362.

[15] Hall T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp*. 41. 95-98.

[16] Thompson T. D., Higgins D. G. and Gibson T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22. 4673-4680.

[17] Rozas J., Librado P., Sanchez-Del Barrio J. C., Messeguer X. and Rozas R. (2010). DnaSP version 5 Help Contents [Help File]. http://www.ub.edu/dnasp/

[18] Kumar S., Stecher G and Tamura K. (2015). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*.

[19] RStudio Team. (2017). RStudio: Integrated Development Environment for R. Boston, MA. RStudio, Inc. Available on: http://www.rstudio.com/.

[20] Excoffier L., Laval G. and Schneider S. (2010). Arlequin ver.3.1: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online*. 1. 45-50.

[21] Kénémé B., Mbaye F., Ka S., Diop B., Dem A. and Sembène M. (2017). Mediator Complex Subunit 12 Gene Polymorphisms in Uterine Fibroids and Breast Fibroadenomas in Senegalese Women. *International Biological Biomedical Journal*. 3 (1). 8-16.

[22] Makinen N., Mehine M., Tolvanen J., Kaasinen E., Li Y., Lehtonen H. J., Gentile M., Yan J., Enge M., Taipale M., Aavikko M., Katainen R., Virolainen E., Böhling T., Koski T. A., Launonen V., Sjöberg J., Taipale J., Vahterista P. and Aaltonen L. A. (2011). MED12 the Mediator Complex Subunit 12 Gene, Is Mutated at High Frequency in Uterine Leiomyomas. *Science*. 334 (6053). 252-255.

[23] Halder S. K., Laknaur A., Miller J., Layman L. C., Diamond M. and Al-Hendy A. (2015). Novel MED12 gene somatic mutations in women from the Southern United States with symptomatic uterine fibroids. *Mol. Genet. Genomics*. 290. 505–511. DOI: 10.1007/s00438-014-0938-x.

[24] Bourbon H-M. (2008). Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Research*. 36 (12). 3993-4008. DOI: 10.1093/nar/gkn349.

[25] Turunen M., Spaeth J. M., Keskitalo S., Park M. J., Kivioja T., Clark A. D., Mäkinen N., Gao F., Palin K., Nurkkala H., Vähärautio A., Aavikko M., Kämpjärvi K., Vahteristo P., Kim C. A., Aaltonen L. A., Varjosalo M., Taipale J. and Boyer T. G. (2014). Uterine leiomyoma-linked MED12 mutations disrupt mediator associated CDK activity. *Cell Reports*. 7 (3). 654-660. DOI: 10.1016/j.celrep.2014.03.047.

[26] Carey A. H., Waterworth D., Patel K., White D., Little J., Novelli P. Franks S. and Williamson R. (1994). Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Human Molecular Genetics*. 3 (10). 1873-1876. DOI: 10.1093/hmg/3.10.1873.

[27] Sharp L. Cardy A. H. Cotton S. C. and Little J. (2004). CYP17 gene polymorphisms: prevalence and associations with hormone levels and related factors. A HuGE review. *American Journal of Epidemiology*. 160 (8). 729-740. DOI: 10.1093/aje/kwh287.

[28] Sata F., Yamada H., Yamada A., Kato E. H., Kataoka S. Saijo Y., Kondo T., Tamaki J., Minakami H. and Kishi R. (2003). A polymorphism in the CYP17 gene relates to the risk of recurrent pregnancy loss. *Molecular Human Reproduction*. 9 (11). 725-728. DOI: https://doi.org/10.1093/molehr/gag089

[29] Li T. C., Tuckerman E. M. and Laird S. M. (2002). Endometrial factors in recurrent miscarriage. *Human Reproduction Update*. 8 (1). 43-52. DOI: https://doi.org/10.1093/humupd/8.1.43.

[30] Sun J., Zhang H. Gao M., Tang Z., Guo D., Zhang X., Li R. Liu Y. Sun W. and Sun X. (2017). Association between CYP17 T-34C rs743572 and breast cancer risk. *Oncotarget*. 9 (3). 4200-4213. DOI: 10.18632/oncotarget.23688.

[31] Kaur R., Kaur T. and Kaur A. (2018). Genetic association study from North India to analyze association of CYP19A1 and CYP17A1 with polycystic ovary syndrome. *Journal of Assisted Reproduction and Genetics*. 35 (6). 1123-1129.

[32] Suzuki Y., Nagashima T., Nomura Y., Onigata K., Nagashima, K. and Morikawa A. (1998). A New Compound Heterozygous Mutation (W17X, 436 + 5G→T) in the Cytochrome P450c17 Gene Causes 17 Alpha-Hydroxylase/17,20-Lyase Deficiency. *The Journal of Clinical Endocrinology and Metabolism* 83 (1): 199-202. DOI: 10.1210/jcem.83.1.4516.

[33] Müssig K., Kaltenbach S., Machicao F., Maser-Gluth C., Hartmann M. F., Wudy S. A., Schnauder G., Häring H-U., Seif F. J. and Gallwitz B. (2005). 17alpha-hydroxylase/17,20-lyase deficiency caused by a novel homozygous mutation (Y27Stop) in the cytochrome CYP17 gene. The Journal of Clinical Endocrinology and Metabolism. 90 (7). 4362-4365. DOI: 10.1210/jc.2005-0136.

[34] Albaugh V. L., Pinzon-Guzman C. and Barbul A. (2017). Arginine Metabolism and Cancer. *Journal of surgical oncology* 115 (3). 273-280.

[35] Razak M. A., Begum P. S., Viswanath B. and Rajagopal S. (2017). Multifarious Beneficial Effect of Nonessential Amino Acid, Glycine: A Review. *Oxidative Medicine and Cellular Longevity*. 2017.8p. DOI: https://doi.org/10.1155/2017/1716701

[36] Jain M., Nilsson R., Sharma S., Madhusudhan N., Kitami T., Souza A. L., Kafri R., Kirschner M. W., Clish C. B. and Mootha V. K. (2012). Metabolite Profiling Identifies a Key Role for Glycine in Rapid Cancer Cell Proliferation. *Science.* 336 (6084). 1040-1044.

[37] Combs J. A. and DeNicola G. M. (2019). The Non-Essential Amino Acid Cysteine Becomes Essential for Tumor Proliferation and Survival. *Cancers.* 11 (5). 678. DOI: 10.3390/cancers11050678.

[38] Mbaye F., Dem A., Fall M. and Sembène M. (2012). Biological evolution of tryptophan and phenylalanine in the occurrence of breast cancer in Senegalese women. *International journal of medicine and medical sciences.* 4 (4). 103-109.

[39] Krall A. S., Xu S., Graeber T. G., Braas D. and Christofk H. R. (2016). Asparagine promotes cancer cell proliferation though use as an amino acid exchange factor. *Nature Communications.* 7. 11457. DOI: 10.1038/ncomms11457.

[40] Sapino A., Bosco M., Cassoni P., Castellano I., Arisio R., Cserni G., Tos A. P. D., Fortunati N., Catalano M. G. and Bussolati G. (2006). Estrogen receptor- β is expressed in stromal cells of fibroadenoma and phyllodes tumors of the breast. *Modern Pathology.* 19 (4). 599-606. DOI: 10.1038/modpathol.3800574.

[41] Bidgoli S. A., Ahmadi R. and Zavarhei M. D. (2010). Role of hormonal and environmental factors on early incidence of breast cancer in Iran. *Science of The Total Environment.* 48 (19). 4056-4061.

[42] Vera-Sempere F. J. and Artes M. J. (1997). [Estrogen receptors and breast fibroadenoma with suspicious traits of malignancy]. *Anales De Medicina Interna.* 14 (7). 332-336.

[43] Darbre P. D. (2006). Environmental oestrogens, cosmetics and breast cancer. *Best Practice & Research Clinical Endocrinology & Metabolism.* 20 (1). 121-143.

[44] Van der Heiden E. Bechoux N., Muller M., Sergent T., Schneider Y-J., Larondelle Y., Maghuin-Rogister G. and Scippo M-L. (2009). Food flavonoid aryl hydrocarbon receptor-mediated agonistic/antagonistic/synergic activities in human and rat reporter gene assays. *Analytica Chimica Acta.* 637 (1). 337-345.

[45] Wong P. S., Li W., Vogel C. F. and Matsumura F. (2009). Characterization of MCF mammary epithelial cells overexpressing the Arylhydrocarbon receptor (AhR). *BMC Cancer.* 9 (234). 1-15. DOI: 10.1186/1471-2407-9-234.