ISSN: 2634-9221
European Journal of Applied Sciences; Vol. 14 No. 01 (2026) (158-175)
https://doi.org/10.14738/aivp.1401.19802

Association of Codon 72 Polymorphism in Exon 4 of the *TP53* Gene in Benign and Malignant Breast Tumors in Senegal

Hijeux Minte Cheikh Mohamed Malainy Samb[1,2], Anna Ndong[1,2], Mbacké Sembene[1,2]

1. Department of Animal Biology, Faculty of Science and Technology (FST), Cheikh Anta Diop University of Dakar (UCAD), Dakar, Senegal.
2. Laboratoire of Genomic, Cheikh Anta Diop University, Animal Biology, Dakar-Fann, Senegal.

Abstract: Background: The Arg72Pro (R72P) polymorphism of the *TP53* tumor suppressor gene has been controversially associated with breast cancer risk, with significant variations depending on ethnic origin. Little data exists for West African populations. This study seeks to evaluate the distribution and association of the R72P polymorphism of exon 4 of the *TP53* gene with benign and malignant breast tumors in the Senegalese population. Methodology: The study was conducted on 48 Senegalese women: 17 with breast cancer (malignant), 12 with benign tumors, and 19 healthy controls. The polymorphism was genotyped by PCR followed by restriction fragment analysis. Polymorphism and allelic diversity, as well as genetic differentiation parameters and correspondence factor analysis, were generated using Genetix software version 4.05.2 and Bayesian inference with STRUCTURE. Hardy-Weinberg equilibrium was tested using GenePop software version 4.3. Results: The analysis revealed a distinct allele distribution, with a predominant frequency of the C allele (Pro72) in controls (77.7%) and an increased frequency of the G allele (Arg72) in patients with tumors (25% malignant, 12.5% benign). Statistically, no significant association was found between genotypes and the risk of developing malignant or benign breast tumors. Hardy-Weinberg equilibrium tests showed a significant imbalance in the patient populations, unlike in the controls. Finally, population genetic analyses (low F_{ST} differentiation indices, negligible genetic distances, and genetic structure in two clusters) indicated high genetic homogeneity between the three groups for this specific locus. Conclusion: Although differences in allele frequency were observed, the R72P polymorphism of *TP53* is not an independent and significant risk factor for breast cancer in this Senegalese cohort. The high genetic homogeneity observed suggests that this variant alone is probably not a key determinant of breast pathology in this population. These results highlight the importance of local studies and the need for broader research incorporating other genetic and environmental factors.

Keywords: Tumor, Benign, Malignant, Breast, *TP53,* Exon 4, codon 72, Senegal

INTRODUCTION

Cancer, a disease characterized by uncontrolled and anarchic cell proliferation, is a major public health issue worldwide. This deregulation, which is found in many cancers, occurs when cells escape the strict mechanisms that control their growth and renewal [1]. Among these diseases, breast cancer remains the most common invasive cancer in women, representing a considerable health burden [2].

Malignant transformation is a complex process, often involving the alteration of critical signaling pathways. Cancer cells thus acquire proliferative and invasive capabilities through various mechanisms, such as overexpression of growth factor receptors (EGF

receptor), loss of sensitivity to inhibitory signals (TGF-β receptor), or internal reconfiguration of the signaling cascades downstream of these receptors [3]. These dysfunctions are frequently induced by genomic instabilities, leading to mutations, deletions, or rearrangements affecting key genes involved in cell cycle control, particularly oncogenes and tumor suppressor genes [4].

Among these guardians of the genome, the TP53 gene, which codes for the p53 protein, plays a central role. Known as the "guardian of the genome," p53 regulates cell cycle arrest, DNA repair, and apoptosis in response to cellular stress [5]. Mutations or polymorphisms (common genetic variations in the population) of *TP53* can therefore compromise these functions and promote the accumulation of alterations, accelerating tumorigenesis. In particular, a non-synonymous polymorphism at codon 72 of exon 4 (Arg/Pro substitution) has been extensively studied. This exon codes for part of the protein's transactivation domain, which is essential for its activity. The R72P polymorphism can thus modify the structure, stability, or apoptotic function of p53, potentially influencing individual susceptibility to cancer or tumor aggressiveness [6].

Therefore, understanding the variation in expression of this polymorphism allows us to explore the molecular mechanisms underlying tumor progression, identify prognostic or predictive biomarkers, and develop targeted therapeutic strategies for cancers associated with p53 dysfunction.

It is in this context that the present study was conducted, with the aim of evaluating the association between the R72P polymorphism of exon 4 of the *TP53* gene and breast tumors in Senegalese women, and more specifically:

- to determine the distribution of allele frequencies (Arg and Pro) and genotypes (Arg/Arg, Arg/Pro, Pro/Pro) in patients with benign and malignant breast tumors.
- to analyze the association between different genotypes and the risk of developing malignant breast tumors.

METHODOLOGY

Study Population

The study was conducted on 48 individuals, including 17 patients with breast cancer, 12 patients with benign breast tumors, and 19 control cases for comparison. These patients were recruited at the Juliot Curie Institute at Aristide Le Dantec Hospital. For each patient who underwent surgery, a sample was taken from the fresh surgical specimen in the middle of the tumor, collected in a dry tube and stored at 20°C, along with their clinical information sheet. The sheet was used to collect information on clinical, pathological, and demographic characteristics, as well as medical and/or family history. Any patient diagnosed with cancer or a benign tumor by the facility's pathologist was included in the study. This data was collected after ethical approval and obtaining informed consent, which was completed and signed by each patient.

After collection, the samples were sent directly to the Genomics Laboratory of the Department of Animal Biology at the Faculty of Science and Technology of the University of Dakar, where the biopsies were stored in 96% alcohol for various molecular analyses.

DNA Extraction and PCR-RFLP of Exon 4 of the *TP53* Gene

DNA extraction was performed using the standard protocol of the Zymo Research kit for biopsies and the Qiagen kit (Qiagen Dneasy Blood kit) for control blood samples. Amplification was performed in a 25 µl reaction volume containing 16.4 µl of MilliQ water, 2.5 µl of buffer (10X), 1 µl of MgCl2, 0.5 µl of dNTP, 1.25 µl of each primer, which are: F 5'-TCCCCCTTGCCGTTCCAA-3' and R 5'-CGTGCAAGTCACAGACTT-3', 0.1 µl of Taq polymerase, and 2 µl of concentrated DNA extract. PCR is performed by repeating cycles, which ensures that the target DNA is doubled in each cycle with initial denaturation conditions at 94°C (5 minutes) followed by 35 cycles starting with denaturation at 94°C (30 seconds), primer hybridization at 58°C (45 seconds), and elongation of complementary DNA strands at 72°C for 40 seconds, and is completed by a final elongation at 72°C for 10 minutes. The PCR products obtained were subjected to migration on a 2% agarose gel to verify the presence of amplicons, then digested with the BstUI restriction enzyme at 37°C for 2 hours in a thermocycler. The BstUI enzyme will recognize the cleavage site only if the mutant allele "G" is present. The digestion products are then analyzed by electrophoresis on a 3% gel, allowing the ¨CC¨ genotype (normal homozygote with 279 bp) to be distinguished from the ¨GC¨ genotype (heterozygote with fragments of 279 bp, 160 bp, and 119 bp) and the ¨GG¨ genotype (mutant homozygote with fragments of 160 bp and 119 bp).

Genetic and Statistical Analyses

Statistical approaches in population genetics contribute to the description of data and the possibility of inferring the evolutionary processes of allele frequency fluctuations. Each species or organism reveals considerable genetic variation that is expressed within individuals, within a population, or between populations. The level of genetic variability within populations and genetic differentiation between populations can be quantified from a set of allele frequencies, in the form of fixation and gene diversity indices, as well as genetic distances. The latter allow us to infer the genetic structure of populations and the evolution of pathology. In the analyses, only one level is considered: population.

Genetic Variability

Locus polymorphism is one of the quantitative descriptions of genetic variability. Some loci have only one allele in the sample studied, while others have several, with varying frequencies. The former are called monomorphic and the latter polymorphic. A population is said to be polymorphic if a portion of its DNA has a sequence variation corresponding to several allelic forms. In fact, it is the probability of observing at least two alleles at the same locus, and this probability depends on the respective frequencies of the alleles and also on the size of the sample [7]. Genetic variability was estimated based on a set of basic genetic polymorphism parameters, including the following:

The polymorphism rate (P) or percentage of polymorphic loci in a sample is the probability of observing at least two alleles at the same locus, and this probability depends on the respective frequencies of the alleles and also on the sample size. In this study, a locus is considered polymorphic when the most frequent allele has a frequency less than or equal to 0.95. The simplest way to describe Mendelian variation is to give the distribution of genotype frequencies in the population. Variation can exist both within a population and

between populations. Generally, instead of genotype frequencies, the frequencies of distinct alleles are used. Allele frequencies are calculated from the genotypes of individuals studied in a population and represent the ratio of the number of copies of an allele in the population to the total number of copies of all alleles. Analyses focus on codominant loci (distinguishing between heterozygotes and homozygotes) with two alleles in a diploid organism [8].

Genetic Diversity

Nei's heterozygosity or genetic diversity [9] characterizes the level of variability of genes within a given population, i.e., it expresses the probability that two variants (alleles) at the same locus randomly selected from a population will be different. On the one hand, there is observed heterozygosity HO (observed frequency of heterozygous individuals), or frequency of occurrence of heterozygotes per locus in the sample, which is the ratio of the number of heterozygous individuals to the total number of individuals genotyped for a locus. It is a measure of polymorphism.

On the other hand, expected or theoretical heterozygosity, or diversity index (HE), corresponds to the heterozygosity expected under the assumptions of Hardy-Weinberg equilibrium [10] and the unbiased expected heterozygosity Hnb proposed by Nei [11] when the number of individuals tested is small. The latter is a good estimator of the genetic variability of a population since it is less sensitive to sampling hazards such as the number of alleles observed, for example [10].

Based on allele frequencies, unbiased expected heterozygosity (Hnb) and observed heterozygosity (HO) were calculated for each population and for the total population in order to test whether these populations were in Hardy-Weinberg equilibrium, i.e., HE = HO or not, and HE< or >HO.

Equilibre d'Hardy-Weinberg

The Hardy-Weinberg equilibrium model is one of the fundamental principles of population genetics. It models the behavior of allele and genotype frequencies for a polymorphism, particularly SNPs, within a population over generations under different conditions. It states that in a diploid population of infinite size where reproduction occurs randomly, without selection, mutation, or migration, the allele and genotype frequencies of a polymorphism are stable within the population over generations. Such a population is said to be in "Hardy-Weinberg equilibrium" (HWE) [8]. However, these assumptions may not be met: in such cases, deviations from the Hardy-Weinberg equilibrium model are observed.

Deviations from Hardy-Weinberg equilibrium are tested for each locus by considering all populations and for each population, all loci combined. These exact tests were performed using Genepop version 4.3 [12] using the Markov chain method with exact estimation of Chi2 p-values using the Fisher method (the parameters set are: dememorization = 10,000, batches = 20, 5,000 iterations per batch).

The null hypothesis is that the populations are in Hardy-Weinberg equilibrium.

Genetic Differentiation

F-statistics parameters for F_{IS} and F_{ST} indices. F_{IS} measures the deficit or excess of heterozygotes relative to Hardy-Weinberg equilibrium within each group formed on the basis of phenotypic characteristics [13]. Also known as the coefficient of co-ancestry, Weir and Cockerham's θ [14], or "fixation index," F_{ST} measures the reduction in heterozygosity in subpopulations linked to differences in average allele frequencies. It is calculated using the expected average heterozygosity of the subpopulations and the expected heterozygosity of the total population. F_{ST} is always positive and ranges from 0 = panmixia (random mating, no genetic divergence within populations, hence no differences between allele frequencies in subpopulations) to 1 = complete isolation (all subpopulations concerned are panmictic and completely isolated). F_{ST} values are thus measured per locus for all populations using Genetix software version 4.05.2 [15]. The significance of the values was deduced based on a 95% confidence interval. Nei's standard genetic distance (DS) estimate [16] is based on the probability of gene identity and is therefore closely related to the definition of the coefficient of relatedness.

DS is the most widely used distance measure according to the infinite alleles model (IAM). It is intended to measure the average number of allele substitutions per locus that have occurred since the divergence of two populations, and is expected to increase linearly with time [17]. DS assumes that the rate of gene substitution per locus is uniform across loci and lineages. The genetic distance between populations was calculated using Genetix software version 4.05.2 [15].

Structuration Génétique

Correspondence factor analysis (CFA) is used to visualize relationships between individuals from different populations based on allele frequencies and to test for possible genetic admixture between populations. It is a multivariate analysis method that considers the allele frequencies of all populations at different loci as variables [18]. According to Meeûs [8], this technique is also used to detect hidden structures in a sample, such as those resulting from the Wahlund effect (deficits in heterozygotes at all loci that cannot be explained by the reproductive system). To this end, a graphical representation is produced from the allele frequencies using the Genetix version 4.05.2 program [15] to estimate the distribution of genetic diversity at all levels (individuals, subpopulations, and total population).

To infer the genetic structure of populations from genomic data, the analysis used the STRUCTURE v2.3.4 software [19]. This program implements a probabilistic mixture model based on a Bayesian Markov chain Monte Carlo (MCMC) approach, allowing the estimation of population structuring parameters while integrating the statistical uncertainty associated with loci. Bayesian inference discriminates genetic clusters by maximizing intragroup similarity of allele frequencies, while estimating individual ancestor coefficients (Q) that reveal historical or recent genetic admixture. It optimizes subsequent analyses by stratifying the sample into genetically homogeneous subgroups, thereby increasing statistical power. These steps are based on the use of neutral unlinked markers and a rigorous choice of the number of clusters (K), validated by Evanno's DeltaK (ΔK) [20] and Puechmaille's MedMeanK [21] methods.

RESULTS

Genotypes of Individuals

The electrophoretic migration profiles of restriction fragments of the 215C>G polymorphism in exon 4 of the *TP53* gene show the presence of the ¨C¨ allele in 10 healthy individuals (T3, T4, T6, T7, T8, T11, T14, T15, T16, and T17), seven individuals have the heterozygous genotype (T1, T5, T9, T10, T12, T13, and T18), and the mutant genotype is found only in one individual (T2) (Figure 1).

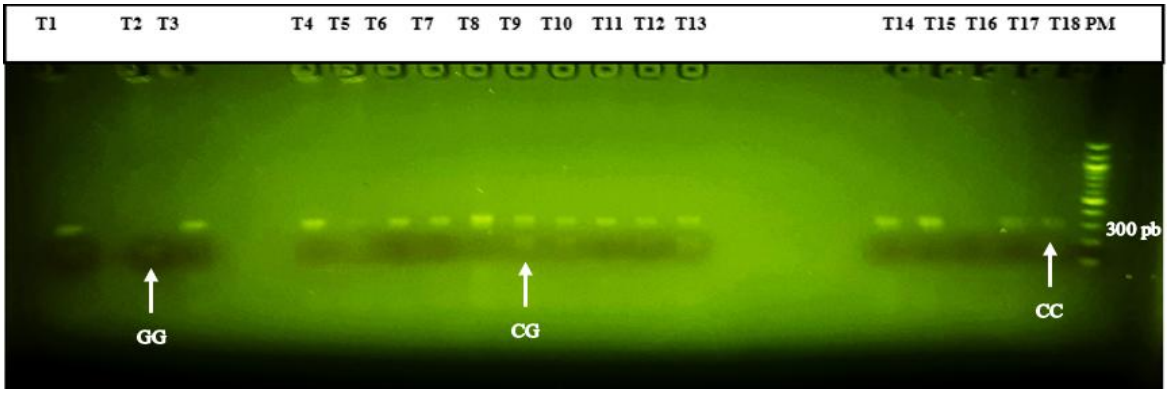

Figure 1: Electrophoretic migration profiles of *TP53* restriction fragments from controls

CC : 279 bp ; CG : 279 bp - 160 bp – 119 bp ; GG : 160 bp – 119 bp

The electrophoretic migration profiles of restriction fragments of exon 4 of the TP53 gene in benign breast tumors (Figure 2) show that seven individuals with benign breast tumors (TB1, TB2, TB3, TB5, TB6, TB8, and TB12) have the CC genotype, as do seven other patients with malignant tumors (TM2, TM3, TM6, TM10, TM11, TM12, and TM13). The mutant genotype “GG” is found only in three individuals with benign tumors (TB7, TB9, and TB10) and three other patients with malignant tumors (TM1, TM4, and TM9). Finally, the heterozygous “CG” genotype is present in two individuals (TB 4 and TB11). The heterozygous “CG” genotype is present in two individuals (TB4 and TB11) with benign tumors and three individuals (TM1, TM4, and TM9) in the population with malignant tumors.

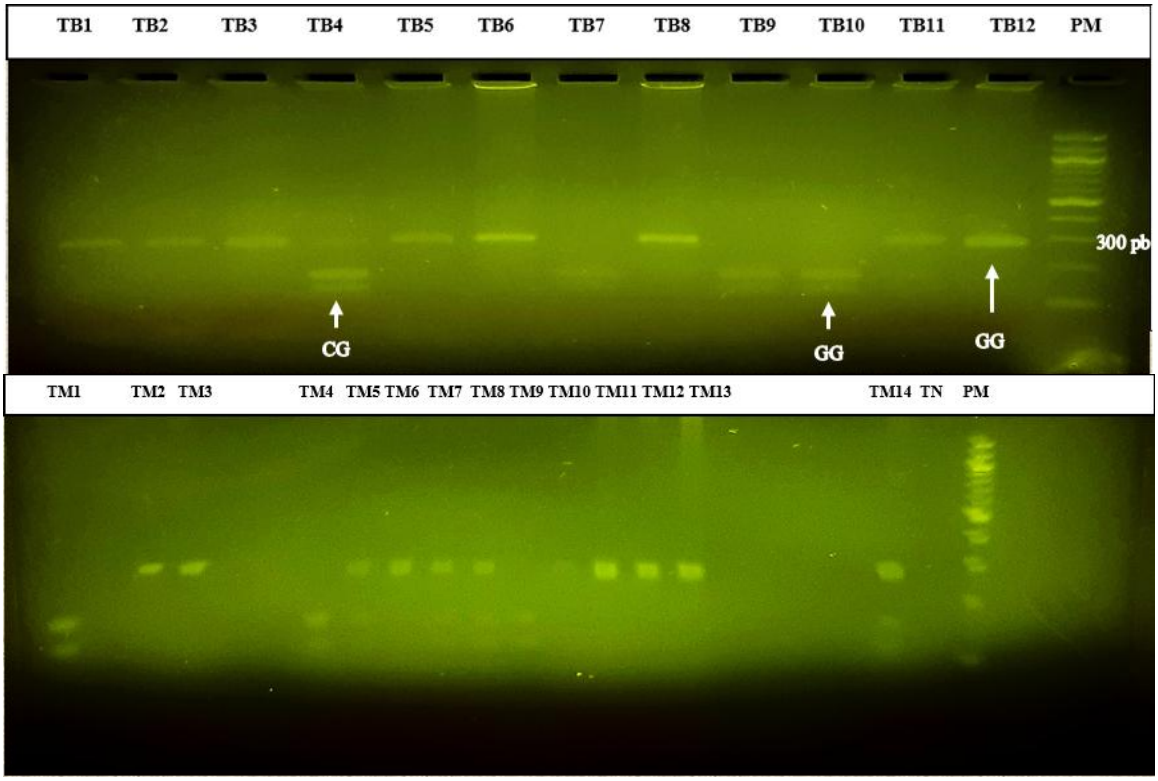

Figure 2: Electrophoretic migration profiles of restriction fragments of the TP53 gene from benign breast tumors

CC : 279 bp ; CG : 279 bp - 160 bp - 119 bp ; GG : 160 bp - 119 bp

The ¨CC¨ genotype is predominantly found in the general population, with a high frequency in controls (10/18) compared to the population with benign and malignant tumors (7/12 and 7/14, respectively). This is also observed for the heterozygous ¨CG¨ genotype, which is more common in controls (7/18) than in the population with benign and malignant tumors (3/12 and 3/14, respectively). However, the mutant allele is more common in the affected population and is found only in the control individual. The genotype of each individual is shown in Table 1.

Table 1: Characterization of individuals based on the mutation profile of polymorphism 215C>G in exon 4 of the *TP53* gene

Population	Individuals	Genotypes		
Controls	T1	ARG - PRO	CG	Het
	T2	ARG - ARG	GG	G
	T3	PRO - PRO	CC	C
	T4	PRO - PRO	CC	C
	T5	ARG - PRO	CG	Het
	T6	PRO - PRO	CC	C
	T7	PRO - PRO	CC	C

	T8	PRO - PRO	CC	C
	T9	ARG - PRO	CG	Het
	T10	ARG - PRO	CG	Het
	T11	PRO - PRO	CC	C
	T12	ARG - PRO	CG	Het
	T13	ARG - PRO	CG	Het
	T14	PRO - PRO	CC	C
	T15	PRO - PRO	CC	C
	T16	PRO - PRO	CC	C
	T17	PRO - PRO	CC	C
	T18	ARG - PRO	CG	Het
Benign tumors	TB1	PRO - PRO	CC	C
	TB2	PRO - PRO	CC	C
	TB3	PRO - PRO	CC	C
	TB4	ARG - PRO	CG	Het
	TB5	PRO - PRO	CC	C
	TB6	PRO - PRO	CC	C
	TB7	ARG - ARG	GG	G
	TB8	PRO - PRO	CC	C
	TB9	ARG - ARG	GG	G
	TB10	ARG - ARG	GG	G
	TB11	ARG - PRO	CG	Het
	TB12	PRO - PRO	CC	C
Malignant tumors	TM1	ARG - ARG	GG	G
	TM2	PRO - PRO	CC	C
	TM3	PRO - PRO	CC	C
	TM4	ARG - ARG	GG	G
	TM5	ARG - PRO	CG	Het
	TM6	PRO - PRO	CC	C
	TM7	ARG - PRO	CG	Het
	TM8	ARG - PRO	CG	Het
	TM9	ARG - ARG	GG	G
	TM10	PRO - PRO	CC	C
	TM11	PRO - PRO	CC	C

	TM12	PRO - PRO	CC	C
	TM13	PRO - PRO	CC	C
	TM14	ARG - PRO	CG	Het

T = controls; TB = Benign tumors; TM = Malignant tumors; PRO = proline ; ARG = arginine ; GG = mutant genotype; CC = wild-type genotype; G = cytosine (mutant allele); C = guanine (wild-type allele); het = heterozygous

Indices of Genetic Variability and Genetic Equilibrium

Codon 72 of exon 4 of the *TP53* gene is highly variable in the Senegalese population. This variability is marked by a higher frequency of the normal allele ¨C¨ in controls and in patients with benign tumors compared to the cancer population, which has a higher frequency of the mutant allele ¨G¨.

Analysis of allele frequencies shows that the prevalence of the ¨C¨ allele is 64.2% in cancer patients, 66.6% in the benign population, and 77.7% in controls.

The mutant allele “G” has a frequency of 25% in cancer patients, 22.2% in the control group, and 12.5% in the benign tumor population.

The CC genotype is the most common in all three populations, with a prevalence of 50% (7/14) in cancer patients, 58.3% (7/12) in patients with benign tumors, and 55.5% (10/18) in controls. The heterozygous “GC” genotype was observed in 28.57% (4/14) of cancer patients, 16.6% (2/12) of patients with benign tumors, and 44.4% (8/18) of controls. The GG genotype, meanwhile, was only detected in cancer patients, with a frequency of 21.4% (3/14), and in benign patients, with a frequency of 25% (3/12).

In the cancer population and in patients with benign tumors, the expected heterozygosity rates are higher than the observed heterozygosity rates. For controls, the observed heterozygosity is higher than the expected heterozygosity.

Statistical analyses show insignificant differences in allele and genotype frequencies between populations with p-values above the 5% threshold, as well as odds ratios (OR) and relative risks (RR) with very wide 95% confidence intervals (95% CI) that include 1, indicating a lack of statistical significance.

The genetic equilibrium test reveals that the cancerous and benign populations show an extremely significant deviation from Hardy-Weinberg equilibrium with respective p-values of (0.0285 and 0.016) below the 5% threshold (Table 2). The control population appears more stable and follows Hardy-Weinberg equilibrium with a p-value of 0.5289.

Table 2: Genetic variability and genetic equilibrium (H-W) parameters

		Malignant	**Control**			
		N=14	N=18	**P-value** 5 %	**OR** (IC à 95 %)	**RR** (IC à 95 %)
Alleles	C	0.642	0.777	1	0.73 (0.00 – 540.87)	0.85 (0.03 – 21.83)
	G	0.250	0.222	1	1.36 (0.00 – 1004.71)	1.17 (0.05 – 29.92)

Genotypes	CC	0.5	0.555	1	0.80 (0.20 – 3.25)	0.88 (0.40 – 1.93)
	GG	0.214	0	0.146	Na	2.91 (1.27 – 2.87)
	CG	0.285	4.444	0.580	0.50 (0.11 – 2.21)	0.67 (0.27 – 1.66)
Heterozygoties	Exp	0.512	0.3457	1	1.32 (0.00 – 485.80)	1.13 (0.09 – 14.82)
	Obs	0.500	0.444	1	0.76 (0.00 – 280.86)	0.89 (0.07 – 11.67)
		Malignant	**Benign**			
		N=14	N=12	**P-value** 5 %	**OR** (IC à 95 %)	**RR** (IC à 95 %)
Alleles	C	0.642	0.666	1	0.87 (0.00 – 668.83)	0.93 (0.04 – 23.18)
	G	0.250	0.125	1	1.34 (0.00 – 974.12)	1.16 (0.05 – 29.97)
Genotypes	CC	0.5	0.583	0.975	0.71 (0.15 – 3.38)	0.86 (0.42 – 1.74)
	GG	0.214	0.25	1	0.82 (0.13 – 5.08)	0.91 (0.37 – 2.22)
	CG	0.285	0.166	0.801	2.00 (0.30 – 13.51)	1.33 (0.65 – 2.73)
Heterozygoties	Exp	0.512	0.496	1	0.86 (0.00 – 249.44)	0.93 (0.06 – 13.57)
	Obs	0.500	0.416	1	0.81 (0.00 – 314.84)	0.91 (0.07 – 11.74)
		Benign	**Control**	**P-value**	**OR**	**RR**
		N=12	N=18	**P-value** 5 %	**OR** (IC à 95 %)	**RR** (IC à 95 %)
		N=12	N=18	5 %	(IC à 95 %)	(IC à 95 %)
Alleles	C	0.666	0.777	1	0.85 (0.00 – 697.33)	0.92 (0.03 – 27.58)
	G	0.125	0.222	1	1.18 (0.00 – 974.97)	1.09 (0.04 – 32.80)
Genotypes	CC	0.583	0.555	1	1.12 (0.26 – 4.91)	1.07 (0.44 – 2.61)
	GG	0.25	0	0.235	Na	3.00 (1.76 – 5.11)
	CG	0.166	4.444	0.106	0.25 (0.04 – 1.48)	0.40 (0.11 – 1.49)
Heterozygoties	Exp	0.496	0.3457	1	1.53 (0.00 – 657.05)	1,22 (0.07 – 20.42)
	Obs	0.416	0.444	1	0.65 (0.00 – 280.19)	0.82 (0.05 – 13.76)

Équilibre de H-W	**Control**	**Benign**	**Malignant**
	0.5289	0.0161*	0.0285*

N = number of individuals; Exp = expected heterozygosity; Obs = observed heterozygosity; OR = odds ratio ; RR = relative risk; IC = confidence interval ; H-W = Hardy-Weinberg ; (*) = significant ($p < 0,05$)

Genetic Differentiation Parameters

The values for individual heterozygosity or inbreeding coefficient (F_{IS}) observed are less than (-1) in the cancer population and in the population with benign breast tumors, showing a

heterozygosity deficit within these two populations. This heterozygosity deficit is also observed when taking into account the total population (F_{IT}).

The index (F_{ST}) or genetic differentiation measurement tool shows that the malignant tumor population on the one hand and the benign tumor population on the other, as well as the controls, have very few genetic differences, with respective F_{ST} values (0.000, 0.006, and -0.042) below 0.05.

At the genotypic level, individual heterozygosity rates (F_{IS}) and total population heterozygosity rates (F_{IT}) are very low, except for the CC genotype, which has a heterozygote deficit with indices (F_{IS} = 0.254 and F_{IT} = 0.242) greater than zero and an F_{ST} index = 0.14845 less than 0.15, indicating moderate genetic differentiation (Table 3).

Very low genetic differentiation (F_{ST}) is observed with indices of (F_{ST}) below 0.05 for both homozygotes ¨CC¨ and ¨GG¨ and for heterozygote ¨CG¨.

Table 3: Genetic differentiation parameters estimated per locus across all populations and per population

	F_{IS}	F_{IT}	F_{ST}
Population			
Control	0.126	0.088	-0.042
Benign	-0.087	-0.080	0.006
Malignant	-0.032	-0.032	0.000
Average	0.0026	-0.007	-0.006
Genotype			
CC	0.254	0.242	-0.016
CG	-0.096	-0.045	0.046
GG	-0.263	-0.293	-0.023
Average	0.002	-0.008	-0.010

The results of genetic differentiation are consistent with genetic distances, which indicate very low genetic diversity between populations.

However, the controls differ more from the malignant population (-0.011) than from the benign population (-0.013). The smallest genetic distance is found between the malignant and benign populations, at (-0.039). These results are presented in Table 4.

Table 4: Estimated genetic distance between populations

	Control	**Benign**
Benign	-0.013	-
Malignant	-0.011	-0.039

Genetic Structuring

Inferring the Genetic Structure of Populations Through Multivariate Analysis: Factorial Correspondence Analysis (FCA)

The total inertia is expressed on axis 1 and axis 2. The first axis expresses 77.90% of the total inertia and accounts for most of the data variability (Figure 3). This axis groups together all the controls and the majority of the cancerous and benign population, where individuals are grouped according to the similarity of their genotypes. The second axis of the PCA of the populations accounts for 22.10% of the total genetic variability, grouping together part of the cancerous and benign populations (Figure 3). Table 5 shows the distribution of individuals on the two axes according to their scores.

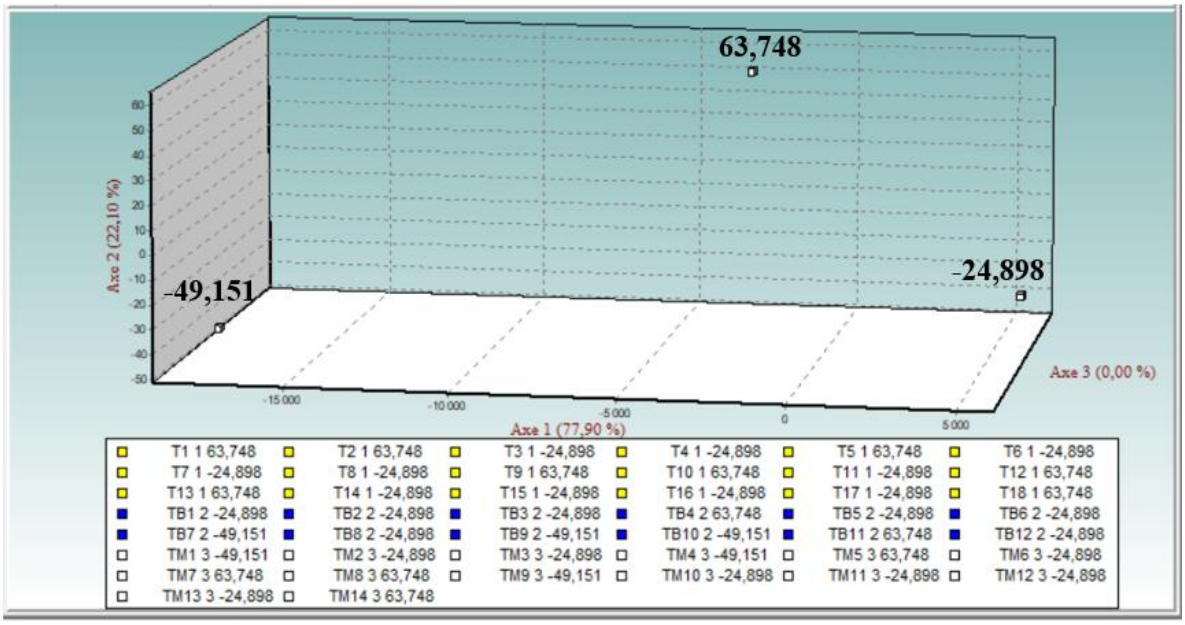

Figure 3: Correspondence factor analysis of individuals in the overall population

Table 5: Distribution of individuals on the axes according to their scores.

	Axe 1					**Axe 2**
Score	-24.898			63.748		-49.151
Individus	T2	TB1	TM2	T1	TB4	TB7
	T3	TB2	TM3	T5	TB11	TB9
	T4	TB3	TM6	T9	TM5	TB10
	T6	TB5	TM10	T10	TM7	TM1
	T7	TB6	TM11	T12	TM8	TM4
	T8	TB8	TM12	T13	TM14	TM9
	T11	TB11	TM13	T19		
	T14					

	T15					
	T16					
	T17					
	11T	7TB	7TM	7T	2TB/4TM	3TB/3TM

T = Control ; TB = Benign tumors; TM = Malignant tumors

Inférence de la Structure Génétique des Populations par Approche Bayesienne

Genetic structuring analysis conducted on a population of 44 individuals (14 cancer cases, 12 benign tumor cases, and 18 controls) identified K=2, which corresponds to the maximum peak and the optimal number of genetic clusters using the Evanno method (ΔK) (Figure 4) and the Puechmaille method (MedMean K) (Figure 5).

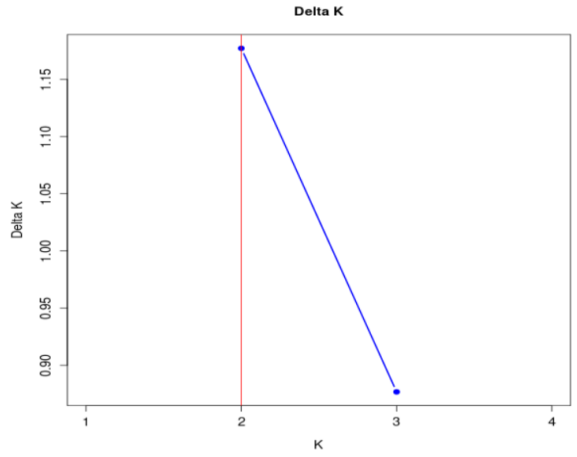

Figure 4: Identification of the number of clusters K using Evanno's method (ΔK)

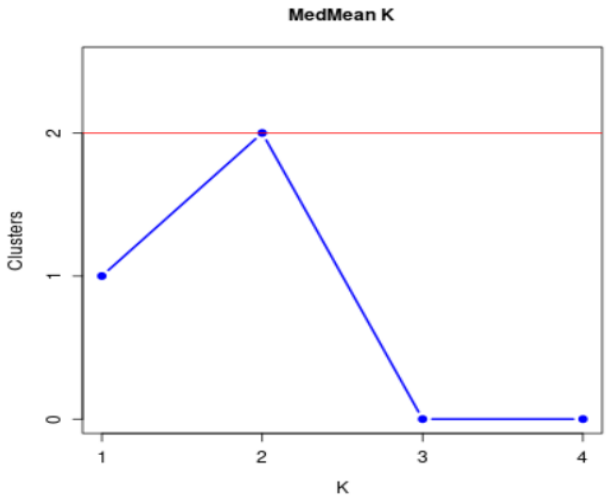

Figure 5: Identification of the number of clusters K using the Puechmaille method (MedMean K)

Inference of the genetic structure of populations revealed a distribution of Q coefficients of cluster membership mainly at extreme values (1.00 and 0.00). Cancer patients are mostly associated with coefficients close to 1.00, suggesting a strong affiliation with a specific genetic cluster, potentially linked to mutations or markers associated with cancer. In contrast, controls have coefficients close to 0.00, correlated with an absence of this genetic signal, thus confirming their status as a control group.Les valeurs intermédiaires (0,50) reflètent un mélange génétique entre les clusters, suggérant une hétérogénéité tumorale ou l'influence de facteurs environnementaux sur l'expression génétique. La figure 6 montre la structure génétique globale et la figure 7 met en exergue celles des individus des deux populations.

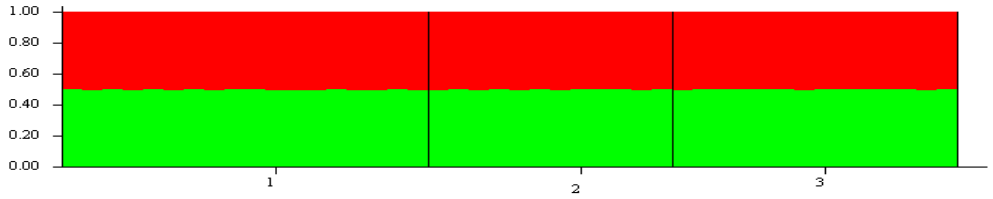

Figure 6: Genetic structure of the total population

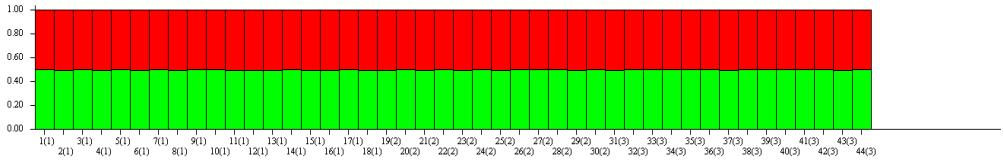

Figure 7: Genetic structuring of individuals by population

DISCUSSION

The polymorphism of codon 72 in exon 4 of the TP53 gene is a genetic variation where the nucleotide in the second position of this codon can be either a C or a G. This variation results in two versions of the p53 protein, one carrying a proline (P72) and the other an arginine (R72). Genetic analyses of codon 72 polymorphism show that this locus is highly variable in the Senegalese population. A very high frequency of C alleles was found in all three populations, and this frequency was higher in the control group. The C allele is associated with more effective triggering of cell cycle arrest in the G1 phase and activation of TP53-dependent DNA repair genes [22]. Furthermore, cells carrying the P72 variant showed reduced micronucleus formation, suggesting greater genomic stability [22], which could explain the high frequency of the normal allele ¨C¨ in the control population. The mutant allele "G," a precursor to the R72 variant, has an increased ability to induce apoptosis and also acts dominantly in the transcriptional regulation of downstream TP53 targets that induce apoptosis or repress the transformation of primary cells [23]. Therefore, the higher frequency of the mutant G allele in the affected population could translate into a defense

system against the process of healthy cells transforming into tumor cells. However, its low frequency shows that arginine appears to be less effective at eliminating transformed cells that are growing. This reduced effectiveness could allow damaged cells to survive more easily, thereby increasing the risk of developing breast cancer in subjects with benign tumors or metastatic cancer in affected subjects. On the other hand, reduced apoptosis induction may indicate the persistence of cells with DNA damage in the face of apoptosis, which can lead to the accumulation of mutations and tumor development [24]. Therefore, the "G" allele confers a form of p53 protein with less protective activity or less effectiveness in eliminating precancerous cells, thereby increasing susceptibility to the development of the disease.

The lack of statistical significance for the odds ratio (OR) and relative risk (RR) between controls and populations with benign and malignant tumors suggests that this polymorphism is not a determining risk factor for these two types of pathologies within the Senegalese population. Furthermore, the TP53 gene is primarily involved in the suppression of malignant (cancerous) tumors by stopping cell growth and inducing apoptosis in damaged cells. Therefore, the absence of a statistical link with benign tumors is not surprising, especially since other biological mechanisms of cell control, unrelated to TP53, specifically regulate benign processes. The low penetrance of this polymorphism in breast tumors was demonstrated in studies by Ma et al. [25] in their meta-analysis of the association between the codon 72 polymorphism of the TP53 gene and breast cancer risk in 24,063 subjects, as well as in research by Gonçalves et al. [26] on the association between TP53 codon 72 polymorphism and breast cancer risk: a meta-analysis. However, these results are often controversial and depend on several factors. The results of the meta-analysis by Diakite et al. [27] show that patients carrying at least one Pro allele had an increased risk of breast cancer compared to those carrying the Arg allele. In a subpopulation analysis, this study showed that the Pro allele was associated with an increased risk of breast cancer in Caucasians and Africans, while the recessive (GG) and additive (CG) models were also associated with an increased risk. These differences could be explained by sample size, allele variant types, and the studies included. These differences could be explained by sample size, types of allele variants, as the effect of polymorphism or population heterogeneity also differs between ethnic groups, and risk may depend on exposure to other factors such as gene-environment interactions that modulate the effect of polymorphism.

Hardy-Weinberg equilibrium describes a theoretical situation in which the frequencies of alleles (versions of a gene) and genotypes (combinations of alleles) remain constant from one generation to the next, indicating that no evolution is occurring in the population. This stable equilibrium is maintained only if several ideal conditions are met, such as a large population size, the absence of mutations, natural selection, migration, and non-random mating. A non-significant p-value in controls shows that the distribution of genotypes follows the proportions predicted by the model, as expected in a healthy population representative of the general population. The significant p-value obtained within the affected population indicates a Hardy-Weinberg imbalance. This indicates that the observed distribution of genotypes (Arg/Arg, Pro/Pro, and Arg/Pro) in benign and malignant populations differs significantly from the distribution expected based on allele frequencies.

This imbalance shows that the development of a tumor (benign or malignant) is not a random event. Certain genetic variants confer an advantage or disadvantage in terms of the risk of developing the disease. However, the Hardy-Weinberg model is based on the

assumption of a theoretically infinite population. Thus, the small size of the population studied may be a contributing factor to the observed deviations from Hardy-Weinberg equilibrium.

Genetic distance measures the differences that remain between individuals or populations and is based on allele or DNA sequence frequencies, providing information on their degree of divergence and evolution. The analysis reveals a low genetic distance between populations with benign tumors and malignant tumors, suggesting that these two populations are genetically very similar for the c.215C/G locus of exon 4 of the *TP53* gene. This low genetic distance also shows that the two groups of patients (benign and malignant tumors) may belong to the same source population, with no significant ethnic or geographic differences. Furthermore, the polymorphism of codon 72 of exon 4 of *TP53* may not be the key factor that causes a tumor to change from benign to malignant. In fact, the transformation of a normal cell into a tumor cell results from a paradigm involving four mutations [28]. As a result, variations in codon 72 polymorphism would be harmless because benign tumor cells are caused by two or three specific cancer mutations, while malignant tumor cells contain four specific cancer mutations and one to three tumor progression mutations [28].

Genetic differentiation represents variations in the genome between individuals or populations and is important in case-control studies to measure the degree of differentiation of specific alleles or genotypes within populations. As a result, differentiation indices can be used to identify genetic markers linked to susceptibility or protection against a disease [29]. The low genetic differentiation expressed by the index ($F_{ST} < 0.05$) found within the three populations indicates a bottleneck or strong natural selection, reducing diversity and promoting high genetic homogeneity, low diversity, and few different alleles. Within the cancerous and benign population, cells evolve towards a more uniform state through cell growth with identical mutations in several individuals (with the presence of the mutant genotype “GG” and heterozygous “CG”). This low genetic differentiation (marked by increased homozygosity or low diversity) at this locus could indicate selection or increased susceptibility in diseased populations, but this requires large-scale, specific studies. A low F_{IS} and F_{IT} value is generally associated with a lower risk of breast cancer and is above the threshold in controls, as it may indicate a less aggressive genetic mutation, as in the case of certain benign tumors, compared to more aggressive breast cancer tumors.

Correspondence factor analysis (CFA) is used to study the links between qualitative variables in the patient population (breast cancer, benign tumor, controls) in order to visualize and prioritize genetic information in a synthetic manner. This provides a better understanding of the profiles of different populations by comparing allele frequencies and identifying the salient features that differentiate or bring them together [30]. AFC and the identification of admixture through genetic structuring analysis reveal a division of the study population into two genetically distinct groups. However, no division was observed between the controls and the affected population, which is consistent with the low genetic differentiation of allele frequencies at the population level. This suggests that some patients with breast cancer and benign tumors share similar genetic profiles with controls, and codon 72 of exon 4 of the TP53 gene may be due to less aggressive molecular subtypes or common mutations that are very frequent in the Senegalese population.

CONCLUSION

This study analyzed the R72P polymorphism of the TP53 gene in Senegalese women with breast tumors. Our results reveal a specific allele distribution, characterized by a high frequency of the C allele (Pro72) in the control population and an increased frequency of the G allele (Arg72) in patients with both benign and malignant tumors. Functionally, this observation is consistent with data in the literature, with the Pro72 variant being associated with greater genomic stability, while the Arg72 variant appears to have an increased propensity to induce apoptosis. However, no strong statistical link was found between a specific genotype and an increased risk of breast cancer in our study. This confirms that this polymorphism alone is not a determining risk factor. Genetic analyses indicate that the distribution of genotypes among patients is not random, suggesting some selection pressure. Furthermore, benign and malignant tumor populations are genetically very similar for this gene, meaning that this polymorphism is probably not the key factor that causes a benign tumor to become cancerous.

REFERENCES

1. Hanahan D., Weinberg R.A., « The hallmarks of cancer », *Cell*, 2000, Jan 7, 100(1), p. 57-70, ainsi que leur actualisation in « Hallmarks of cancer: the next generation », *Cell*, 2011, March 4, 144(5), p. 646-74.
2. Ollivier-Bousquet, M. (2006). Introduction-Les cellules mammaires: du développement normal à la transformation tumorale. *Journal de la société de Biologie*, *200*(2), 179-180.
3. de Cremoux, P., & Robert, J. (2012). Cell signaling and cancer: characterization of therapeutic targets. *Pathology Biology*, 60(4), 217-222.
4. Smeby, J., Sveen, A., Eilertsen, IA, Danielsen, SA, Hoff, AM, Eide, PW, ... & Lothe, RA (2019). Transcriptional and functional consequences of TP53 splicing mutations in colorectal cancer. *Oncogenesis*, 8 (6), 35.
5. Kim, H., & Kim, Y. M. (2018). Pan-cancer analysis of somatic mutations and transcriptomes reveals common functional gene clusters shared by multiple cancer types. *Scientific reports*, 8(1), 6041.
6. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, RL, Torre, LA, & Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394-424.
7. Foulley, J. L., & Ollivier, L. (2006). Estimating allelic richness and its diversity. *Livestock Science*, *101*(1-3), 150-158.
8. De Meeûs TM. Introduction to the genetics of natural populations: Application to parasites and their vectors. *Marseille, IRD Editions, Didactiques Collection*, 335p.
9. Nei, M. (1973). The theory and estimation of genetic distance.
10. Nei M. Molecular Evolutionary Genetics. Columbia University Press. 1987; New York, 506p.
11. Nei, M. (1978). The theory of genetic distance and evolution of human races. *Japanese Journal of Human Genetics*, *23*(4), 341-369.
12. Rousset F. GENEPOP: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources. 2008; 8: 103-106.

13. Lenfant P. Heterozygosity and selective value: the case of the marine fish, Diplodus sargus (Linnaeus, 1758). *Proceedings. Biologies*. 2002; 325(3), 239-252.

14. Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, 1358-1370.

15. Belkhir K, Borsa P, Chikhi L, Raufaste N, & Bonhomme F. GENETIX 4.05, Windows TM software for population genetics. Laboratory Genome, Populations, Interactions, CNRS UMR 5171, University of Montpellier II, Montpellier. 2004.

16. Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, 106(949), 283-292.

17. Takezaki N, & Nei M. Inconsistency of the maximum parsimony method when the rate of nucleotide substitution is constant. *Journal of molecular evolution*. 1994; 39, 210-218.

18. Hoda, Anila, and Paolo Ajmone Marsan. "Genetic characterization of Albanian sheep breeds by microsatellite markers." Analysis of genetic variation in animals. 2012; 3-27.

19. Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

20. Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, *14*(8), 2611-2620.

21. PUECHMAILLE, Sebastien J. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular ecology resources*, 2016, vol. 16, no 3, p. 608-627.

22. Lin, H. Y., Huang, C. H., Wu, W. J., Chang, L. C., & Lung, F. W. (2008). TP53 codon 72 Gene Polymorphism Paradox in Associated with Various Carcinoma Incidences, Invasiveness and Chemotherapy Responses. *International journal of biomedical science : IJBS*, 4(4), 248–254.

23. Siddique, M., & Sabapathy, K. (2006). Trp53-dependent DNA repair is affected by codon 72 polymorphism. *Oncogene*, 25 (25), 3489-3500.

24. Aubrey, B. J., Strasser, A., & Kelly, G. L. (2016). Tumor-Suppressor Functions of the TP53 Pathway. *Cold Spring Harbor perspectives in medicine*, *6*(5), a026062.

25. Ma, Y., Yang, J., Liu, Z., Zhang, P., Yang, Z., Wang, Y., & Qin, H. (2011). No significant association between *TP53* codon 72 polymorphism and breast cancer risk: a meta-analysis of 21 studies involving 24,063 subjects. *Breast Cancer Research and Treatment,* 125 (1), 201-205.

26. Gonçalves, ML, Borja, SM, Cordeiro, JABL, Saddi, VA, Ayres, FM, Vilanova-Costa, CAST, & Silva, AMTC (2014). Association between TP53 codon 72 polymorphism and breast cancer risk: a meta-analysis. *Springerplus*, 3 (1), 749.

27. Diakite, B., Kassogue, Y., Dolo, G., Wang, J., Neuschler, E., Kassogue, O., Keita, M. L., Traore, C. B., Kamate, B., Dembele, E., Nadifi, S., Murphy, R. L., Doumbia, S., Hou, L., & Maiga, M. (2020). p.Arg72Pro polymorphism of P53 and breast cancer risk: a meta-analysis of case-control studies. *BMC medical genetics*, 21(1), 206.

28. Koten, J. W., Neijt, J. P., Zonnenberg, B. A., & Den Otter, W. (1993). The difference between benign and malignant tumours explained with the 4-mutation paradigm for carcinogenesis. *Anticancer research*, 13(4), 1179–1182.

29. Rousseau, F., & Laflamme, N. (2003). Human molecular genetics: from monogenic diseases to complex diseases. M/*S: Medicine Sciences*, 19(10), 950-954.

30. Dozo, B.-O. (2011). 3. Correspondence factor analysis. In Mesures de l'écrivain. Liège: Presses universitaires de Liège.