Association of Homozygous Wild-Type Glutathione S-Transferase M1 Genotype with Increased Breast Cancer Risk

Nady Roodi,1 William D. Dupont,2 Jason H. Moore,3 and Fritz F. Paril1

Departments of 1Pathology, 2Preventive Medicine and Biostatistics, and 3Molecular Physiology and Biophysics, Program in Human Genetics and Vanderbilt Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee

Abstract

More than 500 studies have examined the association of the glutathione S-transferase M1 (GSTM1) genotype with various malignancies yielding inconsistent results. The genotyping was based on a PCR assay that identified the GSTM1 null (−/−) genotype but did not distinguish homozygous wild-type (+/+), and heterozygous (+/−) individuals. We developed an assay that allowed the definition of +/+, +/−, and −/− genotypes by separate identification of wild-type and null alleles, which were found with frequencies of 0.225 and 0.775, respectively, in Caucasian women. We applied the new assay to a breast cancer case-control study and identified the +/+ genotype in 14 (6.9%) of 202 control subjects compared with 37 (18.2%) of 203 patients. Compared with women with the −/− genotype, the relative risk of breast cancer for the +/+ genotype was 2.83 (95% confidence interval, 1.45–5.59; \( P = 0.002 \)), suggesting a protective effect of the GSTM1 deletion.

Introduction

Glutathione S-transferases (GSTM1) constitute a superfamily of ubiquitous, multifunctional enzymes, which play a key role in cellular detoxification (1). The GSTs catalyze the conjugation of the tripeptide glutathione (GSH) to a wide variety of exogenous and endogenous chemicals with electrophilic functional groups (e.g., products of oxidative stress, environmental pollutants, and carcinogens), thereby neutralizing their electrophilic sites and rendering the products more water soluble (2). On the basis of sequence homology and immunologic cross-reactivity, human cytosolic GSTs have been grouped into seven families, designated GST-α, -μ, -π, -σ, -θ, and ζ (1, 3). The GST-μ family is encoded by a 100-kb gene cluster at 1p13.3 arranged as 5′-GSTM4-GSTM2-GSTM1-GSTM5-GSTM3-3′ (4). Deletion of the GSTM1 gene, GSTM1*0, frequently affects both alleles, resulting in the so-called null genotype, GSTM1−/−. A meta-analysis of 30 studies involving >10,000 individuals identified the GSTM1 null genotype in 53% Caucasians, with a 42–62% range for individual studies including our own (5, 6). The frequency of the GSTM1 null genotype was similar in Asians but lower in African-Americans (27%; range, 16–36%). Detailed mapping of the GST-μ gene cluster revealed that two almost identical 4.2-kb regions flank the GSTM1 gene. The GSTM1*0 deletion is caused by homologous recombination involving the 5′ and 3′ 4.2-kb repeats (4). Analysis of 20 GSTM1*0 alleles from 13 unrelated individuals showed the same recombination pattern, which results in a 16-kb deletion containing the entire GSTM1 gene. The GSTM1 gene is excised relatively precisely leaving the adjacent GSTM2 and GSTM5 genes intact. Therefore, one can rule out recombination with neighboring GSTM genes as a possible mechanism for the GSTM1*0 deletion despite extensive homologies in certain regions.

In view of the importance of GSTs in cellular detoxification, the enzyme deficiency associated with the GSTM1 null genotype has attracted considerable attention with regard to cancer epidemiology. A search of the literature published from 1993 to 2003 listed >500 studies of the GSTM1 genotype in relation to lung, breast, colon, brain, and various other types of cancer (7–9). These studies have in common PCR-based genotyping using an assay designed to identify the wild-type allele of GSTM1 (10). The absence of a PCR product (273 bp) indicates the GSTM1 null genotype. Consequently, study participants were categorized as either wild-type or null “genotypes.” This analytical approach has one basic flaw because it does not positively identify the null allele and therefore cannot distinguish homozygous wild-type (+/+), from heterozygous (+/−) individuals. Assuming that the presence of 2, 1, or 0 GSTM1 alleles is associated with a gene-dosage effect resulting in high-, low-, or non-GSTM1 conjugator phenotypes, the current approach oversimplifies phenotypes as all or none. In this study, we have analyzed the GSTM1 gene locus and designed a PCR assay to allow positive identification of the null allele. Combined with the identification of the wild-type allele, we could perform true GSTM1 genotyping and examine the associated inheritance patterns. On the basis of the newly gained information, we reanalyzed a hospital-based case-control study to determine whether the GSTM1 genotype is associated with breast cancer risk (5, 11).

Materials and Methods

Subjects. The hospital-based case-control study group of 203 Caucasian and 59 African-American women with primary invasive breast cancer and their age-matched control subjects has been described previously (5, 11). Genomic DNA was extracted from tumor tissue or WBCs. The DNA samples of one Caucasian control subject and five African-American patients had been depleted in previous studies, leaving 203 patients and 202 control subjects for the Caucasian population and 54 patients and 59 control subjects for the African-American study group.

DNA Analysis. DNA samples were analyzed by three separate PCR reactions. (a) The GSTM1 wild-type allele was identified by using primers M1, 5′-CGCCCTACTTGATGATGGG-3′ and M2, 5′-CTGGATTGTAGCA-GATCATGC-3′ (12) and amplification conditions described previously for the 273-bp PCR product (5). (b) The primers for detection of the GSTM1 null allele were M3, 5′-CTGTTGAGAGCTTATGCTGAA-3′ and M4, 5′-TCTT-GAGGACTTGACTGATC-3′. PCR of the null allele was carried out in a total volume of 50 μl containing 0.5–1.0 μg DNA using the GeneAmp XL PCR kit as specified by the manufacturer (Applied Biosystems, Foster City, CA). Amplification conditions consisted of an initial denaturation step at 92°C for 2 min, followed by 10 cycles of 92°C for 10 s, 54°C for 30 s, and 68°C for 8 min, and then by 29 cycles of 92°C for 10 s, 54°C for 30 s, and 68°C for 8 min plus 10 s for each successive cycle, and final elongation at 68°C for 10 min. The 14-kb PCR product was electrophoresed in 0.5% SeaKem Gold agarose gel (Cambrex, East Rutherford, NJ) and visualized by ethidium bromide staining. Digestion of the 14-kb product with restriction endonuclease SwaI yielded two fragments of ~12.4 and 1.6 kb. Each PCR contained wild-type and null allele internal controls, and random samples were repeated...
samples were classified as of all of the individuals. On the basis of this approach, all of the combined analysis of the two PCR reactions permitted positive identification of the null allele. The expected PCR product of ~30 kb for the wild-type allele could not be amplified because of its length.

All of the DNA samples also were analyzed by the established short-range PCR using the original primers within the GSTM1 gene to obtain a 273-bp product for the wild-type allele (Fig. 2). The combined analysis of the two PCR reactions performed positive identification of wild-type and null alleles, resulting in GSTM1 genotyping of all of the individuals. On the basis of this approach, all of the samples were classified as ++, +/−, or −/− (Fig. 2D). Each PCR contained wild-type and null allele internal controls. The results of the short- and long-range PCR assays confirmed each other in every instance (i.e., samples lacking the wild-type allele always contained null allele and vice versa). As yet another control for the validity of the 14-kb PCR results and the integrity of the DNA samples, we developed an independent long-range PCR. We chose the ERα gene at 6q25.1 and designed primers to amplify a 14-kb fragment in exon 112 bp, whereas the introns vary from 87–2641 bp (see top of diagram). The point of deletion cannot be localized precisely because of the high sequence identity between the repeats. PCR primers 1 and 2 in exons 4 and 5, respectively ( ▶ and ▶), were used to identify the presence of the wild-type allele, yielding a product of 273 bp. Primers 3 and 4 were designed to anneal outside the homology region. The expected PCR product of ~30 kb for the wild-type allele could not be amplified. However, the null allele yielded a 14-kb product, which allowed positive identification of the null allele. The arrowheads are not drawn to scale. The vertical arrows ( ▶) indicate the Swal digestion sites.

Statistical Methods. We performed likelihood ratio tests of Hardy-Weinberg equilibrium for patients and control subjects using the method of Elston and Forthofer (13). Relative risks for breast cancer were estimated by odds ratios and derived using logistic regression (14). These relative risks were adjusted for age by including age as a covariate in the regression models.

Results

To develop a PCR-based strategy for the positive identification of the GSTM1 null allele, we analyzed the GSTM1 gene locus at 1p13.3 (Fig. 1). We initially designed primers near the 4.2-kb repeat regions, trying to use minor sequence differences between the upstream and downstream regions. However, the PCR reactions yielded ambiguous results with multiple bands and inconsistent restriction endonuclease digests. Primers in adjacent homologous regions also failed to produce unambiguous patterns. Finally, we designed primers that annealed clearly outside the homologous regions flanking the GSTM1 gene. This necessitated the use of long-range PCR amplification, which yielded a 14-kb product for the GSTM1 null allele (Fig. 2, A and B).

To confirm the amplified sequence, we selected the restriction endonuclease Swal, an 8-nucleotide cutter, with one recognition site in intron 7 and another upstream of the S′ repeat region, 1570 bp from primer 3 (Fig. 1). Digestion of the 14-kb product yielded the expected two fragments of ~12.4 and 1.6 kb (Fig. 2A). The expected PCR product of ~30 kb for the wild-type allele could not be amplified because of its length.

We performed likelihood ratio tests of Hardy-Weinberg equilibrium for patients and control subjects using the method of Elston and Forthofer (13). Relative risks for breast cancer were estimated by odds ratios and derived using logistic regression (14). These relative risks were adjusted for age by including age as a covariate in the regression models.

Figure 2. Glutathione S-transferase M1 (GSTM1) genotyping. A, PCR analysis of GSTM1 null allele in three women yielded a single 14-kb product, which on digestion with Swal resulted in ~12.4- and 1.6-kb fragments. Lanes are designated as u (undigested) and d (digested) products; M, DNA molecular weight markers. B, long-range PCR analysis of GSTM1 null allele in seven women. The 14-kb PCR product (Lanes 3–7) indicates the presence of the null allele, whereas the lack of a product reveals its absence (Lanes 1 and 2). C, short-range PCR analysis of GSTM1 wild-type allele in the same seven women. The 273-bp PCR product (Lanes 1–5) indicates the presence of the wild-type allele, whereas the lack of a product reveals its absence (Lanes 6 and 7). D, the combined analysis of null and wild-type alleles indicates +/+ (Lanes 1 and 2), +/− (Lanes 3–5), and −/− (Lanes 6 and 7). E, separate long-range PCR analysis of the ERα gene yields a 14-kb fragment in every DNA sample, indicating DNA integrity.

Table 1 Distribution of GSTM1* genotypes in breast cancer patients and control subjects

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Caucasian</th>
<th>African-American</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (n = 203)</td>
<td>Control subjects (n = 202)</td>
<td>Patients (n = 54)</td>
</tr>
<tr>
<td>++/−</td>
<td>37 (18.2)</td>
<td>14 (6.9)</td>
</tr>
<tr>
<td>+/+</td>
<td>49 (24.1)</td>
<td>63 (31.2)</td>
</tr>
<tr>
<td>−/−</td>
<td>117 (57.7)</td>
<td>125 (61.9)</td>
</tr>
</tbody>
</table>

*GSTM1, glutathione S-transferase M1.

Number of individuals followed by percentage in parentheses.

A separate long-range PCR used primers E1, 5′-GGAGCTGGTTCACATGATCAAC-3′ and E2, 5′-CTCGGCCACTCCT-TAGTGCAATG-3′ to yield a 14-kb fragment of exon and intron 4 of the estrogen receptor α (ERα) gene. Amplification conditions were the same as for reaction 2 except the annealing temperature was 50°C.

We performed likelihood ratio tests of Hardy-Weinberg equilibrium for patients and control subjects using the method of Elston and Forthofer (13). Relative risks for breast cancer were estimated by odds ratios and derived using logistic regression (14). These relative risks were adjusted for age by including age as a covariate in the regression models.

We determined the frequency of the GSTM1 wild-type and null alleles in the Caucasian control population to be 0.225 and 0.775, respectively (Table 1). The distribution of homozygous and heterozygous individuals was consistent with Hardy-Weinberg equilibrium. There were 14 (6.9%) of 202 Caucasian control subjects with the ++/− genotype but 37 (18.2%) among the 203 patients. Thus, the Caucasian cancer population showed a conspicuous deviation from the Hardy-Weinberg law with an excess of ++/− individuals (P < 0.0001). The
frequency of the GSTM1 wild-type and null alleles in the African-American control population was 0.407 and 0.593, respectively (Table 1). The distribution of homozygous and heterozygous individuals in the control population again was consistent with Hardy-Weinberg equilibrium, whereas the cancer population deviated with an excess of +/+ individuals ($P = 0.002$).

Compared with Caucasian women with the $+/-$ genotype, the relative risk of breast cancer was 0.83 (95% confidence interval, 0.53–1.30; $P = 0.42$) for the $+/-$ genotype and 2.82 (95% confidence interval, 1.45–5.49; $P = 0.002$; Table 2) for the $+/+$ genotype. There was no evidence for a gene-dose effect ($P$, trend = 0.42). Among African-American women, the $+/+$ genotype was observed in 13 (22.0%) of 59 control subjects compared with 18 (33.4%) of 54 patients, but the increased relative risk associated with the $+/+$ genotype did not reach significance (Table 2).

**Discussion**

The majority of polymorphisms affecting genes involved in carcinogen metabolism are single nucleotide polymorphisms. Deletions are less common, and the complete absence of a gene in form of a null allele is rare. For this reason, the GSTM1$^{−/−}$ genotype has attracted so much attention and become the focus of >500 publications in molecular epidemiology. A limitation of these studies was the PCR assay, which did not truly genotype GSTM1 but only identified $−/−$ homozygosity without being able to separate the $+/+$ and $+/+$ genotypes. The positive identification of the wild-type and null alleles described here allowed definition of the $+/+$, $+/-$, and $−/−$ genotypes and unambiguous assignment of high, low, and no conjugator phenotypes. Thus, we finally could determine the frequency of the GSTM1 wild-type and null alleles in Caucasian and African-American populations and assess the effect of the GSTM1 genotype on cancer risk. We defined the difference between the populations with the GSTM1 wild-type allele as nearly twice as common in African-American (0.407) than in Caucasian (0.225) women.

The PCR-based analysis of genotypes has become standardized and is not a major source of error in molecular epidemiologic investigations, provided proper precautions are taken, such as prevention of DNA sample contamination, inclusion of positive and negative controls in each assay, and repeated analysis of random samples. The majority of PCR assays are short-range amplifications of DNA fragments <3 kb in length. In this study, we combined the result of such a typical short-range assay with that of a long-range assay yielding a 14-kb product. In addition to the usual precautions, we validated the result of the long-range assay by digesting the 14-kb fragment with the restriction endonuclease SmaI, obtaining the expected digestion fragments in every instance. Another form of validation was offered by the consistency of results of the short- and long-range assays, which confirmed each other in every instance (i.e., samples lacking the 14-kb fragment always contained the 273-bp product and vice versa). Finally, to rule out failure of the long-range 14-kb GSTM1 assay resulting from poor DNA quality as a possible cause of erroneous genotyping, we developed a separate long-range assay of the ERα gene. Consistent with earlier studies, every DNA sample tested contained the 14-kb ERα fragment, including $+/+$, $+/-$, and $−/−$ samples, confirming DNA integrity (15, 16).

To examine the association of the GSTM1 genotype with cancer risk, we reanalyzed a breast cancer case-control study that had failed to show any effect of the GSTM1$^{−/−}$ genotype (5, 11). The GSTM1$^{+/−}$ genotype occurred more frequently in Caucasian breast cancer patients and was associated with a significantly higher risk compared with the null $−/−$ genotype. The association between the GSTM1$^{+/−}$ genotype and elevated breast cancer risk was unexpected and requires an explanation, which is speculative at this time, ranging from linkage of GSTM1 with other genes to the substrate GSH and population genetics of the null deletion. Mammalian cells have evolved protective mechanisms, such as GSH conjugation, to minimize injurious events that result from toxic chemicals and normal oxidative products of cellular metabolism. GSH depletion to ~20–30% of total GSH levels can impair the conjugation defense against the toxic actions of such compounds and become detrimental to cellular processes (17). Thus, the combined conjugation activities of all of the GSTs may lead to GSH depletion and thereby become counterproductive. Instead of protecting, the GSTs collectively may expose the cell to injurious effects, such as oxidative DNA damage and associated mutagenic lesions. Although conjecture, this may explain the high frequency of the GSTM1$^{−/−}$ genotype. It seems that the deletion of the GSTM1 gene occurred not only with impunity but also actually may have offered a survival advantage for the cell. We do not know when the deletion of the human GSTM1 gene occurred, but it is interesting to note that the gene is found in African-American women at nearly twice the frequency as in their Caucasian counterparts. Could it be that as humans migrated from Africa, the deletion of the GSTM1 gene became beneficial, leading to the uniquely high absence of this gene in Caucasians and Asians? This idea is not unprecedented. For example, there is a 32-bp deletion in the coding region of the human CCR5 gene that leads to loss of chemokine receptor function. This loss of function mutation is important because the CCR5 protein facilitates the infection of macrophages and monocytes by HIV. Individuals who are homozygous for the CCR5 deletion have nearly complete resistance to HIV infection (18). Interestingly, the functional gene always is present in Africans, whereas the deletion occurs in frequencies of 0–0.14 in a geographic cline across Europe and Asia. The increased frequency of the CCR5 deletion in Caucasians is believed to be the result of strong selection because the mutation originated an estimated 700 years ago (19).

The African-American study group also showed a higher frequency of the GSTM1$^{+/+}$ genotype among patients than control subjects. Although the relative risk of breast cancer associated with the $+/+$ genotype was increased compared with the $−/−$ genotype, the increase was not significant. Possible reasons for the lack of significance are the smaller size of the study group and the different allele frequency in African-Americans. Whatever selection process favored the deletion of the GSTM1 gene in Caucasians may have magnified the difference in risk associated with the wild-type allele in relation to breast cancer. Besides GSTM1, there are other members of the GST superfamily that are expressed in breast tissue, such as GSTP1 and GSTA1 (20, 21). Interestingly, another GST family member, namely the GSTT1 gene at 22q11.2, can be deleted, resulting in the $−/−$ genotype in 20% of Caucasians and 47% of Asians (6). The size and mechanism of the GSTT1 deletion have not been determined, and it is unknown whether the gene is expressed in breast tissue.

In summary, the present study involved a hospital-based breast cancer case-control population of Caucasian and African-American women (5, 11). We took the unusual step of reanalyzing the same study population to clarify the role of a single gene, GSTM1. The

### Table 2: Relative risk of breast cancer based on GSTM1$^{*}$ genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Relative risk*</th>
<th>95% CF</th>
<th>$P$ value</th>
<th>Relative risk*</th>
<th>95% CF</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$−/−$</td>
<td>1.00$^{a}$</td>
<td></td>
<td></td>
<td>1.00$^{a}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$+/-$</td>
<td>0.83</td>
<td>0.53–1.30</td>
<td>0.42</td>
<td>0.88</td>
<td>0.36–2.10</td>
<td>0.77</td>
</tr>
<tr>
<td>$+/+$</td>
<td>2.82</td>
<td>1.45–5.49</td>
<td>0.002</td>
<td>1.66</td>
<td>0.66–4.20</td>
<td>0.28</td>
</tr>
</tbody>
</table>

* GSTM1, glutathione S-transferase M1.

$^{a}$ Adjusted for age.

$^{b}$ 95% Confidence interval.

$^{c}$ Denominator for following relative risks.
previous analysis was based on a PCR assay that identified the GSTM1 null (−/−) genotype but did not distinguish homozygous wild-type (+/+ and heterozygous (+/−) individuals. We developed an analysis that allowed the definition of +/+, +/−, and −/− genotypes by a combination of three separate PCR assays: (a) the short-range 273-bp assay for the GSTM1 wild-type allele; (b) the long-range 14-kb assay for the GSTM1 null allele, digested by SwaI; and (c) the long-range 14-kb assay of the ERα gene to validate DNA integrity. The new analysis revealed an association of +/+/− homozygosity with elevated risk in Caucasian women. Regardless of the explanation underlying the association between the +/+/− genotype and increased breast cancer risk, it will be interesting to apply true GSTM1 genotyping to additional or previously analyzed groups with breast cancer and other malignancies.

References

Association of Homozygous Wild-Type Glutathione S-Transferase M1 Genotype with Increased Breast Cancer Risk


Cancer Res 2004;64:1233-1236.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/64/4/1233

Cited articles
This article cites 18 articles, 9 of which you can access for free at:
http://cancerres.aacrjournals.org/content/64/4/1233.full#ref-list-1

Citing articles
This article has been cited by 7 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/64/4/1233.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/64/4/1233.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.