Mammographic Density and Expression of the Genes Involved in the de novo Cholesterol Biosynthesis

Danila Coradini

Laboratory of Medical Statistics and Biometry, Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy

ARTICLE INFO

Received: 27 May 2024
Revised: 30 June 2024
Accepted: 7 July 2024

ABSTRACT

**Background:** This *in silico* study investigated the association between the local biosynthesis of cholesterol and mammographic density, the major risk of developing breast cancer, as a function of the three cellular components of breast tissue (epithelium, fatty, and non-fatty stroma).

**Methods:** The study compared the expression of 7 genes (*HMGCR, FDPS, FDFT1, GGPS1, SQLE, LSS,* and *SREBF2*) involved in the *de novo* cholesterol biosynthesis, first, according to the radiological density (dense vs. non-dense breast) and, then, according to the cellular components of breast tissue, regardless of the radiological classification.

**Results:** *HMGCR, SQLE,* and *SREBF2* were significantly more frequently expressed in radiologically dense than in non-dense breasts (-1.70 vs. -1.41, *P=0.0028;* -3.63 vs. -3.31 *P=0.0003;* -0.92 vs. -0.76, *P=0.0271,* respectively). When the samples were reclassified based on their cellular components as highly fatty and highly non-fatty, *HMGCR, SQLE,* and *SREBF2* were significantly more frequently expressed in highly non-fatty samples (-1.48 vs. -1.94, *P<0.0001;* -3.39 vs. -4.18, *P<0.0001;* -0.77 vs. -0.94, *P=0.0103,* respectively), whereas *LSS* was overexpressed in high fatty ones (0.28 vs. -0.60, *P<0.0001) . Besides, while in the highly non-fatty subgroup, *SREBF2* was positively associated with both *HMGCR* (*r=0.53, P<0.0001*) and *SQLE* (*r=0.73, P<0.0001*), in the highly fatty subgroup, these positive correlations disappeared (*SREBF2*/*HMGCR: *r=-0.19,* *P=0.3026*) or substantially decreased (*SREBF2*/*SQLE: *r=0.41,* *P=0.0173*).

**Conclusion:** Findings provide a compelling biological explanation for the clinical evidence that women with radiologically dense breasts are at a higher risk of developing cancer compared to those with non-dense breasts because of the prevalence of non-fatty tissue, where the altered expression of genes leading to an increased cholesterol production can contribute to the transformation of epithelial cells, and support the use of mammographic density as a reliable surrogate marker to identify women who may benefit from a preventive treatment aimed at reducing cholesterol production.

INTRODUCTION

Mammographic screening is the primary approach to detecting neoplastic lesions in the breast. It is based on evaluating the mammographic density (MD), which quantifies the radiologically dense breast components (epithelial and non-fatty stromal tissue) compared to the transparent fatty tissue. Epidemiological evidence indicates that MD is a crucial risk factor for non-familial breast cancer.
Women with a breast density greater than 75% are 4 to 6 times more likely to develop breast cancer than women with a breast density lower than 10%.1,2

Early models have been focused on the epithelial component of breast tissue, assuming that the increased breast density was due to the overproliferation of epithelial cells in response to the combined effect of genetic alterations and exposure to exogenous estrogens to explain the association between MD and breast cancer risk.3,4 However, considerable evidence has demonstrated that the stroma, considered just as a “connective” tissue for a long time, plays an essential role in the regulation of the mammary gland morphogenesis through a complex and dynamic interaction with the epithelium that, when dysregulated, can induce and promote tumorigenesis.5,6 Therefore, additional studies, aimed at understanding the biological relationship between MD and the risk of breast cancer, re-evaluated the role played by each breast tissue component in susceptibility to develop cancer instead of only the epithelium.7-10

Cholesterol is an essential structural component of cell membranes, where it cooperates in regulating intracellular trafficking and signaling. Besides, it serves as the precursor for important biomolecules such as steroid hormones and isoprenoids. Because most ingested cholesterol is esterified in the liver and poorly adsorbed, actively proliferating cells respond to the increased need for cholesterol by increasing its de novo biosynthesis.

Previous studies demonstrated that genes coding for the enzymes involved in the essential steps of the de novo cholesterol biosynthesis were overexpressed both in preneoplastic and neoplastic lesions,11,12 and that, in postmenopausal women with estrogen receptor-positive breast cancer, the overexpression of these genes was associated with resistance to endocrine therapy.13

The present study aimed to investigate the expression of the genes involved in the de novo cholesterol biosynthesis in tissue samples from breasts radiologically classified as dense and non-dense, and according to their specific components (epithelium, fatty and non-fatty stroma), evaluated by digital image analysis of the histologic tissue sections. Then, the gene expression profile of the samples with a high non-fatty stroma component was compared with that of the samples from breasts radiologically classified as dense that are expected to be associated with a higher risk of cancer development.

**METHODS**

**Samples**

The study used a publicly accessible dataset from the NCBI Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo/), identified by the GEO accession number GSE49175, the only dataset responding to the specific requirement of available transcriptome data for histologically normal tissue samples with measured mammographic density.

As described in the original article,14 the dataset consisted of 120 snap-frozen samples of normal breast tissue collected at the time of breast surgery from women of ages 20 to 74 years with newly diagnosed in situ or invasive breast carcinoma and associated with a mammographic density measurement of the unaffected breast taken previously. All the participants provided written informed consent under a protocol approved by the U.S. National Cancer Institute and local (Polish) Institutional Review Boards.

**Mammographic density measurement**

The percentage mammographic density was calculated by dividing the absolute dense area by the total breast area multiplied by 100. If the percentage value was less than 25, the breast was classified as non-dense, and if it was 25 or more, as dense.

**Breast tissue composition measurement**

The tissue composition of the samples, measured by digital image analysis, was expressed as the percentage of epithelium, fatty, and non-fatty stroma. The samples were then categorized in tertiles according to the following cutoff points: 7% and 16% for the epithelium, 11% and 34% for the non-fatty stroma, and 47% and 80% for the fatty stroma.15

**Microarray data**

The complete transcriptome of snap-frozen samples was obtained using the Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F (Feature Number version) platform (GEO accession GPL4133) and the Stratagene Universal Human Reference. The expression estimates, filtered and lowess-normalized, were uploaded in GEO database.

**Gene Selection**

Seven genes were selected for the study, six of which, i.e., HMGCR (3-hydroxy-3-methylglutaryl-coenzyme A [HMG-CoA] reductase), FDPS (farnesyl diphosphate synthase), FDFT1 (farnesyl-diphosphate farnesyltransferase 1), GGPS1 (geranylgeranyl diphosphate synthase 1), SQLE (squalene epoxidase), and LSS (lanosterol synthase), code for the enzymes that play an essential role in cholesterol biosynthesis, and one (SREBF2, sterol regulatory element binding transcription factor 2) codes for the transcription factor that regulates the expression of HMGCR and SQLE (Suppl. Figure 1).
Statistical Analysis

The Shapiro-Wilk test, used to check the normality of the distribution of the gene expression values, indicated that not all the genes were normally distributed, except for SREBF2. Therefore, the median value and the inter-quartile range (IQR) were used to describe the expression of the genes and non-parametric tests were applied. Accordingly, the differential expression of the genes between dense and non-dense subgroups was assessed using the unpaired two-sample Wilcoxon test, while the Kruskal-Wallis test was used to evaluate the differential expression of the genes as a function of the epithelium, fatty or non-fatty stroma content. Spearman’s correlation coefficient was calculated to assess the association between the genes. The analyses were performed using the open-source software R Core Team version 4.1.2 (http://www.R-project.org), and the P-value <0.05 was considered statistically significant.

RESULTS

Differential expression of the genes involved in cholesterol biosynthesis according to mammographic density

According to the cutoff defined in the original study,14 56 (47%) breasts were classified as radiologically non-dense, and 64 (53%) as radiologically dense. However, since in a preliminary analysis, two samples (one in each subgroup) showed expression values considered as extreme outliers in most gene distribution, they were excluded from the subsequent statistical analyses.

The unpaired two-sample Wilcoxon test showed that the expression of HMGCR, FDPS, SQLE and SREBF2 was significantly higher in the dense than non-dense subgroup (respectively, -1.70 vs. -1.41, P=0.0028; -1.20 vs. -1.11, P=0.0501; -3.63 vs. -3.31 P=0.0003; -0.92 vs. -0.76, P=0.0271), and the correlation analysis indicated that the positive correlation between SREBF2 and HMGCR or SQLE found in non-dense breast subgroup (SREBF2*HMGCR: r=0.27, P=0.0495; SREBF2*SQLE: r=0.48, P=0.0002), substantially increased in the dense breast subgroup (SREBF2*HMGCR: r=0.44, P=0.0004; SREBF2*SQLE: r=0.74, P<0.0001) (Figure 1).

As shown in Table 1 and Figure 2, when the samples from non-dense and dense breasts were categorized based on their fatty component, a progressive decrease in the expression of HMGCR, SQLE, and SREBF2 was found in both density subgroups following an increase in the fatty content. Specifically, with the increase in fatty content, the median value of HMGCR decreased from -1.31 (0.53) to -1.94 (0.41) (P= 0.0057) in the non-dense breast and from -1.25 (0.47) to -1.93 (0.60) (P=0.0036), the median value of SQLE decreased from -3.33 (0.47) to -4.30 (0.76) (P=0.0001) in the non-dense breast and from -3.20 (0.97) to -4.02 (0.80) (P=0.0042), and the median value of SREBF2 decreased from -0.68 (0.24) to -0.98 (0.26) (P=0.0059) in the non-dense breast and from -0.72 (0.31) to -0.87 (0.40) (P=0.0069) in the dense breast.

Conversely, the expression of LSS significantly increased in the samples with the highest fatty content, with the median value increasing from -0.47 (0.60) to 0.09 (0.85) (P=0.0002) in the non-dense breast and from -0.57 (0.32) to 0.17 (0.53) (P=0.0042).

An opposite trend was found when the samples were categorized based on the amount of the stroma component: the expression of HMGCR, SQLE, and SREBF2 progressively increased with an increase in the stroma content, while the expression of LSS decreased (Figure 3). Specifically, with an increase in the stromal component, the median value of HMGCR increased from -1.95 (0.36) to -1.28 (0.46) (P=...
0.0014) in the non-dense breast and from -1.89 (0.51) to -1.27 (0.45) (P=0.0114), the median value of SREBF2 increased from -4.29 (0.68) to -3.14 (0.97) (P=0.0001) in the non-dense breast and from -3.92 (0.90) to -3.20 (0.89) (P=0.0262), and the median value of SREBF2 increased from -0.98 (0.29) to -0.56 (0.23) (P=0.0107) in the non-dense breast and from -0.84 (0.43) to -0.76 (0.34) (P=0.0595) in the dense breast. Conversely, the median value of LSS decreased from 0.31 (0.74) to -0.64 (0.41) (P=0.0002) in the non-dense breast and from 0.06 (0.72) to -0.53 (0.54) (P=0.0250) in the dense breast.

Table 1. Comparison by Kruskal-Wallis test of the median gene expression in samples from non-dense and dense breasts categorized in tertiles according to their tissue composition

<table>
<thead>
<tr>
<th>Gene</th>
<th>Non-dense breast</th>
<th>Dense breast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fat component</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≤ 47% (N=8)</td>
<td>48-80% (N=15)</td>
</tr>
<tr>
<td>HMGR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.31(0.54)*</td>
<td>-1.59(0.50)</td>
</tr>
<tr>
<td>FDPS</td>
<td>-1.14(0.19)</td>
<td>-1.21(0.20)</td>
</tr>
<tr>
<td>FDFT1</td>
<td>-1.19(0.32)</td>
<td>-1.48(0.36)</td>
</tr>
<tr>
<td>GGPS1</td>
<td>0.23(0.35)</td>
<td>0.17(0.31)</td>
</tr>
<tr>
<td>SREBF2</td>
<td>-0.68(0.24)</td>
<td>-0.75(0.24)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Stroma component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤ 11% (N=22)</td>
</tr>
<tr>
<td>HMGR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.95(0.36)</td>
</tr>
<tr>
<td>FDPS</td>
<td>-1.18(0.40)</td>
</tr>
<tr>
<td>FDFT1</td>
<td>-1.03(0.43)</td>
</tr>
<tr>
<td>GGPS1</td>
<td>0.24(0.26)</td>
</tr>
<tr>
<td>SREBF2</td>
<td>-0.98(0.29)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Epithelium component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤ 7% (N=15)</td>
</tr>
<tr>
<td>HMGR</td>
<td>-1.84(0.47)</td>
</tr>
<tr>
<td>FDPS</td>
<td>-1.14(0.21)</td>
</tr>
<tr>
<td>FDFT1</td>
<td>-1.01(0.33)</td>
</tr>
<tr>
<td>GGPS1</td>
<td>0.29(0.22)</td>
</tr>
<tr>
<td>SREBF2</td>
<td>-4.35(1.41)</td>
</tr>
<tr>
<td>LSS</td>
<td>0.21(0.82)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Epithelium component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤ 7% (N=20)</td>
</tr>
<tr>
<td>HMGR</td>
<td>-1.37(0.60)</td>
</tr>
<tr>
<td>FDPS</td>
<td>-1.09(0.19)</td>
</tr>
<tr>
<td>FDFT1</td>
<td>-1.28(0.59)</td>
</tr>
<tr>
<td>GGPS1</td>
<td>0.37(0.22)</td>
</tr>
<tr>
<td>SREBF2</td>
<td>-3.46(0.64)</td>
</tr>
<tr>
<td>LSS</td>
<td>-0.45(0.64)</td>
</tr>
</tbody>
</table>

*Inter-quartile range
Figure 2. Expression of \textit{HMGCR}, \textit{SQLE}, \textit{LSS}, and \textit{SREBF2} genes in samples from non-dense and dense breasts as a function of their fatty content.

Figure 3. Expression of \textit{HMGCR}, \textit{SQLE}, \textit{LSS}, and \textit{SREBF2} genes in samples from non-dense and dense breasts as a function of their stroma content.
**Differential expression of the genes involved in cholesterol biosynthesis as a function of breast tissue composition, regardless of the radiological classification**

According to the radiological criterion, a breast is classified as dense when the epithelial plus non-fatty stroma compartment is ≥ 25% of the total area. Otherwise, the sample is classified as non-dense. In line with this assumption, the samples were reclassified based on the percentage of the epithelial, fatty and non-fatty stroma components evaluated by digital image analysis. Consequently, samples with a non-fatty stroma content < 11% (I tertiles) and an epithelium content < 16% (I and II tertiles) were reclassified as “highly fatty”, whereas samples with a non-fatty stroma content > 11% (II and III tertiles) and an epithelial content > 7% (II and III tertiles) were reclassified as “highly non-fatty”. According to this new criterion, 35% of the samples were highly fatty, and 65% were highly non-fatty.

All but one (97%) of the highly fatty samples had a fatty content > 80%, but only 67% had been classified as radiologically non-dense. Similarly, only 57% of highly non-fatty samples had been classified as radiologically dense despite a fat content < 80% in 95% of cases (Figure 4).

**Figure 4.** Frequency distribution of highly fatty and highly non-fatty samples according to their fat, stroma, and epithelium content.

Statistical analysis showed that five genes were differentially expressed when highly non-fatty and highly fatty subgroups were compared. *HMGCR*, *SQLE*, and *SREBF2* were more expressed in the highly non-fatty subgroup (respectively, -1.48 vs. -1.94, P<0.0001; -3.39 vs. -4.18, P=0.0001; -0.77 vs. -0.94, P=0.0103), while *FDFT1* and *LSS* were more expressed in highly fatty one (0.28 vs. -0.60, P<0.0001 and -1.05 vs. -1.32, P=0.0459, respectively). Furthermore, the correlation analysis indicated that while in the highly non-fatty subgroup, *SREBF2* was positively associated with both *HMGCR* (r=0.53, P<0.0001) and *SQLE* (r=0.73, P<0.0001), in the highly fatty subgroup, these positive correlations disappeared (*SREBF2*|*HMGCR*: r=-0.19, P=0.3026) or substantially decreased (*SREBF2*|*SQLE*: r=0.41, P=0.0173) (Figure 5).

**DISCUSSION**

The study showed that the genes *HMGCR*, *SQLE*, and *SREBF2* were more frequently expressed in the breasts radiologically classified as dense because of a mammographic density > 25% and that this differential expression was associated with the non-fatty stroma component and not as hypothesized by former studies by the overproliferation of epithelial cells. Indeed, dense and non-dense breasts showed no statistically significant distribution in the class at low (43% vs. 31%, respectively), moderate (37% vs. 35%, respectively), and high (20 vs. 33%, respectively) epithelium content.

The evidence that in dense breasts, *HMGCR*, *SQLE*, and *SREBF2* were overexpressed is of great relevance considering the essential role played by these genes in the biosynthesis of cholesterol, where HMG-CoA reductase coded by *HMGCR* governs the first rate-limiting step, squalene epoxidase coded by *SQLE* regulates the second rate-limiting and irreversible commitment step toward cholesterol, and the SREBP transcription factor coded by *SREBF2* controls the expression of *HMGCR* and *SQLE*. Noteworthy, the overexpression of these genes is associated with a substantial increase in the positive correlation of *SREBF2* with both genes.
Mammographic density and cholesterol biosynthesis

Figure 5. Differential expression of HMGCR, SQLE, SREBF2 (upper panel), FDFT1, and LSS (lower panel), and correlation between SREBF2 and HMGCR or SQLE (middle panel) in highly fatty and highly non-fatty subgroups.

To confirm that the increased expression of HMGCR, SQLE, and SREBF2 in dense breasts was associated with a high presence of non-fatty stroma, the tissue samples were reclassified based on digital image analysis into highly fatty and highly non-fatty subgroups, regardless of their radiological classification. The results showed that highly non-fatty samples expressed significantly high levels of HMGCR, SQLE, and SREBF2 despite the evidence that little more than half of them (56%) came from breasts radiologically classified as dense.

The incomplete correspondence between a high mammographic density and a highly non-fatty content can be explained by the fact that, as described in the original article, the digital evaluation of tissue composition was performed on tissue sections collected at the time of breast surgery, whereas the mammographic density was measured pre-surgery on the unaffected breast. Nevertheless, the increased expression of HMGCR, SQLE, and SREBF2 in highly non-fatty samples corroborates the hypothesis that the overexpression of these genes, found in dense breasts, is due to the high percentage of non-fatty stroma.

The results also showed a significant decrease in the expression level of the LSS gene in highly non-fatty samples. This finding is of interest because LSS codes for lanosterol synthase, which acts as negative feedback on the expression of the upstream SQLE. Considered jointly with the decline in the negative correlation between LSS and SQLE (r= -0.42, P=0.0146), the decrease in LSS expression suggests that, in highly non-fatty samples, the overexpression of HMGCR and SQLE could be the combined effect of inadequate negative feedback of LSS on SQLE expression and the increased expression of SREBF2 which promotes HMGCR and SQLE transcription.

CONCLUSION

Altogether, the present findings suggest that non-fatty stroma can contribute to the development of breast cancer by promoting the epithelial cells growth not only through the recognized paracrine production of growth factors, but also by increasing the local production of cholesterol and its derivatives, especially estrogens, which can stimulate the proliferation of estrogen receptor-positive epithelial cells. Furthermore, the findings explain why women with radiologically dense breasts have a higher risk of developing breast cancer and support the evaluation of mammographic density as an excellent surrogate marker to identify women who may benefit from preventive strategies to reduce cholesterol biosynthesis. One such strategy is the use of statins, which are already in use to reduce breast cancer recurrence and mortality, while several inhibitors of cholesterol biosynthesis, such as allylamines, squalene analogs, natural compounds of selenium and tellurium, whose primary target is squalene synthase, and lapatinib acetate, which acts on lanosterol synthase, are being studied as potential alternatives to statins.

ETHICAL CONSIDERATIONS

All patients consented to provide excess tissues for research purposes, and the study was approved by the U.S. National Cancer Institute and local (Polish) Institutional Review Boards.

ACKNOWLEDGEMENTS

I thank Prof. Federico Ambrogi for the stimulating discussion and useful suggestions.

CONFLICTS OF INTEREST

The author affirms the absence of any conflicts of interest, including both financial and personal relationships with individuals or organizations that could potentially exert undue influence on the study.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sector.
DATA AVAILABILITY
The study used a publicly accessible dataset from the NCBI Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo/), identified by the GEO accession number GSE49175.

REFERENCES