GATA5 Activation of the Progesterone Receptor Gene Promoter in Breast Cancer Cells Is Influenced by the +331G/A Polymorphism

Gordon S. Huggins, Jason Y.Y. Wong, Susan E. Hankinson, and Immaculata De Vivo

Abstract

Previously, a modest association was observed between the progesterone receptor +331G/A gene variant and breast cancer risk. Here, in a larger sample of breast cancer cases and controls (n = 1,322/n = 1,953) nested in the Nurses’ Health Study cohort, we confirm a significant association (odds ratio, 1.41; 95% confidence interval, 1.10-1.79) and suggest a molecular model. The association of the +331G/A variant with breast cancer was particularly strong among obese women (body mass index > 30; odds ratio, 2.87; 95% confidence interval, 1.40-5.90). To help understand the molecular mechanism by which this variant may predispose women to breast cancer, we identified nearby transcription factor binding sites. This search predicted a binding site for the GATA family of transcriptional regulators adjacent to this hPR polymorphism. Importantly, we found GATA3, GATA4, and GATA6 are expressed in normal breast tissue and two breast cancer cell lines, whereas GATA5 is minimally expressed in normal mammary tissue and more strongly expressed in two breast cancer cell lines. This finding was relevant because GATA5 bound the site adjacent to the +331G/A polymorphism, and activated the hPR (~711 to +822)-luciferase reporter plasmid in breast cancer cells. Overexpression of GATA5 increased expression of the endogenous hPR transcript, and GATA5 more strongly activated an hPR promoter construct encoding the PR-B isoform. Finally, hPR promoter constructs including the +331A were more strongly activated by GATA5 than constructs including +331G. Our findings suggest that GATA5 interacts with the +331G/A polymorphism to stimulate hPR-B expression in mammary cells, which may contribute to breast cancer susceptibility. (Cancer Res 2006; 66(3): 1384-90)

Introduction

Progesterone stimulation of its receptor is required for breast development (1) and influences the risk of breast cancer (2). The physiologic effects of progesterone are entirely mediated through the progesterone receptors, PR-A and PR-B, which are products of the same gene through alternative transcriptional initiation. Expression of PR-A and PR-B is regulated in a tissue-specific manner (3), and the balance of PR isoform expression has been considered an important factor contributing to the risk of malignancy (4). PR-B activates transcription more robustly than PR-A (5), and the PR isoforms target unique downstream gene transcriptional targets (6).

The progesterone receptor gene is dynamically expressed in epithelial and stromal breast tissue during sexual development and pregnancy (7). Mice lacking both progesterone receptors have markedly abnormal breast development (8). Although breast development and the response of mammary tissue to progesterone were not altered by the genetic deletion of murine PR-A (9), deletion of PR-B markedly reduced pregnancy-associated ductal and alveolar epithelial cell proliferation (10). Because PR-B is necessary for mammary cell proliferation, we reason that genetic mechanisms that influence PR-B expression may contribute to breast malignancy. Although progesterone receptor gene expression is responsive to estrogen, neither estrogen nor its receptors are required for PR expression in hormonally responsive tissues. In fact, transcriptional regulators that control isoform-specific PR expression in normal mammary and breast cancer cells are poorly understood.

The family of GATA transcription factors regulates cellular differentiation in many tissues (11). In fact, the GATA proteins contribute to gender-specific development and function through their regulation of steroidogenesis genes, including P450 aromatase (Cyp19), Cyp17, and Cyp11A (12). Furthermore, GATA proteins regulate expression of the gonadal hormones inhibin-α and müllrian inhibiting substance (13, 14). GATA3 is required for embryonic neural (15) and T cell differentiation (16), and it is expressed in breast cancer cells (17, 18). GATA4, GATA5, and GATA6 are expressed in mesodermal tissues. GATA4 and GATA6 are necessary for the early stages of mesodermal tissue differentiation and null-mice die during embryonic development (19). By comparison, GATA5-deficient mice are viable with a limited phenotype of abnormal female genital tract morphogenesis (20). Male GATA5-deficient mice are normal. The requirement for GATA5 in the genital development of female mice implies that GATA5 has direct or indirect gender-specific effects that may also regulate breast tissue.

Previously, we identified the hPR +331G/A variant, located between the PR-B (nucleotide 0) and PR-A (nucleotide +751) transcriptional start sites. In the Nurses’ Health Study, the +331G/A variant, found in ~ 10% of subjects, was positively associated with the risk of endometrial cancer (21). Moreover, the risk of endometrial cancer was significantly modified by obesity. Biochemical studies showed that the +331A polymorphism increased promoter activity including expression of PR-B in endometrial cancer cells (21). Previously, we also reported a modest association of the hPR +331G/A variant with breast cancer risk in the Nurse’s Health Study cohort (22). Using a larger number of cases, we now report a significant association between the +331G/A variant with breast cancer risk. Furthermore, we show an interaction of the +331G/A variant with GATA5, which may underlie the genetic mechanism by which this variant contributes to breast cancer risk.
Terms in unconditional logistic regression models. The likelihood ratio test was used to assess the statistical significance of these interactions (nominal likelihood ratio test). We used SAS version 8.0 (SAS Institute, Cary, NC) for all analysis. We tested Hardy-Weinberg agreement by using a χ² test.

### Transcription factor binding search
GATA transcription factor binding sites were identified by using ALIBABA, ver. 2.1 5 to query the TRANSFAC 4.0 database.

### Reverse transcriptase PCR
Total RNA from human mammary adenocarcinoma cell lines MCF-7 and MDA-MB-468 was extracted using the RNeasy Midiprep Kit (Qiagen, Valencia, CA) and DNase I treated following the manufacturer's protocol. Human mammary tissue total RNA was purchased from Clontech (Mountain View, CA). The reverse transcription reaction was done using the One-Step reverse transcription-PCR (RT-PCR) Kit (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. GATA transcription factor expression was analyzed using intron-spanning primers specific for GATA3 (TGAGGAGGAATGCAATGG and TCTCGGTCTTGCTGATGC), GATA4 (CAGCAAAAAGAGGCAG and AGACATGCCTAGCTACTGAGAAC), GATA5 (GAAGAACCGGGAAC), GATA6 (TGGATTGTGCACCTTTTC), and GATA6 (TGGATTGTGCACCTTTTC). The β-actin transcript was amplified using a published protocol (25). The RT-PCR products were resolved on a 2% MetaPhor Agarose gel (Bio Whittaker, Baltimore, MD).

### Electromobility shift assay
Recombinant GATA5 protein expressed in reticulocyte lysate from the pcDNA-GATA5 expression plasmid was mixed with 50,000 cpm [³²P]-labeled aagcttAGTCGGGAGATAAAG-GAGGGCCGTca (in a binding buffer, Pierce, Rockford, IL) that included 100 mmol/L KCl, 2.5% glycerol, 1 mmol/L DTT, and 1 mmol/L EDTA. After 20 minutes, the sample was separated on a 5% acrylamide gel, which was then dried and exposed to Kodak BioMax MR film. The specificity of DNA-protein complexes was determined using a competitive inhibitor oligonucleotide, including a GATA binding site from the atrial natriuretic factor gene promoter or a noncompetitive oligonucleotide lacking a GATA binding site.

### Table 1. Association between the +331 polymorphism and breast cancer risk in the Nurses' Health Study (1989-2000)

<table>
<thead>
<tr>
<th>PR Genotype*</th>
<th>Cases, n (%)</th>
<th>Controls, n (%)</th>
<th>OR</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>+331 G/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td>1,134 (87)</td>
<td>1,560 (90)</td>
<td>1.0&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1.0&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>vt</td>
<td>164 (13)</td>
<td>168 (10)</td>
<td>1.37 (1.08-1.74)</td>
<td>1.41 (1.10-1.79)</td>
</tr>
<tr>
<td>Premenopausal women</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+331 G/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td>101 (91)</td>
<td>102 (89)</td>
<td>1.0&lt;sup&gt;3&lt;/sup&gt;</td>
<td>1.0&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>vt</td>
<td>10 (9)</td>
<td>13 (11)</td>
<td>0.78 (0.33-1.89)</td>
<td>0.82 (0.32-2.14)</td>
</tr>
<tr>
<td>Postmenopausal women</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+331 G/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td>964 (87)</td>
<td>1,381 (90)</td>
<td>1.42 (1.11-1.81)</td>
<td>1.48 (1.16-1.90)</td>
</tr>
<tr>
<td>vt</td>
<td>146 (13)</td>
<td>148 (10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Numbers may vary because of missing genotypes.

<sup>1</sup>Conditional logistic regression adjusted for the matching variables: age, menopausal status, postmenopausal hormone use at blood draw, date at blood draw, and time at blood draw.

<sup>2</sup>Conditional logistic regression adjusted for matching variables and age at menarche, parity, age at first birth, BMI at age 18, weight gain since age 18, benign breast disease, and first-degree family history of breast cancer.

<sup>3</sup>Unconditional logistic regression adjusted for matching variables: age, date at blood draw, time at blood draw, and fasting status.

<sup>4</sup>Unconditional logistic regression adjusted for matching variables in c and age at menarche, parity, age at first birth, BMI at age 18, weight gain since age 18, benign breast disease, first-degree family history of breast cancer and duration of postmenopausal hormone use.

<sup>5</sup>Unconditional logistic regression adjusted for matching variables: age, date at blood draw, time at blood draw and fasting status.

<sup>6</sup>Unconditional logistic regression adjusted for matching variables: age, menopausal status, postmenopausal hormone use at blood draw, date at blood draw, time at blood draw along with age at menarche, parity, age at first birth, BMI at age 18, weight gain since age 18, benign breast disease, first-degree family history of breast cancer, and duration of postmenopausal hormone use.

---

### Materials and Methods

**Study population.** The Nurses' Health Study was initiated in 1976, when 121,700 U.S. registered nurses between the ages of 30 and 55 returned an initial questionnaire reporting medical histories and baseline health-related exposures. Between 1989 and 1990, blood samples were collected from 32,826 women. Incident breast cancers were identified by self-report and confirmed by medical record review. Eligible cases in this study consisted of women diagnosed with pathologically confirmed incident breast cancer after giving a blood specimen up to June 1, 2000. The nested case-control study consists of 1,322 incident breast cancer cases and 1,953 controls matched to cases on year of birth, menopausal status, and postmenopausal hormone use as described previously (23, 24). The protocol was approved by the Committee on Human Subjects, Brigham and Women's Hospital.

**Materials and Methods.**

#### Statistical analysis.
Student's t test and χ² test were used to evaluate differences in breast cancer risk factors between cases and controls. Odds ratios (OR) and 95% confidence intervals (CI) were calculated by using conditional and unconditional logistic regression. In addition to the matching variables, we adjusted for a number of breast cancer risk factors created by using the wild-type genotype as the reference category in conditional and unconditional logistic regression. In addition to the matching variables, we adjusted for a number of breast cancer risk factors.

#### Electromobility shift assay.
Recombinant GATA5 protein expressed in reticulocyte lysate from the pcDNA-GATA5 expression plasmid was mixed with 50,000 cpm [³²P]-labeled aagcttAGTCGGGAGATAAAG-GAGGGCCGTca (in a binding buffer, Pierce, Rockford, IL) that included 100 mmol/L KCl, 2.5% glycerol, 1 mmol/L DTT, and 1 mmol/L EDTA. After 20 minutes, the sample was separated on a 5% acrylamide gel, which was then dried and exposed to Kodak BioMax MR film. The specificity of DNA-protein complexes was determined using a competitive inhibitor oligonucleotide, including a GATA binding site from the atrial natriuretic factor gene promoter or a noncompetitive oligonucleotide lacking a GATA binding site.

---

Cloning. Fragments of the hPR promoter containing the wild-type +331G or variant +331A were amplified from genomic DNA using custom-designed primers, cloned into pCR4 (Invitrogen), and then sequenced. The hPR promoter fragments, and those lacking the isoform A and B transcriptional start sites, were then subcloned into pGLO2-Basic (Promega, Madison, WI) using restriction digestion techniques.

Luciferase reporter assays. MDA-MB-468 cells (courtesy of Dr. R. Parsons, The Institute of Cancer Genetics, Columbia University, New York, NY) were cultured in DMEM supplemented with 10% fetal bovine serum/100 units/mL of penicillin/100 µg/mL of streptomycin in a humidified atmosphere at 37°C with 5% CO2. Twenty-four hours after plating in 12-well dishes (∼115,000 cells/well), 200 ng of luciferase reporter plasmid was cotransfected with 10 ng CMV-β-galactosidase activity, which served as an internal control of transfectional efficiency, and 400 ng of vector control (pCR3, Invitrogen), pCDNA3-mGATA3 (courtesy of Dr. I-Cheng Ho, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA), pCDNA3-mGATA4, pCDNA3-mGATA5, or pCDNA3-mGATA6 (courtesy Dr. Michael Parmacek, Division of Cardiovascular Medicine, University of Pennsylvania, Philadelphia, PA) expression plasmids using Polyfect (Qiagen) reagent. Twenty-four hours following transfection, luciferase and β-galactosidase activity were measured for each sample prepared in reporter lysis buffer (Promega) and the Galacto-Light kits (Applied Biosystems), and a luminometer (Thermo Electron Luminoskan Ascent, Milford, MA). Luciferase activity was normalized to β-galactosidase activity for each sample. Each sample was done in triplicate and normalized to the average luciferase activity of the vector control. Differences in the mean and SE of results from three independent experiments were determined by factorial ANOVA with the STATVIEW program; P < 0.05 was considered significant.

Real-time quantitative reverse transcription-PCR. Approximately 3 × 10^6 MDA-MB-468 cells were split into six-well plates and grown until 85% confluent. Each well was transfected with 300 ng of pcDNA-GATA5 or PC3 using the Qiagen Polyfect reagent. After 24 hours, the cells were harvested and the total RNA was isolated with the Qiagen RNeasy Mini Kit following the manufacturer’s instructions. The total RNA concentration and quality was measured with a Nanodrop Technologies ND-1000 spectrophotometer and by denaturing agarose gel electrophoresis. Real-time reverse transcription-PCR was done with the one-step real-time qRT-PCR supermix (Invitrogen) using template from 200 ng of RNA per reaction and hPR (GGGCGCTTACCTGGACGTCC and TGAATCCGGGCTTCGAGGATG) and RPLP0 (26) gene-specific primers. A four-point standard curve shows a linear slope of −3.3 for both reactions with correlation coefficients (r²) for hPR and RPLP0 of 0.971 and 0.998, respectively. Relative Quantification of the hPR transcript normalized to RPLP0 (36B1) was assessed with the ABI 7300 real-time PCR Systems RQ Study Software using the comparative Ct method.

Results

Progestosterone receptor +331G/A gene variant is associated with breast cancer risk. A modest association between the +331A allele and the risk of breast cancer has been reported (22). To achieve greater statistical power, we genotyped all additional 308 cases and matched controls to confirm this finding. The mean age of cases at blood draw was 57.2 years; for controls, it was 57.9 years. Cases and controls had similar mean body mass index (BMI) at blood draw (25.4 versus 25.5 kg/m²) and weight gain since age 18 years (11.6 versus 11.4 kg). Compared with controls, cases had similar ages at menarche (12.5 versus 12.6 years), first birth (23.0 versus 23.0 years), and age at menopause (48.2 versus 48.0 years). Compared with controls, cases were more likely to have a first-degree family history of breast cancer (21.0% versus 15.0%) and a history of benign breast disease (64.0% versus 51.0%). The prevalence of the variant carriers was similar to a previous report for Caucasian women (22), 13% for the cases and 10% for the controls. The genotype distribution of the +331 polymorphism among the cases and controls was in Hardy-Weinberg equilibrium. There was a modest association between the +331 variant and breast cancer risk among the most recent 308 cases and matched controls (OR, 1.74; 0.99-3.00; data not shown). To give the best estimate, we combined this data with our previous data for a total of 1,134 cases and found that the risk within the Nurses’ Health Study was similar, but the strength of the association was now significant (OR, 1.41; 1.10-1.79; Table 1). Furthermore, we observe the strongest association of the +331G/A variant with breast cancer in postmenopausal women and in women who were obese (>30 kg/m²; Table 2).

The +331G/A polymorphism is adjacent to a GATA protein binding site. Because of the consistent association of the +331G/A polymorphism with breast cancer risk, we sought to understand how this variant might contribute to progesterone receptor expression by searching for nearby binding sites for transcription factors relevant to breast cancer. Interestingly, a binding site for the GATA family of transcription factors was predicted immediately upstream of the +331G/A polymorphism (Fig. 1A). To determine the expression pattern of GATA transcripts in human breast tissue (Clontech) and two human breast cancer cell lines, we did RT-PCR. GATA3 and GATA6 transcripts were detected in breast tissue and cultured MDA-MB-468 and MCF-7 breast cancer cells (Fig. 1B). Interestingly, the GATA5 transcript abundance and percentile rank order are reported to be higher in MDA-MB-436

| Table 2. +331 polymorphism and BMI for postmenopausal women: frequencies and ORs for breast cancer risk |
|--------------|-----------------|---------------|-----------------|---------------|
| Genotype     | BMI < 25 kg/m², n (%) | 25 ≤ BMI ≤ 30 kg/m², n (%) | BMI > 30 kg/m², n (%) |
|             | Cases | Controls | OR (95% CI) | Cases | Controls | OR (95% CI) | Cases | Controls | OR (95% CI) |
| +331 wt⁡     | 529 (89) | 708 (89) | 1.0 | 296 (85) | 453 (91) | 1.04 (0.83-1.30) | 140 (85) | 220 (93) | 1.23 (0.84-1.81) |
| +331 vt⁠     | 68 (11) | 85 (11) | 1.14 (0.81-1.61) | 53 (15) | 46 (9) | 1.91 (1.23-2.95) | 25 (15) | 17 (7) | 2.87 (1.40-5.90) |

NOTE: Unconditional logistic regression adjusted for matching variables and age at menarche, parity, age at first birth, BMI at age 18, weight gain since age 18, benign breast disease, first-degree family history of breast cancer. ORs are independent risks for breast cancer risk following the joint variables were used: +331 wt with BMI (25-30) kg/m², +331 wt with BMI > 30 kg/m², +331 vt with BMI < 25 kg/m², +331 vt with BMI (25-30) kg/m², +331 vt with BMI > 30 kg/m², compared with +331 wt with BMI < 25 kg/m² (reference group).

P for interaction is 0.05.

*Reference category.

Includes heterozygotes.
(ATCC no. HTB-130) cells compared with normal mammary tissue by Entrez GEO (GDS823). We determined that the GATA5 transcript is more abundant in MDA-MB-468 (ATCC no. HTB-132) and MCF-7 cells compared with normal mammary tissue (Fig. 1B). By comparison, GATA4 expression was lower in MCF-7 cells. To determine if GATA5 bound the site adjacent to the +331G/A variant, we incubated a radiolabeled oligonucleotide probe including the GATA binding site with recombinant GATA5 protein expressed in vitro. Recombinant GATA5 retarded the mobility of the radiolabeled oligonucleotide. Formation of this complex was inhibited by a cold oligonucleotide, including a GATA binding site from the atrial natriuretic factor gene and not by a nonspecific oligonucleotide competitor. A radiolabeled oligonucleotide identical to this probe, except including the +331A variant, was also bound by GATA5, suggesting that the +331G/A variant does not abolish GATA5 binding (data not shown). The presence of a GATA5 binding site in the progesterone receptor promoter suggests that GATA5 may regulate promoter activity.

GATA5 activates the human progesterone receptor gene promoter. To understand the significance of the GATA proteins to progesterone receptor gene expression, we expressed GATA proteins in MDA-MB-468 cells with an hPR (−711 to +822)-luciferase reporter plasmid. Although GATA3 and GATA6 expression did not influence hPR activity compared with the vector control, GATA4 and GATA5 significantly increased the activity of the reporter plasmid. Columns, mean of corrected luciferase activity normalized to vector control from three transfections done in triplicate (n = 9); bars, ±SE; *, P < 0.05 compared with vector control. B, GATA5 overexpression increases endogenous hPR transcript abundance. Following transient plasmid transfection of MDA-MB-468 cells with pcDNA3-GATA5 or vector control, hPR and RPLP0 transcript abundance was measured by quantitative real-time reverse transcription-PCR. Columns, mean of relative quantification of hPR normalized to RPLP0 of 18 reactions from three experiments; bars, ±SE; *, P < 0.05 compared with vector control.

Figure 1. A, location of a predicted GATA-protein binding site adjacent to the +331G/A polymorphism in the progesterone receptor gene. B, GATA5 is more strongly expressed in two breast cancer cell lines. GATA and β-actin transcripts were amplified by PCR following cDNA synthesis from normal breast (top), cultured MCF-7 (middle), and MDA-MB-468 (bottom) cell total RNA. C, GATA5 binds the predicted GATA recognition site adjacent to the +331G/A polymorphism. The [32P]-labeled oligonucleotide, shown in (A), was retained by GATA5 (arrow) in the absence of an inhibitor (−) and in the presence of a cold noncompetitive inhibitor (NCI). Inclusion of an unlabeled competitive inhibitor (CI) oligonucleotide, which contains a known GATA binding site, blocked complex formation. The free probe (FP) is shown at the bottom.

Figure 2. A, GATA5 increases hPR promoter activity. MDA-MBA-468 cells were transfected with hPR (−711 to +822)-luc and GATA expression plasmids. Although GATA3 and GATA6 expression did not influence hPR activity compared with the vector control, GATA4 and GATA5 significantly increased the activity of the reporter plasmid. Columns, mean of corrected luciferase activity normalized to vector control from three transfections done in triplicate (n = 9); bars, ±SE; *, P < 0.05 compared with vector control. B, GATA5 overexpression increases endogenous hPR transcript abundance. Following transient plasmid transfection of MDA-MB-468 cells with pcDNA3-GATA5 or vector control, hPR and RPLP0 transcript abundance was measured by quantitative real-time reverse transcription-PCR. Columns, mean of relative quantification of hPR normalized to RPLP0 of 18 reactions from three experiments; bars, ±SE; *, P < 0.05 compared with vector control.
both transcripts was within a linear range. We found that hPR was expressed significantly greater in MDA-MB-468 cells following GATA5 overexpression compared with vector control (Fig. 2B). The expression of GATA5 in breast cancer cells may therefore directly regulate progesterone receptor gene expression.

**GATA5 activates expression of hPR-B isoform.** Through the use of two different transcriptional start sites, the progesterone receptor gene expresses two isoforms (27). We created two hPR promoter deletion constructs that lack either the hPR-A or hPR-B transcriptional start sites. Although GATA5 significantly activated the full-length hPR reporter plasmid (Fig. 3, top construct), it had almost no influence on the hPR reporter plasmid lacking the hPR-B transcriptional start site [middle construct, hPR (−711 to −108, +151 to +833)-luc]. By comparison, removal of sequences including the hPR-A transcriptional start site [bottom construct, hPR (−711 to +447)-luc] increased the response to GATA5 (Fig. 3). Importantly, this result shows that GATA5 has its greatest effect by augmenting expression of an hPR transcript that includes the hPR-B translational start site.

**GATA5 activates the +331A-containing hPR promoter more strongly.** Although the +331A variant does not alter binding of GATA5 to the adjacent hPR GATA binding site, we tested whether the variant altered GATA5-mediated induction of the hPR promoter. Indeed, the +331A variant-containing hPR (−711 to +822)-luc reporter plasmid was activated slightly, but significantly, more strongly by GATA5 compared with the +331G wild-type reporter plasmid (Fig. 4A). To explore the interaction of the GATA proteins with the +331G/A polymorphism more deeply, we also tested their effect on the hPR reporter plasmid lacking the PR-A TSS. The +331A variant-containing hPR (−711 to +447)-luc reporter plasmid was activated significantly more strongly by GATA5 compared with the +331G wild-type reporter plasmid in both MDA-MB-468 cells (data not shown) and Cos7 cells (Fig. 4B). By comparison, GATA4 and GATA6 had no significant effect on the hPR (−711 to +447)-luc reporter plasmid. These findings are consistent with our conclusion that GATA5-dependent regulation of the hPR gene is influenced by the +331G/A variant.

**Discussion**

Functional genetic variants that influence the estrogen and progesterone pathways affect hormone-responsive tissues and their respective malignancies (28, 29). In particular, progesterone and the progesterone receptor pathway have important contributions to the risk of both endometrial and breast cancer. Previously, we showed that the +331G/A polymorphism, lying between the PR-A and PR-B transcriptional start sites, has a direct effect on progesterone receptor promoter activity and isoform expression. In addition to endometrial cancer (21), we found a modest association between the +331G/A variant and breast cancer risk (22). Recently, the association of the +331G/A variant and breast cancer found in the Nurses’ Health Study has been challenged by two studies (30, 31). Two issues may contribute to the differences in our results compared with others. First, the role of ethnicity on the association of the +331G/A polymorphism with breast cancer has not been established. The substantial differences in the ethnic makeup of the Nurses’ Health Study, which is composed primarily of Caucasian women, and the large cohort reported by Gold et al., which has better representation of African-American and Ashkenazi Jewish subjects, might account for our different results. Second, we found the association between the +331G/A polymorphism with breast cancer to be modified by menopausal status (Table 1) and obesity (Table 2). Because Gold et al. could only consider age of >50 years as a surrogate for menopausal status, they had a limited ability to control for the effect of menopausal status. To help resolve this conflict, we chose to reanalyze the association of the +331G/A variant with breast cancer following the inclusion of 308 additional incident cases from the Nurses’ Health Study. The inclusion of additional cases strengthened rather than weakened the association of the +331G/A variant and breast cancer risk (Table 1).

The association of +331G/A with breast cancer was observed primarily in postmenopausal women, suggesting that variant-induced differences in hPR gene expression may influence cancer susceptibility in the setting of reduced hormone levels. This finding is not surprising because nuclear hormone receptors, including the progesterone receptors, could have transcriptional effects in the absence of ligand stimulation (5). In addition to menopausal status and ethnicity, differences between our findings and other published studies might reflect the role of modifiers. Indeed, the risk of breast cancer associated with the +331G/A variant was greater with increasing BMI. Obesity has long been associated with risk of postmenopausal breast cancer, likely in large part through the aromatization of androgens to estrogen in adipose tissue. By comparison, there is no alternative endogenous

**Figure 3.** GATA5 activates hPR-B-containing transcript. Cotransfection of pcDNA3-GATA5 significantly activated the full-length hPR (−711 to +822)-luc reporter plasmid ([top construct] compared with vector control. Deletion of the PR-A transcriptional start site (TSS) caused even greater activation by GATA5 (bottom construct), whereas GATA5 had almost no influence on the hPR reporter plasmid lacking the PR-B TSS (middle construct). Left, a cartoon illustrating each hPR promoter construct. Columns, fold activation relative to full-length hPR (−711 to +822)-luc reporter plasmid transfected with a vector control plasmid; *, P < 0.05 compared with vector control.
transfections done in triplicate (n = 9 for each condition); P*, Columns plasmid. more strongly than the corresponding +331G (open columns) hPR reporter of the +331A polymorphism. Transfection of pcDNA3-GATA5 activated the hPR (∼711 to +447)-luc reporter plasmid lacking the PR-A TSS in the presence of the +331A polymorphism. Transfection of pcDNA3-GATA5 activated the +331A (filled columns) hPR (∼711 to +447)-luc reporter plasmid significantly more strongly than the corresponding +331G (open columns) hPR reporter plasmid. Columns, mean of corrected luciferase activity from three separate transfections done in triplicate (n = 9 for each condition); bars, ± SE; *, P < 0.05 comparing +331G versus +331A.

Because the estrogen receptors were unlikely to account for differences of function of the +331G/A variant, we identified alternative candidate transcription factors using an in silico screen. Finding a nearby consensus binding site for the GATA family of transcription factors was intriguing because GATA5 is known to be expressed in breast cancer cells (17). The GATA family of transcription factors share DNA binding sites and activate transcription of genes important for the differentiation of several cell types (19). We found GATA3 and GATA6 expression in normal mammary tissue and two breast cancer cell lines. By comparison, we found that GATA5 is more strongly expressed in two breast cancer cell lines (Fig. 1B). Whereas GATA4 and GATA6 are required for diverse aspects of mesodermal cell differentiation during embryonic development (32, 33), GATA5 has a more restricted role in female germline development (20). The selective role for GATA5 in female development may reflect its regulation of sex steroids or their receptors. Indeed, GATA5 activates the progesterone receptor gene promoter more than the other GATA factors (Fig. 2A), with a greater effect driving expression of a transcript including the PR-B isoform (Fig. 3). This conclusion is further strengthened by our finding that GATA5 activates expression of the endogenous hPR gene transcript (Fig. 2B). Regulation of hPR-B expression in mammary cells by GATA5 is relevant because the hPR-B isoform is required for ductal and alveolar epithelial cell proliferation (10). Our results are therefore consistent with GATA5 activation of the progesterone receptor gene, with a greater preference for expression of PR-B-containing transcripts, which might contribute to mammary cell proliferation.

The presence of the +331G/A variant adjacent to a site bound by GATA5 suggested a possible interaction. This GATA site is important because despite the prediction of additional GATA binding sites in the hPR promoter, the +331A variant was associated with greater activation by GATA5 driving expression of a transcript encoding PR-B (Fig. 4A and B). Because the +331A variant did not increase GATA5 DNA binding, measured by gel mobility shift assay (data not shown), we hypothesize that the variant influences the recruitment of additional transcriptional regulators or may subtly affect DNA binding. Indeed, nucleotide differences within or adjacent to GATA elements could influence GATA protein binding (34) and GATA element–containing promoters have nonoverlapping patterns of expression.

In summary, we have found that GATA5, which is present in breast cancer cells, potently activates the hPR promoter, driving PR-B expression, and increases the endogenous hPR transcript. This finding is relevant because GATA5-dependent promoter activation is greater in the presence of the +331A variant. Taken together, our work supports screening for interaction of GATA variants and breast malignancies, particularly in conjunction with the hPR +331G/A variant.

Acknowledgments


Grant support: NIH grants CA82838 (I. De Vivo) and CA49449 (S.E. Hankinson), and American Cancer Society grant RSG-00-061-04-CCE (I. De Vivo).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Dr. Ramon Parsons for the MDA-MB-468 cell line and Drs. Ho and Parmacek for GATA expression plasmids, the participants of the Nurses’ Health Study for their exceptional dedication and commitment to the study, Dr. Hardeep Rana for genotyping, Pati Soule for laboratory support, and Carolyn Guo for programming support.

Figure 4. A, the hPR +331A variant increases activation by GATA5. The variant containing the hPR (∼711 to +822)-luc reporter plasmid (+331A) was activated significantly more strongly by GATA5 than the wild-type promoter (+331G) in MDA-MB-468 cells. B, GATA5 caused greater activation of the hPR (∼711 to +447)-luc reporter plasmid lacking the PR-A TSS in the presence of the +331A polymorphism. Transfection of pcDNA3-GATA5 activated the +331A (filled columns) hPR (∼711 to +447)-luc reporter plasmid significantly more strongly than the corresponding +331G (open columns) hPR reporter plasmid. Columns, mean of corrected luciferase activity from three separate transfections done in triplicate (n = 9 for each condition); bars, ± SE; *, P < 0.05 comparing +331G versus +331A.

source of progesterone for postmenopausal woman. This finding suggests that estrogen stimulation in the setting of the +331G/A variant contributes to the risk of endometrial (21) and breast cancer (Table 2). The +331G/A variant, however, does not alter a predicted binding site for an estrogen receptor and estrogen stimulation following estrogen receptor-α or -β overexpression did not alter expression of hPR (∼711 to +822)-luciferase reporter plasmids containing either +331G or +331A (data not shown). The increased risk of breast cancer from the +331A allele in the setting of obesity is therefore unlikely due to a direct effect of estrogen on the hPR promoter. Rather, we consider that this increase in risk is caused by variant-induced differences in progesterone receptor expression that influences progesterone and estrogen pathway cross-talk.

www.aacrjournals.org
References


GATA5 Activation of the Progesterone Receptor Gene Promoter in Breast Cancer Cells Is Influenced by the +331G/A Polymorphism


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/66/3/1384

Cited articles
This article cites 34 articles, 25 of which you can access for free at:
http://cancerres.aacrjournals.org/content/66/3/1384.full.html#ref-list-1

Citing articles
This article has been cited by 5 HighWire-hosted articles. Access the articles at:
/content/66/3/1384.full.html#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, contact the AACR Publications Department at permissions@aacr.org.