**Supplementary Materials**

**Details on Population of Patients Studies**

The median follow-up of patients entered on MA.27 was 4.1 years, which is the longest uniformly collected follow-up that the patients will have, so the results of this study pertain to relatively early breast cancer events with almost all occurring within five years of starting AI adjuvant therapy. There were 6,827 North American patients entered on MA.27, but 2,162 patients declined participation in blood collection and an additional 549 failed quality control procedures as indicated by the Participant Flow Diagram (Supplementary Fig. S1).

To determine if the North American patients included in the GWAS were comparable to to those excluded from the GWAS, we examined the balance of treatment arm (exemestane or anastrozole) and stratification factors (celecoxib use, aspirin use, trastuzumab use, lymph node status, and adjuvant chemotherapy use). Significant imbalances were observed for three of the five stratification factors, trastuzumab use (p<0.001), lymph-node status (p=0.004) and adjuvant chemotherapy (p= 0.01), but not for celecoxib use (p=0.79) or aspirin use (p=0.11). There were no significant differences in treatment arm (exemestane versus anastrozole) (p=0.89), cohort (p=0.59), or race (p=0.89). Details of the patient characteristics for those included and those excluded from the GWAS are given in Supplementary Table 1.

**Details on Quality Control**

**Study Design, Genotyping and Quality Control**

A total of 5207 participants DNA were obtained with clinical outcomes from the NCIC-CTG MA.27 clinical trial of aromatase inhibitors adjuvant therapy in postmenopausal women with estrogen positive breast cancer disease. We were able to obtain genotypes from two previous genotyping efforts for 1811 patients. Cohort 1 was genotyped for a GWAS of musculoskeletal adverse events([1](#_ENREF_1)). Cohort 2 was genotyped for a GWAS with the phenotype of fragility fractures([2](#_ENREF_2)). Insufficient DNA was available for 423 of the 3,396 remaining patients to genotype and hence, these patients, were removed from the study. The remaining 2973 were genotyped for this current study and are designated Cohort 3. Genotype data from all three cohorts were included for the current GWAS analysis. Detailed quality control procedures for Cohorts 1 and 2 are provided in their respective publications([1](#_ENREF_1), [2](#_ENREF_2)). Summaries of the quality control steps for all three genotyping cohorts are given below.

**Cohort 1:** Samples for 888 unique patients were genotyped by RIKEN utilizing the Illumina Human610 Quad BeadChip platform. However, four controls were not plated and, after genotyping, six samples (four from patients with a breast event and two from patients without an event) had call rates < 98% and were excluded from analyses. For 878 patients with clinical data, we received genotyping data for 580955 SNPs. 29,478 SNPs were removed due to low MAF (<1%) and 82 SNPs were removed due to departure from Hardy Weinberg Equilibrium. The remaining 551,395 SNPs from 878 patients were used for analysis. Two of these samples were removed because they were also genotyped in Cohort 2. Thus, 876 samples from Cohort 1 were available for inclusion in this GWAS.

**Cohort 2:** 948 samples were genotyped by RIKEN utilizing the Illumina Human Omni Express platform of which four samples failed genotyping because of a call rate <0.98. Thus, 944 samples were returned by RIKEN which included 3 study duplicates and 20 CEPH trios. An additional 67 samples were removed due to redundancy with Cohort 1 samples and sample mix-ups. A total of 729,758 SNPs were genotyped, but 37,489 were considered failures by the laboratory. Of the remaining 692,269 SNPs, 61,081 SNPs with a minor allele frequency (MAF) <0.01 were excluded because of limited power for association analyses. An additional 1,394 SNPs were removed based due to departure from Hardy Weinberg Equilibrium. For two pairs of duplicate samples, the percentage of discordant genotypes was less than 1.4E-3%, while one duplicate had 42% discordant genotypes, suggesting a sample mix-up and was excluded from further analyses. Thus, 887 samples from Cohort 2 were available for inclusion in this GWAS.

**Cohort 3:** 2973 samples were genotyped by RIKEN using the Illumina Human OmniExpressExome platform. Of these samples, 17 were not returned from RIKEN, 20 samples were removed due to relatedness, three samples were removed due to gender discrepancies and four samples from the cohort were ineligible for imputation. Thus, 2929 samples from Cohort 3 were available for inclusion in this GWAS.

We had genotype data for 964193 SNPs of which 2923 SNPs were removed as they were from chromosome Y, mitochondria or unplaced chromosomes. Additional, 40631 SNPs failed genotyping, 250843 were rare SNPs with MAF<0.01 and 460 SNPs were not in Hardy Weinberg Equilibrium. Finally 669,796 genotyped SNPs were available for inclusion in the analysis.

**Combined Cohort:** To avoid duplication, while creating the final combined cohort for Breast events 876 unique samples from Cohort 1 and 887 unique samples from RIKEN 4 (Omni Express) and 2929 unique samples from RIKEN (Omni ExpressExome) were selected for this breast event GWAS. After removing related samples and merging with available clinical data, a total of 4658 samples, including 252 samples from a patient with a breast event, were available, comprised of 869 samples from Cohort 1, 881 samples from Cohort 2, and 2908 samples from Cohort 3.

**Deep Sequencing Methodology**

We performed two-stage PCR to make a DNA library including the target region (75.4-75.7 Mb of chromosome 8). We performed multiplex PCR with 9 primer pools consisting of 2,130 primer pairs to amplify the target region. As the 2nd PCR, we added barcode and adaptor sequences to each PCR product using the additional sequences at the 5′ end of the primers for the 1st PCR. After purification and qualification of pooled libraries, we analyzed them by 2 × 150-bp paired-end reads on a HiSeq 2500 (Illumina, San Diego, CA)) instrument. Sequence reads were allocated to each individual according to the 8-bp barcode sequences.

**Details on Haplotype Analyses**

An expected design matrix was constructed from these haplotype probabilities and haplotypes that had a frequency less than 0.001% were combined.  The BCFI was modeled using a Cox proportional hazards model with the expected haplotype design matrix combined with covariate adjustment for race, treatment, ER-PR status, TNM stage, ECOG performance score and bisphosphonate use while being stratified for trastuzumab use, prior chemotherapy and nodal status.  Wald tests were conducted to obtain the significance of individual haplotypes versus the most common haplotype as the reference. The hazard ratios were calculated using the most common haplotype as the reference group.

**Details on Functional Genomic Studies**

**Materials**

Anastrozole, exemestane, MG132, bortezomib (Sigma-Aldrich, St. Louis, MO) were dissolved in DMSO (Sigma-Aldrich) as 10 mM stock, N-ethylmaleimide (NEM) was dissolved in 95% ethanol as 12.5% stock (g/vol), 4-androstene-3, 17-dione (Steraloids Inc., Newport, RI), E2 and 4-hyroxy-tamoxifen (4-OH TAM) (Sigma-Aldrich) were dissolved in 100% ethanol for further use. FLAG-ERα plasmid was provided by Thomas Spelsberg, Ph.D. (Mayo Clinic, Rochester, MN). HA-Ub and HA-FOXO3a plasmids were provided by Zhenkun Lou, Ph.D. (Mayo Clinic, Rochester, MN).

**Cell Culture**

CAMA1 cells were cultured in EMEM media (Eagle’s Minimum Essential Medium) with 10% (vol/vol) fetal bovine serum (FBS). HCC1428, BT474, AU565, BT549 and HEK293T cells were cultured in RPMI1640 medium with 10% FBS. MCF7/AC1 cells were cultured in IMEM (Improved MEM, no phenol red (Life Technologies, Carlsbad, CA) with 10% heated inactive FBS (Life Technologies, Carlsbad, CA).

**Lymphoblastoid Cell Lines (LCLs) Culture**

Five LCLs with *MIR2052HG* WT SNPs and five LCLs with *MIR2025HG* variant SNPs were selected. Before E2 or androstenedione treatment, ~ 2×107 cells from each LCL were cultured in RPMI1640 media containing 5% (v/v) charcoal stripped FBS (Invitrogen, Carlsbad, CA) for 24 h, followed with RPMI1640 medium without FBS for additional 24 h. Each LCL was plated into 12 well plates with RPMI 1640 medium containing 10-4, 10-3, 10-2 and 10-1 nM E2 or 10-4, 10-3, and 10-2 androstenedione. After 24 h treatment, increasing concentrations of anastrozole and exemestane were added in LCLs. The exemestane concentrations ranged from 10-4, 10-3, 10-2, 10-1 to 1 μM. The anastrozole concentrations ranged from 10-8, 10-6, 10-4, 10-3 to 10-2 μM. After an additional 24 h, all LCLs were collected for further RNA extraction and qRT-PCR.

**Quantitative real-time PCR assay (qRT-PCR)**

QRT-PCR assays were performed for measuring gene expression using TaqMan RNA-to-Ct 1-Step Kit (Life Technologies). RNA was extracted using miRNeasy mini Kit (Qiagen, Valencia, CA). RNA was measured by NanoDrops300 (Thermo fisher, Rockford, IL). The TaqMan primers for MIR2052HG, ESR1 and GAPDH were purchased from Life Technologies. QRT-PCR reactions were prepared as per the manufacturer’s protocol. Samples were run using StepOnePlus real-time PCR system (ABI).

**Chromatin Immunoprecipitation (ChIP) assays**

ChIP assays were employed using EpiTect ChIP OneDay kit (Qiagen). Approximately 2×107 LCLs per every sample (different SNP genotypes with E2 or E2 plus 4-OH TAM treatment groups) were collected for the ChIP assay. The details for the assay have been described([6](#_ENREF_6)). ChIP results were analyzed by qRT-PCR as well as by running DNA products on standard agarose gels. Primers for SNP rs4470990: forward primer 5’-GGA TTA CAG GCA CCT GGC TA-3’, reverse primer 5’-TAC AGT GGC TCA TGC CTG TC-3’ ; primers for SNP rs3802201: forward primer 5’-ATC TAG ACC TGA GCC CCT GTA CT-3’, reverse primer 5’-GCT AAT GCC GTC TTC AAA GG -3’. PCR products (20 uL) were loaded on 2 % agarose gels, and electrophoresis was performed at 80 V for 1 hour in 1×TAE buffer.

**MIR2052HG knock down and overexpression**

Antisense oligonucleotides (ASOs) were used to knock down and study the functions of MIR2052HG. Ten ASOs were designed for MIR2052HG and produced with locked nucleic acids modification of 5’ and 3’ ends (Exiqon, Woburn, WA). Negative control ASO was obtained from Exiqon (Woburn, MA). Lipofectamine RNAi Max (Invitrogen) and OPTI-MEM (Life Technologies) were used for ASOs transfection. Knock down efficiency was measured using TaqMan q-RT-PCR. Two out of the 10 ASOs showed consistent knocking down effect at 100nM concentration for 72 hours. The sequence of ASOs were as follows: ASO1: 5’-GTTGATTAGATTTGG-3’; ASO2: 5’-ACAGTCCCGATCTTC-3’; negative control: 5’-AACACGTCTATACGC-3’. MIR2052HG overexpression plasmid was obtained from Blue Heron Biotech (Bothell, WA). The gene transcription region was reinserted into the pGFP-C-shLenti shRNA Cloning Plasmid (TR30023, Origene, Rockville, MD). The lentiviral packaging kit (TR30022) (Origene) was used for MIR2052HG virus production in HEK 293T cells. The virus supernatant was harvested and used to infect breast cancer cells. Cells with stable MIR2052HG overexpression were selected using puromycin (Sigma-Aldrich) at a concentration of 4 µg/mL.

**Cell proliferation assays**

Cells transfected with MIR2052HG ASOs or MIR2052HG stably-overexpressed cells were seeded (3000 cells/100 uL/well) in a 96-well plate (Corning, Corning, NY). The MTS assay (Promega, Madison, WI) was used to determine the cell viability in six replicates. For cell proliferation after E2, androstenedione or AI (anastrozole or exemestane) treatment, MCF7/AC1cells stably overexpressing MIR2052HG were starved in IMEM with 10% charcoal stripped FBS (Invitrogen) for 24 hours followed by various treatments. Specifically, 1500 cells were plated per well in 96 well plates, followed by different treatments depending on the experiments: 1 nM androstenedione, 1nM androstenedione plus 1μM anastrozole, 1nM androstenedione plus 200 nM exemestane, 0.01 nM E2 and 0.01 E2 plus 1 μM 4-OH TAM. MTS assays (Promega) were performed to determine the cell viability every two to three days. In the assay, OD 490nm (optical density) represents the absorbance at the wavelength of 490 nm (nanometer).

**Colony forming assays**

Cells transfected with MIR2052HG ASOs or MIR2052HG stably-overexpressed cells were plated (800~ 1000 cells/well) in 6-well culture clusters (Corning) in triplicates. Subsequently, the cells were cultured for up to 14 days. Colonies were washed with cold PBS, fixed with frozen methanol, and then stained with Giemsa stain solution (Sigma-Aldrich). Colonies (>50 cells) were accounted with the Image J software (version 1.42q) and colony-formation rates were calculated.

**Western blotting**

Cells were washed with cold PBS and were lysed in cold NETN buffer (100mM NaCl, 20 mM Tris·HCl pH=0.8, 0.5mM EDTA, NP-40) with Proteasome cocktail inhibitor (Roche, Indianapolis, IN) and phosphatase inhibitor (PhosSTOP EASY pack, Roche). Cell lysates were diluted with SDS loading buffer (SDS, glycerol, bromic acid, 1 M Tris·HCl) and boiled, centrifuged at 10,000 rpm and stored at −20 °C. Equal amounts of protein were subjected to electrophoresis in 4-15% TGX SDS gels (Bio-rad, Hercules, CA) and were transferred to PVDF membranes. Membranes were blocked in TBS with 5% BSA and 0.1% Tween-20 and then incubated overnight at 4 °C with the following primary antibodies: ERα (Abcam, Cambridge, UK), AKT, p-AKT S473, p-AKT T308, FOXO3a, p-FOXO3 S318/S321 antibodies (Cell Signaling Technology, Danvers, MA) and β-actin antibody (Sigma-Aldrich). Membranes were washed with TBS-T (TBS with 0.1% Tween-20) and then incubated with HRP-conjugated anti-mouse IgG (Immuno-Jackson) or HRP-conjugated anti-rabbit IgG (Jackson Immuno, West Grove, PA) for 1 hour at room temperature. All blots were visualized with Supersignal WestPico chemiluminescent ECL kit (Thermo Fisher) and blue X-ray films (Phenix, Candler, NC).

**Ubiquitination assays**

Approximately1.5 µg of HA-ubiquintin plasmid and 1.5 µg of pcDNA 4.1-ERα plasmid with FLAG tag were co-transfected in HEK 293T cells using Lipofectamine 2000 (Invitrogen). Twenty-four hours later, these cells were reversely transfected with the MIR2052HG ASO1, ASO2 or negative control. Approximately 2×105 ASOs transfected and control cells were subsequently seeded into each 60 mm dishes (Corning). After 64 h, MG132 was added at a final concentration of 10 µM for an additional 8 hours Cells were then collected for the ubiquitination assay. Specifically, these cells were washed in cold PBS with NEM (1:100) and lysed in 2% SDS lysis buffer [62.5 mM Tris·HCl pH=6.8, 10% glycerol (v/v), SDS 2% (g/v)]. Immunoprecipitation assays were performed with the anti-FLAG antibody conjugated gels (ANTI-FLAG M2 Affinity gel) (Sigma-Aldrich). After washing, the FLAG gels were dissolved in 2xSDS loading buffer and boiled. These samples were then subjected to western blotting using the anti-ubiquitin antibody (Cell Signaling Technology) and antiFLAG antibody (Sigma).

**Supplementary Table 1 Showing Patients Included and Not Included in Genome-Wide Association Study**

|  | | | |
| --- | --- | --- | --- |
|  | Included in GWAS, n=4658 | Not Included in GWAS, n=2169 | Total, n=6827 |
| **Age at Randomization** |  |  |  |
| Median | 64.4 | 64.4 | 64.4 |
| Q1, Q3 | 58.4, 71.4 | 58.1, 71.7 | 58.1, 71.5 |
| **Treatment Arm** |  |  |  |
| Anastrozole | 2337 (50.2%) | 1084 (50.0%) | 3421 (50.1%) |
| Exemestane | 2321 (49.8%) | 1085 (50.0%) | 3406 (49.9%) |
| **Adjuvant**  **Chemotherapy** |  |  |  |
| No | 3267 (70.1) | 1454 (67.0%) | 4721 (69.2%) |
| Yes | 1391 (29.9%) | 715 (33.0%) | 2106 (30.8%) |
| **ECOG Performance Score** |  |  |  |
| 0 | 3884 (83.4%) | 1767 (81.5%) | 5651 (82.8%) |
| 1  2  3 | 740 (15.9%)  32 (0.7%)  2 (0.0%) | 383 (17.7%)  19 (0.9%)  0 (0.0) | 1123 (16.4%)  51 (0.7%)  2 (0.0%) |
| **Estrogen Receptor, Progesterone Receptor** |  |  |  |
| Positive, Positive  Positive, Negative  Positive, Missing  Negative, Positive  Negative, Negative | 3734 (80.2%)  793 (17.0%)  105 (2.3%)  26 (0.6%)  0 (0.0%) | 1731 (79.8%)  391 (18.0%)  28 (1.3%)  18 (0.8%)  1 (0.0%) | 5465 (80.0%)  1184 (17.3%)  133 (1.9%)  44 (0.6%)  1 (0.0%) |
| **Trastuzumab use**  No  Missing  Yes  **Celecoxib use**  No  Missing  Yes  **Aspirin use**  No  Missing  Yes | 886 (19.0%)  3730 (80.1%)  42 (0.9%)  552 (11.9%)  3544 (76.1%)  562 (12.1%)  1182 (25.4%)  3131 (67.2%)  345 (7.4%) | 265 (12.2%)  1888 (87.0%)  16 (0.7%)  259 (11.9%)  1661 (76.6%)  249 (11.5%)  551 (25.4%)  1487 (68.6%)  131 (6.0%) | 1151 (16.9%)  5618 (82.3%)  58 (0.8%)  1733 (25.4%)  5205 (76.2%)  811 (11.9%)  1733 (25.4%)  4618 (67.7%)  476 (7.0%) |

**Supplementary Table 2. Imputation Quality Values (dosage r squared) of the top six SNPs**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SNP | Cohort 1 dosage  r squared\* | Cohort 2 dosage  r squared | Cohort 3 dosage  r squared | Correlation of Imputed data and variants called from deep sequencing from the Cohorts combined |
| rs2891356 | 0.9958 | 0.9989 | 0.9985 | 0.998 |
| rs746460 | 0.9954 | Observed | Observed | 0.999 |
| rs4476990 | 0.9560 | 0.9583 | 0.9613 | 0.924 |
| rs13260300 | 0.9975 | Observed | Observed | 0.999 |
| rs4735715 | 0.9953 | 0.9903 | 0.9903 | 0.913 |
| rs3802201 | 0.9943 | 0.9828 | 0.9844 | 0.994 |

\*Dosage r-squared is the estimated squared correlation between the estimated allele dosage (0\*P(Hom Ref/first) + 1\*P(AB) + 2\*P(Hom Alt/second)) and the true allele dosage.

**Supplementary Table 3: Haplotype Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| Haplotype | Frequency | HR (95% CI) | Wald p-value (vs. reference) |
| GTCTAC | 0.569 | ref |  |
| ACCCGG | 0.004 | 1.19(0.37,3.76) | 0.766 |
| ACCTAC | 0.00491 | 1.05(0.26,4.22) | 0.942 |
| ACTCAC | 0.07133 | 0.8(0.55,1.15) | 0.2343 |
| **ACTCGG** | **0.311** | **0.58(0.46,0.72)** | **1.23e-06** |
| GTTTAC | 0.03439 | 0.87(0.54, 1.40) | 0.588 |
| Rarer haplotypes (freq < 0.001%) | 0.0053 | 1.089(0.33, 3.5) | 0.880 |

Haplotype reported in position order of: rs2891356, rs746460, rs4476990, rs13260300, rs4735715 and rs3802201.

**Supplementary Figure Legends**

**Supplementary Figure S1.**

Participant Flow Diagram for study([3](#_ENREF_3)).

**Supplementary Figure S 2.**

Quantile-Quantile plot of observed vs. expected p-values under the null hypothesis of no SNP association with breast cancer-free interval for patients treated with anastrozole or exemestane.

**Supplementary Figure S3.**

1. Propensity Score Weighted Survival Plot of breast cancer-free interval for patients with homozygous wild type (CC), heterozygous (CT), and homozygous variant (TT) genotypes for rs4476990. **B.** Propensity Score Weighted SurvivalPlot of breast cancer-free interval for patients with homozygous wild type (CC), heterozygous (GC), and homozygous variant (GG) genotypes for rs3802201. The p-values were obtained from a Stratified Cox-proportional Hazard model that was adjusted for treatment arm, Cohort (1,2,3), race, ERPR status, T-Stage, ECOG performance score, and bisphosphonate use.

**Supplementary Figure S4.**

1. Knock down of MIR2052HGdown-regulates ERα protein levels. **B**. Knock down efficiency is shown by QRT-PCR. Knock down of MIR2052HG decreases cell proliferation in HCC1428 (**C.**) and BT474 (**D.**) cells.

**Supplementary Figure S5.**

Knock down of MIR2052HG decreased androstenedione-induced (**A.**) and E2-induced (**B.**) proliferation in MCF7/AC1 cells. The concentrations of androstenedione and E2 are indicated.

**Supplementary Figure S6.**

Knockdown of MIR2052HG does not affect the proliferation of ERα-negative cell lines AU565 (**A**) and BT549 (**C**). **B** and **D**. Knockdown efficiency is shown by qRT-PCR.

**References for Supplementary Material**

1. Ingle JN, Schaid DJ, Goss PE, Liu M, Mushiroda T, Chapman JA, et al. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. J Clin Oncol. 2010;28:4674-82.

2. Liu M, Goss PE, Ingle JN, Kubo M, Furukawa Y, Batzler A, et al. Aromatase inhibitor-associated bone fractures: a case-cohort GWAS and functional genomics. Mol Endocrinol. 2014;28:1740-51.

3. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med. 2012;9:e1001216.