Cross-ancestry GWAS meta-analysis identifies six new loci for breast cancer in women of African and European ancestry

Background

- Despite the usefulness of GWAS, the majority of the studies have been done among European ancestry populations.
- Some common susceptibility loci are shared across populations, and the shared disease-associated variants are more likely to be causal.
- We discovered new risk loci for breast cancer using a cross-ancestry GWAS approach in African and European ancestry women with an approximate sample size of 248,000 women.

Methods

- Variants associated with breast cancer at \( P < 0.01 \) from African ancestry GWAS meta-analysis (9241 cases and 10192 controls) were meta-analyzed with European ancestry GWAS data (122977 cases and 108974 controls).
- Data: Four consortia of African ancestry populations (BCAC African ancestry samples, AABC, AMBER and ROOT) and the Ghana Breast Health Study (GBHS).
- Meta-analysis: Regression coefficient estimates from the five contributing African ancestry studies were combined in a fixed-effects meta-analysis using METAL.
- The pooled estimates from the five studies were then combined in another meta-analysis with the coefficients from the BCAC European ancestry data.
- SNPs significant genome-wide (5x10^{-8}) from the cross-ancestry meta-analysis, also significant at \( P < 0.005 \) from the African ancestry meta-analysis, and \( >500k \) away from the 180 loci known to be associated with breast cancer risk were identified as novel loci.
- Conditional regression included all variants in the flanking +/-500kb region of the lead SNPs using GCTA with the COJO option; Conditional p value -10^{-4}.
- Functional annotations were done using HaploReg v4.1, ENCODE and the Roadmap Epigenomics Consortium.
- eQTL analysis - All genes within +/-1Mb of index SNP were evaluated and gene expression data from TCGA breast cancer patients (African ancestry, \( n=164 \) and European ancestry, \( n=778 \)). A linear regression model estimated additive effects for each SNP, adjusting for age, ancestry, copy number variation, CNV, and batch effect, and molecular subtype.
- External validation using Latino GWAS (2385 cases and 6416 controls)

Results

Table 1. Novel breast cancer risk loci identified by cross-ancestry meta-analysis of African and European populations

<table>
<thead>
<tr>
<th>Locus</th>
<th>Minor Allele</th>
<th>AF ± 10^-6</th>
<th>Beta (SE) ± 10^{-6}</th>
<th>P value (AF) ± 10^{-6}</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17024629</td>
<td>A</td>
<td>0.26</td>
<td>0.6 (0.11)</td>
<td>0.02359 ± 0.3</td>
</tr>
<tr>
<td>rs2522057</td>
<td>C</td>
<td>0.28</td>
<td>0.92 (0.1)</td>
<td>0.001637 ± 0.8</td>
</tr>
<tr>
<td>rs1869959</td>
<td>T</td>
<td>0.43</td>
<td>0.28 (0.05)</td>
<td>0.002187 ± 0.2</td>
</tr>
<tr>
<td>rs67931591</td>
<td>C</td>
<td>0.28</td>
<td>0.92 (0.1)</td>
<td>0.001637 ± 0.8</td>
</tr>
</tbody>
</table>

Key findings

- 5 SNPs were associated with overall breast cancer and two with ER- breast cancer (Table 1).
- 4 SNPs were within genes and the others were in intergenic regions.
- Conditional regression: 3 additional independent signals at the 1p13.3 locus (rs11633925, rs114351980, and 1:109968974:C:T), two independent signals at 15q24 (rs119399578, rs121975707), and one each at 3q11.3 (5:13214922:G:GCCGCCGCC), and 15q26.3 (rs177793215) for overall breast cancer risk; rs5780828 at1p13.3 was associated with ER-negative breast cancer.
- eQTL analysis significant associations with gene expression: rs17024629 at 1p13.3 on GSTM1, GSTM2, and GSTM4; rs2522057 and IRF1; rs1869959 at 15q24.1 and MPI and ULK3; rs60381548 and 4 genes SIN3A, and PTPN9, SNJNP, SNX33; rs67931591 at 1q41 and PTPN14.
- Functional annotation: overlap with genomic functional biofeatures for rs2522057, rs17024629, rs1869959, and rs60381548 or SNPs in strong LD with these top SNPs in tissue-originated cell lines.
- Replication in Latinos: effect and direction of the association were consistent in 8 out of 11 evaluated variants but none statistically significant.

Summary and Conclusions

- The novel SNPs lie in regions that are close to genes that have been previously implicated in cancer – rs67931591 and KCNK2; rs1869959 and SCAMP2, SIN3A and ULK3; rs2522057 and Rad50, IRF1 and SLC22A5; rs17024629 and AMPD2, GSTM1, GSTM2, and GSTM4; rs1637365 and CASTOR2.
- Loci are shared across ancestry and useful for risk prediction in African and European ancestry populations.
- Further validation in other ancestral populations and functional studies still needed.
- Limitations
  - Discovery driven mainly by European ancestry data.
  - Sample size for ER- negative breast cancer relatively small.
  - External validation was not done in another African ancestry sample.