**Supplementary Files**

**Supplementary Table 1**: Studies participating from Shanghai Breast Cancer Study (SBCS), Shanghai Endometrial Cancer Study (SECS), and Guangzhou Colorectal Cancer Study (GZ), the Shanghai Breast Cancer Survival Study (SBCSS) and the Shanghai Women’s Health Study (SWHS), included in this work.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Cases** | **Controls** | **Study design** | **Age (Years)\*** | **Age < 45 Years** | |  | **Family history (%)&** | |
| **Cases** | **Controls** | **Cases** | **Controls** |
| SBCS | 5664 | 2806 | Population-based | 51 | 1308 | 758 |  | 5.5 | 2.6 |
| SWHS | 1183 | 5464 | Population-based | 56 | 329 | 1438 |  | 3.7 | 2.8 |
| SBCSS | 357 | 0 | Population-based | 53 | 61 | 0 |  | 5.1 | 0 |
| SECS | 0 | 855 | Population-based | 55 | 0 | 106 |  | 0 | 2.4 |
| GZ | 0 | 468 | Hospital-based | 55 | 0 | 106 |  | 0 | 0 |

\* Mean age of cases/controls with available data.

**&** First-degree family history of breast cancer for cases/controls.

**Supplementary Table 2**. Diagnostic age of samples carrying this risk variant in the susceptibility *ATM* gene.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Study** | **Platform** | **Samples** | **Diagnostic age** | **Family history&** |
| SBCS | Whole exome sequencing | Sample 1 | 49 | No |
| SBCS | Whole exome sequencing | Sample 2 | 40 | Yes |
| SBCS | Whole exome-chip | Sample 3 | 73 | No |
| SBCS | Whole exome-chip | Sample 4 | 43 | No |
| SBCS | Whole exome-chip | Sample 5 | 68 | No |
| SBCS | Whole exome-chip | Sample 6 | 66 | No |
| SWHS | Whole exome-chip | Sample 7 | 69 | No |
| SWHS | Whole exome-chip | Sample 8 | 58 | No |
| SBCS | Whole exome-chip | Sample 9 | 50 | No |
| SBCS | Whole exome-chip | Sample 10 | 66 | No |
| SBCS | MEGA | Sample 11 | 73 | No |
| SBCS | MEGA | Sample 12 | 49 | No |
| SBCS | MEGA | Sample 13 | 64 | Yes |

**&** First-degree family history of breast cancer patients.

**Supplementary Figure 1.** Plots of the first two principal components for the samples in the study populations, including A) population I and B) study populations II and III, and for the samples from the 1,000 Genomes Project.

****

**Supplementary Figure 2.** Identification of the variant rs139379666 in *ATM* from whole exome sequencing, whole exome-chip and MEGA.



**Supplementary Video (mpg format)**. Illustration of P2974 locating at the PRD, which is near the protein-protein-interaction surface of the ATM dimer based on the ATM structure prediction.