Genetic variants associated with breast size also influence breast cancer risk

Geoffrey M. Benton[1], Nicholas Eriksson[1], Chuong B. Do[1], Amy K. Kiefer[1], Joanna L. Mountain[1], David A. Hinds[1], Uta Francke[1] and Joyce Y. Tung[1]

[1]23andMe, Inc, Mountain View, CA

Introduction

While some factors of breast morphology, such as density, are directly implicated in breast cancer, the relationship between breast size and cancer is less clear. Breast size is moderately heritable, yet the genetic variants leading to differences in breast size have not been identified.

To investigate the genetic factors underlying breast size, we conducted a genome-wide association study (GWAS) of self-reported bra cup size, controlling for age, genetic ancestry, breast surgeries, pregnancy history and bra band size, in a cohort of 16,175 women of European ancestry.

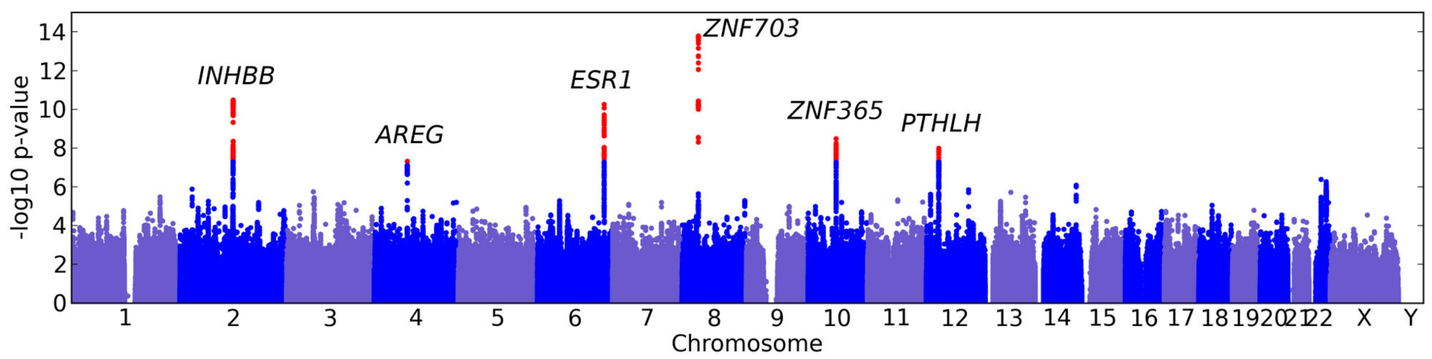

Figure 1. Manhattan plot of association with breast size. -$\log_{10}$ *p*-values across all SNPs tested. SNPs shown in red are genome-wide significant ($p < 5 \times 10^{-8}$). Regions are named with the postulated candidate gene.

Methods

Phenotype Data Collection
Participants were drawn from the customer base of 23andMe, Inc., a consumer genetics company. Participants provided informed consent and participated in the research online. All participants reported bra cup size and bra band size as part of an online body morphology questionnaire. Participants selected a cup size from nine categories and entered band size as an integer.

Genotyping
Participants were genotyped for 586,916 to 1,008,948 SNPs on one of three Illumina-based BeadChips. An additional 7,422,970 imputed SNPs were included in the analysis.

Statistical analysis
Bra size was coded from 0 to 9, corresponding to the categories: Smaller than AAA, AAA, AA, A, B, C, D, DD, DDD, and Larger than DDD, respectively. Genotypes were coded as dosages from 0–2, corresponding to the estimated number of copies of the minor allele present. *p*-values for SNPs were calculated using likelihood ratio tests for linear regressions. As covariates in the analysis, we included the projections onto the first five principal components of genetic ancestry as well as age, bra band size (in inches), and indicator variables for breast augmentation surgery, breast reduction surgery, mastectomy, past pregnancy, and current pregnancy or breastfeeding. We performed conditional analyses within each genome-wide significant region to search for SNPs with independent effects.

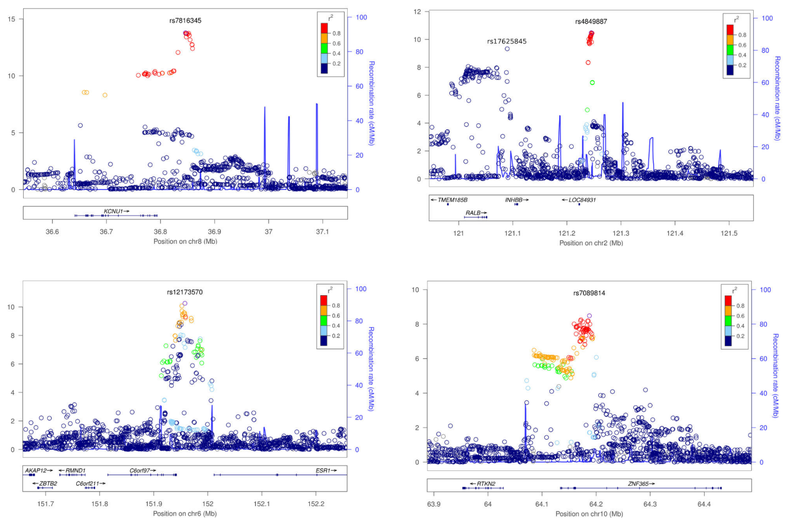

Figure 2. Associations with breast size in four regions with genome-wide significant SNPs. Colors depict the squared correlation (r^2) of each SNP with the most associated SNP (which is shown in purple). Gray indicates SNPs for which r^2 information was missing. For the plot labeled with rs7816345, the gene *ZNF703* lies about 400kb outside the region displayed.

SNP	Chr	Position	Gene	Allele	MAF	HWE	r^2_{V2}	r^2_{V3}	*p*-value	β
rs7816345	8	36846109	*ZNF703*	C/T	0.194	0.08	1.00	1.00	$1.64 \cdot 10^{-14}$	-0.151 (-0.189 – -0.112)
rs4849887	2	121245122	*INHBB*	C/T	0.113	0.97	1.00	0.94	$3.31 \cdot 10^{-11}$	0.166 (0.117 – 0.214)
rs17625845	2	121089731	*INHBB*	T/C	0.205	0.07	0.72	0.99	$4.7 \cdot 10^{-10}$	0.125 (0.086 – 0.164)
rs12173570	6	151957714	*ESR1*	C/T	0.101	0.82	0.98	0.99	$5.58 \cdot 10^{-11}$	0.171 (0.120 – 0.222)
rs7089814	10	64187564	*ZNF365*	T/C	0.375	0.97	0.94	0.98	$3.3 \cdot 10^{-9}$	0.096 (0.064 – 0.128)
rs12371778	12	28156081	*PTHLH*	C/G	0.091	0.72	0.85	0.87	$1.03 \cdot 10^{-8}$	-0.162 (-0.217 – -0.106)
rs62314947	4	75502487	*AREG*	C/T	0.281	0.31	0.73	0.91	$4.79 \cdot 10^{-8}$	-0.101 (-0.137 – -0.065)
rs4820792	22	29161007	*CHEK2*	C/T	0.180	0.01	0.95	0.97	$4.17 \cdot 10^{-7}$	0.105 (0.065 – 0.146)
chr22:40779964	22	40779964	*MKL1*	G/A	0.056	0.05	0.75	0.78	$5.47 \cdot 10^{-7}$	-0.187 (-0.261 – -0.114)
rs61280460	14	94796184	*SERPINA6*	A/T	0.199	0.29	1.00	1.00	$8.3 \cdot 10^{-7}$	-0.095 (-0.132 – -0.057)

Table 1. Index SNPs for regions under $p = 10^{-6}$. The index SNP is defined as the SNP with the smallest *p*-value within a region; or the SNP with the smallest *p*-value in the conditional analysis. The listed gene is our postulated candidate gene near the SNP. For INHBB, conditional analysis revealed two independent SNPs in the region. Alleles are listed as major/minor (in Europeans). The coefficient refers to the average change in breast size (in units of cup size) per copy of the minor allele.

Results

We identified seven single-nucleotide polymorphisms (SNPs) significantly associated with breast size ($p < 5 \times 10^{-8}$): **rs7816345** near *ZNF703*, **rs4849887** and (independently) **rs17625845** flanking *INHBB*, **rs12173570** near *ESR1*, **rs7089814** in *ZNF365*, **rs12371778** near *PTHLH*, and **rs62314947** near *AREG*. Two of these seven SNPs are in linkage disequilibrium (LD) with SNPs associated with breast cancer (those near *ESR1* and *PTHLH*), and a third (*ZNF365*) is near, but not in LD with, a SNP associated with breast cancer.

Discussion

These results provide insight into the genetic factors underlying normal breast development and show that some of these factors are shared with breast cancer. While these results do not directly support any possible epidemiological relationships between breast size and cancer, this study may contribute to a better understanding of the subtle interactions between breast morphology and breast cancer risk.

Acknowledgments

We thank 23andMe's customers who consented to participate in research for enabling this study. We also thank the employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

References

Eriksson, N. et al. *BMC Med. Genet.* 2012 Jun 30;13(1):53. [Epub ahead of print]