Risk factors of Breast Cancer in Rural & Urban India

By

Rajini Nagrani

[HLTH09200904001]

Tata Memorial Centre
Mumbai

A thesis submitted to the Board of Studies in Health Sciences

In partial fulfilment of requirements For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

September, 2014

HOMI BHABHA NATIONAL INSTITUTE

RECOMMENDATIONS OF THE VIVA VOCE BOARD

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Rajini Nagrani entitled "**Risk factors of Breast Cancer in Rural & Urban India"** and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____

Chairman: Dr. Rajendra Badwe **Date:**

_____

Guide / Convener: Dr. Rajesh Dikshit **Date:**

_____

Member 1: Dr. S.V. Chiplunkar **Date:**

_____

Member 2: Dr. R. Mulherkar **Date:**

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date:

Place

Dr.Rajesh Dikshit

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Mumbai

Date:

Rajini Nagrani

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Mumbai

Date:

Rajini Nagrani

List of Publications arising from the thesis

Journal

Rural urban differences in breast cancer in India. **Nagrani RT**, Budukh A, Koyande S, Panse NS, Mhatre SS, Badwe R. Accepted in Indian Journal of Cancer.

Rajini Nagrani

CERTIFICATE

I certify that the thesis titled '**Risk Factors of Breast Cancer in Rural & Urban India** ' submitted for the degree of Doctor of Philosophy by Ms. Rajini Thakur Nagrani is a record of the research carried out by her during the period September 2009 to September 2014 under my supervision. This work has not formed the basis for the award of any degree, diploma, associateship or fellowship at this or any other institute or university.

Mumbai

September 2014.

Dr. Rajesh Dikshit

ACKNOWLEDGEMENTS

One looks back with appreciation to the brilliant teachers, but with gratitude to those who touched our human feelings. The curriculum is so much necessary raw material, but warmth is the vital element for the growing plant and for the soul of the child.

— Carl Jung

I really feel very honored for having come across brilliant teachers who not only taught me the academic subjects but also touched the deepest chords of my heart. I would like to convey my genuine appreciation and heart-felt gratitude to all of them. PhD is not just another degree, but it marks the culmination of very long years of academic education. I would like to take this opportunity to thank one and all who have contributed to my learning process since I was a child in school.

I am extremely glad that I had a chance to do my PhD under someone who truly practices and appreciates deep thinking. I will be forever indebted to Dr. Rajesh Dikshit for all the time he has spent with me in discussions. His stimulating suggestions, experience and encouragement accompanied by the freedom of thought that he granted, has not only helped me to discover my potential but also kept me sparked throughout my Ph.D tenure. I will never forget the generosity that he offered on several occasions. Sir, I thank you for everything and will always be proud to have been mentored by you.

I am thankful to Dr. Rajendra Badwe for providing me an opportunity to work in this institution, excellent infrastructure and chairing my Doctoral Committee. I would also like to thank the members of my Doctoral Committee, Dr. K.M. Mohandas, Dr. S.V Chiplunkar and Dr. Rita Mulherkar, for annually reviewing my work and for giving me positive feedback. Special thanks to Dr. K. S. Sharma (Dean, Academic) for accepting my submission of this thesis.

A special thanks to Dr. Ganesh for boosting my morale when I needed it the most, Dr. Atul Budukh who have been an initiator of one of the most informative weekly session 'Epimet' – an intradepartmental meet and Dr. Sudeep Gupta for his extremely precious clinical inputs in the work.

One person who has always helped me, my friend and colleague Sharayu, it would have been impossible to achieve this feat without you. The scientific conversations with you were a delight. I convey my special and sincere thanks to you. I will remain indebted to you for life.

I am fortunate to have a friend like Garima, with whom I had a memorable year of our PhD days. Garima, you have always been instrumental in projecting a different viewpoint of life.

I want to express my warmest thanks to all past and current members of CCE: Interviewers and data entry operators (Trupti, Samradhini, Sapna, Jhanvi, Swati, Deepika, Trupti, Sapna, Vaishnavi, Shraddha, Rashmi, Kanchan, Aakash, Shweta, Aarti, Vaibhav, Deepika, Shilpa, Geeta, Sakshi, Ishwari, Sheetal, Kruthi and Aarti); Study Co-ordinators, Priyanka and Shraddha have been wearing different hats to co-ordinate activities between staff members and handled the most complicated administrative tasks for the entire department; Ankita, and Rupali, who have made look those numbers beautiful, and easier for analysing data. Vaibhav, Nikita, Bipin and Sushma for making the tiresome long hours in lab so lively and helping me with the experiments. Kishore for helping me with the day-to-day lab requirements.

I owe a lot to all my friends Priyanka, Shital, Shashi, Swati, Lisa, Kamlesh and Shraddha who kept the tough and tense days of my Ph.D as joyous as possible. I am extremely thankful to some of my oldest friends Sunny, Omkar, Deval, Harish, Vaishali, Sonam, Akbar, Mousumi, Pratik, Sagar, Sai, Shweta, Minti, Nimi, Rohit, and Rashmi for being patient listeners and keeping my days fun filled.

A special thanks to my family. Words cannot express how grateful I am to my Mom and Dad for all the sacrifices they have made on my behalf. I know, I always have my family to count

on when times are rough. I feel so proud and blessed to be their daughter. Very special thanks to Dewal family; I have fond memories of the dinner table discussion which were a package of wit and entertainment. The biggest stressbuster in these testing times was my little angel Aarush. Reena-Ashu thanks for surprise treats which acted as mood elevators.

I would like to remember and thank Sakhi Mummy, though not around me anymore has played a pivotal role in whatever I have achieved in life today.

I am also thankful to all others not mentioned here and have been equally helpful to me in various paths of Ph.D.

Finally, I thank almighty God who has helped and blessed me through the entire voyage and have made this day possible.

Table Of Contents

Synopsis

HOMI BHABHA NATIONAL INSTITUTE

Ph. D. PROGRAMME

1. **Name of the Student: Rajini Thakur Nagrani**
2. **Name of the Constituent Institution: Tata Memorial Centre**
3. **Enrolment No. : HLTH09200904001**
4. **Title of the Thesis: Risk Factors of Breast Cancer in Rural & Urban India**
5. **Board of Studies: Health Sciences**

INTRODUCTION

In 2012, 1.7 million women globally were diagnosed with breast cancer and there were 6.3 million women alive who had been diagnosed with breast cancer in the previous five years [1]. In countries with high and medium Human Development Index (HDI) an increased risk in female breast cancer has been observed [2]. Further, the pattern of age adjusted incidence rates as observed in Cancer Incidence in Five Continents report clearly indicate that breast cancer rates are high in developed countries and are much lower in less developed countries including India [3].

Within India, there are likely substantial differences in the incidence rates of breast cancer in rural and urban areas: rates observed in urban registries are in the range of 29 – 35 per 100, 000 whereas those observed in rural registries vary from 10-12 per 100,000. The lowest breast cancer incidence rates are found among women from the rural area of Barshi in Western India, and Dindigul Amblikkai, another rural area in the more developed South of India. Among urban Indian women, breast cancer incidence rates are almost three times higher than in rural women [4]. A twofold increased risk was observed in urban areas and a threefold increased risk was observed in metro areas compared to rural area [5]. The cause of this strong urban rural difference is not known although it is likely to be due to one or more westernised lifestyle related factors such as parity, breastfeeding, age at first birth and obesity, prevalence of which differs strongly between rural and urban women.

Nulliparity and late age at first birth are the most consistently observed risk factors for breast cancer [6]. The risk among women who have their first child after the age of 30 is about twice that of women who have a first child before the age of 20. Similarly, women who start menstruating early in life, or have a late menopause, also have an increased risk of developing breast cancer [7], possibly because of the increased number of ovulatory cycles and exposure

to estrogens and other breast tissue proliferative hormones. It is also possible that extensive breastfeeding reduces the risk of breast cancer by suppressing the number of ovulatory cycles, although the evidence based on studies conducted in western populations is unclear [8]. Overall however, these established risk factors account for only a small part of the large difference in incidence between developed and developing countries and other important risk factors for breast cancer remain to be identified.

Higher body mass index has been found to increase the risk of breast cancer after menopause, although this has not been observed in all cohort studies which have examined the association. Similarly, weight gain during adulthood has been identified as a risk factor for breast cancer in most studies in which it was investigated in post menopausal women. Some studies have also observed that the weight gain at the age of 20 years increases the overall breast cancer risk [9]. Physical activity has also been hypothesized to protect against the development of breast cancer [10]. Green et.al has shown individual height as an independent risk factor in breast cancer [11]. There have been large Genome Wide Association Studies (GWAS) on breast cancer in most developed countries [12-15] showing low to modest associations between common polymorphisms and breast cancer risk. In India, however, there have been no GWAS studies and few properly designed retrospective studies with smaller sample size on genetic susceptibility to study this risk [16-18].

GAPS IN LITERATURE

It has been observed for long time that the rates of breast cancer differ in rural and urban areas. However, there are very few studies in literature to address the reasons for the differences in the breast cancer rates of rural and urban area. Obesity has been observed to be risk factor for postmenopausal breast cancer. However the contribution of different measures of obesity and their role in pre- and postmenopausal women is still not clear. In Indian context, there are no

large studies to address the issue of reproductive factors, obesity, age at last pregnancy, oral contraceptive use and genetic susceptibility in development of breast cancer.

The present thesis proposal is designed to understand more clearly the reasons for rural-urban differences, and role of genetic susceptibility in development of breast cancer.

HYPOTHESIS

Anthropometric and Lifestyle related variables are the cause of large differences in occurrence of breast cancer in rural and urban areas.

AIM

Primary: To study role of anthropometric and other lifestyle related variables in causation of breast cancer in rural and urban areas.

Secondary: To study role of genetic susceptibility in breast cancer.

PRIMARY AIM

To study role of anthropometric and other lifestyle related variables in causation of breast cancer in rural and urban areas.

Study Population: A hospital based case-control study was conducted at Tata Memorial Hospital (TMH), Mumbai during the period of January 2009 to September 2013.

Criteria for enrolment of cases: The cases were female breast cancer patients coming to TMH. Only primary breast cancer cases aged 20-69 were enrolled in the study with date of diagnosis not more than 6 months before the date of interview. All the breast cancer cases enrolled in the study were histologically confirmed.

Criteria for enrolment of controls: All female visitors with no history of cancer coming along with any site cancer patient aged 20-69 were included in the study. Visitor controls coming to various Disease Management Group (DMGs) have been enrolled. Not more than 20% controls have been enrolled from any of the DMGs, to avoid selection bias.

The study has been approved by TMH Institutional Review Board. Written informed consent was obtained from all participants before enrolling them in the study.

Data Collection: In-person interview of each case and control was conducted by trained interviewers using a pre-tested structured questionnaire covering demographic and socioeconomic variables, reproductive history, time spent in household activities on a normal day, residential history, occupational history, personal and family medical history, tobacco and alcohol habits, and diet. Controls were frequency matched to cases on age and region of residence (South, North, East, West and Central India). Anthropometric measurements were taken at the end of the interview.

Blood Collection: A 10ml blood sample was collected from each study participant and centrifuged into plasma and buffy coat. The blood components were then stored at -80^0C.

Definition of Rural and Urban areas: All study participants were asked to list all places of residence where they had lived for at least 1 year, starting with the place of birth. The rural and urban residence status was self reported by study participant. Study participants were stratified into rural and urban using four different definitions as follows:

1. Ever lived in a rural area: If a study participant has lived in a rural area for 1 year or more in life is termed as a "rural participant", whereas any participant who has never lived or lived less than 1 year in a rural area is termed as "urban participant".
2. First 20 years of life lived in a rural area: If a study participant has lived first 20 years of her life in a rural area, i.e., from age 0 to age 20, then participant is classified as "rural participant," whereas any participant who has lived less than 20 years in a rural area is classified as "urban participant".
3. Currently living in a rural area: Any study participant who has a current residence (at the time of enrolment) of 1 year or more in a rural area is termed as "rural participant", versus a current residence in an urban area is an "urban participant".

4. Total years lived in a rural area:
    a. 1-10 years: A minimum of 1 year and a maximum of 10 years lived in a rural area versus never lived in a rural area are categorized as rural and urban participants respectively.
    b. >10 years: If total years lived in a rural area is more than 10 years, study participant is categorized as rural or else urban.

Anthropometric Measurements: Height (without shoes in cm) and weight in light clothing (in kg) of each study participant were measured using standard equipment. Weight was measured with light clothing. Waist size (in cm) was measured using a tape around the narrowest part of the trunk between the lower rib and level of the highest point of the margin of the hipbone, and hip size (in cm) was measured with light clothing at the widest part. All measurements were done twice in succession and averaged for a final value. Waist-to-hip ratio was computed by taking the ratio of waist size (in cm) and hip size (in cm).

Definition of Menopausal status: Women with no history of menstrual period during the last 12 months were classified as postmenopausal. The rest were treated as premenopausal.

Quality Assessment for Questionnaire Based Data

Preparation of Instruction Manual for filling up the Questionnaire in Case-Control Studies: In order to assure the homogeneity of data collection by the Social Investigators, an instruction manual and video recording was prepared. The instruction manual contains detailed guidelines and figures wherever required for better understanding of questions by the social investigator as well as the respondent [19].

Preparation of Instruction Manual for Data Entry: In order to assure the homogeneity while entering the data, clear and precise instructions with predefined logical checks have been listed in the form of Manual [20].

Monitoring of Daily Work: All forms were regularly checked for errors after conducting the interviews and after the data has been entered in the database. Weekly meetings were conducted to understand and resolve the problems of data collection. Training program was conducted every quarter so as to ensure the quality of interviews. The questionnaire was checked daily for completeness of information.

Quality Checks on Data Entry: Logical Checks were prepared to identify errors in the data entry. The data was entered twice and corrected for errors between 2 entries, if any, occurred while entering the data.

Reproducibility of Questionnaire: Abbreviated questionnaire was designed. This questionnaire contained constant (non changing in recent time) variables such as number of pregnancies, height, vegetarian /non-vegetarian status. The reproducibility questionnaire was completed for 249 study participants (approx 8% of total enrolled in study). Details of main measured exposures are shown in Table 1.

Table 1: Reproducibility of Measured Exposure

Variable	**Study Mean (Reproducibility Mean)**	**Coefficient of Correlation (%)**
Age	46.90 (47.17)	92.25
Number of Pregnancies	4.06 (3.99)	91.07
Height	156.92 (157.18)	96.51

Correction of differences between Data entry 1 & 2: There were 207 variables which were corrected for any differences observed between the 2 data entries.

Statistical Analysis: The odds ratios (OR)s of developing breast cancer and their 95% confidence intervals (CI)s for anthropometric measurements and reproductive factors were estimated separately by residential status (Rural/Urban) and menopausal status. Unconditional logistic regression models were adjusted for potential confounders (age, region of residence, rural-urban status, education, age at first full-term pregnancy, waist-to-hip ratio (WHR), height, menopausal status, number of abortion and miscarriage). The ORs for age at first full-

term pregnancy were estimated after adjusting for number of full-term pregnancies in addition to above mentioned variables. Weight in kg as continuous variable was entered in the model for estimating OR for WHR and height in addition to above variables. All analysis were performed using the statistical package Stata version 12 [21]. Tests for linear trend across levels of exposure categories were performed on the continuous categorical variables entered as ordered, quantitative variables into the models.

Result & Discussion: Questionnaire data was collected on 1637 breast cancer cases and 1515 female controls. All the results were adjusted for the confounding variables unless mentioned otherwise.

A protective association was observed using all the definitions of "rural" [Ever lived in rural area – OR=0.81; 95% CI - 0.71-0.94); More than 10 years lived in rural area – OR=0.81; 95% CI - 0.70-0.93); first 20 years of life lived in rural area – OR=0.65; 95% CI - 0.56-0.76)] except in women who were currently residing in rural area at the time of enrolment after adjusting for confounding variables such as age and region of residence. On further adjustment for additional risk factors viz. age, region of residence, education, age at first full-term pregnancy, height, WHR and menopausal status; only women who lived first 20 years of life in rural area showed protection against breast cancer (OR=0.77; 95% CI - 0.65-0.92). However, most of the etiological studies have used current area of residence as definition for rural [22-23] and limited studies which have taken early years of life spent [24] or place of birth [25-26] in rural areas as definition for "rural". The current residence as demonstrated in the present study is not a good marker for studying the effect of rural environment on the risk of breast cancer, as exposures in early life may be more important in the development of breast cancer compared to current exposures [27]. For instance, strenuous physical activity at younger age can delay both menarche and onset of regular menstrual cycle [28]. Further, the individuals migrating from rural area to urban area in adulthood might not change their

lifestyle and adhere to rural life; therefore they may continue to get protection from breast cancer even if they are currently residing in urban areas which have been clearly demonstrated in the present study. Therefore, in all the further analysis women who lived first 20 years of life in rural area were designated as 'rural participants' else stratified as urban.

Prevalence of ER/PR negative (60.9%) cases was higher in rural area compared to urban area where the prevalence was observed to be 54.3%. A statistically significant difference (P = 0.018) in the prevalence of Triple Negative Breast Cancer (TNBC) tumours was observed in the rural area (44.2%) compared to urban area (34.3%).

Women who had 4 or more live births showed a protective association with OR = 0.66 (95% CI - 0.49-0.87) as compared to women with 1 live birth after adjusting for confounding variables and without stratification for rural-urban status. The significant protection was observed only in rural women (OR = 0.42; 95% CI-0.24-0.75). Age at first full-term pregnancy proved to be an important risk factor in the development of breast cancer. Women who had their first full-term pregnancy after age 25 had a significantly elevated risk of breast cancer compared with women who had first full-term pregnancy below 20 years of age (OR = 1.83; 95% CI-1.41-2.36). This protection was observed in both rural areas (OR = 2.24; 95% CI-1.13-4.43) and urban areas (OR = 1.78; 95% CI-1.32-2.41). A statistically significant linear trend was observed among the categories of age at first full-term pregnancy. The lifestyle patterns among women living in urban areas has changed considerably, with women attaining higher level of education, postponing marriage, postponing their first child to an older age, and having fewer pregnancies over time [29-30]. An indication has been observed in this study that use of OC may increase the risk of breast cancer particularly for women residing in urban area (OR = 1.28; 95% CI-0.93-1.76). Two or more than 2 induced abortions has been observed to be a risk factor of breast cancer overall (OR = 1.65; 95% CI-1.25-2.17), urban (OR = 1.58; 95% CI-1.15-2.16) and rural women (OR = 2.08; 95% CI-1.16-3.72). Even

a single miscarriage provides a protection against breast cancer in rural (OR = 0.62; 95% CI-0.41-0.95) and urban women (OR = 0.79; 95% CI-0.58-1.06) possibly due to its protection acquired by pregnancy. However the results observed for abortion and miscarriage has to be interpreted considering the possibility of recall bias, a limitation of case-control studies. A time difference of 10 years or more between age at menarche and age at first full-term pregnancy was observed to be significantly associated in urban women (OR = 1.36; 95% CI-1.11-1.68), but not in rural women (OR = 1.43; 95% CI-0.91-2.24).

Height has been consistently associated with an increase in breast cancer risk [11, 31]. In the present study, for every 5 cm increase in height the OR of 1.10 (95% CI-1.02-1.19) was observed in the urban area, but not in rural area (OR = 1.05; 95% CI-0.93-1.19). The increased risk of breast cancer for WHR of ≥0.95 when compared to WHR of <0.85 was observed to be OR = 3.78 (95% CI-2.92-4.89) without stratifying on rural-urban status; in urban women (OR = 4.07; 95% CI-3.00-5.53) and rural women (OR = 3.00; 95% CI-1.84-4.90). A significant positive association with WHR has been reported in both pre- and postmenopausal women, a result similar to that observed in two meta-analysis report [32-33].

SECONDARY AIM

To study role of genetic susceptibility in breast cancer.

DNA Preparation: Buffy coat samples were available for 1214 cases and 1293 controls. Genomic DNA was extracted from buffy coat using Qiagen QiAamp Blood DNA MidiKit and Macherey Nagel Nucleomag Blood kit. Concentration of each DNA sample was determined by the optical density (OD) at 260 nm and the purification was evaluated by OD 260/280 ratio. All DNA samples were also quantitated using Quant-iT PicoGreen dsDNA reagent, and purity was assessed by measuring the UV absorbance for accuracy. The quality of genomic DNA was assessed on 5% samples using 0.8% agarose gel. 1204 cases and 1212 controls had sufficient yield to proceed with genotyping. DNA concentrations were adjusted

to 50ng/µl and verified using Quant-iT PicoGreen dsDNA reagent. The aliquots of DNA were stored at -20^{0}C.

Design of Custom SNP Panel: A customized panel of 384 single nucleotide polymorphisms (SNPs) was designed using a mixture of 3 strategies which are as follows

1. Candidate SNP Studies: All candidate SNP studies which have been significantly associated with breast cancer and suggested by collaborator on basis of animal experiments were included under this criterion using HuGE Navigator [34]. Total SNPs selected from this category were 96.

2. GWAS: The GWAS snps were identified using HuGE Navigator [35] and NIH GWAS Catalog [36]. The SNPs which were positively associated (p value < 10^{-5}) with following diseases or traits (the number of snps selected in the respective category mentioned in parenthesis) were included:

Body Mass Index (37); Breast Cancer (51); Insulin Like Growth Factors (1); Menstruation and Menopause (41); Obesity (29); Waist-to-Hip Ratio (2). A total of 161 snps were identified using this strategy.

3. Bioinformatics Tool: 127 tag snps were selected using this strategy. Obesity search term was used in Gene evidence [37] tab of HuGE Navigator. Thirty three genes had a score of 0.05 or more which were uploaded on the Candidate gene SNP selection (Genepipe) pipeline of "SNPinfo" a web-based SNP selection tool [38]. The algorithm used for selecting SNPs is as follows: Five kb upstream and 1 kb downstream of the gene coordinate were included in the selection. Only SNPs showing a minor allele frequency (MAF) of 0.05 or greater were included. Tagging proportion cut-off to filter gene was kept 0.8 and LD threshold cut off was kept 0.8. Minimum number of snps tagged by a tag snp was kept as 3. In order to ensure that each gene has some coverage a minimum of 1 tag snp to a maximum of 100 tag snps per gene were included. Further SNPs were filtered using the functional SNP prediction in "Genepipe"

that cause an amino acid change, those that may alter the functional or structural properties of the translated protein, disrupt transcription factor binding sites, disrupt splice sites or other functional sites.

Quality Assessment of Genotyping: Genotyping was performed on the Illumina Hi-Scan using GoldenGate Genotyping (GGGT) Custom SNP Panel assay (Illumina Inc., San Diego, CA). GGGT assay was performed on 1204 cases and 1212 controls (Total: 2416) for 384 custom selected SNPs. Intraplate and interplate replicates (7% approx.) were included on all plates and in all batches. Blinded duplicates were also included on all plates as another QC measure. The reproducibility rate of all the replicate samples (n=160) for all the assays was >98%. Also negative controls were run in some of the assays to check for any inter sample contamination. After excluding 17 samples with call rate <90%, a total of 2399 samples were included in final analysis. Further, 16 SNPs with diffused clusters, 6 SNPs with call frequency <95%, 4 SNPs with MAF<1% and 6 SNPs with substantial deviation from Hardy-Weinberg Equilibrium ($p<0.001$) were excluded to have a list of 352 SNPs for final analysis. All SNPs had a Gen train score value of 0.4 and above leading to no exclusions of SNPs due to poor cluster quality.

Statistical Analysis: A chi-square test was used to verify whether the observed genotype frequencies were in Hardy-Weinberg equilibrium. Principal Component Analysis was conducted to evaluate the potential effects of population structure between the samples. There was no significant difference in eigenvector loadings for the first five factors showing that the regional differences in structure were a minor source of population variability. Therefore, the analysis were not conditioned on region. Unconditional logistic regression was used to estimate OR and corresponding 95% CI between genotypes and case status. The genotypes were coded as 0=wild type, 1=heterozygous and 2=homozygous variant. The models fitted were additive (continuous effect of increasing number of variant alleles - 0 versus 1 versus 2),

dominant (0 versus 1 and 2), recessive (0 and 1 versus 2) and genotypic (0 versus 1, 0 versus 2). Positive associations were defined as an OR larger than 1, whereas an inverse association was specified by an OR below 1. To limit the probability of false-positives due to multiple testing, a false discovery rate method of Benjamini and Hochberg [39] was used to calculate *q*-value. A false discovery rate cut-off of 0.05 was applied to select the top SNPs, which limited the probability of false-positives due to multiple tests that were carried out. All the analysis were performed using the statistical software Stata version 12.0 and PLINK v1.07 [40-41].

Results and Discussion: Out of 384 SNPs genotyped a total of 32 SNPs were excluded from final analysis due to various reasons mentioned above. From 352 SNPs which were analysed 4 SNPs in FGFR2 gene using genotypic model (homozygous dominant v/s homozygous recessive) i.e. rs1219648, rs2420946, rs2981575 and rs2981582 showed positive association having OR 1.32 (1.02-1.70), 1.42 (1.10-1.82), 1.33 (1.04-1.70), 1.31 (1.02-1.68), 1.47 (1.11-1.94) respectively with breast cancer. FGFR2, fibroblast growth factor receptor 2, encodes a receptor tyrosine kinase that is amplified and over expressed in breast cancers. Polymorphisms in FGFR2 associated with breast cancer conferred a 20% increased risk of breast cancer among heterozygotes and a 60% increased risk among homozygotes with variant allele when compared to wild type homozygotes [42-43]. Recently a meta analysis has also observed a similar association of rs1219648, rs2420946 and rs2981582 in Caucasians and East Asians in ER+/PR+ tumours of breast cancer [44].

rs374748 on FBN2 (Fibrillin) which had been associated with obesity in previous studies [45] have been found to be positively associated with breast cancer in this study which may be due to well known association of obesity and breast cancer. Some of the other SNPs which had shown association with body mass index (BMI) [46] and obesity [47-48], weight gain or overweight and showed positive association with breast cancer in this study are rs2922763

Hepatocyte Nuclear Factor 4-Gamma (HNF4G), rs2116830 (KCNMA1 - Potassium Channel, Calcium-Activated, Large Conductance, Subfamily M, Alpha Member 1) and rs10953454 (LHFPL3 - Lipoma HMGIC Fusion Partner-Like 3). A positive association was observed with SNPs rs11121832, rs16886165, rs11594610 and rs2274459 in genes MTHFR (Mitochondrial Carrier Homolog 2), MAP3K1 (Mitogen-Activated Kinase Kinase Kinase 1), TCF7L2 (Transcription Factor 7-Like 2) and MLN (Motilin) respectively.

SUMMARY AND CONCLUSIONS

The strongest risk factors associated with breast cancer after adjusting for confounding variables are as follows:

1. For every 2 year increase in the age at first full-term pregnancy there is a 10% increase in risk of breast cancer.
2. For every 5 cm increase in height there is an increase of breast cancer with OR = 1.09 (95% CI-1.02-1.17).
3. WHR showed strong significant positive association with breast cancer in both rural-urban areas and in pre- and postmenopausal women. The risk was more than 3-fold in highest category (≥0.95) as compared to lowest category (<0.85).
4. Four SNPs selected from FGFR2 gene were positively associated with breast cancer. Some of the other SNPs identified in this study are rs11121832, rs16886165, rs11594610, rs2116830 and rs2274459 in genes MTHFR, MAP3K1, TCF7L2, KCNMA1 and MLN respectively. These are SNPs related to inflammation, obesity and signal transduction pathway.

The current study demonstrates that protection observed for breast cancer by living in a rural area is possibly because of less prevalence of risk factors viz. late age at first full-term pregnancy and central obesity which are observed to be strongly involved in the disease etiology. It's therefore possible to adopt public health strategies to prevent/reverse increasing

trends of breast cancer by monitoring the lifestyle. The strategies to reduce central obesity (and not only BMI) should be evolved as this will be helpful not only in the prevention of breast cancer but also other non communicable diseases. Efforts should be made to prevent late age at first pregnancy by proper counselling and informing about the risk associated with it.

BIBLIOGRAPHY

[1] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11. http://globocan.iarc.fr 2013.

[2] Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. Lancet Oncol 2012;13:790–801.

[3] Forman D, Bray F, Brewster D, Gombe MC, Kohler B, Piñeros M, et al. Cancer Incidence in Five Continents. vol. X. Lyon: IARC; 2013. http://ci5.iarc.fr last accessed on Last accessed on [2014 Feb 10]

[4] National Cancer Registry Programme. Three years Report of Population Based Cancer Registries 2009-11. New Delhi: Indian Council of Medical Research; 2013.

[5] Nagrani R, Budukh A, Koyande S, Panse N, Mhatre S, Badwe RA. Rural Urban differences in Breast Cancer in India. Accepted in Indian J Cancer 2014.

[6] Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. Epidemiol Rev 1993;15:36.

[7] Shin A, Song Y-M, Yoo K-Y, Sung J. Menstrual factors and cancer risk among Korean women. Int J Epidemiol 2011;40:1261–8.

[8] Lipworth L, Bailey LR, Trichopoulos D. History of breast-feeding in relation to breast cancer risk: a review of the epidemiologic literature. J Natl Cancer Inst 2000;92:302–12.

[9] Suzuki R, Iwasaki M, Inoue M, Sasazuki S, Sawada N, Yamaji T, et al. Body weight at age 20 years, subsequent weight change and breast cancer risk defined by estrogen and progesterone receptor status—the Japan public health center-based prospective study. Int J Cancer 2011;129:1214–24.

[10] Jung MM, Colditz GA, Collins LC, Schnitt SJ, Connolly JL, Tamimi RM. Lifetime physical activity and the incidence of proliferative benign breast disease. Cancer Causes Control 2011;22:1297–305.

[11] Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. Lancet Oncol 2011;12:785–94.

[12] Cai Q, Long J, Lu W, Qu S, Wen W, Kang D, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. Hum Mol Genet 2011;20:4991–9.

[13] Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11. 2 and 14q24. 1 (RAD51L1). Nat Genet 2009;41:579–84.

[14] Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet 2013;45:392–398, 398e1–2.

[15] Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45:353–361, 361e1–2.

[16] Joshi NN, Kale MD, Hake SS, Kannan S. Transforming growth factor β signaling pathway associated gene polymorphisms may explain lower breast cancer risk in western Indian women. PloS One 2011;6:e21866.

[17] Kohaar I, Tiwari P, Kumar R, Nasare V, Thakur N, Das BC, et al. Association of single nucleotide polymorphisms (SNPs) in TNF-LTA locus with breast cancer risk in Indian population. Breast Cancer Res Treat 2009;114:347–55.

[18] Tulsyan S, Agarwal G, Lal P, Agrawal S, Mittal RD, Mittal B. CD44 gene polymorphisms in breast cancer risk and prognosis: a study in North Indian population. PloS One 2013;8:e71073.

[19] Dikshit RP, Nagrani R, Mhatre S. Guidelines and Working Manual for conducting interviews for Multi-site Case Control studies. Mumbai, India: Tata Memorial Centre; 2011.

[20] Dikshit RP, Nagrani R, Mhatre S. Guidelines and Working Manual of data entry for Multi-site Case Control studies. Mumbai, India: Tata Memorial Centre; 2012.

[21] StataCorp. Stata Statistical Software. College Station, TX: StataCorp LP; 2011.

[22] Dey S, Soliman AS, Hablas A, Seifeldin IA, Ismail K, Ramadan M, et al. Urban-rural differences in breast cancer incidence by hormone receptor status across 6 years in Egypt. Breast Cancer Res Treat 2010;120:149–60.

[23] Mathew A, Gajalakshmi V, Rajan B, Kanimozhi V, Brennan P, Mathew BS, et al. Anthropometric factors and breast cancer risk among urban and rural women in South India: a multicentric case–control study. Br J Cancer 2008;99:207–13.

[24] Feng B-J, Jalbout M, Ayoub WB, Khyatti M, Dahmoul S, Ayad M, et al. Dietary risk factors for nasopharyngeal carcinoma in Maghrebian countries. Int J Cancer J Int Cancer 2007;121:1550–5.

[25] Cowgill KD, Loffredo CA, Eissa SA-L, Mokhtar N, Abdel-Hamid M, Fahmy A, et al. Case-control study of non-Hodgkin's lymphoma and hepatitis C virus infection in Egypt. Int J Epidemiol 2004;33:1034–9.

[26] Falk RT, Fears TR, Hoover RN, Pike MC, Wu AH, Nomura AMY, et al. Does place of birth influence endogenous hormone levels in Asian-American women? Br J Cancer 2002;87:54–60.

[27] Okasha M, McCarron P, Gunnell D, Smith GD. Exposures in childhood, adolescence and early adulthood and breast cancer risk: a systematic review of the literature. Breast Cancer Res Treat 2003;78:223–76.

[28] Willett W. Nutritional epidemiology. vol. 40. Oxford University Press; 2012.

[29] National Family Health Survey India (NFHS-1): 1992-93. Maharashtra, India.: International Institute for Population Sciences (IIPS); 1995.

[30] National Family Health Survey India (NFHS-3): 2005-06. Maharashtra, India.: International Institute for Population Sciences (IIPS); 2008.

[31] Friedenreich CM. Review of anthropometric factors and breast cancer risk. Eur J Cancer Prev 2001;10:15–32.

[32] Amadou A, Ferrari P, Muwonge R, Moskal A, Biessy C, Romieu I, et al. Overweight, obesity and risk of premenopausal breast cancer according to ethnicity: a systematic review and dose-response meta-analysis. Obes Rev 2013;14:665–78.

[33] Connolly BS, Barnett C, Vogt KN, Li T, Stone J, Boyd NF. A meta-analysis of published literature on waist-to-hip ratio and risk of breast cancer. Nutr Cancer 2002;44:127–38.

[34] Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. Nat Genet 2008;40:124–5.

[35] Yu W, Yesupriya A, Wulf A, Hindorff LA, Dowling N, Khoury MJ, et al. GWAS Integrator: a bioinformatics tool to explore human genetic associations reported in published genome-wide association studies. Eur J Hum Genet EJHG 2011;19:1095–9.

[36] Hindorff L, MacArthur J (European BI, Morales J (European BI, Junkins H, Hall P, Klemm A, et al. A Catalog of Published Genome-Wide Association Studies. Www.genome.gov/gwastudies n.d.

[37] Yu W, Wulf A, Liu T, Khoury MJ, Gwinn M. Gene Prospector: an evidence gateway for evaluating potential susceptibility genes and interacting risk factors for human diseases. BMC Bioinformatics 2008;9:528.

[38] Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 2009;37:W600–W605.

[39] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 1995:289–300.

[40] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.

[41] Purcell S. http://pngu.mgh.harvard.edu/purcell/plink/ n.d.

[42] Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93.

[43] Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4.

[44] Wang H, Yang Z, Zhang H. Assessing interactions between the associations of fibroblast growth factor receptor 2 common genetic variants and hormone receptor status with breast cancer risk. Breast Cancer Res Treat 2013;137:511–22.

[45] Cotsapas C, Speliotes EK, Hatoum IJ, Greenawalt DM, Dobrin R, Lum PY, et al. Common body mass index-associated variants confer risk of extreme obesity. Hum Mol Genet 2009;18:3502–7.

[46] Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010;42:937–48.

[47] Jiao H, Arner P, Hoffstedt J, Brodin D, Dubern B, Czernichow S, et al. Genome wide association study identifies KCNMA1 contributing to human obesity. BMC Med Genomics 2011;4:51.

[48] Wang K, Li W-D, Zhang CK, Wang Z, Glessner JT, Grant SF, et al. A genome-wide association study on obesity and obesity-related traits. PloS One 2011;6:e18939.

List of Abbreviations

BC	Breast Cancer
BMI	Body Mass Index
CI	Confidence Interval
EPIC	European Prospective Investigation into Cancer and Nutrition
ER	Estrogen Receptor
ER-	Estrogen Receptor Negative
ER+	Estrogen Receptor Positive
FDR	False Discovery Rate
GC	Gen Call
GWAS	Genome Wide Association Study
HC	Hip Circumference
HER2	Human Epidermal Growth Factor Receptor 2
HER2-	Human Epidermal Growth Factor Receptor 2 Negative
HER2+	Human Epidermal Growth Factor Receptor 2 Positive
HR	Hormone Receptor
HR-	Hormone Receptor Negative
HR+	Hormone Receptor Positive
HRT	Hormone Replacement Therapy
HuGE	Human Genome Epidemiology
HWE	Hardy-Weinberg Equilibrium
IARC	International Agency for Research on Cancer
LD	Linkage Disequilibrium
NIH	National Institute of Health
OC	Oral Contraceptive
OR	Odds Ratio
PR	Progesterone Receptor
PR-	Progesterone Receptor Negative
PR+	Progesterone Receptor Positive
RNA	Ribose Nucleotide Polymorphism
RR	Relative Risk
SNP	Single Nucleotide Polymorphism
TMH	Tata Memorial Hospital
TNBC	Triple Negative Breast Cancer
UTR	Untranslated Region
WC	Waist Circumference
WCRF	World Cancer Research Fund
WHO	World Health Organization
WHI	Women's Health Initiative
WHR	Waist-to-hip ratio

List of Genes

ACE	Angiotensin I-Converting Enzyme
ADIPOQ	Adiponectin, C1Q And Collagen Domain Containing
AGT	Angiotensinogen
BRCA1	Breast Cancer 1, Early Onset
BRCA2	Breast Cancer 2, Early Onset
COX-2	Cyclooxygenase-2
CSDE1	Cold Shock Domain Containing E1, RNA-binding
CYP1B1	Cytochrome P450 subfamily I dioxin-inducible polypeptide 1
ESR1	Estrogen receptor 1
GNPDA2	Glucoseamine-6-Phosphate Deaminase 2
IL6	Interleukin 6
INSIG2	Insulin-Induced Gene 2
KCNMA1	Potassium Channel, Calcium-Activated, Large Conductance, Subfamily M, Alpha Member 1
LEP	Leptin
MAP3K1	Mitogen-Activated Kinase Kinase Kinase 1
MAT2B	Methionine Adenosyltransferase II, Beta
MC4R	Melanocortin 4 Receptor
MTHFR	Methylenetetrahydrofolate reductase
PHB	Prohibitin
TCF7L2	Transcription Factor 7-Like 2
TMEM18	Transmembrane Protein 18
XRCC2	X-ray repair complementing defective repair in Chinese hamster cells 2
XRCC3	X-ray repair complementing defective repair in Chinese hamster cells 3
ZNF577	Zinc Finger Protein 577

List of Figures

List of Tables

Chapter 1

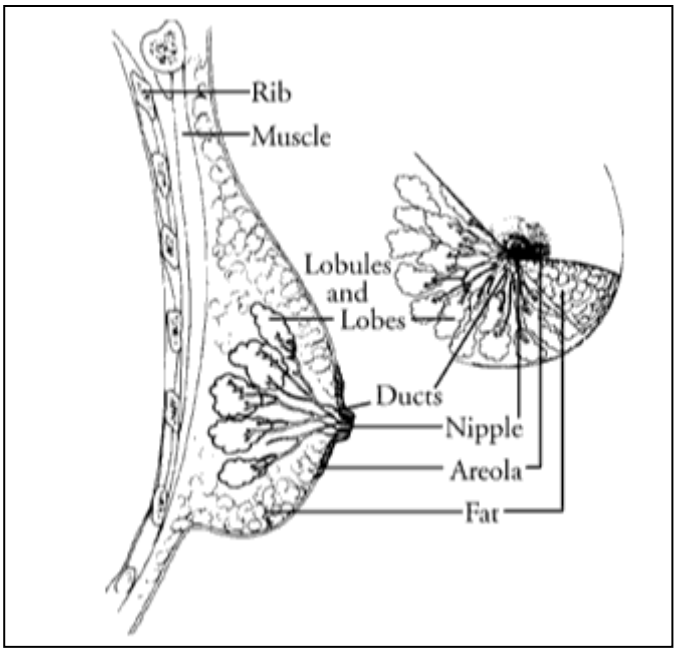
Introduction

1.1 Biology of Breast Cancer

1. 1.1 Anatomy of Breast

The adult breast sits atop the pectoralis muscle, atop the ribcage. The breast tissue extends horizontally (side-to-side) from the edge of the sternum out to the midaxillary line. There are about 15 to 20 lobes in each breast [1]. Each lobe has 20 to 40 lobules. Small ducts are attached to the lobules. These ducts join together like branches of grape stems into increasingly larger ducts. There are about 10 duct systems in each breast, each with its own opening at the nipple [2] (Figure 1.1).

Figure 1.1: Structure of the Breast



The breasts can be divided into quadrants for purposes of location of abnormalities. The four quadrants are the:

- UIQ: Upper Inner Quadrant
- LIQ: Lower Inner Quadrant
- UOQ: Upper Outer Quadrant
- LOQ: Lower Outer Quadrant

The exact locations within the quadrants can be represented by viewing each breast separately as a clock face. The majority of Breast Cancer (BC) occur in the upper outer quadrant of the breast.

The blood supply to the breast is derived from 3 sources. The predominant supply of blood comes from the perforating branches of the internal mammary arteries, derived from the internal thoracic artery. The breast is further supplied by the lateral thoracic and thoracoacromial arteries (branches of the axillary artery) as well as posterior intercostal arteries (branches of the thoracic aorta).

Venous drainage of the breast is mainly accomplished by the axillary vein. The subclavian, intercostal, and internal thoracic veins also aid in returning blood to the heart.

The lymphatic drainage of the breast deserves special attention, due to its role in the metastasis of cancer cells. The majority of lymph (>75%), particularly from the lateral quadrants, drains to the axillary lymph nodes. The remainder of lymph drains to either the parasternal nodes or the opposite breast (medial quadrants) or the inferior phrenic nodes (lower quadrants). With the exception of the nipple and areola, lymph from the skin of the breast drains into the axially, inferior deep cervical, infraclavicular, and parasternal nodes (depending on the location of the vessel) [3].

1.1.2 Tumour size and lymph node involvement

Tumour size is defined as the largest diameter of the tumour and is a prognostic factor for BC death regardless of other tumour characteristics [4,5]. Lymph node involvement is another important independent prognostic factor [5]. Women with lymph node involvement have poorer prognosis compared to women without lymph node involvement, and increasing number of affected lymph nodes are associated with poorer prognosis. Tumour size and lymph node involvement are correlated, and none of them seems to be predictive of treatment effect [6].

1.1.3 Morphology

Breast tumours are almost exclusively adenocarcinomas. Rarely, sarcomas or lymphomas develop, but these tumours are generally excluded when studying BC. The morphology of the breast tumour has been in clinical use for a long time, and current classifications are modifications from the classification made by Fraser in 1927 [7]. They are simply classified by their morphological appearance in the microscope. Still, the underlying carcinogenesis resulting in different histological types is largely unknown, and combinations between different types are common.

The two most common histological types of BC are derived from the breast glandular ducts and lobules, respectively. Ductal tumours make up the majority of BCs, and lobular cancers compose 5-15% of BCs [8]. Compared to ductal cancer, lobular cancer is more common among older women and is more often ER+, multifocal and bilateral. The metastatic pattern is also somewhat different. Despite these differences, ductal and lobular BC have similar prognosis [9,10]. There are also other rarer but well-defined histological types of BC; mucinous, medullary, papillary and tubular cancers. Tubular cancers are by definition of low grade, and correctly classified of having excellent prognosis [10,11].

1.1.4 Estrogen and Progesterone Receptors

Estrogen Receptors (ER) and Progesterone Receptors (PR) belong to the nuclear receptor super family.

The classic mechanism of these receptors is to be activated by ligands (estrogen and progesterone) that bind to the receptor. The ligand-bound receptor then binds another ligand-receptor complex. Together with coactivators, corepressors, and other transcription factors in the cell nucleus this dimer binds to promoter regions of the DNA thereby influencing gene transcription [12]. In normal breast tissue, the concentrations of ERα are low, and expressed in cells in the tubulo-lobular alveolar unit of the breast. The cells expressing ERα almost

never simultaneously express proliferation markers. They are instead expressed in adjacent cells. In premalignant breast tissue, ERα is expressed at higher concentrations in a larger proportion of the cells, and often together with proliferation markers. In BC, 60-80% of tumours express ERα, often at high levels [13]. The proportion of tumours expressing ER increases with increasing age [14]. PRs exist in two variants, PRA and PRB. The two variants come from the same gene, but are regulated by two different estrogen-regulated promoters [15]. In normal breast tissue, PRA and PRB are similarly expressed, while in atypical hyperplasia, non-invasive and invasive BC, PRA and PRB are heterogeneously expressed in adjacent cells, and PRA is often much more expressed than PRB in noninvasive and invasive cancers [16].

Currently, receptor status is assessed with immunohistochemical methods, with at least 10% positive nuclei as a common cutoff [17]. ER and PR are correlated to each other. Absence of PR in Estrogen Receptor Positive (ER+) tumours has been found to be correlated to tamoxifen resistance, and proposed to be an indication of nonfunctioning ER. However, recent data indicate that these tumours are not resistant to aromatase inhibitors, and that absence of PR instead indicates increased growth factor signaling [18,19].

ER and PR are prognostic factors, in that the survival pattern differs between receptor positive and negative tumours [14]. Estrogen Receptor Negative (ER-) and Progesterone Receptor Negative (PR-) tumours have a high mortality peaking around two years after diagnosis, then crossing the receptor positive curves to a much lower mortality rate. On the other hand, ER+ tumours have a rather constant mortality. Consequently, ER+ tumours have a better survival in the first years after diagnosis, but 15 years after diagnosis, the BC survival is unrelated to ER status [6,20]. ER and PR are treatment predictive factors. A majority of tumours expressing ER and PR respond to anti-estrogenic therapy, both in the adjuvant and metastatic setting. ER+PR- tumours respond to tamoxifen, but not as good as ER+/PR+ tumours [21],

and recent data indicate that these tumours are more likely to respond to aromatase inhibitors [18]. ER- tumours do not respond to anti-estrogenic therapy [6] Human Epidermal Growth Factor Receptor 2 (HER2) proto-oncogene encodes a tyrosine kinase situated in the cell membrane. It is over-expressed in approximately 30% of BCs and associated with more aggressive tumour characteristics and poorer survival [5].

1.1.5 Molecular Subtypes

Most studies divide BC into four major molecular subtypes:

- Luminal A
- Luminal B
- Triple negative/basal-like
- HER2 type

Other less common molecular subtypes have also been described including normal breast-like, apocrine molecular type and claudin-low type. BCs that do not fall into any of these subtypes are often listed as unclassified.

At this time, molecular subtypes are used mostly in research settings and are not included in pathology reports. Prognosis and treatment decisions are still guided by tumour stage, hormone receptor status and HER2 status.

The complex profile of each subtype is determined using molecular and genetic information from tumour cells. However, some characteristics (including hormone receptor status, HER2 status and proliferation rate) can be used to roughly define the four major subtypes (Table 1.1). Much of what is known about the four subtypes is related to these characteristics that are already well understood. Most BCs are luminal tumours. Luminal tumour cells look the most like the cells of BCs that start in the inner (luminal) cells lining the mammary ducts.

Table 1.1: Molecular Subtypes of Breast Cancer

Subtype	These tumours tend to be[a]	Prevalence (approximate)
Luminal A	ER+ and/or PR+, HER2-, low Ki67	40%
Luminal B	ER+ and/or PR+, HER2+ (or HER2- with high Ki67)	20%
Triple negative/basal-like	ER-, PR-, HER2-	15-20%
HER2 type	ER-, PR-, HER2+	10-15%

Abbreviations:ER+ Estrogen Receptor Positive, ER-, Estrogen Receptor Negative; HER2+, Human Epidermal Growth Factor Receptor Positive; HER2-, Human Epidermal Growth Factor Receptor Negative; PR+, Progesterone Receptor Positive; PR- = Progesterone Receptor Negative

[a]These are the most common profiles for each subtype. However, not all tumours within each subtype will have all these features.

Adapted from selected sources [22,23].

1.1.5A Luminal A

Luminal A tumours tend to be ER+ and/or Progesterone Receptor Positive (PR+), Human Epidermal Growth Factor Receptor 2 Negative (HER2-) and are tumour grade 1 or 2. Fewer than 15% of luminal A tumours have *p53* mutations, a factor linked with poorer prognosis [22].

Of the four subtypes, luminal A tumours tend to have the best prognosis, with fairly high survival rates and fairly low recurrence rates [23,24]. Because luminal A tumours tend to be ER+, treatment for these tumours often includes hormone therapy.

1.1.5B Luminal B

Luminal B tumours tend to be either ER+ and/or PR+. They are highly positive for Ki67 (have a high number of cancer cells actively dividing) and/or Human Epidermal Growth Factor Receptor Positive (HER2+)

Women with luminal B tumours are often diagnosed at a younger age than those with luminal A tumours [25] and, compared to luminal A tumours, they tend to have factors that lead to a poorer prognosis including [26] poorer tumour grade, larger tumour size, lymph node-positive

and *p53* gene mutations (about 30%). In some studies, women with luminal B tumours have fairly high survival rates, although not as high as those with luminal A tumours [22,24].

1.1.5C Triple negative/basal-like

Triple negative breast cancers (TNBC) are ER-, PR-, HER2-. There are several subsets of TNBC. One subset is referred to as basal-like because the tumours have cells with features similar to those of the outer (basal) cells surrounding the mammary ducts. Most basal-like tumours contain *p53* mutations [22].

Most triple negative tumours are basal-like and most basal-like tumours are triple negative. However, not all triple negative tumours are basal-like and not all basal-like tumours are triple negative. About 15-20% of breast cancers are triple negative or basal-like [22,23]. These tumours tend to occur more often in younger women and African-American women [22,25,27,28]. And, most Breast Cancer 1, Early Onset (BRCA1) BCs are both triple negative and basal-like [27,29,30]. Triple negative/basal-like tumours are often aggressive and have a poorer prognosis (at least within the first five years after diagnosis) compared to the ER+ subtypes (luminal A and luminal B tumours) [31].

1.1.5D HER2 type

The molecular subtype HER2 is not the same as HER2+ and is not used to guide treatment. Although most HER2 type tumors are HER2+ (and named for this reason), about 30% are HER2-. HER2 type tumors tend to be ER-, PR-, Lymph node-positive and poorer tumour grade [24,32].

About 10-15% of BCs have this molecular profile [22,23]. About 75% of HER2 type tumours contain *p53* mutations [32].

HER2 type tumours have a fairly poor prognosis and are prone to early and frequent recurrence and metastases [28,33,34]. Women with HER2 type tumours appear to be diagnosed at a younger age than those with luminal A and luminal B tumours [23].

1.2 Descriptive Epidemiology

Every day, thousands of women around the world from all walks of life are diagnosed with BC. It is by far the most common cancer amongst females worldwide with nearly one million new cases each year, representing one in five of all female tumours. Overall BC accounts for 21% of all cancer diagnoses in women. BC is the most common cancer in women in high-, middle- and low-income countries [35].

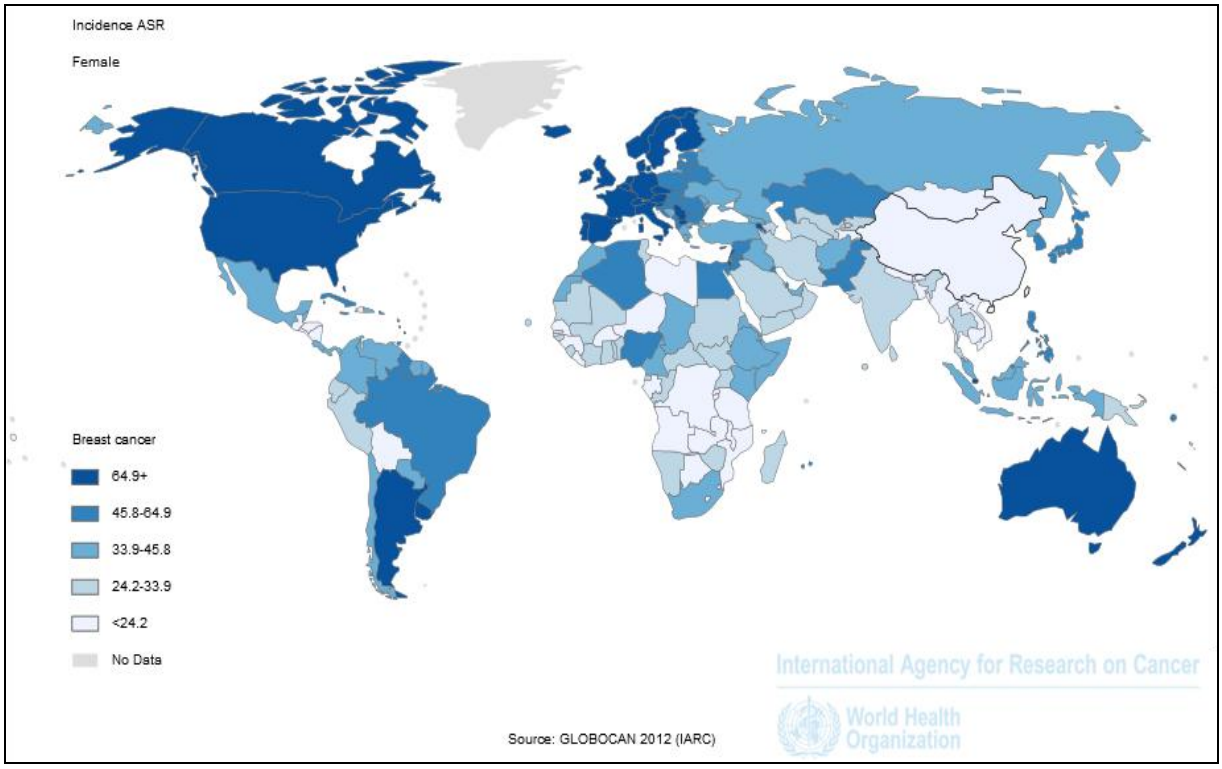
1.2.1 Burden of Disease

Worldwide 1,676,633 women were diagnosed with BC. The burden of BC is higher in less developed regions with 882, 949 cases than in more developed regions with 793, 684 cases estimated by Globocan, 2012. India itself has burden 144,937 BC cases. This implies that, though, the percentage of total women affected seems less, the BC burden in India has almost reached about 2/3rds of some of the developed nations and is steadily rising [36].

1.2.2 Incidence

BC incidence is fast increasing in economically transiting countries though incidence rates in high income countries are nearly three times higher than in middle- to low-income countries. Around the world, age adjusted incidence rates range from 75-100 per 100, 000 women in North America, Northern Europe, and Australia, to less than 20 per 100, 000 in parts of Africa and Asia [37] (Figure 1.2). The adaptation of a western lifestyle – an increased prevalence of ill-defined series of reproductive, hormonal and dietary determinants in the population – has been postulated as a primary reason for the increasing BC incidence rates observed among Asian and Asian American women [38].

Figure 1.2: Age standardized (world) Incidence rate (per 100,000) of Female Breast Cancer (All ages).



1.2.3 Survival

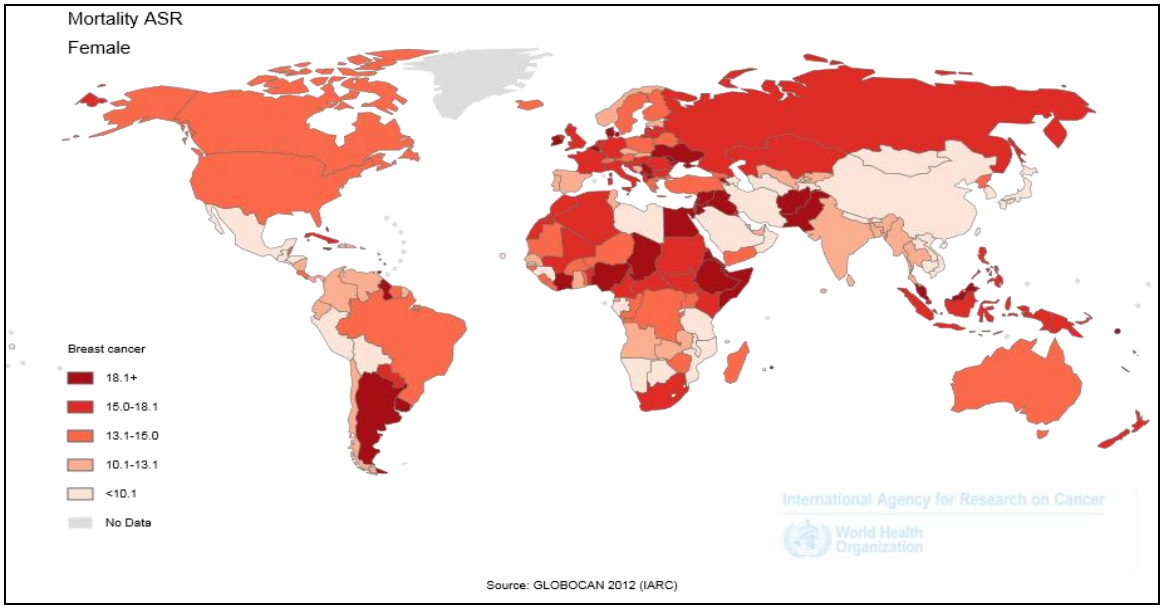
The 5-year survival for female BC is higher than for most other types of cancer. Five year survival ranges from 90 to less than 50%, depending on the characteristics of the tumour, its size and spread, and the availability of treatment [39]. A considerable difference has also been reported in average 5-year survival in low to middle income countries having less than 60% in Brazil and Slovakia and less than 40% in Algeria as compared to high income countries having average 5-year survival proportion of more than 80% in North America, Sweden, Japan, Finland and Australia [39]. The low survival at the end of 5 years in middle- and low-income countries can be explained mainly by a lack of early detection programmes, resulting in a high proportion of women presenting with late-stage disease, as well as by a lack of adequate diagnosis and treatment facilities [40–43]. Educational and cultural barriers also exist for women in less developed countries which often lead to late presentation, such as lack

of awareness of BC, an incorrect belief that the disease is incurable or contagious, the stigma of having a mastectomy and fear of rejection by their partner or community [44,45].

1.2.4 Mortality

The World Health Organization (WHO) has estimated that female BC resulted in a total of 5,884,000 years of life lost globally during 2004. This represented just over 1% of all premature mortality amongst females, but there was a large amount of variation in this proportion between regions, ranging from around 8% in parts of Europe to less than 0.5% in Africa [46]. There is a three-fold variation in mortality by regions of the world, with rates in excess of 20 deaths per 100,000 in Southern Africa, Western Africa and Northern Europe in contrast to 7 to 9.4 deaths per 100,000 in Eastern and Southern Asia [37,47] (Figure 1.3). Rapid increases in mortality have been reported in parts of Asia, Africa and Central/South America [48,49], which have been attributed to rising incidence in conjunction with lower survival. This contrasts with widespread decreasing trends in BC mortality rates of between 2.0% and 3.0% per year throughout North America, parts of Europe and Australia that generally commenced around the late 1980s/early 1990s [50,51].

Figure 1.3: Age Standardized (world) Mortality Rate (per 100,000) of Female Breast Cancer (All ages).



1.2.5 Time trends in Incidence

The most recent incidence data indicate signs of a plateau in time trends in developed world during the mid-1990s, particularly in the Netherlands, Sweden and in England and Wales [52]. Differential trends among pre- and postmenopausal women has been observed in many European countries, where the increases in incidence were observed to be relatively minor in 35–49 years old women but greater in women 50–69 years old. In the U.S. and Canada, BC incidence among postmenopausal women increased in the 1980s and 1990s, stabilized in the late 1990s [53] and declined around 2003 [54], most likely due to saturation of mammography screening [55].

In economically-transiting countries like India and China, incidence rates are increasing, and are predicted to increase further in the next few decades [56,57]. Most registries in India have exhibited rising incidence rates of BC in recent years [58]. As an example, reproductive lifestyle factors appear to be changing in India, with the percentage of women married by the age of 18 declining from 54.2% in 1992-93 to 44.5% in 2005-06. Similarly parity has reduced from 3.39 live-born children per woman delivered in 1992-93 to 2.68 by 2005-06. The use of contraceptive pill has increased from 1.2% to 3.1% [59]. Further the observations may also be explained by differences in the prevalence of specific risk factors in India that increase the risk of pre- or postmenopausal BC, such as obesity [60,61].

1.2.6 Incidence rates in Rural and Urban India

The rates are fast increasing in developing countries like India [62]. However, there are substantial differences in the incidence rates of BC within rural and urban areas of India. Rates observed in metro registries are in the range of 29 – 35 per 100, 000 whereas those observed in rural registries vary from 11 -12 per 100,000 [58]. The lowest BC incidence rates are found among women from the rural area of Barshi in Western India, and Dindigul Amblikkai, another rural area in the more developed South of India (Table 1.2). An increasing

order of rate ratio was observed in the present study from rural to urban to metro regions, clearly suggesting the underlying differences in the incidence rates between rural and urban regions (Table 1.3). A twofold increased risk was observed in urban areas and a threefold increased risk was observed in metro areas compared to rural areas [63]. The cause of this strong urban: rural difference is not known although it is likely to be due to one or more lifestyle factors whose prevalence differs strongly between rural and urban women.

Table 1.2: Breast Cancer incidence in South Asian and Western Population (AAR)

Rural Population	**Town/Small City Population[a]**	**Urban Population[a]**	**Western Population[b]**
Barshi Rural(12.30) [a]	Barshi Town (14.40)	Bangalore (36.60)	US - SEER 9 White (91.8)
Ahmedabad – rural (11.10)[a]	Aurangabad (18.80)	Bhopal (27.40)	UK, England Thames (82.6)
Dindigul Amblikkai (13.80)[c]		Chennai (32.60)	
		Mumbai (31.00)	

Abbreviations: AAR, Age Adjusted Incidence Rate (world) per 100,000 population
[a] NCRP (2009 – 2011)
[b] Cancer Incidence in five Continents Vol X (2003 – 2007)
[c] Personal Communication

Table 1.3: Incidence rate and Rate ratio of developing Breast Cancer in selected cancer registries stratified by Rural, Urban, and Metro regions

Regions	Indian registry	Year	AAR	Rate Ratio	95% CI
Rural	Barshi	2009-2010	12.30	Reference	
	Ahmedabad rural	2009-2010	11.08	0.90	0.64-1.26
Urban	Aurangabad	2009-2010	18.78	1.53	1.13-2.06
	Bhopal	2009-2010	27.39	2.23	1.76-2.82
	Wardha	2010-2011	18.26	1.48	1.11-1.98
Metro	Bangalore	2008-2009	36.65	2.98	2.51-3.54
	Chennai	2009	32.63	2.65	2.17-3.25
	New Delhi	2008-2009	32.18	2.62	2.19-3.12
	Mumbai	2008-2009	30.97	2.52	2.10-3.01
	Nagpur	2009-2010	32.46	2.64	2.15-3.24
	Pune	2009-2010	23.27	1.89	1.52-2.36
	Thiruvananthapuram	2009-2011	35.07	2.85	2.34-3.47
North East regions	Aizawl District	2009-2010	30.33	2.47	1.67-3.64
	Cachar District	2009-2010	16.44	1.34	1.00-1.78
	Dibrugarh District	2009-2011	10.63	0.86	0.62-1.21
	East Khasi Hills	2010-2011	12.10	0.98	0.67-1.44
	Imphal West District	2009-2010	14.36	1.17	0.80-1.69
	Manipur (excluding Imphal West)	2009-2010	7.59	0.62	0.42-0.91
	Kamrup Urban District	2009-2011	22.76	1.85	1.43-2.39
	Manipur	2009-2010	9.14	0.74	0.52-1.06
	Meghalaya	2010-2011	9.10	0.74	0.51-1.07
	Mizoram	2009-2010	16.40	1.33	0.97-1.84
	Mizoram (excluding Aizawl)	2009-2010	8.54	0.69	0.45-1.06
	Nagaland	2010	9.52	0.77	0.42-1.43
	Sikkim	2009-2011	8.56	0.70	0.46-1.04
	Tripura	2010	7.16	0.58	0.39-0.87

Abbreviations: AAR, Age adjusted incidence rate (world) per 100,000; CI, Confidence interval; NCRP, National Cancer Registry Program.
Source: NCRP (2009-2011).

1.3 Etiology

1.3.1 Reproductive Factors

Some reproductive factors modify sex hormone levels; reduction in overall estrogen exposure may partly explain the link between reproductive factors and BC risk.

1.3.1A Age at Menarche and Menopause

Menarche and menopause are markers of onset and cessation, respectively, of ovarian and related endocrine activity associated with reproduction. During women's reproductive years (broadly the time between menarche and menopause) the ovary produces steroid hormones that directly affect development and function of the breast. Early menarche and late menopause are known to increase women's risk of developing BC. BC risk increases by 5% for each year younger at menarche and a 3% increase has been observed for each year increase in menopause, a meta-analysis has shown [64] . Most of the cohort studies have shown direct relation between age at menopause and BC risk [65–69].

An early age at menarche is thought to be associated with an increased risk of BC because a higher number of lifetime ovulatory cycles, and hence greater exposure to ovarian hormones, has been shown to confer an elevated risk of BC [65]. The association between age at menarche and BC is stronger for ER+ and PR+ tumours than for ER- and PR- tumours [70].

1.3.1B Age at first full-term birth and Parity

Compared to nulliparous women, mothers with their first full-term birth before 20 years of age had a 50% reduced risk of BC. On the other hand, those who had their first baby after age 35 had a 22% increased risk. BC risk decreases by 7% with each live birth [70–73] and increases by 3% for each year older a woman is when she first gives birth, meta- and pooled analyses have shown [71]. These relative risk (RR)s were comparable across countries. The protective effect of early age of first full-term birth in parous women was similarly observed

in other studies [74] except for one study from Japan [75]. The association of age at first full-term pregnancy has been found to be similarly associated with pre- and postmenopausal women [76,77]. Many reports observed a protective effect of early age at first full-term birth on Hormone Receptor Positive (HR+) cancers [34,78,79]. A meta-analysis has also revealed a reduced risk among patients with HR+ cancers [70]. When stratified on receptor status, the association between parity and age at first full-term birth and BC risk however was shown to be limited to ER+/PR+ tumours [34,70,80].

1.3.1C Interval between age at menarche and age at first full-term pregnancy

Given the susceptibility of the undifferentiated nulligravid breast to carcinogenic insults, the duration of time between age at menarche and age at first full-term birth may be independently related to BC risk. However, few epidemiologic studies have evaluated this relation. Clavel-Chapelon [65] addressed this issue to some extent in the French E3N cohort by evaluating the relation between the number of menstrual cycles women had before their first full-term birth and BC risk. Compared with women in the lowest quartile, women in the highest quartile of cumulative number of cycles before their first full-term birth had a 1.42-fold [95% Confidence Interval (CI): 1.20–1.67] elevated risk of BC. This risk was essentially the same when women who had used Oral Contraceptives (OC)s were excluded from the analysis. In a combined analysis of 7 case-control studies, Andrieu et al. found similar results. BC risk for women with 21 or more years between menarche and first childbirth was 1.45-fold higher (95 % CI: 1.17–1.82) than that for women with 10 years or less between these two events [81]. A longer duration between age at menarche and age at first full-term birth was associated with an elevated risk of BC, except among premenopausal African-American women. The elevations in risk observed were largely confined to women with HR+ tumours [82]. The large body of data indicates that the risk of BC overall increases with the increase in interval between age at menarche and age at first full-term pregnancy.

1.3.1D Breastfeeding

Breastfeeding has been hypothesized to reduce the risk of BC. However, findings haven't been consistent for the association between BC risk and ever breastfeeding or cumulative breastfeeding duration [83,84]. However, a reduction has been seen most consistently observed among premenopausal women who breastfed for an extended period, but even here the magnitude of the observed effect has varied substantially [83]. Breastfeeding appears to lower the risk of both ER+ and PR- BCs [70]. A pooled analysis from 47 epidemiologic studies, including 50,302 cases and 96,973 controls, showed a significant, 4.3% reduction in BC risk for every 12 months of breastfeeding [71]. A systematic review carried out by Berrino et al. for the World Cancer Research Fund/ American Institute for Cancer Research (WCRF/AICR) included 80 epidemiologic studies. The meta-analysis on four cohort studies as well as that on 37 case-control studies showed a 2% reduction of risk per 5 months of breastfeeding [85] . In a systematic review on Japanese population, cohort studies failed to find a significant inverse association between breastfeeding and the risk of BC and most of the case-control studies observed a statistically significant or non-significant risk reduction for women who ever had breastfed or for women with a longer duration of breastfeeding [86]. In a case-control study conducted in India, where longer duration of breastfeeding is more common as compared to the western population showed an inverse association with BC in premenopausal women, whereas no such protective effect was observed in postmenopausal women [87]. The current literature indicates a weak protection in the development of BC in women who have breast fed for a longer duration.

1.3.1E Induced and Spontaneous Abortions

The relationship between induced abortion and the subsequent development of BC has been the subject of a substantial amount of debate in epidemiologic studies. In contrast to a recent

meta-analysis conducted in Chinese women [88] which largely included retrospective studies (34 case-control studies and 2 cohort studies), prospective studies conclude there is no association between induced abortion and BC [89–95]. A worldwide meta-analysis of 83,000 women examined the relationship between induced abortion and BC and found a significant difference between the overall estimate of RR from studies that had recorded information on induced abortion prospectively (RR = 0.93; 95% CI: 0.89–0.96) and the overall estimate of RR from studies that had recorded such information retrospectively (RR = 1.11; 95% CI: 1.09–1.14), suggesting that reporting bias was probably present in studies using retrospective reporting of abortion history [96].

Findings from cohort studies and a large pooled analysis have shown spontaneous abortion (also known as miscarriage) does not increase the risk of BC [90,93,94,96]. On the other hand premenopausal BC appeared to be less frequent in women who had repeated miscarriages suggesting BC association with spontaneous abortion is possible and may depend on menopausal status [97].

The current literature is divided on the association of spontaneous abortion and BC risk, whereas the results are inconclusive for the association of induced abortion with BC risk

1.3.1F Oral Contraceptives

Studies show that current or recent use of OCs (birth control pills) slightly increases the risk of BC [98–100]. The Women's CARE Study examined the risk of BC associated with OCs among different subgroups of women. In this study, there was no increased risk of BC among current users (RR = 1.0, 95% CI: 0.8–1.0) or former users (RR = 0.9, 95% CI: 0.8–1.0). This study found no increased risk among women with a family history or those who initiated use at an early age. In addition, the risk of BC did not appear to vary by duration, dose or type of progestin [101]. Similarly, a recent systematic review showed that, the RR of BC declines after OC cessation, such that 10 years after cessation no excess risk remains. BC risk does not

appear to increase with longer duration of OC use [102]. A meta-analysis has shown that the risk associated with OC is similar across OC formulations (which have changed considerably over time), family history, and ethnicity [103].

Studies that evaluated the risk by ethnicity observed effect estimates greater for black women [101,104] than for white women [105]. In a follow-up study of Norwegian women, the RR estimate was 1.6 (95% CI: 1.2–2.1) for women who were current or recent OC users at baseline [99]. Another follow-up study in the Netherlands, showed long duration OC use was associated with increased BC risk among women aged 55 years or older but not younger women [106]. In a Long Island case-control study of BC, recent OC use and long duration OC use were associated with increased BC risk among premenopausal women but not among postmenopausal women [107]. In the population based Carolina BC Study, results were close to the null for white women, but OC use within the previous 5 years was associated with increased risk among black women [104]. With regard to the hormone status of the tumour, some studies have found stronger associations of OC use with ER- cancer than with ER+ cancer [108,109], but others have found no difference [110–114]. The current literature suggests that OC use increases the risk of BC in current long term users.

1.3.1G Non-oral hormonal contraceptives

Hormonal contraception is also available as injections, implants and patches. There is substantially less evidence on cancer risk associated with these preparations than there is on cancer risk associated with the OCs. BC risk is increased among users of injectable contraceptives in some studies [111,112], while other study showed no association [113]. In a case-control study a significantly increased association between BC risk and implants was observed [115]. The literature for the association non-oral hormonal contraceptives with BC risk is inconsistent and more studies with larger sample size will be required to estimate a true association.

1.3.1H Tubal Ligation

The US Collaborative Review of Sterilization reported reduced menstrual bleeding and pain and increased cycle irregularity after tubal ligation [116]. These findings provided evidence against a 'post tubal ligation syndrome' that included dysmenorrhoea and menorrhagia, but could not address long term outcomes, such as altered menopausal age [117], symptoms [118–120], or BC risk. A recent meta-analysis reported no association between tubal ligation and BC, however, substantial heterogeneity was observed. Effect estimates among eight studies ranged from 0.37 (95% CI: 0.19–0.68) to 1.20 (95% CI: 1.00–1.30) [121]. This variability may be partly due to incomplete information on subsequent gynaecologic surgeries and tumour subtypes. Few studies have evaluated variation by tumours that express ER or PR and may therefore be more sensitive to hormonal exposures [122,123]. Similarly, in a recently conducted case-control study, tubal ligation did not have an impact on BC overall (Hazards Ratio = 0.95; 95% CI: 0.85–1.06), but had a suggested inverse relation with ER+/PR+ invasive tumours (Hazards Ratio = 0.84; 95% CI: 0.70–1.01), possibly because of subsequent hysterectomy/bilateral oophorectomy [124]. The current literature does not show any association of tubal ligation with BC risk.

1.3.1I Age at last full-term pregnancy

Age at last full-term pregnancy did not show an association [Odds Ratio (OR) = 1.01; 95% CI: 0.97–1.06] with BC [125].

In European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, age at last full-term pregnancy was not associated with the risk of ER-/PR- malignancies but was associated with ER+/PR+ tumours, however no statistical heterogeneity between the BC subtypes was observed [80].

1.3.1J Twin Pregnancy

Twin pregnancies differ from singleton pregnancies in both hormone levels and perinatal changes [73]. Some studies have suggested that twin births may be associated with lower BC risk [126,127]. Although in pooled results of all 17 published studies did not show a reduced maternal risk of BC for twin births (Hazards Ratio = 0.94; 95% CI: 0.87–1.02; *P* = 0.127), a trend toward reduced maternal risk of BC was identified in a subgroup analysis of cohort studies (Hazards Ratio = 0.91; 95% CI: 0.83–1.01; *P* = 0.068). The results of the only meta-analysis suggest that twin pregnancy does not significantly decrease the maternal risk of BC [73]. The current literature is inconclusive on the relationship between twin pregnancies and risk of BC.

1.3.1K Duration since last birth

Liu et al. [128] from the Swedish Fertility Register, with over 30, 000 BC case subjects available for study documented a small increase in the risk of BC for each of the first few years after birth, with adjustment for age at delivery in 1-year increments. Other studies, with considerably smaller numbers of white women, have produced mixed results: some observed an increased risk for shorter interval since last birth [129–131] and a few found no association [132–134]. Duration since last birth has been shown to be associated with ER+/PR+ tumours and not with ER-/PR- tumours [80].

1.3.2 Anthropometric Measurements

1.3.2A Height

Height, representing intrauterine, early childhood as well as the level of adolescent growth spurt, likely relates to factors such as nutrition, genetic growth potential, and hormones thus influencing BC occurrence [135–138]. A positive association between adult height and BC has been found in a large number of studies [61,135,139,140]. In a review of seven large prospective cohort studies, the multivariate-adjusted RR of BC per 5 cm increment of height

was 1.02 (95% CI: 0.96–1.10) in premenopausal women and 1.07 (95% CI: 1.03–1.12) among women of postmenopausal status [139]. A meta-analysis conducted on premenopausal women found an overall weak association with each increment of 10cm in height [141]. Another study showed a positive association of BC risk with postmenopausal women [142]. However some studies have found no association at all with height in pre- or postmenopausal women of European descent [143]. Previous studies have generally not shown any clear differences for overall height associations by ER/PR status of the BC cases [144–148]. Previous studies have consistently associated tallness with increased risk of BC overall.

1.3.2B Body Mass Index (BMI)

Most available studies and meta-analyses have focused on BMI as a marker of general obesity [112,139,149–153]. Several studies supported the hypothesis that higher level of BMI may be associated with a decrease in the risk of premenopausal BC. This hypothesis is supported by results from several case-control studies [109,143,154,155] and cohort studies [156,157]. However, others studies did not observe a statistically significant association when comparing highest versus lowest levels of BMI [61,158,159]. Ethnicity appears to modify this association because while the inverse association between BMI and risk of premenopausal BC is well documented in Caucasians, the association among Asian women is inconsistent. Several studies among Asian women suggest that higher BMI may be associated with an increased risk of premenopausal BC [149,150,160,161]. A prospective study including 11,889 women from Taiwan reported that higher BMI was moderately associated with an increased risk of premenopausal BC [161], with an OR of 1.90 (95% CI: 1.00–3.4) for BMI > 26.2kg/m^2 versus 21.6kg/m^2. In contrast, other studies among Asian women did not detect a significant association between BMI and the risk of premenopausal BC [162,163]. In a recent meta-analysis it has been shown that premenopausal BMI does not relate to BC risk [164].

An overall increase in the risk of postmenopausal BC in overweight or obese women among all ethnic groups has been indicated. The association between BMI and risk of postmenopausal BC was found to be stronger among women who did not use hormone replacement therapy (HRT) compared to women who did use hormones [165]. A dose-response meta-analysis (9 cohorts: 22 case-control studies) showed that the BMI-BC association is stronger for ER+/PR+ tumours (33% increase per 5kg/m^2 increment for postmenopausal BC), while there were no significant BMI-cancer associations for ER-/PR- tumours [166].

1.3.2C Waist-to-Hip ratio (WHR)

WHR is commonly used as a measure of central obesity [167,168]. It has not been consistently associated with increased BC risk in premenopausal women, for whom both null [61,155] and increased risk have been reported [143,167–169]. Two meta-analyses [168,169] have reported that a greater WHR was associated with about 1.5-fold increased risk of premenopausal BC. A pooled analysis on 7 cohorts and 4 case-control studies reported a summary risk estimate of 1.79 (95% CI: 1.22–2.62) [169] but the strength of the association varied according to ethnic groups [170–173]. Other studies [146,161,174] did not find a statistically significant association. Overall, this increased risk associated with larger WHR among premenopausal women is found to be stronger amongst Asian women compared to other ethnic groups.

A WHR of above 0.85 for females has often been associated with the risk of developing postmenopausal BC. However, while most studies have reported a significant increased risk [155,169,175], some studies are inconclusive [61,143,156]. A meta-analysis with 6 case-control and 5 cohort studies observed a summary risk estimate of 1.50 (95% CI: 1.10–2.04) for postmenopausal women [169]. These associations tend to be stronger in Asian women than other ethnic groups. In contrast, some studies conducted in the US did not detect a

significant association. Hall et al. reported a non increased RR of 1.62 (95% CI: 0.70–3.79) in African-American women and of 1.64 (95% CI: 0.88–3.07) for Caucasian American women when comparing highest versus lowest quintiles (0.86–1.34 versus 0.6–0.77) [143]. However the power of the study was limited by small number of cases (179 cases and 182 controls in African women). Regarding Hispanic women, only one study has assessed the association between WHR and BC risk. This study found no significant association between WHR and postmenopausal BC risk [176]. Current literature largely suggests that a high WHR is associated with increased risk of premenopausal and postmenopausal BCs.

1.3.2D Waist Circumference (WC)

Among premenopausal women, WC is generally not related to risk of BC in most studies but positive associations have been found when adjusted for BMI [165]. Recent results from the Nurses' Health Study II showed a strong increase in the risk of ER- BC among premenopausal women with increasing WC (RR = 2.75; 95% CI: 1.15–6.54) [167]. In postmenopausal women, studies that did not adjust for BMI showed a 7% increased risk per 8cm increase in WC and those that did, showed a 4% increased risk [85]. In the Women's Health Initiative (WHI) study, WC was associated with BC risk among postmenopausal women, but only in those who never used HRT [177].

1.3.2E Hip Circumference (HC)

An inverse association in premenopausal women with HC adjusted for BMI was found in some studies [61,178]. An inverse association was also observed in Nigerian BC Study with an OR of 0.36 for the highest quartile (95% CI: 0.24–0.55). The association existed in both pre- and postmenopausal women [179]. In contrast, other studies showed a positive association between HC and BC risk [177,180]. Again, In WHI, HC was positively associated with both ER+/PR+ and ER-/PR- subtypes of premenopausal BC [181]. The evidence of the association between HC and BC risk has been largely inconsistent [61,177–181].

1.3.2F Adult Body Weight

A number of epidemiological studies have reported that both early adult body weight [157,182,183] and a subsequent change in body weight [157,183–185] are associated with BC risk. Several of these have reported an inverse association between body weight in early adulthood and the incidence of BC [183,185].

It has been postulated that the association between body weight and BC risk may be heterogeneous according to the tumour's ER and PR status. Cumulative epidemiological evidence [139,186,187] also suggests that the impact of body weight on BC risk differs across women's menopausal status. Recent meta-analysis of cohort and case-control studies could clarify that overweight is not significantly related to risk of premenopausal BCs [164]. A positive association among postmenopausal women has been observed. Large weight gain since age 20 has been shown to be associated with increased risk of BC [149,188], particularly in postmenopausal women aged >60 years [189–191]. Any weight change since the age of 18 seems not to be related to premenopausal BCs [192]. A large body of data suggests that early adult body weight and a subsequent change in body weight are associated with BC risk.

1.3.3 Other Factors

1.3.3A Physical Activity

Physical activity is a modifiable factor that is associated with a decreased risk for both premenopausal and postmenopausal BC [193,194].

BC risk is around 25% lower in the most active women compared with the least [195]. BC risk decreases by 5% for every 2 hours per week increment in recreational activity (moderate and vigorous), a meta-analysis showed [196]. Light intensity activity may be insufficient to reduce BC risk, a Canadian case-control study indicated [197]. Further BC risk declined with increasing time spent on household activities, a factor which is more prevalent in rural women

as compared to urban women [198]. Thus it can be concluded that there is sufficient evidence for the role of physical activity in preventing BC [199].

1.3.3B Occupation

Villeneuve et al. [200] in a case-control study (1230 cases) observed a statistically significant BC excess after 10 years duration in motor vehicle manufacturing (obs/exp= 18/7=2.6 (95% CI: 1.00–6.30). Labrèche et al. [201] found significant excesses of postmenopausal cancer for polycyclic aromatic hydrocarbons (PAHs), and several polymeric fibers. Clapp et al found risk was elevated among postmenopausal women whose husbands used specific pesticides [202]. A recent study found that young women exposed to DDT before the age of 14 had an excess BC risk before age 50 [203]. Band et al. [204] found in pre- and postmenopausal cases (combined) elevated BC risk in fruit and other vegetable farming (OR = 3.11, 90% CI: 1.24–7.81).

In meta-analysis of 13 observational studies found a 48% (RR = 1.48; 95% CI: 1.36–1.61) increased risk of BC among shift workers [205]. Exposure to light at night is associated with higher levels of sex hormones, because it disturbs the circadian system, which suppresses melatonin production, and melatonin is thought to reduce circulating estrogen [206,207]. This may partly explain the link between shift work and BC risk, but confounding by other lifestyle factors such as tobacco use, BMI and physical activity is possible [208,209].

1.3.3C Ionizing Radiation

Exposure to ionizing radiation is a well established cause of somatic DNA mutations. BC risk is increased after several types of previous cancer, with radiotherapy an important factor in this association. BC risk is nonsignificantly increased in survivors of childhood solid cancer who received radiotherapy, compared with those who did not receive radiotherapy [210]. BC risk is 9-11% higher in women who received radiotherapy for cancer in the opposite breast, compared with women who had surgery [211,212].

Diagnostic radiology involves much lower radiation doses than radiotherapy. An estimated 0.1% of BC in women aged 75 and under are caused by exposure to diagnostic x-rays [213]. X-ray-associated BC risk is further elevated in women with BRCA1 or Breast Cancer 2, Early Onset (BRCA2) mutation [214]. Mammograms are associated with a very small number of BC: of 10,000 women who are screened every three years between the ages of 47 and 73, between three and six will develop cancer during their lifetime because of mammogram radiation [215]. Exposure to computed tomography (CT) scans in childhood or adolescence does not appear to be linked with increased BC risk [216]. The ionizing radiation thus has been consistently associated with increased risk of BC.

1.3.3D Diet

BC risk decreases with higher consumption of fruit and vegetables [217], dietary fibre (at least 25g/per day) [218], some carotenoids [219], lignans (postmenopausal women) [220], soya-based foods (Asian populations only) [221,222], flavonols and flavones (postmenopausal women) [223], and marine omega-3 polyunsaturated fatty acids (PUFA) [224]. BC risk is not associated with consumption of red meat [225,226], green or black tea [227], use of vitamin supplements [228], or vitamin D levels [229].

BC risk may be slightly increased with higher consumption of eggs, but no dose-response has been shown and confounding may be likely [230]. There was no evidence of an association between traditional dietary patterns and risk of BC [231], and only one study showed a significant increase in risk associated with the western dietary pattern [232]. Diets that include alcoholic beverages may be associated with increased risk [231]. Though links between BC risk and diet have been extensively studied, WCRF/IARC deems the evidence insufficient (due to quality, consistency and amount) to derive classifications as to the breast carcinogenicity of any dietary exposure except total dietary fat [85].

1.3.3E Family history and genetic factors

BC risk is around doubled in women with one first degree relative with BC, compared with women with no first degree relatives, meta- and pooled analyses have shown. The risk is further increased with a larger number of affected first degree relatives, or relatives affected aged under 50 [233]. The risk increase is similar for first degree relatives with ER+ or ER- BC [234].

Environmental and lifestyle factors explain around three-quarters of BC risk, with hereditary factors explaining only around a quarter [235]. The reasons for BC clustering in families remain largely unclear, but a small proportion of families share BC predisposition genes, some of which are discussed below.

High Penetrance Gene Mutations

BRCA1 and BRCA2 mutations confer a high risk of BC in carriers (high penetrance). Women with a BRCA1 or BRCA2 mutation have a 45-65% chance of developing BC by age 70 [236]. BRCA2 negative women with a BRCA2 carrying first degree relative may also have increased BC risk, a small UK cohort study showed [237]. Higher sex hormone levels in BRCA mutation carriers may explain some of the increased risk [238]. Early onset BC risk may be increased in BRCA mutation carriers born in the 1950s or later, suggesting possible interactions with lifestyle factors [239].

Other breast cancer predisposition genes

Li Fraumeni syndrome caused by Tumour Protein 53 (TP53) mutation and Cowden syndrome caused by Phosphatase and Tensin Homolog (PTEN) mutation are high-penetrance BC predisposition genes, but they are both rare and so account for a very low proportion of BC cases overall and among cases with first degree family history. Mutations in Checkpoint Kinase 2 (CHEK2), Ataxia Telangiectasia Mutated (ATM**)**, BRCA1 interacting protein C-

terminal helicase 1 (BRIP1), and Partner and Localizer of BRCA2 (PALB2) confer an intermediate risk of BC in carriers, but again are rare. Mutations in a number of other genes are more common but confer a lower risk of BC. BC risk in some other rare genetic mutation syndromes, such as Peutz-Jeghers syndrome caused by Serine/Threonine kinase 11(STK11) mutation, and hereditary diffuse gastric cancer syndrome [caused by cadherin 1, type 1, E-cadherin (epithelial) (CDH1) mutations], remains unclear [240,241].

Low Penetrance Polymorphisms

Several common Single Nucleotide Polymorphism (SNP)s associated with BC have been identified primarily through Genome Wide Association Study (GWAS) of very large case-control populations. These alleles occur with high frequency in the general population, although the increased BC risk associated with each is very small relative to the general population risk. GWAS on BC has largely been conducted in most developed countries [242–245] showing low to modest associations between common polymorphisms and BC risk. Susceptibility locus on Estrogen receptor alpha (ESR1) gene – a key mediator of ER in mammary tissue have consistently shown its association with BC [246]. A meta-analysis confirmed the association of polymorphisms rs1219648 (A > G), rs2420946 (C > T), and rs2981582 (C > T) in Fibroblast Growth Factor Receptor 2 (FGFR2) suggesting that FGFR2 is likely an important genetic marker contributing to susceptibility of BC [247].

1.3.3F Smoking

Tobacco smoking is classified by IARC as a probable cause of BC, based on limited evidence [248]. Tobacco smoking is associated with higher levels of sex hormones, which may partly explain the link between tobacco and BC risk [249].

BC risk is 12% higher in current smokers, and 9% higher in former smokers, both compared with never smokers, a meta-analysis has shown [250]. BC risk increases with amount, duration, and starting age of smoking [250,251]. The effect of smoking may be limited to

premenopausal BC and non obese women [251,252], and ER+ (not triple negative) BC [149,253].

1.3.3G Alcohol

In 2007, the IARC concluded that there is sufficient evidence that alcohol causes cancer of the female breast [254]. A meta-analysis has shown that even light drinkers (up to one alcoholic drink per day, or around 1.5 units) have a 5% higher BC risk compared with non-drinkers [255]. Studies have consistently demonstrated a linear dose-response relation between alcohol consumption and BC risk, with increases observed to be around 7-12% per unit of alcohol per day [256–258]. Although the exact mechanism for the association between alcohol consumption and BC is not known, one probable explanation would involve alcohol's effects on circulating estrogen levels. Most large studies have shown a stronger association with ER+ BCs [259–263]. Alcohol intake is thus the dietary factor most consistently associated with BC risk, although the relationship observed has generally been modest.

Table 1.4: Epidemiology of Breast Cancer: Risk Factor Summary

Risk Factor	Direction of Effect	
	Premenopausal	**Postmenopausal**
Well confirmed Risk Factors		
Family history	↑↑	↑↑
Benign breast diseases	↑↑	↑↑
Mammographically dense breast	↑↑	↑↑
Age at first >30 years versus, <20	↑↑	↑↑
Menopause at > 54 years versus, <45	-	↑↑
High endogenous estrogen levels	↑	↑↑
Postmenopausal hormone use	-	↑
Ionizing radiation exposure	↑↑	↑↑
Menarche at <12 years versus, >14	↑	↑
Alcohol use	↑	↑
High Body mass index	↓	↑
Probable relationship exists, Based on substantial data		
High endogenous androgen level	↑↑	↑↑
Current oral contraceptive use	↑	-
Physical activity	↓	↓
Lactation (longer duration)	↓	↓
Folate	↓	↓
Carotenoids	↓	↓
Weak, if any , Relationship exist , Based on substantial data		
Total dietary fat intake during childhood	-	-
Induced or spontaneous abortion	-	-
Cigarette smoking	-	-
Past oral contraceptive use	-	-
Exposure to electromagnetic field	-	-

↑↑: Moderate to large extent in risk, ↑: Slight increase in risk,↓: moderate to large decrease in risk ,-: no association.

Source: Textbook of Cancer Epidemiology, Second Edition, Hans Olav Adami, David Hunter, Dimitrios Trichopoulos

1.4 Gaps in Literature

It has been observed for long time that the rates of BC differ in rural and urban areas. However, there are very few studies in literature to address the reasons for the differences in BC rates of rural and urban area. Obesity has been observed to be risk factor for postmenopausal BC. However the contribution of different measures of obesity and their role in pre- and postmenopausal women is still not clear. In Indian context, there are no large studies to address the issue of reproductive factors, obesity, age at last pregnancy, OC use in development of BC. Though there has been large GWAS on BC in most developed countries [242–245] showing low to modest associations between common polymorphisms and BC risk. In India, however, there have been no GWAS and few properly designed retrospective studies with smaller sample size on genetic susceptibility to study this risk [264–267].

The present thesis proposal is designed to understand more clearly the reasons for rural-urban differences, and role of genetic susceptibility in development of BC.

HYPOTHESIS

Anthropometric and Lifestyle related variables are the cause of large differences in occurrence of BC in rural and urban areas.

AIM

Primary: To study role of anthropometric and other lifestyle related variables in causation of BC in rural and urban areas.

Secondary: To study role of genetic susceptibility in BC.

Chapter 2

Lifestyle & Breast Cancer Risk

2.1 Introduction

BC risk is largely driven by lifestyle and related factors. BC has largely been the disease of developed world, however India – an economically transiting country has been showing an increase in the incidence of BC. This increase has been largely observed by the cancer registries in the metro cities as compared to the rural registries, which show around 3-fold increased incidence rate as compared to rural registries. Further, there are large differences in pre- and postmenopausal cancer in urban registries [63].

This chapter addresses the reasons for difference in the incidence of BC in rural and urban India, by highlighting on the protective lifestyle of rural women. The differences in risk factors for pre- and postmenopausal BC are also presented to understand etiology of BC

Following sections are described in detail:

- Methodology and quality measures for questionnaire based study.
- Reproductive and BC risk.
- Anthropometric measurements and BC risk.

2.2 Study Design

A hospital based case-control study was conducted at Tata Memorial Hospital (TMH), Mumbai during the period of January 2009 to September 2013.

2.2.1 Criteria for enrolment of cases

The cases were female BC patients coming to TMH. Only primary BC cases aged 20-69 were enrolled in the study with date of diagnosis not more than 6 months from the date of interview. All the BC cases enrolled in the study were histologically confirmed.

2.2.2 Criteria for enrolment of controls

All female visitors with no history of cancer coming along with any site cancer patient aged 20-69 were included in the study. None of the cancer site patients with which visitors came

along with, constituted more than 20% of enrolled controls. The maximum number of enrolled visitor control came along with head and neck cancer patients while the minimum number of enrolled visitor controls came along with bone and soft tissue tumours. Out of 1515 visitor controls enrolled in the study 50% were first degree relatives (parents, siblings and children) while remaining were other relatives, friends and neighbours (first cousins, grandparents, friend, neighbour, in-laws, other distance relatives) of different cancer site patients. The selected controls thus belonged to the same study base from which cases were coming to TMH and the selection bias was kept at minimum.

2.2.3 Matching

Controls were frequency matched to cases on age (± 10 years) and region of residence at the time of enrolment. For region matching, India was divided into five regions which are as follows:

1. North (Haryana, Uttar Pradesh, Himachal Pradesh, Delhi, Punjab, Uttarakhand, Rajasthan, Bihar, Chandigarh, Jammu & Kashmir)
2. South (Kerala, Tamil Nadu, Andhra Pradesh, Karnataka, Puducherry, Lakshadweep, Andaman & Nicobar Islands)
3. East (Arunachal Pradesh, Assam, Meghalaya, Manipur, Mizoram, Nagaland, Tripura, West Bengal, Orissa and Jharkhand)
4. West (Maharashtra, Goa, Gujarat, Dadra Nagar Haveli, Daman & Diu)
5. Central (Chattisgarh and Madhya Pradesh).

The cases and controls were recruited simultaneously during the study period**.** The study has been approved by TMH Institutional Review Board. Written informed consent was obtained from all study participants before enrolling them in the study.

2.3 Data Collection

2.3.1 Questionnaire Data

Questionnaires were designed and tested by conducting mock and group interviews. The questions were reframed if required. The final questionnaire consisted of demographic and socioeconomic status, reproductive history, time spent in household activities on a normal day, residential history, occupational history, personal and family medical history, diet, tobacco and alcohol habits.

Anthropometric measurements were taken at the end of the interview.

2.3.2 Blood Collection

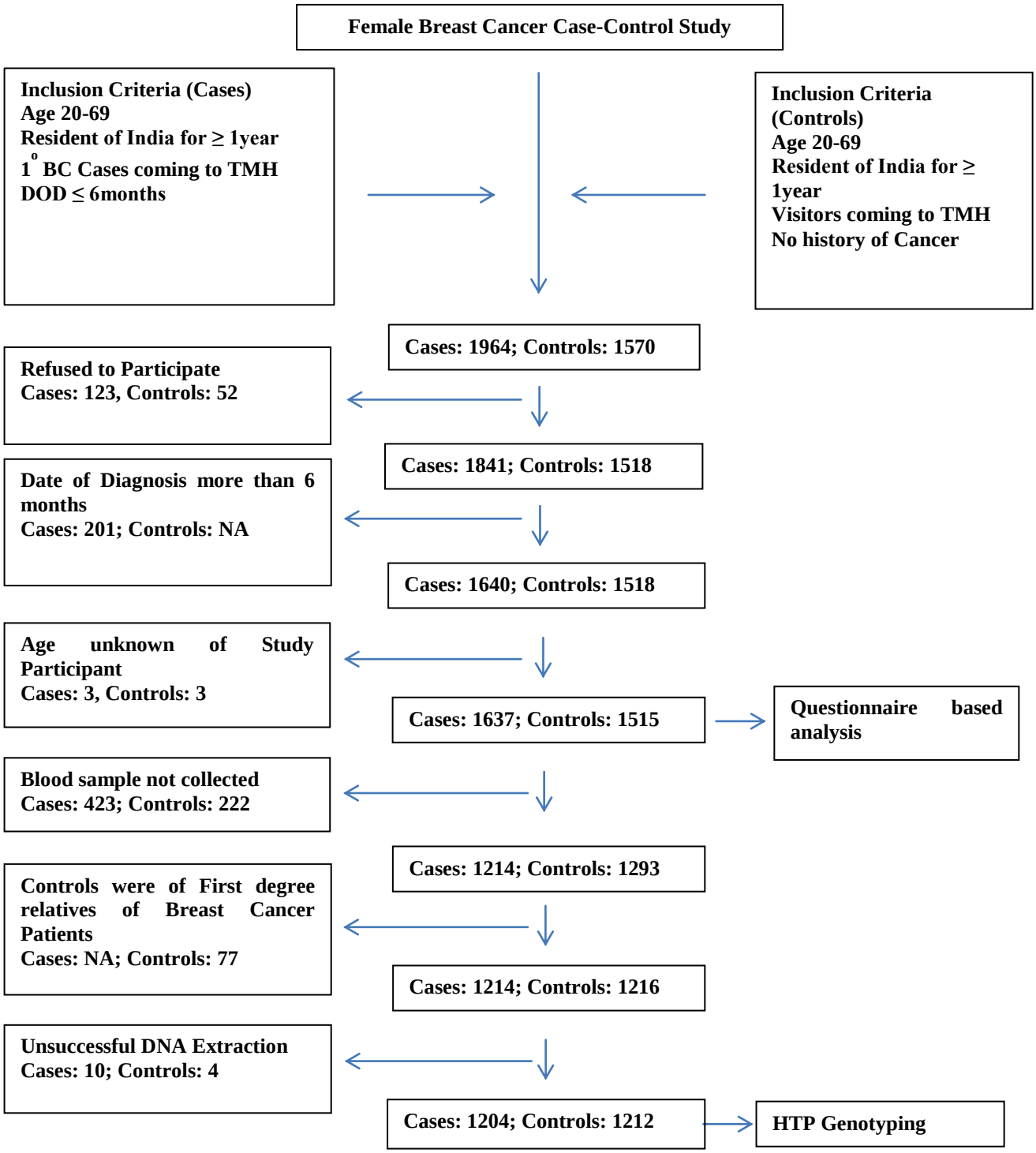
A 10ml blood sample was collected from each study participant and centrifuged into plasma and buffy coat. After separation the blood components were then stored at -80^0C immediately and transferred to Liquid Nitrogen Cylinder for long-term storage.

A flowchart describing the enrolment of study participants is shown in Figure 2.1

2.3.3 Hormone Receptor (HR) Status

ER, PR and HER2 status were obtained from hospital pathology records.

Figure 2.1: Flowchart representing enrolment of study participants in Case-Control Study



2.4 Quality Assessment for Questionnaire Based Data

2.4.1 Preparation of Instruction Manual for filling up the Questionnaire in Case-Control Studies

In order to assure the homogeneity of data collection by the social investigators, an instruction manual and video recording has been prepared. The instruction manual contains detailed guidelines and figures wherever required for better understanding of questions by the social investigator as well as the respondent [268].

2.4.2 Preparation of Instruction Manual for Data Entry

In order to assure the homogeneity while entering the data, clear and precise instructions with predefined logical checks have been listed in the form of Manual [269].

2.4.3 Monitoring of Daily Work

The forms were checked at three levels for completeness of information, first by the interviewer, immediately after taking the interview, second by the study co-ordinator, on following day of the interview and finally by the data entry operators, before entering the data. Weekly meetings were conducted to understand and resolve the problems of data collection. Training programs were conducted every quarter so as to ensure the quality of interviews. The questionnaire was checked daily for completeness of information.

2.4.4 Quality Checks on Data Entry

Logical Checks were prepared to identify errors in the data entry. The data was entered twice and corrected for errors between the 2 entries referring the hard copy of the questionnaire (Table 2.1), if any, occurred while entering the data.

Table 2.1: Example of Corrected Differences of Variables between Data Entry 1 & 2

Total number of Pregnancies			
Study ID	First Data Entry	Second Data Entry	Corrected Entry
820337	3	6	6
820514	4	3	3
820545	3	2	3
820589	7	9	7
820744	10	3	10
820810	3	1	3
820957	5	3	5
840574	3	2	3
840584	2	3	2
840630	2	1	2
840718	6	5	5
840893	5	4	4
840988	3	4	4
840989	2	22	2
841134	5	4	5
841198	7	6	7
841341	4	3	4
841411	6	5	6

2.4.5 Reproducibility of Questionnaire

Abbreviated questionnaire was designed. This questionnaire contains constant (non changing in recent time) variables such as number of pregnancies, height, age at menarche, age at first full-term pregnancy. The reproducibility questionnaire was completed for 249 study participants (approx. 8% of total enrolled in study). The interval between main questionnaire and reproducibility questionnaire was minimum of 7 days. The main questionnaire and reproducibility questionnaire was interviewed by 2 different interviewers. Details of main measured exposures are shown in Table 2.2.

Table 2.2: Reproducibility of Measured Exposure

Variable	Study Mean (Reproducibility Mean) N=249	Coefficient of Correlation (%)
Age	46.90 (47.17)	92.25
Number of Pregnancies	4.06 (3.99)	91.07
Height	156.92 (157.18)	96.51
Age at Menarche	13.95 (14.27)	76.86
Age at first full-term pregnancy	21.97 (21.75)	81.88
Age at last full term pregnancy	27.30 (28.57)	69.43
Current Residence	NA	90.64
Education	NA	87.15

2.4.6 Calibration of Study Instruments

To ensure study reliability, regular calibration process was performed on weighing balance, measuring tape, wall mounted stadiometer and centrifuge machine. Weighing balance, measuring tape and wall mounted stadiometer were calibrated twice a year using an unused weighing balance which was used as standard and a difference of ±1kg was considered acceptable. Similarly an unused measuring tape was used to calibrate the measuring tape and wall mounted stadiometer and a difference of ±1cm was considered acceptable. A yearly calibration was conducted for centrifuge machine by the supplier.

2.5 Exposure Assessment

2.5.1 Rural and Urban Status

All study participants were asked to list all places of residence where they had lived for at least 1 year, starting with the place of birth. The rural and urban residence status was self reported by study participant for each of the residence they mentioned to the interviewer. Study participants were then stratified into matrix to classify women into rural and urban using four different definitions as follows:

1. Ever lived in a rural area: If a study participant had ever lived in a rural area for 1 year or more in life were termed as a “rural participant”, otherwise the participant was termed as

"urban participant".

2. First 20 years of life lived in a rural area: If a study participant had lived first 20 years of her life in a rural area, i.e., from age 0 to age 20, then participant was classified as "rural participant," whereas any participant who had lived <20 years in a rural area in her entire life was classified as "urban participant".
3. Currently living in a rural area: Any study participant who has a current residence (at the time of enrolment) of 1 year or more in a rural area is termed as "rural participant", versus a current residence in an urban area is an "urban participant".
4. Total years lived in a rural area:
    a. 1-10 years: A minimum of 1 year and a maximum of 10 years lived in a rural area versus never lived in a rural area are categorized as rural and urban participants respectively.
    b. >10 years: If total years lived in a rural area is >10 years, study participant was categorized as rural or else urban.

2.5.2 Menopausal Status

Women whose menstrual period had stopped either naturally, or due to oophorectomy, hysterectomy or any other reason for 12 months or more from the date of enrolment were classified as postmenopausal. The rest were treated as premenopausal.

2.5.3 Hormone Receptor Status

The information on HR status i.e. ER, PR and HER2 was available on 1273 BC cases. The study participants were stratified into ER+/PR+, ER-/PR-, HER2+ and TNBC.

2.5.4 Reproductive Factors

Cases and controls were interviewed in-person by trained interviewers using a pre-tested structured questionnaire.

With respect to our primary exposures of interest, age at menarche was grouped as ages ≤12 (reference), 13–14, and 15-20 years. The interval between age at menarche and first full-term pregnancy was grouped as <10 years (reference) and ≥ 10years. Total number of pregnancies was classified as never (reference) and ever (inclusive of abortion, miscarriage, still birth and full-term pregnancies). Total number of full-term pregnancy was estimated by categorizing into 1 (reference), 2, 3 and ≥4. Age at first full-term pregnancy was categorized into <20 (reference), 20-21, 22-23, 24-25 and ≥26 years. The other categories that were used for age at first full-term pregnancy are ≤25 (reference) and >25 years. Age at last full-term pregnancy was grouped into women who had their last pregnancies at age ≤24 (reference), 25-29, 30-34 and ≥35. Duration since last birth (in years) was stratified into ≤10 (reference) and >10. Ever breastfeeding was defined as breastfeeding for at least one month or else categorized as never, which was used as reference. Total duration of breastfeeding was defined as duration of breastfeeding in life measured in months and was grouped into the following categories: never (reference), ≤12, 13-24, 25-36 and >36 months. Average duration breastfed per child (in months) was obtained by dividing total duration of breastfeeding in life measured in months with total number of live births. Average duration breastfed per child was categorized into ≤6, 7-12, 13-18, 19-24 and >24 months. Twin pregnancy and maternal risk of BC was estimated using never (reference) and ever categories. OC use was grouped into never (reference) and study participant who had used OC at least once in lifetime were grouped in ever category. The participants who had ever used OC were stratified into past users and current users keeping never users as reference. The current users were women who used OC in last 5 years whereas past users were women who had stopped OC use more than 5 years ago from the date of enrolment. Duration of OC use in current users were classified into Short-term users (≤ 2 years of OC usage) and Long-term users (> 2years of OC usage), using never users as reference. Age of OC use was stratified into ≤24 (reference), 25-29 and ≥30

years. Total duration of OC used (including past and current users) measured in years was categorised as never (reference), <1, 1-4 and ≥5 years. Any intentional expulsion of foetus was classified as "Induced abortion" whereas a naturally occurring expulsion of foetus was termed as "Spontaneous abortion" in the study. Number of induced and spontaneous abortion were grouped as 0 (reference), 1 and ≥2.

2.5.5 Anthropometric Measurements

Height (without shoes in cm) as shown in Figure 2.2 and weight in light clothing (in kg) of each study participant were measured using standard equipment. Weight was measured with light clothing. WC was measured halfway between the costal edge and iliac crest and HC was measured as the greatest circumference around the buttocks [270]. All measurements were done twice in succession and averaged for a final value. WHR was computed by taking the ratio of WC (in cm) and HC (in cm) and grouped into three categories, namely ≤ 0.84 (reference), 0.85-0.94 and ≥0.95. BMI (kg/m^2) was calculated by dividing weight in kg with square of height (in m^2). BMI was divided into 5 categories, in accordance with the WHO classification for Asian population [271]: < 18.5, 18.5–22.9 (reference), 23.0–24.9, 25–29.9, and ≥30. To test the effect of BMI in accordance with the WHO classification for world population [272], following stratifications were made: <18.5, 18.5-24.9 (reference), 25.0-29.9 and ≥30. Postmenopausal women were divided into two groups to analyse the association of BMI using both world and Asian categories. The two groups of postmenopausal women were those who had attained menopause <10 years ago and those who had attained menopause ≥ 10 years ago from the day of enrolment. Height was grouped into ≤ 150 (reference), 151-155, 156-160 and ≥161cm. Weight was categorized into ≤ 60 (reference), 61-65 and >65kgs. WC was grouped into ≤79 (reference), 80-85 and ≥86cm whereas categories for HC include ≤90 (reference), 91-99 and ≥100cm. Furthermore, a total of nine different body size pictogram (Figure 2.3) were shown to each study participant to indicate their body sizes at different

periods of life (at 10 years, 20 years and at the time of enrolment). Body size pictogram at age 10, 20 and at the time of enrolment were categorized into <3 (reference), 3-4, ≥5 figure. Increase in body size was estimated at two stages i.e. from age 10 to age 20 and from age 20 to age at the time of enrolment using the pictogram with categories of no increase (reference), moderate increase and drastic increase. No increase was defined when the body size of the study participant remained between 1 and 2. Moderate increase was defined when the body size of the study participant increased from 1-2 to 3-4. Drastic increase was defined when the body size of the study participant increased from 1-2 to 5-9.

Figure 2.2: Pictogram for Measuring Height

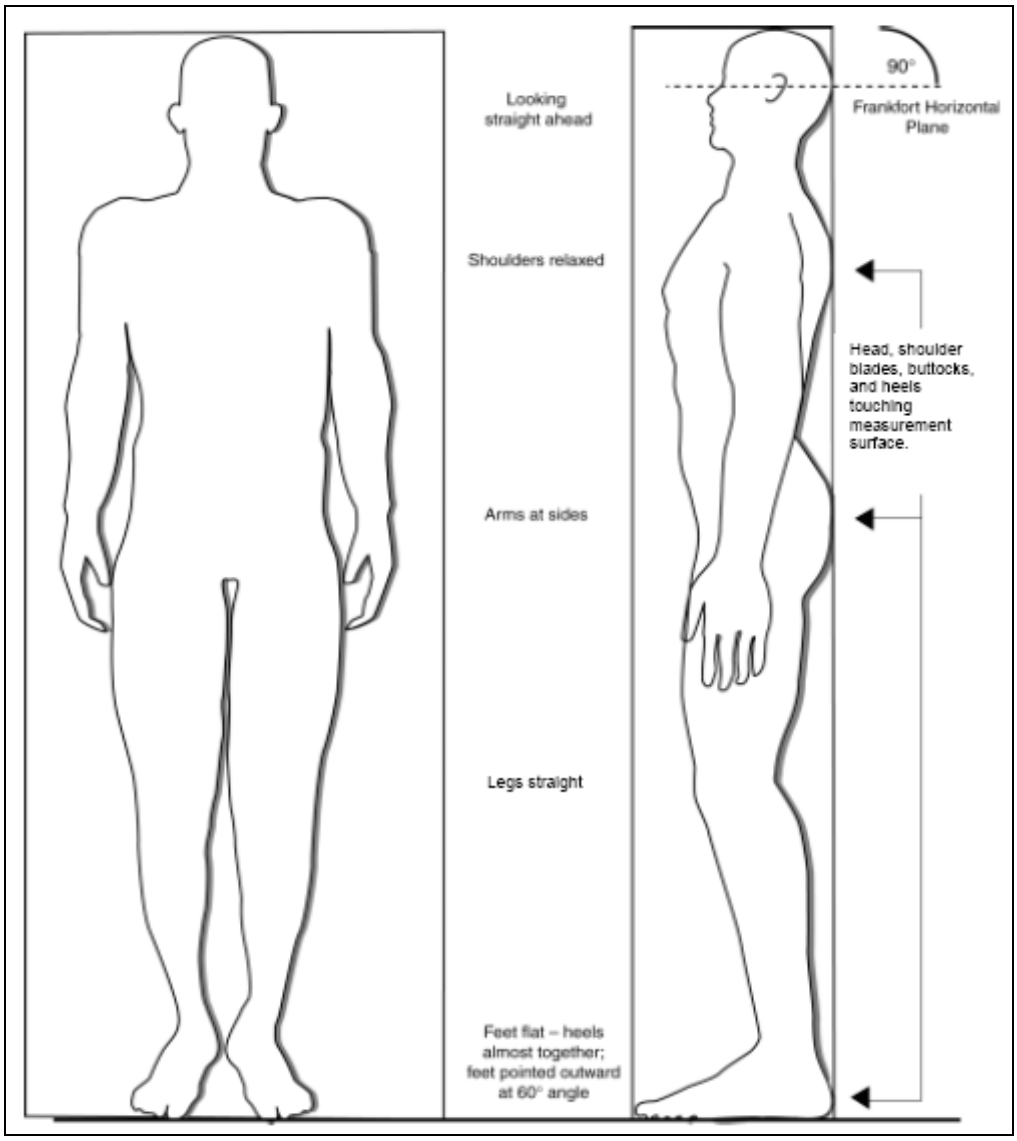
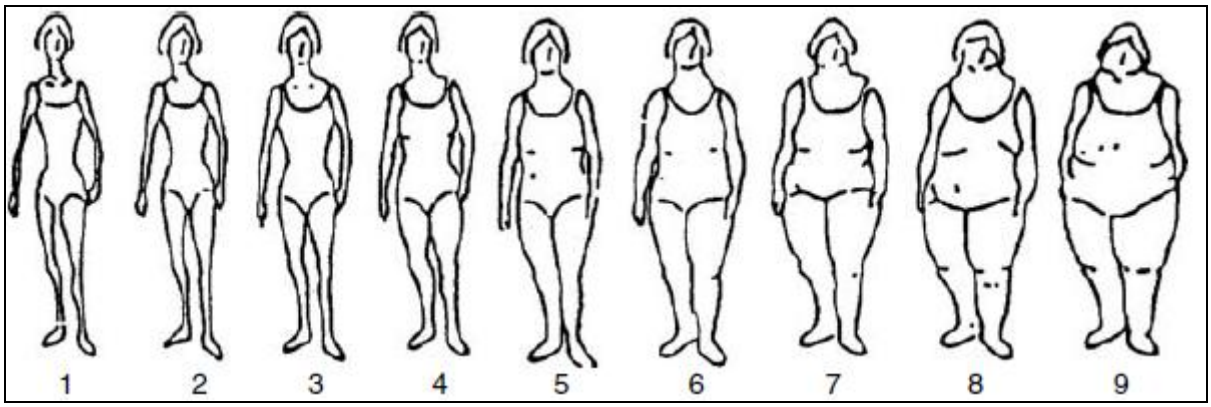


Figure 2.3: Body Size Pictogram



2.6 Statistical Analysis

Crude and adjusted ORs of developing BC and their 95% CI [273] for anthropometric measurements and reproductive factors were estimated separately by residential status (Rural/Urban) and menopausal status. Unconditional logistic regression models were adjusted for potential confounders such as age (continuous variable), region of residence (North, South, East, West and Central India), rural-urban status (Rural, Urban), education (Less than 5 years of schooling, ≥5 year of education), age at first full-term pregnancy (continuous variable), WHR (continuous variable), height (continuous variable), menopausal status (premenopausal, postmenopausal), number of induced abortion (continuous variable) and spontaneous abortion (continuous variable). The ORs for interval between menarche and first full-term pregnancy, age at first full-term pregnancy, age at last full-term pregnancy, duration since last birth and total duration of breastfeeding were estimated after adjusting for total number of pregnancies (continuous variable) instead of age at first full-term pregnancy. Weight in kg as continuous variable was entered in the model for estimating OR for WHR, WC, HC instead of WHR and weight was replaced for height for estimating the risk of latter. Increase in body size from age 10 to 20 and age 20 to age at enrolment were adjusted for BMI (continuous variable) instead of WHR. Study participants for whom values for one or more of the variables in the models were missing (0.82-6.73%) were eliminated from the analyses.

Women with ER, PR or HER2 status that was unknown or could not be assessed were excluded from the analyses. The proportion of women with particular HR status was computed using different definitions of rural-urban status (mentioned earlier). Test for linear trend for ordered variables were performed by assigning the score j to the jth exposure level of a categorical variable (where j = 1, 2 . . .) and treating it as a continuous predictor in unconditional logistic regression. Test of heterogeneity to estimate differences in stratum specific odds ratio for rural-urban and pre and postmenopausal women was performed by comparing models with and without interaction term using likelihood ratio test. All analysis were performed using the statistical package Stata version 12.0 [274].

2.7 Results

A total of 1637 cases and 1515 controls were enrolled in the study. Distribution of cases and controls with respect to age and region of residence at enrolment, education and menopausal status are given in Table 2.3. The mean age at enrolment of cases and controls was 46.18 and 45.02 respectively. 49.79% of women were postmenopausal in cases whereas 43.76% women were postmenopausal in controls. 22.30% women cases were Graduate. Maximum cases and controls were from western region of the country with 48.56% and 51.55% respectively.

2.7.1 Rural and Urban Status

A statistically significant protection in risk of BC was observed in women who lived for first twenty years of life in rural area as compared to women who lived less than 20 years in rural area in their entire life. The protection wasn't significant when stratified on menopausal status. Other definitions used for describing the rural status of women were not significantly protective for the risk of BC. A dose-response relationship suggestive of protection was observed (Table 2.4). In the analysis of anthropometric measurements, reproductive factors and BC risk, women who lived first 20 years of life in rural area were designated as 'rural'.

Women who had lived less than 20 years in rural area in their entire life were categorized as 'urban'.

2.7.2 Hormone Receptor Status

Among 1273 cases on which the information on ER, PR and HER2 were available, it was observed that the ER+/PR+ cases were higher in urban population as compared to rural in all the definitions of rural and urban status. A higher proportion of ER-/PR- cases were observed in women who lived first twenty years in rural area (60.98%) as compared to those lived less than 20 years in rural area (54.29%). A statistically significant difference (P = 0.018) (Data not shown) in the prevalence of TNBC tumours was observed in women who have lived first twenty years of life in rural area (44.21%) as compared to women who have lived less than 20 years in rural area (34.39%) in their entire life (Table 2.5).

2.7.3 Reproductive Factors

Age at menarche was not significantly related to BC risk in all three instances, i.e. without stratification, and when stratified on rural-urban status and menopausal status. No association was observed among women who were ever pregnant as compared to women who were never pregnant. Rural women who had 4 or more live births showed a protective association with OR = 0.42 (95% CI: 0.24–0.75) as compared to women with 1 live birth. A protective association in premenopausal women (OR = 0.58; 95% CI: 0.39–0.86) and a suggestive protection in postmenopausal women (OR = 0.64; 95% CI: 0.41–1.00) was observed when women who had four or more live births were compared with women who had one live birth. Age at first full-term pregnancy proved to be an important risk factor in the development of BC. Women who had their first full-term pregnancy after age 25 had a significantly elevated risk of BC compared with women who had first full-term pregnancy below 20 years of age (OR = 1.83; 95% CI: 1.41–2.36). A 5% increase in risk was observed with every 2 year delay in age at first full-term pregnancy after age 25. The interval between age at menarche and age

at first full-term pregnancy was positively related to BC risk in women from urban areas and in premenopausal women.

An increasing trend in age at last full-term pregnancy was shown to increase the risk of BC across all categories. However, when adjusted for age at first full-term pregnancy the association was no longer significant (Data not shown). A duration of more than 10 years since last childbirth is protective in urban and premenopausal women as compared to a recent (≤ 10 years) childbirth in BC risk. No association was observed between women who ever breastfed compared to women who haven't breastfed even for one month. On further stratification in duration of breastfeeding, a similar result of no association was observed even in the highest category of breastfeeding (>36 months) when compared to lowest category of breastfeeding of ≤ 12months. A high maternal risk of BC has been observed in women with twin pregnancies as compared to singleton pregnancy in premenopausal women (OR = 3.45; 95% CI: 1.07–11.04) and rural women (OR = 6.28; 95% CI: 1.34–29.26)

Current users of OC were at increased risk of BC as compared to never users in urban women (OR = 2.26; 95% CI: 1.16–4.37). The current OC users using OC for more than 2 years showed increased risk for urban women compared to women without use of OC (OR = 2.46; 95% CI: 0.99–6.12). Two or more than 2 induced abortions were observed to be a risk factor of BC overall (OR = 1.65; 95% CI: 1.25–2.17), urban (OR = 1.58; 95% CI: 1.15–2.16) and rural women (OR = 2.08; 95% CI: 1.16–3.72) and in premenopausal women (OR = 2.04; 95% CI: 1.42–2.94). Even a single miscarriage showed a protection from BC in rural (OR = 0.62; 95% CI: 0.41–0.95) and premenopausal women (OR = 0.68; 95% CI: 0.48–0.96) (Tables 2.6, 2.7 and 2.8).

2.7.4 Anthropometric Measurements

Tables 2.9, 2.10, 2.11, 2.12 and 2.13 represent the risk of developing BC and anthropometric measurements including body size at different ages in women overall, stratified by residential status (Rural/urban) and menopausal status respectively.

Risk of BC increased in underweight women (BMI <18.5kg/m^2) when compared to women with normal BMI of Asian category (18.5–22.9 kg/m^2). The increased risk was observed overall and in stratified analysis for residential status (Rural/Urban) and menopausal status. The OR in urban women was 1.62 (95% CI: 1.03–2.52), similarly in rural women an OR of 1.80 (95% CI: 1.09–2.99) was observed. The risk of developing BC in premenopausal women was 1.75 (95% CI: 1.16–2.65) whereas in postmenopausal women was (OR = 1.89; 95% CI: 1.09–3.29). The risk persisted even when compared with normal BMI of world classification (18.5–24.9 kg/m^2) and when the analysis were limited to women who did not report any weight change in last one year of enrolment (Data not shown). A protective association observed in premenopausal women (OR = 0.93; 95% CI: 0.91–0.95) with per unit increase in BMI (world) continued in women who had attained menopause less than 10 years ago (OR = 0.95; 95% CI: 0.92–0.98). However the risk of BC increased in women in highest category of BMI (world) who had attained menopause ≥10 years ago from the date of enrolment (OR = 1.85; 95% CI: 1.05–3.28). In the present study, for every 5 cm increase in height the OR of 1.10 (95% CI: 1.02–1.19) was observed in the urban area and in premenopausal women (OR = 1.24; 95% CI: 1.12–1.37), but not in rural area (OR = 1.05; 95% CI: 0.93–1.19). A significant increase in risk was observed with linear trend in WC in stratified analysis of residential status and menopausal status. An inverse relationship has been observed between HC and BC risk across all strata. The inverse association persisted even after adjusting for total number of full term pregnancies (Data not shown). Adult body weight of more than 65 kg was found to be associated with premenopausal BC risk (OR = 0.56; 95% CI: 0.41–0.75)

and those living in urban area (OR = 0.70; 95% CI: 0.56–0.89). With every 0.1 unit increase in WHR the risk of BC increased (OR = 1.76; 95% CI: 1.55–2.01), (OR = 1.55; 95% CI: 1.29–1.86) (OR = 1.69; 95% CI: 1.47–1.96) and (OR = 1.71; 95% CI: 1.41–2.00) in urban, rural, premenopausal and postmenopausal women respectively. Body size at ages 10, 20 and at enrolment were analysed for their association with BC risk after adjusting for current WHR. Body size at age 10 was not found to be associated with BC; body size at age 20 increased the risk of premenopausal cancer when adjusted for BMI (Data not shown). The risk, however, disappears when adjusted for WHR. Body size at the time of enrolment was shown to decrease the risk of premenopausal (OR = 0.82; 95% CI: 0.69–0.97) and postmenopausal cancer (OR = 0.83; 95% CI: 0.69–1.0002) when adjusted for WHR. An increased risk was observed in postmenopausal women with an increase in body size from age 10 to age 20 (OR_{trend} = 1.25; P = 0.058). Any changes observed by women between age 20 and at enrolment were not associated with the risk of BC.

Table 2.3: Distribution of selected Characteristics among Cases and Controls

Parameters	Categories	Cases (n=1637)		Controls (n=1515)	
		Number	%	Number	%
Age at enrolment	20-29	53	3.24	68	4.49
	30-39	357	21.81	364	24.03
	40-49	597	36.47	534	35.25
	50-59	440	26.88	401	26.47
	60-69	190	11.61	147	9.70
	Missing	0	0.00	1	0.07
	Mean (±SD)	46.18 (±9.73)		45.02 (±10.07)	
Region of residence at enrolment	North	359	21.93	301	19.87
	West	795	48.56	781	51.55
	Central	98	5.99	85	5.61
	East	367	22.42	322	21.25
	South	18	1.10	26	1.72
	Missing	0	0.00	0	0.00
Education	No formal schooling	355	21.69	271	17.89
	Less than 5 years of schooling	101	6.17	99	6.53
	5-8 years of schooling	361	22.05	365	24.09
	High School	451	27.55	454	29.97
	College graduation and more	365	22.30	323	21.32
	Missing	4	0.24	3	0.20
Menopausal Status	Premenopausal	818	49.97	841	55.51
	Postmenopausal (Total)	815	49.79	663	43.76
	Natural menopause	623	76.44	499	76.30
	Menopause due to other reasons	177	21.70	155	23.37
	Missing	4	0.24	11	0.73

Table 2.4: Association of different time periods lived in rural area and risk of breast cancer stratified by menopausal status

Time spent (in years) in rural area	Total (Cases=1637, Controls=1515)					Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
	Case/ Control	OR[a] (95%CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95%CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95%CI)	p-value	OR[b] (95%CI)	p-value
Never	754/623	Reference		Reference		381/354	Reference		Reference		372/265	Reference		Reference	
Ever	883/892	0.81 (0.71-0.94)	0.006	0.93 (0.79-1.09)	0.377	440/487	0.82 (0.68-1.00)	0.060	1.09 (0.87-1.36)	0.443	440/398	0.78 (0.63-0.96)	0.022	0.78 (0.61-0.98)	0.040
1-10 years	61/54	0.93 (0.63-1.36)	0.721	1.00 (0.67-1.50)	0.985	29/32	0.83 (0.49-1.41)	0.507	0.98 0.55-1.72	0.945	32/22	1.03 0.58-1.82	0.904	1.04 (0.58-1.89)	0.876
>10 years	822/838	0.81 (0.70-0.93)	0.005	0.92 (0.78-1.08)	0.347	411/455	0.82 (0.68-1.01)	0.063	1.10 (0.87-1.38)	0.409	408/376	0.76 (0.62-0.95)	0.015	0.76 (0.60-0.97)	0.027
Risk per year lived in rural area		0.90 (0.83-0.96)	0.004	0.96 (0.88-1.04)	0.344		0.91 (0.82-1.00)	0.064	1.04 (0.93-1.17)	0.405		0.87 (0.78-0.97)	0.015	0.87 (0.77-0.98)	0.026
Lived First 20 years of life	442/543	0.67 0.57-0.79	<0.001	0.81 (0.67-0.99)	0.042	200/305	0.59 (0.47-0.75)	<0.001	0.87 (0.66-1.15)	0.351	241/235	0.72 (0.57-0.92)	0.010	0.76 (0.57-1.00)	0.057
Currently residing in rural	555/497	0.92 (0.79-1.09)	0.379	1.06 (0.88-1.28)	0.503	290/293	0.91 (0.73-1.13)	0.423	1.18 (0.91-1.52)	0.200	262/199	0.92 (0.72-1.18)	0.531	0.95 (0.71-1.26)	0.731
Lived < 20 years in life	1195/972	Reference		Reference		621/536	Reference		Reference		571/428	Reference		Reference	
Lived First 20 years of life	442/543	0.65 (0.56-0.76)	<0.001	0.77 (0.65-0.92)	0.004	200/305	0.55 (0.44-0.68)	<0.001	0.74 (0.58-0.95)	0.019	241/235	0.76 (0.61-0.95)	0.018	0.80 (0.62-1.04)	0.101
Currently residing in urban	1080/1018	Reference		Reference		530/548	Reference		Reference		549/464	Reference		Reference	
Currently residing in rural	557/497	1.07 (0.92-1.24)	0.365	1.18 (1.006-1.39)	0.041	291/293	1.03 (0.84-1.26)	0.761	1.24 (0.99-1.55)	0.059	263/199	1.11 (0.89-1.39)	0.331	1.14 (0.89-1.46)	0.268

Abbreviations: CI, Confidence Interval; OR, Odds ratio

[a] Adjusted for age and region of residence

[b] Adjusted for age, region of residence, education, height waist-to-hip ratio, age at first full-term pregnancy, menopausal status (where appropriate)

Missing values were excluded from analysis

Table 2.5: Prevalence of Hormone Receptors in Rural and Urban India

Time spent (in years) in rural area	ER+/PR+ (N=569)		ER-/PR- (N=725)		HER2+ (N=479)		TNBC (N=470)		Total
	N	%	N	%	N	%	N	%	N
Never	279	46.73	323	53.27	187	31.32	201	33.67	597
Ever	290	41.57	402	58.43	203	30.03	269	39.79	676
Lived <20 years	432	45.71	513	54.29	302	31.96	325	34.39	945
Lived first 20 years of life	128	39.02	200	60.98	88	26.83	145	44.21	328
Currently living in urban	386	45.84	456	54.16	255	30.29	292	34.68	842
Currently living in rural	174	40.37	257	59.63	135	31.32	178	41.30	431

Abbreviations: ER+, Estrogen Receptor Positive ; ER-, Estrogen Receptor Negative; HER2+, Human epidermal growth factor receptor 2; PR+, Progesterone Receptor Positive; PR-, Progesterone Receptor Negative; TNBC, Triple Negative Breast Cancer

Table 2.6: Association of Reproductive Factors and Risk of Breast Cancer

Parameters	Categories	Total (Cases=1637, Controls=1515)				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Age at menarche (in years) Total number of pregnancies[c]	≤12	251/233	Reference		Reference	
	13-14	858/811	0.97 (0.79-1.19)	0.783	0.93 (0.75-1.16)	0.551
	15-20	508/451	1.01 (0.81-1.27)	0.870	0.99 (0.78-1.26)	0.987
	Missing	20/20				
	Trend test		1.01 (0.91-1.13)	0.769	1.01 (0.90-1.13)	0.854
	Risk per year increase in age at menarche		0.99 (0.95-1.04)	0.996	0.99 (0.94-1.04)	0.868
	0 (Never)	68/59	Reference		Reference	
	Ever	1548/1441	0.88 (0.61-1.26)	0.500	0.89 (0.61-1.29)	0.564
	Missing	21/15				
No. of Full-term Pregnancies[c]	1	230/171	Reference		Reference	
	2	555/496	0.80 (0.64-1.02)	0.076	0.76 (0.59-0.97)	0.030
	3	358/394	0.63 (0.49-0.81)	<0.001	0.62 (0.48-0.81)	0.001
	≥4	388/361	0.71 (0.55-0.91)	0.009	0.66 (0.49-0.87)	0.004
	Missing	21/15				
	Trend test		0.89 (0.82-0.96)	0.004	0.87 (0.80-0.95)	0.003
	Risk per increase in number of full-term pregnancy		0.92 (0.88-0.98)	0.008	0.91 (0.86-0.97)	0.007
Age at first full-term pregnancy (in years)[d]	<20 yrs	422/494	Reference		Reference	
	20-21	316/331	1.11 (0.91-1.36)	0.289	1.16 (0.94-1.44)	0.154
	22-23	261/234	1.30 (1.04-1.62)	0.017	1.32 (1.04-1.68)	0.022
	24-25	187/151	1.45 (1.13-1.87)	0.003	1.31 (0.99-1.73)	0.056
	≥26	335/209	1.87 (1.50-2.32)	<0.001	1.83 (1.41-2.36)	<0.001
	Missing	116/96				
	Trend test		1.16 (1.10-1.22)	<0.001	1.14 (1.08-1.21)	<0.001
	Per 2 year increase in age at first full-term pregnancy		1.11 (1.07-1.15)	<0.001	1.10 (1.05-1.15)	<0.001

Parameters	Categories	Total (Cases=1637, Controls=1515)				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Age at first full-term pregnancy (in years)[d]	≤25	1186/1210	Reference		Reference	
	>25	335/209	1.63 (1.34-1.97)	<0.001	1.53 (1.23-1.89)	<0.001
	Missing	116/96				
	Risk Per year increase in age at first full-term pregnancy		1.05 (1.03-1.07)	<0.001	1.05 (1.02-1.07)	<0.001
Interval between menarche and first full-term pregnancy (in years)[d]	<10	984/1031	Reference		Reference	
	≥10	522/370	1.48 (1.26-1.74)	<0.001	1.33 (1.11-1.60)	0.002
	Missing	131/114				
Age at last full-term pregnancy (in years)[d]	≤24	336/400	Reference		Reference	
	25-29	600/573	1.22 (1.02-1.47)	0.029	1.25 (1.02-1.52)	0.025
	30-34	427/337	1.46 (1.19-1.80)	<0.001	1.45 (1.16-1.81)	0.001
	≥35	180/126	1.62 (1.22-2.13)	0.001	1.78 (1.31-2.41)	<0.001
	Missing	94/79				
	Trend test		1.18 (1.09-1.28)	<0.001	1.20 (1.10-1.31)	<0.001
Duration since last birth (in years)[d]	≤10	414/378				
	>10	1127/1057	0.75 (0.61-0.93)	0.010	0.70 (0.56-0.88)	0.002
	Missing	96/80				
History of Breastfeeding	Never	107/102	Reference		Reference	
	Ever	1503/1394	1.01 (1.004-1.018)	0.001	1.05 (0.57-1.94)	0.855
	Missing	27/19				
Duration of Breastfeeding (in months)[d]	≤12	165/144	Reference		Reference	
	13-24	195/162	1.05 (0.77-1.43)	0.738	1.06 (0.77-1.45)	0.719
	25-36	218/197	0.98 (0.72-1.31)	0.899	0.96 (0.70-1.31)	0.829
	> 36	943/912	0.88 (0.69-1.12)	0.319	0.96 (0.73-1.26)	0.770
	Missing	115/100				
	Trend test		0.94 (0.88-1.01)	0.111	0.97 (0.89-1.06)	0.574
	Risk per month increase in breastfeeding		0.999 (0.997-1.00)	0.334	1.00 (0.99-1.003)	0.525
Average duration breastfed per child (in months)	≤6	112/93	Reference		Reference	
	7-12	287/247	0.96 (0.69-1.33)	0.825	0.89 (0.63-1.26)	0.538
	13-18	286/282	0.84 (0.61-1.16)	0.304	0.85 (0.60-1.20)	0.378
	19-24	357/375	0.78 (0.57-1.07)	0.132	0.81 (0.58-1.41)	0.239
	>24	455/394	0.96 (0.70-1.31)	0.818	1.00 (0.72-1.38)	0.989
	Missing	140/124				
	Trend test		0.98 (0.93-1.04)	0.708	1.01 (0.95-1.07)	0.646
	Risk per month increase in breastfeeding		1.003 (0.99-1.009)	0.265	1.00 (0.99-1.01)	0.160
Twin pregnancy	Never	1496/1407	Reference		Reference	
	Ever	36/16	2.06 (1.13-3.73)	0.017	1.81 (0.96-3.40)	0.065
	Missing	NA				

Parameters	Categories	Total (Cases=1637, Controls=1515)				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
OC use	Never	1430/1367	Reference		Reference	
	Ever	155/122	1.29 (1.008-1.67)	0.042	1.17 (0.90-1.54)	0.232
	Past users[e]	101/91	1.10 (0.82-1.48)	0.517	1.01 (0.74-1.38)	0.910
	Current users[f]	40/24	1.86 (1.11-3.14)	0.019	1.68 (0.97-2.90)	0.059
	Short term current users[g]	17/10	1.89 (0.85-4.18)	0.113	1.45 (0.64-3.30)	0.364
	Long term current users[h]	23/14	1.79 (0.91-3.53)	0.088	1.75 (0.86-3.54)	0.118
	Missing	52/26				
Age OC use started (in years)	Never	1430/1367	Reference		Reference	
	≤24	66/54	1.28 (0.88-1.86)	0.184	1.27 (0.86-1.89)	0.226
	25-29	45/41	1.10 (0.71-1.69)	0.660	0.95 (0.60-1.49)	0.832
	≥30	32/20	1.59 (0.90-2.80)	0.109	1.33 (0.73-2.44)	0.345
	Missing	64/33				
	Trend test		1.13 (0.98-1.29)	0.070	1.06 (0.92-1.23)	0.368
Total duration of OC use (in years)	Never	1430/1367	Reference		Reference	
	<1	77/61	1.29 (0.91-1.83)	0.142	1.14 (0.79-1.65)	0.467
	1-4	45/35	1.29 (0.82-2.03)	0.259	1.18 (0.74-1.89)	0.468
	≥5	18/19	0.98 (0.51-1.88)	0.959	1.04 (0.53-2.05)	0.895
	Missing	67/33				
	Trend test		1.09 (0.94-1.26)	0.223	1.06 (0.91-1.24)	0.430
	Risk per month increase in duration of OC use		0.999 (0.994-1.005)	0.874	0.999 (0.993-1.005)	0.824
	Risk per year increase in duration of OC use		0.99 (0.93-1.06)	0.874	0.99 (0.92-1.06)	0.824
No. of Induced Abortions	0	1187/1172	Reference		Reference	
	1	275/240	1.15 (0.94-1.39)	0.152	1.10 (0.89-1.35)	0.361
	≥2	175/103	1.70 (1.31-2.20)	<0.001	1.65 (1.25-2.17)	<0.001
	Missing	NA				
	Trend		1.25 (1.12-1.40)	<0.001	1.23 (1.09-1.39)	0.001
	Risk per increase in abortion		1.24 (1.12-1.37)	<0.001	1.22 (1.10-1.36)	<0.001
No. of Spontaneous Abortions	0	1431/1220	Reference		Reference	
	1	157/195	0.68 (0.54-0.85)	0.001	0.73 (0.57-0.92)	0.010
	≥2	49/100	0.42 (0.29-0.60)	<0.001	0.44 (0.30-0.64)	<0.001
	Missing	NA				
	Trend		0.66 (0.57-0.76)	<0.001	0.68 (0.59-0.80)	<0.001
	Risk per increase in miscarriage		0.72 (0.63-0.81)	<0.001	0.73 (0.64-0.84)	<0.001

Abbreviations: CI, Confidence Interval; OR, Odds ratio; NA, Not Applicable; OC, Oral Contraceptive.

[a] Adjusted for age & region of residence.

[b] Adjusted on age, region of residence, education, rural-urban status, menopausal status, induced & spontaneous abortion, age at first full-term pregnancy, height, waist-to-hip ratio.

[c] Not adjusted for age at first full-term pregnancy.

[d] Adjusted for total number of pregnancies instead of age at first full-term pregnancy.

[e] Past Users: Women who had stopped OC use more than 5 years ago from the date of enrolment.

[f] Current Users: Women who had used OC in last 5 years from the date of enrolment.

[g] Short-term Current Users: Current OC users with no more than 2 years of usage.

[h] Long-term Current Users: Current OC users with more than 2 years of usage.

Missing values were excluded from analysis

Table 2.7: Association of Reproductive Factors and risk of Breast Cancer stratified by Rural-Urban Status

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Age at menarche (in years)	≤12	190/148	Reference		Reference		61/85	Reference		Reference	
	13-14	637/525	0.93 (0.73-1.19)	0.600	0.88 (0.67-1.15)	0.381	221/286	1.06 (0.73-1.54)	0.750	1.06 (0.72-1.58)	0.738
	15-20	353/287	0.92 (0.70-1.21)	0.580	0.91 (0.68-1.22)	0.561	155/164	1.29 (0.86-1.92)	0.212	1.20 (0.78-1.84)	0.395
	Missing	15/12					5/8				
	Trend test		0.96 (0.84-1.10)	0.619	0.97 (0.84-1.11)	0.675		1.15 (0.95-1.39)	0.149	1.10 (0.89-1.35)	0.354
	Risk per year increase in age at menarche		0.97 (0.92-1.03)	0.422	0.97 (0.91-1.03)	0.379		1.06 (0.97-1.15)	0.157	1.03 (0.94-1.13)	0.415
	P heterogeneity		0.583								
Total number of pregnancies[c]	0 (Never)	62/49	Reference		Reference		6/10	Reference		Reference	
	Ever	1113/908	0.92 (0.62-1.35)	0.683	0.82 (0.55-1.23)	0.350	435/533	1.39 (0.50-3.89)	0.522	1.46 (0.52-4.14)	0.468
	Missing	20/15					1/0				
	P heterogeneity		0.396								
No. of Full-term Pregnancies[c]	1	190/141	Reference		Reference		40/30	Reference		Reference	
	2	443/369	0.87 (0.67-1.13)	0.318	0.80 (0.61-1.05)	0.117	112/127	0.62 (0.36-1.07)	0.088	0.57 (0.32-1.00)	0.051
	3	232/209	0.79 (0.59-1.06)	0.121	0.69 (0.51-0.94)	0.019	126/185	0.44 (0.26-0.76)	0.003	0.41 (0.23-0.72)	0.002
	≥4	232/176	0.91 (0.67-1.24)	0.568	0.72 (0.52-1.02)	0.066	156/185	0.50 (0.29-0.86)	0.013	0.42 (0.24-0.75)	0.004
	Missing	20/15					1/0				
	Trend test		0.96 (0.88-1.06)	0.504	0.89 (0.80-0.99)	0.042		0.84 (0.72-0.97)	0.021	0.80 (0.68-0.93)	0.006
	Risk per increase in number of full-term pregnancy		1.00 (0.93-1.07)	0.906	0.95 (0.88-1.03)	0.267		0.87 (0.79-0.95)	0.004	0.84 (0.76-0.93)	0.001
	P heterogeneity		0.313								

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Age at first full-term pregnancy (in years)[d]	<20	221/195	Reference		Reference		201/299	Reference		Reference	
	20-21	186/190	0.86 (0.65-1.13)	0.297	1.03 (0.76-1.38)	0.841	130/141	1.37 (1.01-1.85)	0.037	1.34 (0.98-1.84)	0.061
	22-23	213/193	0.97 (0.74-1.28)	0.860	1.19 (0.88-1.60)	0.239	48/41	1.70 (1.08-2.69)	0.021	1.71 (1.07-2.76)	0.025
	24-25	162/122	1.17 (0.87-1.59)	0.287	1.32 (0.94-1.84)	0.101	25/29	1.26 (0.71-2.22)	0.418	1.10 (0.61-1.99)	0.737
	≥26	307/193	1.41 (1.08-1.84)	0.010	1.78 (1.32-2.41)	<0.001	28/16	2.59 (1.36-4.92)	0.004	2.24 (1.13-4.43)	0.020
	Missing	106/79					10/17				
	Trend test		1.10 (1.04-1.17)	0.001	1.15 (1.07-1.24)	<0.001		1.21 (1.08-1.36)	0.001	1.17 (1.04-1.33)	0.010
	Per 2 year increase in age at first full-term pregnancy		1.07 (1.03-1.11)	0.001	1.10 (1.05-1.16)	<0.001		1.16 (1.06-1.26)	0.001	1.14 (1.04-1.25)	0.005
	P heterogeneity		0.216								
Age at first full-term pregnancy (in years)[d]	≤25	782/700	Reference		Reference		404/510	Reference		Reference	
	>25	307/193	1.43 (1.16-1.76)	0.001	1.57 (1.25-1.97)	<0.001	28/16	2.21 (1.17-4.15)	0.014	1.88 (0.97-3.66)	0.061
	Missing	106/79					10/17				
	Risk Per year increase in age at first full-term pregnancy		1.03 (1.01-1.05)	0.001	1.05 (1.02-1.07)	<0.001		1.07 (1.03-1.12)	0.001	1.06 (1.02-1.11)	0.005
	P heterogeneity		0.211								
Interval between menarche and first full-term pregnancy (in years)[d]	<10	612/556	Reference		Reference		372/475	Reference		Reference	
	≥10	467/327	1.31 (1.09-1.57)	0.004	1.36 (1.11-1.68)	0.003	55/43	1.60 (1.05-2.45)	0.028	1.43 (0.91-2.24)	0.114
	Missing	116/89					15/25				
	P heterogeneity		0.307								
Age at last full-term pregnancy (in years)[d]	≤24	209/200	Reference		Reference		127/200	Reference		Reference	
	25-29	413/362	1.08 (0.85-1.38)	0.494	1.13 (0.87-1.46)	0.334	187/211	1.33 (0.99-1.81)	0.057	1.56 (1.12-2.16)	0.007
	30-34	332/254	1.23 (0.95-1.60)	0.100	1.29 (0.98-1.69)	0.065	95/83	1.68 (1.16-2.44)	0.006	2.13 (1.40-3.23)	0.000
	≥35	155/87	1.67 (1.20-2.33)	0.002	1.87 (1.30-2.68)	0.001	25/39	0.85 (0.48-1.50)	0.588	1.42 (0.76-2.68)	0.267
	Missing	86/69					8/10				
	Trend test		1.16 (1.05-1.28)	0.002	1.20 (1.08-1.33)	0.001		1.11 (0.95-1.28)	0.166	1.28 (1.07-1.52)	0.005
	P heterogeneity		0.068								

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Duration since last birth (in years)[d]	≤10	320/249	Reference		Reference		94/129	Reference		Reference	
	>10	787/653	0.78 (0.60-1.02)	0.072	0.68 (0.52-0.89)	0.006	340/404	0.83 (0.56-1.21)	0.341	0.68 (0.45-1.03)	0.070
	Missing	88/70					8/10				
	P heterogeneity						0.301				
History of Breastfeeding	Never	96/79	Reference		Reference		11/23	Reference		Reference	
	Ever	1073/876	1.00 (0.99-1.01)	0.071	1.02 (0.51-2.03)	0.944	430/518	1.02 (1.00-1.03)	0.003	0.98 (0.25-3.81)	0.981
	Missing	26/17					½				
	P heterogeneity						0.934				
Duration of Breastfeeding (in months)[d]	≤12	143/121	Reference		Reference		22/23	Reference		Reference	
	13-24	156/120	1.09 (0.77-1.53)	0.599	1.07 (0.75-1.52)	0.692	39/42	0.96 (0.46-2.01)	0.931	0.93 (0.43-1.99)	0.864
	25-36	168/140	1.01 (0.72-1.41)	0.926	0.95 (0.67-1.35)	0.794	50/57	0.93 (0.46-1.87)	0.842	0.86 (0.41-1.80)	0.697
	> 36	620/511	1.00 (0.76-1.32)	0.95	0.92 (0.68-1.25)	0.628	323/401	0.78 (0.42-1.44)	0.438	0.85 (0.44-1.65)	0.648
	Missing	107/80					8/20				
	Trend test		0.99 (0.91-1.07)	0.826	0.96 (0.87-1.05)	0.440		0.90 (0.77-1.06)	0.220	0.95 (0.80-1.13)	0.616
	Risk per month increase in breastfeeding		1.001 (0.99-1.003)	0.309	1.000 (0.99-1.003)	0.501		0.997 (0.994-1.00)	0.128	1.000 (0.99-1.003)	0.999
	P heterogeneity						0.846				
Average duration breastfed per child (in months)	≤6	95/77	Reference		Reference		17/16	Reference		Reference	
	7- 12	227/171	1.05 (0.73-1.52)	0.754	0.95 (0.65-1.39)	0.817	60/76	0.78 (0.36-1.69)	0.541	0.66 (0.30-1.48)	0.326
	13-18	197/172	0.91 (0.63-1.32)	0.649	0.87 (0.59-1.28)	0.486	89/110	0.78 (0.37-1.64)	0.514	0.71 (0.32-1.56)	0.402
	19- 24	228/223	0.81 (0.57-1.16)	0.262	0.77 (0.53-1.12)	0.185	129/152	0.81 (0.39-1.68)	0.581	0.75 (0.35-1.60)	0.461
	>24	320/232	1.11 (0.79-1.57)	0.525	1.06 (0.74-1.53)	0.730	135/162	0.81 (0.39-1.67)	0.572	0.74 (0.34-1.57)	0.436
	Missing	128/97					12/27				
	Trend test		1.00 (0.94-1.07)	0.840	1.01 (0.94-1.08)	0.777		0.99 (0.89-1.11)	0.935	1.00 (0.89-1.12)	0.978
	Risk per month increase in breastfeeding		1.00 (0.99-1.01)	0.123	1.00 (0.99-1.01)	0.145		1.00 (0.98-1.01)	0.877	0.99 (0.98-1.01)	0.961
	P heterogeneity						0.547				
Twin pregnancy	Never	1073/882	Reference		Reference		423/525	Reference		Reference	
	Ever	25/13	1.55 (0.79-3.06)	0.199	1.21 (0.59-2.46)	0.590	11/3	4.27 (1.17-15.53)	0.027	6.28 (1.34-29.26)	0.019
	Missing	NA					NA				
	P heterogeneity						0.030				

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
OC use	Never	1031/872	Reference		Reference		399/495	Reference		Reference	
	Ever	125/81	1.38 (1.02-1.85)	0.034	1.28 (0.94-1.76)	0.115	30/41	1.01 (0.61-1.67)	0.950	0.89 (0.52-1.52)	0.685
	Past users[e]	82/60	1.19 (0.84-1.69)	0.312	1.13 (0.78-1.63)	0.498	19/31	0.82 (0.45-1.49)	0.519	0.73 (0.39-1.36)	0.329
	Current users[f]	34/14	2.39 (1.26-4.52)	0.008	2.26 (1.16-4.37)	0.016	6/10	0.87 (0.30-2.47)	0.800	0.68 (0.22-2.06)	0.502
	Short-term current users[g]	16/7	2.24 (0.91-5.53)	0.079	1.92 (0.75-4.90)	0.169	Not estimated because of small number (n ≤5 among controls)				
	Long-term current users[h]	18/7	2.48 (1.02-6.03)	0.043	2.46 (0.99-6.12)	0.052					
	Missing	39/19					13/7				
	P heterogeneity		0.252								
Age OC use started (in years)	Never	1031/872	Reference		Reference		399/495	Reference		Reference	
	≤24	49/32	1.40 (0.88-2.21)	0.151	1.42 (0.87-2.30)	0.155	17/22	1.09 (0.56-2.12)	0.787	0.97 (0.48-1.96)	0.940
	25-29	40/29	1.21 (0.74-1.97)	0.438	1.09 (0.66-1.81)	0.722	5/12	0.56 (0.19-1.63)	0.295	0.50 (0.16-1.51)	0.233
	≥30	29/14	1.86 (0.97-3.57)	0.059	1.67 (0.84-3.33)	0.139	3/6	0.65 (0.16-2.67)	0.559	0.48 (0.11-2.06)	0.333
	Missing	46/25					18/8				
	Trend test		1.19 (1.02-1.39)	0.026	1.15 (0.97-1.35)	0.097		0.86 (0.63-1.18)	0.368	0.79 (0.56-1.10)	0.169
	P heterogeneity		0.251								
Total duration of OC use (in years)	Never	1031/872	Reference		Reference		399/495	Reference		Reference	
	<1	65/36	1.62 (1.06-2.47)	0.023	1.42 (0.91-2.22)	0.120	12/25	0.67 (0.32-1.36)	0.270	0.62 (0.29-1.32)	0.219
	1-4	37/27	1.22 (0.73-2.03)	0.432	1.22 (0.72-2.06)	0.445	8/8	1.27 (0.47-3.43)	0.635	1.01 (0.35-2.89)	0.978
	≥5	1412	1.04 (0.48-2.28)	0.906	1.21 (0.54-2.70)	0.638	4/7	0.87 (0.24-3.05)	0.830	0.77 (0.21-2.81)	0.703
	Missing	48/25					19/8				
	Trend test		1.13 (0.95-1.34)	0.156	1.12 (0.94-1.34)	0.188		0.95 (0.70-1.28)	0.758	0.89 (0.64-1.22)	0.479
	Risk per month increase in duration of OC use		1.000 (0.99-1.007)	0.826	1.001 (0.99-1.007)	0.687		0.99 (0.98-1.00)	0.343	0.99 (0.97-1.00)	0.259
	Risk per year increase in duration of OC use		1.00 (0.93-1.08)	0.826	1.01 (0.93-1.10)	0.687		0.92 (0.78-1.08)	0.343	0.90 (0.75-1.07)	0.259
	P heterogeneity		0.290								

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
No. of Induced Abortions	0	836/719	Reference		Reference		351/453	Reference		Reference	
	1	219/175	1.08 (0.86-1.35)	0.472	1.10 (0.87-1.40)	0.407	56/65	1.15 (0.78-1.69)	0.475	1.10 (0.73-1.66)	0.629
	≥2	140/78	1.55 (1.15-2.08)	0.004	1.58 (1.15-2.16)	0.004	35/25	1.94 (1.13-3.31)	0.015	2.08 (1.16-3.72)	0.013
	Missing	NA					NA				
	Trend		1.20 (1.05-1.36)	0.006	1.21 (1.05-1.40)	0.006		1.31 (1.04-1.65)	0.019	1.32 (1.03-1.70)	0.025
	Risk per increase in abortion		1.20 (1.07-1.34)	0.002	1.21 (1.07-1.37)	0.002		1.26 (1.03-1.54)	0.019	1.29 (1.04-1.60)	0.016
	P heterogeneity		0.651								
No. of Spontaneous Abortions	0	1046/795	Reference		Reference		385/425	Reference		Reference	
	1	115/118	0.73 (0.56-0.96)	0.029	0.79 (0.58-1.06)	0.120	42/77	0.60 (0.40-0.90)	0.015	0.62 (0.41-0.95)	0.029
	≥2	34/59	0.44 (0.28-0.67)	<0.001	0.47 (0.29-0.75)	0.002	15/41	0.41 (0.22-0.76)	0.005	0.42 (0.22-0.80)	0.002
	Missing	NA					NA				
	Trend		0.69 (0.58-0.82)	<0.001	0.72 (0.59-0.87)	0.001		0.63 (0.49-0.81)	<0.001	0.64 (0.49-0.83)	0.001
	Risk per increase in miscarriage		0.74 (0.64-0.86)	<0.001	0.76 (0.64-0.89)	0.001		0.68 (0.55-0.85)	0.001	0.70 (0.56-0.89)	0.004
	P heterogeneity		0.681								

Abbreviations: CI, Confidence Interval; NA, Not Applicable; OR, Odds ratio; OC, Oral Contraceptive.

[a] Adjusted for age and region of residence.

[b] Adjusted on age, region of residence, education, menopausal status, induced and spontaneous abortion, age at first full-term pregnancy, height, waist-to-hip ratio.

[c] Not adjusted for age at first full-term pregnancy.

[d] Adjusted for total number of pregnancies instead of age at first full-term pregnancy.

[e] Past Users: Women who had stopped OC use more than 5 years ago from the date of enrolment.

[f] Current Users: Women who had used OC in last 5 years from the date of enrolment.

[g] Short-term Current Users: Current OC users with no more than 2 years of usage.

[h] Long-term Current Users: Current OC users with more than 2 years of usage.

Missing values were excluded from analysis.

Table 2.8: Association of Reproductive Factors and risk of Breast Cancer stratified by Menopausal Status

Parameters	Categories	Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Age at menarche (in years)	≤12	132/141	Reference		Reference		118/92	Reference		Reference	
	13-14	431/455	0.99 (0.75-1.30)	0.962	0.99 (0.73-1.34)	0.964	426/350	0.94 (0.69-1.28)	0.735	0.87 (0.62-1.20)	0.414
	15-20	250/233	1.09 (0.81-1.48)	0.548	1.11 (0.80-1.56)	0.509	257/214	0.93 (0.67-1.30)	0.707	0.88 (0.62-1.25)	0.492
	Missing	8/12					11/7				
	Trend test		1.05 (0.91-1.22)	0.472	1.06 (0.90-1.25)	0.420		0.97 (0.83-1.13)	0.739	0.95 (0.80-1.13)	0.589
	Risk per year increase in age at menarche		1.03 (0.97-1.10)	0.276	1.04 (0.97-1.11)	0.248		0.96 (0.90-1.02)	0.269	0.95 (0.88-1.02)	0.185
	P heterogeneity		0.319								
Total number of pregnancies[c]	0 (Never)	35/42	Reference		Reference		33/17	Reference		Reference	
	Ever	776/789	1.08 (0.67-1.72)	0.738	1.13 (0.69-1.84)	0.622	769/641	0.62 (0.34-1.12)	0.116	0.63 (0.34-1.17)	0.146
	Missing	10/10					10/5				
	P heterogeneity		0.103								
No. of Full-term Pregnancies[c]	1	145/122	Reference		Reference		85/48	Reference		Reference	
	2	340/299	0.91 (0.68-1.22)	0.589	0.90 (0.66-1.21)	0.502	215/196	0.62 (0.41-0.93)	0.021	0.58 (0.38-0.89)	0.013
	3	173/215	0.61 (0.44-0.84)	0.005	0.64 (0.45-0.91)	0.013	184/174	0.60 (0.39-0.90)	0.015	0.58 (0.37-0.90)	0.015
	≥4	108/144	0.54 (0.38-0.78)	0.004	0.58 (0.39-0.86)	0.008	278/213	0.75 (0.50-1.12)	0.171	0.64 (0.41-1.00)	0.052
	Missing	10/10					10/5				
	Trend test		0.79 (0.70-0.88)	<0.001	0.83 (0.74-0.94)	0.004		0.99 (0.89-1.10)	0.880	0.94 (0.83-1.06)	0.329
	Risk per increase in number of full-term pregnancy		0.83 (0.76-0.91)	<0.001	0.85 (0.77-0.94)	0.002		0.98 (0.92-1.06)	0.761	0.95 (0.88-1.03)	0.265
	P heterogeneity		0.019								
Age at first full-term pregnancy (in years)[d]	<20	176/277	Reference		Reference		244/212	Reference		Reference	
	20-21	156/189	1.27 (0.95-1.69)	0.097	1.26 (0.93-1.70)	0.126	160/141	0.98 (0.73-1.32)	0.930	1.06 (0.78-1.45)	0.668
	22-23	134/126	1.67 (1.23-2.28)	0.001	1.46 (1.04-2.05)	0.027	127/104	1.06 (0.77-1.46)	0.693	1.16 (0.82-1.65)	0.384
	24-25	110/86	2.00 (1.42-2.81)	<0.001	1.57 (1.07-2.30)	0.019	76/65	1.01 (0.69-1.48)	0.927	1.01 (0.67-1.53)	0.934
	≥26	184/100	2.91 (2.13-3.97)	<0.001	2.45 (1.71-3.53)	<0.001	151/108	1.21 (0.89-1.65)	0.213	1.33 (0.92-1.92)	0.122
	Missing	61/63					54/33				
	Trend test		1.29 (1.20-1.39)	<0.001	1.22 (1.12-1.33)	<0.001		1.04 (0.97-1.12)	0.236	1.06 (0.97-1.15)	0.178
	Risk per 2 year increase in age at first full-term pregnancy		1.21 (1.14-1.27)	<0.001	1.16 (1.09-1.24)	<0.001		1.02 (0.97-1.08)	0.252	1.04 (0.98-1.10)	0.186
	P heterogeneity		0.006								

Parameters	Categories	Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Age at first full-term pregnancy (in years)[d]	≤25	576/678	Reference		Reference		607/522	Reference		Reference	
	>25	184/100	2.18 (1.66-2.86)	<0.001	1.86 (1.38-2.50)	<0.001	151/108	1.20 (0.91-1.58)	0.188	1.25 (0.91-1.70)	0.156
	Missing	61/63					54/33				
	Risk Per year increase in age at first full-term pregnancy		1.10 (1.07-1.12)	<0.001	1.08 (1.04-1.11)	<0.001		1.01 (0.98-1.03)	0.252	1.02 (0.99-1.05)	0.186
	P heterogeneity						0.004				
Interval between menarche and first full-term pregnancy (in years)[d]	<10	460/575	Reference		Reference		522/448	Reference		Reference	
	≥10	293/192	1.93 (1.54-2.42)	<0.001	1.50 (1.16-1.94)	0.002	228/176	1.11 (0.87-1.40)	0.380	1.16 (0.89-1.52)	0.263
	Missing	68/74					62/39				
	P heterogeneity						0.012				
Age at last full-term pregnancy (in years)[d]	≤24	176/260	Reference		Reference		158/137	Reference		Reference	
	25-29	305/318	1.41 (1.10-1.80)	0.013	1.34 (1.03-1.75)	0.029	295/249	1.03 (0.77-1.37)	0.822	1.09 (0.80-1.47)	0.576
	30-34	207/170	1.78 (1.34-2.37)	<0.001	1.65 (1.21-2.25)	0.001	219/167	1.14 (0.84-1.55)	0.381	1.19 (0.85-1.66)	0.305
	≥35	84/38	3.21 (2.07-4.97)	<0.001	3.16 (1.94-5.13)	<0.001	96/86	0.98 (0.67-1.43)	0.927	1.17 (0.77-1.76)	0.446
	Missing	49/55					44/24				
	Trend test		1.40 (1.24-1.57)	<0.001	1.36 (1.20-1.55)	<0.001		1.01 (0.91-1.14)	0.731	1.06 (0.94-1.20)	0.325
	P heterogeneity						0.005				
Duration since last birth (in years)[d]	≤10	383/349	Reference		Reference		31/29	Reference		Reference	
	>10	389/436	0.59 (0.46-0.77)	<0.001	0.58 (0.44-0.77)	<0.001	735/610	1.17 (0.67-2.03)	0.565	1.03 (0.58-1.84)	0.903
	Missing	49/56					46/24				
	P heterogeneity						0.320				
History of Breastfeeding	Never	58/61	Reference		Reference		49/41	Reference		Reference	
	Ever	751/768	0.96 (0.66-1.41)	0.874	0.74 (0.30-1.83)	0.526	749/615	1.02 (0.66-1.57)	0.918	1.36 (0.58-3.22)	0.472
	Missing	12/12					14/7				
	P heterogeneity						0.182				
Duration of Breastfeeding (in months)[d]	≤12	93/86	Reference		Reference		72/58	Reference		Reference	
	13-24	114/91	1.16 (0.77-1.74)	0.456	1.23 (0.80-1.87)	0.337	81/70	0.93 (0.58-1.49)	0.771	0.87 (0.53-1.42)	0.585
	25-36	130/132	0.92 (0.62-1.34)	0.672	0.95 (0.63-1.42)	0.829	88/64	1.10 (0.69-1.78)	0.666	1.02 (0.61-1.68)	0.935
	>36	425/467	0.82 (0.59-1.13)	0.234	1.02 (0.71-1.47)	0.891	515/436	0.95 (0.65-1.38)	0.805	0.89 (0.59-1.36)	0.615
	Missing	58/65					56/35				
	Trend test		0.90 (0.82-0.99)	0.043	0.98 (0.87-1.09)	0.751		0.90 (0.82-0.99)	0.043	0.97 (0.86-1.10)	0.715
	Risk per month increase in breastfeeding		0.996 (0.994-0.999)	0.027	1.000 (0.99-1.004)	0.730		0.996 (0.994-0.999)	0.027	1.000 (0.99-1.003)	0.776
	P heterogeneity						0.341				

Parameters	Categories	Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Average duration breastfed per child (in months)	≤6	54/52	Reference		Reference		58/40	Reference		Reference	
	7-12	139/140	0.93 (0.59-1.46)	0.780	0.98 (0.60-1.58)	0.948	148/107	0.95 (0.59-1.53)	0.845	0.79 (0.48-1.31)	0.372
	13-18	147/160	0.86 (0.55-1.34)	0.522	0.99 (0.61-1.59)	0.973	138/120	0.79 (0.49-1.27)	0.335	0.68 (0.41-1.13)	0.145
	19-24	168/196	0.80 (0.52-1.24)	0.331	0.96 (0.60-1.53)	0.879	188/176	0.73 (0.46-1.15)	0.186	0.65 (0.40-1.06)	0.088
	>24	240/218	1.04 (0.68-1.59)	0.841	1.21 (0.77-1.90)	0.402	214/171	0.86 (0.54-1.35)	0.515	0.77 (0.47-1.24)	0.289
	Missing	7375					66/49				
	Trend test		1.01 (0.94-1.10)	0.671	1.05 (0.97-1.14)	0.209		1.01 (0.94-1.10)	0.671	0.96 (0.87-1.05)	0.392
	Risk per month increase in breastfeeding		1.00 (0.99-1.01)	0.349	1.00 (0.99-1.01)	0.172		1.00 (0.99-1.01)	0.349	1.00 (0.99-1.01)	0.621
	P heterogeneity		0.976								
Twin pregnancy	Never	753/776	Reference		Reference		740/620	Reference		Reference	
	Ever	14/4	3.72 (1.21-11.38)	0.021	3.45 (1.07-11.04)	0.037	22/12	1.54 (0.75-3.14)	0.233	1.34 (0.62-2.87)	0.447
	Missing	NA					NA				
	P heterogeneity		0.215								
OC use	Never	669/723	Reference		Reference		758/633	Reference		Reference	
	Ever	110/95	1.30 (0.97-1.76)	0.075	1.14 (0.82-1.57)	0.422	45/27	1.37 (0.84-2.25)	0.198	1.36 (0.80-2.28)	0.246
	Missing	39/23					12/3				
	P heterogeneity		0.730								
Age OC use started (in years)	Never	669/723	Reference		Reference		758/633	Reference		Reference	
	≤24	47/45	1.23 (0.80-1.88)	0.341	1.26 (0.80-2.00)	0.314	19/9	1.73(0.78-3.87)	0.176	1.78 (0.76-4.19)	0.182
	25-29	31/32	1.06 (0.64-1.76)	0.805	0.83 (0.48-1.42)	0.500	14/9	1.30 (0.56-3.03)	0.539	1.37 (0.57-3.28)	0.479
	≥30	23/13	1.92 (0.96-3.84)	0.064	1.45 (0.69-3.04)	0.316	9/7	1.04 (0.38-2.84)	0.929	0.88 (0.30-2.60)	0.825
	Missing	48/28					15/5				
	Trend test		1.15 (0.98-1.36)	0.081	1.05 (0.88-1.25)	0.546		1.11 (0.86-1.43)	0.400	1.09 (0.83-1.43)	0.502
	P heterogeneity		0.404								

Parameters	Categories	Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Duration of OC use (in years)	Never	669/723	Reference		Reference		758/633	Reference		Reference	
	<1	53/47	1.29 (0.85-1.94)	0.221	1.09 (0.70-1.69)	0.690	24/14	1.41 (0.72-2.76)	0.310	1.35 (0.66-2.79)	0.402
	1-4	31/26	1.33 (0.78-2.26)	0.293	1.13 (0.64-1.98)	0.656	14/9	1.28 (0.55-2.99)	0.558	1.34 (0.56-3.23)	0.505
	≥5	14/17	0.93 (0.45-1.91)	0.847	1.06 (0.50-2.25)	0.870	4/2	1.66 (0.30-9.11)	0.558	1.64 (0.28-9.42)	0.577
	Missing	51/28					15/5				
	Trend test		1.08 (0.91-1.28)	0.357	1.04 (0.87-1.25)	0.615		1.20 (0.88-1.63)	0.242	1.20 (0.87-1.65)	0.263
	Risk per month increase in duration of OC use		0.998 (0.992-1.004)	0.681	0.998 (0.992-1.004)	0.661		1.00 (0.99-1.02)	0.361	1.00 (0.99-1.02)	0.323
	Risk per year increase in duration of OC use		0.98 (0.91-1.05)	0.681	0.98 (0.91-1.06)	0.661		1.09 (0.90-1.31)	0.361	1.10 (0.90-1.35)	0.323
	P heterogeneity		0.959								
No. of Induced Abortions	0	555/631	Reference		Reference		629/530	Reference		Reference	
	1	157/152	1.17 (0.91-1.50)	0.217	1.00 (0.76-1.32)	0.948	117/88	1.12 (0.83-1.51)	0.455	1.21 (0.88-1.67)	0.237
	≥2	109/58	2.12 (1.51-2.98)	<0.001	2.04 (1.42-2.94)	<0.001	66/45	1.23 (0.83-1.83)	0.299	1.27 (0.83-1.95)	0.267
	Missing	NA					NA				
	Trend		1.37 (1.18-1.59)	<0.001	1.30 (1.11-1.53)	0.001		1.11 (0.93-1.32)	0.224	1.15 (0.95-1.39)	0.146
	Risk per increase in abortion		1.32 (1.16-1.51)	<0.001	1.28 (1.11-1.47)	0.001		1.13 (0.97-1.31)	0.109	1.16 (0.98-1.37)	0.070
	P heterogeneity		0.104								
No. of Spontaneous Abortions	0	715/680	Reference		Reference		712/532	Reference		Reference	
	1	72/106	0.63 (0.46-0.87)	0.006	0.68 (0.48-0.96)	0.032	85/87	0.72 (0.52-1.00)	0.050	0.77 (0.54-1.08)	0.135
	≥2	34/55	0.57 (0.37-0.89)	0.015	0.61 (0.38-1.00)	0.053	15/44	0.25 (0.13-0.46)	<0.001	0.27 (0.14-0.51)	<0.001
	Missing	NA					NA				
	Trend		0.71 (0.59-0.86)	0.001	0.75 (0.61-0.92)	0.007		0.59 (0.47-0.73)	<0.001	0.62 (0.49-0.78)	<0.001
	Risk per increase in miscarriage		0.79 (0.68-0.93)	0.005	0.82 (0.69-0.97)	0.024		0.62 (0.51-0.75)	<0.001	0.65 (0.52-0.80)	<0.001
	P heterogeneity		0.085								

Abbreviations: CI, Confidence Interval; OR, Odds ratio; NA, Not Applicable; OC, Oral Contraceptive.

[a] Adjusted for age and region of residence.

[b] Adjusted on age, region of residence, education, rural-urban status, induced and spontaneous abortion, age at first full-term pregnancy, height, waist-to-hip ratio.

[c] Not adjusted for age at first full-term pregnancy.

[d] Adjusted for total number of pregnancies instead of age at first full-term pregnancy.

Missing values were excluded from analysis.

Table 2.9: Association of Anthropometric measurements, Body size and Risk of Breast Cancer

Parameters	Categories	Total (Cases=1637, Controls=1515)				
		Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
BMI (Kg/m^2)-world[c]	<18.5	138/58	1.53 (1.14-2.05)	0.004	1.84 (1.34-2.52)	<0.001
	18.5-24.9	748/690	Reference		Reference	
	25.0-29.9	540/513	0.95 (0.81-1.11)	0.548	0.90 (0.75-1.08)	0.278
	≥30	200/215	0.82 (0.66-1.03)	0.092	0.70 (0.55-0.89)	0.004
	Missing	11/12				
	Trend test		0.95 (0.89-1.01)	0.121	0.91 (0.85-0.98)	0.014
	Risk per unit increase in BMI		0.97 (0.95-0.98)	0.001	0.95 (0.94-0.97)	<0.001
BMI (Kg/m^2)-asia[c]	<18.5	138/85	1.56 (1.15-2.11)	0.004	1.78 (1.29-2.48)	<0.001
	18.5-22.9	458/432	Reference		Reference	
	23.0-24.9	290/258	1.04 (0.84-1.29)	0.658	0.92 (0.73-1.16)	0.512
	25.0-29.9	540/513	0.96 (0.81-1.16)	0.739	0.87 (0.72-1.07)	0.206
	≥30	200/215	0.84 (0.66-1.06)	0.154	0.68 (0.52-0.88)	0.004
	Missing	11/12				
	Trend test		0.95 (0.91-1.00)	0.102	0.91 (0.86-0.96)	0.002
	Risk per unit increase in BMI		0.97 (0.95-0.98)	0.001	0.95 (0.94-0.97)	<0.001
Height (in cm)[d]	≤150	648/658	Reference		Reference	
	151-155	560/438	1.31 (1.11-1.55)	0.001	1.48 (1.23-1.77)	<0.001
	156-160	284/312	0.93 (0.77-1.14)	0.527	1.04 (0.84-1.30)	0.684
	≥161	135/95	1.48 (1.11-1.98)	0.006	1.62(1.17-2.25)	0.003
	Missing	10/12				
	Trend test		1.06 (0.98-1.14)	0.111	1.10 (1.01-1.20)	0.018
	Risk for every 1 cm increase in height		1.00 (0.99-1.02)	0.092	1.01 (1.004-1.03)	0.008
	Risk for every 5 cm increase in height		1.05 (0.99-1.11)	0.092	1.09 (1.02-1.17)	0.008
Waist circumference (in cm)[e]	≤79	612/666	Reference		Reference	
	80-85	333/288	1.24 (1.02-1.51)	0.026	1.63 (1.31-2.04)	<0.001
	≥86	678/549	1.29 (1.10-1.52)	0.002	2.34 (1.84-2.96)	<0.001
	Missing	14/12				
	Trend test		1.13 (1.05-1.23)	0.002	1.53 (1.35-1.72)	<0.001
	Risk per 1 cm increase in waist circumference		1.009 (1.003-1.01)	0.002	1.04 (1.03-1.05)	<0.001
Hip circumference (in cm)[e]	≤90	565/393	Reference		Reference	
	91-99	516/529	0.65 (0.54-0.78)	<0.001	0.65 (0.53-0.80)	<0.001
	≥100	542/581	0.61 (0.51-0.73)	<0.001	0.60 (0.45-0.79)	<0.001
	Missing	14/12				
	Trend test		0.78 (0.72-0.86)	<0.001	0.77 (0.67-0.88)	<0.001
	Risk per 1 cm increase in hip circumference		0.98 (0.97-0.99)	<0.001	0.97 (0.96-0.98)	<0.001

Parameters	Categories	Total (Cases=1637, Controls=1515)				
		Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Waist-to-hip ratio[e]	≤0.84	541/825	Reference		Reference	
	0.85-0.94	787/550	2.18 (1.86-2.54)	<0.001	2.38 (2.01-2.81)	<0.001
	≥0.95	294/128	3.52 (2.77-4.46)	<0.001	3.78 (2.92-4.89)	<0.001
	Missing	15/12				
	Trend test		1.96 (1.76-2.19)	<0.001	2.06 (1.83-2.33)	<0.001
	Risk per 0.1 unit increase in WHR		1.65 (1.50-1.82)	<0.001	1.70 (1.53-1.89)	<0.001
Weight (in Kg)	≤60	1063/949	Reference		Reference	
	61-65	215/216	0.87 (0.71-1.07)	0.214	0.86 (0.68-1.10)	0.206
	>65	351/341	0.89 (0.75-1.06)	0.230	0.76 (0.62-0.93)	0.008
	Missing	8/9				
	Trend test		0.94 (0.86-1.02)	0.168	0.87 (0.79-0.96)	0.006
	Risk for every 1 Kg increase in weight		0.99 (0.98-0.99)	0.011	0.981 (0.97-0.989)	<0.001
	Risk for every 5 Kg increase in weight		0.96 (0.93-0.99)	0.011	0.91 (0.87-0.94)	<0.001
Increase in body size from age 10 to 20 using Pictogram[f]	No increase[g]	302/320	Reference		Reference	
	Moderate increase[h]	328/273	1.28 (1.02-1.61)	0.029	1.38 (1.08-1.76)	0.008
	Drastic increase[i]	97/96	1.07 (0.77-1.48)	0.672	1.18 (0.83-1.67)	0.336
	Missing	39/38				
	Trend test		1.09 (0.94-1.27)	0.228	1.15 (0.98-1.36)	0.071
Increase in body size from age 20 to current age using pictogram[f]	No increase[g]	55/50	Reference		Reference	
	Moderate increase[h]	227/208	0.99 (0.65-1.53)	0.998	1.23 (0.77-1.95)	0.372
	Drastic increase[i]	423/415	0.89 (0.59-1.34)	0.586	1.25 (0.76-2.04)	0.362
	Missing	41/36				
	Trend test		0.92 (0.77-1.09)	0.354	1.08 (0.86-1.34)	0.486
Body size at age 10 (using pictogram)	<3	732/695	Reference		Reference	
	3-4	473/431	1.04 (0.88-1.24)	0.568	1.06 (0.88-1.27)	0.523
	≥5	408/372	1.02 (0.86-1.22)	0.747	1.09 (0.90-1.32)	0.357
	Missing	24/17				
	Trend test		1.01 (0.93-1.10)	0.690	1.04 (0.95-1.14)	0.336
Body size at age 20 (using pictogram)	<3	338/351	Reference		Reference	
	3-4	675/600	1.18 (0.98-1.42)	0.078	1.19 (0.97-1.45)	0.085
	≥5	589/531	1.14 (0.94-1.38)	0.159	1.19 (0.96-1.46)	0.096
	Missing	35/33				
	Trend test		1.06 (0.96-1.16)	0.223	1.08 (0.97-1.19)	0.129

Parameters	Categories	Total (Cases=1637, Controls=1515)				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Body size at current age (using pictogram)	<3	144/117	Reference		Reference	
	3-4	470/433	0.89 (0.67-1.17)	0.413	0.74 (0.55-1.01)	0.060
	≥5	1002/951	0.83 (0.64-1.08)	0.188	0.64 (0.48-0.86)	0.003
	Missing	21/14				
	Trend test		0.92 (0.82-1.03)	0.171	0.82 (0.73-0.93)	0.003

Abbreviations: BMI, Body Mass Index; CI, Confidence Interval; OR, Odds Ratio.
[a]Adjusted for age and region of residence.
[b]Adjusted on age, region of residence, education, rural-urban status, menopausal status, induced and spontaneous abortion, age at first full-term pregnancy, height, waist-to-hip ratio.
[c] Not adjusted for height.
[d] Adjusted for weight instead of height.
[e] Adjusted for weight instead of waist-to-hip ratio.
[f] Adjusted for Body Mass Index instead of waist-to-hip ratio.
[g] No increase: Body size (pictogram) remained between 1 and 2.
[h] Moderate increase: Body size (pictogram) increased from 1-2 to 3-4.
[I] Drastic increase: Body size (pictogram) increased from 1-2 to 5-9.
Missing values were excluded from analysis.

Table 2.10: Association of Anthropometric Measurements, Body Size and Breast Cancer Risk stratified by Rural-Urban Status

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
BMI (Kg/m^2)-world[c]	<18.5	91/43	1.64 (1.11-2.41)	0.012	1.78 (1.17-2.72)	0.007	47/42	1.52 (0.96-2.41)	0.070	1.86 (1.14-3.04)	0.012
	18.5-24.9	528/409	Reference		Reference		220/281	Reference		Reference	
	25.0-29.9	411/354	0.88 (0.72-1.07)	0.214	0.87 (0.70-1.08)	0.233	129/159	1.01 (0.75-1.36)	0.911	1.00 (0.73-1.37)	0.969
	≥30	154/156	0.74 (0.57-0.95)	0.023	0.67 (0.50-0.90)	0.007	46/59	0.94 (0.61-1.45)	0.804	0.83 (0.52-1.32)	0.448
	Missing	11/10					0/2				
	Trend test		0.91 (0.84-0.98)	0.020	0.90 (0.82-0.97)	0.013		0.99 (0.88-1.11)	0.940	0.98 (0.86-1.10)	0.521
	Risk per unit increase in BMI		0.96 (0.94-0.97)	<0.001	0.95 (0.93-0.97)	<0.001		0.98 (0.95-1.01)	0.290	0.97 (0.94-1.00)	0.085
	P heterogeneity		0.765								
BMI (Kg/m^2)-asia[c]	<18.5	91/43	1.65 (1.11-2.47)	0.013	1.62 (1.03-2.52)	0.033	47/42	1.51 (0.94-2.42)	0.086	1.80 (1.09-2.99)	0.021
	18.5-22.9	311/244	Reference		Reference		147/188	Reference		Reference	
	23.0-24.9	217/165	1.02 (0.78-1.33)	0.848	0.95 (0.70-1.29)	0.788	73/93	0.97 (0.66-1.41)	0.889	0.90 (0.60-1.33)	0.604
	25.0-29.9	411/354	0.89 (0.71-1.11)	0.322	0.84 (0.65-1.10)	0.215	129/159	1.00 (0.73-1.38)	0.962	0.96 (0.68-1.36)	0.861
	≥30	154/156	0.74 (0.56-0.99)	0.043	0.61 (0.43-0.86)	0.005	46/59	0.93 (0.60-1.46)	0.781	0.80 (0.49-1.30)	0.380
	Missing	11/10					0/2				
	Trend test		0.92 (0.87-0.98)	0.016	0.90 (0.84-0.96)	0.004		0.98 (0.89-1.07)	0.675	0.95 (0.87-1.05)	0.385
	Risk per unit increase in BMI		0.96 (0.94-0.97)	<0.001	0.95 (0.93-0.97)	<0.001		0.98 (0.95-1.01)	0.290	0.97 (0.94-1.00)	0.085
	P heterogeneity		0.874								
Height (in cm)[d]	≤150	457/401	Reference		Reference		191/257	Reference		Reference	
	151-155	414/290	1.26 (1.03-1.55)	0.021	1.46 (1.16-1.82)	0.001	146/148	1.34 (0.99-1.80)	0.051	1.50 (1.09-2.06)	0.012
	156-160	204/204	0.88 (0.70-1.12)	0.329	0.99 (0.76-1.30)	0.974	80/108	1.02 (0.72-1.44)	0.909	1.09 (0.74-1.61)	0.641
	≥161	110/67	1.48 (1.06-2.07)	0.020	1.78 (1.21-2.62)	0.003	25/28	1.23 (0.69-2.19)	0.473	1.23 (0.66-2.31)	0.501
	Missing	10/10					0/2				
	Trend test		1.05 (0.96-1.14)	0.281	1.11 (1.003-1.23)	0.043		1.05 (0.91-1.20)	0.471	1.08 (0.92-1.26)	0.328
	Risk for every 1 cm increase in height		1.00 (0.99-1.02)	0.188	1.02 (1.004-1.03)	0.014		1.00 (0.98-1.02)	0.599	1.01 (0.98-1.03)	0.378
	Risk for every 5 cm increase in height		1.04 (0.97-1.11)	0.188	1.10 (1.02-1.19)	0.014		1.02 (0.92-1.14)	0.599	1.05 (0.93-1.19)	0.378
	P heterogeneity		0.700								

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Waist circumference (in cm)[e]	≤79	418/380	Reference		Reference		194/286	Reference		Reference	
	80-85	237/189	1.14 (0.89-1.44)	0.278	1.53 (1.17-2.01)	0.002	96/99	1.37 (0.98-1.93)	0.063	1.81 (1.23-2.67)	0.003
	≥86	526/393	1.19 (0.97-1.44)	0.080	2.25 (1.70-2.98)	<0.001	152/156	1.34 (0.99-1.80)	0.052	2.42 (1.55-3.80)	<0.001
	Missing	14/10					0/2				
	Trend test		1.09 (0.98-1.20)	0.082	1.50 (1.30-1.72)	<0.001		1.16 (1.006-1.35)	0.041	1.57 (1.25-1.96)	<0.001
	Risk per 1 cm increase in waist circumference		1.00 (0.99-1.01)	0.070	1.04 (1.02-1.05)	<0.001		1.01 (0.99-1.02)	0.078	1.04 (1/02-1.06)	<0.001
	P heterogeneity		0.322								
Hip circumference (in cm)[e]	≤90	396/223	Reference		Reference		169/170	Reference		Reference	
	91-99	383/321	0.66 (0.53-0.82)	<0.001	0.68 (0.52-0.88)	0.004	133/208	0.59 (0.43-0.81)	0.001	0.58 (0.41-0.84)	0.004
	≥100	403/418	0.52 (0.41-0.64)	<0.001	0.54 (0.39-0.75)	<0.001	139/163	0.77 (0.56-1.07)	0.123	0.75 (0.45-1.26)	0.285
	Missing	13/10					1/2				
	Trend test		0.72 (0.65-0.80)	<0.001	0.73 (0.62-0.86)	<0.001		0.87 (0.74-1.02)	0.110	0.83 (0.64-1.07)	0.163
	Risk per 1 cm increase in hip circumference		0.979 (0.971-0.98)	<0.001	0.97 (0.96-0.98)	<0.001		0.99 (0.97-1.00)	0.149	0.98 (0.96-1.00)	0.223
	P heterogeneity		0.014								
Waist-to-Hip ratio[e]	≤0.84	363/508	Reference		Reference		178/317	Reference		Reference	
	0.85-0.94	584/366	2.26 (1.87-2.73)	<0.001	2.50 (2.03-3.07)	<0.001	203/184	1.90 (1.44-2.50)	<0.001	2.13 (1.58-2.87)	<0.001
	≥0.95	234/88	3.83 (2.88-5.09)	<0.001	4.07 (3.00-5.53)	<0.001	60/40	2.52 (1.61-3.95)	<0.001	3.00 (1.84-4.90)	<0.001
	Missing	14/10					1/2				
	Trend test		2.04 (1.78-2.33)	<0.001	2.14 (1.85-2.47)	<0.001		1.69 (1.39-2.07)	<0.001	1.87 (1.51-2.33)	<0.001
	Risk per 0.1 unit increase in WHR		1.73 (1.53-1.94)	<0.001	1.76 (1.55-2.01)	<0.001		1.43 (1.22-1.69)	<0.001	1.55 (1.29-1.86)	<0.001
	P heterogeneity		0.773								

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Weight (in Kg)	≤60	755/560	Reference		Reference		308/389	Reference		Reference	
	61-65	161/153	0.77 (0.60-0.98)	0.039	0.82 (0.62-1.08)	0.171	54/63	1.08 (0.73-1.61)	0.685	1.08 (0.70-1.66)	0.703
	>65	271/251	0.79 (0.64-0.96)	0.024	0.70 (0.56-0.89)	0.004	80/90	1.07 (0.76-1.50)	0.694	0.97 (0.66-1.42)	0.895
	Missing	8/8					0/1				
	Trend test		0.87 (0.79-0.97)	0.013	0.84 (0.74-0.94)	0.003		1.03 (0.88-1.22)	0.640	0.99 (0.82-1.20)	0.972
	Risk for every 1 Kg increase in weight		0.98 (0.97-0.99)	0.001	0.980 (0.96-0.989)	<0.001		0.99 (0.98-1.00)	0.441	0.98 (0.97-1.00)	0.086
	Risk for every 5 Kg increase in weight		0.93 (0.90-0.97)	0.001	0.90 (0.86-0.94)	<0.001		0.97 (0.92-1.03)	0.441	0.94 (0.87-1.00)	0.086
	P heterogeneity		0.354								
Increase in body size from age 10 to 20 using Pictogram[f]	No increase[g]	224/213	Reference		Reference		78/107	Reference		Reference	
	Moderate increase[h]	241/181	1.28 (0.97-1.68)	0.070	1.39 (1.04-1.86)	0.025	87/92	1.29 (0.85-1.95)	0.225	1.37 (0.89-2.12)	0.149
	Drastic increase[i]	70/56	1.18 (0.79-1.76)	0.401	1.22 (0.80-1.87)	0.346	27/40	0.94 (0.53-1.68)	0.858	1.17 (0.64-2.16)	0.599
	Missing	31/24					8/14				
	Trend test		1.14 (0.95-1.37)	0.156	1.18 (0.97-1.43)	0.097		1.03 (0.79-1.35)	0.778	1.14 (0.86-1.52)	0.342
	P heterogeneity		0.952								
Increase in body size from age 20 to current age using pictogram[f]	No increase[g]	37/30	Reference		Reference		18/20	Reference		Reference	
	Moderate increase[h]	162/128	1.04 (0.61-1.78)	0.872	1.36 (0.75-2.44)	0.664	65/80	0.89 (0.43-1.84)	0.772	1.09 (0.51-2.33)	0.811
	Drastic increase[i]	321/285	0.89 (0.53-1.49)	0.682	1.39 (0.75-2.57)	0.208	102/130	0.81 (0.40-1.62)	0.557	1.10 (0.48-2.50)	0.808
	Missing	32/23					9/13				
	Trend test		0.90 (0.73-1.11)	0.366	1.11 (0.85-1.45)	0.084		0.90 (0.66-1.21)	0.500	1.03 (0.71-1.51)	0.846
	P heterogeneity		0.830								
Body size at age 10 (using pictogram)	<3	538/454	Reference		Reference		194/241	Reference		Reference	
	3-4	353/279	1.07 (0.87-1.31)	0.476	1.09 (0.87-1.36)	0.441	120/152	0.98 (0.72-1.33)	0.921	1.00 (0.72-1.38)	0.992
	≥5	284/227	1.04 (0.84-1.29)	0.705	1.09 (0.86-1.38)	0.473	124/145	1.06 (0.78-1.45)	0.670	1.06 (0.77-1.46)	0.700
	Missing	20/12					4/5				
	Trend test		1.02 (0.92-1.14)	0.626	1.04 (0.93-1.17)	0.419		1.03 (0.88-1.19)	0.699	1.03 (0.87-1.20)	0.716
	P heterogeneity		0.946								

Parameters	Categories	Lived less than 20 years in life in rural area (Urban) Cases=1195, Controls=972					Lived first 20 years of life in rural area (Rural) Cases=442, Controls=543				
		Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/ Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Body size at age 20 (using pictogram)	<3	251/228	Reference		Reference		87/123	Reference		Reference	
	3-4	501/386	1.19 (0.95-1.49)	0.120	1.23 (0.96-1.57)	0.095	174/214	1.14 (0.81-1.61)	0.436	1.14 (0.80-1.64)	0.455
	≥5	416/338	1.11 (0.88-1.40)	0.362	1.11 (0.86-1.43)	0.408	173/193	1.27 (0.90-1.80)	0.167	1.38 (0.96-1.99)	0.075
	Missing	27/20					8/13				
	Trend test		1.04 (0.93-1.16)	0.470	1.04 (0.91-1.17)	0.536		1.12 (0.95-1.33)	0.165	1.18 (0.98-1.41)	0.066
	P heterogeneity		0.325								
Body size at current age (using pictogram)	<3	97/61	Reference		Reference		47/56	Reference		Reference	
	3-4	333/265	0.80 (0.56-1.15)	0.236	0.66 (0.44-0.98)	0.042	137/168	0.96 (0.61-1.50)	0.862	0.87 (0.54-1.41)	0.596
	≥5	749/638	0.73 (0.52-1.03)	0.076	0.57 (0.39-0.83)	0.004	253/313	0.91 (0.59-1.40)	0.690	0.80 (0.50-1.27)	0.353
	Missing	16/8					5/6				
	Trend test		0.88 (0.76-1.01)	0.073	0.79 (0.68-0.93)	0.014		0.95 (0.79-1.15)	0.645	0.90 (0.73-1.10)	0.329
	P heterogeneity		0.512								

Abbreviations: BMI, Body Mass Index; CI, Confidence Interval; OR, Odds Ratio.
[a] Adjusted for age and region of residence.
[b] Adjusted on age, region of residence, education, rural-urban status, menopausal status, induced and spontaneous abortion, age at first full-term pregnancy, height, waist-to-hip ratio.
[c] Not adjusted for height.
[d] Adjusted for weight instead of height.
[e] Adjusted for weight instead of waist-to-hip ratio.
[f] Adjusted for Body Mass Index instead of waist-to-hip ratio.
[g] No increase: Body size (pictogram) remained between 1 and 2.
[h] Moderate increase: Body size (pictogram) increased from 1-2 to 3-4.
[I] Drastic increase: Body size (pictogram) increased from 1-2 to 5-9
Missing values were excluded from analysis.

Table 2.11: Association of BMI (kg/m^2) and Breast Cancer Risk in Premenopausal women

Parameters	Categories	Case/ Control	OR (95% CI)[a]	p-value	OR (95% CI)[b]	p-value	OR (95%CI)[c]	p-value
BMI (world) category	<18.5	81/58	1.40 (0.97-2.02)	0.067	1.48 (1.009-2.17)	0.045	1.82 (1.22-2.70)	0.003
	18.5-24.9	399/400	Reference		Reference		Reference	
	25.0-29.9	261/270	0.93 (0.75-1.17)	0.582	0.92 (0.73-1.17)	0.537	0.85 (0.67-1.09)	0.222
	≥30	75/108	0.66 (0.47-0.91)	0.014	0.59 (0.41-0.85)	0.005	0.52 (0.36-0.76)	0.001
	Missing	5/5						
	Trend test		0.91 (0.84-1.00)	0.055	0.90 (0.82-0.99)	0.033	0.87 (0.78-0.95)	0.005
	Risk per unit increase in BMI		0.95 (0.93-0.98)	<0.001	0.95 (0.92-0.97)	<0.001	0.93 (0.91-0.95)	<0.001
BMI (asia) category	<18.5	81/58	1.40 (0.96-2.06)	0.078	1.47 (0.98-2.19)	0.060	1.75 (1.16-2.65)	0.007
	18.5-22.9	251/255	Reference		Reference		Reference	
	23.0-24.9	148/145	1.00 (0.75-1.33)	0.984	0.97 (0.72-1.33)	0.894	0.90 (0.65-1.23)	0.521
	25.0-29.9	261/270	0.94 (0.73-1.20)	0.627	0.91 (0.70-1.20)	0.541	0.82 (0.62-1.08)	0.168
	≥30	75/108	0.66 (0.46-0.93)	0.020	0.59 (0.40-0.86)	0.007	0.50 (0.34-0.74)	0.001
	Missing	5/5						
	Trend test		0.93 (0.87-1.00)	0.073	0.91 (0.85-0.98)	0.023	0.87 (0.81-0.94)	0.001
	Risk per unit increase in BMI		0.95 (0.93-0.98)	<0.001	0.95 (0.92-0.97)	<0.001	0.93 (0.91-0.95)	<0.001

Abbreviations: BMI, Body Mass Index; CI, Confidence Interval; OR, Odds Ratio.

[a] Adjusted for age and region of residence.

[b] Adjusted on age, region of residence, education, menopausal status, induced and spontaneous abortion, age at first full-term pregnancy, height.

[c] Adjusted on age, region of residence, education, menopausal status, induced and spontaneous abortion, age at first full-term pregnancy, height, waist-to-hip ratio.

Missing values were excluded from analysis.

Table 2.12: Association of BMI (kg/m^2) and Breast Cancer Risk in Postmenopausal women

Parameters		Categories	Case/Control	OR (95% CI)[a]	p-value	OR (95%CI)[b]	p-value	OR (95%CI)[c]	p-value
BMI (world) category	Attained menopause <10yrs ago	<18.5	37/14	1.90 (0.99-3.66)	0.052	1.71 (0.87-3.36)	0.116	2.07 (1.04-4.14)	0.038
		18.5-24.9	221/157	Reference		Reference		Reference	
		25.0-29.9	168/141	0.84 (0.62-1.14)	0.275	0.94 (0.67-1.30)	0.711	0.87 (0.62-1.22)	0.432
		≥30	72/77	0.65 (0.44-0.96)	0.032	0.70 (0.46-1.05)	0.085	0.60 (0.39-0.91)	0.018
		Missing	3/3						
		Trend test		0.88 (0.78-0.98)	0.029	0.91 (0.80-1.03)	0.144	0.87 (0.77-0.99)	0.037
		Risk per unit increase in BMI		0.95 (0.93-0.98)	0.005	0.96 (0.93-0.99)	0.041	0.95 (0.92-0.98)	0.008
	Attained menopause ≥10yrs ago	<18.5	18/12	1.60 (0.74-3.49)	0.229	1.44 (0.60-3.41)	0.407	1.76 (0.72-4.26)	0.21
		18.5-24.9	117/129	Reference		Reference		Reference	
		25.0-29.9	103/99	1.16 (0.80-1.69)	0.423	1.33 (0.89-1.97)	0.156	1.19 (0.79-1.80)	0.382
		≥30	48/29	1.90 (1.12-3.23)	0.017	2.04 (1.17-3.57)	0.012	1.85 (1.05-3.28)	0.033
		Missing	2/3						
		Trend test		1.16 (1.004-1.34)	0.044	1.21 (1.04-1.42)	0.014	1.16 (0.99-1.37)	0.057
		Risk per unit increase in BMI		1.02 (0.98-1.06)	0.205	1.04 (0.99-1.08)	0.061	1.02 (0.97-1.06)	0.321
		P heterogeneity		0.042					
	Total menopausal women	<18.5	56/26	1.77 (1.08-2.90)	0.022	1.60 (0.95-2.71)	0.076	1.93 (1.12-3.30)	0.016
		18.5-24.9	347/286	Reference		Reference		Reference	
		25.0-29.9	278/238	0.96 (0.76-1.21)	0.738	1.07 (0.83-1.37)	0.564	0.98 (0.76-1.26)	0.889
		≥30	125/106	0.97 (0.71-1.31)	0.855	1.01 (0.74-1.39)	0.913	0.89 (0.64-1.23)	0.489
		Missing	6/7						
		Trend test		0.98 (0.89-1.07)	0.673	1.01 (0.92-1.11)	0.748	0.97 (0.88-1.07)	0.566
		Risk per unit increase in BMI		0.98 (0.96-1.00)	0.161	0.99 (0.97-1.01)	0.635	0.98 (0.95-1.00)	0.122
		P heterogeneity		0.450					

	Parameters	Categories	Case/Control	OR (95% CI)[a]	p-value	OR (95%CI)[b]	p-value	OR (95%CI)[c]	p-value
BMI (asia) category	Attained menopause <10yrs ago	<18.5	37/14	2.04 (1.04-4.03)	0.038	1.84 (0.91-3.71)	0.085	2.09 (1.02-4.28)	0.042
		18.5-22.9	118/90	Reference		Reference		Reference	
		23.0-24.9	103/67	1.17 (0.77-1.77)	0.448	1.18 (0.76-1.82)	0.445	1.02 (0.65-1.59)	0.918
		25.0-29.9	168/141	0.90 (0.63-1.29)	0.588	1.01 (0.69-1.48)	0.939	0.88 (0.59-1.30)	0.537
		≥30	72/77	0.70 (0.46-1.08)	0.109	0.75 (0.48-1.18)	0.225	0.61 (0.38-0.97)	0.038
		Missing	3/3						
		Trend test		0.91 (0.82-1.00)	0.056	0.94 (0.84-1.04)	0.24	0.89 (0.80-0.99)	0.035
		Risk per unit increase in BMI		0.95 (0.93-0.98)	0.005	0.96 (0.93-0.99)	0.041	0.95 (0.92-0.98)	0.008
	Attained menopause ≥10yrs ago	<18.5	18/12	1.44 (0.65-3.19)	0.364	1.31 (0.54-3.17)	0.538	1.55 (0.63-3.82)	0.337
		18.5-22.9	84/83	Reference		Reference		Reference	
		23.0-24.9	33/46	0.71 (0.41-1.22)	0.217	0.73 (0.41-1.30)	0.3	0.65 (0.36-1.17)	0.152
		25.0-29.9	103/99	1.04 (0.69-1.57)	0.836	1.20 (0.77-1.86)	0.405	1.03 (0.65-1.63)	0.882
		≥30	48/29	1.70 (0.98-2.98)	0.059	1.84 (1.01-3.32)	0.043	1.58 (0.86-2.91)	0.136
		Missing	2/3						
		Trend test		1.06 (0.95-1.19)	0.256	1.11 (0.98-1.25)	0.096	1.06 (0.93-1.20)	0.36
		Risk per unit increase in BMI		1.02 (0.98-1.06)	0.205	1.04 (0.99-1.08)	0.061	1.02 (0.97-1.06)	0.321
		P heterogeneity		0.050					
	Total menopausal women	<18.5	56/26	1.82 (1.10-3.04)	0.02	1.66 (0.96-2.85)	0.066	1.89 (1.09-3.29)	0.023
		18.5-22.9	205/174	Reference		Reference		Reference	
		23.0-24.9	142/112	1.07 (0.77-1.47)	0.672	1.08 (0.78-1.52)	0.614	0.94 (0.67-1.33)	0.767
		25.0-29.9	278/238	0.98 (0.75-1.28)	0.929	1.11 (0.83-1.48)	0.454	0.96 (0.71-1.28)	0.788
		≥30	125/106	0.99 (0.71-1.38)	0.998	1.05 (0.74-1.49)	0.762	0.87 (0.60-1.24)	0.452
		Missing	6/7						
		Trend test		0.98 (0.91-1.05)	0.592	1.01 (0.93-1.09)	0.773	0.96 (0.88-1.04)	0.333
		Risk per unit increase in BMI		0.98 (0.96-1.00)	0.161	0.99 (0.97-1.01)	0.635	0.98 (0.95-1.00)	0.122
		P heterogeneity		0.617					

Abbreviations: BMI, Body Mass Index; CI, Confidence Interval; OR, Odds Ratio.

[a] Adjusted for age and region of residence.

[b] Adjusted on age, region of residence, education, menopausal status, induced and spontaneous abortion, age at first full-term pregnancy, height.

[c] Adjusted on age, region of residence, education, menopausal status, induced and spontaneous abortion, age at first full-term pregnancy, height, waist-to-hip ratio.

Missing values were excluded from analysis.

Table 2.13: Association of Anthropometric Measurements, Body Size and Breast Cancer Risk stratified by Menopausal Status

Parameters	Categories	Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
		Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Height (in cm)[d]	≤150	267/359	Reference		Reference		281/293	Reference		Reference	
	151-155	293/255	1.53 (1.22-1.94)	<0.001	1.77 (1.37-2.29)	<0.001	266/179	1.14 (0.89-1.45)	0.287	1.27 (0.97-1.65)	0.074
	156-160	164/175	1.24 (0.95-1.62)	0.102	1.43 (1.06-1.93)	0.019	117/137	0.65 (0.48-0.87)	0.004	0.71 (0.52-0.99)	0.047
	≥161	93/47	2.68 (1.82-3.95)	<0.001	3.03 (1.94-4.74)	<0.001	42/47	0.68 (0.43-1.06)	0.093	0.72 (0.44-1.19)	0.208
	Missing	4/5					6/7				
	Trend test		1.25 (1.13-1.39)	<0.001	1.32 (1.17-1.49)	<0.001		0.85 (0.76-0.96)	0.008	0.89 (0.78-1.01)	0.082
	Risk for every 1 cm increase in height		1.03 (1.01-1.05)	<0.001	1.04 (1.02-1.06)	<0.001		0.98 (0.96-0.99)	0.041	0.99 (0.97-1.01)	0.361
	Risk for every 5 cm increase in height		1.18 (1.09-1.28)	<0.001	1.24 (1.12-1.37)	<0.001		0.91 (0.84-0.99)	0.041	0.95 (0.86-1.05)	0.361
	P heterogeneity		0.0002								
Waist circumference (in cm)[e]	≤79	347/419	Reference		Reference		264/241	Reference		Reference	
	80-85	198/157	1.49 (1.16-1.93)	0.002	1.91 (1.41-2.68)	<0.001	133/131	0.92 (0.68-1.24)	0.611	1.31 (0.94-1.83)	0.106
	≥86	267/260	1.21 (0.96-1.52)	0.104	1.96 (1.40-2.74)	<0.001	410/284	1.31 (1.04-1.66)	0.020	2.67 (1.90-3.76)	<0.001
	Missing	9/5					5/7				
	Trend test		1.11 (0.99-1.24)	0.067	1.41 (1.19-1.67)	<0.001		1.15 (1.02-1.29)	0.015	1.64 (1.38-1.94)	<0.001
	Risk per 1 cm increase in waist circumference		1.010 (1.001-1.019)	0.016	1.04 (1.02-1.05)	<0.001		1.00 (0.99-1.01)	0.109	1.04 (1.02-1.05)	<0.001
	P heterogeneity		0.018								
Hip circumference (in cm)[e]	≤90	320/257	Reference		Reference		243/134	Reference		Reference	
	91-99	257/287	0.69 (0.55-0.88)	0.003	0.65 (0.49-0.86)	0.003	257/235	0.60 (0.45-0.79)	<0.001	0.62 (0.45-0.85)	0.003
	≥100	235/292	0.60 (0.47-0.77)	<0.001	0.58 (0.40-0.85)	0.006	307/287	0.58 (0.45-0.76)	<0.001	0.57 (0.38-0.86)	0.008
	Missing	9/5					5/7				
	Trend test		0.77 (0.68-0.87)	<0.001	0.75 (0.62-0.91)	0.004		0.78 (0.68-0.89)	<0.001	0.75 (0.61-0.92)	0.007
	Risk per 1 cm increase in hip circumference		0.98 (0.97-0.99)	<0.001	0.97 (0.96-0.99)	0.004		0.98 (0.97-0.99)	<0.001	0.97 (0.95-0.98)	0.002
	P heterogeneity		0.938								

Parameters	Categories	Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
		Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Waist-to-Hip ratio[e]	≤0.84	310/509	Reference		Reference		231/310	Reference		Reference	
	0.85-0.94	372/275	2.26 (1.82-2.80)	<0.001	2.39 (1.89-3.03)	<0.001	412/273	2.03 (1.61-2.55)	<0.001	2.31 (1.81-2.96)	<0.001
	≥0.95	130/52	4.28 (2.99-6.13)	<0.001	4.11 (2.78-6.08)	<0.001	163/73	3.02 (2.18-4.18)	<0.001	3.43 (2.42-4.85)	<0.001
	Missing	9/5					6/7				
	Trend test		2.14 (1.83-2.50)	<0.001	2.15 (1.81-2.56)	<0.001		1.80 (1.54-2.11)	<0.001	1.95 (1.65-2.30)	<0.001
	Risk per 0.1 unit increase in WHR		1.70 (1.49-1.94)	<0.001	1.69 (1.47-1.96)	<0.001		1.61 (1.39-1.85)	<0.001	1.71 (1.47-2.00)	<0.001
	P heterogeneity		0.830								
Weight (in Kg)	≤60	553/547	Reference		Reference		507/393	Reference		Reference	
	61-65	115/108	1.02 (0.76-1.36)	0.882	0.91 (0.66-1.26)	0.576	99/106	0.72 (0.53-0.97)	0.036	0.84 (0.60-1.18)	0.327
	>65	149/181	0.78 (0.61-1.00)	0.059	0.56 (0.41-0.75)	<0.001	202/160	0.97 (0.76-1.24)	0.856	1.02 (0.77-1.35)	0.849
	Missing	4/5					4/4				
	Trend test		0.89 (0.79-1.01)	0.085	0.76 (0.66-0.88)	<0.001		0.96 (0.85-1.09)	0.586	1.00 (0.87-1.15)	0.971
	P heterogeneity										
	Risk for every 1 Kg increase in weight		0.99 (0.98-1.00)	0.089	0.97 (0.96-0.98)	<0.001		0.989 (0.980-0.99)	0.026	0.99 (0.97-1.00)	0.084
	Risk for every 5 Kg increase in weight		0.96 (0.92-1.00)	0.089	0.87 (0.82-0.92)	<0.001		0.94 (0.90-0.99)	0.026	0.95 (0.90-1.00)	0.084
	P heterogeneity		0.090								
Increase in body size from age 10 to 20 using Pictogram[f]	No increase[g]	143/172	Reference		Reference		159/146	Reference		Reference	
	Moderate increase[h]	177/155	1.40 (1.02-1.91)	0.034	1.54 (1.10-2.17)	0.012	151/116	1.18 (0.85-1.64)	0.317	1.29 (0.91-1.84)	0.145
	Drastic increase[i]	48/59	0.97 (0.62-1.52)	0.922	1.01 (0.62-1.65)	0.938	49/37	1.20 (0.74-1.94)	0.457	1.54 (0.91-2.60)	0.103
	Missing	19/20					19/18				
	Trend test		1.07 (0.87-1.32)	0.468	1.12 (0.89-1.40)	0.307		1.12 (0.89-1.39)	0.316	1.25 (0.99-1.59)	0.058
	P heterogeneity		0.180								

Parameters	Categories	Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
		Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Increase in body size from age 20 to current age using pictogram[f]	No increase[g]	29/28	Reference		Reference		26/22	Reference		Reference	
	Moderate increase[h]	131/137	0.91 (0.51-1.63)	0.772	1.07 (0.58-2.00)	0.808	96/69	1.16 (0.60-2.22)	0.643	1.60 (0.79-3.25)	0.188
	Drastic increase[i]	199/211	0.87 (0.49-1.52)	0.628	1.25 (0.64-2.47)	0.503	224/202	0.90 (0.49-1.64)	0.736	1.33 (0.64-2.76)	0.430
	Missing	18/18					22/18				
	Trend test		0.93 (0.74-1.18)	0.596	1.13 (0.84-1.54)	0.402		0.87 (0.68-1.12)	0.308	1.02 (0.74-1.40)	0.894
	P heterogeneity		0.515								
Body size at age 10 (using pictogram)	<3	372/389	Reference		Reference		360/302	Reference		Reference	
	3-4	247/253	1.03 (0.82-1.30)	0.739	1.09 (0.85-1.41)	0.459	225/176	1.07 (0.83-1.37)	0.581	1.01 (0.77-1.32)	0.914
	≥5	192/191	1.04 (0.81-1.33)	0.754	1.26 (0.96-1.64)	0.092	214/176	1.02 (0.79-1.31)	0.877	0.96 (0.73-1.26)	0.782
	Missing	10/8					13/9				
	Trend test		1.02 (0.90-1.15)	0.724	1.12 (0.98-1.27)	0.093		1.01 (0.89-1.14)	0.820	0.98 (0.85-1.12)	0.810
	P heterogeneity		0.594								
Body size at age 20 (using pictogram)	<3 fig	166/194	Reference		Reference		172/155	Reference		Reference	
	3-4 fig	353/347	1.21 (0.94-1.57)	0.130	1.31 (0.98-1.73)	0.059	321/250	1.15 (0.88-1.52)	0.293	1.09 (0.81-1.47)	0.542
	≥5 fig	285/284	1.17 (0.90-1.53)	0.235	1.32 (0.99-1.77)	0.056	302/241	1.12 (0.85-1.48)	0.386	1.07 (0.79-1.44)	0.629
	Missing	16/16					18/17				
	Trend test		1.07 (0.93-1.22)	0.305	1.13 (0.98-1.31)	0.077		1.05 (0.91-1.20)	0.451	1.08 (0.94-1.25)	0.261
	P heterogeneity		0.834								

Parameters	Categories	Premenopausal (Cases=818; Controls=841)					Postmenopausal (Cases=815, Controls=663)				
		Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value	Case/Control	OR[a] (95% CI)	p-value	OR[b] (95%CI)	p-value
Body size at current age (using pictogram)	<3	72/69	Reference		Reference		71/48	Reference		Reference	
	3-4	258/269	0.92 (0.63-1.34)	0.674	0.70 (0.47-1.06)	0.082	211/162	0.87 (0.57-1.33)	0.546	0.82 (0.52-1.30)	0.409
	≥5	481/500	0.88 (0.62-1.27)	0.520	0.62 (0.41-0.92)	0.011	520/442	0.79 (0.53-1.16)	0.236	0.68 (0.44-1.06)	0.091
	Missing	10/3					10/11				
	Trend test		0.95 (0.81-1.10)	0.516	0.82 (0.69-0.97)	0.025		0.89 (0.75-1.05)	0.175	0.83 (0.69-1.0002)	0.050
	P heterogeneity		0.819								

Abbreviations: CI, Confidence Interval; N, Number; OR, Odds ratio.

[a] Adjusted for age and region of residence.

[b] Adjusted on age, region of residence, education, menopausal status, induced and spontaneous abortion, age at first full-term pregnancy, height, waist-to-hip ratio.

[c] Not adjusted for height.

[d] Adjusted for weight instead of height.

[e] Adjusted for weight instead of waist-to-hip ratio.

[f] Adjusted for Body Mass Index instead of waist-to-hip ratio.

[g] No increase: Body size (pictogram) remained between 1 and 2.

[h] Moderate increase: Body size (pictogram) increased from 1-2 to 3-4.

[i] Drastic increase: Body size (pictogram) increased from 1-2 to 5-9.

Missing values were excluded from analysis.

2.8 Discussion

In the present study, a detailed analysis was conducted to understand the cause and estimate the risk of developing BC in women stratified by residence in rural-urban region and menopausal status. This is the only study to assess the risk of BC using different definitions for rural-urban status.

The strength of the study is that large number of study participants (1637 cases and 1515 controls) were enrolled in relatively short period of time with interviews conducted in a close room by trained investigators under constant supervision of senior staff. All the cases were histologically confirmed. The controls have been enrolled from a pool of visitors coming to TMH along with various cancer site patients. As the BC cases were enrolled from TMH, the selection of visitors as control, visiting TMH along with all cancer site patients in different units representing both first degree relative and other relatives, friends & neighbours in the ratio of 1:1, group matched on age and area of residence ensured that the selection bias is minimal. Anthropometric measurements have been taken by trained social investigators rather than relying on self-reported weights and heights by the cases and controls. Anthropometric studies usually rely on self-reported measures and the evidence suggests that obese women tend to underestimate their weight gain as compared to lean adolescents [275].

There was good correlation between main questionnaire and abbreviated questionnaire which was administered on approximately 8% of study participants (correlation coefficient range 69%-96%) indicating reliability of measured exposures. Only incident cases diagnosed not more than 6 months before the date of enrolment were enrolled to ensure that information on exposure given by the cases is not influenced because of long duration of illness and exposures related to survival. The constant monitoring of data at three levels helped to keep the missing information at minimum. The continuous training and preparation of manual ensured that information is collected similarly between cases and controls by different

interviewers. Further, as interviews were conducted in the closed room it was possible to collect information on reproductive variables with greater accuracy.

The case-control study design has however; inherited weakness as study participants are interviewed after the outcome has occurred and controls are not randomly selected from known population, leading to a possibility of recall and selection bias. In the present study, enrolment of controls from same study base, enrolling only incident cases, conducting interviews in closed room by trained investigators, constant monitoring, 8% resampling to measure reproducibility, as discussed, have helped to keep these biases at minimum.

2.8.1 Rural and Urban Status

Most of the etiological studies have used current area of residence as a definition for rural [150,276] and limited studies which have taken early years of life spent [277] or place of the birth [278,279] in rural areas as a definition for "rural." In the present study four definitions for rural and urban status were assigned to each study participant to estimate its association with BC risk. A current residence in a rural area compared to current residence in an urban area does not give protection against BC. A "dose-response" relationship between numbers of years lived in rural areas and increase in protection from BC development has been observed indicating that women living for many years in rural areas adhere to lifestyles, which are protective against BC.

The residence in first 20 years of life in rural areas was observed to be a protective factor against BC, As exposures in early life may be more important in the development of BC compared to current exposures [280]. For instance, strenuous physical activity at a younger age can delay both menarche and onset of regular menstrual cycle [281]. Further, the individuals migrating from rural area to urban area in adulthood might not change their lifestyle and continue to adhere to rural life and therefore they may continue to get protection from BC even if they are currently residing in urban areas which have been clearly

demonstrated in the present study. Therefore, while estimating the risk associated with anthropometric measurements and reproductive history, women who lived first 20 years of life in rural area were designated as 'rural' else categorized into urban.

2.8.2 Hormone Receptor Status

ER-/PR- BCs were more common in rural areas as compared to urban areas. Concurrently ER+/PR+ tumours were more prevalent in urban areas. These results are consistent with previous reports which have observed 2-4 times higher ER+ and PR+ malignancies in urban areas [276]. Another study conducted in a semi urban town in India observed more of Hormone Receptor negative (HR-) tumours [282], which clearly suggests that the proportion of HR- tumours increase from urban to rural areas. The reasons for higher incidence of HR+ tumours in urban areas are multi-factorial. It is quite possible that women in urban areas have better nutrition and development which leads to early menarche. They might be more educated resulting in higher age of marriage, lesser number of children and reduced breastfeeding [283]. All of these reproductive factors result in higher lifetime exposure of women to endogenous estrogens and thus can increase HR+ cancer in urban women.

2.8.3 Reproductive Factors

The results of the reproductive variables by doing stratification on rural-urban and pre and postmenopausal women showed that there were no stratum specific differences for women living in rural areas except that observed for twin pregnancies. However the risk of number of full-term pregnancies, age at first full-term pregnancy, interval between menarche & first full-term pregnancy and age at last pregnancy were significantly different among pre and postmenopausal women.

The results of the present study regarding age at first full-term pregnancy and the interval between age at menarche and first full-term pregnancy were consistent with previous results [65,71,74,81]. The significant effect between interval of age at menarche and first full-term

pregnancy, possibly further confirms association of age at first full-term pregnancy with increase in BC risk. However, the findings in present study with respect to age at menarche are not consistent with previous literature [64]. Epidemiologic studies of both premenopausal and postmenopausal women have consistently found that BC risk is reduced 5–20% for each year that menarche is delayed [82], but in the present study an older age at menarche was not associated with a reduced risk of BC. The reason for this inconsistency could be due to difficulties in obtaining exact age at menarche, as the information recorded is many years later than the occurrence of actual exposure, leading to a non-differential misclassification which could have diluted the estimates of OR [284,285]. No protection was observed in women who were ever pregnant as compared to women who were never pregnant which is in contrast to other studies [286]. This could be attributed to the fact that an "ever pregnant" woman comprises of abortion and miscarriage in addition to full-term pregnancies. The study population has lower prevalence of women with no history of pregnancy which could have been responsible in neutralizing the association. However when comparison was made among parous women the findings of present study were consistent with previous literature, showing protection with increase in number of pregnancies [70]. The results were when adjusted for age at first full-term pregnancy, the protection was no longer observed losing statistical significance at 5% level except in rural women, indicating that age at first full-term pregnancy is more important than number of full-term pregnancies. The protection observed in rural women could be attributed to the fact that higher number of pregnancies are more prevalent in rural areas as compared to urban areas with rural women having on an average 0.9 children more compared to urban women [59]. The risk of BC significantly increased with every 2 year delay in pregnancy after age 25 in women from both urban and rural areas. On stratifying into menopausal status, statistically significant association was observed only in premenopausal women, while the postmenopausal women showed increase but statistically

non significant risk. This is possibly because of homogeneity of population in relation to age at first pregnancy in older cohort where the median age at first child birth is lower as compared to the women from newer cohort [59]. The possible mechanism responsible for increase in risk is that mammary glands become fully differentiated at pregnancy, and that less differentiated ducts are more susceptible to carcinogens [287]. As carcinogen exposures accumulate with increasing age, later age at first full-term pregnancy then places the breast at greater risk [288,289].

Given the clear biological evidence that the interval between menarche and first full-term pregnancy is relevant because of the susceptibility of undifferentiated breast tissue [82,287], women with this longer window have shown to have an increased risk of BC in the present and previous studies [65,81,82].

In the present study the association was no longer significant when adjusted for age at first full-term pregnancy instead of total number of pregnancies (Data not shown). This is similar to observation by Chie et.al, which did not observe any association of BC with age at last full-term pregnancy after adjusting for age at other full-term pregnancies [125]. In a recently published report of EPIC cohort, which studied the reproductive factors according to receptor status, ER+/PR+ malignancies but did not find any statistical heterogeneity between BC subtypes [80].

Duration since last birth was associated with increased risk of BC in women from urban areas and those who were premenopausal. However, after adjusting for age at first full-term pregnancy, the association no longer exists (Data not shown). From the Swedish Fertility Register, with over 30 000 BC cases available for study, Liu et al. [128] documented a small increase in the risk of BC for each of the first few years after a birth, with adjustment for age at delivery in 1-year increments. Other studies, with considerably smaller numbers of white

women, have produced mixed results: some observed an increased risk for shorter interval since last birth [129–131] and a few found no association [132–134].

No association between breastfeeding and BC was observed in present study whereas few other studies have observed protective association with breastfeeding [83,109]. The protection was not observed for ever breastfed v/s never breastfed; average duration breastfed per child and total duration of breastfeeding. Breastfeeding is a universal practice in India, which has a very high prevalence of 95.7% and 96.0% in general population in both urban and rural areas respectively [59]. As population is homogenous, it is difficult to observe the protection associated with breastfeeding in India. A study done in South of India [87] had found a significant protection (without adjustment for parity) in premenopausal women who had breastfed for 6 years or more. Further, as total duration of breast feeding is associated with number of pregnancies, not adjusting for it might have resulted in confounding by protection given by number of pregnancies. The present study had only 0.6% of total women who have breastfed for more than 5 years (Data not shown) resulting in no association between BC and breastfeeding. A similar observation of no association was reported in a systematic review on Japanese population, where cohort studies failed to find a significant inverse association between breastfeeding and the risk of BC [86].

A positive association was seen between BC and maternal risk of twin pregnancy in rural and premenopausal women. A similar association was observed with the increased risk confined to the first five years following a multiple birth [290]. In contrast few studies observed an protective association [126,127] whereas another meta-analysis suggest that twin pregnancy does not significantly decrease the maternal risk of BC [73]. The question of a differential effect of multiple births compared to singleton on BC risk is of great importance from an etiological perspective, since many pregnancy hormones are increased in twin pregnancies compared to singleton [291–293]. Although statistically significant at 5% level the finding of

the present study needs further replication and could not exclude completely the possibility of chance, given the contradictory literature.

An association was observed between current OC users and BC. This association was stronger and reached statistical significance (P = 0.016) in urban women. When the current OC users were stratified by duration of use, an increased risk (OR = 2.46; CI: 0.99–6.12; P = 0.052) was observed for women who had used OC for more than 2 years. Other studies have also observed an increased risk with current use of OC [98–101]. Given the literature on OC use and BC and the strong significant association observed in the present study, the association may be interpreted as causal in urban India. As the association was significant only for current users, it can be interpreted that OC may act as late-stage promoters.

Other contraceptive methods like tubal ligation and Intrauterine devices were not associated with BC in the present study (Data not shown).

An increased risk of BC was observed in women who had undergone 2 or more than 2 induced abortions in their lifetime. On stratification, the risk was observed in urban, rural and premenopausal women. In the past, most case-control studies have observed an increased positive association [88] whereas most cohort studies have not reported an association [89–92,94,95]. This suggests that increased risk observed in the case-control study might be due to recall bias and its possibility cannot be excluded in the present study.

A consistent protection was observed across all strata (Rural, urban, pre and postmenopausal women) for women undergoing spontaneous abortion. The risk was highly significant and showed dose-response relationship across all strata. Given that the spontaneous abortion can occur after 1st trimester [294], it's plausible that the observed protection may be because of the protection acquired by pregnancy. Many studies have not observed an association between spontaneous abortion and risk of BC [90,93,94,96]. One cohort study however have observed

reduced risk in premenopausal and not postmenopausal women [97]. Future studies from India should indicate the trimester at which spontaneous abortion has occurred.

2.8.4 Anthropometric Measurements

The results of the anthropometric variables by doing stratification on rural-urban and pre and postmenopausal women showed that there were no stratum specific differences for women living in rural areas except that observed for hip circumference. However the risk of BMI, height and WC were significantly different among pre and postmenopausal women.

There are several tests that can be performed to determine obesity. Hydrostatic body fat test and Dual energy X-ray absorptiometry (DEXA) are more accurate than others. Unfortunately, these methods, however accurate, are expensive and complex to utilize in field epidemiological studies. As a result, easier methods have been developed to determine obesity, including Skin fold thickness (using callipers), WC, HC, Adult body weight, WHR [295] and they are shown to accurately predict the obesity. A correlation of more than 70% is observed between DEXA and BMI or WC [296].

Lower BMI (<18.5 kg/m^2) increased the risk of BC compared to women with BMI in normal range. This phenomenon was observed across strata irrespective of their residential or menopausal status. Low BMI is associated with undernutrition and metabolic syndrome [297–299]. It is well known that even at low BMI, Indians are at higher risk of developing type 2 diabetes mellitus and metabolic syndrome [300–306]. Indians with low BMI have higher central obesity [307–310]. Even in this study 17.6% of controls with low BMI (<18.5 kg/m^2) had higher WHR (≥0.85). Reverse causality is unlikely as weight loss in incident BC cases is least frequent phenomenon [311]. The increased risk for low BMI is therefore suggestive of risk related to metabolic syndrome. This is the first study to observe increased risk of BC with low BMI. Given the association of low BMI with metabolic syndrome this seems to be plausible.

The current study observed a protection from premenopausal BC with increase in the BMI, similar to observations of other investigators [312,313]. A possible mechanism suggestive of any protective effect of high BMI in young women could be because obese premenopausal women have a higher number of anovulatory cycles, resulting in decreased estradiol and progesterone levels [314], which causes reduced risk of BC [315].

In the present study, the average menopausal age in controls was 44.9 which is lower than the western countries [316]. No increase in risk in postmenopausal BC could be observed possibly because of carry-over protective effect. Other studies have observed a weak/no relationship between BMI and risk in postmenopausal women [317–319] possibly because the reduction in BC risk due to overweight in early adulthood appears to continue into the postmenopausal women. Pike et.al have argued that menopausal transition shifts BMI from a protective factor to a risk factor of BC in almost a decade. This effect was modelled to demonstrate that it takes a decade for a BMI of 30 kg/m^2 in a premenopausal woman (at age 50, RR of 0.75) to become a risk factor (RR of 1.20 at age 62) [320].

Therefore, in the present study the postmenopausal women were stratified by the duration since menopause is attained. In consistent with the above argument no increased risk with increase in BMI was observed in women who attained menopause less than 10 years ago and the observed risk was similar to that of premenopausal women. A strong significant increase (OR =1.85; P = 0.03) with dose-response was observed for women with high BMI who attained menopause ≥ 10 years ago. The finding of the present study is consistent to the previous data where the association tended to be stronger in studies from Asia Pacific (RR = 1.31; 95% CI: 1.15–1.48) than studies from North America, Europe, and Australia (RR = 1.15; 95% CI: 1.08–1.23 and 1.09; 95% CI: 1.00–1.14, resp.) In most case-control and prospective studies conducted in Asian women have observed a positive and strong relationship [112,149,162]. Pooled data from seven cohort studies including 337,819 women

and 4,385 incident BC cases found a 26% increase in postmenopausal BC risk with BMIs greater or equal to 28 kg/m^2 [139]. In a cohort study of 10,106 women, conducted in Japan, the RR for developing postmenopausal BC was 2.54; 94% CI (1.16–5.55) in women with BMI of 25 kg/m^2 or above compared to those with less 20.5 kg/m^2 [149].

Given the literature of protection observed in premenopausal women and increased risk in postmenopausal women seems to be causal. However, the paradox of low BMI and increased risk of BC needs further validation. Further studies with measurement of abdominal fat and correlation of inflammatory markers with low BMI are required to further confirm the finding of present study.

WHR in contrast to BMI is commonly used as a measure of central obesity [167,168]. A strong significant risk with dose-response was consistently observed in all the strata (Rural, urban, pre- and postmenopausal women) with larger WHR. The studies from literature have observed an increased risk associated with larger WHR among pre- and postmenopausal women particularly in Asian women compared to other ethnic groups [155,169,175], however, some studies are inconclusive [61,143,156]. In the present study the prevalence of WHR was higher in women residing in urban as well as rural areas. Even with women for normal BMI, 34.4% had higher WHR in the present study. This is consistent with thin-fat hypothesis of Asian –Indians [321]. The WC also showed increased in risk in rural, urban, pre- and postmenopausal women similar to other studies [313,322]. Given the high prevalence of central obesity in the Indian population [323] and strong risk observed for both pre- and postmenopausal BC, central obesity (WHR, WC) is main contributor to the risk of BC.

An inverse association was observed throughout with HC even after adjusting for number of full-term pregnancies, which was also observed by another Nigerian BC Study [179]. A large prospective cohort study of 11,889 women conducted in Asian population found that central adiposity reflected by HC was a significant predictor of BC [161]. Other studies showed an

inverse association in premenopausal women with HC [61,178]. In contrast, other studies showed a positive association between HC and BC risk [177,180]. The mechanism with which the association of HC with BC could be explained by the fact that the highest aromatase activity associated with estrogen metabolism was found in hip [324]. In addition, WC and HCs have been found to be associated with higher levels of androgens, insulin, and reduced levels of sex-hormone binding globulin [325], factors that have been previously linked with BC risk [326]. Overall, these observations may suggest a different role of body/build measurements in BC etiology between populations, and this may merit future exploration.

Height was observed to be positively associated with premenopausal BCs, which is consistent to previous reports [139,142,154,178]. No such increase was observed for postmenopausal women which could be due to lower prevalence of taller women in older cohort [59]. Similarly height was significantly associated with increased risk in urban women but not in rural women possibly because of the low prevalence of taller women in rural areas. NFHS 3 report also shows that the percentage of short stature women (below 145 cm) in rural areas is 12.1%, possibly due to poor nutrition status as compared to 9.8% in urban women [59]. Height, represents intrauterine, early childhood as well as the level of the adolescent growth spurt, which likely relates to factors such as nutrition, genetic growth potential, and hormones thus influencing BC occurrence [135–138]. Another possible explanation is that taller women may have higher levels of Insulin-like Growth Factors (IGF), which may underlie the link to BC risk [135].

The result of present study showed an increase in risk of BC with increase in body size (using pictogram) from age 10 to age 20 for postmenopausal women (OR = 1.25; 95% CI 0.99–1.59) after adjusting for current BMI. Body size at age 20 showed increase in the risk for premenopausal women (OR = 1.37; 95% CI: 1.03–1.83) and non statistical significant

increase for postmenopausal women when adjusted for current BMI (Data not shown). The risk was however attenuated for pre- and postmenopausal women ($OR_{premenopausal}$ = 1.32, P = 0.056) when adjusted for WHR.

Data on body shape evolution and BC risk are limited; in contrast weight gain has been found to be associated with BC risk [191,327]. Most of the Indian women have low birth weight, and higher weight at age 20 (as indicated by pictogram) or increase in body size from age 10 to 20 is indicative of accelerated childhood which may have attributed to increased adiposity and insulin resistance which might have influenced BC risk. No association was observed in increase in body size from age 20 to current age. To date, few studies have examined the relationship between body size at young age and BC risk. In contrast to other studies which showed a protective association, no association was observed between body sizes at age 10 and pre & postmenopausal BCs. The larger body size at the time of enrolment was observed to be associated with decreased risk of BC in premenopausal women. This is similar to the observation mentioned previously with BMI in the current study.

To conclude, it was observed that the risk factor for developing BC were similar between rural and urban area (no significant heterogeneity observed in the two groups) indicating that prevalence of these risk factors and not the risk factors per se may be responsible for observed differences in the incidence rates of BC in rural and urban areas. The BC showed heterogeneity in rural and urban area in reference to receptor status (P = 0.018) and non significant differences in mean age at menopause in rural (44.11) and urban area (44.73) and age at enrolment (rural women = 44.61; urban women = 45.24). However it cannot be inferred that differences in hormone receptor status also exist in rural-urban population as there could be referral biases towards reference to tertiary care hospital like TMH.

2.9 Summary

The strongest risk factors associated with BC after adjusting for confounding variables are as

follows:

1. For every 2 year increase in the age at first full-term pregnancy there is a 10% increase in risk of BC. The age at first full-term pregnancy after age 25 almost doubled the risk in rural and urban India.
2. Current users of OC with a longer duration increased the risk of BC in urban women and not in rural women, possibly due to its lower prevalence in rural areas.
3. Given the possibility of recall bias in the case-control study design the observed association of increase in risk of BC with induced abortion is difficult to interpret.
4. Spontaneous abortion may confer protection in both urban and rural women against the development of BC possibly because of the protection acquired from pregnancy lasting more than first trimester.
5. For every 5 cm increase in height there is an increase of BC in urban women but not in rural women possibly because of less prevalence of taller women in rural areas.
6. WHR showed strong significant positive association with BC in both rural - urban areas and in pre and postmenopausal women. The risk was more than 3-fold in highest category (≥ 0.95) as compared to lowest category (<0.85).
7. Low BMI seems to increase the risk for pre and post menopausal women. In the light of association of low BMI with metabolic syndrome, this finding observed for the first time in literature, needs to be replicated in other parts of India and in geographical areas where low BMI is highly prevalent.
8. High BMI protects from premenopausal BC but increases the risk in women who had attained menopause a decade ago.
9. ER-/PR- and TNBC cases are more prevalent in women from rural areas compared to those from urban areas.

The current study demonstrates that protection observed for BC by living in a rural area is

possibly because of less prevalence of risk factors viz. late age at first full-term pregnancy, OC use and tallness. The central obesity common to both rural and urban is a strong risk factor for BC.

Chapter 3

Single Nucleotide Polymorphisms & Breast Cancer Risk

3.1 Introduction

In addition to the lifestyle related factors BC risk can also be attributed to genetic factors involving both high penetrance gene mutations and low penetrance polymorphisms. High penetrance genes occur less than 1% in population whereas low penetrance polymorphisms are more common in general population. This leads to the conclusion that there must be an influence from low penetrant genes [328] and can be deduced that they may be important from public health perspective. SNPs explain up to 95% of all variant DNA sites [329]. There have been numerous studies on SNPs and risk of BC using various approaches viz Candidate SNP selection, GWAS and Sequencing. The identified SNPs have largely been reported from western population using GWAS [244,245,330–333]. There have been some SNPs identified in GWAS conducted in Asian population [242,334,335]. However, there has been no GWAS from India so far.

This chapter focuses mainly on replicating SNPs which have shown to be associated in already conducted large epidemiologic studies. SNPs associated with obesity have also been studied in detail and its association with BC risk have been determined.

3.2 Methodology

Female BC cases and visitor controls enrolled in the study were matched for age and region of residence as mentioned in Chapter 2. The controls unrelated to BC cases were utilized for Genotyping.

3.2.1 DNA Preparation & Assay Performance

Buffy coat samples were available for 1214 cases and 1293 controls. Genomic DNA was extracted from buffy coat using Qiagen QiAamp Blood DNA MidiKit and Macherey Nagel Nucleomag Blood kit. Concentration of each DNA sample was determined by the optical density (OD) at 260 nm and the purification was evaluated by OD 260/280 ratio. All DNA samples were also quantitated using Quant-iT PicoGreen dsDNA reagent, and purity was

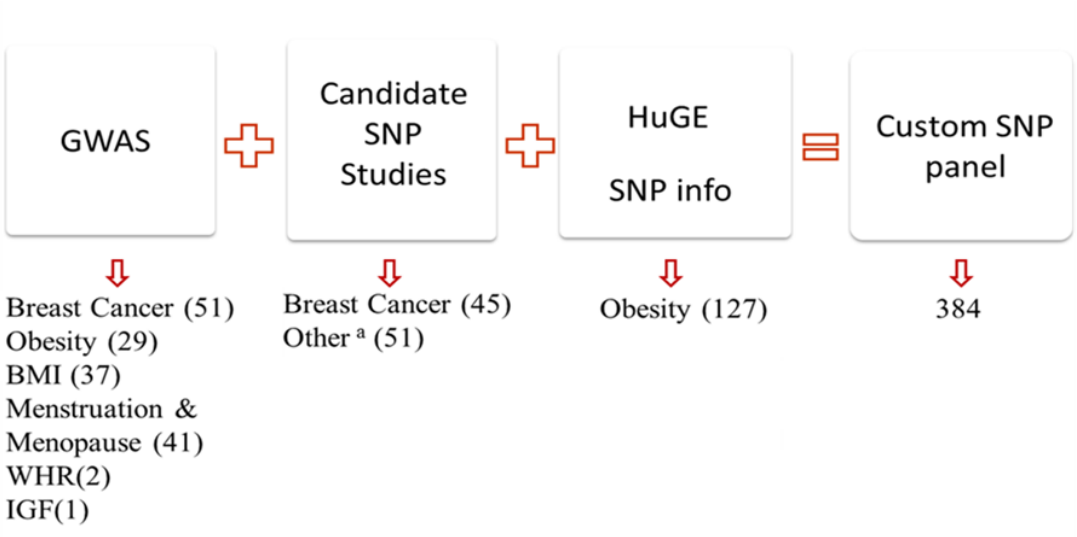
assessed by measuring the UV absorbance for accuracy. Mean total yield of 15.12 µg was obtained for 2416 samples (Cases: 1204; Controls: 1212) having 260/280 ratio in the range of 1.8 to 2.0. The yield was sufficient to proceed with genotyping. The quality of genomic DNA was assessed on 5% samples using 0.8% agarose gel. Single and intact bands of DNA were observed indicating that the isolated DNA was not fragmented and was of good quality. DNA concentrations were normalized to 50ng/µl and verified using Quant-iT PicoGreen dsDNA reagent before using for Genotyping assay. The aliquots of DNA were stored at -20^{0}C.

A total of 250ng DNA was applied to SNP typing using GoldenGate Genotyping (GGGT) Custom SNP Panel assay (Illumina Inc., San Diego, CA) [336]. GGGT assay was performed on 1204 cases and 1212 controls (Total: 2416) for 384 custom selected SNPs. Plates were prepared containing randomly mixed cases and controls. Intraplate and interplate replicates (7% approx.) were included on all plates and in all batches.

3.2.2 Design of Custom SNP Panel

A customized panel of 384 SNPs was designed using a mixture of 3 strategies which are as follows. A summary of the selection strategy in Figure 3.1.

Figure 3.1 Design of Custom SNP Panel



Abbreviations: GWAS, Genome Wide Association Study; IGF, Insulin-like Growth Factor ; SNP, Single Nucleotide Polymorphism; WHR, Waist-to-Hip Ratio.
[a] Suggested by Collaborator from animal experiments
Numbers of SNPs selected in each criteria are mentioned in parenthesis

3.2.2A SNPs selected from GWAS

The GWAS SNPs significantly associated with BC were identified using Human Genome Epidemiology (HuGE) Navigator [337]. National Institute of Health (NIH) GWAS Catalog [338] was used to identify significantly associated SNPs for BC, obesity, menstruation and menopause and other traits. The SNPs with p value < 10^{-5} were included in the custom panel. Duplicate SNPs between HuGE Navigator and NIH GWAS Catalog were removed. The number of SNPs selected in the respective category is mentioned in Table 3.1. A total of 161 SNPs were identified using this strategy.

Table 3.1: SNPs included using GWAS approach in different categories

Category (Number of SNPs)	**Trait/Disease**
Breast Cancer (**51**)	Breast Cancer
Obesity (**29**)	Obesity + Obesity (extreme)
Body Mass Index (**37**)	Body Mass (lean) +Body Mass Index
Mensturation and Menopause (**41**)	Menarche (age at onset) +Menopause (age at onset) + Menarche and menopause (onset)
Waist-to-hip ratio (**2**)	Waist-to-hip ratio
Insulin-like growth factor (**1**)	Insulin-like growth factor

Abbreviations: GWAS, Genome Wide Association Study; SNP Single Nucleotide Polymorphism
Number of SNPs selected in each category is mentioned in paranthesis

3.2.2B SNPs selected from Candidate Studies

All candidate SNP studies which have been significantly associated with BC (total number = 45) and other SNPs (suggested by collaborator on the basis of animal experiments, total number = 51) were included under this criterion using (HuGE) Navigator [339]. Total SNPs selected from this category were 96.

3.2.2C SNPs selected using Bioinformatics Tool

127 tag SNPs were selected using this strategy. Obesity search term was used in Gene evidence [340] tab of HuGE Navigator (Figure 3.2). Thirty three genes had a score of 0.05 or more which were uploaded on the Candidate gene SNP selection (Genepipe) pipeline of "SNPinfo" a web-based SNP selection tool [341]. Linkage Disequilibrium (LD) relationships were evaluated between SNPs using a pairwise LD data calculated from HapMap genotype data for ethnic populations of Utah residents with ancestry from northern and western Europe (CEU), Han Chinese in Beijing, China (CHB) and Gujarati Indian from Houston, Texas (GIH). The algorithm used for selecting SNPs from the provided list of genes of obesity is as follows: Five kb upstream and 1 kb downstream of the gene coordinate only were included in the selection. Only SNPs showing a minor allele frequency (MAF) of 0.05 or greater were included. Tagging proportion cut-off to filter gene was kept 0.8 and LD threshold cut off was kept 0.8. Minimum number of SNPs tagged by a tag SNP was kept as 3. In order to ensure that each gene has some coverage a minimum of 1 tag SNP to a maximum of 100 tag SNPs per gene were included. Further SNPs were filtered using the functional SNP prediction in "Genepipe" that cause an amino acid change, those that may alter the functional or structural properties of the translated protein, disrupt transcription factor binding sites, disrupt splice sites or other functional sites (Figure 3.3).

Figure 3.2: Snapshot of Gene Evidence tab of HuGE Navigator used for identifying genes associated with obesity

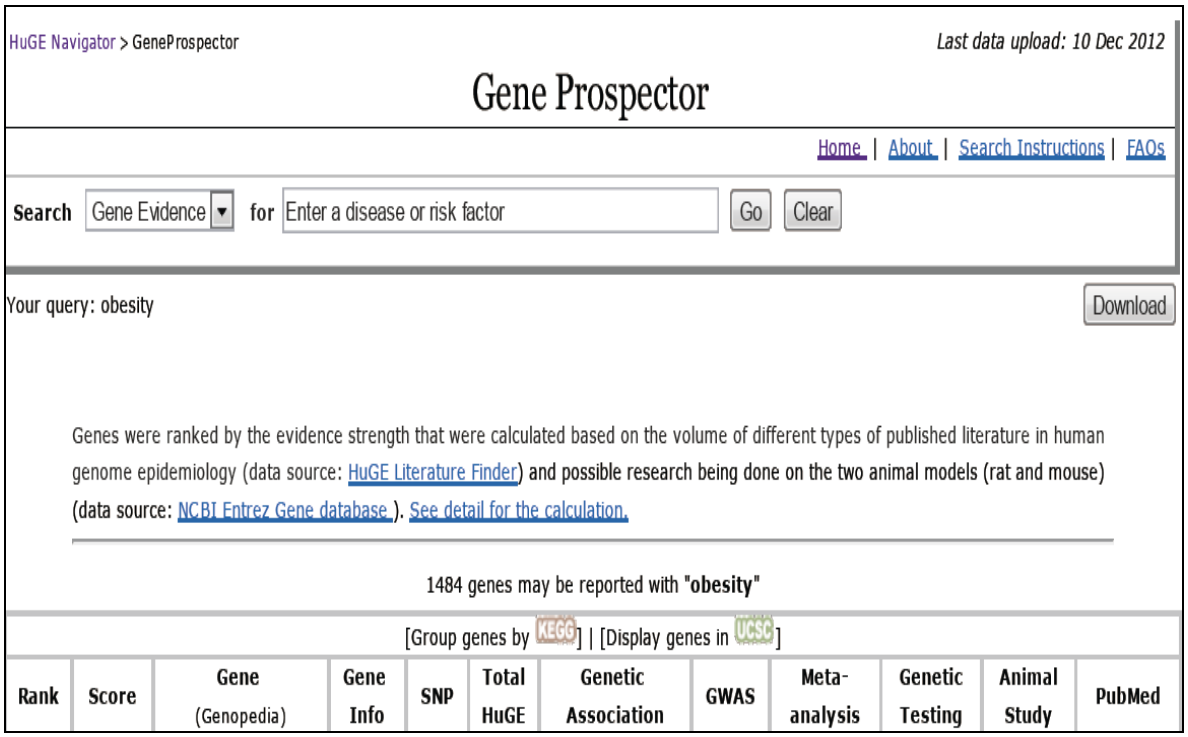
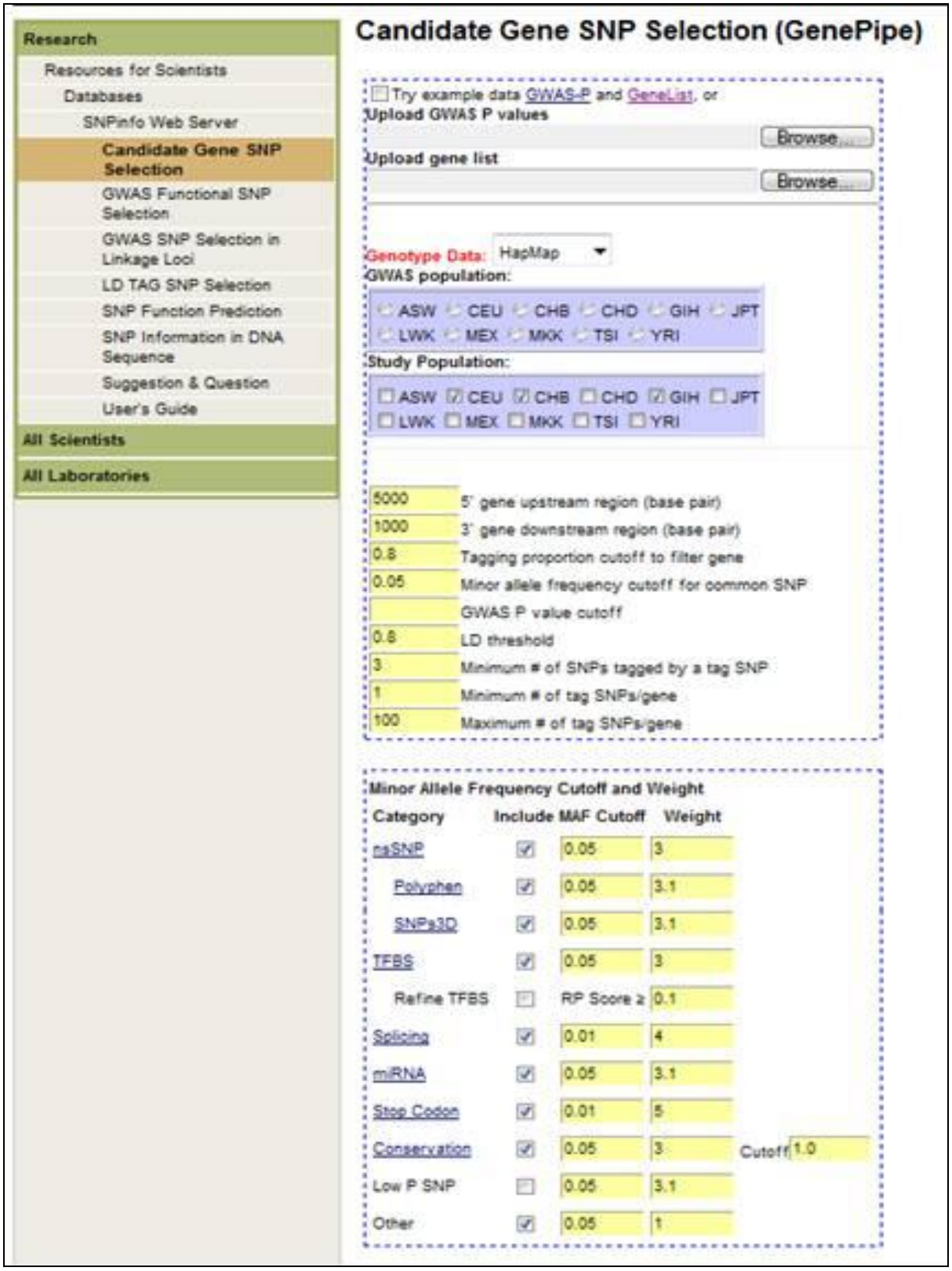


Figure 3.3: Snapshot of selecting SNPs from Candidate genes using GenePipe tab of bioinformatics tool SNPinfo



3.3 Quality Assessment

Blinded intra and interplate duplicates were included in all plates as a QC measure. The reproducibility rate of all the replicate samples (n=160) for all the assays was >98% (Table 3.2). No inter sample contamination was observed which was determined by including negative controls in some of the assays. A designability rank score (0-1.0) was calculated for each SNP by Illumina for the conversion of the SNP into a successful GoldenGate Assay. Of the 384 SNPs, 347 had a score of 1.0 (designability score =1.0, high success rate). Following completion of the assay, data cleaning was done using Illumina GenomeStudio software version 1.9.4. The automatic allele calling was done using a GenCall (GC) threshold of 0.25. The software assigned three clusters on a graph based on the fluorescence obtained. The GC score, a confidence score of the genotyping of each point, depends on the intensity of the fluorescence and the distance of the point from the centre of the cluster on the graph. Call rate, a quality assessment used for samples, defined as number of SNPs worked for a given sample divided by total number of SNPs genotyped. Seventeen samples had a call rate <90% (Figure 3.4), a total of 2399 samples were included in final analysis. Similarly to assess the quality of SNPs call frequency is used, which is number of samples worked for a given SNP divided by total number of samples genotyped. Six SNPs with call frequency <95% were excluded from final analysis (Figure 3.5). Further, 16 SNPs with diffused clusters, 4 SNPs with MAF<1%, 6 SNPs with call frequency below 95% and 6 SNPs with substantial deviation from Hardy-Weinberg Equilibrium (HWE) ($p<0.001$) were excluded to have a list of 352 SNPs for final analysis. All SNPs had a Gen train score value of 0.4 and above leading to no exclusions of SNPs due to poor cluster quality. The exclusion criteria has been summarised in flowchart (Figure 3.6) and the full list of SNP loci and corresponding gene is given in Annexure 1.

Table 3.2: Inter & Intra-assay reproducibility of Genotyping Assay

Reproducibility	Number of samples			Reproducibility Frequency		
	Cases	Controls	Total	Min-Max (Cases)	Min-Max (Controls)	Min-Max (Total)
Inter-assay	12	32	44	0.99-1.00	0.99-1.00	0.99-1.00
Intra-assay	57	59	116	0.98-1.00	0.99-1.00	0.98-1.00
Total	69	91	160	0.98-1.00	0.99-1.00	0.98-1.00

Figure 3.4: Histogram for Call Rate

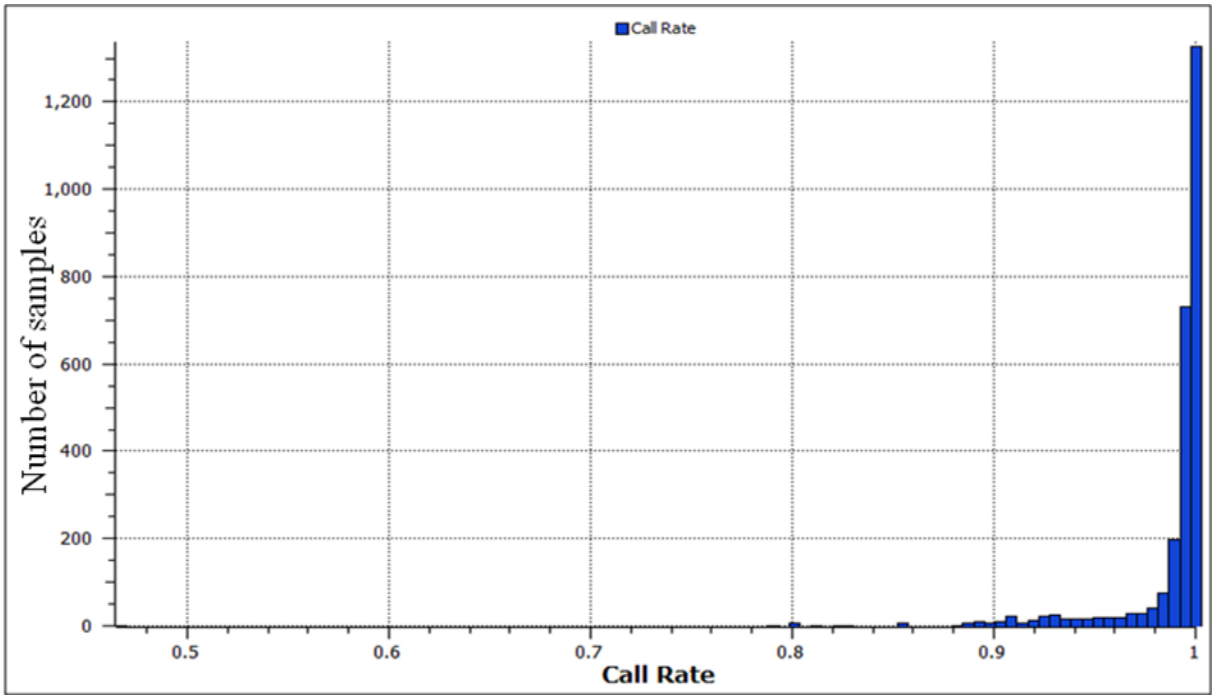


Figure 3.5: Histogram of Call Frequency

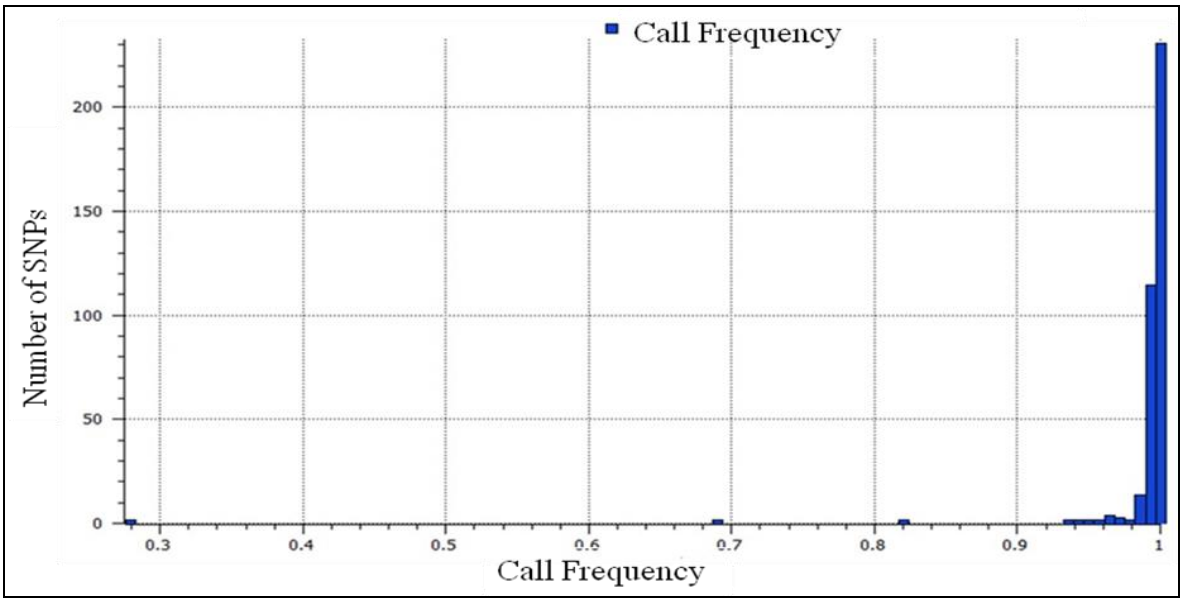
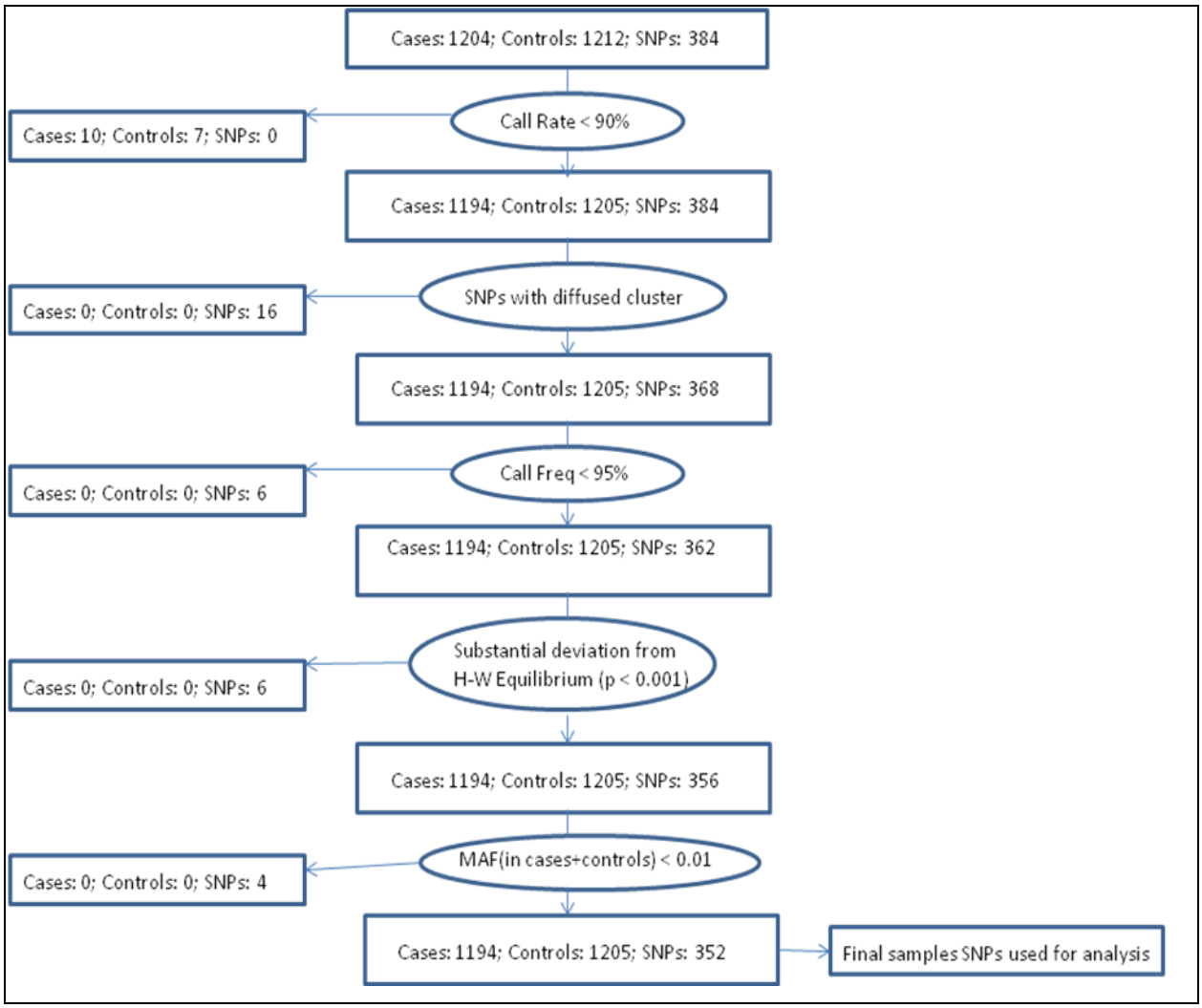


Figure 3.6: Flow chart of exclusion criteria in Genotyping data

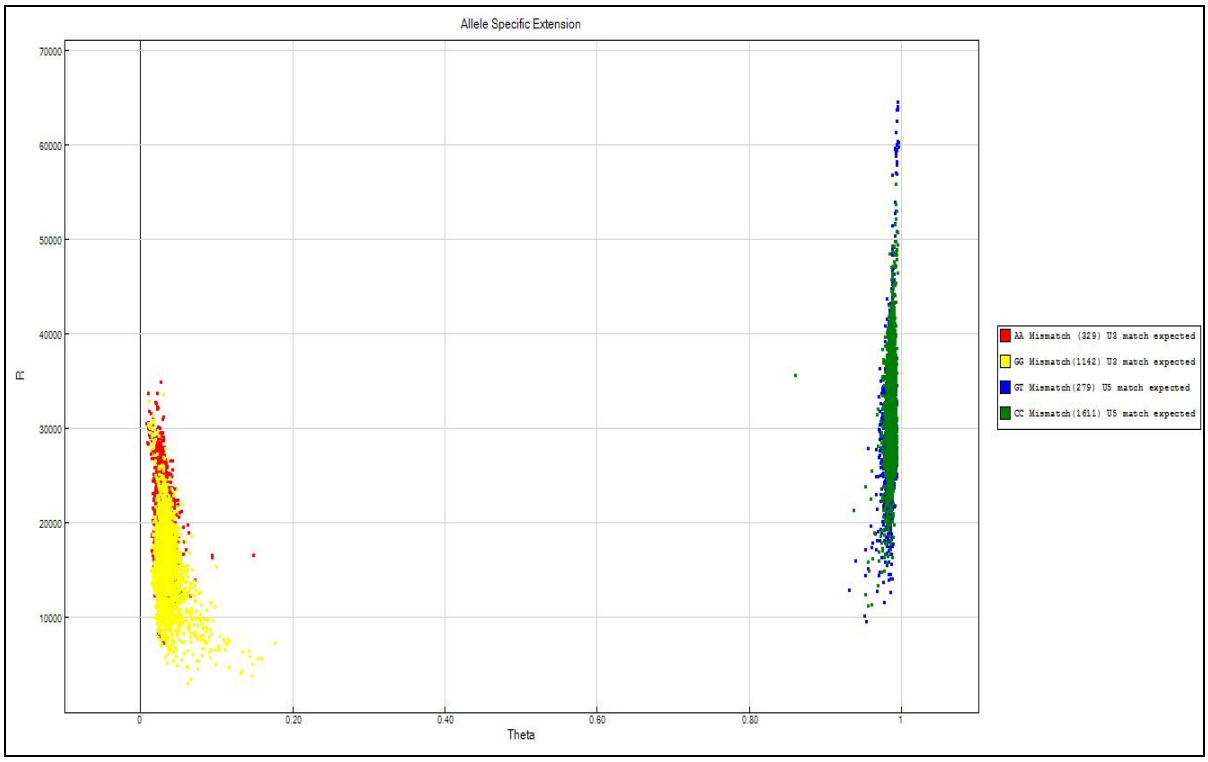


Following quality control dashboards provided in Genome Studio Software showed that the quality of the assays were satisfactory.

3.3.1 Allele Specific Extension

It is a test of efficiency of the extension step. The points have clustered on the sides of the graph. If the points were located in the centre of the graph, the hybridization of the probes would not have been specific, indicating something wrong during the Extension and Ligation Master Mix (MEL) step. This step shows that extension step had been properly conducted during the experiment (Figure 3.7).

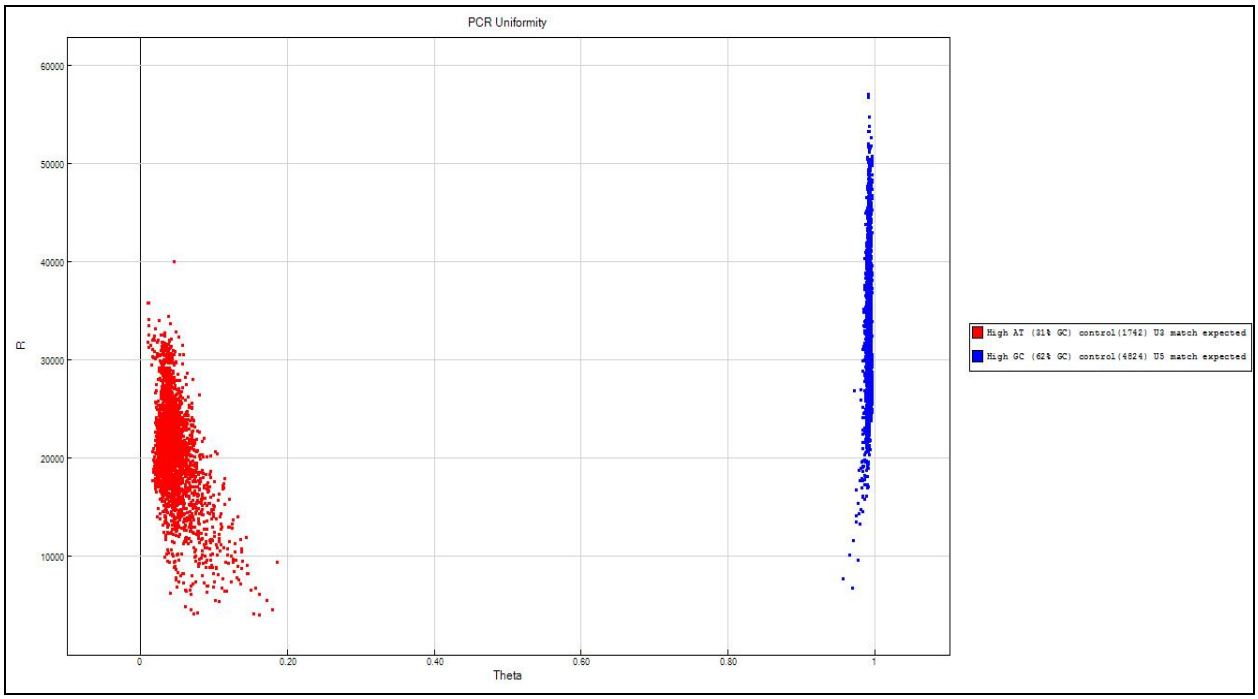
Figure 3.7: Allele Specific Extension



3.3.2 PCR Uniformity

The signal intensity of CY3 (blue) is visible on the right and CY5 (red) on the left panel of the graphs. If the PCR would have failed, the signal intensity would have been very low. The graph displayed below meets the quality criteria as mentioned in the Genome Studio software (Figure 3.8).

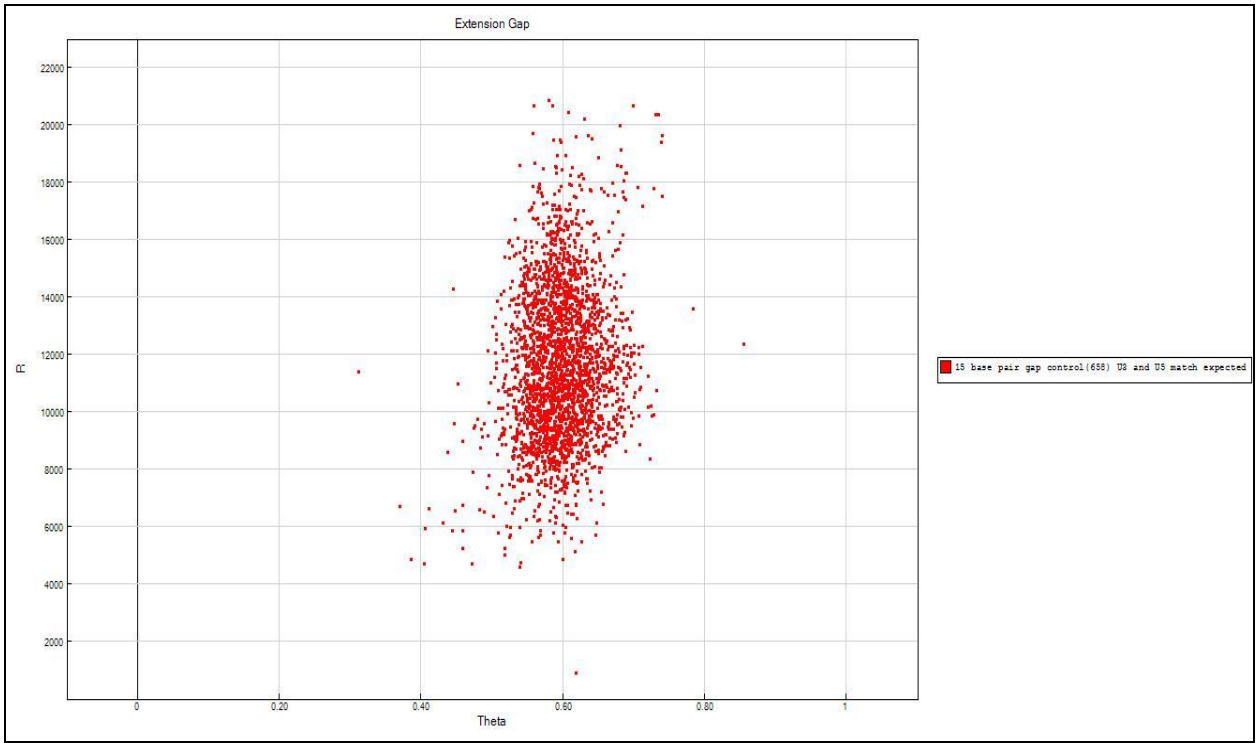
Figure 3.8: PCR Uniformity



3.3.3 Extension Gap

It tests the efficiency of extension. The signal intensity was high and centred as suggested in the protocol (Figure 3.9).

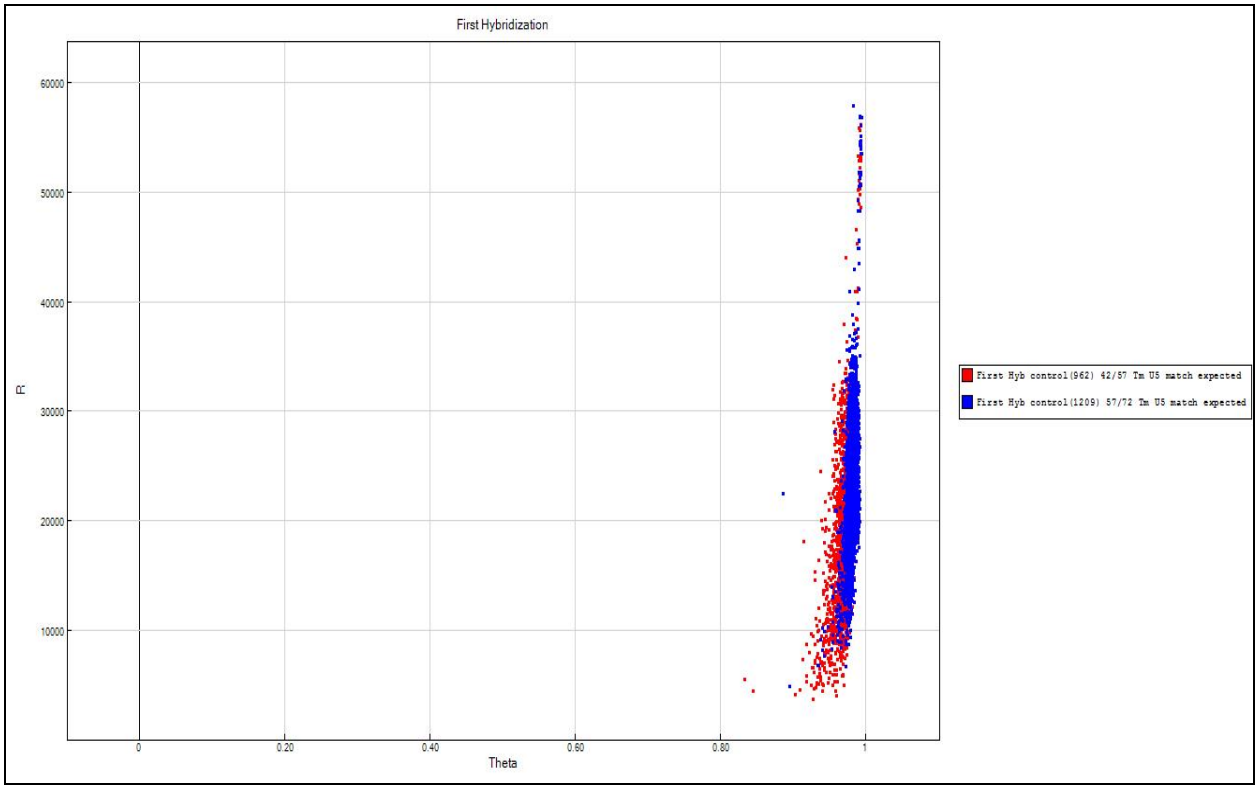
Figure 3.9: Extension Gap



3.3.4 First Hybridization

It tests the efficiency of the adjacent oligos to hybridize during the first hybridization step test when the temperature decreases from 70 to 30^0C. If the panel of points is located on the right of the graph, then the control is ok. This is rightly shown in the graph below (Figure 3.10)

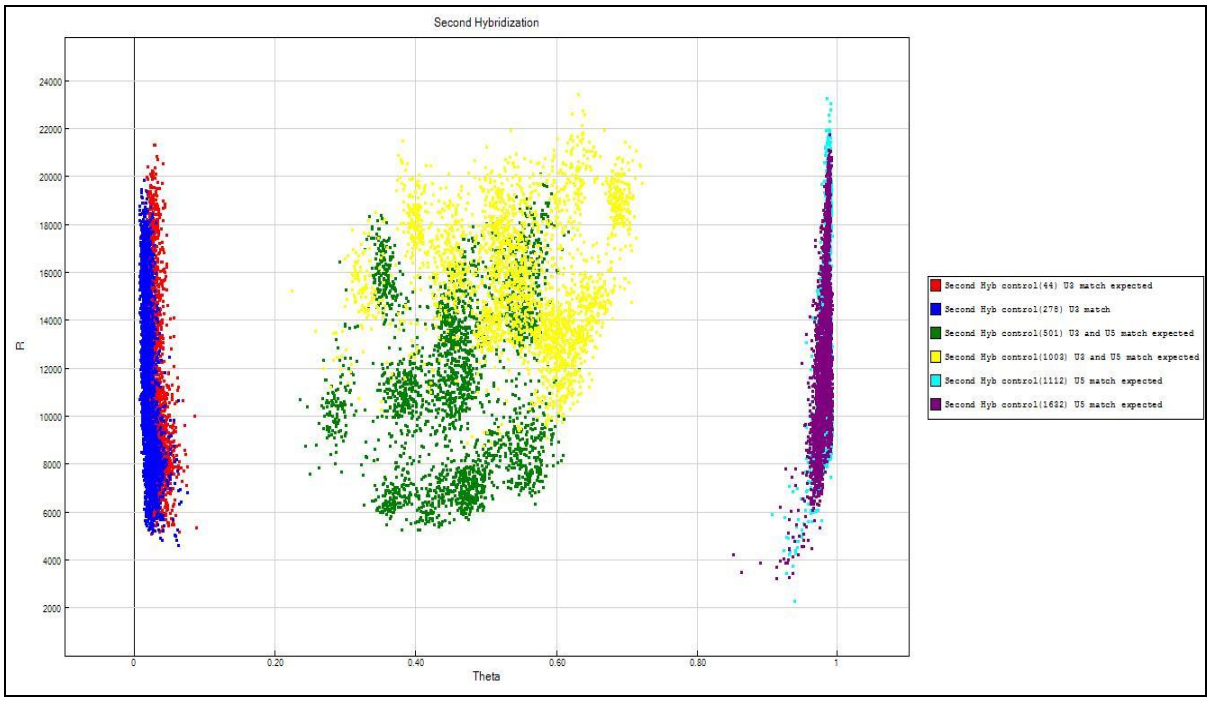
Figure 3.10: First Hybridization



3.3.5 Second Hybridization

It tests the efficiency of the hybridization. If the second hybridization failed, the signal intensity is low, reflecting a competition during hybridization. The intensity of all the samples are sufficiently high to include in the analysis as depicted in the below graph (Figure 3.11).

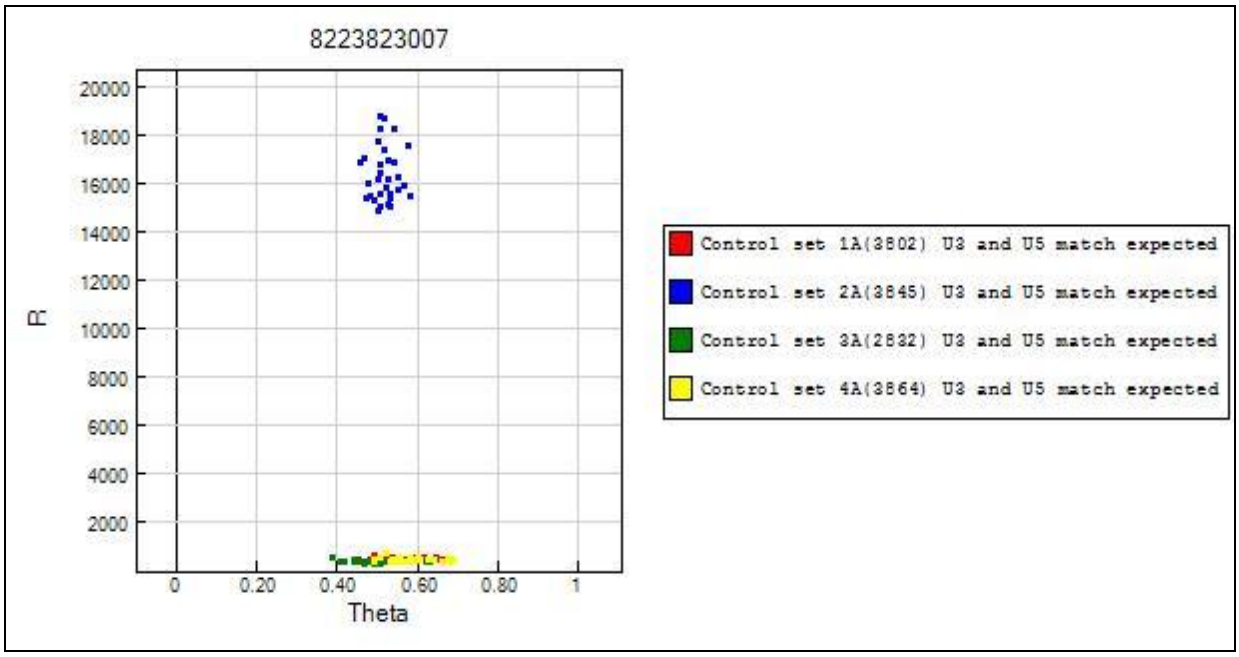
Figure 3.11: Second Hybridization



3.3.6 Contamination Dashboard

There is only one contamination control (one colour) per OPA tube. There are 4 different colours, so 4 contamination controls are possible. When checked for contamination via the dashboard, the graph must have only one colour (per OPA tube used). If a contamination occurred, 2 or more colours will be seen for samples that have been processed with the same OPA tube which reflects contamination of PCR products from the previous experiments. Since only one colour is observed in the graph below, it shows there was no inter sample contamination (Figure 3.12).

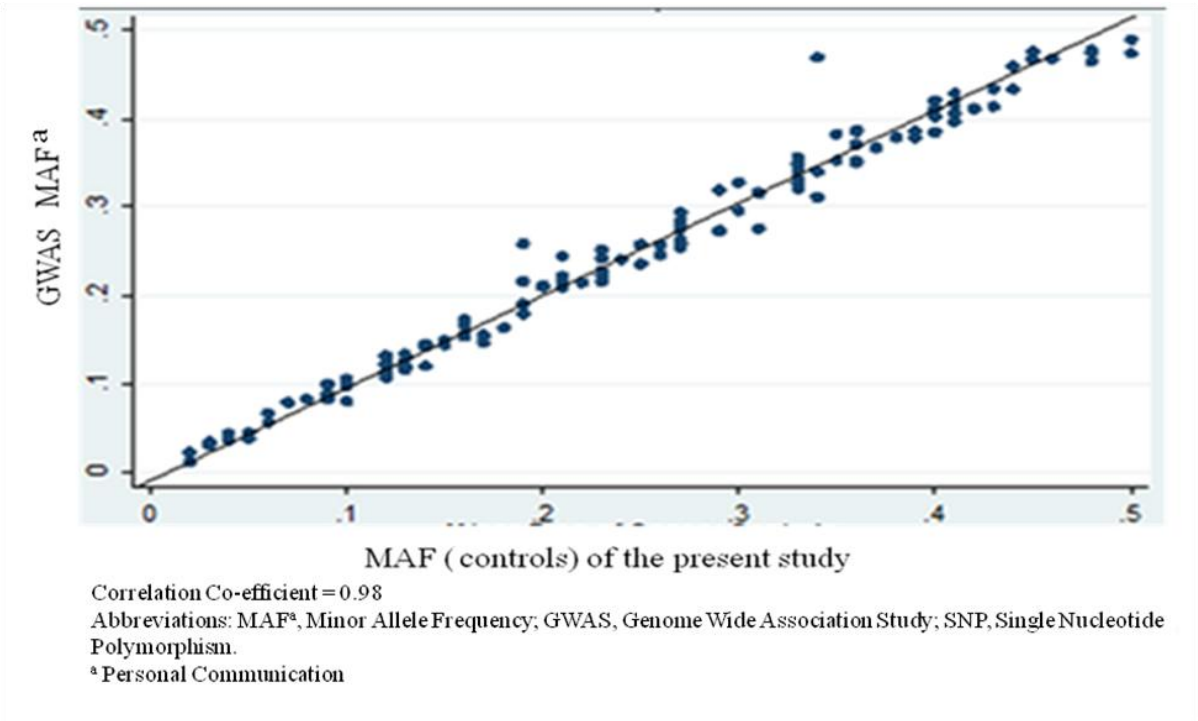
Figure 3.12: An example of Contamination Dashboard for a Beadchip of 32 samples



3.3.7 MAF Comparison with another study

MAF of the present study was compared with another ongoing GWAS in South India, a correlation of 98% was observed between the MAFs of two studies (Figure 3.13).

Figure 3.13: Correlation of MAFs between common SNPs in two Indian studies



3.4 Statistical Analysis

A chi-square test was used to verify whether the observed genotype frequencies were in HWE. Principal Component Analysis was conducted to evaluate the potential effects of population structure between the samples. Each principal component explains a certain percentage of the total variance in the SNPs. The first principal component explains the most SNP variation, and the second principal component uncorrelated with the first principal component explains the second most SNP variation. The number of principal components needed to describe common variation across a locus was defined a priori as the number of principal components needed to explain at least 80% of the sequence variance. Unconditional logistic regression was used to estimate OR and corresponding 95% CI between genotypes and case status. The genotypes were coded as 0=wild type, 1=heterozygous and 2=homozygous variant. The models fitted were additive (continuous effect of increasing number of variant alleles - 0 versus 1 versus 2), dominant (0 versus 1 and 2), recessive (0 and 1 versus 2) and genotypic (0 versus 1, 0 versus 2). Positive associations were defined as an OR larger than 1, whereas an inverse association was specified by an OR below 1. The results were considered significant at 0.05 level of significance. Allele and its corresponding frequency obtained from GWAS were recalculated for comparison of the SNPs replicated from BC GWAS in the present study. Additive model were presented in the main body of analysis whereas genotypic, dominant and recessive model were presented in Annexure 2.

A false discovery rate (FDR) method of Benjamini and Hochberg [342] was used to calculate *q-value*. A FDR cut-off of 0.05 was applied to select the top SNPs, which limited the probability of false-positives due to multiple tests that were carried out.

There was no significant difference in eigenvector loadings for the first five factors showing that the regional differences in structure were a minor source of population variability.

Therefore, the analysis were not conditioned on region. Similarly age was not associated with SNPs and therefore was not considered as the confounding factor in the model.

In the absence of any SNPs being statistically significant after adjusting for multiple comparisons, the results have been discussed in terms of possible false positive association versus real associations. SNPs significantly identified in BC GWAS and in present study were further analysed to estimate the association with BC stratified on receptor and menopausal status. The strongest SNPs observed in the present study which were previously related to obesity using bioinformatics tools and GWAS were selected to estimate interaction with WHR. Stratum specific estimates in the different strata of WHR for these SNPs were presented. Statistical significance of multiplicative interaction between SNPs and WHR were tested using the likelihood ratio test by comparing the logistic regression models with and without an interaction term.

All the analysis were performed using the statistical software Stata version 12.0 [274] and PLINK v1.07 [343,344].

3.5 Results

Out of 384 SNPs genotyped a total of 32 SNPs were excluded from final analysis due to various reasons mentioned above. A total of 27 SNPs were observed to be significantly associated with BC. However, neither of the associations remained statistically significant after adjustment for multiple comparisons.

Table 3.3 summarizes the number of significantly associated SNPs with BC risk using different categories i.e. GWAS, Candidate Studies and Bioinformatics tool. A total of 12 SNPs were associated out of 89 which had shown a previous association with BC. SNPs associated with Obesity and BMI were also associated with BC risk in the present study.

3.5.1 SNPs selected from GWAS

3.5.1A Breast Cancer

Table 3.4 shows 7 SNPs which were replicated from previous studies on BC GWAS. The risk of SNP in FGFR2 were 1.13 (95% CI: 1.01–1.27) and 1.16 (95% CI: 1.03–1.31) in rs2981575 and rs2981582 respectively in the additive model. Table 3.5 presents the MAF of SNPs which were significantly identified in the GWAS of BC but were not associated in the present study. The prevalence of certain SNPs in the present study varied with that observed in GWAS. Out of 40 SNPs which did not show an association with BC, 15 SNPs had a MAF below 20% and 25 SNPs had prevalence of 20% and above.

The risk of SNPs which were significantly identified in BC GWAS and in present study were also studied in BCs stratified on menopausal status and hormone receptor status. All SNPs were observed to be significantly associated with postmenopausal BCs whereas none of the SNPs analysed showed an association with premenopausal BCs. The SNP rs2046210 in ESR1 showed an increased risk for BCs which were ER-/PR- and triple negative but not ER+/PR+. The SNPs rs10411161 in Zinc Finger Protein 577 (ZNF577) gene showed significant protection in the development of only ER+/PR+ BCs whereas SNPs in FGFR2 (rs2981575, rs2981582), Mitogen-Activated Kinase Kinase Kinase 1 (MAP3K1) (rs889312) and 9q31.2 (rs865686) increased the risk of hormone receptor positive BC (Table 3.6 and 3.7).

3.5.1B Other-traits

Three SNPs associated with obesity and 2 associated with BMI showed a significant association with BC (Table 3.8). rs2274459 protected from the development of BC [0.74 (95% CI: 0.59–0.93)] and rs2116830 located in the untranslated region of Potassium Channel, Calcium-Activated, Large Conductance, Subfamily M, Alpha Member 1 (KCNMA1) gene showed an increased association of 1.51 (95% CI: 1.11–2.06). rs16867321 (P = 0.036) in Motilin (MLN) showed a significant protection in development of BC. SNPs previously

related to BMI showed a positive association with SNPs rs987237 and rs2287019 in genes Transcription Factor Ap2-Beta (TFAP2B) and Glutaminyl-Peptide Cyclotransferase-Like (QPCTL) respectively.

3.5.2 SNPs selected from Candidate Studies

3.5.2A Breast Cancer

Table 3.9 displays the association of 5 SNPs that were selected using Candidate Studies on BC. Statistical analysis on the individual effect of SNPs on BC showed an increased association with rs3218408, rs1641535, rs1641536 and rs861539. Another strongly associated SNP, rs2420946 in FGFR2 gene showed an increased risk of 1.17 (95% CI: 1.04–1.32).

3.5.2B Other SNPs

SNP rs2233660 included in final analysis suggested by Collaborators were also associated with BC risk (Table 3.9). The risk observed for rs2233660 in PHB gene was observed to be 1.21 (95% CI: 1.02-1.44)

3.5.3 SNPs selected using Bioinformatics Tool

The most promising SNP identified using the Bioinformatics tool is rs4362, present in the coding region of ACE (Angiotensin I-Converting Enzyme). The protective association was observed to be 0.84; 95% CI: 0.74–0.95. Other SNPs which had a protective association were rs4293, rs11196219 and rs11872992 whereas rs11121832, rs1474347 and rs3774261 showed an increased association with BC risk. The SNPs which had a p-value below 0.01 were rs1159460 and rs2161829 in the intronic region of genes Transcription Factor 7-Like 2 (TCF7L2) and Insulin-Induced Gene 2 (INSIG2) respectively (Table 3.10).

3.5.4 Gene-Environment Interaction

SNPs associated with obesity (using Bioinformatics tool and GWAS) previously and BC in present study with lowest p-values were selected to study the interaction between SNPs and

WHR. None of the SNPs showed statistically significant multiplicative interaction. However, there was an increase in the OR with increase in WHR for all selected SNPs (Table 3.11).

Table 3.3: Summary of significantly associated SNPs (Additive model) with Breast Cancer risk

Phenotype	GWAS	Candidate Studies	Bioinformatics tool	Total
Breast Cancer	7/47	5/42		12/89
Obesity	3/27		9/116	12/143
BMI	2/33			2/33
Menstruation and menopause	0/37			0/37
WHR	0/2			0/2
IGF	0/1			0/1
Others[a]	0/1	1/47		1/48
Significant association	12/147	6/89	9/116	27/352
Excluded	14	7	11	32
Total	161	96	127	384

Abbreviations: BMI, Body Mass Index; GWAS, Genome Wide Association Studies; IGF, Insulin-like Growth FactorWHR, Waist-to-Hip Ratio.
[a]Suggested by Collaborator on the basis of animal experiments
All SNPs are represented as Total significant/Total included in final analysis.

Table 3.4: Association of SNPs significantly identified in BC GWAS and risk of Breast Cancer

SNP ID	Chr	Gene Symbol	Major: Minor Allele	MAF [a]	Case/Control	OR (95% CI) [b]	p-value	Details of Previous GWAS[d]				
								Ethnicity (Ref. Study)	Minor allele (MAF)	Model	OR (95%CI)	p- value
rs16886165	5	MAP3K1	T:G	0.34	**1188/1195**	**1.20 (1.06-1.35)**	**0.003**	Not Given	G (0.15)	TT v/s TG	1.23 (1.12-1.35)	5.0×10^{-7}
rs865686	9	9q31.2	T:G	0.14	**1191/1199**	**0.80 (0.68-0.95)**	**0.010**	British, European	G (0.39)	Allelic	0.89 (0.85-0.92)	2.0×10^{-10}
rs889312	5	MAP3K1	A:C	0.39	**1178/1178**	**1.16 (1.03-1.30)**	**0.011**	British, European	C (0.28)	Allelic	1.22 (1.14-1.30)	5.0×10^{-9}
								Multiple[c]	A (0.28)	Allelic	1.13 (1.10-1.16)	7.0×10^{-20}
rs2981582	10	FGFR2	C:T	0.33	**1194/1204**	**1.16 (1.03-1.31)**	**0.019**	Multiple[c]	A (0.38)	Allelic	1.26 (1.23-1.30)	2.0×10^{-76}
rs10411161	19	ZNF577	C:T	0.48	**1182/1195**	**0.88 (0.79-0.99)**	**0.029**	European	T (0.13)	Add	1.42 (1.22-1.65)	7.0×10^{-6}
rs2046210	6	ESR1	C:T	0.35	**1189/1198**	**1.13 (1.009-1.28)**	**0.034**	Chinese	A (0.37)	Allelic	1.29 (1.21-1.37)	2.0×10^{-15}
rs2981575	10	FGFR2	T:C	0.38	**1187/1193**	**1.13 (1.01-1.27)**	**0.040**	European	T (0.42)	Allelic	1.28 (1.18-1.39)	1.0×10^{-8}
rs1219648	10	FGFR2	A:G	0.37	1188/1196	1.13 (1.00-1.27)	0.051	British	G (0.42)	Allelic	1.31(1.25-1.37)	1.0×10^{-30}
								European	G (0.42)		1.32 (1.22-1.42)	2.0×10^{-13}
								Multiple[c]	G (0.40)		1.2(1.07-1.42)	1.0×10^{-10}
rs10871290	16	GLG1	T:C	0.26	1192/1199	0.88 (0.77-1.00)	0.055	Not Given	C(0.34)	Not Given	Not Given	4.0×10^{-7}
rs7107217	11	BARX2	C:A	0.35	1181/1203	0.88 (0.78-1.00)	0.057	East Asian	C (0.36)	Allelic	1.08 (1.05-1.11)	5.0×10^{-7}
								Chinese	C (0.32)	Allelic	1.08(1.03–1.12)	2.2×10^{-4}
								Korean	C (0.38)	Allelic	1.10(1.04–1.16)	7.1×10^{-4}
								Japanese	C (0.47)	Allelic	1.05(0.96–1.15)	0.33
rs1011970	9	CDKN2BAS	G:T	0.26	1187/1201	1.13 (0.99-1.29)	0.062	Multiple[c]	T (0.17)	Allelic	1.09 (1.04-1.14)	3.0×10^{-8}
rs2981579	10	FGFR2	C:T	0.40	1192/1205	1.11 (0.98-1.24)	0.08	British, European	A (0.42)	Allelic	1.43 (1.35-1.53)	4.0×10^{-31}
								Multiple[c]	T (0.41)	CC v/s CT	1.17 (1.07-1.27)	2.0×10^{-10}
rs981782	5	HCN1	T:G	0.21	1192/1204	0.88 (0.76-1.01)	0.081	British	A (0.47)	Allelic	0.96 (0.93-0.99)	8.0×10^{-5}
								Multiple[c]	A (0.37)	Allelic	0.96 (0.93-0.99)	9.0×10^{-6}
rs1876206	15	FBN1	A:G	0.14	1189/1193	0.86 (0.72-1.02)	0.099	Framingham	Not Given	Not Given	Not Given	6.0×10^{-6}
rs614367	11	CCND1	C:T	0.15	1183/1189	1.13 (0.97-1.32)	0.112	British, European	T (0.15)	CC v/s CT	1.15 (1.10-1.20)	3.0×10^{-15}
rs3734805	6	C6orf97	A:C	0.07	1192/1200	1.17 (0.94-1.46)	0.135	British, European	C (0.08)	Allelic	1.19 (1.11-1.27)	1.0×10^{-7}
rs4784227	16	TOX3	C:T	0.22	1189/1201	1.10 (0.96-1.26)	0.154	Multiple[c]	T (0.24)	Allelic	1.24 (1.20-1.29)	1.0×10^{-28}
								Chinese	T (0.24)	Allelic	1.24 (1.19-1.30)	2.7×10^{-20}
								Japanese	T (0.24)	Allelic	1.34 (1.19-1.50)	7.3×10^{-7}
								East Asians	T (0.24)	Allelic	1.25 (1.20-1.31)	3.2×10^{-25}
								Multiple[c]	T (0.24)	Allelic	1.19 (1.09-1.31)	1.3×10^{-4}

SNP ID	Chr	Gene Symbol	Major: Minor allele	MAF [a]	Case/Control	OR (95% CI) [b]	p-value	Details of Previous GWAS[d]				
								Ethnicity (Ref. Study)	Minor allele (MAF)	Model	OR (95%CI)	p- value
rs13387042	2	TNP1	A:G	0.48	1188/1196	1.07 (0.95-1.20)	0.223	British, European	G (0.48)	Allelic	0.86 (0.82-0.90)	1.8×10^{-10}
								European	Not Given (0.53)		1.18 (1.10-1.27)	9.0×10^{-6}
								British, European	A (0.49)	Allelic	1.21 (1.14-1.29)	2.0×10^{-10}
								Multiple[c]	A (0.51)		1.25 (1.15-1.37)	2.0×10^{-8}
								Multiple[c]	A (0.49)	Allelic	1.2 (1.14-1.26)	1.0×10^{-13}
rs3803662	16	TOX3	C:T	0.28	1190/1195	1.07 (0.95-1.22)	0.235	European	A (0.30)		1.22 (1.13-1.32)	4.0×10^{-7}
								British, European	A (0.26)	Allelic	1.3 (1.22-1.39)	3.0×10^{-15}
								Multiple[c]	T (0.27)	CC v/s CT	1.16 (1.07-1.27)	1.0×10^{-9}
								Multiple[c]	T (0.27)	Allelic	1.28 (1.21-1.35)	6.0×10^{-19}
								Multiple[c]	C (0.25)	Allelic	1.2 (1.16-1.24)	1.0×10^{-36}
rs6504950	17	STXBP4	G:A	0.16	1190/1203	1.09 (0.94-1.27)	0.236	Multiple[c]	A (0.27)	Allelic	0.95 (0.92-0.97)	1.4×10^{-8}
rs9485372	6	UST	G:A	0.21	1183/1199	0.91 (0.79-1.05)	0.237	Chinese	A (0.43)	GG v/s AA	0.83 (0.76-0.90)	3.5×10^{-6}
								Korean	A (0.48)	GG v/s AA	0.76 (0.68-0.85)	6.0×10^{-7}
								Japanese	A (0.47)	GG v/s AA	0.84 (0.66-1.07)	0.15
								East Asian	A (0.45)	GG v/s AA	0.80 (0.75-0.86)	3.8×10^{-12}
rs10995190	10	ZNF365	G:A	0.08	1194/1205	0.88 (0.71-1.09)	0.269	British, European	A (0.15)	Allelic	0.86 (0.82-0.91)	5.0×10^{-15}
								British	A (0.14)	Allelic	0.76 (0.70-0.84)	6.1×10^{-8}
rs704010	10	ZMIZ1	G:A	0.29	1192/1198	1.06 (0.93-1.20)	0.339	British, European	A (0.39)	Allelic	1.07 (1.03-1.11)	4.0×10^{-9}
rs3817198	11	LSP1	T:C	0.36	1189/1194	1.05 (0.94-1.19)	0.342	Multiple[c]	C (0.30)	Allelic	1.07 (1.04-1.11)	3.0×10^{-9}
rs1926657	13	ABCC4	C:T	0.32	1187/1201	0.94 (0.83-1.06)	0.350	Framingham	Not Given	Not Given	Not Given	2.0×10^{-6}
rs13281615	8	FAM84B	G:A	0.5	1183/1189	0.94 (0.84-1.06)	0.365	Multiple[c]	C (0.42)	Allelic	1.08 (1.05-1.11)	5.0×10^{-12}
rs10263639	7	AUTS2	T:C	0.13	1190/1200	1.08 (0.92-1.27)	0.366	Framingham	Not Given	Not Given	Not Given	3.0×10^{-6}
rs1978503	18	TCF4	A:G	0.11	1193/1205	1.08 (0.90-1.29)	0.370	Framingham	Not Given	Not Given	Not Given	1.0×10^{-6}
rs2180341	6	RNF146	A:G	0.41	1182/1192	0.95 (0.84-1.06)	0.383	Ashkenazi Jews	G (0.21)	Allelic	1.41 (1.25-1.59)	3.0×10^{-8}
rs999737	14	RAD51L1	C:T	0.12	1194/1204	1.07 (0.90-1.28)	0.389	Multiple[c]	C (0.76)	Genotypic	1.06 (1.01-1.14)	2.0×10^{-7}
rs10069690	5	TERT	C:T	0.29	1187/1197	0.96 (0.85-1.09)	0.500	Multiple[c]		Allelic	1.18 (1.13-1.25)	1.0×10^{-10}
								African-American	T (0.57)	Allelic	1.32 (1.18–1.48)	1.3×10^{-6}
								European	T (0.27)	Allelic	1.18 (1.07–1.30)	1.0×10^{-3}
rs4973768	3	SLC4A7	C:T	0.45	1189/1198	1.03 (0.92-1.16)	0.536	British, European	C (0.49)	Allelic	1.14 (1.09-1.19)	2.0×10^{-8}
								British, European	T (0.47)	Allelic	1.16 (1.10-1.24)	6.0×10^{-7}
rs3112612	16	TOX3	C:T	0.46	1187/1197	1.03 (0.92-1.15)	0.541	British, European	T (0.43)	Allelic	1.15 (1.10-1.21)	4.0×10^{-10}

SNP ID	Chr	Gene Symbol	Major: Minor Allele	MAF [a]	Case/Control	OR (95% CI) [b]	p-value	Details of Previous GWAS[d]				
								Ethnicity (Ref. Study)	Minor allele (MAF)	Model	OR (95%CI)	p- value
rs8170	19	C19orf62	C:T	0.10	1190/1204	1.05 (0.87-1.27)	0.545	White	A (0.17)	Allelic	1.26 (1.17-1.35)	2.0×10^{-9}
rs458685	21	GRIK1	T:C	0.13	1193/1204	0.95 (0.80-1.12)	0.567	Framingham	Not Given	Not Given	Not Given	6.0×10^{-6}
rs4415084	5	FGF10	T:C	0.48	1184/1197	0.96 (0.86-1.08)	0.601	British, European	T (0.42)	Allelic	1.17 (1.11-1.22)	8.0×10^{-11}
rs909116	11	TNNT3	T:C	0.37	1189/1199	0.97 (0.86-1.09)	0.642	British, European	T (0.53)	Allelic	1.17 (1.10-1.24)	7.0×10^{-7}
rs11249433	1	LOC647121	T:C	0.17	1174/1190	1.03 (0.88-1.20)	0.682	Multiple[c]	C (0.39)		1.16 (1.09-1.24)	7.0×10^{-10}
rs10490113	2	BCL11A	A:C	0.16	1185/1198	0.98 (0.83-1.14)	0.754	Framhingham	Not Given	Not Given	Not Given	5.0×10^{-6}
rs10510102	10	ATE1	A:G	0.10	1190/1202	0.97 (0.80-1.17)	0.779	British, European	G (0.17)	Allelic	1.12 (1.07-1.17)	2.0×10^{-6}
rs10822013	10	ZNF365	T:C	0.50	1188/1193	0.98 (0.87-1.11)	0.833	East Asian	T (0.47)	Not Given	1.12 (1.06-1.18)	6.0×10^{-9}
rs2048672	7	LOC647017	G:T	0.44	1183/1196	1.01 (0.90-1.13)	0.852	East Asian	C (0.45)	AA v/s AC	1.11(1.05-1.17)	6.0×10^{-6}
								Chinese	C (0.41)	AA v/s AC	1.12 (1.05-1.20)	2.4×10^{-5}
								Japanese	C (0.5)	AA v/s AC	1.15 (0.96-1.37)	0.1738
								Korean	C (0.48)	AA v/s AC	1.05 (0.95-1.17)	0.1445
rs6556756	5	MAT2B	T:G	0.14	1175/1197	1.02 (0.86-1.20)	0.864	Framingham	Not Given	Not Given	Not Given	5.0×10^{-7}
rs1562430	8	FAM84B	A:G	0.23	1185/1204	1.00 (0.87-1.15)	0.920	British, European	G (0.40)	Allelic	0.86 (0.82-0.90)	3.1×10^{-11}
								British, European	T (0.58)	Allelic	1.17 (1.10-1.25)	5.8×10^{-7}
rs3757318	6	C6orf97	G:A	0.07	1193/1202	0.99 (0.79-1.24)	0.964	British, European	A (0.07)	Allelic	1.30 (1.17-1.46)	3.0×10^{-6}
rs1092913	5	ROPN1L	G:A	0.33	1189/1201	0.99 (0.88-1.12)	0.976	European	T (0.13)	Additive	1.45 (1.24-1.69)	2.0×10^{-6}
rs10941679	5	FGF10	A:G	0.39	1190/1200	1.00 (0.89-1.12)	0.981				1.27	2.5×10^{-12}

Abbreviations: ABCC4, ATP-Binding Cassette, Sub-Family C, Member 4; ATE1, Arginyltransferase 1; AUTS2, Autism Susceptibility Candidate 2; BARX2, BARX Homeobox 2; BCL11A, B-Cell CLL/Lymphoma 11A (Zinc Finger Protein); C19orf62, Chromosome 19 Open Reading Frame 62; C6orf97, Chromosome 6 Open Reading Frame 97; CCND1, Cyclin D1; CDK, Cyclin-Dependent Kinase; CDKN2BAS, Cyclin-Dependent Kinase Inhibitor 2B, Antisense; Chr, Chromosome; ESR1, Estrogen receptor 1; FAM48B, Family With Sequence Similarity 84, Member B; FBN1, Fibrillin 1; FGF10, Fibroblast Growth Factor 10; FGFR2, Fibroblast Growth Factor Receptor 2; GLG1, Golgi Apparatus Protein 1; GRIK1, Glutamate Receptor, Ionotropic, Kainate 1; GWAS, Genome Wide Association Studies; HCN1, Hyperpolarization-Activated Cyclic Nucleotide-Gated Potassium Channel 1; LSP1, Lymphocyte-Specific Protein; MAF, Minor Allele Frequency; MAP3K1, Mitogen-Activated Kinase Kinase Kinase 1; MAT2B; Methionine Adenosyltransferase II, Beta; RAD51L1, RAD51 Paralog B; RNF146, Ring Finger Protein 146; ROPN1L, rhophilin associated tail protein 1-like; SLC4A7, Solute Carrier Family 4 (Sodium Bicarbonate Cotransporter), MEMBER 7; SNP, Single Nucleotide Polymorphism; STXBP4, Syntaxin-Binding Protein 4; TCF4, Transcription Factor 4; TERT, Telomerase Reverse Transcriptase; TNNT3, Troponin T3, Fast Skeletal; TNP1, Transition Protein 1; TOX3, Tox High Mobility Group Box Family Member 3; UST, Uronyl 2-Sulfotransferase;, ZMIZ1, Zinc Finger Miz-Domain Containing 1; ZNF365, Zinc Finger Protein 365; ZNF577, Zinc Finger Protein 577.

[a] Minor Allele Frequency in controls.

[b] Unadjusted Odds Ratio fitted for Additive model.

[c] Risk estimated for more than one ethnicity.

[d] Adapted from HuGE Navigator.

Total number per SNP may vary because of missing values.

Significant association in the present study were shown in bold.

Table 3.5: Comparison of allele frequencies observed in present study & GWAS for SNPs which did not show association in present study

Allele Frequency reported in GWAS	MAF reported in the present Case-Control study[a]	
	0-19%	20-49%
0-19%	6[b]	2[c]
20-49%	3[d]	22[c]
Not Known	6[e]	1[c]
Total	15	25

Abbreviations: GWAS, Genome Wide Association Studies; MAF, Minor Allele Frequency.
[a] Minor Allele Frequency in controls of study population
[b] Not powered to detect association.
[c] Not a risk factor of breast cancer in the study population.
[d] Attributable risk may be low due to low prevalence.
[e] Not Known

Table 3.6: Association of SNPs significantly identified in BC GWAS & present study and risk of Breast Cancer stratified by menopausal status

SNP ID[a]	Gene Symbol	Premenopausal (Cases=818; Controls=841)			Postmenopausal (Cases=815, Controls=663)		
		Case/ Control	OR[b] (95% CI)	p-value	Case/ Control	OR[b] (95% CI)	p-value
rs10411161	ZNF577	599/643	0.94 (0.80-1.10)	0.455	**582/538**	**0.82 (0.69-0.96)**	**0.017**
rs16886165	MAP3K1	605/643	1.13 (0.96-1.32)	0.138	**582/538**	**1.27 (1.07-1.52)**	**0.006**
rs2046210	ESR1	604/645	1.09 (0.93-1.28)	0.267	**584/539**	**1.22 (1.01-1.46)**	**0.031**
rs2981575	FGFR2	604/641	1.04 (0.89-1.23)	0.572	**582/538**	**1.23 (1.03-1.46)**	**0.018**
rs2981582	FGFR2	607/649	1.05 (0.89-1.24)	0.539	**586/541**	**1.29 (1.07-1.55)**	**0.006**
rs865686	9q31.2	606/646	0.82 (0.64-1.04)	0.105	**584/539**	**0.76 (0.63-0.97)**	**0.030**
rs889312	MAP3K1	599/635	1.10 (0.94-1.30)	0.620	**578/529**	**1.24 (1.04-1.47)**	**0.008**

Abbreviations: BC, Breast Cancer; CI, Confidence Interval; ESR1, Estrogen Receptor 1; FGFR2, Fibroblast Growth Factor Receptor 2; GWAS, Genome Wide Association Studies; MAP3K1, Mitogen-Activated Kinase Kinase Kinase 1; OR, Odds Ratio; SNP, Single Nucleotide Polymorphism; TNBC, Triple Negative Breast Cancer, ZNF577, Zinc Finger Protein 577.
[a] SNPs significantly associated in the present study and in BC GWAS.
[b] Unadjusted Odds Ratio
Significant associations were shown in bold.
Total number per SNP may vary because of missing values.

Table 3.7: Association of SNPs significantly identified in BC GWAS & in present study and risk of Breast Cancer stratified by hormone receptor

SNP ID[a]	Gene Symbol	ER+/PR+ (N=569)			ER-/PR-(N=725)			TNBC (N=470)		
		Case/Control	OR (95% CI)[b]	p-value	Case/Control	OR (95% CI)[b]	p-value	Case/Control	OR (95% CI)[b]	p-value
rs10411161	ZNF577	**401/1190**	**0.77 (0.66-0.91)**	**0.002**	527/1190	0.92 (0.80-1.06)	0.274	339/1190	0.94 (0.80-1.12)	0.534
rs16886165	MAP3K1	**405/1190**	**1.27 (1.08-1.50)**	**0.003**	**526/1190**	**1.18 (1.01-1.37)**	**0.028**	338/1190	1.17 (0.98-1.40)	0.066
rs2046210	ESR1	407/1193	1.13 (0.95-1.33)	0.145	**527/1193**	**1.20 (1.03-1.40)**	**0.017**	**339/1193**	**1.23 (1.03-1.47)**	**0.020**
rs2981575	FGFR2	**406/1188**	**1.25 (1.06-1.47)**	**0.007**	528/1188	1.02 (0.87-1.18)	0.781	339/1188	1.05 (0.88-1.25)	0.550
rs2981582	FGFR2	**408/1199**	**1.23 (1.06-1.47)**	**0.014**	529/1199	1.07 (0.91-1.25)	0.374	340/1188	1.08 (0.90-1.31)	0.376
rs865686	9q31.2	407/1194	0.80 (0.63-1.02)	0.078	528/1194	0.85 (0.68-1.05)	0.145	339/1194	0.79 (0.61-1.03)	0.090
rs889312	MAP3K1	**404/1173**	**1.27 (1.08-1.49)**	**0.004**	524/1173	1.11 (0.96-1.29)	0.136	337/1173	1.04 (0.87-1.24)	0.620

Abbreviations: BC, Breast Cancer; CI, Confidence Interval; ER+, Estrogen Receptor Positive ; ER-, Estrogen Receptor Negative; ESR1, Estrogen receptor 1; FGFR2, Fibroblast Growth Factor Receptor 2; GWAS, Genome Wide Association Studies; MAP3K1, Mitogen-Activated Kinase Kinase Kinase 1; OR, Odds Ratio; PR+, Progesterone Receptor Positive; PR-, Progesterone Receptor Negative; SNP, Single Nucleotide Polymorphism; TNBC, Triple Negative Breast Cancer, ZNF577, Zinc Finger Protein 577.

[a] SNPs significantly associated in the present study and in BC GWAS.

[b] Unadjusted Odds Ratio fitted for Additive Model

Significant associations were shown in bold.

Total number per SNP may vary because of missing values.

Table 3.8: Association of SNPs significantly identified in GWAS on obesity related traits and risk of Breast Cancer

SNP ID	Chr	Gene Symbol	SNP Location	Major: Minor allele	MAF[a]	Phenotype	Case/Control	OR[b]	95% CI	p-value
rs2116830	10	KCNMA1	UTR	C:A	0.03	Obesity	1192/1205	1.51	1.11-2.06	0.009
rs2274459	6	MLN	Intergenic	G:A	0.07	Obesity	1184/1203	0.74	0.59-0.93	0.010
rs16867321	2	CWC22	Intergenic	C:T	0.35	Obesity	1189/1195	0.88	0.78-0.99	0.036
rs987237	6	TFAP2B	Intron	A:G	0.20	BMI	1182/1198	1.18	1.03-1.36	0.020
rs2287019	19	QPCTL	Intron	C:T	0.14	BMI	1193/1204	1.19	1.02-1.40	0.031

Abbreviation: BMI, Body Mass Index; Chr, Chromosome; CI, Confidence Interval; CWC22, CWC22 spliceosome-associated protein homolog; D; GWAS, Genome Wide Association Studies; KCNMA1, Potassium Channel, Calcium-Activated, Large Conductance, Subfamily M, Alpha Member 1; MAF, Minor Allele Frequency; MLN, Motilin; OR, Odds ratio; QPCTL, Glutaminyl-Peptide Cyclotransferase-Like; SNP, Single Nucleotide Polymorphism; TFAP2B, Transcription Factor Ap2-Beta; UTR, Untranslated Region.

[a] Minor Allele Frequency in controls.

[b] Unadjusted Odds Ratio fitted for Additive model.

Total number per SNP may vary because of missing values.

Table 3.9: Association of SNPs selected from Candidate Studies and risk of Breast Cancer

SNP ID	Chr	Gene Symbol	SNP Location	Major: Minor allele	MAF[a]	Phenotype	Case/Control	OR[b]	95% CI	p-value
rs2420946	10	FGFR2	Intron	C:T	0.38	Breast cancer	1187/1194	1.17	1.04-1.32	0.010
rs3218408	7	XRCC2	Intron	T:G	0.19	Breast cancer	1181/1197	1.17	1.01-1.35	0.031
rs1641535	17	ATP1B2	Intergenic	G:A	0.21	Breast cancer	1187/1193	1.16	1.01-1.33	0.039
rs1641536	17	ATP1B2	Intergenic	G:A	0.21	Breast cancer	1189/1197	1.15	1.01-1.33	0.043
rs861539	14	XRCC3	Coding	C:T	0.18	Breast cancer	1191/1184	1.16	1.00-1.34	0.044
rs2233660	17	PHB	Intergenic	T:C	0.11	Others	1194/1204	1.21	1.02-1.44	0.034

Abbreviations: ATP1B2, ATPase, Na+/K+ transporting, beta 2 polypeptide; Chr, Chromosome; CI; Confidence Interval; FGFR2, Fibroblast Growth Factor Receptor 2; MAF, Minor Allele Frequency; OR, Odds Ratio; PHB, Prohibitiin; SNP, Single Nucleotide Polymorphism XRCC3, X-Ray Repair Cross Complimenting Defective gene 3.
[a] Minor Allele Frequency in controls.
[b] Unadjusted Odds Ratio fitted for Additive model.
Total number per SNP may vary because of missing values.

Table 3.10: Association of Obesity related SNPs selected using Bioinformatics tool and risk of Breast Cancer

SNP ID	Chr	Gene Symbol	SNP Location	Major: Minor allele	MAF[a]	Case/Control	OR [b]	95% CI	p-value
rs11594610	10	TCF7L2	Intron	G:A	0.08	1194/1204	0.72	0.58-0.91	0.005
rs2161829	2	INSIG2	Intron	C:T	0.44	1190/1198	1.18	1.05-1.33	0.005
rs4362	17	ACE	Coding	C:T	0.36	1172/1180	0.84	0.74-0.95	0.005
rs11121832	1	MTHFR	Intron	C:T	0.27	1194/1201	1.16	1.02-1.32	0.024
rs1474347	7	IL6	Intron	T:G	0.12	1193/1202	1.21	1.02-1.43	0.030
rs4293	17	ACE	Intron	A:G	0.42	1185/1200	0.88	0.78-0.99	0.031
rs3774261	3	ADIPOQ	Intron	G:A	0.34	1184/1199	1.14	1.01-1.28	0.037
rs11196219	10	TCF7L2	Intron	G:A	0.33	1187/1198	0.88	0.78-0.99	0.044
rs11872992	18	MC4R	Intergenic	G:A	0.18	1186/1201	0.86	0.73-0.99	0.050

Abbreviations: ACE, Angiotensin I-Converting Enzyme; ADIPOQ, Adiponectin, C1Q And Collagen Domain Containing; Chr, Chromosome; CI, Confidence Interval; IL6, Interleukin 6; INSIG2, Insulin-Induced Gene 2; MAF, Minor Allele Frequency; MC4R, Melanocortin 4 Receptor; MTHFR, Methylenetetrahydrofolate reductase; OR, Odds Ratio; TCF7L2, Transcription Factor 7-Like 2.
[a] Minor Allele Frequency in controls.
[b] Unadjusted Odds Ratio fitted for Additive model.
Total number per SNP may vary because of missing values.

Table 3.11: Association of selected SNPs[a] and risk of BC stratified by WHR

SNP ID	Gene Name	WHR				P interaction
		0.85-0.94		≥0.95		
		OR[b](95% CI)	p-value	OR[b] (95% CI)	p-value	
rs11594610	TCF7L2	1.72 (1.22-2.43)	0.002	1.99 (1.0009-3.98)	0.050	0.535
rs2161829	INSIG2	2.40 (1.88-3.07)	<0.001	3.63 (2.64-4.99)	<0.001	0.745
rs4362	ACE	1.78 (1.41-2.24)	<0.001	2.71 (1.98-3.72)	<0.001	0.696
rs2116830	KCNMA1	2.77 (1.72-4.47)	<0.001	3.44 (1.67-7.09)	0.001	0.480
rs2274459	MLN	1.58 (1.13-2.21)	0.007	2.39 (1.13-5.04)	0.021	0.756

Abbreviations: ACE, Angiotensin I-Converting Enzyme; BC, Breast Cancer; CI, Confidence Interval; INSIG2, Insulin-Induced Gene 2; KCNMA1, Potassium Channel, Calcium-Activated, Large Conductance, Subfamily M, Alpha Member 1; MLN, Motilin; OR, Odds Ratio; SNP, Single Nucleotide Polymorphism; TCF7L2, Transcription Factor 7-Like 2; WHR, Waist-to-Hip Ratio.

[a] SNPs significantly associated with BC in present study and related to obesity (Bioinformatics tool & GWAS) previously.

[b] Unadjusted Odds Ratio fitted with WHR ≤0.84 and homozygous dominant with 1 df as reference. Estimated using multiplicative interaction logistic regression model between SNP and WHR

3.6 Discussion

This is the first study in Indian population to study large number of SNPs using large sample size with a main aim to replicate SNPs which were observed to be associated in GWAS of BC in other populations.

The present case-control study was conducted using the GoldenGate assay to evaluate 384 SNPs in about 200 candidate genes in 1204 cases and 1212 controls. The study replicated the SNPs associated with BC, which had been previously identified in GWAS and Candidate Studies. As obesity is an important trait associated with BC, the study also includes the SNPs related to obesity identified by GWAS or Bioinformatics tool.

The strength of the study is that the GGGT assay used for genotyping is more flexible method in terms of multiplexing than Taqman SNP genotyping assay [345]. All quality control steps recommended by manufacturer were accomplished successfully. It is reported to be highly accurate in humans with error rates in order of 0.3-0.4% [346]. The DNA before genotyping was quantitated using Picogreen dye. A correlation co-efficient of 98% was observed between MAF of the present study and another ongoing study in South India. The sample size

was however, moderate and the study was not powered to detect a weak association for the SNPs which are less prevalent.

3.6.1 SNPs selected from GWAS

3.6.1A Breast Cancer

Seven SNPs from BC GWAS were replicated in the present study. rs2420946, rs2981575, rs2981582 (significantly identified in additive model) and rs1219648 (significantly identified in recessive model) in intronic region of FGFR2 have shown to increase the risk of BC. SNPs in FGFR2 gene have attracted considerable attention for BC since it was first identified through genome-wide association approach [330,331]. FGFR2 is a member of a receptor tyrosine kinase gene superfamily, FGFR2 is a tumour suppressor gene that is amplified and overexpressed in 10–15% of breast tumours [347,348]. FGFR2 can transform human mammary epithelial cells [348], and inhibition of FGFR2 signalling can inhibit breast tumour cell proliferation [349].

Three meta-analyses and one case-control study showed that significantly increased BC risk was associated with rs2981582 and rs2420946 polymorphisms [247,350–352]. The association was observed in Caucasian, Asian [247], Arabic [353], African-American [354,355] and Non-Hispanic white women [356] .

Studies have indicated that rs2981582 would upregulate FGFR2 expression in BC tissues by altering Runx2 and/or C/EBP β binding affinity, thereby influencing the propensity for tumour formation [357]. However, homozygotes of minor alleles at rs2981582 were significantly correlated with decreasing FGFR2 expression level in normal breast tissue [358]. Therefore, the precise mechanism of how FGFR2 risk alleles that span the putative enhancer region within intron 2 induce FGFR2 overexpression remains to be determined. rs2981582 [247,352,356,359], rs1219648 [247,351,352,354–356,359], rs2420946 [247,351,352] have been significantly identified in many studies including case-control and GWAS across various

ethnicities. rs2981575 had the strongest association with BC risk (per allele Hazards Ratio = 1.28, 95% CI: 1.18–1.39) [360]. rs2981575 (OR = 1.25; P = 0.007) and rs2981582 (OR = 1.23, P =0.014) were significantly associated with ER+/PR+ BC. A similar association was observed in previous studies [356,361–363]. Thus FGFR2 is likely an important genetic marker contributing to susceptibility of BC in Indian population and it is recommended that these SNPs to be included in functional assays.

rs10411161 located in the intronic region of ZNF577, attenuated the risk of BC in the present study in co-dominant CC v/s TT (OR = 0.78; 95% CI: 0.63–0.98), dominant (OR = 0.81; 95% CI: 0.68–0.96) and additive model (OR = 0.88; 95% CI: 0.79–0.99). However the GWAS showed an increased association in stage I of the analysis (OR = 1.42; 95% CI: 1.22–1.65). In stage II rs10411161 showed deviation from HWE [333]. Therefore, SNP rs10411161 would require further scrutiny and validation from independent studies

rs10871290, a T to C polymorphism in the intergenic region of GLG1 (Golgi Apparatus Protein 1) was observed to be protective against the risk of BC in recessive model and genotypic model (TT v/s CC). A similar protective association was observed in haplotype association tests using GWA case-control pilot study. However, in single marker analysis, rs10871290 was not statistically significant after correction for multiple testing [364]. Nevertheless, without systematic replication, association of rs10871290 in relation to BC risk should be interpreted with caution.

rs2046210, located 180 kb upstream of the transcription initiation site of the first coding exon of the ESR1 gene, was successfully replicated and associated with an increased risk of BC in the additive and dominant model. This association was first reported by Zheng et al. [365]. However, several subsequent replication studies could not reach consistent results; for example, Stacey et al. [366] failed to validate the association in Europeans, and similarly, Campa et al. [361] were also unable to replicate the findings in Asians. A meta-analysis

showed that the polymorphism had a larger effect on Asians than on Europeans or Africans [367], whereas another recent meta-analysis showed significant association in Asians and Europeans and not in Africans [368]. Potential explanations for the discrepancy could be the modest effect of this SNP and the diverse genetic backgrounds of the different ethnic groups.

Considering the relative vicinity of rs2046210 to the ESR1 gene, it was speculated that the SNP itself or causal variants in LD with it might alter ESR1 gene expression, thus affecting the susceptibility to BC. However, the functional genomic analyses and in vitro functional experiments conducted by Cai et. al [369] provided no support for the potential involvement of this polymorphism in the regulation of ESR1. The function of this SNP therefore is still unclear; future fine-mapping of the BC susceptibility loci tagged by rs2046210 is warranted and the underlying biological mechanism of this polymorphism still needs further investigation. On stratification, rs2046210 showed an increased association with ER-/PR- and TNBC and in postmenopausal women. This association is consistent with literature suggested that rs2046210 tended to increase BC risk in ER- tumours by a greater magnitude compared to ER+ tumours [368,370,371]. The association of rs2046210 with TNBC and not with hormone receptor tumours was also observed previously in many studies [372,373].

TNBC is defined in part by the absence of expression of ERs, it can be speculated that inherited variation may downregulate ESR1 expression and promote formation of ERα -negative tumours. However, studies in mice have shown that the mammary stem cell compartment can be regulated by 17β-estradiol and progesterone through a paracrine-signaling mechanism from steroid receptor positive luminal cells to steroid receptor − negative stem [354,355]. Thus, SNPs in the ESR1 locus may promote expansion of receptor-negative precursors and subsequent development of triple-negative tumours.

rs6556756 located in the intergenic region of Methionine Adenosyltransferase II, Beta (MAT2B) showed a protective association in TT v/s GG genotype (OR = 0.50; 95% CI: 0.25–0.97) which was successfully replicated as suggested in previous reports [374,375].

The minor allele (G) frequency of rs865686 (MAF=0.13) obtained in the study was comparable to the Asians (MAF =0.09) which warrants the similarity with the sub-ethnicities of Asia [0.12 for controls from the Indian subcontinent (N = 99); 0.09 for controls from South East Asia (N = 660); P = 0.21]. The MAF was significantly higher among European (MAF = 0.38) as compared to the Asians [376].

The association of rs865686 was replicated in the present study using the co-dominant, dominant and additive model which suggests a strong evidence of inverse association. Large scale GWAS studies have warranted the association of rs865686 at 9q31.2 with risk of BC [244,376,377]. The mechanism behind this association is not known, however, it could be attributed to the fact that rs865686 (9q31.2) lies more than 600 kb from the nearest genes, KLF4 and RAD23B which are both attractive candidates for mediating an effect on BC risk. Recent functional studies of rs6983267 (a colorectal cancer risk locus mapping to 8q24.21.) have shown that physical interaction between a causal variant and its target (the MYC proto-oncogene) can occur over a large distance (~335 kb) [378,379].

The common variants rs889312 (P = 0.011) and rs16886165 (P = 0.003) using the additive model in MAP3K1 increased the risk of BC. These SNPs lie in a LD block of approximately 280 kb which includes MAP3K1 gene [352]. The MAP3K1 gene encodes a 196-kDa serine/threonine protein kinase that activates the extracellular signal regulated kinase (ERK), c-Jun NH2-terminal kinase (JNK) and nuclear factor-kB (NF-kB) pathways [380]. The downstream signal transductions regulate the survival, differentiation, proliferation and apoptosis of cell, and appear to be involved in tumour development and tumour progression [381–383]. The causal variant may be closer to rs16886165 than rs889312 [362] thus showing

a positive association with BC. These 2 SNPs may have effects on the modulation of MAP3K1 expression and therefore the tuning of MAPK signal transductions. The SNP rs889312 in MAP3K1 was identified to be associated with BC risk by GWAS [330], with confirmation of the association in European ancestry population by another study [384].

Previous studies suggested that rs889312 in MAP3K1 gene was related to ER+ BCs, which was confirmed by stratified analysis by ER/PR in present study [385,386]. The association of rs889312 significantly increased the risk of postmenopausal BC but not premenopausal BC.

A recent meta-analysis demonstrated that the rs889312 and rs16886165 SNPs in MAP3K1 were associated with increased BC susceptibility. When stratified by ethnicity, the rs889312-C allele showed to be a risk factor for the development of BC in European and Asian ancestry populations, but not in Africans [386]. rs16886165 have been identified as a low-penetrance risk factor for BC in European ancestry population by GWAS [245,386]. The pooled result of the recent meta-analysis indicated that the rs16886165-G allele conferred BC risk in Asian and African ancestry population [386].

Forty SNPs out of 47 included in the final analysis replicated from BC GWAS weren't associated in the present study, as the study was not powered to detect the association of SNPs with prevalence below 20% (N=6). The SNPs (N=25) which had MAF ≥ 20% and still did not show association implied that such SNPs may not be a risk factor in the study population. Three SNPs (rs11249433, rs6504950 and rs999737) which were previously reported to have MAF ≥ 20% but below 20% in the present study suggests that their attributable risk is low and their contribution to the disease development would be less important due to low prevalence.

3.6.1B Other-traits

Loci rs10980926 (Recessive model) and rs6575793 (Dominant model) which have been associated with age at menarche previously [387] have shown to be associated with modest

increase in BC risk. Further a Chinese study evaluated the SNPs associated with age at menarche identified in European population in BC cases and found an increased association which comprised of approx. 2000 BC cases [388].

The SNPs which had been associated with BMI and obesity in GWAS have shown an association with BC susceptibility in the present study. BC risk is largely driven by obesity [112,139,149–153]. Many studies have shown that body weight and obesity are strongly influenced by genetic factors, with heritability estimates in the range of 65–80% [389]. Genetic variants in several genes are known to influence BMI, but these mutations are rare and often cause severe monogenic syndromes with obesity [390]. Large-scale meta-analysis of multiple GWAS identified additional genes harboring common SNPs that associate with BMI [391–394]. GWASs have also found associations with measures of body fat distribution [391,395].

Loci rs2287019 (additive) [396–401], rs2867125 (recessive), rs2922763 (recessive) and rs987237 (additive, dominant, genotypic) which have been previously observed as a risk factor for BMI [396–401] have increased the risk of BC in the present study. Other obesity SNPs such as rs2116830 (additive), rs374748 (recessive) and rs988712 recessive and genotypic) [402,403] also showed an increased risk, whereas rs10953454 (recessive), rs16867321 (additive, dominant), rs2274459 (additive, dominant) and rs925642 (recessive) [350,402] showed an inverse relation with BC.

The mechanism of obesity/BMI SNPs in the BC susceptibility is not known however the genes in which these belong or are near the gene can help in elucidating their action. Transmembrane Protein 18 (TMEM18) gene which may be participating in the appetite signalling system [397] partly explain the risk associated with rs2867125 and rs12990777. KCNMA1 channels enhances proliferation of human pre-adipocytes in vitro [404]. It is intriguing since it has been shown that there is a high rate of adipocyte turnover in vivo; with

about 10% of fat cells being renewed annually. Furthermore, adipocyte number is a major determinant for the fat mass in adults [405]. Thus, KCNMA1 (rs2116830) could hypothetically contribute to obesity and hence BC. The Brain-Derived Neurotrophic Factor (BDNF) (rs988712) is involved in catabolic pathway [406].

3.6.2 SNPs selected from Candidate Studies

3.6.2A Breast Cancer

rs1056836 (Val432Leu) located in the Untranslated Region (UTR) of Cytochrome P450 subfamily I dioxin-inducible polypeptide 1 (CYP1B1) with a C to G polymorphism had a prevalence of 0.21. The risk increased to 1.58 (95% CI: 1.07–2.34) for CC to GG variant in the present study. The change in amino acid from valine to leucine has shown to increase the activity of the CYP1B1 enzyme on a variety of substrates, including procarcinogens and gonadal steroid hormones [407]. The increased formation of 4-hydroxyestrone induced by the CYP1B1 enzyme could be a possible risk factor for BC [408]. Paracchini et.al showed [409] that women with the leucine substitution had higher 2-/16-hydroxyestrone metabolites than women carrying the other CYP1B1 genotypes, which implies the association of rs1056836 in the study population. This increased risk was also observed in African-American population, however, no significant association could be estimated for Asian population in other study [410].

rs1641535 and rs1641536 in the promoter region of ATPase, Na+/K+ transporting, beta 2 polypeptide (ATP1B2) showed a slight increased risk of BC in additive model [OR = 1.16; 95% CI: 1.008–1.33 and OR = 1.15; 95% CI: 1.01–1.33 respectively]. The 2 SNPs reported to be in strong LD have shown an inverse association which could be attributed to the different minor allele observed in the present study and the case-control study conducted in Norwegian and Polish population [411].

Allele frequency of rs1695 was 26% which is almost comparable with the previously reported studies in Asian population. A protective association was observed in heterozygous mutants in the present study. An A to G polymorphism at nucleotide 313 in the Glutathione S-Transferase Pi (GSTP1) gene leads to an amino acid change (*Ile105Val,* rs1695). However a meta-analysis showed no significant association [412]. A pooled analysis also did not show any association between polymorphism and risk of BC [413]. This suggests that the association of SNPs might be different in different populations.

An increased risk was observed in homozygous variant (P = 0.008) and recessive model (P = 0.008) in rs2070744 in the promoter region of NOS3 gene A meta-analysis of 3 studies on Asian population have shown reduced risk of BC which is inconsistent to the results of present study [414]. However largely the results from various case-control studies have been inconclusive suggesting that the association of rs2070744 needs further confirmation from large genome wide analysis.

The SNP rs2287499 in WRAP53 showed a protective association for homozygous mutants (P = 0.027) and in recessive model (P = 0.01). This is inconsistent with results from a case-control study on Norwegian and Polish population which showed a different association among ER+/PR+ (OR = 1.02; 95%CI: 0.87–1.19) and ER-/PR- (OR = 1.42; CI: 1.18–1.71) tumours [411]. Therefore, possibility of chance could not be excluded in the association of rs2287499 and BC risk

An increased risk (OR = 1.17; 95% CI: 1.01–1.35) was observed in the additive model for rs3218408 in intronic region of X-ray Repair Complementing defective repair in Chinese hamster cells 2 (XRCC2). The mechanism through which the polymorphism in XRCC2 alters the risk of BC is that XRCC2-deficient cells [415] show a greater than 100 fold reduction in homologous recombination repair compared to XRCC2-proficient cells [416–418]. As meta-

analysis also showed an overall recessive OR of 1.33 (95% CI: 1.12–1.57) [419], the association observed in the present study could be real.

The results of the present study indicate an increase in association with rs5275 in Cyclooxygenase-2 (COX-2) (OR = 1.20; 95% CI: 1.008– 1.43; P = 0.040) for genotype TT v/s CT. rs5275 located in the UTR of COX-2 could influence the risk of BC, through altering the levels of expression or activity of the Prostaglandin-endoperoxide Synthase 2 (PTGS2) enzyme, which is responsible for transforming arachidonic acid into prostaglandins. There are conflicting results in the literature with respect to the role of rs5275 in which one study has shown positive association [420] and others have shown no association [421–423]. There were some studies which indicated that women homozygous for the rs5275 C allele have a 20% lower risk of BC than those homozygous for the T allele (OR = 0.80, 95% CI: 0.66–0.97) [424]. This reduced risk was confirmed by Zhu et al. [425] in a meta-analysis. Taken together, these studies appear to suggest no strong influence of rs5275 SNP on BC risk. The present work indicates that variant in the 3' UTR of COX-2 do not appear to greatly influence BC risk, as the apparent risk association found for rs5275 SNP was limited to heterozygotes with a low OR value and borderline significance. However, the apparently negative results do not exclude potential low risks (i.e., OR < 1.5), whose detection with high level of statistical significance (P < 0.001) would require large individual studies or meta-analysis (N > 6000).

The present study showed an increased association of BC with rs861539 (OR = 1.16; P = 0.044) in the additive model. rs861539, a Thr241Met substitution is the most thoroughly investigated polymorphism in X-ray Repair Complementing defective repair in Chinese hamster cells 3 (XRCC3) gene due to a (C to T) transition at exon 7 codon 241. A modest association between the homozygous variant genotype of the T241M allele of *XRCC3* and BC risk was first reported in a study in the United Kingdom [426]; however, most subsequent studies in Caucasian populations [427–433] have not confirmed this association. Three meta-

analyses of published data suggested a very small increase in risk among women homozygous for the methionine allele [434–436].

In contrast, a review which included 5 studies on XRCC3 polymorphisms and BC association consortium which included participants from Europe, the United States, Australia and Asia did not show any association with BC risk [437,438]. A recent meta-analysis confirmed that the T allele was associated with elevated BC risk mainly following a recessive model (pooled OR = 1.064, 95% CI: 1.007–1.124, fixed effects), given that the effect was more pronounced in homozygous carriers (pooled OR = 1.073, 95% CI: 1.010–1.140, fixed effects) [439]. In conclusion, the present study supports the fact that the XRCC3 could not be a major increased risk factor for BC but it might represent a low-penetrance susceptible gene.

3.6.2B Other SNPs

The present study had also performed genotyping on the SNPs which had been suggested by the Collaborators on the basis of animal experiments. These SNPs haven't been studied in epidemiologic studies and the causality of their association with risk of BC should be verified using large scale studies.

rs10489525 in the Cold Shock Domain Containing E1, RNA-binding (CSDE1) gene showed an increased risk of BC in heterozygous mutants and in dominant model (P = 0.033 and P = 0.023 respectively). rs2233660 located in the intergenic region of Prohibitin (PHB) gene had also shown an increased risk in 3 of the 4 models fitted (TT v/s CT, P = 0.018). PHB is evolutionarily conserved, and its product is proposed to play a role in human cellular senescence and tumour suppression. Antiproliferative activity is reported to be localized to the 3' UTR, which is proposed to function as a trans-acting regulatory Ribose Nucleic Acid (RNA). rs2240123 (P = 0.044; Recessive model) in Chromobox protein homolog 1 (CBX1); rs274586 (P = 0.029; Recessive model) and rs489990 (P = 0.036; AA v/s GG) located in

intronic region of Troponin I Type 3 Interacting Kinase (TNNI3K) [440] have shown to be associated with BC risk.

3.6.3 SNPs selected using Bioinformatics Tool

The total number of SNPs selected using this criterion were 127. The findings of the present study suggest an association of SNPs using four different models namely genotypic, additive, recessive and dominant in 12 obesity genes with BC namely E26 transformation-specific variant 5 (ETV5), Mitochondrial Carrier Homolog 2 (MTHFR), TCF7L2, Melanocortin 4 Receptor (MC4R), TMEM18, Interleukin 6 (IL6), Glucoseamine-6-Phosphate Deaminase 2 (GNPDA2), INSIG2, Adiponectin, C1Q and Collagen Domain Containing (ADIPOQ), Leptin (LEP), ACE and Angiotensinogen (AGT). The SNPs were identified for the first time in the BC etiology and hence need further validation, particularly as they were not significant after correction for multiple comparisons. The SNPs in genes MTHFR (rs11121832; P = 0.007; CC v/s CT), TMEM18 (rs12990777; P = 0.030; AA v/s AG), IL6 (rs1474347; P = 0.030; Additive model), INSIG2 (rs2161829; P = 0.007; CC v/s TT), ADIPOQ (rs3774261; P = 0.037; Additive model) LEP (rs4236625; P = 0.020; AA v/s TT) and AGT (rs7079; P = 0.048; CC v/s AA) were positively associated with risk of BC whereas the SNPs in TCF7L2 (rs11196219; P = 0.042 GG v/s AG and rs11594610; P = 0.001; GG v/s AG), MC4R (rs11872992; P = 0.050; Additive model), GNPDA2 (rs16857402; P = 0.033; TT v/s CT), ACE gene (rs4293; P = 0.031; Additive model and rs4362; P = 0.006; CC v/s TT) and ETV5 (rs9831938; P = 0.036; CC v/s TT) reduced the risk of BC.

rs2161829 present in the intronic region of INSIG2 which has been functionally linked to obesity due to its role in cholesterol and fatty acid synthesis feedback inhibition [441,442].

rs3774261 in ADIPOQ identified as a risk factor of BC in present study, may play a regulatory role in the expression of metabolic traits in obesity-associated chronic disease [443,444]. rs3774261 is known to strongly associate with serum adiponectin level [445,446].

Introns are non-coding regions of a gene; however, there is evidence that introns of the protein-coding gene transcripts can affect gene expression by repressing translation or cleaving RNA transcripts [447]. In particular, rs3774261 is an intronic enhancer, and thus could affect protein levels via enhancing transcription and thus increasing the risk of BC. No association studies have been conducted on SNPs selected from Bioinformatics tool and risk of BC suggesting that these associations can more likely be due to chance. Therefore more epidemiologic studies are required to interpret the results of observed associations between obesity genes and BC risk.

3.6.4 Gene-Environment Interaction

Increase in the point estimates for SNPs with increase in WHR indicates that independent association observed for these SNPs might not be false positive. However whether the joint effect of SNPs and WHR is more (or less) than multiplicative needs to be replicated with larger sample size as multiplicative interaction was not significant.

3.7 Summary

The present study could replicate 7 SNPs from BC GWAS. Five SNPs were successfully replicated out of 42 Candidate SNPs of BC used for analysis. SNPs selected from FGFR2 gene were positively associated with BC. 25 SNPs which were identified as a risk factor for BC in the GWAS conducted in other populations did not replicate in the Indian population, even though their prevalence was high (≥ 20%) indicating that they may not be a risk factor in Indian population. The 3 SNPs which were highly prevalent in GWAS population could not be replicated even if they were associated the attributable risk of the SNPs remains low due to their low prevalence in the present study population. The SNPs associated with ER+/PR+ and ER-/PR- BCs were observed to be different suggesting that the cancer stratified on hormone receptor status may differ due to different biological pathways.

The genetic susceptibility of SNPs associated with BMI and obesity in various GWAS were

associated with BC risk suggesting that BC is mainly driven by genes related to obesity. Fourteen SNPs selected using Bioinformatics tool from candidate genes associated with obesity were first time identified to be strongly associated with the BC risk and hence need further validation.

Chapter 4

Conclusion & Future Perspectives

4.1 Conclusion

The BC incidence is increasing in India and recent estimates suggest that it is the most common malignancy among females. The incidence of BC in rural areas is reported to be less than half of that observed in urban areas. The current study was therefore undertaken to understand the lifestyle and genetic factors with a focus to identify the reasons for differences in rural and urban regions. As the risk factors differ in pre- and postmenopausal women, a focus was to identify the similarity and differences in risk factors by menopausal status.

The major highlights of the work are as follows (Table 4.1):

The study enrolled 1637 cases and 1515 controls and data was obtained on reproductive factors and anthropometric measurements among other important lifestyle related variables.

The lifestyle related risk factors associated with BC after adjusting for confounding variables are as follows:

1. The first twenty years of life spent in rural area is protective for BC after adjustement for well known risk factors including age, region of residence, education, height, WHR, age at first full-term pregnancy and menopausal status. This suggests that exposures in early life may be more important in the development of BC compared to current exposures as current residence does not show any protection after controlling for important risk factors.
2. This is the first study to observe the prevalence of HR status in women who were currently residing, lived for first twenty years and never lived in rural areas. Prevalence of TNBC is higher in rural area (44.21%) as compared to urban area (34.39%).
3. For every 2 year increase in the age at first full-term pregnancy there is a 10% increase in risk of BC.
4. Longer duration of current OC use increases the risk of BC cases in urban women.
5. No association could be observed with breastfeeding possibly due to homogeneity in the

study population.

6. Spontaneous abortion showed a protective association against the development of BC which could have attributed to the protection acquired from pregnancy.
7. Central obesity measured by WC and WHR is more important than BMI in increasing risk of pre- and postmenopausal BC in India women.
8. Increase in body size from age 10 to age 20 increases the risk of BC indicating accelerated growth in teenage years play an important role in BC etiology.
9. Low BMI increases the risk of both pre- and postmenopausal BC. This new finding from the study needs further replication.
10. For every 5 cm increase in height there is an increase of premenopausal BC with OR = 1.24 (95% CI: 1.12–1.37).

The study thus identified the reasons for differences in BC in rural and urban India. The main reasons being low prevalence of following risk factors which were identified as strong predictors in the present study

1. The mean age at first full-term pregnancy in rural areas was 20.59 whereas it was 22.68 in urban areas ($P < 0.0001$)
2. Mean WC was 79.50 cm in rural areas and in urban areas it was found to be 83.85cm ($P < 0.0001$).
3. WHR had a mean of 0.84 in urban women and 0.83 in rural women ($P = 0.0016$).
4. BMI in urban women who had attained menopause $\geq$ 10 years ago had a mean of 25.83kg/m^2 whereas the mean was 24.42 kg/m^2 in rural women ($P = 0.0002$).

The attributable risk for developing BC was estimated to be 9% and 2% for urban and rural women respectively if first full-term pregnancy occurs after age 25. Similarly the attributable risk for developing BC for rural and urban women with WHR $\geq$0.95 respectively. In addition to lifestyle factors, genetic susceptibility also plays an important role in development of BC.

Therefore a large scale case-control study was conducted, a first in India, which attempts to replicate BC GWAS SNPs identified in the developed countries.

The susceptibility of genetic factors in Indian population was estimated using GoldenGate genotyping assay performed on 1204 cases and 1212 controls for 384 SNPs.

7 SNPs which were identified in GWAS on BC of other population were replicated in the present study while 25 SNPs were not identified as a risk factor even though they had a high prevalence (> 20%). SNPs in FGFR2 and MAP3K1 showed a strong positive association with BC risk. The association of other SNPs identified in BC GWAS could not be confirmed because of low prevalence of these SNPs in Indian population (N=15). The findings show that common genetic variants influence the pathological subtype of BC and provide further support for the hypothesis that ER+/PR+ and ER-/PR- disease are biologically distinct.

Previously identified BC Candidate SNPs (N=5) were confirmed as risk factor for BC. rs1056836 (P = 0.019, CC v/s GG), rs2287499 (P = 0.027; CC v/s GG) are some of the SNPs identified from Candidate Studies. The study was successful in identifying new SNPs in obesity genes identified using Bioinformatics tool, however their association would have to be replicated in other studies particularly because they were not significant after adjusting for multiple comparisons. This also indicates that the lifestyle factors are more important than the genetic markers from public health point of view.

Table 4.1: Risk Factor Summary of Breast Cancer derived from Present Study

Risk Factor	Direction of Effect
Association possibly causal	
Age at first full-term pregnancy >25 years versus, <20	↑↑
Waist-to-hip ratio	↑↑
Tallness (Premenopausal)	↑↑
Large Waist Circumference	↑↑
High Body mass index (Postmenopausal-after 10 years of attaining menopause)	↑
High Body mass index (Premenopausal)	↓
Spontaneous abortion	↓
Current Oral contraceptive use	↑
BC GWAS SNPs replicated in genes ESR1, FGFR2, MAP3K1 (N=5)	↑
BC GWAS SNPs replicated in genes ZNF577, 9q31.2 (N=2)	↓
BC Candidate SNPs in genes ATP1B2, FGFR2, XRCC2, XRCC3 (N=5)	↑
Weak association possibly causal	
Large Hip Circumference	↓
Increased number of full-term pregnancies	↓
Increased duration between menarche and first full-term pregnancy	↑
Association difficult to interpret causality	
Low BMI	↑
Induced abortion	↑
Increased duration since last birth	↓
Twin Pregnancy	↑
Obesity SNPs identified using Bioinformatics tool ($p \leq 0.03$) in genes IL6, INSIG2, MTHFR (N=3)	↑
Obesity SNPs identified using Bioinformatics tool ($p \leq 0.03$) in genes ACE, TCF7L2 (N=3)	↓
No association	
Menarche at <12 years versus, >14	-
Breastfeeding	-
BC GWAS SNPs not replicated (N=40)	-

Abbreviations: ACE, Angiotensin I-Converting Enzymee; ATP1B2, ATPase, Na+/K+ transporting, beta 2 polypeptide; BC, Breast Cancer; COX-2, Cyclooxygenase-2; CYP1B1, Cytochrome P450 subfamily I dioxin-inducible polypeptide 1; ESR1,Estrgen Receptor alpha; FGFR2, Fibroblast Growth Factor Receptor 2, GSTP1, Glutathione S-Transferase Pi; IL6, Interleukin 6; INSIG2, Insulin-Induced Gene 2; LEP, Leptin, MAP3K1, Mitogen-Activated Kinase Kinase Kinase 1; MAT2B, Methionine Adenosyltransferase II, Beta; MTHFR, Methylenetetrahydrofolate Reductase; TCF7L2, Transcription Factor 7-Like 2; TMEM18, Transmembrane Protein 18; WDR79, WD repeat containing, antisense to TP53; XRCC2, X-Ray Repair Cross Complimenting gene 2; XRCC3, X-Ray Repair Cross Complimenting Defective gene 3. ↑↑: Moderate to large extent in risk, ↑: Slight increase in risk, ↓: moderate to large decreases in risk ,-: no association

4.2 Future Perspectives

The study demonstrates that BC is preventable in India and is possible to reverse the increasing trend of BC. As living in a rural area protects from development of BC, public health authorities should spread the message about harmful effects of increase in central obesity and later age at first child-birth - factors that are less prevalent in rural areas. The strategies to reduce BMI should consider the cut-off of 18.5 kg/m^2 as BMI lower than this might increase the risk for pre- and postmenopausal BC. There is urgency to develop policies and the necessary infrastructure for early detection and improved medical care before BC reaches the "epidemic" proportions seen in many high-resource settings.

Future studies on BC should include much more sophisticated measurement of central obesity and total body fat. Studies should also focus on role of nutrition and accelerated growth in teenagers and risk of BC. Given the role of obesity in BC etiology, possible role of inflammatory markers can also be studied.

A relatively high proportion of TNBC represents an important feature of the study. With the difference in receptor status in rural and urban population, an obvious analysis on various lifestyle related factors stratified on receptor status should be followed to understand the etiologic differences which may operate differently across strata.

Prevalence of certain SNPs are different in Indian population as compared to the West. Large-scale GWAS are thus imperative as they could identify new loci for Indian population. Fourteen SNPs which were identified for the first time need replication from other Indian studies. Deep sequencing would be helpful to better understand the mechanism and identification of new loci in FGFR2 and MAP3K1.

Bibliography

[1] Getting Your Pathology Report. Breastcancer.org n.d. http://www.breastcancer.org/symptoms/diagnosis/getting_path_report (accessed September 17, 2014).
[2] Harris JR, Lippman ME, Osborne CK, Morrow M. Diseases of the Breast. Lippincott Williams & Wilkins; 2012.
[3] Breast Anatomy. Univ Conn Health Cent n.d. http://fitsweb.uchc.edu/student/selectives/Luzietti/Breast_anatomy.htm.
[4] Mirza AN, Mirza NQ, Vlastos G, Singletary SE. Prognostic factors in node-negative breast cancer: a review of studies with sample size more than 200 and follow-up more than 5 years. Ann Surg 2002;235:10.
[5] Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. The Oncologist 2004;9:606–16.
[6] CTSU RI. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365:1687–717.
[7] Fraser J. A study of the malignant breast by whole section and key block section methods. Surg Gynecol Obstet 1927;45:64.
[8] Li CI, Anderson BO, Daling JR, Moe RE. Trends in incidence rates of invasive lobular and ductal breast carcinoma. Jama 2003;289:1421–4.
[9] Arpino G, Bardou VJ, Clark GM, Elledge RM. Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. Breast Cancer Res 2004;6:R149–56.
[10] Li CI, Moe RE, Daling JR. Risk of mortality by histologic type of breast cancer among women aged 50 to 79 years. Arch Intern Med 2003;163:2149–53.
[11] Goldstein NS, Kestin LL, Vicini FA. Refined morphologic criteria for tubular carcinoma to retain its favorable outcome status in contemporary breast carcinoma patients. Am J Clin Pathol 2004;122:728–39.
[12] Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. N Engl J Med 2006;354:270–82.
[13] Allred DC, Brown P, Medina D. The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer. Breast Cancer Res 2004;6:240–57.
[14] Anderson WF, Jatoi I, Devesa SS. Distinct breast cancer incidence and prognostic patterns in the NCI's SEER program: suggesting a possible link between etiology and outcome. Breast Cancer Res Treat 2005;90:127–37.
[15] Kastner P, Bocquel M-T, Turcotte B, Garnier J-M, Horwitz KB, Chambon P, et al. Transient expression of human and chicken progesterone receptors does not support alternative translational initiation from a single mRNA as the mechanism generating two receptor isoforms. J Biol Chem 1990;265:12163–7.
[16] Mote PA, Bartow S, Tran N, Clarke CL. Loss of co-ordinate expression of progesterone receptors A and B is an early event in breast carcinogenesis. Breast Cancer Res Treat 2002;72:163–72.
[17] Chebil G, Bendahl P-O, Ferno M. Estrogen and Progesterone Receptor Assay in Paraffin-Embedded Breast Cancer 2003.

[18] Cui X, Schiff R, Arpino G, Osborne CK, Lee AV. Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. J Clin Oncol 2005;23:7721–35.
[19] Osborne CK, Schiff R, Arpino G, Lee AS, Hilsenbeck VG. Endocrine responsiveness: understanding how progesterone receptor can be used to select endocrine therapy. The Breast 2005;14:458–65.
[20] Anderson WF, Chen BE, Jatoi I, Rosenberg PS. Effects of estrogen receptor expression and histopathology on annual hazard rates of death from breast cancer. Breast Cancer Res Treat 2006;100:121–6.
[21] Bardou V-J, Arpino G, Elledge RM, Osborne CK, Clark GM. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. J Clin Oncol 2003;21:1973–9.
[22] Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama 2006;295:2492–502.
[23] Voduc KD, Cheang MC, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol 2010;28:1684–91.
[24] Metzger-Filho O, Sun Z, Viale G, Price KN, Crivellari D, Snyder RD, et al. Patterns of Recurrence and Outcome According to Breast Cancer Subtypes in Lymph Node–Negative Disease: Results From International Breast Cancer Study Group Trials VIII and IX. J Clin Oncol 2013:JCO – 2012.
[25] Lund MJ, Butler EN, Hair BY, Ward KC, Andrews JH, Oprea-Ilies G, et al. Age/race differences in HER2 testing and in incidence rates for breast cancer triple subtypes. Cancer 2010;116:2549–59.
[26] Haque R, Ahmed SA, Inzhakova G, Shi J, Avila C, Polikoff J, et al. Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. Cancer Epidemiol Biomarkers Prev 2012;21:1848–55.
[27] Turner NC, Reis-Filho JS. Basal-like breast cancer and the BRCA1 phenotype. Oncogene 2006;25:5846–53.
[28] Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev 2007;16:439–43.
[29] Atchley DP, Albarracin CT, Lopez A, Valero V, Amos CI, Gonzalez-Angulo AM, et al. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. J Clin Oncol 2008;26:4282–8.
[30] Hartman A-R, Kaldate RR, Sailer LM, Painter L, Grier CE, Endsley RR, et al. Prevalence of BRCA mutations in an unselected population of triple-negative breast cancer. Cancer 2012;118:2787–95.
[31] Millar EK, Graham PH, O'Toole SA, McNeil CM, Browne L, Morey AL, et al. Prediction of local recurrence, distant metastases, and death after breast-conserving therapy in early-stage invasive breast cancer using a five-biomarker panel. J Clin Oncol 2009;27:4701–8.

[32] Kim M-J, Ro JY, Ahn S-H, Kim HH, Kim S-B, Gong G. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. Hum Pathol 2006;37:1217–26.
[33] Millikan RC, Newman B, Tse C-K, Moorman PG, Conway K, Smith LV, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109:123–39.
[34] Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst 2011;103:250–63.
[35] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69–90.
[36] Bray F, Ren J-S, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer 2013;132:1133–45. doi:10.1002/ijc.27711.
[37] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11. Http://globocan.iarc.fr 2013. http://globocan.iarc.fr (accessed May 22, 2014).
[38] Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. Eur J Cancer Oxf Engl 1990 2001;37 Suppl 8:S4–66.
[39] Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. Cancer Epidemiol 2012;36:237–48.
[40] Brenner H, Francisci S, De Angelis R, Marcos-Gragera R, Verdecchia A, Gatta G, et al. Long-term survival expectations of cancer patients in Europe in 2000–2002. Eur J Cancer 2009;45:1028–41.
[41] Coleman MP, Quaresma M, Berrino F, Lutz J-M, De Angelis R, Capocaccia R, et al. Cancer survival in five continents: a worldwide population-based study (CONCORD). Lancet Oncol 2008;9:730–56.
[42] Hortobagyi GN, de la Garza Salazar J, Pritchard K, Amadori D, Haidinger R, Hudis CA, et al. The global breast cancer burden: variations in epidemiology and survival. Clin Breast Cancer 2005;6:391–401.
[43] Sankaranarayanan R, Swaminathan R, Brenner H, Chen K, Chia KS, Chen JG, et al. Cancer survival in Africa, Asia, and Central America: a population-based study. Lancet Oncol 2010;11:165–73.
[44] El Saghir NS, Adebamowo CA, Anderson BO, Carlson RW, Bird PA, Corbex M, et al. Breast cancer management in low resource countries (LRCs): consensus statement from the Breast Health Global Initiative. The Breast 2011;20:S3–11.
[45] Shulman LN, Willett W, Sievers A, Knaul FM. Breast cancer in developing countries: opportunities for improved survival. J Oncol 2010;2010.
[46] Mathers C, Fat DM, Boerma JT. The global burden of disease: 2004 update. World Health Organization; 2008.
[47] Dikshit R, Gupta PC, Ramasundarahettige C, Gajalakshmi V, Aleksandrowicz L, Badwe R, et al. Cancer mortality in India: a nationally representative survey. The Lancet 2012;379:1807–16. doi:10.1016/S0140-6736(12)60358-4.

[48] Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev 2010;19:1893–907.
[49] Shin H-R, Boniol M, Joubert C, Hery C, Haukka J, Autier P, et al. Secular trends in breast cancer mortality in five East Asian populations: Hong Kong, Japan, Korea, Singapore and Taiwan. Cancer Sci 2010;101:1241–6.
[50] Autier P, Boniol M, LaVecchia C, Vatten L, Gavin A, Héry C, et al. Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. BMJ 2010;341.
[51] Smigal C, Jemal A, Ward E, Cokkinides V, Smith R, Howe HL, et al. Trends in breast cancer by race and ethnicity: update 2006. CA Cancer J Clin 2006;56:168–83.
[52] Botha JL, Bray F, Sankila R, Parkin DM. Breast cancer incidence and mortality trends in 16 European countries. Eur J Cancer 2003;39:1718–29.
[53] Althuis MD, Dozier JM, Anderson WF, Devesa SS, Brinton LA. Global trends in breast cancer incidence and mortality 1973–1997. Int J Epidemiol 2005;34:405–12.
[54] Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, et al. The decrease in breast-cancer incidence in 2003 in the United States. N Engl J Med 2007;356:1670–4.
[55] Li CI, Daling JR. Changes in breast cancer incidence rates in the United States by histologic subtype and race/ethnicity, 1995 to 2004. Cancer Epidemiol Biomarkers Prev 2007;16:2773–80.
[56] Dikshit RP, Yeole BB, Nagrani R, Dhillon P, Badwe R, Bray F. Increase in breast cancer incidence among older women in Mumbai: 30-Year trends and predictions to 2025. Cancer Epidemiol 2012;36:e215–20.
[57] Linos E, Spanos D, Rosner BA, Linos K, Hesketh T, Qu JD, et al. Effects of reproductive and demographic changes on breast cancer incidence in China: a modeling analysis. J Natl Cancer Inst 2008;100:1352–60.
[58] National Cancer Registry Programme. Three years Report of Population Based Cancer Registries 2009-11. New Delhi: Indian Council of Medical Research; 2013.
[59] National Family Health Survey India (NFHS-3): 2005-06. Maharashtra, India.: International Institute for Population Sciences (IIPS); 2008.
[60] Carmichael AR, Bates T. Obesity and breast cancer: a review of the literature. The Breast 2004;13:85–92.
[61] Lahmann PH, Hoffmann K, Allen N, Van Gils CH, Khaw K-T, Tehard B, et al. Body size and breast cancer risk: findings from the European Prospective Investigation into Cancer And Nutrition (EPIC). Int J Cancer 2004;111:762–71.
[62] Dhillon PK, Yeole BB, Dikshit R, Kurkure AP, Bray F. Trends in breast, ovarian and cervical cancer incidence in Mumbai, India over a 30-year period, 1976–2005: an age–period–cohort analysis. Br J Cancer 2011;105:723–30.
[63] Nagrani R, Budukh A, Koyande S, Panse N, Mhatre S, Badwe RA. Rural Urban differences in Breast Cancer in India. Indian J Cancer 2014.
[64] Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol 2012;13:1141–51. doi:10.1016/S1470-2045(12)70425-4.

[65] Clavel-Chapelon F. Cumulative number of menstrual cycles and breast cancer risk: results from the E3N cohort study of French women. Cancer Causes Control 2002;13:831–8.
[66] Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. Am J Epidemiol 2000;152:950–64.
[67] Horn J, \AAsvold BO, Opdahl S, Tretli S, Vatten LJ. Reproductive factors and the risk of breast cancer in old age: a Norwegian cohort study. Breast Cancer Res Treat 2013;139:237–43.
[68] Lacey JV, Kreimer AR, Buys SS, Marcus PM, Chang S-C, Leitzmann MF, et al. Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort. BMC Cancer 2009;9:84.
[69] Mertens AJ, Sweeney C, Shahar E, Rosamond WD, Folsom AR. Physical activity and breast cancer incidence in middle-aged women: a prospective cohort study. Breast Cancer Res Treat 2006;97:209–14.
[70] Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. Breast Cancer Res BCR 2006;8:R43. doi:10.1186/bcr1525.
[71] Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet 2002;360:187–95. doi:10.1016/S0140-6736(02)09454-0.
[72] Ewertz M, Duffy SW, Adami HO, Kvåle G, Lund E, Meirik O, et al. Age at first birth, parity and risk of breast cancer: a meta-analysis of 8 studies from the Nordic countries. Int J Cancer J Int Cancer 1990;46:597–603.
[73] Kim HS, Woo OH, Park KH, Woo SU, Yang DS, Kim A-R, et al. The relationship between twin births and maternal risk of breast cancer: a meta-analysis. Breast Cancer Res Treat 2012;131:671–7. doi:10.1007/s10549-011-1779-5.
[74] Lee SH, Akuete K, Fulton J, Chelmow D, Chung MA, Cady B. An increased risk of breast cancer after delayed first parity. Am J Surg 2003;186:409–12.
[75] Iwasaki M, Otani T, Inoue M, Sasazuki S, Tsugane S, Group JPHCPS. Role and impact of menstrual and reproductive factors on breast cancer risk in Japan. Eur J Cancer Prev 2007;16:116–23.
[76] Clavel-Chapelon F, E3N-EPIC Group. Differential effects of reproductive factors on the risk of pre- and postmenopausal breast cancer. Results from a large cohort of French women. Br J Cancer 2002;86:723–7. doi:10.1038/sj.bjc.6600124.
[77] Lagiou A, Lagiou P, Vassilarou D-S, Stoikidou M, Trichopoulos D. Comparison of age at first full-term pregnancy between women with breast cancer and women with benign breast diseases. Int J Cancer 2003;107:817–21. doi:10.1002/ijc.11476.
[78] Lord SJ, Bernstein L, Johnson KA, Malone KE, McDonald JA, Marchbanks PA, et al. Breast cancer risk and hormone receptor status in older women by parity, age of first birth, and breastfeeding: a case-control study. Cancer Epidemiol Biomarkers Prev 2008;17:1723–30.

[79] Ursin G, Bernstein L, Lord SJ, Karim R, Deapen D, Press MF, et al. Reproductive factors and subtypes of breast cancer defined by hormone receptor and histology. Br J Cancer 2005;93:364–71.
[80] Ritte R, Tikk K, Lukanova A, Tjønneland A, Olsen A, Overvad K, et al. Reproductive factors and risk of hormone receptor positive and negative breast cancer: a cohort study. BMC Cancer 2013;13:584. doi:10.1186/1471-2407-13-584.
[81] Andrieu N, Prevost T, Rohan TE, Luporsi E, Le MG, Gerber M, et al. Variation in the interaction between familial and reproductive factors on the risk of breast cancer according to age, menopausal status, and degree of familiality. Int J Epidemiol 2000;29:214–23.
[82] Li CI, Malone KE, Daling JR, Potter JD, Bernstein L, Marchbanks PA, et al. Timing of menarche and first full-term birth in relation to breast cancer risk. Am J Epidemiol 2008;167:230–9.
[83] Lipworth L, Bailey LR, Trichopoulos D. History of breast-feeding in relation to breast cancer risk: a review of the epidemiologic literature. J Natl Cancer Inst 2000;92:302–12.
[84] Yang L, Jacobsen KH. A systematic review of the association between breastfeeding and breast cancer. J Womens Health 2008;17:1635–45.
[85] World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR; 2007.
[86] Nagata C, Mizoue T, Tanaka K, Tsuji I, Tamakoshi A, Wakai K, et al. Breastfeeding and breast cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. Jpn J Clin Oncol 2012;42:124–30.
[87] Gajalakshmi V, Mathew A, Brennan P, Rajan B, Kanimozhi VC, Mathews A, et al. Breastfeeding and breast cancer risk in India: A multicenter case-control study. Int J Cancer 2009;125:662–5.
[88] Huang Y, Zhang X, Li W, Song F, Dai H, Wang J, et al. A meta-analysis of the association between induced abortion and breast cancer risk among Chinese females. Cancer Causes Control 2014;25:227–36. doi:10.1007/s10552-013-0325-7.
[89] Braüner CM, Overvad K, Tjønneland A, Attermann J. Induced abortion and breast cancer among parous women: A Danish cohort study. Acta Obstet Gynecol Scand 2013;92:700–5.
[90] Henderson KD, Sullivan-Halley J, Reynolds P, Horn-Ross PL, Clarke CA, Chang ET, et al. Incomplete pregnancy is not associated with breast cancer risk: the California Teachers Study. Contraception 2008;77:391–6.
[91] Lazovich D, Thompson JA, Mink PJ, Sellers TA, Anderson KE. Induced abortion and breast cancer risk. Epidemiology 2000;11:76–80.
[92] Michels KB, Xue F, Colditz GA, Willett WC. Induced and spontaneous abortion and incidence of breast cancer among young women: a prospective cohort study. Arch Intern Med 2007;167:814–20.
[93] Palmer JR, Wise LA, Adams-Campbell LL, Rosenberg L. A prospective study of induced abortion and breast cancer in African-American women. Cancer Causes Control 2004;15:105–11.

[94] Reeves GK, Kan S-W, Key T, Tjønneland A, Olsen A, Overvad K, et al. Breast cancer risk in relation to abortion: Results from the EPIC study. Int J Cancer 2006;119:1741–5.

[95] Rosenblatt KA, Gao DL, Ray RM, Rowland MR, Nelson ZC, Wernli KJ, et al. Induced abortions and the risk of all cancers combined and site-specific cancers in Shanghai. Cancer Causes Control 2006;17:1275–80.

[96] Beral V, Bull D, Doll R, Peto R, Reeves G, Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and abortion: collaborative reanalysis of data from 53 epidemiological studies, including 83?000 women with breast cancer from 16 countries. Lancet 2004;363:1007–16. doi:10.1016/S0140-6736(04)15835-2.

[97] Paoletti X, Clavel-Chapelon F. Induced and spontaneous abortion and breast cancer risk: results from the E3N cohort study. Int J Cancer J Int Cancer 2003;106:270–6. doi:10.1002/ijc.11203.

[98] Hunter DJ, Colditz GA, Hankinson SE, Malspeis S, Spiegelman D, Chen W, et al. Oral contraceptive use and breast cancer: a prospective study of young women. Cancer Epidemiol Biomarkers Prev 2010;19:2496–502.

[99] Kumle M, Weiderpass E, Braaten T, Persson I, Adami H-O, Lund E. Use of Oral Contraceptives and Breast Cancer Risk The Norwegian-Swedish Women's Lifestyle and Health Cohort Study. Cancer Epidemiol Biomarkers Prev 2002;11:1375–81.

[100] Nelson HD, Zakher B, Cantor A, Fu R, Griffin J, O'Meara ES, et al. Risk Factors for Breast Cancer for Women Aged 40 to 49 YearsA Systematic Review and Meta-analysis. Ann Intern Med 2012;156:635–48.

[101] Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, et al. Oral contraceptives and the risk of breast cancer. N Engl J Med 2002;346:2025–32.

[102] Gierisch JM, Coeytaux RR, Urrutia RP, Havrilesky LJ, Moorman PG, Lowery WJ, et al. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2013;22:1931–43. doi:10.1158/1055-9965.EPI-13-0298.

[103] Moorman PG, Havrilesky LJ, Gierisch JM, Coeytaux RR, Lowery WJ, Peragallo Urrutia R, et al. Oral contraceptives and risk of ovarian cancer and breast cancer among high-risk women: a systematic review and meta-analysis. J Clin Oncol Off J Am Soc Clin Oncol 2013;31:4188–98. doi:10.1200/JCO.2013.48.9021.

[104] Moorman PG, Millikan RC, Newman B. Oral contraceptives and breast cancer among African-american women and white women. J Natl Med Assoc 2001;93:329.

[105] Hall IJ, Moorman PG, Millikan RC, Newman B. Comparative analysis of breast cancer risk factors among African-American women and White women. Am J Epidemiol 2005;161:40–51.

[106] Van Hoften C, Burger H, Peeters PH, Grobbee DE, Van Noord PA, Leufkens HG. Long-term oral contraceptive use increases breast cancer risk in women over 55 years of age: The DOM cohort. Int J Cancer 2000;87:591–4.

[107] Shantakumar S, Terry MB, Paykin A, Teitelbaum SL, Britton JA, Moorman PG, et al. Age and menopausal effects of hormonal birth control and hormone replacement therapy in relation to breast cancer risk. Am J Epidemiol 2007;165:1187–98.

[108] Althuis MD, Brogan DD, Coates RJ, Daling JR, Gammon MD, Malone KE, et al. Breast cancers among very young premenopausal women (United States). Cancer Causes Control 2003;14:151–60.
[109] Ma H, Bernstein L, Ross RK, Ursin G. Hormone-related risk factors for breast cancer in women under age 50 years by estrogen and progesterone receptor status: results from a case-control and a case-case comparison. Breast Cancer Res 2006;8:R39.
[110] Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor–defined breast cancer: a systematic review of the literature. Cancer Epidemiol Biomarkers Prev 2004;13:1558–68.
[111] Cotterchio M, Kreiger N, Theis B, Sloan M, Bahl S. Hormonal factors and the risk of breast cancer according to estrogen-and progesterone-receptor subgroup. Cancer Epidemiol Biomarkers Prev 2003;12:1053–60.
[112] Huang W-Y, Newman B, Millikan RC, Schell MJ, Hulka BS, Moorman PG. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. Am J Epidemiol 2000;151:703–14.
[113] McCredie MRE, Dite GS, Southey MC, Venter DJ, Giles GG, Hopper JL. Risk factors for breast cancer in young women by oestrogen receptor and progesterone receptor status. Br J Cancer 2003;89:1661–3.
[114] Rosenberg L, Zhang Y, Coogan PF, Strom BL, Palmer JR. A case-control study of oral contraceptive use and incident breast cancer. Am J Epidemiol 2009;169:473–9.
[115] Sweeney C, Giuliano AR, Baumgartner KB, Byers T, Herrick JS, Edwards SL, et al. Oral, injected and implanted contraceptives and breast cancer risk among U.S. Hispanic and non-Hispanic white women. Int J Cancer J Int Cancer 2007;121:2517–23. doi:10.1002/ijc.22970.
[116] Peterson HB, Jeng G, Folger SG, Hillis SA, Marchbanks PA, Wilcox LS. The risk of menstrual abnormalities after tubal sterilization. N Engl J Med 2000;343:1681–7.
[117] Pokoradi AJ, Iversen L, Hannaford PC. Factors associated with age of onset and type of menopause in a cohort of UK women. Am J Obstet Gynecol 2011;205:34–e1.
[118] Nelson DB, Sammel MD, Freeman EW, Gracia CR, Liu L, Langan E. Tubal ligation does not affect hormonal changes during the early menopausal transition. Contraception 2005;71:104–10.
[119] Whiteman MK, Miller KP, Tomic D, Langenberg P, Flaws JA. Tubal sterilization and hot flashes. Fertil Steril 2004;82:502–4.
[120] Wyshak G. Menopausal symptoms and psychological distress in women with and without tubal sterilization. Psychosomatics 2004;45:403–13.
[121] Gaudet MM, Patel AV, Sun J, Teras LR, Gapstur SM. Tubal sterilization and breast cancer incidence: results from the Cancer Prevention Study II Nutrition Cohort and meta-analysis. Am J Epidemiol 2013;177:492–9.
[122] Eliassen AH, Colditz GA, Rosner B, Hankinson SE. Tubal sterilization in relation to breast cancer risk. Int J Cancer 2006;118:2026–30.
[123] Press DJ, Sullivan-Halley J, Ursin G, Deapen D, McDonald JA, Strom BL, et al. Breast cancer risk and ovariectomy, hysterectomy, and tubal sterilization in the Women's Contraceptive and Reproductive Experiences Study. Am J Epidemiol 2011;173:38–47.

[124] Nichols HB, Baird DD, DeRoo LA, Kissling GE, Sandler DP. Tubal ligation in relation to menopausal symptoms and breast cancer risk. Br J Cancer 2013;109:1291–5.
[125] Chie W-C, Hsieh C, Newcomb PA, Longnecker MP, Mittendorf R, Greenberg ER, et al. Age at any full-term pregnancy and breast cancer risk. Am J Epidemiol 2000;151:715–22.
[126] Ji J, Fôrsti A, Sundquist J, Hemminki K. Risks of breast, endometrial, and ovarian cancers after twin births. Endocr Relat Cancer 2007;14:703–11.
[127] Neale RE, Darlington S, Murphy MF, Silcocks P, Purdie DM, Talbäck M. The effects of twins, parity and age at first birth on cancer risk in Swedish women. Twin Res Hum Genet 2005;8:156–62.
[128] Liu Q, Wuu J, Lambe M, Hsieh S-F, Ekbom A, Hsieh C-C. Transient increase in breast cancer risk after giving birth: postpartum period with the highest risk (Sweden). Cancer Causes Control 2002;13:299–305. doi:10.1023/A:1015287208222.
[129] Bruzzi P, Negri E, La Vecchia C, Decarli A, Palli D, Parazzini F, et al. Short term increase in risk of breast cancer after full term pregnancy. Br Med J 1988;297:1096–8. doi:10.1136/bmj.297.6656.1096.
[130] Kalache A, Maguire A, Thompson SG. Age at last full-term pregnancy and risk of breast cancer. The Lancet 1993;341:33–6. doi:10.1016/0140-6736(93)92497-H.
[131] Williams EM, Jones L, Vessey MP, McPherson K. Short term increase in risk of breast cancer associated with full term pregnancy. Br Med J 1990;300:578–9. doi:10.1136/bmj.300.6724.578.
[132] Adami HO, Bergström R, Lund E, Meirik O. Absence of association between reproductive variables and the risk of breast cancer in young women in Sweden and Norway. Br J Cancer 1990;62:122–6.
[133] Palmer JR, Wise LA, Horton NJ, Adams-Campbell LL, Rosenberg L. Dual Effect of Parity on Breast Cancer Risk in African-American Women. J Natl Cancer Inst 2003;95:478–83. doi:10.1093/jnci/95.6.478.
[134] Vatten LJ, Kvinnsland S. Pregnancy-related factors and risk of breast cancer in a prospective study of 29 981 Norwegian women. Eur J Cancer 1992;28:1148–53. doi:10.1016/0959-8049(92)90476-I.
[135] Gunnell D. Height, insulin-like growth factors and cancer risk. Growth Horm IGF Res 2000;10:S39–40.
[136] Li CI, Littman AJ, White E. Relationship between age maximum height is attained, age at menarche, and age at first full-term birth and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2007;16:2144–9.
[137] Opdahl S, Nilsen TIL, Romundstad PR, Vanky E, Carlsen SM, Vatten LJ. Association of size at birth with adolescent hormone levels, body size and age at menarche: relevance for breast cancer risk. Br J Cancer 2008;99:201–6.
[138] Ruder EH, Dorgan JF, Kranz S, Kris-Etherton PM, Hartman TJ. Examining breast cancer growth and lifestyle risk factors: early life, childhood, and adolescence. Clin Breast Cancer 2008;8:334–42.
[139] Van den Brandt PA, Spiegelman D, Yaun S-S, Adami H-O, Beeson L, Folsom AR, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol 2000;152:514–27.

[140] Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. Lancet Oncol 2011;12:785–94.
[141] Amadou A, Ferrari P, Muwonge R, Moskal A, Biessy C, Romieu I, et al. Overweight, obesity and risk of premenopausal breast cancer according to ethnicity: a systematic review and dose-response meta-analysis. Obes Rev 2013;14:665–78.
[142] Kabat GC, Heo M, Kamensky V, Miller AB, Rohan TE. Adult height in relation to risk of cancer in a cohort of Canadian women. Int J Cancer 2013;132:1125–32.
[143] Hall IJ, Newman B, Millikan RC, Moorman PG. Body Size and Breast Cancer Risk in Black Women and White Women The Carolina Breast Cancer Study. Am J Epidemiol 2000;151:754–64.
[144] Borgquist S, Jirström K, Anagnostaki L, Manjer J, Landberg G. Anthropometric factors in relation to different tumor biological subgroups of postmenopausal breast cancer. Int J Cancer 2009;124:402–11.
[145] Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk factors for breast cancer according to estrogen and progesterone receptor status. J Natl Cancer Inst 2004;96:218–28.
[146] John EM, Sangaramoorthy M, Phipps AI, Koo J, Horn-Ross PL. Adult Body Size, Hormone Receptor Status, and Premenopausal Breast Cancer Risk in a Multiethnic Population The San Francisco Bay Area Breast Cancer Study. Am J Epidemiol 2011;173:201–16.
[147] Mellemkjær L, Christensen J, Frederiksen K, Baker JL, Olsen A, Sørensen TIA, et al. Leg length, sitting height and postmenopausal breast cancer risk. Br J Cancer 2012;107:165–8. doi:10.1038/bjc.2012.244.
[148] Rosenberg LU, Einarsdóttir K, Friman EI, Wedrén S, Dickman PW, Hall P, et al. Risk factors for hormone receptor-defined breast cancer in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2006;15:2482–8.
[149] Kawai M, Minami Y, Kuriyama S, Kakizaki M, Kakugawa Y, Nishino Y, et al. Adiposity, adult weight change and breast cancer risk in postmenopausal Japanese women: the Miyagi Cohort Study. Br J Cancer 2010;103:1443–7.
[150] Mathew A, Gajalakshmi V, Rajan B, Kanimozhi V, Brennan P, Mathew BS, et al. Anthropometric factors and breast cancer risk among urban and rural women in South India: a multicentric case–control study. Br J Cancer 2008;99:207–13.
[151] Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. The Lancet 2008;371:569–78.
[152] Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status—A meta-analysis. Int J Cancer 2009;124:698–712.
[153] Tehard B, Clavel-Chapelon F. Several anthropometric measurements and breast cancer risk: results of the E3N cohort study. Int J Obes 2006;30:156–63.
[154] Ogundiran TO, Huo D, Adenipekun A, Campbell O, Oyesegun R, Akang E, et al. Case-control study of body size and breast cancer risk in Nigerian women. Am J Epidemiol 2010;172:682–90.

[155] Wu AH, Yu MC, Tseng C-C, Pike MC. Body size, hormone therapy and risk of breast cancer in Asian–American women. Int J Cancer 2007;120:844–52.
[156] Palmer JR, Adams-Campbell LL, Boggs DA, Wise LA, Rosenberg L. A prospective study of body size and breast cancer in black women. Cancer Epidemiol Biomarkers Prev 2007;16:1795–802.
[157] Weiderpass E, Braaten T, Magnusson C, Kumle M, Vainio H, Lund E, et al. A prospective study of body size in different periods of life and risk of premenopausal breast cancer. Cancer Epidemiol Biomarkers Prev 2004;13:1121–7.
[158] Adebamowo CA, Ogundiran TO, Adenipekun AA, Oyesegun RA, Campbell OB, Akang EU, et al. Obesity and height in urban Nigerian women with breast cancer. Ann Epidemiol 2003;13:455–61.
[159] Enger SM, Ross RK, Paganini-Hill A, Carpenter CL, Bernstein L. Body size, physical activity, and breast cancer hormone receptor status: results from two case-control studies. Cancer Epidemiol Biomarkers Prev 2000;9:681–7.
[160] Kuriyama S, Tsubono Y, Hozawa A, Shimazu T, Suzuki Y, Koizumi Y, et al. Obesity and risk of cancer in Japan. Int J Cancer 2005;113:148–57.
[161] Wu M-H, Chou Y-C, Yu J-C, Yu C-P, Wu C-C, Chu C-M, et al. Hormonal and body-size factors in relation to breast cancer risk: a prospective study of 11,889 women in a low-incidence area. Ann Epidemiol 2006;16:223–9.
[162] Chow LW, Lui KL, Chan JCY, Chan TC, Ho PK, Lee WY, et al. Association between body mass index and risk of formation of breast cancer in Chinese women. Asian J Surg 2005;28:179–84.
[163] Iwasaki M, Otani T, Inoue M, Sasazuki S, Tsugane S. Body size and risk for breast cancer in relation to estrogen and progesterone receptor status in Japan. Ann Epidemiol 2007;17:304–12.
[164] Cheraghi Z, Poorolajal J, Hashem T, Esmailnasab N, Irani AD. Effect of body mass index on breast cancer during premenopausal and postmenopausal periods: a meta-analysis. PloS One 2012;7:e51446.
[165] Pischon T, Nöthlings U, Boeing H. Obesity and cancer. Proc Nutr Soc 2008;67:128–45.
[166] De Pergola G, Silvestris F. Obesity as a major risk factor for cancer. J Obes 2013;2013.
[167] Harris HR, Willett WC, Terry KL, Michels KB. Body fat distribution and risk of premenopausal breast cancer in the Nurses' Health Study II. J Natl Cancer Inst 2011;103:273–8. doi:10.1093/jnci/djq500.
[168] Harvie M, Hooper L, Howell AH. Central obesity and breast cancer risk: a systematic review. Obes Rev 2003;4:157–73.
[169] Connolly BS, Barnett C, Vogt KN, Li T, Stone J, Boyd NF. A meta-analysis of published literature on waist-to-hip ratio and risk of breast cancer. Nutr Cancer 2002;44:127–38.
[170] Ehtisham S, Crabtree N, Clark P, Shaw N, Barrett T. Ethnic differences in insulin resistance and body composition in United Kingdom adolescents. J Clin Endocrinol Metab 2005;90:3963–9.

[171] Liu A, Byrne NM, Kagawa M, Ma G, Poh BK, Ismail MN, et al. Ethnic differences in the relationship between body mass index and percentage body fat among Asian children from different backgrounds. Br J Nutr 2011;106:1390.
[172] Rush EC, Goedecke JH, Jennings C, Micklesfield L, Dugas L, Lambert EV, et al. BMI, fat and muscle differences in urban women of five ethnicities from two countries. Int J Obes 2007;31:1232–9.
[173] Rush EC, Freitas I, Plank LD. Body size, body composition and fat distribution: comparative analysis of European, Maori, Pacific Island and Asian Indian adults. Br J Nutr 2009;102:632–41.
[174] Jordan S, Lim L, Vilainerun D, Banks E, Sripaiboonkij N, Seubsman S, et al. Breast cancer in the Thai Cohort Study: an exploratory case-control analysis. The Breast 2009;18:299–303.
[175] Friedenreich CM, Courneya KS, Bryant HE. Case-control study of anthropometric measures and breast cancer risk. Int J Cancer 2002;99:445–52.
[176] Slattery ML, Sweeney C, Edwards S, Herrick J, Baumgartner K, Wolff R, et al. Body size, weight change, fat distribution and breast cancer risk in Hispanic and non-Hispanic white women. Breast Cancer Res Treat 2007;102:85–101.
[177] Morimoto LM, White E, Chen Z, Chlebowski RT, Hays J, Kuller L, et al. Obesity, body size, and risk of postmenopausal breast cancer: the Women's Health Initiative (United States). Cancer Causes Control 2002;13:741–51.
[178] Amadou A, Torres Mejia G, Fagherazzi G, Ortega C, Angeles-Llerenas A, Chajes V, et al. Anthropometry, Silhouette Trajectory, and Risk of Breast Cancer in Mexican Women. Am J Prev Med 2014;46:S52–64.
[179] Ogundiran TO, Huo D, Adenipekun A, Campbell O, Oyesegun R, Akang E, et al. Body fat distribution and breast cancer risk: findings from the Nigerian breast cancer study. Cancer Causes Control 2012;23:565–74.
[180] Bandera EV, Chandran U, Zirpoli G, Gong Z, McCann SE, Hong C-C, et al. Body fatness and breast cancer risk in women of African ancestry. BMC Cancer 2013;13:475. doi:10.1186/1471-2407-13-475.
[181] Fagherazzi G, Chabbert-Buffet N, Fabre A, Guillas G, Boutron-Ruault M-C, Mesrine S, et al. Hip circumference is associated with the risk of premenopausal ER-/PR- breast cancer. Int J Obes 2005 2012;36:431–9. doi:10.1038/ijo.2011.66.
[182] Michels KB, Terry KL, Willett WC. Longitudinal study on the role of body size in premenopausal breast cancer. Arch Intern Med 2006;166:2395–402.
[183] Vasconcelos AB de, Sichieri R. Height, weight, weight change and risk of breast cancer in Rio de Janeiro, Brazil. Sao Paulo Med J 2001;119:62–6.
[184] Eliassen AH, Colditz GA, Rosner B, Willett WC, Hankinson SE. Adult weight change and risk of postmenopausal breast cancer. Jama 2006;296:193–201.
[185] Lahmann PH, Schulz M, Hoffmann K, Boeing H, Tjønneland A, Olsen A, et al. Long-term weight change and breast cancer risk: the European prospective investigation into cancer and nutrition (EPIC). Br J Cancer 2005;93:582–9.
[186] Bergström A, Pisani P, Tenet V, Wolk A, Adami H-O. Overweight as an avoidable cause of cancer in Europe. Int J Cancer 2001;91:421–30.

[187] Marmot M, Atinmo T, Byers T, Chen J, Hirohata T, Jackson A, et al. Food, nutrition, physical activity, and the prevention of cancer: a global perspective 2007.
[188] Suzuki R, Iwasaki M, Inoue M, Sasazuki S, Sawada N, Yamaji T, et al. Body weight at age 20 years, subsequent weight change and breast cancer risk defined by estrogen and progesterone receptor status—the Japan public health center-based prospective study. Int J Cancer 2011;129:1214–24.
[189] Suzuki S, Kojima M, Tokudome S, Mori M, Sakauchi F, Wakai K, et al. Obesity/weight gain and breast cancer risk: findings from the Japan collaborative cohort study for the evaluation of cancer risk. J Epidemiol 2013;23:139.
[190] Teras LR, Goodman M, Patel AV, Diver WR, Flanders WD, Feigelson HS. Weight loss and postmenopausal breast cancer in a prospective cohort of overweight and obese US women. Cancer Causes Control 2011;22:573–9.
[191] White KK, Park S-Y, Kolonel LN, Henderson BE, Wilkens LR. Body size and breast cancer risk: the Multiethnic Cohort. Int J Cancer 2012;131:E705–16.
[192] Michels KB, Terry KL, Eliassen AH, Hankinson SE, Willett WC. Adult weight change and incidence of premenopausal breast cancer. Int J Cancer 2012;130:902–9.
[193] Maruti SS, Willett WC, Feskanich D, Rosner B, Colditz GA. A prospective study of age-specific physical activity and premenopausal breast cancer. J Natl Cancer Inst 2008;100:728–37.
[194] Monninkhof EM, Elias SG, Vlems FA, van der Tweel I, Schuit AJ, Voskuil DW, et al. Physical activity and breast cancer: a systematic review. Epidemiology 2007;18:137–57.
[195] Friedenreich CM, Neilson HK, Lynch BM. State of the epidemiological evidence on physical activity and cancer prevention. Eur J Cancer Oxf Engl 1990 2010;46:2593–604. doi:10.1016/j.ejca.2010.07.028.
[196] Wu Y, Zhang D, Kang S. Physical activity and risk of breast cancer: a meta-analysis of prospective studies. Breast Cancer Res Treat 2013;137:869–82. doi:10.1007/s10549-012-2396-7.
[197] Kobayashi LC, Janssen I, Richardson H, Lai AS, Spinelli JJ, Aronson KJ. A case-control study of lifetime light intensity physical activity and breast cancer risk. Cancer Causes Control CCC 2014;25:133–40. doi:10.1007/s10552-013-0312-z.
[198] Mathew A, Gajalakshmi V, Rajan B, Kanimozhi VC, Brennan P, Binukumar BP, et al. Physical activity levels among urban and rural women in south India and the risk of breast cancer: a case-control study. Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP 2009;18:368–76. doi:10.1097/CEJ.0b013e32832e1c46.
[199] Vainio H, Kaaks R, Bianchini F. Weight control and physical activity in cancer prevention: international evaluation of the evidence. Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP 2002;11:S94.
[200] Villeneuve S, Févotte J, Anger A, Truong T, Lamkarkach F, Gaye O, et al. Breast cancer risk by occupation and industry: Analysis of the CECILE study, a population-based case–control study in France. Am J Ind Med 2011;54:499–509.
[201] Labrèche F, Goldberg MS, Valois M-F, Nadon L. Postmenopausal breast cancer and occupational exposures. Occup Environ Med 2010;67:263–9. doi:10.1136/oem.2009.049817.

[202] Clapp RW, Jacobs MM, Loechler EL. Environmental and occupational causes of cancer: new evidence 2005-2007. Rev Environ Health 2008;23:1–38.
[203] Cohn BA, Wolff MS, Cirillo PM, Sholtz RI. DDT and breast cancer in young women: new data on the significance of age at exposure. Environ Health Perspect 2007;115:1406–14. doi:10.1289/ehp.10260.
[204] Band PR, Le ND, Fang R, Deschamps M, Gallagher RP, Yang P. Identification of occupational cancer risks in British Columbia: a population-based case-control study of 995 incident breast cancer cases by menopausal status, controlling for confounding factors. J Occup Environ Med 2000;42:284–310.
[205] Megdal SP, Kroenke CH, Laden F, Pukkala E, Schernhammer ES. Night work and breast cancer risk: a systematic review and meta-analysis. Eur J Cancer 2005;41:2023–32.
[206] Schernhammer ES, Hankinson SE. Urinary melatonin levels and postmenopausal breast cancer risk in the Nurses' Health Study cohort. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2009;18:74–9. doi:10.1158/1055-9965.EPI-08-0637.
[207] Stevens RG, Blask DE, Brainard GC, Hansen J, Lockley SW, Provencio I, et al. Meeting Report: The Role of Environmental Lighting and Circadian Disruption in Cancer and Other Diseases. Environ Health Perspect 2007;115:1357–62. doi:10.1289/ehp.10200.
[208] Jia Y, Lu Y, Wu K, Lin Q, Shen W, Zhu M, et al. Does night work increase the risk of breast cancer? A systematic review and meta-analysis of epidemiological studies. Cancer Epidemiol 2013;37:197–206. doi:10.1016/j.canep.2013.01.005.
[209] Kolstad HA. Nightshift work and risk of breast cancer and other cancers--a critical review of the epidemiologic evidence. Scand J Work Environ Health 2008;34:5–22.
[210] Reulen RC, Frobisher C, Winter DL, Kelly J, Lancashire ER, Stiller CA, et al. Long-term risks of subsequent primary neoplasms among survivors of childhood cancer. JAMA J Am Med Assoc 2011;305:2311–9. doi:10.1001/jama.2011.747.
[211] Berrington de Gonzalez A, Curtis RE, Gilbert E, Berg CD, Smith SA, Stovall M, et al. Second solid cancers after radiotherapy for breast cancer in SEER cancer registries. Br J Cancer 2010;102:220–6. doi:10.1038/sj.bjc.6605435.
[212] Neta G, Anderson WF, Gilbert E, Berrington A. Variation in the risk of radiation-related contralateral breast cancer by histology and estrogen receptor expression in SEER. Breast Cancer Res Treat 2012;131:1021–7. doi:10.1007/s10549-011-1820-8.
[213] Berrington de González A, Darby S. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. Lancet 2004;363:345–51. doi:10.1016/S0140-6736(04)15433-0.
[214] Pijpe A, Andrieu N, Easton DF, Kesminiene A, Cardis E, Noguès C, et al. Exposure to diagnostic radiation and risk of breast cancer among carriers of BRCA1/2 mutations: retrospective cohort study (GENE-RAD-RISK). BMJ 2012;345:e5660.
[215] Berrington de González A. Estimates of the potential risk of radiation-related cancer from screening in the UK. J Med Screen 2011;18:163–4. doi:10.1258/jms.2011.011073.

[216] Mathews JD, Forsythe AV, Brady Z, Butler MW, Goergen SK, Byrnes GB, et al. Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. BMJ 2013;346:f2360–f2360. doi:10.1136/bmj.f2360.

[217] Aune D, Chan DSM, Vieira AR, Rosenblatt DAN, Vieira R, Greenwood DC, et al. Fruits, vegetables and breast cancer risk: a systematic review and meta-analysis of prospective studies. Breast Cancer Res Treat 2012;134:479–93. doi:10.1007/s10549-012-2118-1.

[218] Aune D, Chan DSM, Greenwood DC, Vieira AR, Rosenblatt DAN, Vieira R, et al. Dietary fiber and breast cancer risk: a systematic review and meta-analysis of prospective studies. Ann Oncol Off J Eur Soc Med Oncol ESMO 2012;23:1394–402. doi:10.1093/annonc/mdr589.

[219] Eliassen AH, Hendrickson SJ, Brinton LA, Buring JE, Campos H, Dai Q, et al. Circulating carotenoids and risk of breast cancer: pooled analysis of eight prospective studies. J Natl Cancer Inst 2012;104:1905–16. doi:10.1093/jnci/djs461.

[220] Velentzis LS, Cantwell MM, Cardwell C, Keshtgar MR, Leathem AJ, Woodside JV. Lignans and breast cancer risk in pre- and post-menopausal women: meta-analyses of observational studies. Br J Cancer 2009;100:1492–8. doi:10.1038/sj.bjc.6605003.

[221] Dong J-Y, Qin L-Q. Soy isoflavones consumption and risk of breast cancer incidence or recurrence: a meta-analysis of prospective studies. Breast Cancer Res Treat 2011;125:315–23. doi:10.1007/s10549-010-1270-8.

[222] Fritz H, Seely D, Flower G, Skidmore B, Fernandes R, Vadeboncoeur S, et al. Soy, red clover, and isoflavones and breast cancer: a systematic review. PloS One 2013;8:e81968. doi:10.1371/journal.pone.0081968.

[223] Hui C, Qi X, Qianyong Z, Xiaoli P, Jundong Z, Mantian M. Flavonoids, flavonoid subclasses and breast cancer risk: a meta-analysis of epidemiologic studies. PloS One 2013;8:e54318. doi:10.1371/journal.pone.0054318.

[224] Zheng J-S, Hu X-J, Zhao Y-M, Yang J, Li D. Intake of fish and marine n-3 polyunsaturated fatty acids and risk of breast cancer: meta-analysis of data from 21 independent prospective cohort studies. BMJ 2013;346:f3706.

[225] Alexander DD, Morimoto LM, Mink PJ, Cushing CA. A review and meta-analysis of red and processed meat consumption and breast cancer. Nutr Res Rev 2010;23:349–65. doi:10.1017/S0954422410000235.

[226] Missmer SA, Smith-Warner SA, Spiegelman D, Yaun S-S, Adami H-O, Beeson WL, et al. Meat and dairy food consumption and breast cancer: a pooled analysis of cohort studies. Int J Epidemiol 2002;31:78–85.

[227] Wu Y, Zhang D, Kang S. Black tea, green tea and risk of breast cancer: an update. SpringerPlus 2013;2:240. doi:10.1186/2193-1801-2-240.

[228] Misotti AM, Gnagnarella P. Vitamin supplement consumption and breast cancer risk: a review. Ecancermedicalscience 2013;7. doi:10.3332/ecancer.2013.365.

[229] Kim Y, Je Y. Vitamin D intake, blood 25(OH)D levels, and breast cancer risk or mortality: a meta-analysis. Br J Cancer 2014. doi:10.1038/bjc.2014.175.

[230] Si R, Qu K, Jiang Z, Yang X, Gao P. Egg consumption and breast cancer risk: a meta-analysis. Breast Cancer Tokyo Jpn 2014;21:251–61. doi:10.1007/s12282-014-0519-1.

[231] Albuquerque RCR, Baltar VT, Marchioni DML. Breast cancer and dietary patterns: a systematic review. Nutr Rev 2014;72:1–17. doi:10.1111/nure.12083.

[232] Cottet V, Touvier M, Fournier A, Touillaud MS, Lafay L, Clavel-Chapelon F, et al. Postmenopausal breast cancer risk and dietary patterns in the E3N-EPIC prospective cohort study. Am J Epidemiol 2009;170:1257–67. doi:10.1093/aje/kwp257.

[233] Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001;358:1389–99. doi:10.1016/S0140-6736(01)06524-2.

[234] Mavaddat N, Pharoah PD, Blows F, Driver KE, Provenzano E, Thompson D, et al. Familial relative risks for breast cancer by pathological subtype: a population-based cohort study. Breast Cancer Res BCR 2010;12:R10. doi:10.1186/bcr2476.

[235] Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85. doi:10.1056/NEJM200007133430201.

[236] Antoniou A, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003;72:1117–30. doi:10.1086/375033.

[237] Evans DGR, Ingham SL, Buchan I, Woodward ER, Byers H, Howell A, et al. Increased rate of phenocopies in all age groups in BRCA1/BRCA2 mutation kindred, but increased prospective breast cancer risk is confined to BRCA2 mutation carriers. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2013;22:2269–76. doi:10.1158/1055-9965.EPI-13-0316-T.

[238] Widschwendter M, Rosenthal AN, Philpott S, Rizzuto I, Fraser L, Hayward J, et al. The sex hormone system in carriers of BRCA1/2 mutations: a case-control study. Lancet Oncol 2013;14:1226–32. doi:10.1016/S1470-2045(13)70448-0.

[239] Tea M-KM, Kroiss R, Muhr D, Fuerhauser-Rappaport C, Oefner P, Wagner TM, et al. Central European BRCA2 mutation carriers: birth cohort status correlates with onset of breast cancer. Maturitas 2014;77:68–72. doi:10.1016/j.maturitas.2013.09.012.

[240] Turnbull C, Rahman N. Genetic predisposition to breast cancer: past, present, and future. Annu Rev Genomics Hum Genet 2008;9:321–45. doi:10.1146/annurev.genom.9.081307.164339.

[241] Zhang B, Beeghly-Fadiel A, Long J, Zheng W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Lancet Oncol 2011;12:477–88. doi:10.1016/S1470-2045(11)70076-6.

[242] Cai Q, Long J, Lu W, Qu S, Wen W, Kang D, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. Hum Mol Genet 2011;20:4991–9. doi:10.1093/hmg/ddr405.

[243] Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet 2013;45:392–8, 398e1–2. doi:10.1038/ng.2561.

[244] Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45:353–61, 361e1–2. doi:10.1038/ng.2563.

[245] Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009;41:579–84. doi:10.1038/ng.353.

[246] Li N, Dong J, Hu Z, Shen H, Dai M. Potentially functional polymorphisms in ESR1 and breast cancer risk: a meta-analysis. Breast Cancer Res Treat 2010;121:177–84. doi:10.1007/s10549-009-0532-9.

[247] Zhang J, Qiu L-X, Wang Z-H, Leaw S-J, Wang B-Y, Wang J-L, et al. Current evidence on the relationship between three polymorphisms in the FGFR2 gene and breast cancer risk: a meta-analysis. Breast Cancer Res Treat 2010;124:419–24. doi:10.1007/s10549-010-0846-7.

[248] Cogliano VJ, Baan R, Straif K, Grosse Y, Lauby-Secretan B, El Ghissassi F, et al. Preventable exposures associated with human cancers. J Natl Cancer Inst 2011;103:1827–39. doi:10.1093/jnci/djr483.

[249] Endogenous Hormones and Breast Cancer Collaborative Group, Key TJ, Appleby PN, Reeves GK, Roddam AW, Helzlsouer KJ, et al. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. Br J Cancer 2011;105:709–22. doi:10.1038/bjc.2011.254.

[250] Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ. Active smoking and breast cancer risk: original cohort data and meta-analysis. J Natl Cancer Inst 2013;105:515–25. doi:10.1093/jnci/djt023.

[251] Xue F, Willett WC, Rosner BA, Hankinson SE, Michels KB. Cigarette smoking and the incidence of breast cancer. Arch Intern Med 2011;171:125–33. doi:10.1001/archinternmed.2010.503.

[252] Luo J, Horn K, Ockene JK, Simon MS, Stefanick ML, Tong E, et al. Interaction between smoking and obesity and the risk of developing breast cancer among postmenopausal women: the Women's Health Initiative Observational Study. Am J Epidemiol 2011;174:919–28. doi:10.1093/aje/kwr192.

[253] Kabat GC, Kim M, Phipps AI, Li CI, Messina CR, Wactawski-Wende J, et al. Smoking and alcohol consumption in relation to risk of triple-negative breast cancer in a cohort of postmenopausal women. Cancer Causes Control CCC 2011;22:775–83. doi:10.1007/s10552-011-9750-7.

[254] Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Bouvard V, et al. Carcinogenicity of alcoholic beverages. Lancet Oncol 2007;8:292–3.

[255] Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, et al. Light alcohol drinking and cancer: a meta-analysis. Ann Oncol Off J Eur Soc Med Oncol ESMO 2013;24:301–8. doi:10.1093/annonc/mds337.

[256] Allen NE, Beral V, Casabonne D, Kan SW, Reeves GK, Brown A, et al. Moderate Alcohol Intake and Cancer Incidence in Women. J Natl Cancer Inst 2009;101:296–305. doi:10.1093/jnci/djn514.

[257] Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW Jr, et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53

epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer 2002;87:1234–45. doi:10.1038/sj.bjc.6600596.
[258] Key J, Hodgson S, Omar RZ, Jensen TK, Thompson SG, Boobis AR, et al. Meta-analysis of studies of alcohol and breast cancer with consideration of the methodological issues. Cancer Causes Control CCC 2006;17:759–70. doi:10.1007/s10552-006-0011-0.
[259] Deandrea S, Talamini R, Foschi R, Montella M, Dal Maso L, Falcini F, et al. Alcohol and breast cancer risk defined by estrogen and progesterone receptor status: a case-control study. Cancer Epidemiol Biomarkers Prev 2008;17:2025–8.
[260] Li CI, Malone KE, Porter PL, Weiss NS, Tang M-TC, Daling JR. The relationship between alcohol use and risk of breast cancer by histology and hormone receptor status among women 65–79 years of age. Cancer Epidemiol Biomarkers Prev 2003;12:1061–6.
[261] Li CI, Chlebowski RT, Freiberg M, Johnson KC, Kuller L, Lane D, et al. Alcohol consumption and risk of postmenopausal breast cancer by subtype: the women's health initiative observational study. J Natl Cancer Inst 2010;102:1422–31.
[262] Suzuki R, Ye W, Rylander-Rudqvist T, Saji S, Colditz GA, Wolk A. Alcohol and postmenopausal breast cancer risk defined by estrogen and progesterone receptor status: a prospective cohort study. J Natl Cancer Inst 2005;97:1601–8.
[263] Terry MB, Zhang FF, Kabat G, Britton JA, Teitelbaum SL, Neugut AI, et al. Lifetime alcohol intake and breast cancer risk. Ann Epidemiol 2006;16:230–40.
[264] Chattopadhyay S, Siddiqui S, Akhtar MS, Najm MZ, Deo SVS, Shukla NK, et al. Genetic polymorphisms of ESR1, ESR2, CYP17A1, and CYP19A1 and the risk of breast cancer: a case control study from North India. Tumor Biol 2014;35:4517–27. doi:10.1007/s13277-013-1594-1.
[265] Joshi NN, Kale MD, Hake SS, Kannan S. Transforming growth factor β signaling pathway associated gene polymorphisms may explain lower breast cancer risk in western Indian women. PloS One 2011;6:e21866. doi:10.1371/journal.pone.0021866.
[266] Kohaar I, Tiwari P, Kumar R, Nasare V, Thakur N, Das BC, et al. Association of single nucleotide polymorphisms (SNPs) in TNF-LTA locus with breast cancer risk in Indian population. Breast Cancer Res Treat 2009;114:347–55.
[267] Tulsyan S, Agarwal G, Lal P, Agrawal S, Mittal RD, Mittal B. CD44 gene polymorphisms in breast cancer risk and prognosis: a study in North Indian population. PloS One 2013;8:e71073. doi:10.1371/journal.pone.0071073.
[268] Dikshit RP, Nagrani R, Mhatre S. Guidelines and Working Manual for conducting interviews for Multi-site Case Control studies. Mumbai, India: Tata Memorial Centre; 2011.
[269] Dikshit RP, Nagrani R, Mhatre S. Guidelines and Working Manual of data entry for Multi-site Case Control studies. Mumbai, India: Tata Memorial Centre; 2012.
[270] Lee R, Nieman D. Nutritional Assessment. 6th ed. 2012.
[271] WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004;363:157–63. doi:10.1016/S0140-6736(03)15268-3.

[272] World Health Organization. Obesity: Preventing and Managing the Global Epidemic. World Health Organization; 2000.
[273] Breslow NE, Day NE. Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC Sci Publ 1980:5–338.
[274] StataCorp. Stata Statistical Software. College Station, TX: StataCorp LP; 2011.
[275] Must A, Willett WC, Dietz WH. Remote recall of childhood height, weight, and body build by elderly subjects. Am J Epidemiol 1993;138:56–64.
[276] Dey S, Soliman AS, Hablas A, Seifeldin IA, Ismail K, Ramadan M, et al. Urban–rural differences in breast cancer incidence by hormone receptor status across 6 years in Egypt. Breast Cancer Res Treat 2010;120:149–60.
[277] Feng B-J, Jalbout M, Ayoub WB, Khyatti M, Dahmoul S, Ayad M, et al. Dietary risk factors for nasopharyngeal carcinoma in Maghrebian countries. Int J Cancer J Int Cancer 2007;121:1550–5. doi:10.1002/ijc.22813.
[278] Cowgill KD, Loffredo CA, Eissa SA-L, Mokhtar N, Abdel-Hamid M, Fahmy A, et al. Case-control study of non-Hodgkin's lymphoma and hepatitis C virus infection in Egypt. Int J Epidemiol 2004;33:1034–9.
[279] Falk RT, Fears TR, Hoover RN, Pike MC, Wu AH, Nomura AMY, et al. Does place of birth influence endogenous hormone levels in Asian-American women? Br J Cancer 2002;87:54–60. doi:10.1038/sj.bjc.6600339.
[280] Okasha M, McCarron P, Gunnell D, Smith GD. Exposures in childhood, adolescence and early adulthood and breast cancer risk: a systematic review of the literature. Breast Cancer Res Treat 2003;78:223–76.
[281] Willett W. Nutritional epidemiology. vol. 40. Oxford University Press; 2012.
[282] Pawar S, Pawar R, Gandham S, Prabhudesai S, Singh R, Gupta S. Evaluation of ER, PR and HER-2 receptor expression in breast cancer patients presenting to a semi urban cancer centre in Western India. J Cancer Res Ther 2014;10:26. doi:10.4103/0973-1482.131348.
[283] Kelsey JL, Bernstein L. Epidemiology and prevention of breast cancer. Annu Rev Public Health 1996;17:47–67. doi:10.1146/annurev.pu.17.050196.000403.
[284] Cairns BJ, Liu B, Clennell S, Cooper R, Reeves GK, Beral V, et al. Lifetime body size and reproductive factors: comparisons of data recorded prospectively with self reports in middle age. BMC Med Res Methodol 2011;11:7. doi:10.1186/1471-2288-11-7.
[285] Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, et al. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. Am J Epidemiol 1999;150:341–53.
[286] Ma H, Henderson KD, Sullivan-Halley J, Duan L, Marshall SF, Ursin G, et al. Pregnancy-related factors and the risk of breast carcinoma in situ and invasive breast cancer among postmenopausal women in the California Teachers Study cohort. Breast Cancer Res 2010;12:R35. doi:10.1186/bcr2589.
[287] Russo J, Mailo D, Hu Y-F, Balogh G, Sheriff F, Russo IH. Breast differentiation and its implication in cancer prevention. Clin Cancer Res Off J Am Assoc Cancer Res 2005;11:931s – 6s.
[288] Henderson BE, Feigelson HS. Hormonal carcinogenesis. Carcinogenesis 2000;21:427–33.

[289] Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. Lancet Oncol 2001;2:133–40. doi:10.1016/S1470-2045(00)00254-0.
[290] Wohlfahrt J, Melbye M. Maternal risk of breast cancer and birth characteristics of offspring by time since birth. Epidemiology 1999;10:441–4.
[291] Kappel B, Hansen K, Moller J, Faaborg-Andersen J. Human placental lactogen and dU-estrogen levels in normal twin pregnancies. Acta Genet Med Gemellol (Roma) 1984;34:59–65.
[292] Noble PL, Snijders RJM, Abraha HD, Sherwood RA, Nicolaides KH. Maternal serum free β-hCG at 10 to 14 weeks of gestation in trisomic twin pregnancies. BJOG Int J Obstet Gynaecol 1997;104:741–3.
[293] WALD N, CUCKLE H, WU T, GEORGE L. Maternal serum unconjugated oestriol and human chorionic gonadotrophin levels in twin pregnancies: implications for screening for Down's syndrome. BJOG Int J Obstet Gynaecol 1991;98:905–8.
[294] Ammon Avalos L, Galindo C, Li D-K. A systematic review to calculate background miscarriage rates using life table analysis. Birt Defects Res A Clin Mol Teratol 2012;94:417–23. doi:10.1002/bdra.23014.
[295] Han TS, Sattar N, Lean M. Assessment of obesity and its clinical implications. BMJ 2006;333:695–8. doi:10.1136/bmj.333.7570.695.
[296] Sun Q, Dam RM van, Spiegelman D, Heymsfield SB, Willett WC, Hu FB. Comparison of Dual-Energy X-Ray Absorptiometric and Anthropometric Measures of Adiposity in Relation to Adiposity-Related Biologic Factors. Am J Epidemiol 2010;172:1442–54. doi:10.1093/aje/kwq306.
[297] Batch BC, Shah SH, Newgard CB, Turer CB, Haynes C, Bain JR, et al. Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. Metabolism 2013;62:961–9. doi:10.1016/j.metabol.2013.01.007.
[298] McCourt HJ, Hunter SJ, Cardwell CR, Young IS, Murray LJ, Boreham CA, et al. Adiponectin multimers, body weight and markers of cardiovascular risk in adolescence: Northern Ireland Young Hearts Project. Int J Obes 2005 2013;37:1247–53. doi:10.1038/ijo.2012.214.
[299] Delisle H, Ntandou G, Sodjinou R, Couillard C, Després J-P. At-risk serum cholesterol profile at both ends of the nutrition spectrum in West African adults? The Benin study. Nutrients 2013;5:1366–83. doi:10.3390/nu5041366.
[300] Banerji MA, Faridi N, Atluri R, Chaiken RL, Lebovitz HE. Body Composition, Visceral Fat, Leptin, and Insulin Resistance in Asian Indian Men 1. J Clin Endocrinol Metab 1999;84:137–44.
[301] Deepa R, Sandeep S, Mohan V. Abdominal obesity, visceral fat and Type 2 diabetes-"Asian Indian Phenotype." Type 2006;2:138–52.
[302] Kamath SK, Hussain EA, Amin D, Mortillaro E, West B, Peterson CT, et al. Cardiovascular disease risk factors in 2 distinct ethnic groups: Indian and Pakistani compared with American premenopausal women. Am J Clin Nutr 1999;69:621–31.
[303] Raji A, Seely EW, Arky RA, Simonson DC. Body fat distribution and insulin resistance in healthy Asian Indians and Caucasians. J Clin Endocrinol Metab 2001;86:5366–71.

[304] Sharp PS, Mohan V, Levy JC, Mather HM, Kohner EM. Insulin resistance in patients of Asian Indian and European origin with non-insulin dependent diabetes. Horm Metab Res Horm Stoffwechselforschung Horm Metab 1987;19:84–5.
[305] Unni US, Ramakrishnan G, Raj T, Kishore RP, Thomas T, Vaz M, et al. Muscle mass and functional correlates of insulin sensitivity in lean young Indian men. Eur J Clin Nutr 2009;63:1206–12.
[306] Yajnik CS, Fall CHD, Coyaji KJ, Hirve SS, Rao S, Barker DJP, et al. Neonatal anthropometry: the thin–fat Indian baby. The Pune Maternal Nutrition Study. Int J Obes 2003;27:173–80. doi:10.1038/sj.ijo.802219.
[307] Misra A, Khurana L. Obesity-related non-communicable diseases: South Asians vs White Caucasians. Int J Obes 2011;35:167–87.
[308] Misra A, Vikram NK. Clinical and pathophysiological consequences of abdominal adiposity and abdominal adipose tissue depots. Nutrition 2003;19:457–66.
[309] Misra A, Sharma R, Pandey RM, Khanna N. Adverse profile of dietary nutrients, anthropometry and lipids in urban slum dwellers of northern India. Eur J Clin Nutr 2001;55:727–34.
[310] Misra A, Athiko D, Sharma R, Pandey RM, Khanna N. Non-obese hyperlipidemic Asian northern Indian males have adverse anthropometric profile. Nutr Metab Cardiovasc Dis NMCD 2002;12:178–83.
[311] Bruera E. ABC of palliative care. Anorexia, cachexia, and nutrition. BMJ 1997;315:1219.
[312] Friedenreich CM. Review of anthropometric factors and breast cancer risk. Eur J Cancer Prev 2001;10:15–32.
[313] Shu X-O, Jin F, Dai Q, Shi JR, Potter JD, Brinton LA, et al. Association of body size and fat distribution with risk of breast cancer among Chinese women. Int J Cancer 2001;94:449–55.
[314] Key TJA, Pike MC. The role of oestrogens and progestagens in the epidemiology and prevention of breast cancer. Eur J Cancer Clin Oncol 1988;24:29–43.
[315] Henderson BE, Ross RK, Judd HL, Krailo MD, Pike MC. Do regular ovulatory cycles increase breast cancer risk? Cancer 1985;56:1206–8.
[316] Lahmann PH, Lissner L, Gullberg B, Olsson H akan, Berglund G. A prospective study of adiposity and postmenopausal breast cancer risk: the Malmö Diet and Cancer Study. Int J Cancer 2003;103:246–52.
[317] Männistö S, Pietinen P, Pyy M, Palmgren J, Eskelinen M, Uusitupa M. Body-size indicators and risk of breast cancer according to menopause and estrogen-receptor status. Int J Cancer J Int Cancer 1996;68:8–13. doi:10.1002/(SICI)1097-0215(19960927)68:1<8::AID-IJC2>3.0.CO;2-V.
[318] Berstad P, Coates RJ, Bernstein L, Folger SG, Malone KE, Marchbanks PA, et al. A Case-Control Study of Body Mass Index and Breast Cancer Risk in White and African American Women. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2010;19:1532–44. doi:10.1158/1055-9965.EPI-10-0025.

[319] Kaaks R, Van Noord PA, Den Tonkelaar I, Peeters PH, Riboli E, Grobbee DE. Breast-cancer incidence in relation to height, weight and body-fat distribution in the Dutch DOM cohort. Int J Cancer 1998;76:647–51.
[320] Pike MC, Wu AH, Spicer DV, Lee S, Pearce CL. Estrogens, progestins, and risk of breast cancer. Progestins Mammary Gland, Springer; 2008, p. 127–50.
[321] Enas EA, Mohan V, Deepa M, Farooq S, Pazhoor S, Chennikkara H. The metabolic syndrome and dyslipidemia among Asian Indians: a population with high rates of diabetes and premature coronary artery disease. J Cardiometab Syndr 2007;2:267–75.
[322] Wu MH, Chou YC, Chou WY, Hsu GC, Chu CH, Yu CP, et al. Circulating levels of leptin, adiposity and breast cancer risk. Br J Cancer 2009;100:578–82.
[323] Misra A, Shrivastava U. Obesity and dyslipidemia in South Asians. Nutrients 2013;5:2708–33.
[324] Nelson LR, Bulun SE. Estrogen production and action. J Am Acad Dermatol 2001;45:S116–24.
[325] Hunter DJ, Willett WC. Diet, body size, and breast cancer. Epidemiol Rev 1993;15:110–32.
[326] Thomas HV, Key TJ, Allen DS, Moore JW, Dowsett M, Fentiman IS, et al. A prospective study of endogenous serum hormone concentrations and breast cancer risk in post-menopausal women on the island of Guernsey. Br J Cancer 1997;76:401–5.
[327] Wenten M, Gilliland FD, Baumgartner K, Samet JM. Associations of weight, weight change, and body mass with breast cancer risk in Hispanic and non-Hispanic white women. Ann Epidemiol 2002;12:435–44.
[328] Dapic V, Carvalho MA, Monteiro AN. Breast cancer susceptibility and the DNA damage response. Cancer Control 2005;12:127–36.
[329] Meyer UA. Pharmacogenetics–five decades of therapeutic lessons from genetic diversity. Nat Rev Genet 2004;5:669–76.
[330] Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93. doi:10.1038/nature05887.
[331] Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4. doi:10.1038/ng2075.
[332] Sapkota Y, Yasui Y, Lai R, Sridharan M, Robson PJ, Cass CE, et al. Identification of a breast cancer susceptibility locus at 4q31.22 using a genome-wide association study paradigm. PloS One 2013;8:e62550. doi:10.1371/journal.pone.0062550.
[333] Sehrawat B, Sridharan M, Ghosh S, Robson P, Cass CE, Mackey JR, et al. Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. Hum Genet 2011;130:529–37. doi:10.1007/s00439-011-0973-1.
[334] Kim H, Lee J-Y, Sung H, Choi J-Y, Park SK, Lee K-M, et al. A genome-wide association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study. Breast Cancer Res BCR 2012;14:R56. doi:10.1186/bcr3158.

[335] Long J, Cai Q, Sung H, Shi J, Zhang B, Choi J-Y, et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. PLoS Genet 2012;8:e1002532. doi:10.1371/journal.pgen.1002532.
[336] Steemers FJ, Gunderson KL. Illumina, Inc. Pharmacogenomics 2005;6:777–82. doi:10.2217/14622416.6.7.777.
[337] Yu W, Yesupriya A, Wulf A, Hindorff LA, Dowling N, Khoury MJ, et al. GWAS Integrator: a bioinformatics tool to explore human genetic associations reported in published genome-wide association studies. Eur J Hum Genet EJHG 2011;19:1095–9. doi:10.1038/ejhg.2011.91.
[338] Hindorff L, MacArthur J (European BI, Morales J (European BI, Junkins H, Hall P, Klemm A, et al. A Catalog of Published Genome-Wide Association Studies. Www.genome.gov/gwastudies n.d. www.genome.gov/gwastudies (accessed February 10, 2012).
[339] Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. Nat Genet 2008;40:124–5. doi:10.1038/ng0208-124.
[340] Yu W, Wulf A, Liu T, Khoury MJ, Gwinn M. Gene Prospector: an evidence gateway for evaluating potential susceptibility genes and interacting risk factors for human diseases. BMC Bioinformatics 2008;9:528. doi:10.1186/1471-2105-9-528.
[341] Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 2009;37:W600–5.
[342] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 1995:289–300.
[343] Purcell S. Http://pngu.mgh.harvard.edu/purcell/plink/ n.d. http://pngu.mgh.harvard.edu/purcell/plink/ (accessed December 16, 2013).
[344] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75. doi:10.1086/519795.
[345] Hoffman JI, Tucker R, Bridgett SJ, Clark MS, Forcada J, Slate J. Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: a case study in the Antarctic fur seal. Mol Ecol Resour 2012;12:861–72.
[346] Fan J-B, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, et al. Highly parallel SNP genotyping. Cold Spring Harb. Symp. Quant. Biol., vol. 68, Cold Spring Harbor Laboratory Press; 2003, p. 69–78.
[347] Grose R, Dickson C. Fibroblast growth factor signaling in tumorigenesis. Cytokine Growth Factor Rev 2005;16:179–86.
[348] Moffa AB, Ethier SP. Differential signal transduction of alternatively spliced FGFR2 variants expressed in human mammary epithelial cells. J Cell Physiol 2007;210:720–31.
[349] Koziczak M, Holbro T, Hynes NE. Blocking of FGFR signaling inhibits breast cancer cell proliferation through downregulation of D-type cyclins. Oncogene 2004;23:3501–8.

[350] Wang K, Li W-D, Zhang CK, Wang Z, Glessner JT, Grant SF, et al. A genome-wide association study on obesity and obesity-related traits. PloS One 2011;6:e18939.

[351] Jia C, Jia C, Cai Y, Ma Y, Fu D. Quantitative assessment of the effect of FGFR2 gene polymorphism on the risk of breast cancer. Breast Cancer Res Treat 2010;124:521–8. doi:10.1007/s10549-010-0872-5.

[352] Jara L, Gonzalez-Hormazabal P, Cerceño K, Di Capua GA, Reyes JM, Blanco R, et al. Genetic variants in FGFR2 and MAP3K1 are associated with the risk of familial and early-onset breast cancer in a South-American population. Breast Cancer Res Treat 2013;137:559–69. doi:10.1007/s10549-012-2359-z.

[353] Shan J, Mahfoudh W, Dsouza SP, Hassen E, Bouaouina N, Abdelhak S, et al. Genome-Wide Association Studies (GWAS) breast cancer susceptibility loci in Arabs: susceptibility and prognostic implications in Tunisians. Breast Cancer Res Treat 2012;135:715–24. doi:10.1007/s10549-012-2202-6.

[354] Zheng W, Cai Q, Signorello LB, Long J, Hargreaves MK, Deming SL, et al. Evaluation of 11 breast cancer susceptibility loci in African-American women. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2009;18:2761–4. doi:10.1158/1055-9965.EPI-09-0624.

[355] Long J, Zhang B, Signorello LB, Cai Q, Deming-Halverson S, Shrubsole MJ, et al. Evaluating genome-wide association study-identified breast cancer risk variants in African-American women. PloS One 2013;8:e58350. doi:10.1371/journal.pone.0058350.

[356] Slattery ML, John EM, Stern MC, Herrick J, Lundgreen A, Giuliano AR, et al. Associations with growth factor genes (FGF1, FGF2, PDGFB, FGFR2, NRG2, EGF, ERBB2) with breast cancer risk and survival: the Breast Cancer Health Disparities Study. Breast Cancer Res Treat 2013;140:587–601. doi:10.1007/s10549-013-2644-5.

[357] Meyer KB, Maia A-T, O'Reilly M, Teschendorff AE, Chin S-F, Caldas C, et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. PLoS Biol 2008;6:e108.

[358] Sun C, Olopade OI, Di Rienzo A. rs2981582 is associated with FGFR2 expression in normal breast. Cancer Genet Cytogenet 2010;197:193.

[359] Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, Brennan DJ, et al. FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. Carcinogenesis 2010;31:1417–23. doi:10.1093/carcin/bgq128.

[360] Gaudet MM, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C, et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. PLoS Genet 2010;6:e1001183. doi:10.1371/journal.pgen.1001183.

[361] Campa D, Kaaks R, Le Marchand L, Haiman CA, Travis RC, Berg CD, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. J Natl Cancer Inst 2011;103:1252–63.

[362] Chan M, Ji SM, Liaw CS, Yap YS, Law HY, Yoon CS, et al. Association of common genetic variants with breast cancer risk and clinicopathological characteristics in a Chinese population. Breast Cancer Res Treat 2012;136:209–20. doi:10.1007/s10549-012-2234-y.

[363] Cen Y-L, Qi M-L, Li H-G, Su Y, Chen L-J, Lin Y, et al. Associations of polymorphisms in the genes of FGFR2, FGF1, and RBFOX2 with breast cancer risk by estrogen/progesterone receptor status. Mol Carcinog 2013;52 Suppl 1:E52–9. doi:10.1002/mc.21979.

[364] Kibriya MG, Jasmine F, Argos M, Andrulis IL, John EM, Chang-Claude J, et al. A pilot genome-wide association study of early-onset breast cancer. Breast Cancer Res Treat 2009;114:463–77. doi:10.1007/s10549-008-0039-9.

[365] Zheng W, Long J, Gao Y-T, Li C, Zheng Y, Xiang Y-B, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009;41:324–8. doi:10.1038/ng.318.

[366] Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, et al. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. PLoS Genet 2010;6:e1001029.

[367] Guo H, Ming J, Liu C, Li Z, Zhang N, Cheng H, et al. A common polymorphism near the ESR1 gene is associated with risk of breast cancer: evidence from a case-control study and a meta-analysis. PloS One 2012;7:e52445. doi:10.1371/journal.pone.0052445.

[368] Yang Z, Shen J, Cao Z, Wang B. Association between a novel polymorphism (rs2046210) of the 6q25.1 locus and breast cancer risk. Breast Cancer Res Treat 2013;139:267–75. doi:10.1007/s10549-013-2494-1.

[369] Cai Q, Wen W, Qu S, Li G, Egan KM, Chen K, et al. Replication and functional genomic analyses of the breast cancer susceptibility locus at 6q25. 1 generalize its importance in women of chinese, Japanese, and European ancestry. Cancer Res 2011;71:1344–55.

[370] Zheng W, Zhang B, Cai Q, Sung H, Michailidou K, Shi J, et al. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. Hum Mol Genet 2013;22:2539–50. doi:10.1093/hmg/ddt089.

[371] Hein R, Maranian M, Hopper JL, Kapuscinski MK, Southey MC, Park DJ, et al. Comparison of 6q25 breast cancer hits from Asian and European Genome Wide Association Studies in the Breast Cancer Association Consortium (BCAC). PloS One 2012;7:e42380. doi:10.1371/journal.pone.0042380.

[372] Stevens KN, Vachon CM, Lee AM, Slager S, Lesnick T, Olswold C, et al. Common breast cancer susceptibility loci are associated with triple-negative breast cancer. Cancer Res 2011;71:6240–9. doi:10.1158/0008-5472.CAN-11-1266.

[373] Han W, Woo JH, Yu J-H, Lee M-J, Moon H-G, Kang D, et al. Common genetic variants associated with breast cancer in Korean women and differential susceptibility according to intrinsic subtype. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2011;20:793–8. doi:10.1158/1055-9965.EPI-10-1282.

[374] Murabito JM, Rosenberg CL, Finger D, Kreger BE, Levy D, Splansky GL, et al. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. BMC Med Genet 2007;8 Suppl 1:S6. doi:10.1186/1471-2350-8-S1-S6.

[375] Chen W, Song H, Zhong R, Zhu B, Guo H, Lou J, et al. Risk of GWAS-identified genetic variants for breast cancer in a Chinese population: a multiple interaction analysis. Breast Cancer Res Treat 2013;142:637–44. doi:10.1007/s10549-013-2775-8.
[376] Warren H, Dudbridge F, Fletcher O, Orr N, Johnson N, Hopper JL, et al. 9q31.2-rs865686 as a Susceptibility Locus for Estrogen Receptor-Positive Breast Cancer: Evidence from the Breast Cancer Association Consortium. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2012;21:1783–91. doi:10.1158/1055-9965.EPI-12-0526.
[377] Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, et al. Novel Breast Cancer Susceptibility Locus at 9q31.2: Results of a Genome-Wide Association Study. J Natl Cancer Inst 2011;103:425–35. doi:10.1093/jnci/djq563.
[378] Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet 2009;41:882–4.
[379] Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet 2009;41:885–90.
[380] Mokhtari D, Myers JW, Welsh N. MAPK kinase kinase-1 is essential for cytokine-induced c-Jun NH2-terminal kinase and nuclear factor-kappaB activation in human pancreatic islet cells. Diabetes 2008;57:1896–904. doi:10.2337/db07-1670.
[381] Dent P. ERK plays the baddie (again). Cancer Biol Ther 2013;14:997–8.
[382] Fan Y, Mao R, Yang J. NF-κB and STAT3 signaling pathways collaboratively link inflammation to cancer. Protein Cell 2013;4:176–85.
[383] Seki E, Brenner DA, Karin M. A liver full of JNK: signaling in regulation of cell function and disease pathogenesis, and clinical approaches. Gastroenterology 2012;143:307–20. doi:10.1053/j.gastro.2012.06.004.
[384] Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet 2008;4:e1000054. doi:10.1371/journal.pgen.1000054.
[385] Rebbeck TR, DeMichele A, Tran TV, Panossian S, Bunin GR, Troxel AB, et al. Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. Carcinogenesis 2009;30:269–74. doi:10.1093/carcin/bgn247.
[386] Zheng Q, Ye J, Wu H, Yu Q, Cao J. Association between Mitogen-Activated Protein Kinase Kinase Kinase 1 Polymorphisms and Breast Cancer Susceptibility: A Meta-Analysis of 20 Case-Control Studies. PLoS ONE 2014;9. doi:10.1371/journal.pone.0090771.
[387] Elks CE, Perry JRB, Sulem P, Chasman DI, Franceschini N, He C, et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nat Genet 2010;42:1077–85. doi:10.1038/ng.714.
[388] Delahanty RJ, Beeghly-Fadiel A, Long JR, Gao YT, Lu W, Xiang YB, et al. Evaluation of GWAS-identified genetic variants for age at menarche among Chinese women. Hum Reprod 2013;28:1135–43.

[389] Malis C, Rasmussen EL, Poulsen P, Petersen I, Christensen K, Beck-Nielsen H, et al. Total and regional fat distribution is strongly influenced by genetic factors in young and elderly twins. Obes Res 2005;13:2139–45.
[390] Farooqi IS. Genetic aspects of severe childhood obesity. Pediatr Endocrinol Rev PER 2006;3:528–36.
[391] Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, et al. Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. PLoS Genet 2009;5:e1000508. doi:10.1371/journal.pgen.1000508.
[392] Scherag A, Dina C, Hinney A, Vatin V, Scherag S, Vogel CI, et al. Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. PLoS Genet 2010;6:e1000916.
[393] Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 2009;41:18–24.
[394] Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet 2009;41:25–34.
[395] Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 2010;42:949–60.
[396] Burgdorf KS, Gjesing AP, Grarup N, Justesen JM, Sandholt CH, Witte DR, et al. Association studies of novel obesity-related gene variants with quantitative metabolic phenotypes in a population-based sample of 6,039 Danish individuals. Diabetologia 2012;55:105–13. doi:10.1007/s00125-011-2320-4.
[397] Hotta K, Nakamura M, Nakamura T, Matsuo T, Nakata Y, Kamohara S, et al. Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population. J Hum Genet 2009;54:727–31. doi:10.1038/jhg.2009.106.
[398] Kitamoto A, Kitamoto T, Mizusawa S, Teranishi H, So R, Matsuo T, et al. NUDT3 rs206936 is associated with body mass index in obese Japanese women. Endocr J 2013;60:991–1000.
[399] Mitchell JA, Hakonarson H, Rebbeck TR, Grant SFA. Obesity-susceptibility loci and the tails of the pediatric BMI distribution. Obes Silver Spring Md 2013;21:1256–60. doi:10.1002/oby.20319.
[400] Qi Q, Bray GA, Hu FB, Sacks FM, Qi L. Weight-loss diets modify glucose-dependent insulinotropic polypeptide receptor rs2287019 genotype effects on changes in body weight, fasting glucose, and insulin resistance: the Preventing Overweight Using Novel Dietary Strategies trial. Am J Clin Nutr 2012;95:506–13. doi:10.3945/ajcn.111.025270.
[401] Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010;42:937–48. doi:10.1038/ng.686.

[402] Cotsapas C, Speliotes EK, Hatoum IJ, Greenawalt DM, Dobrin R, Lum PY, et al. Common body mass index-associated variants confer risk of extreme obesity. Hum Mol Genet 2009;18:3502–7.
[403] Jiao H, Arner P, Hoffstedt J, Brodin D, Dubern B, Czernichow S, et al. Genome wide association study identifies KCNMA1 contributing to human obesity. BMC Med Genomics 2011;4:51. doi:10.1186/1755-8794-4-51.
[404] Hu H, He M-L, Tao R, Sun H-Y, Hu R, Zang W-J, et al. Characterization of ion channels in human preadipocytes. J Cell Physiol 2009;218:427–35. doi:10.1002/jcp.21617.
[405] Spalding KL, Arner E, Westermark PO, Bernard S, Buchholz BA, Bergmann O, et al. Dynamics of fat cell turnover in humans. Nature 2008;453:783–7. doi:10.1038/nature06902.
[406] Hofker M, Wijmenga C. A supersized list of obesity genes. Nat Genet 2009;41:139–40. doi:10.1038/ng0209-139.
[407] Shimada T, Watanabe J, Kawajiri K, Sutter TR, Guengerich FP, Gillam EM, et al. Catalytic properties of polymorphic human cytochrome P450 1B1 variants. Carcinogenesis 1999;20:1607–14.
[408] Thyagarajan B, Brott M, Mink P, Folsom AR, E Anderson K, Oetting WS, et al. CYP1B1 and CYP19 gene polymorphisms and breast cancer incidence: no association in the ARIC study. Cancer Lett 2004;207:183–9.
[409] Paracchini V, Pedotti P, Raimondi S, Garte S, Bradlow HL, Sepkovic DW, et al. A common CYP1B1 polymorphism is associated with 2-OHE1/16-OHE1 urinary estrone ratio. Clin Chem Lab Med 2005;43:702–6.
[410] Paracchini V, Raimondi S, Gram IT, Kang D, Kocabas NA, Kristensen VN, et al. Meta-and pooled analyses of the cytochrome P-450 1B1 Val432Leu polymorphism and breast cancer: a HuGE–GSEC review. Am J Epidemiol 2007;165:115–25.
[411] Garcia-Closas M, Kristensen V, Langerød A, Qi Y, Yeager M, Burdett L, et al. Common genetic variation in TP53 and its flanking genes, WDR79 and ATP1B2, and susceptibility to breast cancer. Int J Cancer 2007;121:2532–8.
[412] Egan KM, Cai Q, Shu X-O, Jin F, Zhu T-L, Dai Q, et al. Genetic Polymorphisms in GSTM1, GSTP1, and GSTT1 and the Risk for Breast Cancer Results from the Shanghai Breast Cancer Study and Meta-Analysis. Cancer Epidemiol Biomarkers Prev 2004;13:197–204.
[413] Vogl FD, Taioli E, Maugard C, Zheng W, Pinto LFR, Ambrosone C, et al. Glutathione S-transferases M1, T1, and P1 and breast cancer: a pooled analysis. Cancer Epidemiol Biomarkers Prev 2004;13:1473–9.
[414] Yao L, Fang F, Zhong Y, Yu L. The association between two polymorphisms of eNOS and breast cancer risk: a meta-analysis. Breast Cancer Res Treat 2010;124:223–7.
[415] Johnson RD, Liu N, Jasin M. Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination. Nature 1999;401:397–9.
[416] Deans B, Griffin CS, O'Regan P, Jasin M, Thacker J. Homologous recombination deficiency leads to profound genetic instability in cells derived from Xrcc2-knockout mice. Cancer Res 2003;63:8181–7.

[417] Griffin CS, Simpson PJ, Wilson CR, Thacker J. Mammalian recombination-repair genes XRCC2 and XRCC3 promote correct chromosome segregation. Nat Cell Biol 2000;2:757–61.

[418] Huang H, Fletcher L, Beeharry N, Daniel R, Kao G, Yen TJ, et al. Abnormal cytokinesis after X-irradiation in tumor cells that override the G2 DNA damage checkpoint. Cancer Res 2008;68:3724–32.

[419] Lin W-Y, Camp NJ, Cannon-Albright LA, Allen-Brady K, Balasubramanian S, Reed MW, et al. A role for XRCC2 gene polymorphisms in breast cancer risk and survival. J Med Genet 2011;48:477–84.

[420] Langsenlehner U, Yazdani-Biuki B, Eder T, Renner W, Wascher TC, Paulweber B, et al. The cyclooxygenase-2 (PTGS2) 8473T>C polymorphism is associated with breast cancer risk. Clin Cancer Res Off J Am Assoc Cancer Res 2006;12:1392–4. doi:10.1158/1078-0432.CCR-05-2055.

[421] Abraham JE, Harrington P, Driver KE, Tyrer J, Easton DF, Dunning AM, et al. Common polymorphisms in the prostaglandin pathway genes and their association with breast cancer susceptibility and survival. Clin Cancer Res 2009;15:2181–91.

[422] Piranda DN, Festa-Vasconcellos JS, Amaral LM, Bergmann A, Vianna-Jorge R. Polymorphisms in regulatory regions of cyclooxygenase-2 gene and breast cancer risk in Brazilians: a case-control study. BMC Cancer 2010;10:613. doi:10.1186/1471-2407-10-613.

[423] Vogel U, Christensen J, Nexø BA, Wallin H akan, Friis S, Tjønneland A. Peroxisome profilerator-activated receptorγ2 Pro12Ala, interaction with alcohol intake and NSAID use, in relation to risk of breast cancer in a prospective study of Danes. Carcinogenesis 2006;28:427–34.

[424] Cox DG, Buring J, Hankinson SE, Hunter DJ. A polymorphism in the 3' untranslated region of the gene encoding prostaglandin endoperoxide synthase 2 is not associated with an increase in breast cancer risk: a nested case-control study. Breast Cancer Res 2007;9:R3. doi:10.1186/bcr1635.

[425] Zhu W, Wei B, Shan X, Liu P. -765G>C and 8473T>C polymorphisms of COX-2 and cancer risk: a meta-analysis based on 33 case-control studies. Mol Biol Rep 2010;37:277–88. doi:10.1007/s11033-009-9685-1.

[426] Kuschel B, Auranen A, McBride S, Novik KL, Antoniou A, Lipscombe JM, et al. Variants in DNA double-strand break repair genes and breast cancer susceptibility. Hum Mol Genet 2002;11:1399–407.

[427] Figueiredo JC, Knight JA, Briollais L, Andrulis IL, Ozcelik H. Polymorphisms XRCC1-R399Q and XRCC3-T241M and the risk of breast cancer at the Ontario site of the Breast Cancer Family Registry. Cancer Epidemiol Biomarkers Prev 2004;13:583–91.

[428] Försti A, Angelini S, Festa F, Sanyal S, Zhang Z, Grzybowska E, et al. Single nucleotide polymorphisms in breast cancer. Oncol Rep 2004;11:917–22.

[429] Han J, Hankinson SE, Ranu H, De Vivo I, Hunter DJ. Polymorphisms in DNA double-strand break repair genes and breast cancer risk in the Nurses' Health Study. Carcinogenesis 2004;25:189–95.

[430] Jacobsen NR, Nexø BA, Olsen A, Overvad K, Wallin H akan, Tjønneland A, et al. No association between the DNA repair gene XRCC3 T241M polymorphism and risk of skin cancer and breast cancer. Cancer Epidemiol Biomarkers Prev 2003;12:584–5.
[431] Smith TR, Levine EA, Perrier ND, Miller MS, Freimanis RI, Lohman K, et al. DNA-repair genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2003;12:1200–4.
[432] Smith TR, Miller MS, Lohman K, Lange EM, Case LD, Mohrenweiser HW, et al. Polymorphisms of< i> XRCC1</i> and< i> XRCC3</i> genes and susceptibility to breast cancer. Cancer Lett 2003;190:183–90.
[433] Webb PM, Hopper JL, Newman B, Chen X, Kelemen L, Giles GG, et al. Double-strand break repair gene polymorphisms and risk of breast or ovarian cancer. Cancer Epidemiol Biomarkers Prev 2005;14:319–23.
[434] García-Closas M, Egan KM, Newcomb PA, Brinton LA, Titus-Ernstoff L, Chanock S, et al. Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. Hum Genet 2006;119:376–88.
[435] Han S, Zhang H-T, Wang Z, Xie Y, Tang R, Mao Y, et al. DNA repair gene XRCC3 polymorphisms and cancer risk: a meta-analysis of 48 case–control studies. Eur J Hum Genet 2006;14:1136–44.
[436] Lee S-A, Lee K-M, Park SK, Choi J-Y, Kim B, Nam J, et al. Genetic polymorphism of XRCC3 Thr241Met and breast cancer risk: case-control study in Korean women and meta-analysis of 12 studies. Breast Cancer Res Treat 2007;103:71–6.
[437] Consortium BCA. Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. J Natl Cancer Inst 2006;98:1382–96.
[438] Manuguerra M, Saletta F, Karagas MR, Berwick M, Veglia F, Vineis P, et al. XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review. Am J Epidemiol 2006;164:297–302.
[439] Economopoulos KP, Sergentanis TN. XRCC3 Thr241Met polymorphism and breast cancer risk: a meta-analysis. Breast Cancer Res Treat 2010;121:439–43.
[440] Zhao Y, Meng X-M, Wei Y-J, Zhao X-W, Liu D-Q, Cao H-Q, et al. Cloning and characterization of a novel cardiac-specific kinase that interacts specifically with cardiac troponin I. J Mol Med Berl Ger 2003;81:297–304. doi:10.1007/s00109-003-0427-x.
[441] Engelking LJ, Liang G, Hammer RE, Takaishi K, Kuriyama H, Evers BM, et al. Schoenheimer effect explained–feedback regulation of cholesterol synthesis in mice mediated by Insig proteins. J Clin Invest 2005;115:2489–98.
[442] Goldstein JL, DeBose-Boyd RA, Brown MS. Protein sensors for membrane sterols. Cell 2006;124:35–46.
[443] Beckers S, Peeters AV, De Freitas F, Mertens IL, Verhulst SL, Haentjens D, et al. Association study and mutation analysis of adiponectin shows association of variants in APM1 with complex obesity in women. Ann Hum Genet 2009;73:492–501.

[444] Specchia C, Scott K, Fortina P, Devoto M, Falkner B. Association of a polymorphic variant of the adiponectin gene with Insulin resistance in African Americans. Clin Transl Sci 2008;1:194–9.
[445] Peters KE, Beilby J, Cadby G, Warrington NM, Bruce DG, Davis WA, et al. A comprehensive investigation of variants in genes encoding adiponectin (ADIPOQ) and its receptors (ADIPOR1/R2), and their association with serum adiponectin, type 2 diabetes, insulin resistance and the metabolic syndrome. BMC Med Genet 2013;14:15.
[446] Wassel CL, Pankow JS, Jacobs DR, Steffes MW, Li NA, Schreiner PJ. Variants in the adiponectin gene and serum adiponectin: the Coronary Artery Development in Young Adults (CARDIA) Study. Obesity 2010;18:2333–8.
[447] Ying S-Y, Lin S-L. Intron-derived microRNAs—fine tuning of gene functions. Gene 2004;342:25–8.

Annexure

Annexure 1: Summary of 384 Custom SNP Panel

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
10	ACTA2	rs2234767	Intergenic	0.21		G:A	Candidate SNPs	Breast Cancer
3	AGTR1	rs5186	UTR	0.06		A:C	Candidate SNPs	Breast Cancer
7	AKAP9	rs6964587	Coding	0.39		G:T	Candidate SNPs	Breast Cancer
2	ALS2CR12	rs17468277	Coding	0.04		C:T	Candidate SNPs	Breast Cancer
11	ATM	rs1801516	Complex	0.07		G:A	Candidate SNPs	Breast Cancer
17	ATP1B2	rs1641535 [b]	Intergenic	0.22		G:A	Candidate SNPs	Breast Cancer
17	ATP1B2	rs1641536 [b]	Intergenic	0.22		G:A	Candidate SNPs	Breast Cancer
20	AURKA	rs2273535	Coding	0.32		A:T	Candidate SNPs	Breast Cancer
2	CASP8	rs1045485	Complex	0.04		G:C	Candidate SNPs	Breast Cancer
6	CDKN1A	rs1801270	Coding	0.13		C:A	Candidate SNPs	Breast Cancer
22	COMT	rs4680	Complex	0.42		G:A	Candidate SNPs	Breast Cancer
2	CTLA4	rs231775	Coding	0.33		A:G	Candidate SNPs	Breast Cancer
10	CXCL12	rs1801157	UTR	0.23		G:A	Candidate SNPs	Breast Cancer
10	CYP17A1	rs4919682	Intergenic	0.07		C:T	Candidate SNPs	Breast Cancer
10	CYP17A1	rs4919687	Intron	0.1		G:A	Candidate SNPs	Breast Cancer
2	CYP1B1	rs1056836 [b]	UTR	0.21		C:G	Candidate SNPs	Breast Cancer
18	ENOSF1	rs34489327	Complex	NA	Diffused cluster	NA	Candidate SNPs	Breast Cancer
17	ERBB2	rs1136201	Complex	0.13		A:G	Candidate SNPs	Breast Cancer
16	ERCC4	rs1800067	Coding	0.03		G:A	Candidate SNPs	Breast Cancer
14	ESR2	rs4986938	UTR	0.27		G:A	Candidate SNPs	Breast Cancer
10	FGFR2	rs2420946 [b]	Intron	0.4		C:T	Candidate SNPs	Breast Cancer
10	GATA3	rs570613	Intron	0.23		A:G	Candidate SNPs	Breast Cancer
11	GSTP1	rs1695 [b]	Coding	0.26		A:G	Candidate SNPs	Breast Cancer
12	HOXC9	rs11614913	Intergenic	0.27		C:T	Candidate SNPs	Breast Cancer
12	IFNG	rs2430561	Intron	NA	Diffused cluster	NA	Candidate SNPs	Breast Cancer
7	LOC100129619	rs2854744	Intergenic	0.49		C:A	Candidate SNPs	Breast Cancer
1	MTHFR	rs1801133	Coding	0.15		C:T	Candidate SNPs	Breast Cancer
8	NBN	rs1805794	Coding	0.43		C:G	Candidate SNPs	Breast Cancer
7	NOS3	rs1799983	Coding	0.18		G:T	Candidate SNPs	Breast Cancer
7	NOS3	rs2070744 [b]	Intron	0.23		T:C	Candidate SNPs	Breast Cancer
16	NQO1	rs1800566	Coding	0.33		C:T	Candidate SNPs	Breast Cancer
11	PGR	rs1042838	Coding	0.07		G:T	Candidate SNPs	Breast Cancer
7	POR	rs10262966	Coding	0.07		A:G	Candidate SNPs	Breast Cancer
1	PTGS2	rs5275 [b]	UTR	0.38		T:C	Candidate SNPs	Breast Cancer
15	RAD51	rs1801320	UTR	0.13		G:C	Candidate SNPs	Breast Cancer
9	TGFBR1	rs11568785	Intron	0.03		A:G	Candidate SNPs	Breast Cancer
6	TNF	rs1800629	Intergenic	0.05		G:A	Candidate SNPs	Breast Cancer
17	TP53	rs1042522	Coding	0.48		C:G	Candidate SNPs	Breast Cancer

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
17	TP53	rs1625895	Intron	0.18		G:A	Candidate SNPs	Breast Cancer
17	TP53	rs17878362	Intron	NA	Diffused cluster	NA	Candidate SNPs	Breast Cancer
17	WDR79	rs2287499 [b]	Coding	0.26		C:G	Candidate SNPs	Breast Cancer
7	XRCC2	rs3218408 [b]	Intron	0.2		T:G	Candidate SNPs	Breast Cancer
14	XRCC3	rs1799794	UTR	0.41		A:G	Candidate SNPs	Breast Cancer
14	XRCC3	rs1799796	Intergenic	0.19		A:G	Candidate SNPs	Breast Cancer
14	XRCC3	rs861539 [b]	Coding	0.19		C:T	Candidate SNPs	Breast Cancer
11	ATM	rs3092829	Intron	NA	MAF < 1%	NA	Candidate SNPs	Others
13	BIVM	rs4743	UTR	0.39		C:G	Candidate SNPs	Others
17	CBX1	rs2240121	Intron	0.11		T:C	Candidate SNPs	Others
17	CBX1	rs2240123 [b]	Intron	0.25		C:T	Candidate SNPs	Others
17	CBX1	rs7215582	Intron	0.36		G:A	Candidate SNPs	Others
17	CBX1	rs8065670	Intron	0.19		A:G	Candidate SNPs	Others
1	CSDE1	rs10489525 [b]	Intron	0.48		A:G	Candidate SNPs	Others
1	CSDE1	rs6668128	Intron	0.21		G:A	Candidate SNPs	Others
2	DNAJC27	rs1172294	UTR	0.47		A:G	Candidate SNPs	Others
2	DNAJC27	rs13388020	Intergenic	0.05		G:C	Candidate SNPs	Others
2	DNAJC27	rs17046742	Intergenic	0.18		G:A	Candidate SNPs	Others
2	DNAJC27	rs17046751	UTR	0.07		A:G	Candidate SNPs	Others
2	DNAJC27	rs6545841	Intergenic	NA	Not in HWE	NA	Candidate SNPs	Others
13	ERCC5	rs1047768	Coding	0.46		T:C	Candidate SNPs	Others
13	ERCC5	rs4150355	Intron	0.34		C:T	Candidate SNPs	Others
13	ERCC5	rs751402	UTR	0.3		C:T	Candidate SNPs	Others
13	ERCC5	rs873601	UTR	0.27		A:G	Candidate SNPs	Others
3	GPX1	rs3811699	Intergenic	0.2		A:G	Candidate SNPs	Others
6	GSTA1	rs4715326	Intron	0.35		T:C	Candidate SNPs	Others
12	IFNG	rs2069705	Intergenic	0.3		T:C	Candidate SNPs	Others
12	IFNG	rs2069727	Intergenic	0.38		A:G	Candidate SNPs	Others
12	IFNG	rs3181032	Intergenic	0.22		T:G	Candidate SNPs	Others
5	IL13	rs2243250	Intergenic	0.19		C:T	Candidate SNPs	Others
2	IL1B	rs1143627	Intergenic	0.4		C:T	Candidate SNPs	Others
2	IL1B	rs1143633	Intron	0.27		G:A	Candidate SNPs	Others
2	IL1B	rs12621220	Intergenic	0.36		C:T	Candidate SNPs	Others
2	IL1B	rs3136558	Intron	0.16		T:C	Candidate SNPs	Others
2	IL8RA	rs1008563	Intergenic	0.35		C:T	Candidate SNPs	Others
2	IL8RA	rs16858808	Coding	0.03		C:T	Candidate SNPs	Others
2	IL8RA	rs16858811	Coding	0.04		T:G	Candidate SNPs	Others
2	IL8RB	rs2854386	Intergenic	0.16		G:C	Candidate SNPs	Others
14	INSM2	rs2233406	Intergenic	0.28		C:T	Candidate SNPs	Others

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
14	INSM2	rs3138045	Intergenic	0.19		A:G	Candidate SNPs	Others
14	NFKBIA	rs3138055	Intergenic	0.27		A:G	Candidate SNPs	Others
14	NFKBIA	rs4982269	Intergenic	0.29		T:C	Candidate SNPs	Others
14	NFKBIA	rs8904	UTR	0.45		C:T	Candidate SNPs	Others
1	NRAS	rs6671984	Intron	0.15		T:C	Candidate SNPs	Others
17	PHB	rs2233660 [b]	Intergenic	0.12		T:C	Candidate SNPs	Others
17	PHB	rs2233669	Intron	0.3		A:G	Candidate SNPs	Others
17	PHB	rs2277637	Intron	0.03		A:G	Candidate SNPs	Others
17	PHB	rs2898883	Intron	0.12		G:A	Candidate SNPs	Others
17	PHB	rs6917	UTR	NA	Diffused cluster	NA	Candidate SNPs	Others
20	SNRPB	rs2143862	Intergenic	0.19		C:T	Candidate SNPs	Others
20	SNRPB	rs6049212	Intergenic	0.17		G:A	Candidate SNPs	Others
20	SNRPB	rs6049288	UTR	0.2		G:T	Candidate SNPs	Others
20	SNRPB	rs6138178	Intron	NA	Not in HWE	NA	Candidate SNPs	Others
1	TNNI3K	rs11581900	Intron	0.33		A:C	Candidate SNPs	Others
1	TNNI3K	rs274586 [b]	Intron	0.49		G:C	Candidate SNPs	Others
1	TNNI3K	rs489990 [b]	Intron	0.14		A:G	Candidate SNPs	Others
1	TNNI3K	rs7515072	Intron	0.24		G:A	Candidate SNPs	Others
1	TNNI3K	rs7553158	Intron	0.41		C:T	Candidate SNPs	Others
9	9q31.2	rs865686 [b]	Intergenic	0.13		T:G	GWAS	Breast Cancer
13	ABCC4	rs1926657	Intron	0.31		C:T	GWAS	Breast Cancer
10	ANKRD16	rs2380205	Intergenic	NA	Diffused cluster	NA	GWAS	Breast Cancer
10	ATE1	rs10510102	Intron	0.1		A:G	GWAS	Breast Cancer
7	AUTS2	rs10263639	Intergenic	0.13		T:C	GWAS	Breast Cancer
11	BARX2	rs7107217	Intergenic	0.34		C:A	GWAS	Breast Cancer
2	BCL11A	rs10490113	Intergenic	0.16		A:C	GWAS	Breast Cancer
19	C19orf62	rs8170	Coding	0.1		C:T	GWAS	Breast Cancer
6	C6orf97	rs3734805	UTR	0.08		A:C	GWAS	Breast Cancer
6	C6orf97	rs3757318	Intron	0.07		G:A	GWAS	Breast Cancer
11	CCND1	rs614367	Intergenic	0.16		C:T	GWAS	Breast Cancer
9	CDKN2BAS	rs1011970	Intron	0.27		G:T	GWAS	Breast Cancer
17	COL1A1	rs2075555	Intron	NA	MAF < 1%	NA	GWAS	Breast Cancer
6	ESR1	rs2046210 [b]	Intergenic	0.37		C:T	GWAS	Breast Cancer
6	ESR1	rs9383951	Intron	NA	MAF < 1%	NA	GWAS	Breast Cancer
8	FAM84B	rs13281615	Intergenic	0.49		G:A	GWAS	Breast Cancer
8	FAM84B	rs1562430	Intergenic	0.23		A:G	GWAS	Breast Cancer
15	FBN1	rs1876206	Intron	0.13		A:G	GWAS	Breast Cancer
5	FGF10	rs10941679	Intergenic	0.39		A:G	GWAS	Breast Cancer
5	FGF10	rs4415084	Intergenic	0.48		T:C	GWAS	Breast Cancer

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
10	FGFR2	rs1219648 [b]	Intron	0.39		A:G	GWAS	Breast Cancer
10	FGFR2	rs2981575 [b]	Intron	0.4		T:C	GWAS	Breast Cancer
10	FGFR2	rs2981579	Intron	0.41		C:T	GWAS	Breast Cancer
10	FGFR2	rs2981582 [b]	Intron	0.34		C:T	GWAS	Breast Cancer
16	GLG1	rs10871290 [b]	Intergenic	0.25		T:C	GWAS	Breast Cancer
21	GRIK1	rs458685	Intergenic	0.12		T:C	GWAS	Breast Cancer
5	HCN1	rs981782	Intron	0.2		T:G	GWAS	Breast Cancer
7	LOC647017	rs2048672	Intron	0.44		G:T	GWAS	Breast Cancer
1	LOC647121	rs11249433	Intron	0.17		T:C	GWAS	Breast Cancer
11	LSP1	rs3817198	Intron	0.37		T:C	GWAS	Breast Cancer
5	MAP3K1	rs16886165 [b]	Intergenic	0.36		T:G	GWAS	Breast Cancer
5	MAP3K1	rs889312 [b]	Intergenic	0.41		A:C	GWAS	Breast Cancer
5	MAT2B	rs6556756 [b]	Intergenic	0.14		T:G	GWAS	Breast Cancer
14	RAD51L1	rs10483813	Intron	NA	Diffused cluster	NA	GWAS	Breast Cancer
14	RAD51L1	rs999737	Intron	0.12		C:T	GWAS	Breast Cancer
6	RNF146	rs2180341	Intron	0.41		A:G	GWAS	Breast Cancer
5	ROPN1L	rs1092913	Intergenic	0.33		G:A	GWAS	Breast Cancer
3	SLC4A7	rs4973768	UTR	0.45		C:T	GWAS	Breast Cancer
17	STXBP4	rs6504950	Intron	0.16		G:A	GWAS	Breast Cancer
18	TCF4	rs1978503	Intergenic	0.12		A:G	GWAS	Breast Cancer
5	TERT	rs10069690	Intron	0.29		C:T	GWAS	Breast Cancer
11	TNNT3	rs909116	Intron	0.37		T:C	GWAS	Breast Cancer
2	TNP1	rs13387042	Intergenic	0.49		A:G	GWAS	Breast Cancer
16	TOX3	rs3112612	Intergenic	0.47		C:T	GWAS	Breast Cancer
16	TOX3	rs4784227	Intergenic	0.23		C:T	GWAS	Breast Cancer
16	TOX3	rs3803662	Intergenic	0.29		C:T	GWAS	Breast Cancer
6	UST	rs9485372	Intergenic	0.2		G:A	GWAS	Breast Cancer
10	ZMIZ1	rs704010	Intron	0.29		G:A	GWAS	Breast Cancer
10	ZNF365	rs10822013	Intron	0.5		T:C	GWAS	Breast Cancer
10	ZNF365	rs10995190	Intron	0.08		G:A	GWAS	Breast Cancer
19	ZNF577	rs10411161 [b]	Intron	0.46		C:T	GWAS	Breast Cancer
16	ADCY9	rs2444217	Intron	0.4		C:T	GWAS	BMI
11	BDNF	rs925946	Intergenic	0.38		G:T	GWAS	BMI
3	CADM2	rs13078807	Intron	0.08		A:G	GWAS	BMI
2	DNAJC27	rs713586	Intergenic	0.45		T:C	GWAS	BMI
3	ETV5	rs7647305	Intergenic	0.23		C:T	GWAS	BMI
3	ETV5	rs9816226	Intergenic	NA	Call frequency < 95%	NA	GWAS	BMI
12	FAIM2	rs7138803	Intergenic	0.36		G:A	GWAS	BMI
2	FANCL	rs887912	Intergenic	0.13		G:A	GWAS	BMI

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
16	FTO	rs6499640	Intron	0.46		G:A	GWAS	BMI
14	G2E3	rs11847697	Intergenic	0.09		C:T	GWAS	BMI
4	GNPDA2	rs10938397	Intergenic	0.38		A:G	GWAS	BMI
5	HMGCR	rs2112347	Intergenic	0.42		G:T	GWAS	BMI
8	HNF4G	rs2922763 [b]	Intergenic	0.3		A:C	GWAS	BMI
19	KCTD15	rs11084753	Intergenic	NA	Not in HWE	NA	GWAS	BMI
19	KCTD15	rs29941	Intergenic	0.44		C:T	GWAS	BMI
2	LRP1B	rs2890652	Intergenic	NA	Diffused cluster	NA	GWAS	BMI
15	MAP2K5	rs2241423	Intron	0.33		G:A	GWAS	BMI
18	MC4R	rs12970134	Intergenic	0.33		G:A	GWAS	BMI
18	MC4R	rs17782313	Intergenic	0.33		T:C	GWAS	BMI
18	MC4R	rs571312	Intergenic	0.34		G:T	GWAS	BMI
1	NEGR1	rs2568958	Intergenic	0.35		A:G	GWAS	BMI
1	NEGR1	rs2815752	Intergenic	0.36		T:C	GWAS	BMI
14	NRXN3	rs10150332	Intron	0.11		T:C	GWAS	BMI
6	NUDT3	rs206936	Intron	0.42		A:G	GWAS	BMI
1	PTBP2	rs1555543	Intergenic	0.45		C:A	GWAS	BMI
19	QPCTL	rs2287019 [b]	Intron	0.15		C:T	GWAS	BMI
1	SEC16B	rs10913469	Intron	0.18		T:C	GWAS	BMI
1	SEC16B	rs543874	Intergenic	0.15		A:G	GWAS	BMI
16	SH2B1	rs7359397	Intron	0.17		C:T	GWAS	BMI
4	SLC39A8	rs13107325	Complex	NA	MAF < 1%	NA	GWAS	BMI
11	STK33	rs4929949	Intron	0.34		T:C	GWAS	BMI
6	TFAP2B	rs987237 [b]	Intron	0.21		A:G	GWAS	BMI
2	TMEM18	rs2867125 [b]	Intergenic	0.2		G:A	GWAS	BMI
2	TMEM18	rs7561317	Intergenic	0.13		G:A	GWAS	BMI
1	TNNI3K	rs1514175	Intron	0.48		T:C	GWAS	BMI
8	TRHR	rs7832552	Intron	0.41		C:T	GWAS	BMI
19	ZC3H4	rs3810291	UTR	0.37		G:A	GWAS	BMI
14	BEGAIN	rs6575793 [b]	Intron	0.48		T:C	GWAS	Menstruation & menopause
19	BRSK1	rs1172822	Intron	0.47		T:C	GWAS	Menstruation & menopause
11	BSX	rs6589964	Intergenic	0.45		A:C	GWAS	Menstruation & menopause
17	CA10	rs9635759	Intergenic	0.29		G:A	GWAS	Menstruation & menopause
3	CADM2	rs7642134	Intergenic	0.41		G:A	GWAS	Menstruation & menopause
2	CCDC85A	rs17268785	Intron	0.14		A:G	GWAS	Menstruation & menopause
6	CENPW	rs1361108	Intergenic	0.3		T:C	GWAS	Menstruation & menopause
19	CRTC1	rs10423674	Intron	NA	Diffused cluster	NA	GWAS	Menstruation & menopause
18	FUSSEL18	rs1398217	Intron	0.45		C:G	GWAS	Menstruation & menopause
11	GAB2	rs10899489	UTR	0.27		C:A	GWAS	Menstruation & menopause

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
3	GSF11	rs6438424	Intergenic	0.42		A:C	GWAS	Menstruation & menopause
6	HACE1	rs7759938	Intergenic	0.29		T:C	GWAS	Menstruation & menopause
7	INHBA	rs1079866	Intergenic	0.12		C:G	GWAS	Menstruation & menopause
2	KCNJ3	rs17188434	Intergenic	0.03		T:C	GWAS	Menstruation & menopause
5	KDM3B	rs757647	Intron	0.31		C:T	GWAS	Menstruation & menopause
3	KLHDC8B	rs7617480	Intron	0.06		C:A	GWAS	Menstruation & menopause
6	LIN28B	rs314277	Intron	0.14		C:A	GWAS	Menstruation & menopause
6	LIN28B,	rs314280	Intron	0.38		C:T	GWAS	Menstruation & menopause
6	MCHR2	rs4840086	Intergenic	0.32		A:G	GWAS	Menstruation & menopause
20	MCM8	rs16991615	Complex	0.02		G:A	GWAS	Menstruation & menopause
20	MCM8	rs236114	Intron	0.19		G:A	GWAS	Menstruation & menopause
16	NFAT5	rs1364063	Intergenic	0.3		T:C	GWAS	Menstruation & menopause
16	PARN	rs1659127	Intergenic	0.32		G:A	GWAS	Menstruation & menopause
20	PCSK2	rs852069	Intergenic	NA	Diffused cluster	NA	GWAS	Menstruation & menopause
5	PHF15	rs13187289	Intergenic	0.11		C:G	GWAS	Menstruation & menopause
11	PHF21A	rs16938437	Intron	0.1		C:T	GWAS	Menstruation & menopause
8	PKIA	rs7821178	Intergenic	0.47		C:A	GWAS	Menstruation & menopause
2	PLCL1	rs12617311	Intergenic	0.49		G:A	GWAS	Menstruation & menopause
11	RASSF10	rs900145	Intergenic	NA	Diffused cluster	NA	GWAS	Menstruation & menopause
3	RBM6	rs6762477	Intron	0.17		A:G	GWAS	Menstruation & menopause
1	RXRG	rs466639	Intron	0.23		C:T	GWAS	Menstruation & menopause
1	SEC16B/	rs633715	Intergenic	0.14		T:C	GWAS	Menstruation & menopause
13	SOX1	rs7333181	Intergenic	0.09		G:A	GWAS	Menstruation & menopause
13	SOX1	rs9555810	Intergenic	0.26		C:G	GWAS	Menstruation & menopause
6	SYCP2L	rs2153157	Intron	0.44		C:T	GWAS	Menstruation & menopause
3	TMEM108	rs6439371	Intergenic	0.43		A:G	GWAS	Menstruation & menopause
2	TMEM18	rs2947411	Intergenic	0.21		C:T	GWAS	Menstruation & menopause
9	TMEM38B	rs2090409	Intergenic	0.16		G:T	GWAS	Menstruation & menopause
9	TMEM38B	rs7861820	Intergenic	0.47		T:C	GWAS	Menstruation & menopause
11	TRIM66	rs4929923	UTR	NA	Diffused cluster	NA	GWAS	Menstruation & menopause
9	ZNF483	rs10980926 [b]	Intron	0.48		G:A	GWAS	Menstruation & menopause
6	ARG1	rs2807278	Intergenic	0.26		T:C	GWAS	Obesity
11	BDNFOS	rs988712 [b]	Intron	0.27		G:T	GWAS	Obesity
2	CWC22	rs16867321 [a]	Intergenic	0.33		C:T	GWAS	Obesity
16	DYNLRB2	rs1424233	Intergenic	0.31		A:G	GWAS	Obesity
2	EML6	rs6726292	Intron	0.48		A:G	GWAS	Obesity
4	FAT1	rs925642 [b]	Intergenic	0.45		A:G	GWAS	Obesity
5	FBN2	rs374748 [b]	Intron	0.09		A:G	GWAS	Obesity
16	FTO	rs1121980	Intron	0.4		C:T	GWAS	Obesity

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
2	INHBB	rs7581710	Intergenic	0.33		A:G	GWAS	Obesity
6	ITPR3	rs999943	Intron	0.28		T:C	GWAS	Obesity
10	KCNMA1	rs2116830 [b]	UTR	0.03		C:A	GWAS	Obesity
7	LHFPL3	rs10953454 [b]	Intron	0.08		G:A	GWAS	Obesity
20	MACROD2	rs6110577	Intron	0.19		T:C	GWAS	Obesity
18	MC4R	rs10871777	Intergenic	NA	Diffused cluster	NA	GWAS	Obesity
6	MLN	rs2274459 [b]	Intergenic	0.06		G:A	GWAS	Obesity
11	MUC15	rs12295638	Intergenic	0.08		T:C	GWAS	Obesity
8	NAT1	rs17126232	Intergenic	0.09		C:T	GWAS	Obesity
21	NCAM2	rs11088859	Intron	0.04		G:A	GWAS	Obesity
18	NPC1	rs1805081	Coding	0.27		A:G	GWAS	Obesity
2	NRP2	rs7603514	Intergenic	0.19		G:A	GWAS	Obesity
14	NRXN3	rs11624704	Intergenic	0.17		A:C	GWAS	Obesity
13	PCDH9	rs17081231	Intron	0.11		A:G	GWAS	Obesity
10	PRF1	rs10999409	Intergenic	0.49		T:C	GWAS	Obesity
10	PTER	rs10508503	Intergenic	0.02		C:T	GWAS	Obesity
3	RARB	rs1435703	Intron	0.01		G:T	GWAS	Obesity
3	RFTN1	rs12635698	Intron	0.13		T:C	GWAS	Obesity
4	TRAM1L1	rs10433903	Intergenic	0.46		T:C	GWAS	Obesity
16	WWOX	rs9923451	Intron	NA	Call frequency < 95%	NA	GWAS	Obesity
10	ZNF248	rs7474896	Intergenic	0.05		C:T	GWAS	Obesity
7	TNS3	rs700752	Intergenic	0.26		C:G	GWAS	IGF
12	ITPR2	rs718314	Intergenic	0.25		T:C	GWAS	WHR
3	MAGI1	rs6795735	Intergenic	0.24		T:C	GWAS	WHR
17	ACE	rs4293 [b]	Intron	0.41		A:G	Bioinformatics tool	Obesity
17	ACE	rs4331	Coding	0.4		G:A	Bioinformatics tool	Obesity
17	ACE	rs4362 [b]	Coding	0.35		C:T	Bioinformatics tool	Obesity
3	ADIPOQ	rs16861194	Intergenic	0.07		A:G	Bioinformatics tool	Obesity
3	ADIPOQ	rs3774261 [b]	Intron	0.35		G:A	Bioinformatics tool	Obesity
3	ADIPOQ	rs3774262	Intron	0.13		G:A	Bioinformatics tool	Obesity
3	ADIPOQ	rs710445	Intron	0.37		A:G	Bioinformatics tool	Obesity
3	ADIPOQ	rs822396	Intron	0.2		A:G	Bioinformatics tool	Obesity
5	ADRB2	rs1042718	Coding	0.31		C:A	Bioinformatics tool	Obesity
5	ADRB2	rs11168066	Intergenic	0.25		A:C	Bioinformatics tool	Obesity
1	AGT	rs1926723	Intron	0.12		A:G	Bioinformatics tool	Obesity
1	AGT	rs2071404	Intergenic	0.21		G:T	Bioinformatics tool	Obesity
1	AGT	rs4762	Coding	0.13		C:T	Bioinformatics tool	Obesity
1	AGT	rs699	Coding	0.36		C:T	Bioinformatics tool	Obesity
1	AGT	rs7079 [b]	UTR	0.28		C:A	Bioinformatics tool	Obesity

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
19	APOE	rs405509	Intergenic	0.43		A:C	Bioinformatics tool	Obesity
19	APOE	rs7412	Coding	0.04		C:T	Bioinformatics tool	Obesity
11	BDNF	rs10767664	Intron	0.31		A:T	Bioinformatics tool	Obesity
11	BDNF	rs10835210	Intergenic	0.25		C:A	Bioinformatics tool	Obesity
11	BDNF	rs11030119	Intron	0.39		G:A	Bioinformatics tool	Obesity
11	BDNF	rs12273363	Intergenic	0.09		T:C	Bioinformatics tool	Obesity
11	BDNFOS	rs6265	UTR	0.22		G:A	Bioinformatics tool	Obesity
1	C1orf167	rs1537514	Coding	0.23		C:G	Bioinformatics tool	Obesity
6	ENPP1	rs1044498	Coding	0.18		A:C	Bioinformatics tool	Obesity
6	ENPP1	rs10457576	Intron	0.17		T:C	Bioinformatics tool	Obesity
6	ENPP1	rs1830971	Intron	0.4		T:C	Bioinformatics tool	Obesity
6	ENPP1	rs4997284	Intron	0.25		A:G	Bioinformatics tool	Obesity
6	ENPP1	rs7767111	Intron	0.04		G:A	Bioinformatics tool	Obesity
3	ETV5	rs10513801	Intron	0.15		T:G	Bioinformatics tool	Obesity
3	ETV5	rs10937240	Intron	0.29		C:T	Bioinformatics tool	Obesity
3	ETV5	rs4686733	Intergenic	0.17		A:G	Bioinformatics tool	Obesity
3	ETV5	rs6780296	Intergenic	0.5		C:T	Bioinformatics tool	Obesity
3	ETV5	rs9831938 [b]	Intron	0.25		C:T	Bioinformatics tool	Obesity
1	F5	rs13306334	Coding	0.17		C:T	Bioinformatics tool	Obesity
1	F5	rs2213872	Intron	0.36		G:A	Bioinformatics tool	Obesity
1	F5	rs7534848	Intron	0.19		T:C	Bioinformatics tool	Obesity
16	FTO	rs12448529	Intron	NA	Call frequency < 95%	NA	Bioinformatics tool	Obesity
16	FTO	rs13338113	Intron	NA	Not in HWE	NA	Bioinformatics tool	Obesity
16	FTO	rs1420318	Intron	0.35		C:T	Bioinformatics tool	Obesity
16	FTO	rs2111116	Intron	0.34		C:T	Bioinformatics tool	Obesity
16	FTO	rs3751812	Intron	0.3		G:T	Bioinformatics tool	Obesity
4	GNPDA2	rs1128553	UTR	0.48		C:T	Bioinformatics tool	Obesity
4	GNPDA2	rs12640665	Intron	0.08		T:C	Bioinformatics tool	Obesity
4	GNPDA2	rs16857402 [b]	Intron	0.31		T:C	Bioinformatics tool	Obesity
7	IL6	rs1474347 [b]	Intron	0.14		T:G	Bioinformatics tool	Obesity
7	IL6	rs2066992	Intron	0.41		G:T	Bioinformatics tool	Obesity
7	IL6	rs2069840	Intron	0.15		C:G	Bioinformatics tool	Obesity
7	IL6	rs2069843	Intron	0.05		G:A	Bioinformatics tool	Obesity
2	INSIG2	rs10490625	Intron	0.06		C:T	Bioinformatics tool	Obesity
2	INSIG2	rs2161829 [b]	Intron	0.46		C:T	Bioinformatics tool	Obesity
2	INSIG2	rs4848492	Intergenic	0.14		T:C	Bioinformatics tool	Obesity
2	INSIG2	rs9308762	Intron	0.27		T:C	Bioinformatics tool	Obesity
2	IRS1	rs10181778	Intron	0.13		T:A	Bioinformatics tool	Obesity
2	IRS1	rs10498210	Intron	0.09		G:A	Bioinformatics tool	Obesity

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
2	IRS1	rs1560251	Intron	0.28		G:T	Bioinformatics tool	Obesity
2	IRS1	rs16822640	Intron	0.25		G:T	Bioinformatics tool	Obesity
2	IRS1	rs4675096	Intergenic	0.08		G:A	Bioinformatics tool	Obesity
19	KCTD15	rs14810	UTR	0.46		G:C	Bioinformatics tool	Obesity
19	KCTD15	rs2056180	Complex	0.15		T:C	Bioinformatics tool	Obesity
19	KCTD15	rs287104	Intron	0.43		T:C	Bioinformatics tool	Obesity
7	LEP	rs10249476	Intergenic	0.3		G:T	Bioinformatics tool	Obesity
7	LEP	rs10487506	Intergenic	0.5		A:G	Bioinformatics tool	Obesity
7	LEP	rs11763517	Intron	0.38		T:C	Bioinformatics tool	Obesity
7	LEP	rs12706832	Intron	0.48		G:A	Bioinformatics tool	Obesity
7	LEP	rs4236625 [b]	Intron	0.09		A:T	Bioinformatics tool	Obesity
1	LEPR	rs1137100	Coding	0.17		A:G	Bioinformatics tool	Obesity
1	LEPR	rs1137101	Coding	0.48		G:A	Bioinformatics tool	Obesity
1	LEPR	rs1186403	Intergenic	0.19		T:C	Bioinformatics tool	Obesity
1	LEPR	rs7554485	Intergenic	0.48		C:T	Bioinformatics tool	Obesity
1	LEPR	rs8179183	Coding	0.11		G:C	Bioinformatics tool	Obesity
12	LEPREL2	rs1047776	UTR	0.19		G:A	Bioinformatics tool	Obesity
12	LEPREL2	rs1129649	Coding	0.35		T:C	Bioinformatics tool	Obesity
12	LEPREL2	rs4963516	Intron	0.18		A:C	Bioinformatics tool	Obesity
12	LEPREL2	rs5439	UTR	NA	Diffused cluster	NA	Bioinformatics tool	Obesity
8	LPL	rs15285	UTR	NA	Diffused cluster	NA	Bioinformatics tool	Obesity
8	LPL	rs264	Intron	0.2		G:A	Bioinformatics tool	Obesity
8	LPL	rs3200218	UTR	0.13		A:G	Bioinformatics tool	Obesity
8	LPL	rs328	Coding	0.11		C:G	Bioinformatics tool	Obesity
8	LPL	rs330	Intron	NA	Not in HWE	NA	Bioinformatics tool	Obesity
6	LTA	rs1041981	Coding	0.24		C:A	Bioinformatics tool	Obesity
6	LTA	rs2071590	Intergenic	NA	Not in HWE	NA	Bioinformatics tool	Obesity
6	LTA	rs2229094	Coding	0.37		T:C	Bioinformatics tool	Obesity
6	LTA	rs3093542	Intron	0.01		G:C	Bioinformatics tool	Obesity
18	MC4R	rs11872992 [b]	Intergenic	0.16		G:A	Bioinformatics tool	Obesity
18	MC4R	rs1943217	Intergenic	0.27		T:G	Bioinformatics tool	Obesity
11	MTCH2	rs1064608	Coding	0.27		C:G	Bioinformatics tool	Obesity
11	MTCH2	rs4752856	Intron	0.27		G:A	Bioinformatics tool	Obesity
11	MTCH2	rs7118178	Intron	0.26		G:A	Bioinformatics tool	Obesity
1	MTHFR	rs11121832 [b]	Intron	0.28		C:T	Bioinformatics tool	Obesity
1	MTHFR	rs1801131	Coding	NA	Call frequency < 95%	NA	Bioinformatics tool	Obesity
1	MTHFR	rs2274976	Coding	0.19		G:A	Bioinformatics tool	Obesity
1	MTHFR	rs3753584	Intron	0.21		A:G	Bioinformatics tool	Obesity
1	NEGR1	rs1016126	Intron	0.4		G:A	Bioinformatics tool	Obesity

Chr	Gene Symbol	SNP ID	SNP location	MAF [a]	Reason for Exclusion	Major:Minor Allele	Approach for selection	Phenotype
1	NEGR1	rs17092041	Intron	0.19		A:T	Bioinformatics tool	Obesity
1	NEGR1	rs2153929	Intron	0.3		T:C	Bioinformatics tool	Obesity
1	NEGR1	rs2821255	Intron	NA	Call frequency < 95%	NA	Bioinformatics tool	Obesity
1	NEGR1	rs9326098	Intron	0.33		C:G	Bioinformatics tool	Obesity
3	PPARG	rs11128597	Intron	0.27		A:G	Bioinformatics tool	Obesity
3	PPARG	rs11128599	Intron	0.21		G:A	Bioinformatics tool	Obesity
3	PPARG	rs1151996	Intron	0.21		T:G	Bioinformatics tool	Obesity
3	PPARG	rs1801282	Complex	0.12		C:G	Bioinformatics tool	Obesity
3	PPARG	rs1822825	Intron	0.3		C:T	Bioinformatics tool	Obesity
19	RETN	rs3745367	Intron	0.36		G:A	Bioinformatics tool	Obesity
19	RETN	rs7408174	Intergenic	0.1		T:C	Bioinformatics tool	Obesity
1	SEC16B	rs3813649	Coding	NA	Diffused cluster	NA	Bioinformatics tool	Obesity
1	SEC16B	rs7413442	Coding	0.13		G:A	Bioinformatics tool	Obesity
16	SH2B1	rs7201929	Intergenic	0.31		C:T	Bioinformatics tool	Obesity
16	SH2B1	rs7498665	Coding	0.19		A:G	Bioinformatics tool	Obesity
10	TCF7L2	rs10787472	Intron	NA	Diffused cluster	NA	Bioinformatics tool	Obesity
10	TCF7L2	rs11196219 [b]	Intron	0.32		G:A	Bioinformatics tool	Obesity
10	TCF7L2	rs11594610 [b]	Intron	0.06		G:A	Bioinformatics tool	Obesity
10	TCF7L2	rs3814570	Intergenic	0.21		C:T	Bioinformatics tool	Obesity
2	TMEM18	rs11127493	Intergenic	0.33		G:T	Bioinformatics tool	Obesity
2	TMEM18	rs12990777 [b]	Intron	0.3		A:G	Bioinformatics tool	Obesity
2	TMEM18	rs2966398	Intergenic	0.19		A:C	Bioinformatics tool	Obesity
2	TMEM18	rs3187671	UTR	0.12		G:A	Bioinformatics tool	Obesity
19	TOMM40	rs405697	Intron	0.41		C:T	Bioinformatics tool	Obesity
11	UCP2	rs17132534	Intergenic	0.05		T:C	Bioinformatics tool	Obesity
11	UCP3	rs11235972	Intron	0.23		G:A	Bioinformatics tool	Obesity
11	UCP3	rs15763	UTR	0.21		C:T	Bioinformatics tool	Obesity
11	UCP3	rs2075577	Complex	0.49		C:T	Bioinformatics tool	Obesity
11	UCP3	rs647126	UTR	0.48		A:G	Bioinformatics tool	Obesity
11	UCP3	rs7109266	Intergenic	NA	Call frequency < 95%	NA	Bioinformatics tool	Obesity
11	UCP3	rs7930460	Intergenic	0.23		A:G	Bioinformatics tool	Obesity
12	GNB3	rs2301339	Intron	0.28		G:A	Bioinformatics tool	Obesity
19	KCTD15	rs2303174	Intron	0.33		C:G	Bioinformatics tool	Obesity
2	TMEM18	rs2293084	Intron	0.49		A:C	Bioinformatics tool	Obesity

Abbreviations: BMI, Body Mass Index; Chr, Chromosome; GWAS, Genome Wide Association Studies; HWE, Hardy-Weinberg Equilibrium; MAF, Minor Allele Frequency; UTR, Untranslated Region; WHR, Waist-to-Hip Ratio

[a] MAF is estimated in cases and controls

[b] Significant association

Annexure2: Association of SNPs significantly identified either using Genotypic, Dominant or Recessive model and Breast Cancer Risk

SNP ID	Chr	Gene Symbol	SNP location	MAF [a]	Method of Selection	Phenotype	Model	N (Case/Control)	OR[b]	95% CI	p-value
rs10411161	19	ZNF577	Intron	0.48	GWAS	Breast cancer	CC	381/331	Reference		
							CT	545/582	**0.81**	**0.67-0.98**	**0.031**
							TT	256/282	**0.78**	**0.63-0.98**	**0.038**
							D		**0.81**	**0.68-0.96**	**0.016**
							R		0.9	0.74-1.09	0.259
rs10871290	16	GLG1	Intergenic	0.26	GWAS	Breast cancer	TT	685/655	Reference		
							CT	451/465	0.92	0.78-1.09	0.380
							CC	56/79	**0.67**	**0.47-0.97**	**0.034**
							D		0.89	0.76-1.05	0.162
							R		**0.7**	**0.49-0.99**	**0.046**
rs2046210	6	ESR1	Intergenic	0.35	GWAS	Breast cancer	CC	448/503	Reference		
							CT	581/552	1.18	0.99-1.40	0.058
							TT	160/143	1.25	0.96-1.62	0.084
							D		**1.20**	**1.02-1.41**	**0.032**
							R		1.15	0.90-1.46	0.265
rs1219648	10	FGFR2	Intron	0.37	GWAS	Breast cancer	AA	420/451	Reference		
							AG	583/595	1.05	0.88-1.25	0.570
							GG	185/150	**1.32**	**1.02-1.70**	**0.030**
							D		1.11	0.94-1.31	0.233
							R		**1.29**	**1.02-1.62**	**0.034**
rs2981575	10	FGFR2	Intron	0.38	GWAS	Breast cancer	TT	407/442	Reference		
							CT	579/585	1.07	0.90-1.28	0.424
							CC	201/166	**1.31**	**1.02-1.68**	**0.029**
							D		1.13	0.95-1.33	0.160
							R		**1.26**	**1.01-1.58**	**0.042**
rs2981582	10	FGFR2	Intron	0.33	GWAS	Breast cancer	CC	479/519	Reference		
							CT	571/579	1.06	0.90-1.26	0.444
							TT	144/106	**1.47**	**1.11-1.94**	**0.007**
							D		1.13	0.96-1.33	0.138
							R		**1.42**	**1.09-1.85**	**0.009**

SNP ID	Chr	Gene Symbol	SNP location	MAF[a]	Method of Selection	Phenotype	Model	N (Case/Control)	OR[b]	95% CI	p-value
rs6556756	5	MAT2B	Intergenic	0.14	GWAS	Breast cancer	TT	864/898	Reference		
							GT	298/272	1.13	0.94-1.37	0.178
							GG	13/27	**0.50**	**0.25-0.97**	**0.042**
							D		1.08	0.90-1.30	0.407
							R		**0.48**	**0.25-0.94**	**0.033**
rs865686	9	9q31.2	Intergenic	0.14	GWAS	Breast cancer	TT	920/875	Reference		
							GT	255/298	**0.81**	**0.67-0.98**	**0.035**
							GG	16/26	0.58	0.31-1.09	0.095
							D		**0.80**	**0.66-0.96**	**0.016**
							R		0.61	0.33-1.15	0.128
rs16886165	5	MAP3K1	Intergenic	0.34	GWAS	Breast cancer	TT	456/528	Reference		
							GT	557/522	**1.23**	**1.03-1.46**	**0.017**
							GG	175/145	**1.39**	**1.08-1.80**	**0.010**
							D		**1.27**	**1.08-1.50**	**0.004**
							R		1.25	0.99-1.59	0.063
rs889312	5	MAP3K1	Intergenic	0.39	GWAS	Breast cancer	AA	377/439	Reference		
							AC	586/548	**1.24**	**1.03-1.49**	**0.017**
							CC	215/191	**1.31**	**1.03-1.66**	**0.026**
							D		**1.26**	**1.07-1.50**	**0.007**
							R		1.15	0.93-1.43	0.191
rs2867125	2	TMEM18	Intergenic	0.19	GWAS	BMI	GG	757/772	Reference		
							AG	383/395	0.98	0.83-1.17	0.899
							AA	52/32	**1.65**	**1.05-2.60**	**0.028**
							D		1.04	0.88-1.23	0.654
							R		**1.66**	**1.06-2.60**	**0.026**
rs2922763	8	HNF4G	Intergenic	0.28	GWAS	BMI	AA	582/609	Reference		
							AC	490/501	1.02	0.86-1.21	0.788
							CC	116/90	**1.34**	**1.001-1.81**	**0.049**
							D		1.07	0.91-1.26	0.390
							R		**1.34**	**1.00-1.78**	**0.049**
rs987237	6	TFAP2B	Intron	0.20	GWAS	BMI	AA	708/779	Reference		
							AG	418/367	**1.25**	**1.05-1.49**	**0.011**
							GG	56/52	1.18	0.80-1.75	0.395
							D		**1.25**	**1.05-1.47**	**0.010**
							R		1.10	0.74-1.61	0.642

SNP ID	Chr	Gene Symbol	SNP location	MAF[a]	Method of Selection	Phenotype	Model	N (Case/Control)	OR[b]	95% CI	p-value
rs10980926	9	ZNF483	Intron	0.47	GWAS	Menstruation and menopause	GG	312/322	Reference		
							AG	582/622	0.96	0.79-1.17	0.722
							AA	293/253	1.19	0.95-1.50	0.127
							D		1.03	0.86-1.24	0.734
							R		**1.22**	**1.01-1.48**	**0.039**
rs6575793	14	BEGAIN	Intron	0.48	GWAS	Menstruation and menopause	TT	301/347	Reference		
							CT	616/560	**1.26**	**1.04-1.53**	**0.015**
							CC	275/294	1.07	0.86-1.35	0.512
							D		**1.20**	**1.00-1.44**	**0.045**
							R		0.93	0.77-1.12	0.418
rs10953454	7	LHFPL3	Intron	0.08	GWAS	Obesity	GG	1016/1020	Reference		
							AG	168/166	1.01	0.80-1.28	0.893
							AA	7/17	0.41	0.17-1.00	0.051
							D		0.96	0.77-1.20	0.722
							R		**0.41**	**0.16-0.99**	**0.047**
rs16867321	2	CWC22	Intergenic	0.35	GWAS	Obesity	CC	547/500	Reference		
							CT	528/563	0.85	0.72-1.01	0.075
							TT	114/132	0.78	0.59-1.04	0.096
							D		**0.84**	**0.72-0.99**	**0.041**
							R		0.85	0.66-1.11	0.242
rs2116830	10	KCNMA1	UTR	0.03	GWAS	Obesity	CC	1097/1142	Reference		
							AC	91/60	**1.57**	**1.12-2.21**	**0.008**
							AA	4/3	1.38	0.30-6.21	0.668
							D		**1.57**	**1.13-2.18**	**0.007**
							R		1.35	0.30-6.04	0.696
rs2274459	6	MLN	Intergenic	0.07	GWAS	Obesity	GG	1060/1031	Reference		
							AG	117/166	**0.68**	**0.53-0.88**	**0.003**
							AA	7/6	1.13	0.38-3.38	0.821
							D		**0.70**	**0.55-0.90**	**0.005**
							R		1.19	0.40-3.54	0.759
rs374748	5	FBN2	Intron	0.08	GWAS	Obesity	AA	989/1009	Reference		
							AG	184/188	0.99	0.80-1.24	0.990
							GG	17/7	**2.47**	**1.02-6.00**	**0.044**
							D		1.05	0.85-1.31	0.647
							R		**2.48**	**1.02-6.00**	**0.044**

SNP ID	Chr	Gene Symbol	SNP location	MAF[a]	Method of Selection	Phenotype	Model	N (Case/Control)	OR[b]	95% CI	p-value
rs925642	4	FAT1	Intergenic	0.46	GWAS	Obesity	AA	353/368	Reference		
							AG	622/557	1.16	0.96-1.40	0.108
							GG	217/272	0.83	0.66-1.04	0.117
							D		1.06	0.89-1.26	0.548
							R		**0.76**	**0.62-0.92**	**0.006**
rs988712	11	BDNF	Intron	0.26	GWAS	Obesity	GG	612/646	Reference		
							GT	473/478	1.04	0.88-1.23	0.612
							TT	103/76	**1.43**	**1.04-1.96**	**0.027**
							D		1.10	0.93-1.29	0.257
							R		**1.40**	**1.03-1.91**	**0.031**
rs1056836	2	CYP1B1	UTR	0.20	Candidate SNPs	Breast cancer	CC	735/751	Reference		
							CG	381/401	0.97	0.81-1.15	0.737
							GG	70/45	**1.58**	**1.07-2.34**	**0.019**
							D		1.03	0.88-1.22	0.699
							R		**1.61**	**1.09-2.36**	**0.015**
rs1695	11	GSTP1	Coding	0.28	Candidate SNPs	Breast cancer	AA	665/624	Reference		
							AG	434/485	**0.83**	**0.70-0.99**	**0.043**
							GG	90/86	0.98	0.71-1.34	0.910
							D		0.86	0.73-1.01	0.069
							R		1.05	0.77-1.43	0.728
rs2070744	7	NOS3	Intron	0.22	Candidate SNPs	Breast cancer	TT	681/711	Reference		
							CT	436/447	1.01	0.86-1.20	0.833
							CC	74/46	**1.67**	**1.14-2.46**	**0.008**
							D		1.08	0.92-1.27	0.353
							R		**1.67**	**1.14-2.43**	**0.008**
rs2287499	17	WDR79	Coding	0.26	Candidate SNPs	Breast cancer	CC	649/670	Reference		
							CG	476/432	1.13	0.96-1.34	0.135
							GG	63/95	**0.68**	**0.48-0.95**	**0.027**
							D		1.06	0.90-1.24	0.509
							R		**0.65**	**0.47-0.90**	**0.010**
rs2420946	10	FGFR2	Intron	0.38	Candidate SNPs	Breast cancer	CC	400/443	Reference		
							CT	592/599	1.09	0.91-1.30	0.316
							TT	195/152	**1.42**	**1.10-1.82**	**0.006**
							D		1.16	0.98-1.37	0.083
							R		**1.35**	**1.07-1.70**	**0.011**

SNP ID	Chr	Gene Symbol	SNP location	MAF[a]	Method of Selection	Phenotype	Model	N (Case/Control)	OR[b]	95% CI	p-value
rs3218408	7	XRCC2	Intron	0.19	Candidate SNPs	Breast cancer	TT	718/784	Reference		
							GT	415/368	**1.23**	**1.03-1.46**	**0.018**
							GG	48/45	1.16	0.76-1.77	0.476
							D		**1.22**	**1.04-1.45**	**0.018**
							R		1.09	0.72-1.64	0.701
rs5275	1	COX-2	UTR	0.38	Candidate SNPs	Breast cancer	TT	428/469	Reference		
							CT	585/533	**1.20**	**1.008-1.43**	**0.040**
							CC	155/181	0.93	0.72-1.20	0.620
							D		1.14	0.96-1.34	0.134
							R		0.85	0.67-1.07	0.160
rs10489525	1	CSDE1	Intron	0.47	Candidate SNPs	Others	AA	288/341	Reference		
							AG	613/588	**1.23**	**1.01-1.49**	**0.033**
							GG	281/268	1.24	0.98-1.56	0.065
							D		**1.24**	**1.03-1.49**	**0.023**
							R		1.08	0.89-1.31	0.423
rs2233660	17	PHB	Intergenic	0.11	Candidate SNPs	Others	TT	904/959	Reference		
							CT	274/229	**1.26**	**1.04-1.54**	**0.018**
							CC	16/16	1.06	0.52-2.13	0.868
							D		**1.26**	**1.04-1.52**	**0.021**
							R		1.01	0.50-2.03	0.981
rs2240123	17	CBX1	Intron	0.26	Candidate SNPs	Others	CC	665/657	Reference		
							CT	468/456	1.01	0.85-1.19	0.872
							TT	56/79	0.70	0.48-1.00	0.052
							D		0.97	0.82-1.14	0.690
							R		**0.70**	**0.49-0.99**	**0.044**
rs274586	1	TNNI3K	Intron	0.50	Candidate SNPs	Others	GG	307/315	Reference		
							CG	614/570	1.10	0.91-1.34	0.312
							CC	254/302	0.86	0.68-1.08	0.208
							D		1.02	0.85-1.23	0.821
							R		**0.81**	**0.67-0.98**	**0.029**
rs489990	1	TNNI3K	Intron	0.14	Candidate SNPs	Others	AA	995/879	Reference		
							AG	259/300	0.84	0.70-1.02	0.090
							GG	34/18	**1.85**	**1.03-3.30**	**0.036**
							D		0.90	0.75-1.09	0.287
							R		**1.93**	**1.08-3.44**	**0.026**

SNP ID	Chr	Gene Symbol	SNP location	MAF[a]	Method of Selection	Phenotype	Model	N (Case/Control)	OR[b]	95% CI	p-value
rs11121832	1	MTHFR	Intron	0.27	Bioinformatics tool	Obesity	CC	584/653	Reference		
							CT	517/459	**1.25**	**1.06-1.49**	**0.007**
							TT	93/89	1.16	0.85-1.59	0.327
							D		**1.25**	**1.06-1.46**	**0.008**
							R		1.06	0.78-1.43	0.727
rs11196219	10	TCF7L2	Intron	0.33	Bioinformatics tool	Obesity	GG	583/535	Reference		
							AG	489/535	**0.83**	**0.70-0.99**	**0.042**
							AA	115/128	0.82	0.62-1.08	0.173
							D		**0.84**	**0.71-0.98**	**0.029**
							R		0.90	0.69-1.17	0.422
rs11594610	10	TCF7L2	Intron	0.08	Bioinformatics tool	Obesity	GG	1068/1026	Reference		
							AG	118/172	**0.65**	**0.51-0.84**	**0.001**
							AA	8/6	1.28	0.44-3.70	0.648
							D		**0.68**	**0.53-0.87**	**0.002**
							R		1.35	0.47-3.89	0.583
rs11872992	18	MC4R	Intergenic	0.18	Bioinformatics tool	Obesity	GG	853/828	Reference		
							AG	308/333	0.89	0.74-1.07	0.246
							AA	25/40	0.60	0.36-1.00	0.054
							D		0.87	0.73-1.03	0.111
							R		0.63	0.38-1.04	0.069
rs12990777	2	TMEM18	Intron	0.29	Bioinformatics tool	Obesity	AA	547/600	Reference		
							AG	530/482	**1.20**	**1.01-1.42**	**0.030**
							GG	106/110	1.05	0.79-1.41	0.709
							D		**1.18**	**1.00-1.38**	**0.046**
							R		0.97	0.73-1.28	0.820
rs16857402	4	GNPDA2	Intron	0.32	Bioinformatics tool	Obesity	TT	589/543	Reference		
							CT	495/548	**0.83**	**0.70-0.98**	**0.033**
							CC	106/111	0.88	0.65-1.17	0.390
							D		**0.84**	**0.72-0.99**	**0.034**
							R		0.96	0.73-1.27	0.781

SNP ID	Chr	Gene Symbol	SNP location	MAF[a]	Method of Selection	Phenotype	Model	N (Case/Control)	OR[b]	95% CI	p-value
rs2161829	2	INSIG2	Intron	0.44	Bioinformatics tool	Obesity	CC	303/362	Reference		
							CT	634/618	**1.22**	**1.01-1.48**	**0.034**
							TT	253/218	**1.38**	**1.09-1.75**	**0.007**
							D		**1.27**	**1.06-1.52**	**0.010**
							R		1.21	0.99-1.49	0.060
rs3774261	3	ADIPOQ	Intron	0.34	Bioinformatics tool	Obesity	GG	471/525	Reference		
							AG	556/536	1.15	0.97-1.37	0.098
							AA	157/138	1.26	0.97-1.64	0.074
							D		**1.18**	**1.00-1.39**	**0.048**
							R		1.18	0.92-1.50	0.195
rs4236625	7	LEP	Intron	0.09	Bioinformatics tool	Obesity	AA	1000/988	Reference		
							AT	171/208	0.81	0.65-1.01	0.065
							TT	15/4	**3.70**	**1.22-11.20**	**0.020**
							D		0.86	0.69-1.07	0.194
							R		**3.83**	**1.26-11.57**	**0.017**
rs4293	17	ACE	Intron	0.42	Bioinformatics tool	Obesity	AA	438/395	Reference		
							AG	576/608	0.85	0.71-1.02	0.082
							GG	171/197	0.78	0.61-1.00	0.051
							D		**0.84**	**0.71-0.99**	**0.038**
							R		0.86	0.69-1.07	0.180
rs4362	17	ACE	Coding	0.36	Bioinformatics tool	Obesity	CC	515/464	Reference		
							CT	552/575	0.86	0.72-1.02	0.097
							TT	105/141	**0.67**	**0.50-0.88**	**0.006**
							D		**0.83**	**0.70-0.97**	**0.023**
							R		**0.73**	**0.56-0.95**	**0.018**
rs7079	1	AGT	UTR	0.27	Bioinformatics tool	Obesity	CC	599/637	Reference		
							AC	483/478	1.07	0.90-1.27	0.403
							AA	102/79	**1.37**	**1.002-1.87**	**0.048**
							D		1.12	0.95-1.31	0.178
							R		1.33	0.98-1.80	0.067

SNP ID	Chr	Gene Symbol	SNP location	MAF[a]	Method of Selection	Phenotype	Model	N (Case/Control)	OR[b]	95% CI	p-value
rs9831938	3	ETV5	Intron	0.26	Bioinformatics tool	Obesity	CC	670/662	Reference		
							CT	448/444	0.99	0.84-1.18	0.972
							TT	65/92	**0.69**	**0.49-0.97**	**0.036**
							D		0.95	0.80-1.11	0.499
							R		**0.70**	**0.50-0.97**	**0.032**

Abbreviations: ACE, Angiotensin I-Converting Enzyme; ADIPOQ, Adiponectin, C1Q And Collagen Domain Containing; AGT, Angiotensinogen; BDNF, Brain-Derived Neurotrophic Factor; BEGAIN, Brain Enriched Guanylate Kinase Associated; CBX1, Chromobox Homolog 1; Chr, Chromosome; CI, Confidence Interval; COX-2, Cyclooxygenase-2; CSDE1, Cold Shock Domain Containing E1, RNA-binding; CWC22, CWC22 spliceosome-associated protein homolog; CYP1B1, Cytochrome P450 subfamily I dioxin-inducible polypeptide 1; D, Dominant; ESR1, Estrogen receptor 1; ETV5, E26 transformation-specific variant 5; FAT1, FAT atypical cadherin 1; FBN2, Fibrillin 2; FGFR2, Fibroblast Growth Factor Receptor 2; GLG1, Golgi Apparatus Protein 1; GNPDA2, Glucoseamine-6-Phosphate Deaminase 2; GSTP1, Glutathione S-Transferase Pi; GWAS, Genome Wide Association Studies; HNF4G; Hepatocyte Nuclear factor 4 Gamma; INSIG2, Insulin-Induced Gene 2; KCNMA1, Potassium Channel, Calcium-Activated, Large Conductance, Subfamily M, Alpha Member 1; LEP, Leptin; LHFPL3, Lipoma High mobility group protein isoform I-C Fusion Partner like Protein 3; MAF, Minor Allele Frequency; MAP3K1, Mitogen-Activated Kinase Kinase Kinase 1; MAT2B; Methionine Adenosyltransferase II, Beta; MC4R, Melanocortin 4 Receptor ; MLN, Motilin; N, Number; MTHFR, Methylenetetrahydrofolate reductase; NOS3, Nitric Oxide Synthase 3; OR, Odds Ratio; PHB, Prohibitiin; R, Recessive; SNP, Single Nucleotide Polymorphism; TCF7L2, Transcription Factor 7-Like 2; TFAP2B, Transcription Factor Ap2-Beta; TMEM18, Transmembrane Protein 18; TNNI3K, Troponin I Type 3 Interacting Kinase; UTR, Untranslated Region; WDR79, WD repeat containing, antisense to TP53; XRCC2 X-Ray Repair Cross Complimenting gene 2; XRCC3 X-Ray Repair Cross Complimenting Defective gene 3; ZNF483, Zinc Finger Protein 483; ZNF577, Zinc Finger Protein 577.

[a] Minor Allele Frequency in controls

[b] Not adjusted

Total number per SNP may vary because of missing values

Significant models were shown in bold