Published in final edited form as:

Future Oncol. 2009 June ; 5(5): 689–701. doi:10.2217/fon.09.29.

Polygenic susceptibility to breast cancer: current state-of-the-art

Maya Ghoussaini, PhD[†] and

Cancer Research UK, Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK, Tel.: +44 122 374 0272, maya@srl.cam.ac.uk

Paul DP Pharoah, MD, PhD

Cancer Research UK, Department of Oncology/Public Health Genetics Unit, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK, Tel.: +44 122 340 166, Fax: +44 122 340 147, paul.pharoah@srl.cam.ac.uk

Abstract

Breast cancer is the most commonly occurring invasive cancer among women and its estimated incidence rate worldwide is approximately 1 million cases annually. Although several breast cancer susceptibility genes have been identified over the past two decades, it is likely that many genes with modest effects are yet to be found. In this review, we discuss the progress that has been recently made with the emergence of empirical genome-wide association studies for breast cancer and the identification of several common, low-penetrance disease alleles at loci that had not been previously implicated as candidates for breast cancer susceptibility. We also discuss the implications of these recent findings for risk prediction, targeted screening and public health interventions, and conclude by discussing the strengths and weaknesses of these studies, and the strategies required to identify additional risk factors for breast cancer.

Keywords

association studies; breast cancer; clinical relevance; frequent variants; genetic susceptibility; genome-wide association; low penetrance; risk prediction

Breast cancer is the most commonly occurring cancer in women worldwide. With an estimated 1,150,000 new cases each year, female breast cancer accounts for over 400,000 cancer deaths per year [1]. Its incidence rates have increased in all regions of the world, but are nearly threefold higher in developed than less-developed countries [1]. However, breast cancer death rates have been dropping steadily since 1990, because of a better understanding of the disease etiology, earlier detection and better treatments [101].

[†]Author for correspondence: Cancer Research UK, Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK Tel.: +44 122 374 0272 maya@srl.cam.ac.uk.

Financial & competing interests disclosure

Dr Paul DP Pharoah is a Cancer Research UK Senior Clinical Research Fellow. Dr Maya Ghoussaini is funded through a program grant from Cancer Research UK. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Observational epidemiology has shown that breast cancer is approximately twice as common among first-degree relatives of breast cancer patients as compared with women in the general population [2,3]. In parallel, twin studies reported that the relative risk of breast cancer is much higher in monozygotic twins compared with dizygotic [4,5]. Together, these observations strongly suggest that the predominant component of familial aggregation in breast cancer is genetic rather than environmental.

To date, several approaches have been used to identify breast cancer susceptibility alleles, and these have revealed loci that can be categorized into three groups according to their population frequency and the risks conferred; high-penetrance alleles, which are very rare in the population (minor allele frequency [MAF] < 0.005) that confer a risk greater than tenfold; rare, moderate-penetrance alleles (MAF: 0.005–0.01), that confer a two- to four-fold increase in risk; and low-penetrance alleles (MAF > 0.05) that confer a less than 1.5-fold increase in risk. In the 1990s, two major autosomal-dominant susceptibility genes for *BRCA1* and *BRCA2* were identified through linkage analysis and positional cloning approaches [6–9]. These genes confer a risk of breast cancer among carriers that is ten- to 30-times as high as the risk among women in the general population with a carrier frequency of approximately 0.1% in the UK population [10]. Other highly penetrant mutations have been found in inherited syndromes that include breast cancer as part of the complex phenotype – *TP53* mutations in Li–Fraumeni syndrome [11] and *PTEN* mutations in Cowden syndrome [12–14] – but these alleles are extremely rare and they account for less than 1% of cases of breast cancer.

Combined, this category of rare, high-penetrance alleles account for approximately 20% of the genetic component of breast cancer risk. This assumes that the genetic component of breast cancer has a log-normal distribution of relative risk with variance of 1.44 [15]

The second category of breast cancer susceptibility alleles includes certain rare variants in *ATM* [16–19], *CHEK2* [20,21], *PALB2* [22,23] and *BRIP1* [19]. These were identified by candidate gene resequencing approaches, in which genes interacting with *BRCA1* and *BRCA2* or acting in the same DNA repair pathway were sequenced in breast cancer cases with a strong family history. The risk alleles are all rare (frequency less than 0.4%) and confer a two- to four-fold increased risk of breast cancer among carriers in comparison with noncarrier individuals. Moderate risk alleles account for approximately 5% of the genetic component of risk.

Despite extensive efforts, linkage analyses and resequencing of candidate genes failed to identify further genes affecting susceptibility to breast cancer [24]. Moreover, the genes mentioned above account for less than 30% of the familial risk of breast cancer [25], indicating that the majority of the familial clustering is still unexplained (Figure 1). This suggests that the residual genetic variance is likely due to a large number of variants, each of which confers a moderate effect on breast cancer risk.

Candidate gene-association studies and, more importantly, genome-wide association scans (GWAS) have been successful at identifying a third category of susceptibility genes/loci. These variants are frequent in the general population (frequency up to 50%) and confer a

relative risk of breast cancer of less than 1.5. A nonsynonymous variant in the *CASP8* gene was identified through the association study approach [26], while 11 other variants were identified by GWAS and follow-up studies [27–33]. These two approaches and the variants that emerged from them are detailed in the sections of the review that follow.

Candidate gene-association studies

Association studies (also called case–control studies) are used to search for common (usually $MAF > 0.05$) low-penetrance susceptibility alleles and compare the frequency of genotypes or alleles of a particular variant between cases and unrelated controls. This approach is cheaper and simpler than the complete resequencing of candidate genes, and has generally focused on candidate genes or pathways that are important in breast cancer biology, including cell-cycle regulation, carcinogenesis, apoptosis, DNA repair, carcinogen metabolism and steroid hormone metabolism.

The most convincing example that emerged from this approach is the *CASP8*D302H [26,34], where a common nonsynonymous coding variant in the *CASP8* gene was associated with a moderate reduction in breast cancer risk in approximately 16,000 cases and 17,000 controls from the Breast Cancer Association Consortium (BCAC) study, and the association achieved a significance level close to that required for GWAS (p-value: 1.1×10^{-7} ; odds ratio [OR]: 0.88; 95% CI: 0.84–0.92) [26]. *CASP8* is a regulator of apoptosis [35], and its impaired expression or function was reported to promote tumor formation, tumor progression and treatment resistance in several types of cancers [36]. A weaker evidence of association was found for the nonsynonymous variant L10P in *TGFB1* (p-value: 2.8×10^{-5} ; OR: 1.08; 95% CI: 1.04–1.11) [26]. *TGFB1* is a polypeptide cytokine that regulates normal mammary gland development and its high circulating levels are associated with poor prognosis in cancer patients [37].

Several other positive associations have been reported with breast cancer, but very few of these have been convincingly replicated [38,39]. Studies were either too small or subject to bias (e.g., population stratification, clinical heterogeneity in cases, genotyping errors, nonstringent p-values, etc.). Moreover, the candidate gene approach is limited by its reliance on predicting the identity of the correct gene on the basis of biological hypotheses. This means that when the fundamental molecular disease mechanisms are not fully understood, the strategy will be inadequate to fully identify the genetic basis of the disease. Therefore, an alternative and more robust approach that does not rely on existing knowledge of biological pathways and gene function was required to identify more common susceptibility variants for breast cancer susceptibility.

Genome-wide association

A genome-wide association approach is an association study that scans most of the genome for causal genetic variants without prior assumption of biological function or location. The dbSNP database currently contains more than 10 million SNPs, including most of the 11 million SNPs that are estimated to exist in the human genome with MAFs of 1% or greater [40]. The emergence of this high-density map of SNPs, the determination of linkage

disequilibrium (LD) patterns on a genome-wide scale through the HapMap project and the availability of affordable genotyping platforms has allowed a genome-wide approach to become a realistic option in association studies. GWAS do not require direct testing of each of the approximately 10 million common variants for association with the disease. Because of the high LD (high correlation between variants), a panel of a few hundred thousand SNPs can be used as ‘tags’ for the majority of the millions of common variants in the genome. One of the key issues that has emerged from genetic association studies is the problem of false positives. This is a particular issue where the prior probability of association is small (only a small proportion of all possible SNPs will truly be associated with any one phenotype). There has been much debate regarding the level of statistical significance that is appropriate for GWAS – a p-value of less than 10^{-7} has been proposed [41,42], but some published studies have used more stringent criteria ($p < 5 \times 10^{-8}$) [43,44].

In contrast to candidate-gene studies, the outcome of GWAS in breast cancer and other multifactorial diseases has been very promising. Previously unsuspected loci have been identified, pointing out new potential pathways involved in breast cancer pathogenesis. To date, three GWAS of breast cancer have been reported in European populations, and identified nine well-validated new breast cancer susceptibility loci at highly significant levels (Table 1) [27,29–31]. A fourth GWAS performed in an Ashkenazi Jewish population reported a new locus [28]; however, the results of this study have yet to be validated in independent populations. More recently, a fifth GWAS was performed in Chinese women and identified a well-validated new breast cancer susceptibility locus.

Easton and colleagues performed a three-stage GWAS [27]. In the first stage, approximately 206,586 SNPs were successfully genotyped in 390 familial breast cancers and in 364 controls from the UK using the Perlegen GV4 genotyping platform. In the second stage, 12,711 variants (approximately 5% of those typed in stage 1 and showing the strongest evidence for association) were then evaluated in a further 3990 invasive breast cancer cases and 3916 controls. In the third stage, the 30 most significant SNPs from stages 1 and 2 were tested for association in 22 replication case–control studies from the international BCAC, comprising 21,860 cases of invasive breast cancer, 988 cases of carcinoma *in situ* and 22,578 controls. SNPs at five loci (10q26, 16q12, 5q11, 11p15 and 8q24) showed associations with breast cancer at a genome-wide significance level ($10^{-7.6}$ – 10^{-12}). Four of the SNPs were within known genes. More recently, 814 promising associations from the original two-staged genome scan have been tested by Ahmed *et al.* in a further sample of 37,012 cases and 40,069 controls from 33 BCAC studies, evidence has been found for association of two additional loci (3p24 and 17q22) [32].

Hunter and colleagues conducted a GWAS, as a part of the Cancer Genetic Markers of Susceptibility (CGEMS) project, by genotyping over 500,000 SNPs using the Illumina HumanHap500 array in 1145 women of European ancestry with postmenopausal invasive breast cancer and 1142 controls [29]. Of the six SNPs with most extreme p-values, two were at the 10q26 locus/intron 2 of *FGFR2* and the associations were replicated in a further cohort comprising 1776 cases and 2072 controls. The remaining four variants at the other chromosomal loci were not associated in the further replication sets.

Stacey and coworkers genotyped 311,524 SNPs using the Illumina (CA, USA) HumanHap300 array in 1600 Icelandic breast cancer cases and 11,560 controls [30]. The ten most associated variants were genotyped in an independent sample of Icelandic individuals of 584 cases and 1341 controls and four replication sets of non-Icelandic samples comprising 2350 cases and 4626 controls. Two variants at the 2q35 and 16q12 loci showed significant association signals in all five of these replication sets. An additional variant at the 5p12 locus was reported in a subsequent paper [31]. This locus had also been identified in the study of Easton *et al.*, but it had not quite reached genome-wide significance in these data, and it is not clear if the two results reflect the same association [27].

Gold and colleagues performed a three-stage GWAS in a Ashkenazi Jewish population [28]. In the first stage, over 150,000 variants were successfully genotyped in 249 cases and 299 controls. In the second stage, 343 SNPs selected from the top-ranking 123 chromosomal regions were further tested in 950 cases and 979 controls. A third cohort comprising 243 cases and 187 controls was used as a replication set. In addition to the 10q26/*FGFR2* locus, a new risk locus at chromosome 6q22 was identified.

More recently, a GWAS has been conducted in Chinese women from Shanghai [33]. Over 600,000 SNPs were genotyped in stage 1, comprising 1505 cases and 1522 controls. In stage 2, the 29 variants with the strongest evidence for association were then evaluated in an independent case-control set comprising 1554 cases and 1576 controls. In stage 3, the four most significant variants from stage 1 and 2 were replicated in a further set of 3472 cases and 900 controls. Two variants (rs2046210 and rs10872676) out of the four were found to be associated with breast cancer. They both lie on the 6q25.1 locus and are in LD with each other ($r^2 = 0.68$).

In total ten loci for breast cancer susceptibility have been identified at genome-wide significance with replication in multiple, independent studies (Table 1). An eleventh locus, identified by a candidate-gene approach, did not quite reach the threshold for genome-wide significance. However, because they confer small increases in risk, these ten alleles only account for 5.4% of the genetic component of breast cancer. Thus, approximately 30% of the genetic component of breast cancer can be explained by the known susceptibility loci (Figure 1).

Biology of known low-penetrance breast cancer susceptibility loci

The biological mechanisms underlying these associations are as yet unexplained and the process to identify them by functional studies will be difficult. Moreover, despite their strong associations with breast cancer, it is not clear that any of the variants identified here are the causal ones within their LD blocks. Therefore, the causal variants remain unknown to date. Once identified, it will still be hard to determine how these variants affect gene function/ expression as they may not even lie within genes. Even when they are intragenic, the expression effect may not be restricted to the gene itself, but to nearby ones.

It is notable that most of the associated SNPs lie in LD blocks that are confined to introns or intergenic regions rather than coding regions of the genes. It is therefore unlikely that

variation in the coding sequence can explain the functional effects of these SNPs. Indeed, two-thirds of the evolutionary conserved part of the human genome (5% of the whole genome – and likely to be functional) consists of noncoding sequences [45]. One example is the 8q24 locus, which is a multicancer susceptibility region, where associated variants are located hundreds of kilobases away from the closest gene and separated by recombination hotspots.

10q26

The most strongly associated gene variants with breast cancer susceptibility were reported by two independent studies at the 10q locus, rs2981582 and rs1219648 [27,29] (rs2981582 in the UK study) (p -value = 2×10^{-76} ; OR: 1.26; 95% CI: 1.23–1.30) [27]. These variants lie in intron 2 of *FGFR2*, known to be highly conserved among mammals. Associations of SNPs in intron 2 of *FGFR2* were further confirmed in Asian, African–American and Israeli populations [33,46–48]. The rs2981582 variant was also reported to be associated with an increased risk of breast cancer in *BRCA2* mutation carriers, but not in *BRCA1* carriers [49]. Fine-scale mapping of the 25 kb LD block in Asian and UK populations reduced the association to a haplotype of six variants spanning a region of 7.5 kb [27]. These six variants were too strongly correlated to be separated using statistical approaches, even with very large sample sizes. However, rs7895676 and rs2981578 have been found to alter the binding affinity of transcription factors Oct1/Runx2 and C/EBP β , giving rise to increased *FGFR2* expression [50]. More recently, rs2981578 was found to lie in a region of the chromatin that is in an open configuration, and thus accessible to regulatory factors [48]. Taken together, the genetic, biochemical and conservation data indicate that rs2981578 could be the causal variant in this gene.

FGFR2 is a tyrosine kinase receptor playing a crucial role in mammary gland development [51] and known to be amplified and over expressed in 5–10% of breast tumors [52–54]. Moreover, somatic missense mutations of this gene have been found in several tumors and they induce overactivity of the protein [55,56]. In human breast cancer, the expression of *FGFR2* is known to be elevated in estrogen receptor (ER)-positive tumors [57]. Furthermore, Stacey and collaborators found that the risk conferred by these variants was only found in patients with ER-positive tumors [31]. However, these findings were not confirmed by BCAC collaborators, who reported a significant risk for both ER-positive and ER-negative disease, with a greater risk associated with ER-positive breast cancer [58].

16q12

Two variants (rs12443621 and rs8051542) associated with breast cancer were originally identified at this locus by the BCAC study ($p = 2 \times 10^{-19}$; OR: 1.11; 95% CI: 1.08–1.14 and $p = 10^{-12}$; OR: 1.09; 95% CI: 1.06–1.13, respectively) [27]. To further map this region, 19 additional variants were examined for association with breast cancer, and identified an even stronger association originated by rs3803662, a synonymous coding variant in the hypothetical gene *LOC643714* ($p = 10^{-36}$; OR: 1.20; 95% CI: 1.16–1.24) [27]. The same variant was identified by the DECODE group to be associated with breast cancer ($p = 5.9 \times 10^{-19}$; OR: 1.28; 95% CI: 1.21–1.35), particularly with ER-positive tumors [30]. More recently, this variant was associated with an increased breast cancer risk in both *BRCA1* and

BRCA2 mutation carriers [49]. rs3803662 is located 8 kb upstream of *TOX3* (previously known as *TNRC9*), a gene of unknown function expected to be acting as a transcription factor owing to the presence of a high-mobility group box motif in its sequence [27,59]. Moreover, the increased expression of *TOX3* seems to be predictive of metastasis of breast cancer to bone [60].

11p15

Evidence of association was found for rs3817198 at the 11p locus ($p = 3 \times 10^{-9}$; OR: 1.07; 95% CI: 1.04–1.11); however, these findings were not replicated in the Chinese population [33]. rs3817198 lies in intron 10 of *LSP1*, an F-actin-binding cytoskeletal protein expressed in hematopoietic and endothelial cells [27]. Evidence of association was also found for rs2107425 in the neighboring gene *H19*, an imprinted maternally expressed untranslated mRNA closely involved in regulation of the IGF2 ($p = 2 \times 10^{-5}$; OR: 0.96; 95% CI: 0.93–0.99) [27]. No fine mapping of this locus has been reported and there are no clues as to how it might affect gene function.

5q11

Variant rs889312 at the 5q11 locus was reported to be associated with breast cancer ($p = 7 \times 10^{-20}$; OR: 1.13; 95% CI: 1.10–1.16) [27]. This variant was associated with an increased risk of breast cancer in *BRCA2* carriers, but not in *BRCA1* [49]. This variant lies in a 298 kb LD block containing *MAP3K1*. The MAPK pathway is known for its involvement in cell proliferation and apoptosis [61]; however, no direct evidence linking *MAP3K1* to cancer has been reported to date. Two other genes map to the region: *MGC33648* and *MIER3* [62].

8q24

Association with breast cancer risk was found for rs13281615 at the 8q24 locus ($p = 5 \times 10^{-12}$, OR: 1.08; 95% CI: 1.05–1.11) [27]. This association was stronger for ER-positive compared with ER-negative tumors [58]. This SNP tags a 160 kb LD block in a region of 600 kb containing no known genes but considered to be highly amplified in different cancers. The rs13281615 SNP is located approximately 800 kb downstream of the *FAM84B* (a breast cancer membrane protein 101) and 400 kb upstream of the pro-oncogene *MYC*. Overexpression of *MYC* occurs in different cancers [63,64], and knockdown of c-MYC expression by RNAi inhibits tumor growth both *in vivo* and *in vitro* [65]. Several studies reported variants within the 600 kb region but outside the LD block to be associated with prostate, colorectal, ovarian and bladder cancers [66–74]. The mechanism underlying these associations remains unknown, as these variants do not seem to affect the expression level of *FAM84B* and *MYC* [66,68]. A potential long-range mode of action of these variants is possible.

3p24

The association of variant rs4973768 was reported after a follow-up of the 925 best associations from the original two-staged genome scan reported by Easton and coworkers ($p = 4.1 \times 10^{-23}$; OR: 1.11; 95% CI: 1.08–1.13 in approximately 30,000 cases and 35,000 controls) [32]. rs4973768 lies in the 3p24 locus that contains two genes, *NEK10* and

SLC4A7. *NEK10* is one of a family of 11 involved in cell-cycle control. The function of *NEK10* is unknown, but evidence shows that *NEK2*, 6, 7 and 9 are involved in mitosis [75]. *SLC4A7* is a potential tyrosine kinase substrate where expression is reduced in breast cancer tumors, and is predicted to affect the pH of the microenvironment around breast tumor cells [76].

17q22

Evidence of association was also found for rs6504950 after a follow-up of the best variants from the BCAC study ($p = 1.4 \times 10^{-8}$; OR: 0.95; 95% CI: 0.92–0.97) [32]. rs6504950 tags a 300 kb LD block and lies in intron 1 of *STXBP4*, an insulin-regulated *STX4*-binding protein involved in glucose transport and GLUT4 vesicle translocation [77]. The *COX11* and *TOM1L1* genes also lie in this LD block. Interestingly, using a dataset of gene expression quantified in B lymphocytes from the HapMap samples, we found the risk allele rs6504950 to be associated with increased levels of *COX11*, with no effect on *STXBP4* and *TOM1L1*, pointing out that *COX11* might be the plausible gene [32].

2q35

Stacey and coworkers identified rs13387042 at the 2q35 locus to be significantly associated with breast cancer ($p = 1.3 \times 10^{-13}$; OR: 1.20; 95% CI: 1.14–1.26) [30]. No known genes or expressed sequence tag (EST) map to this LD block. The closest genes to this region are *TNPI* (181 kb), *IGFBP5* (345 kb) and *IGFBP2* (376 kb) upstream of the variant and *TNSI* (761 kb distal) downstream [30].

5p12

The three GWAS performed in European populations independently reported variants at the 5p locus to be associated with increased risk of breast cancer, but none of these associations reached a genome-wide significance level (rs7703618 in the deCODE study, rs981782 in the BCAC study and rs4866929 in the CGEMS study) [27,29,30]. While rs4866929 and rs981782 were in strong LD ($r^2 = 0.96$), rs7703618, located approximately 370 kb upstream of rs981782, showed a weak correlation with both of these SNPs ($r^2 < 0.2$). To follow up this association, Stacey and coworkers performed fine mapping of the region in European and African ancestry populations by typing 21 and 38 variants, respectively [31]. rs10941679 ($p = 2.9 \times 10^{-11}$; OR: 1.19; 95% CI: 1.13–1.26) and rs4415084 ($p = 6.4 \times 10^{-10}$; OR: 1.16; 95% CI: 1.10–1.21) showed the strongest association in European populations, with a greater risk in patients with ER-positive tumors [31]. *MRPS30*, also known as *PDCD9*, maps to this region and is a major component in the apoptotic signaling pathways in cells [78,79]. *FGF10*, a mammary tumor oncogene gene is located nearby. It has been shown to be highly mitogenic upon expression in the immortalized mouse mammary epithelial cell line *in vitro*, and renders cells tumorigenic *in vivo*. It is also overexpressed in approximately 10% of human breast carcinomas [80].

6q22

This association has been identified by the GWAS performed in Ashkenazi Jews [28]. Four variants in high LD were found to be associated with breast cancer at the 6q22 locus with

the strongest association ascribed to the rs2180341 variant ($p = 2.9 \times 10^{-8}$; OR: 1.41; 95% CI: 1.25–1.59). In the Chinese population, a trend toward association was found for this variant and breast cancer risk, but in the opposite direction ($p = 0.068$) [33]. This association has yet to be replicated in other populations. Genes in this region include *RNF146* and *ECHDC1*. *RNF146* encodes a polypeptide involved in neurodegenerative diseases [28] and *ECHDC1* has a related domain to a mitochondrial trifunctional protein involved in mitochondrial fatty acid oxidation. The latter is known to be abnormally elevated in a biologically aggressive subset of breast carcinomas [81].

6q25.1

The two variants rs2046210 and rs10872676 were reported to be associated with breast cancer in Chinese women, but the signal was stronger for rs2046210 ($p = 2 \times 10^{-15}$; OR: 1.29; 95% CI: 1.21–1.37) [33]. This association was replicated in a population of European ancestry comprising 1591 cases and 1466 controls, but the signal was weaker ($p = 0.01$) [33]. The two variants are in strong LD with each other, and thus reflect the same signal of association. These variants lie on the 6q25.1 locus, upstream of the *ESR1* gene (which maps 29 kb downstream rs2046210). *ESR1* encodes the estrogen receptor- α (ER α), a transcription factor capable of binding estrogen, a sex hormone playing a pivotal role in the development and growth of breast cancer [82]. Circulating levels of endogenous estrogens have been associated with an increased risk of breast cancer [83,84], and *ESR1* amplification is a frequent event in proliferating breast disease and breast cancer [85].

Clinical relevance of the low-penetrance susceptibility variants

The breast cancer susceptibility SNPs that have been identified to date could be used in a clinical setting to predict future risk of breast cancer. The clinical utility of such a predictive test depends on a variety of factors. These factors include characteristics of the predictive test such as sensitivity, specificity and positive predictive value, as well as characteristics of any intervention that might be used to modify risk. In general, individual SNPs are likely to have extremely limited clinical utility. The potential clinical utility of polygenic prediction has been extensively debated [86–89].

The published per-allele relative risk of breast cancer for *FGFR2* rs2981582 is 1.26 [27]. This can be converted to a genotype-specific relative risks of 0.83 for carriers of two low-risk alleles (common allele homozygotes), 1.05 for carriers of one high- and one low-risk allele (heterozygotes), and 1.38 for carriers of two high-risk alleles (rare allele homozygotes) compared with the average population risk. These correspond to a lifetime risk of breast cancer of 9.8% for women who carry one risk allele and 12.7% for women who carry two risk alleles, compared with a population average of 9.4%. The estimated population frequencies of these groups are 47 and 14%, respectively. A test that classifies an individual with one or two copies of the risk allele as ‘at risk’ would have a sensitivity of 68% (probability that someone who will get the disease carries one or two copies of the risk allele) and a specificity of just 39% (probability that someone who will not get the disease does not carry the risk allele), with a positive-predictive value of 10.5% (probability that someone who carries one or two risk alleles will get breast cancer).

It would be possible to use multiple SNPs in combination, which would be expected to improve the clinical utility. It is likely that the per-allele relative risks have a multiplicative effect, as there is little evidence for interaction between SNPs. Based on the ten loci described above, there are almost 60,000 possible different genotype combinations. Some of these will be very rare in the population – only one in 10 million women carry two high-risk alleles at each locus. The polygenic model of disease susceptibility predicts that the distribution of relative risk based on these variants will follow a log-normal distribution. The polygenic model of disease susceptibility predicts that the distribution of relative risk based on these variants will follow a log-normal distribution assuming a log-additive model of interaction. Based on the predicted distribution, women on the lowest centile of risk will have a 50% reduction in risk (relative risk = 0.5) compared with the population average, whereas those on the top centile of risk will have an 84% increase in risk (relative risk = 1.84). It is predicted that 90% of women – those between the 5th and 95th centiles – will have relative risks between 0.67 and 1.52. Women above the top centile would account for 2% of all cases, women above the 20th centile would account for 29% of cases and women in the top half of the risk distribution would account for 61% of all cases. Figure 2 shows the proportion of cases accounted for by the proportion of population at highest risk. This is known as the receiver operator characteristic curve (ROC), and the area under this curve (AUC) gives a measure of the utility of a predictive test, where an AUC of 0.5 is no better than tossing a coin as a predictor of disease. The area under the curve for polygenic risk profiling is 0.58. For comparison, the figure shows the ROC curve for known lifestyle risk factors for breast cancer (AUC = 0.61) [90], and the ROC curve that would be generated if all possible susceptibility genes were known (AUC = 0.81) [38].

Thus, even using a combination of ten loci the clinical utility of a polygenic test will be limited in predicting future disease for the individual – most women are at only a slight increased risk, and the risk is only modest, even for the very small number at highest risk.

Nevertheless, polygenic risk profiles may provide useful risk stratification in the context of population-based screening programmes, particularly if combined with lifestyle risk factors [88]. For example, all women aged 50 years and above are eligible for screening under the UK National Health Service breast screening program.

Eligibility is irrespective of family history or other risk factors. A 50-year-old woman in the UK general population has a 2.3% chance of breast cancer within the next 10 years of her life. As the benefit of an intervention is generally proportional to the absolute risk, it can be assumed that 2.3% is the threshold at which the screening program has a net benefit. It would therefore be reasonable to offer screening to all women with that level of risk, irrespective of age.

Thus, a 40-year-old woman with a 10-year risk of 2.3% would be offered screening, whereas a 55-year-old woman with a 10-year risk of 1% would not. Targeting screening in this way to those at highest risk would increase the efficiency of the screening program as the cost of a genetic risk profiling would be minimal compared with the costs of a lifetime screening program. A risk-based approach is advocated by the UK National Institute for Health and Clinical Excellence (NICE) guidelines for women with a family history of breast cancer in

recommending mammographic screening for women over the age of 40 years if their 10-year risk is over 3% based on family history alone.

A personalized screening program could be implemented in which the age at starting screening depends on the polygenic risk profile. Based on ten loci, women in the top 15% of the risk distribution would have a relative risk of 1.55 or greater, which corresponds to a 10-year risk of 2.3% at age 40 years. Such women may benefit from additional screening outside the current NHS screening programme. Many issues and questions need to be addressed before such an approach becomes standard practice. In particular, the relationship between benefits and harms/cost of screening may not be simply related to absolute risk, as screening is less sensitive in younger women. Furthermore, the improvement in efficiency of a personalized prevention program may not be worth the inherent additional complexity. There would also be issues of patient and professional acceptance. Nevertheless, some companies are already offering polygenic testing and risk prediction, and the age of personalized prevention based on personalized risk has already arrived.

Discussion & future perspective

GWAS identify previously unsuspected loci

To date, GWAS have revealed ten genes/loci involved in the polygenic susceptibility of breast cancer, and another possible one is awaiting replication. None of these loci were previously detected by candidate gene-association studies, emphasizing the limitations of the candidate gene approach.

Functional analysis will help identify the causal variant

The mechanisms by which these common variants affect breast cancer susceptibility remain largely unknown. To date, only the association signal found in the *FGFR2* gene has been narrowed down by epidemiological genetic approaches to six potential causal variants, and *in vitro* characterization studies successfully reduced the number of variants affecting *FGFR2* expression to two [50]. The characterization of the remaining common variants has yet to be completed. Prior to this, resequencing and fine-mapping of each of the associated loci are required in order to identify all possible causative variants per region. Once such a set of SNPs is identified, functional studies can evaluate the potential impact of each of them on the expression and/or function of nearby genes. These functional analyses are essential in identifying the causal variant and unraveling its molecular mechanism in cancer susceptibility and pathogenesis. Recently, a dataset of gene expression quantified in immortalized B-lymphocytes from all 270 HapMap individuals (CEU [Utah residents with ancestry from northern and western Europe], CHB [Han Chinese in Beijing, China], JPT [Japanese in Tokyo, Japan] and YRI [Yoruba, Ibadan, Nigeria] populations [102]) has been made publicly available [103]. This offers a convenient tool to easily test the correlation between SNP genotypes and gene expression levels. With such data, it becomes possible to determine if any of the variants identified by the GWAS (or SNPs highly correlated with them) affects the transcription of nearby genes in immortalized lymphocytes. This tool has the potential to identify potential genes or mechanisms and to enable a targeted approach to resequencing and fine-mapping. However, the effects of SNPs on gene expression may be

tissue-specific, and it would be ideal to have the equivalent information of gene-expression data in a wider range of tissues and cell lines such as normal and breast cancer tissues.

New strategies to find novel genes/loci

To date, GWAS have suffered from limited statistical power, and thus many loci may have been missed. One of the next challenges would be to reach sufficient statistical power without increasing the sample size and experiment costs. This could be reached by using cases enriched for genetic predisposition, such as cases with family history and early age of onset of disease – an approach used by Easton *et al.* [27], but with a modest sample size. Larger studies combining multiple scans and joint analysis of data also provide a low-cost approach to enhance power and test for interactions between genes. In addition, GWAS performed in homogeneous subgroups of breast cancer cases (e.g., *BRCA1* and 2 positive/negative status, histopathological subgroups) can be very useful at identifying new susceptibility variants, as it will prevent the dilution of an individual gene effect that occurs when all breast cancer patients are considered jointly if multiple genes are contributing to the broad phenotype.

Despite the recent findings made by the GWAS, the current sets of loci identified together explain only a small fraction of the overall familial risk of breast cancer (approximately 5% of breast cancer risk). However, it is important to highlight that the locus-attributable risk will often be underestimated, as SNPs used in the GWAS are only tags of the actual functional variants that have driven the original association. These data also suggest that GWAS have missed many common and rare variants of similar or smaller effect on the disease. Many more loci may be identified through GWAS performed in different ethnic groups and population isolates where allele frequency and correlation of susceptibility variants, together with environmental factors and behaviors, may differ from Europeans [91]. With the exception of one study [33], almost all GWAS reported to date in breast cancer have been performed in case–control sets of European ancestry.

Whole-genome re-sequencing approach

The most comprehensive strategy towards understanding the architecture of breast cancer or any other complex disease would be a full resequencing of the genome in a large cohort of cases and controls. This approach would not be restricted by the choice of candidate genes, it would capture the complete range of coding and noncoding variants and, unlike GWAS, it would be able to test the effect of both rare and common variants on breast cancer susceptibility. Technological advances are making this highly thorough methodology become feasible.

With the discovery of novel genes/loci and their characterization, new etiological pathways will be revealed that should lead to a better understanding of cancer pathogenesis and help potential innovative therapies, better risk prediction and public health interventions.

Acknowledgments

The authors would like to acknowledge Mel Maranian for her comments.

Bibliography

Papers of special note have been highlighted as:

■ of interest

■ ■ of considerable interest

1. Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol.* 2006; 24(14):2137–2150. [PubMed: 16682732] **■ ■ Worldwide cancer statistics.**
2. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst.* 1994; 86(21):1600–1608. [PubMed: 7932824]
3. Hemminki K, Vaitinen P. Familial breast cancer in the family-cancer database. *Int J Cancer.* 1998; 77(3):386–391. [PubMed: 9663600]
4. Lichtenstein P, Holm NV, Verkasalo PK, 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(2):78–85. [PubMed: 10891514]
5. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet.* 2000; 26(4):411–414. [PubMed: 11101836]
6. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science.* 1990; 250(4988):1684–1689. [PubMed: 2270482]
7. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science.* 1994; 266(5182):66–71. [PubMed: 7545954]
8. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, *BRCA2*, to chromosome 13q12–13. *Science.* 1994; 265(5181):2088–2090. [PubMed: 8091231]
9. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene *BRCA2*. *Nature.* 1995; 378(6559):789–792. [PubMed: 8524414]
10. Antoniou A, Pharoah PD, Narod S, 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(5):1117–1130. [PubMed: 12677558]
11. Malkin D, Li FP, Strong LC, et al. Germ line *p53* mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science.* 1990; 250(4985):1233–1238. [PubMed: 1978757]
12. Nelen MR, Padberg GW, Peeters EA, et al. Localization of the gene for Cowden disease to chromosome 10q22–23. *Nat Genet.* 1996; 13(1):114–116. [PubMed: 8673088]
13. Nelen MR, van Staveren WC, Peeters EA, et al. Germline mutations in the *PTEN/MMAC1* gene in patients with Cowden disease. *Hum Mol Genet.* 1997; 6(8):1383–1387. [PubMed: 9259288]
14. Waite KA, Eng C. Protean PTEN: form and function. *Am J Hum Genet.* 2002; 70(4):829–844. [PubMed: 11875759]
15. Antoniou AC, Hardy R, Walker L, et al. Predicting the likelihood of carrying a *BRCA1* or *BRCA2* mutation: validation of BOADICEA, BRCAPRO, IBIS, Myriad and the Manchester scoring system using data from UK genetics clinics. *J Med Genet.* 2008; 45(7):425–431. [PubMed: 18413374]
16. Easton DF. Cancer risks in A-T heterozygotes. *Int J Radiat Biol.* 1994; 66(Suppl. 6):S177–S182. [PubMed: 7836845]
17. Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet.* 2006; 38(8):873–875. [PubMed: 16832357]
18. Savitsky K, Bar-Shira A, Gilad S, et al. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science.* 1995; 268(5218):1749–1753. [PubMed: 7792600]

19. Seal S, Thompson D, Renwick A, et al. Truncating mutations in the Fanconi anemia J gene *BRIP1* are low-penetrance breast cancer susceptibility alleles. *Nat Genet.* 2006; 38(11):1239–1241. [PubMed: 17033622]
20. Bell DW, Varley JM, Szydlo TE, et al. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science.* 1999; 286(5449):2528–2531. [PubMed: 10617473]
21. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2 (*) 1100delC in noncarriers of *BRCA1* or *BRCA2* mutations. *Nat Genet.* 2002; 31(1):55–59. [PubMed: 11967536]
22. Erkkö H, Dowty JG, Nikkila J, et al. Penetrance analysis of the PALB2 c.1592delT founder mutation. *Clin Cancer Res.* 2008; 14(14):4667–4671. [PubMed: 18628482]
23. Rahman N, Seal S, Thompson D, et al. *PALB2*, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet.* 2007; 39(2):165–167. [PubMed: 17200668]
24. Smith P, McGuffog L, Easton DF, et al. A genome wide linkage search for breast cancer susceptibility genes. *Genes Chromosomes Cancer.* 2006; 45(7):646–655. [PubMed: 16575876]
25. Easton DF. How many more breast cancer predisposition genes are there? *Breast Cancer Res.* 1999; 1(1):14–17. [PubMed: 11250676]
26. Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in *CASP8* is associated with breast cancer risk. *Nat Genet.* 2007; 39(3):352–358. [PubMed: 17293864] [■ ■ ■ **Describes the association that has been convincingly replicated using the candidate gene approach.**]
27. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007; 447(7148):1087–1093. [PubMed: 17529967] [■ ■ ■ **Reports new genes/loci involved in the polygenic susceptibility of breast cancer using genome-wide association (GWA)/follow-up studies.**]
28. Gold B, Kirchhoff T, Stefanov S, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci USA.* 2008; 105(11):4340–4345. [PubMed: 18326623] [■ ■ ■ **Reports new genes/loci involved in the polygenic susceptibility of breast cancer using GWA/follow-up studies.**]
29. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat Genet.* 2007; 39(7):870–874. [PubMed: 17529973] [■ ■ ■ **Reports new genes/loci involved in the polygenic susceptibility of breast cancer using GWA/follow-up studies.**]
30. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet.* 2007; 39(7):865–869. [PubMed: 17529974] [■ ■ ■ **Reports new genes/loci involved in the polygenic susceptibility of breast cancer using GWA/follow-up studies.**]
31. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet.* 2008; 40(6):703–706. [PubMed: 18438407] [■ ■ ■ **Reports new genes/loci involved in the polygenic susceptibility of breast cancer using GWA/follow-up studies.**]
32. Ahmed S, Thomas G, Ghoussaini M, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet.* 2009; 41(5):585–590. [PubMed: 19330027] [■ ■ ■ **Reports new genes/loci involved in the polygenic susceptibility of breast cancer using GWA/follow-up studies.**]
33. Zheng W, Long J, Gao YT, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet.* 2009; 41(3):324–328. [PubMed: 19219042] [■ ■ ■ **Reports new genes/loci involved in the polygenic susceptibility of breast cancer using GWA/follow-up studies.**]
34. Garcia-Closas M, Malats N, Silverman D, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet.* 2005; 366(9486):649–659. [PubMed: 16112301]
35. Barnhart BC, Lee JC, Alappat EC, Peter ME. The death effector domain protein family. *Oncogene.* 2003; 22(53):8634–8644. [PubMed: 14634625]
36. Fulda S. Caspase-8 in cancer biology and therapy. *Cancer Lett.* 2008 Epub ahead of print.

37. Saunier EF, Akhurst RJ. TGF β inhibition for cancer therapy. *Curr Cancer Drug Targets*. 2006; 6(7):565–578. [PubMed: 17100564]
38. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet*. 2002; 31(1):33–36. [PubMed: 11984562]
39. Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA. Association between common variation in 120 candidate genes and breast cancer risk. *PLoS Genet*. 2007; 3(3):e42. [PubMed: 17367212]
40. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008; 322(5903):881–888. [PubMed: 18988837]
41. Yang Q, Cui J, Chazaro I, Cupples LA, Demissie S. Power and type I error rate of false discovery rate approaches in genome-wide association studies. *BMC Genet*. 2005; 6(Suppl. 1):S134. [PubMed: 16451593]
42. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA*. 2008; 299(11):1335–1344. [PubMed: 18349094]
43. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*. 2008; 32(4):381–385. [PubMed: 18348202]
44. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996; 273(5281):1516–1517. [PubMed: 8801636]
45. Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002; 420(6915):520–562. [PubMed: 12466850]
46. Liang J, Chen P, Hu Z, et al. Genetic variants in fibroblast growth factor receptor 2 (*FGFR2*) contribute to susceptibility of breast cancer in Chinese women. *Carcinogenesis*. 2008; 29(12):2341–2346. [PubMed: 18845558]
47. Raskin L, Pinchev M, Arad C, et al. *FGFR2* is a breast cancer susceptibility gene in Jewish and Arab Israeli populations. *Cancer Epidemiol Biomarkers Prev*. 2008; 17(5):1060–1065. [PubMed: 18483326]
48. Udler MS, Meyer KB, Pooley KA, et al. *FGFR2* variants and breast cancer risk: fine-scale mapping using African-American studies and analysis of chromatin conformation. *Hum Mol Genet*. 2009; 18(9):1692–1703. [PubMed: 19223389] **■■■ Describes the fine-mapping of the *FGFR2* locus in African-Americans and performs functional analysis (chromatin conformation) on several candidate causal single nucleotide polymorphisms in order to identify the functional variant at this locus.]**
49. Antoniou AC, Spurdle AB, Sinilnikova OM, et al. Common breast cancer-predisposition alleles are associated with breast cancer risk in *BRCA1* and *BRCA2* mutation carriers. *Am J Hum Genet*. 2008; 82(4):937–948. [PubMed: 18355772]
50. Meyer KB, Maia AT, O'Reilly M, et al. Allele-specific up-regulation of *FGFR2* increases susceptibility to breast cancer. *PLoS Biol*. 2008; 6(5):e108. [PubMed: 18462018] **■■■ Carried out functional analysis (gel shift assays/expression analysis) on several candidate causal variants at the *FGFR2* locus in order to identify the functional variant.]**
51. Dickson C, Spencer-Dene B, Dillon C, Fantl V. Tyrosine kinase signalling in breast cancer: fibroblast growth factors and their receptors. *Breast Cancer Res*. 2000; 2(3):191–196. [PubMed: 11250709]
52. Adnane J, Gaudray P, Dionne CA, et al. BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene*. 1991; 6(4):659–663. [PubMed: 1851551]
53. Jang JH, Shin KH, Park JG. Mutations in fibroblast growth factor receptor 2 and fibroblast growth factor receptor 3 genes associated with human gastric and colorectal cancers. *Cancer Res*. 2001; 61(9):3541–3543. [PubMed: 11325814]
54. Moffa AB, Tannheimer SL, Ethier SP. Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. *Mol Cancer Res*. 2004; 2(11):643–652. [PubMed: 15561780]

55. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature*. 2007; 446(7132):153–158. [PubMed: 17344846]
56. Turnbull C, Rahman N. Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet*. 2008; 9:321–345. [PubMed: 18544032] [■ ■ ■ **Interesting review discussing the known genetic predisposition factors to breast cancer (high-, moderate- and low-penetrance alleles).**]
57. Luqmani YA, Graham M, Coombes RC. Expression of basic fibroblast growth factor, FGFR1 and FGFR2 in normal and malignant human breast, and comparison with other normal tissues. *Br J Cancer*. 1992; 66(2):273–280. [PubMed: 1380281]
58. Garcia-Closas M, Hall P, Nevanlinna H, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet*. 2008; 4(4):e1000054. [PubMed: 18437204] [■ ■ ■ **Investigated whether the associations between known susceptibility alleles and breast cancer risk varied by clinically important tumor characteristics (e.g., estrogen receptor [ER]⁺, ER⁻).**]
59. O'Flaherty E, Kaye J. TOX defines a conserved subfamily of HMG-box proteins. *BMC Genomics*. 2003; 4(1):13. [PubMed: 12697058]
60. Smid M, Wang Y, Klijn JG, et al. Genes associated with breast cancer metastatic to bone. *J Clin Oncol*. 2006; 24(15):2261–2267. [PubMed: 16636340]
61. Dhanasekaran DN, Johnson GL. MAPKs: function, regulation, role in cancer and therapeutic targeting. *Oncogene*. 2007; 26(22):3097–3099. [PubMed: 17496908]
62. Easton DF, Eeles RA. Genome-wide association studies in cancer. *Hum Mol Genet*. 2008; 17(R2):R109–R115. [PubMed: 18852198]
63. Buttyan R, Sawczuk IS, Benson MC, Siegal JD, Olsson CA. Enhanced expression of the c-myc protooncogene in high-grade human prostate cancers. *Prostate*. 1987; 11(4):327–337. [PubMed: 2446300]
64. Sears RC. The life cycle of C-myc: from synthesis to degradation. *Cell Cycle*. 2004; 3(9):1133–1137. [PubMed: 15467447]
65. Wang YH, Liu S, Zhang G, et al. Knockdown of c-Myc expression by RNAi inhibits MCF-7 breast tumor cells growth *in vitro* and *in vivo*. *Breast Cancer Res*. 2005; 7(2):R220–R228. [PubMed: 15743499]
66. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet*. 2006; 38(6):652–658. [PubMed: 16682969] [■ ■ **Describes several regions on the 8q24 locus that confer susceptibility to different cancers.**]
67. Ghoussaini M, Song H, Koessler T, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst*. 2008; 100(13):962–966. [PubMed: 18577746] [■ ■ **Describes several regions on the 8q24 locus that confer susceptibility to different cancers.**]
68. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet*. 2007; 39(5):631–637. [PubMed: 17401366] [■ ■ **Describes several regions on the 8q24 locus that confer susceptibility to different cancers.**]
69. Haiman CA, Le ML, Yamamoto J, et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet*. 2007; 39(8):954–956. [PubMed: 17618282] [■ ■ **Describes several regions on the 8q24 locus that confer susceptibility to different cancers.**]
70. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet*. 2007; 39(5):638–644. [PubMed: 17401364] [■ ■ **Describes several regions on the 8q24 locus that confer susceptibility to different cancers.**]
71. Kiemeny LA, Thorlacius S, Sulem P, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet*. 2008; 40(11):1307–1312. [PubMed: 18794855] [■ ■ **Report describes several regions on the 8q24 locus that confer susceptibility to different cancers.**]
72. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet*. 2007; 39(8):984–988. [PubMed: 17618284] [■ ■ **Report describes several regions on the 8q24 locus that confer susceptibility to different cancers.**]

73. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007; 39(5):645–649. [PubMed: 17401363] **■ Report describes several regions on the 8q24 locus that confer susceptibility to different cancers.**
74. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2007; 39(8):989–994. [PubMed: 17618283] **■ Report describes several regions on the 8q24 locus that confer susceptibility to different cancers.**
75. Quarmby LM, Mahjoub MR. Caught Nek-ing: cilia and centrioles. *J Cell Sci.* 2005; 118(Pt 22): 5161–5169. [PubMed: 16280549]
76. Chen Y, Choong LY, Lin Q, et al. Differential expression of novel tyrosine kinase substrates during breast cancer development. *Mol Cell Proteomics.* 2007; 6(12):2072–2087. [PubMed: 17855441]
77. Min J, Okada S, Kanzaki M, et al. Synip: a novel insulin-regulated syntaxin 4-binding protein mediating GLUT4 translocation in adipocytes. *Mol Cell.* 1999; 3(6):751–760. [PubMed: 10394363]
78. Cavdar KE, Burkhardt W, Blackburn K, Moseley A, Spremulli LL. The small subunit of the mammalian mitochondrial ribosome. Identification of the full complement of ribosomal proteins present. *J Biol Chem.* 2001; 276(22):19363–19374. [PubMed: 11279123]
79. Cavdar KE, Ranasinghe A, Burkhardt W, et al. A new face on apoptosis: death-associated protein 3 and PDCD9 are mitochondrial ribosomal proteins. *FEBS Lett.* 2001; 492(1–2):166–170. [PubMed: 11248257]
80. Theodorou V, Boer M, Weigelt B, Jonkers J, van der Valk N, Hilken J. Fgf10 is an oncogene activated by MMTV insertional mutagenesis in mouse mammary tumors and overexpressed in a subset of human breast carcinomas. *Oncogene.* 2004; 23(36):6047–6055. [PubMed: 15208658]
81. Hashimoto T, Shindo Y, Souri M, Baldwin GS. A new inhibitor of mitochondrial fatty acid oxidation. *J Biochem.* 1996; 119(6):1196–1201. [PubMed: 8827458]
82. Clemons M, Goss P. Estrogen and the risk of breast cancer. *N Engl J Med.* 2001; 344(4):276–285. [PubMed: 11172156]
83. Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst.* 2002; 94(8): 606–616. [PubMed: 11959894]
84. Toniolo PG, Levitz M, Zeleniuch-Jacquotte A, et al. A prospective study of endogenous estrogens and breast cancer in postmenopausal women. *J Natl Cancer Inst.* 1995; 87(3):190–197. [PubMed: 7707406]
85. Holst F, Stahl PR, Ruiz C, et al. Estrogen receptor α (ESR1) gene amplification is frequent in breast cancer. *Nat Genet.* 2007; 39(5):655–660. [PubMed: 17417639]
86. Gartner CE, Barendregt JJ, Hall WD. Multiple genetic tests for susceptibility to smoking do not outperform simple family history. *Addiction.* 2009; 104(1):118–126. [PubMed: 19133896]
87. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. *Hum Mol Genet.* 2008; 17(R2):R166–R173. [PubMed: 18852206] **■ This report mainly describes individual genetic risk prediction in common diseases.**
88. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med.* 2008; 358(26):2796–2803. [PubMed: 18579814] **■ This report mainly describes individual genetic risk prediction in common diseases.**
89. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev.* 2008; 18(3):257–263. [PubMed: 18682292] **■ This report mainly describes individual genetic risk prediction in common diseases.**
90. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst.* 2008; 100(14):1037–1041. [PubMed: 18612136] **■■ This report mainly describes individual genetic risk prediction in common diseases.**
91. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet.* 2008; 17(R2):R156–R165. [PubMed: 18852205] **■■■ Interesting review discussing the next steps that may help accelerate progress from genome-wide studies to biological application.**

Websites

101. American Cancer Society – Statistics for 2009. www.cancer.org/docroot/STT/STT_0.asp
102. International HapMap Project. www.hapmap.org/hapmappopulations.html.en
103. Sanger Institute: GENEVAR - GENE Expression VARIation. www.sanger.ac.uk/humgen/genevar/

Executive summary

Three categories of breast cancer susceptibility alleles have been identified to date according to their population frequency and the risk they confer

- Rare, high-penetrance alleles that confer a risk greater than tenfold in carriers.
- Rare, moderate-penetrance alleles that confer a two- to four-fold increase in risk.
- Frequent, low-penetrance alleles (minor allele frequency > 0.05) that confer a less than 1.5-fold increase in risk.

Although it started much earlier than genome-wide association studies, the candidate gene approach has been relatively disappointing

- The candidate gene approach relies on existing knowledge of biological pathways and gene function.
- Several positive associations have been reported with breast cancer, but only one has been convincingly replicated and achieved a significance level close to that required for genome-wide studies (*CASP8*).

The outcome of genome-wide association studies in breast cancer and other multifactorial diseases has been very promising

- To date, genome-wide association scans have revealed ten genes/loci involved in the polygenic susceptibility of breast cancer, and another possible one is awaiting replication.
- None of these loci were previously detected by candidate gene-association studies, emphasizing the limitations of the candidate gene approach.
- The current sets of loci explain only a small fraction of the overall familial risk of breast cancer (approximately 5% of breast cancer risk).

Clinical relevance of the low-penetrance susceptibility variants

- In general, individual SNPs have extremely limited clinical utility.
- Polygenic risk profiles may provide useful risk stratification in the context of population-based screening programmes, particularly if combined with lifestyle risk factors.
- It would be possible to genotype every woman at all known susceptibility loci and, based on her breast-cancer risk profile, offer a personalized screening program in which starting age would vary.

The need of new strategies to find novel genes/loci

- There is a need for larger studies combining multiple scans and joint analysis of data in order to enhance power and test for interactions between genes.
- Many more loci may be identified through genome-wide studies performed in different ethnic groups and population isolates where allele frequency and correlation of susceptibility variants, together with environmental factors and behaviors, may differ from Europeans.
- The most comprehensive strategy towards understanding the architecture of breast cancer or any other complex disease would be a full resequencing of the genome in a large cohort of cases and controls.

Figure 1. Fraction of the excess familial risk of breast cancer explained by the currently known susceptibility genes/loci.

This assumes that most of the familial risk is due to shared genes rather than shared environment.

Figure 2. The proportion of cases of breast cancer accounted for by the proportion of the population at highest risk.

Table 1

Summary of the eleven well-validated common low-penetrance variants associated with breast cancer to date.

Gene(s) in/nearby the region	Locus	Variant	MAF	Relative risk per allele(95% CI)	p-value	Ref.
<i>FGFR2</i>	10q26	rs2981582	0.38	1.26 (1.23–1.30)	2×10^{-76}	[27]
	10q26	rs1219648	0.39	1.32 (1.17–1.49)	1.1×10^{-10}	[29]
<i>TNRC9/LOC643714</i>	16q12	rs3803662	0.25	1.20 (1.16–1.24)	1×10^{-36}	[27]
	16q12	rs3803662	0.27	1.28 (1.21–1.35)	5.9×10^{-19}	[30]
<i>NEK10/SLC4A7</i>	3p24	rs4973768	0.46	1.11 (1.08–1.13)	4.1×10^{-23}	[32]
<i>MAP3K1/MGC33648/MIER3</i>	5q11	rs889312	0.28	1.13 (1.10–1.16)	7×10^{-20}	[27]
<i>ESR1</i>	6q25	rs2046210	0.36	1.29 (1.21–1.37)	2.0×10^{-15}	[33]
<i>TNP1/IGFBP5/IGFBP2/TNS1</i>	2q35	rs13387042	0.50	1.20 (1.14–1.26)	1.3×10^{-13}	[30]
<i>FAM84B/c-MYC</i>	8q24	rs13281615	0.40	1.08 (1.05–1.11)	5×10^{-12}	[27]
<i>MRPS30/FGFR10</i>	5p12	rs10941679	0.24	1.19 (1.13–1.26)	2.9×10^{-11}	[31]
<i>LSP1/H19</i>	11p15	rs3817198	0.30	1.07 (1.04–1.11)	3×10^{-9}	[27]
<i>COX11/STXBP4/TOM1L1</i>	17q22	rs6504950	0.27	0.95 (0.92–0.97)	1.4×10^{-8}	[32]
<i>CASP8</i>	10p14	rs1053485 (D302H)	0.13	0.88 (0.84–0.92)	1.1×10^{-7}	[26]

The odds ratio per allele is relative to the common homozygous p-value (per allele effect under multiplicative model).

MAF: Minor allele frequency.