npg

REVIEW

Breast cancer genome-wide association studies: there is strength in numbers

D Fanale[1,3], V Amodeo[1,3], LR Corsini[1,3], S Rizzo[1], V Bazan[1,2] and A Russo[1,2]

[1]Department of Surgical and Oncological Sciences, Section of Medical Oncology, University of Palermo, Palermo, Italy and
[2]Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, College of Science and Technology, Temple University, Philadelphia, PA, USA

Breast cancer (BC) is a heterogeneous disease that exhibits familial aggregation. Family linkage studies have identified high-penetrance genes, *BRCA1*, *BRCA2*, *PTEN* and *TP53*, that are responsible for inherited BC syndromes. Moreover, a combination of family-based and population-based approaches indicated that genes involved in DNA repair, such as *CHEK2*, *ATM*, *BRIP* and *PALB2*, are associated with moderate risk. Therefore, all of these known genes account for only 25% of the familial aggregation cases. Recently, genome wide association studies (GWAS) in BC revealed single nucleotide polymorphisms (SNPs) in five novel genes associated to susceptibility: *TNRC9*, *FGFR2*, *MAP3K1*, *H19* and lymphocyte-specific protein 1 (*LSP1*). The most strongly associated SNP was in intron 2 of the *FGFR2* gene that is amplified and overexpressed in 5–10% of BC. rs3803662 of *TNRC9* gene has been shown to be the SNP with the strongest association with BC, in particular, this polymorphism seems to be correlated with bone metastases and estrogen receptor positivity. Relevant data indicate that SNP rs889312 in *MAP3K1* is correlated with BC susceptibility only in *BRCA2* mutation carriers, but is not associated with an increased risk in *BRCA1* carriers. Finally, different SNPs in *LSP1* and *H19* and in minor genes probably were associated with BC risk. New susceptibility allelic variants associated with BC risk were recently discovered including potential causative genes involved in regulation of cell cycle, apoptosis, metabolism and mitochondrial functions. In conclusion, the identification of disease susceptibility loci may lead to a better understanding of the biological mechanism for BC to improve prevention, early detection and treatment.
Oncogene advance online publication, 26 September 2011; doi:10.1038/onc.2011.408

Correspondence: Professor A Russo, Department of Surgical and Oncological Sciences, Section of Medical Oncology, Università di Palermo, via del Vespro 129, 90127 Palermo, Italy.
E-mail: lab-oncobiologia@usa.net
[3]These authors contributed equally to this work

Introduction

Breast cancer (BC) is the most common cancer and the second leading cause of cancer death among women (Parkin *et al.*, 2005).

The family history is the main risk factor for BC, indicating that the genetic factors are very important in the development of disease (Antoniou and Easton, 2006).

In the 1990s, linkage studies in multiple case families have identified two major susceptibility genes in BC *BRCA1* and *BRCA2* (Miki *et al.*, 1994; Wooster *et al.*, 1995).

Germline mutations in *BRCA1* and *BRCA2* genes occur rarely in the general population but confer high risks of breast and ovarian cancer and a lower risk for other cancers (Antoniou *et al.*, 2003; Thompson and Easton, 2004).

TP53 and *PTEN* mutations are also present in the population at low frequency and lead to very high BC risk associated with rare cancer syndrome, however, population-based studies have estimated that alterations in these genes account only the 15% of the familial risk of BC (Sidransky *et al.*, 1992; FitzGerald *et al.*, 1998; Peto *et al.*, 1999; Dite *et al.*, 2003).

Further, genetic linkage analyses failed to identify additional high-penetrance susceptibility genes and the identification of rare variants of genes involved in DNA repair, such as *CHEK2*, *ATM*, *BRIP* and *PALB2* in families lacking *BRCA* mutations (Meijers-Heijboer *et al.*, 2002; Thompson *et al.*, 2005; Rahman *et al.*, 2007; Hollestelle *et al.*, 2010), associated with a moderate risk of disease, can explain only a small portion of familial risk.

Therefore, all of these known genes account for only 25% of the familial aggregation cases (Thompson and Easton, 2004), suggesting that most of the familial risk of BC can plausibly involve a combination of multiple low-penetrance susceptibility alleles, each conferring a small effect on BC risk (Antoniou and Easton, 2006; Table 1).

According to this model defined 'polygenic', proposed to explain the genetic susceptibility to BC, a large number of low-risk variants occurs with high frequency in populations, therefore, it may have a multiplicative effect in determining the overall risk of disease (Pharoah *et al.*, 2002; Figure 1). A significant part of polygenic contribute to low-penetrance susceptibility may rise by non-conservative missense mutations in evolutionarily conserved domains.

Table 1 Genetic loci implicated in breast cancer susceptibility

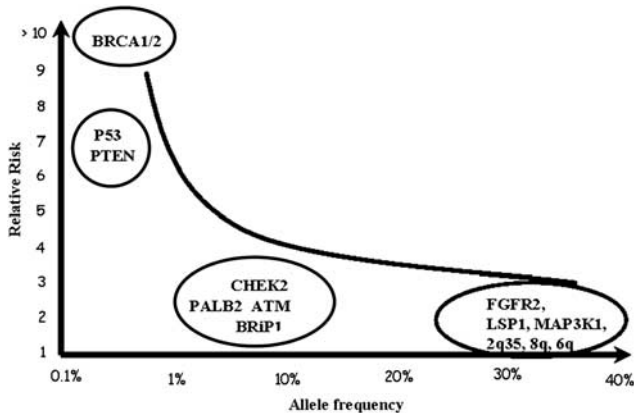| High penetrance, low frequency	Low penetrance, low frequency	Low penetrance, high frequency
BRCA1	CHEK2	FGFR2
BRCA2	ATM	TNRC9
p53	PALB2	LSP1
PTEN	BRIP1	MAP3K1
		SLC4A7
		COX11

Figure 1 Allelic variants at low frequency tend to be associated with higher relative risk of BC (for example, BRCA1, BRCA2), high-frequency allelic variants are associated with lower RR (for example, FGFR2, LSP1, and so on) configuring an inverse correlation.

Genome-wide association studies (GWAS)

In recent years the research of low-penetrance allelic variants was conducted mainly through GWAS. These studies use a large number of common genetic single nucleotide polymorphisms (SNPs) to identify associations with disease that rely upon patterns of linkage disequilibrium (LD) in the human genome (Hirshfield et al., 2010). The power of GWAS is to evaluate the association of genetic variants at different loci on different chromosomes (LD) in large series of cases versus controls, analyzing a panel of hundred thousand SNPs simultaneously, to identify new alleles of susceptibility to BC (Orr and Chanock, 2008). In the human genome has been estimated that there are seven million of common SNPs that have a minor allele frequency (m.a.f.), >5% and because recombination occurs in different hot-spots, the nascent polymorphisms are often strongly correlated.

These studies therefore provide a powerful tool to identify novel markers for susceptibility and prognosis of disease (Peto, 2002; Houlston and Peto, 2004; Easton and Eeles, 2008). In the GWA studies the accumulation of a large number of data is crucial. Houlston and Peto, (2004) have estimated the number of cases required to identify low-penetrance alleles conferring a relative risk of two both in an unselected population and in families with first-degree relatives affected. In an unselected

population the identification of a susceptibility allele with a frequency of 5% requires over 800 cases. In the same population, the identification of a susceptibility allele with a frequency of 1% requires over 3700 unselected cases, whereas about 700 would be enough if three affected families are selected. Therefore, the power of association studies can be significantly increased using selected cases with a family history of cancer because less cases are required to demonstrate the association with disease (Houlston and Peto, 2004).

The potential of the association studies of cases with a family history to identify low-penetrance alleles conferring a relative risk of 2 has been demonstrated by the mutation CHEK2 1100delC in patients with BC. This variant carried by 1% of the population confers an increased risk of 1.7-fold. The frequency was not significantly increased in unselected cases (1.4%), but it was strongly increased in familial cases without BRCA1 and BRCA2 mutations (5.1%; Meijers-Heijboer et al., 2002).

In the past years several novel risk alleles for BC were identified by four recent GWA studies: Breast Cancer Association Consortium, Cancer Genetic Markers of Susceptibility, DeCode Islanda, Memorial Sloan–Kettering Cancer Center (Easton et al., 2007; Hunter et al., 2007; Gold et al., 2008; Stacey et al., 2008).

In each of them the association study was shared into three phases: the first phase identifies the common SNPs in cases and controls, the second phase evaluates how many of the above SNPs are common to a greater number of cases and controls and, finally, the third phase aims to identify new alleles of susceptibility of BC.

Easton et al., in their study, identified five independent loci associated with increased susceptibility to BC ($P < 10^{-7}$). This multistage study involved in the first stage 390 BC cases with a strong family history and 364 controls, and 3990 cases and 3916 controls in the second stage.

To define the risk associated with the 30 most significant SNPs, a third stage of the study was conducted involving 21 860 cases and 22 578 controls from 22 additional studies in the Breast Cancer Association Consortium.

These combined analyses allowed to observe that the SNPs showing a stronger statistical evidence of association with an increased familial risk were: rs2981582 lies in intron 2 of FGFR2, rs12443621 and rs8051542 within TNRC9, rs889312 lies in a region that contain MAP3K1 gene, rs3817198 lies in intron 10 of lymphocyte-specific protein 1 (LSP1) and rs2107425 within the H19 gene.

Unlike other BC susceptibility genes previously identified that are involved in DNA repair and sex hormone synthesis, in this work three of the five loci reported contain genes involved in regulation of cell growth and cell signaling (Easton et al., 2007).

Starting from this study of Easton et al., in attempt to identify further loci associated with BC risk, Ahmed et al. have genotyped in a third stage further 814 SNPs, involving 3878 cases and 3928 controls from three studies of the Cancer Genetic Markers of Susceptibility.

These analyses allowed to identify three additional SNPs (rs4973768, rs4132417 and rs6504950) that have been evaluated in a fourth stage from 27 studies in the Breast Cancer Association Consortium.

rs4132417 showed no evidence of association in the fourth stage, it was probably a false positive. rs4973768 showed strong association with age that was higher for ER-positive than ER-negative disease, moreover there was no evidence of association with a positive family history of BC.

Similarly, rs6504950 showed statistical evidence of association with ER-positive disease and no association with the family history, and unlike of rs4973768 no association with the age.

Moreover, in this study additional association analyses showed that another SNP rs1357245 is located in the same LD block as rs4973768 in the 3p24 region (Ahmed et al., 2009). Genotyping the 28 SNPs, known to be present in this region, in 2301 cases and 2256 controls, it presents a further SNP rs2307032 that was correlated with both SNPs.

Hunter et al., in a recent work, have identified alleles in FGFR2 associated with risk of sporadic postmenopausal BC.

In this study, the National Cancer Institute Cancer Genetic Markers of Susceptibility identified four SNPs, two (rs1219648 and rs2420946) in intron 2 of FGFR2 and two (rs11200014 and rs28981579) in FGFR2 gene, using GWAS of BC by genotyping 528 173 SNPs in 1145 postmenopausal women of European ancestry with invasive BC and 1142 controls.

These polymorphic variants showed a strong association with the risk of disease (Hunter et al., 2007).

A similar experimental design was conducted by Gold et al. in 249 Ashkenazi Jewish women, containing multiple cases of BC but lacking BRCA1 or BRCA2 mutations, presented at the Memorial Sloan–Kettering Cancer Center.

This study confirmed the BC association with the FGFR2 locus, identified by Easton and Hunter studies, but showed an association with the RNF146 and ECHDC1 region at 6q22.33 not seen in the previous works (Gold et al., 2008).

ECHDC1 gene encodes for a trifunctional protein involved in mitochondrial fatty acid oxidation (Hashimoto et al., 1996) and RNF146 encodes for a protein, called actylidin, that functions as a ubiquitin protein ligase (E3; Mani and Gelmann, 2005), that could have a role in breast tumorigenesis.

To identify new risk variants associated with BC susceptibility, Stacey et al. have carried out a GWA study genotyping 1600 Icelandic individuals with BC and 11 563 controls.

Two SNPs showed statistically significant association with BC: the A allele of rs13387042 on chromosome 2q35 and the T allele of rs3803662 on 16q12.

The 25% of individuals of European descent are homozygous for allele A of rs13387042 and have an estimated 1.44-fold greater risk than noncarriers and about 7% are homozygous for allele T of rs3803662 and have a 1.64-fold greater risk.

These risk alleles were not associated with histopathological subtype, stage and grade of tumors, but confer preferential risk for estrogen receptor (ER)-positive BC.

In the LD block where lies rs13387042 there are no known genes, but there are proximally and distally TNP1, IGFBP5, IGFBP2 and TNS1 genes (Stacey et al., 2007).

rs3803662 is near the 5′ end of TNRC9, whose increased expression is highly predictive of metastasis to bone of BC (Smid et al., 2006).

Comparing the results obtained from four major studies of GWA, it has been highlighted a correlation of allele frequency of some SNPs located on the genes: FGFR2 encoding a receptor tyrosine kinase, TNRC9 encoding a high-mobility group chromatin-associated protein, MAP3K1, which encodes the signaling protein mitogen-activated protein kinase 1 (MAPK1), LSP1 encoding LSP1 and H19 an untranslated messenger RNA involved in regulation of the insulin growth factor gene 2 (Table 2).

These new discovered susceptibility genes are differentially expressed between the five distinct molecular subtypes of BC, based on differential gene expression profiles: luminal A, luminal B, basal like, ErbB2 + and normal like (Sorlie et al., 2003). These distinct molecular subtypes of BC are associated with different clinical outcomes (Sorlie et al., 2001).

If the probability to develop a given subtype of BC is genetically determined, we would expect to find that the

Table 2 Comparative analysis of the SNPs identified in the four studies (BCAC, CGEMS, MSKCC and DeCode Islanda)

Gene	Location	BCAC	CGEMS	MSKCC	DeCode Islanda
FGFR2	Chr 10q	rs2981582 (in intron2)	rs1219648 rs2420946 (in intron2) rs11200014 and rs28981579	rs2981582	No
TNRC9	Chr 16q	rs12443621 rs8051542	rs8049226 (within 200 kb of TNRC9)	rs3803662 and rs3112625	rs3803662 (near TNRC9)
MAP3K1	Chr 5q	rs889312	rs726501	No	No
LSP1	Chr 11p	rs3817198 rs498337	rs7120258	No	No
H19	Chr 11p	rs2107425	rs7120258 rs7578974	No	No

Abbreviation: BCAC, Breast Cancer Association Consortium; CGEMS, Cancer Genetic Markers of Susceptibility; MSKCC, Memorial Sloan–Kettering Cancer Center; SNPs, single nucleotide polymorphisms.

newly discovered susceptibility genes (Easton et al., 2007) are differentially expressed in the various tumor subtypes.

Recently a significantly differential mRNA expression of TNRC9, FGFR2, MAP3K1, H19 and LSP1 from 112 breast tumor samples, representing all five subtypes, has been identified by analysis of variance (Nordgard et al., 2007).

These data show the necessity to conduct stratified SNP disease association studies and to select patients by their molecular subtypes, to confer more power to the GWA studies.

FGFR2

FGFR2 is a member of a receptor tyrosine kinase gene superfamily, which contributes to the cell growth, invasiveness, motility and angiogenesis (Ricol et al., 1999). Overexpression of FGFR2, one of the common low-penetrance susceptibility genes, is observed in breast tumor tissues (Adnane et al., 1991) and in BC cell lines (Tannheimer et al., 2000). Its expression is associate with ER + tumors (Luqmani et al., 1992), suggesting a hormone-dependent action of this gene. Recently, gene expression studies have shown increasing FGFR2 expression levels associated with the rare homozygote genotype and functional studies identified the OCT1/RUNX2-binding site as the main determinant of the increased expression levels (Meyer et al., 2008). Aberrant expression of alternatively spliced isoforms of FGFR2 transforms BC cells by sustained signal transduction (Moffa and Ethier, 2007). The FGFR2 gene, located at chromosome 10q26, contains at least 22 exons (Ingersoll et al., 2001).

Several mutations and common SNPs within or flanking the FGFR2 gene have been identified. A number of studies have been conducted to investigate the association between FGFR2 polymorphisms and the risk of BC in humans. The association is restricted to SNPs in the LD block covering intron 2. In particular, three polymorphic variants, rs1219648 (A > G), rs2420946 (C > T) and rs2981582 (C > T) are more investigated for their closed correlation with BC. Easton et al. (2007) showed that rs2981582 had a clear relevance to BC.

Gold et al. (2008) confirmed the previously reported results for FGFR2 locus. Recently, a further GWAS study confirmed the correspondence between FGFR2 susceptibility loci and BC risk. In particular, the per-allele odds ratio was higher for ER-positive rather than for ER-negative BC (Ahmed et al., 2009). This finding is consistent with the involvement of FGFR2 in estrogen-related breast carcinogenesis (Tamaru et al., 2004), and with higher levels of FGFR2 expression in ER + than ER− cell lines and tumors (Zhang et al., 1999). Stacey et al. genotyped ∼300 000 SNPs in 1600 Icelandic individuals with BC and 11 563 controls. They found that 25% of individuals of European descent are homozygous for allele A of rs13387042 on chromosome

2q35 and have an estimated 1.44-fold greater risk than noncarriers. Risk from both alleles was confined to ER-positive tumors.

The variant in the 5p12 region, which is close to the FGFR2 ligand FGF10, also shows strong evidence of an association primarily with ER + tumors (Stacey et al., 2008).

TNRC9

The locus on 16q includes a gene TNRC9 and a hypothetical gene LOC643714. TNCR9 (also known as TOX3) is a gene of uncertain function containing a tri-nucleotide repeat motif and encoding a member of the high-mobility group family of non-histone chromatin proteins. The presence of a putative high-mobility group box motif suggests that it might function as a transcription factor (Easton et al., 2007). Several studies have shown that susceptibility loci at TNCR9 predispose to sporadic BC. Rs3803662, located near the 5′ end of TNRC9, has been shown to be the SNP with the strongest association with BC. The SNP rs3803662 is related to both ER + and ER− tumors (McInerney et al., 2009).

The associations of rs3803662 with other SNPs seem to be not significant. Other two SNPs (rs12443621 and rs8051542) with important evidence of association are located in an LD block containing the 5′ end of TNRC9. Furthermore, Hunter et al. (2007) showed that there is only one SNP significant (rs8049226) within 200 kb of TNRC9. In contrast, the coding region of TNCR9 contains SNPs showing no evidence of association. The A allele of rs13387042 located on chromosome 2q35 (A-rs13387042) and the T allele of rs3803662 on 16q12 (T-rs3803662) confer increased risk of BC for ER-positive tumors. Any interaction was observed between the 2q35 and 16q12 loci. Moreover, no known gene or human RNA was found for the LD block containing rs13387042. The rs3803662 SNP located on 16q12 occurs in the fourth exon of a poorly characterized mRNA. In BC, the q arm of chromosome 16 is frequently lost, therefore, it is likely that one or more tumor suppressor genes are present in the same region. Differences in stage, grade or histopathological subtype were not significantly correlated with the low-penetrance susceptibility alleles, and there was no significant difference in allele frequencies between in situ and invasive carcinoma. In African Americans, T allele of the SNP rs3803662 was significantly protective and, thus, it was not associated with increased BC risk (Stacey et al., 2007). Three susceptibility alleles (rs2981582, rs3803662 and rs13281615) also have shown an evidence of association with family history of BC. In fact, each of these SNPs was more frequent in women with a first-degree relative with the disease than in those without. Furthermore, an evidence of association with breast in situ carcinoma has been shown by three SNPs (rs2981582, rs3803662 and rs889312; Easton et al., 2007). Increased expression of TNRC9 indicates a major

susceptibility to metastasis of BC to bone. ER positivity is predictive of bone metastases. The possible effects of the correlation between rs3803662, *TNRC9*, bone metastases and ER positivity remain to be explicated.

Many association studies have shown that SNPs in *FGFR2*, *TNRC9* and *MAP3K1* increase the BC risk in *BRCA2* mutation carriers with a similar relative risk to that seen in the general population. In contrast, in *BRCA1* mutation carriers only the rs3803662 SNP was associated with an increased BC risk (Easton and Eeles, 2008).

MAP3K1

MAP3K1 (MEKK1) encodes the MAPK protein that phosphorylates and activates the MAPK kinase (MAPK2) that in turn phosphorylates the MAPK/ERK to produce downstream signaling effects on a variety of cancer genes. MAP3K1 forms part of the MAPK cell signaling pathway implicated in cellular response to mitogens. The MAPK pathway is strongly linked to HER2 receptor activity and activating mutations in the MAPK pathway have been associated with HER2 + breast tumors (Bild *et al.*, 2006; Creighton *et al.*, 2006). *MAP3K1* was identified by Easton *et al.* (2007) to have a per-allele odds ratio effect of 1.13 (95% confidence interval: 1.09–1.18). MAP3K1 effects were found to be relevant in ER + and PR + tumors to a greater degree than in ER− or PR− tumors (Garcia-Closas and Chanock, 2008). MAP3K1 is differentially expressed in different BC subtypes (Nordgard *et al.*, 2007). Hunter *et al.* (2007) found only one SNP (rs726501) with a *P* value in the range of $P < 0.01$ by allele test. Gold in the Memorial Sloan–Kettering Cancer Center study and Stacey, in the DeCode study, did not see significant SNPs between individuals with BC and controls. GWA studies conducted by Garcia-Closas and Chanock identified that the rs889312 variant is in a LD block containing the *MAP3K* (Garcia-Closas *et al.*, 2008).

The Consortium of Investigators of Modifiers of BRCA1/2 has recently evaluated whether variants in *FGFR2* (rs2981582), *TNRC9* (rs3803662) and *MAP3K1* (rs889312) are associated with the risk of BC in over 10 000 *BRCA1* and *BRCA2* mutation carriers from 23 studies (Antoniou *et al.*, 2008). The evidence of association with SNP rs889312 in *MAP3K1* was weaker and was restricted to *BRCA2* mutation carriers, however, this SNP was not associated with an increased risk in *BRCA1* carriers.

LSP1 and H19

LSP1 gene (also known as WP43) encodes an F-actin bundling cytoskeletal protein expressed in hematopoietic and endothelial cells. *LSP1* has been implicated in malignant lymphoma and Hodgkin's disease (Marafioti *et al.*, 2003), and other variants in this gene have been

associated with risk of developing non-Hodgkin's lymphoma (Cerhan *et al.*, 2007). The most important GWASs reported different conclusion about the role of *LSP1* gene in BC susceptibility. Easton *et al.* (2007) reported one SNP (rs3817198) lies in intron 10 of *LSP1* gene with *P* values in the range 10^{-5}–10^{-9}; Gold *et al.* (2008) found two SNPs (rs3817198, rs498337), near the *LSP1* region, with *P* values in the range of $P < 0.01$ by allele test, where Hunter *et al.* (2007) provided evidence for one SNP (rs7120258) in the region with a *P* value 0.01. Recent study identified that *LSP1* minor allele of rs3817198 was associated with increased BC risk only for BRCA2 mutation carriers (Antoniou *et al.*, 2009). A further SNP, rs2107425, located just 110 kb from rs3817198, was also identified (overall $P = 0.00002$). rs2107425 is within the *H19* gene, an imprinted maternally expressed untranslated messenger RNA closely involved in regulation of the insulin growth factor gene 2 (Easton *et al.*, 2007). In *H19* region on chromosome 11p, Easton *et al.* (2007) reported *P* values in the range 0.01–10^{-5}, Gold *et al.* saw no signal, whereas Hunter *et al.* (2007) found two SNPs (rs7120258, rs7578974), with association *P* values in the range of 0.01, with one additional SNP, rs217228, with a *P* value in the range of 0.02.

Recently discovered BC susceptibility loci

New susceptibility allelic variants associated with BC risk were recently discovered through large replication studies in combination with the original GWAS data. The combination between these studies and GWAS data allowed to identify three SNPs: rs4973768, rs4132417 and rs6504950. There is a strong evidence for additional susceptibility loci on 3p and 17q. The region 3p24 includes two potential causative genes, *SLC4A7* and *NEK10*. SLC4A7 (solute carrier family 4, sodium carbonate cotransporter, member 7) is a potential tyrosine kinase protein whose expression is reduced in BC specimens and cell lines. NEK10 (never-in mitosis-related kinase 10) belongs to a family of 11 never in mitosis a-related kinases that are involved in cell cycle regulation. However, unlike other NEKs, no role has been associated to NEK10. A 300-kb LD block on 17q23.2 includes rs6504950 that lies in intron 1 of *STXBP4* (syntaxin-binding protein 4), codifying an insulin-regulated STX4-binding protein involved in the regulation of GLUT4 vesicle translocation and glucose transport. The rs6504950 allelic variant showed no association with age or family history, but a stronger association in ER-positive disease versus ER-negative disease. The same LD block includes other genes as *COX11* (cytocrome C assembly protein 11), that is located 10 kb upstream of rs6504950, and *TOM1L1* (target of myb-1-like1). In lymphocytes, the rs6504950 risk allele is correlated with higher expression of *COX11* levels, but no association has been shown with expression levels of either *TOM1L1* or *STXBP4*. Allele frequency studies in European populations have revealed that rs4973768 and rs6504950 could explain

6

respectively 0.4% and 0.07% of the familial risk of BC. These susceptibility loci together with those previously identified in original GWAS would give rise to 5.9% of BC familial risk (Ahmed et al., 2009). Further genome-wide linkage studies have revealed three putative BC susceptibility regions of interest, located on 3q25, 6q24 and 21q22. Moreover, it has been observed that the allelic variants on both chromosomes 21 and 3 were correlated with a higher percentage of bilaterality and a higher number of familial cases (Rosa-Rosa et al., 2009).

A recent GWAS has identified a novel polymorphic variant rs11249433 within the 1p11.2 region associated with BC risk.

This association is stronger in ER-positive than ER-negative tumors, is correlated with mRNA expression of NOTCH2 gene and is highest in breast tumors without TP53 mutations.

Further studies are needed to evaluate the possible role of rs11249433 in NOTCH2 regulation and BC development (Fu et al., 2010). Other variants can significantly modify the BC risk in BRCA1 and BRCA2 mutations carriers. The rs6138178 in SNRPB and rs6602595 in CAMK1D show a strongest association in BRCA1 carriers, whereas rs9393597 in LOC134997 and rs12652447 in FBXL7 in BRCA2 carriers.

These loci appeared to interact multiplicatively for BC risk in BRCA1/BRCA2 carriers, therefore, these SNPs together with other genetic and environmental factors may improve the BC risk assessment in these populations (Wang et al., 2010).

Recently, a new study (Black Women's Health Study) has been conducted in a population of African American Women (886 BC cases versus 1089 controls) to identify genetic variants associated with risk of BC. As in the original study (Stacey et al., 2008), it has been confirmed the strong association of rs4415084 on 5p12 with overall risk and ER-positive tumors. No association was observed for ER- and PR-negative tumors. Other susceptibility allelic variants identified from BWHS are rs6451770, rs12515012, rs13156930 and rs16901937. A 21% increase in risk was associated with each copy of the rs16901937 G-allele. The closest gene to these regions is MRPS30, involved in apoptosis and encoding a component of the mitochondrial ribosome. Moreover, MRPS30 is involved in a gene expression profile that allows to discriminate ER-positive from ER-negative tumors (Ruiz-Narvaez et al., 2010). Other BC susceptibility alleles could be identified through large-scale replication studies in combination with previous GWAS. However, these analyses have still a limited power.

In a recent work, the BC risk association with eight susceptibility loci identified by GWAS was investigated in relation to specific breast tumor subtypes. A strong association was identified between ER + tumors and six of eight loci identified by GWAS: rs2981582 (10q26), rs3803662 (16q12), rs13281615 (8q24), rs13387042 (2q35), rs4973768 (3p24) and rs6504950 (17q23). A most strongly relation was observed between two candidate loci, CASP8 (rs1045485, rs17468277) and TGFB1 (rs1982073) and PR tumors. Four loci were associated with triple negative tumors ($P \leqslant 0.016$):

rs3803662 (16q12), rs889312 (5q11), rs3817198 (11p15) and rs13387042 (2q35) but only two of them (16q12 and 2q35) were associated with tumors with the core basal phenotype ($P \leqslant 0.002$). This study identifying novel risk factors associated with BC subtypes could allow a better tumor stratification resulting in prevention, early detection and treatment improvement (Broeks et al., 2011).

The power of GWAS

The GWAS represents a new powerful approach to identify lower penetrance alleles that cannot be detected by genetic linkage studies. The risk conferred by these alleles individually is too weak, generally 1.3-fold or less, but the combined effects may be useful for risk prediction (Easton et al., 2007). This would promote the development of novel methodologies for analysis of data generated by large-scale SNP studies. In recent years, the research and identification of low-penetrance susceptibility loci played a key role in the etiology of BC and, in particular, of those BCs that have estrogen and progesterone receptors. The combination of BC susceptibility alleles together with other risk factors may be important clinically and it may explain an appreciable fraction of the genetic variance in BC risk. The identification of the causative variants can be extremely problematic but the use of GWAS from multiple populations with different patterns of LD can reduce the difficulty of analysis. The power of GWAS may be increased by enlarging the number of samples in both the cases and the controls, and by identifying clinical and molecular subtypes (Kristensen and Borresen-Dale, 2008). However, the GWA experiments need the effort of several research groups to collect a sufficient number of patients for large multistage studies and they require large amounts of money. The allele frequency of the variant and the risk conferred by it will determine the number of cases to be genotyped. There is a common scepticism toward these new approaches because it is not known the mechanism by which the novel allelic variants cause the susceptibility. Furthermore, some differences were found between different studies. This could be due to population stratification, sample-size differences or genetic heterogeneity in the setting of different genotyping platforms (Perlegen, Mountain View, CA, USA; Affymetrix, Santa Clara, CA, USA; Illumina, San Diego, CA, USA) and different algorithms to filter data (Gold et al., 2008).

Recently, a novel multi-SNP GWAS analysis method called Pathways of Distinction Analysis was developed. This method uses GWAS data and pathway-gene and gene-SNP associations to identify pathways that could permit the distinction of cases from controls. Therefore, relating a pathway with the disease risk, for the SNPs associated with a pathway, the cases will be similar to other cases than to controls. This method provides a new analytical tool that could enrich the power of GWAS in BC risk prediction (Braun and Buetow, 2011).

In conclusion, the recently discovered data could open up new streets for basic research. In future, a new

generation of large-scale association studies, in combination with replication analyses and multiple scans could be able to identify many more loci.

Conflict of interest

The authors declare no conflict of interest.

References

Adnane J, Gaudray P, Dionne CA, Crumley G, Jaye M, Schlessinger J et al. (1991). BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene* **6**: 659–663.

Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R et al. (2009). Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41**: 585–590.

Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL et al. (2003). 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* **72**: 1117–1130.

Antoniou AC, Easton DF. (2006). Models of genetic susceptibility to breast cancer. *Oncogene* **25**: 5898–5905.

Antoniou AC, Sinilnikova OM, McGuffog L, Healey S, Nevanlinna H, Heikkinen T et al. (2009). Common variants in LSP1, 2q35 and 8q24 and breast cancer risk for BRCA1 and BRCA2 mutation carriers. *Hum Mol Genet* **18**: 4442–4456.

Antoniou AC, Spurdle AB, Sinilnikova OM, Healey S, Pooley KA, Schmutzler RK et al. (2008). Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am J Hum Genet* **82**: 937–948.

Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**: 353–357.

Braun R, Buetow K. (2011). Pathways of distinction analysis: a new technique for multi-SNP analysis of GWAS data. *PLoS Genet* **7**: e1002101.

Broeks A, Schmidt MK, Sherman ME, Couch FJ, Hopper JL, Dite GS et al. (2011). Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum Mol Genet* **20**: 3289–3303.

Cerhan JR, Ansell SM, Fredericksen ZS, Kay NE, Liebow M, Call TG et al. (2007). Genetic variation in 1253 immune and inflammation genes and risk of non-Hodgkin lymphoma. *Blood* **110**: 4455–4463.

Creighton CJ, Hilger AM, Murthy S, Rae JM, Chinnaiyan AM, El-Ashry D. (2006). Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells *in vitro* induces an *in vivo* molecular phenotype of estrogen receptor alpha-negative human breast tumors. *Cancer Res* **66**: 3903–3911.

Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MR et al. (2003). Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. *J Natl Cancer Inst* **95**: 448–457.

Easton DF, Eeles RA. (2008). Genome-wide association studies in cancer. *Hum Mol Genet* **17**: R109–R115.

Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**: 1087–1093.

FitzGerald MG, Marsh DJ, Wahrer D, Bell D, Caron S, Shannon KE et al. (1998). Germline mutations in PTEN are an infrequent cause of genetic predisposition to breast cancer. *Oncogene* **17**: 727–731.

Fu YP, Edvardsen H, Kaushiva A, Arhancet JP, Howe TM, Kohaar I et al. (2010). NOTCH2 in breast cancer: association of SNP rs11249433 with gene expression in ER-positive breast tumors without TP53 mutations. *Mol Cancer* **9**: 113.

Garcia-Closas M, Chanock S. (2008). Genetic susceptibility loci for breast cancer by estrogen receptor status. *Clin Cancer Res* **14**: 8000–8009.

Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA et al. (2008). Heterogeneity of breast cancer

associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* **4**: e1000054.

Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, Garber J et al. (2008). Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci USA* **105**: 4340–4345.

Hashimoto T, Shindo Y, Souri M, Baldwin GS. (1996). A new inhibitor of mitochondrial fatty acid oxidation. *J Biochem* **119**: 1196–1201.

Hirshfield KM, Rebbeck TR, Levine AJ. (2010). Germline mutations and polymorphisms in the origins of cancers in women. *J Oncol* **2010**: 297671.

Hollestelle A, Wasielewski M, Martens JW, Schutte M. (2010). Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* **20**: 268–276.

Houlston RS, Peto J. (2004). The search for low-penetrance cancer susceptibility alleles. *Oncogene* **23**: 6471–6476.

Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE et al. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**: 870–874.

Ingersoll RG, Paznekas WA, Tran AK, Scott AF, Jiang G, Jabs EW. (2001). Fibroblast growth factor receptor 2 (FGFR2): genomic sequence and variations. *Cytogenet Cell Genet* **94**: 121–126.

Kristensen VN, Borresen-Dale AL. (2008). SNPs associated with molecular subtypes of breast cancer: on the usefulness of stratified Genome-wide Association Studies (GWAS) in the identification of novel susceptibility loci. *Mol Oncol* **2**: 12–15.

Luqmani YA, Graham M, Coombes RC. (1992). Expression of basic fibroblast growth factor, FGFR1 and FGFR2 in normal and malignant human breast, and comparison with other normal tissues. *Br J Cancer* **66**: 273–280.

Mani A, Gelmann EP. (2005). The ubiquitin-proteasome pathway and its role in cancer. *J Clin Oncol* **23**: 4776–4789.

Marafioti T, Jabri L, Pulford K, Brousset P, Mason DY, Delsol G. (2003). Leucocyte-specific protein (LSP1) in malignant lymphoma and Hodgkin's disease. *Br J Haematol* **120**: 671–678.

McInerney N, Colleran G, Rowan A, Walther A, Barclay E, Spain S et al. (2009). Low penetrance breast cancer predisposition SNPs are site specific. *Breast Cancer Res Treat* **117**: 151–159.

Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R et al. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**: 55–59.

Meyer KB, Maia AT, O'Reilly M, Teschendorff AE, Chin SF, Caldas C et al. (2008). Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol* **6**: e108.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**: 66–71.

Moffa AB, Ethier SP. (2007). Differential signal transduction of alternatively spliced FGFR2 variants expressed in human mammary epithelial cells. *J Cell Physiol* **210**: 720–731.

Nordgard SH, Johansen FE, Alnaes GI, Naume B, Borresen-Dale AL, Kristensen VN. (2007). Genes harbouring susceptibility SNPs are differentially expressed in the breast cancer subtypes. *Breast Cancer Res* **9**: 113.

Orr N, Chanock S. (2008). Common genetic variation and human disease. *Adv Genet* **62**: 1–32.

Parkin DM, Bray F, Ferlay J, Pisani P. (2005). Global cancer statistics, 2002. *CA Cancer J Clin* **55**: 74–108.

Peto J. (2002). Breast cancer susceptibility—A new look at an old model. *Cancer Cell* **1**: 411–412.

Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N et al. (1999). Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst* **91**: 943–949.

Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. (2002). Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* **31**: 33–36.

Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A et al. (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**: 165–167.

Ricol D, Cappellen D, El Marjou A, Gil-Diez-de-Medina S, Girault JM, Yoshida T et al. (1999). Tumour suppressive properties of fibroblast growth factor receptor 2-IIIb in human bladder cancer. *Oncogene* **18**: 7234–7243.

Rosa-Rosa JM, Pita G, Urioste M, Llort G, Brunet J, Lazaro C et al. (2009). Genome-wide linkage scan reveals three putative breast-cancer-susceptibility loci. *Am J Hum Genet* **84**: 115–122.

Ruiz-Narvaez EA, Rosenberg L, Rotimi CN, Cupples LA, Boggs DA, Adeyemo A et al. (2010). Genetic variants on chromosome 5p12 are associated with risk of breast cancer in African American women: the Black Women's Health Study. *Breast Cancer Res Treat* **123**: 525–530.

Sidransky D, Tokino T, Helzlsouer K, Zehnbauer B, Rausch G, Shelton B et al. (1992). Inherited p53 gene mutations in breast cancer. *Cancer Res* **52**: 2984–2986.

Smid M, Wang Y, Klijn JG, Sieuwerts AM, Zhang Y, Atkins D et al. (2006). Genes associated with breast cancer metastatic to bone. *J Clin Oncol* **24**: 2261–2267.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* **98**: 10869–10874.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* **100**: 8418–8423.

Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA et al. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**: 865–869.

Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF et al. (2008). Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **40**: 703–706.

Tamaru N, Hishikawa Y, Ejima K, Nagasue N, Inoue S, Muramatsu M et al. (2004). Estrogen receptor-associated expression of keratinocyte growth factor and its possible role in the inhibition of apoptosis in human breast cancer. *Lab Invest* **84**: 1460–1471.

Tannheimer SL, Rehemtulla A, Ethier SP. (2000). Characterization of fibroblast growth factor receptor 2 overexpression in the qhuman breast cancer cell line SUM-52PE. *Breast Cancer Res* **2**: 311–320.

Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A et al. (2005). Cancer risks and mortality in heterozygous ATM mutation carriers. *J Natl Cancer Inst* **97**: 813–822.

Thompson D, Easton D. (2004). The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia* **9**: 221–236.

Wang X, Pankratz VS, Fredericksen Z, Tarrell R, Karaus M, McGuffog L et al. (2010). Common variants associated with breast cancer in genome-wide association studies are modifiers of breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Hum Mol Genet* **19**: 2886–2897.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J et al. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**: 789–792.

Zhang Y, Gorry MC, Post JC, Ehrlich GD. (1999). Genomic organization of the human fibroblast growth factor receptor 2 (FGFR2) gene and comparative analysis of the human FGFR gene family. *Gene* **230**: 69–79.