

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



GWAS in Breast Cancer

Paulo C.M. Lyra-Junior, Nayara G. Tessarollo,
Isabella S. Guimarães, Taciane B. Henriques,
Diandra Z. dos Santos, Marcele L.M. de Souza,
Victor Hugo M. Marques, Laura F.R.L. de Oliveira,
Krislayne V. Siqueira, Ian V. Silva,
Leticia B.A. Rangel and Alan T. Branco

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/67223>

Abstract

Breast cancer is the most diagnosed cancer in women, and the second cause of cancer-related deaths among women worldwide. It is expected that more than 240,000 new cases and 40,450 deaths related to the disease will occur in 2016. It is well known that inherited genetic variants are drivers for breast cancer development. There are many mechanisms through which germline genetic variation affects prognosis, such as BRCA1 and BRCA2 genes, which account for approximately 20% of the increased hereditary risks. Therefore, it is evident that the genetic pathways that underlie cancer development are complex in which networks of multiple alleles confer disease susceptibility and risks. Global analyses through genome-wide association studies (GWAS) have revealed several *loci* across the genome are associated with the breast cancer. This chapter compiles all breast GWAS released since 2007, year of the first article published in this area, and discuss the future directions of this field. Currently, hundreds of genetic markers are linked to breast cancer, and understanding the underlying mechanisms of these variants might lead to the discover of biomarkers and targets for therapy in patients.

Keywords: breast cancer, genome-wide association studies (GWAS), susceptibility, *Loci*, SNPs

1. Introduction

One of the main goals of human genetics is to understand genetic pathways underlying traits. It has been highly successful the gene mapping of disorders with a Mendelian pattern of inheritance using the tendency of genes and other genetic markers to be inherited together. It is well known that genetic variants underlying these single-gene Mendelian disorders are rare in the population and tend to be highly penetrant, which means that a high percentage of carriers of the genotype will manifest the phenotype. On the other hand, mapping of non-Mendelian (or complex) traits, cases in which variants in multiple genes contribute to the phenotype, was only possible after sequencing and study of the human genome. Inherited variants underlying complex diseases, opposing the Mendelian disorders, have modest penetrance but higher frequency in the population [1–4] (**Figure 1**). Thus, efforts have been made to identify genes and pathways that control human traits, and, in the future, predict illness and establish more appropriated methods of treatment.

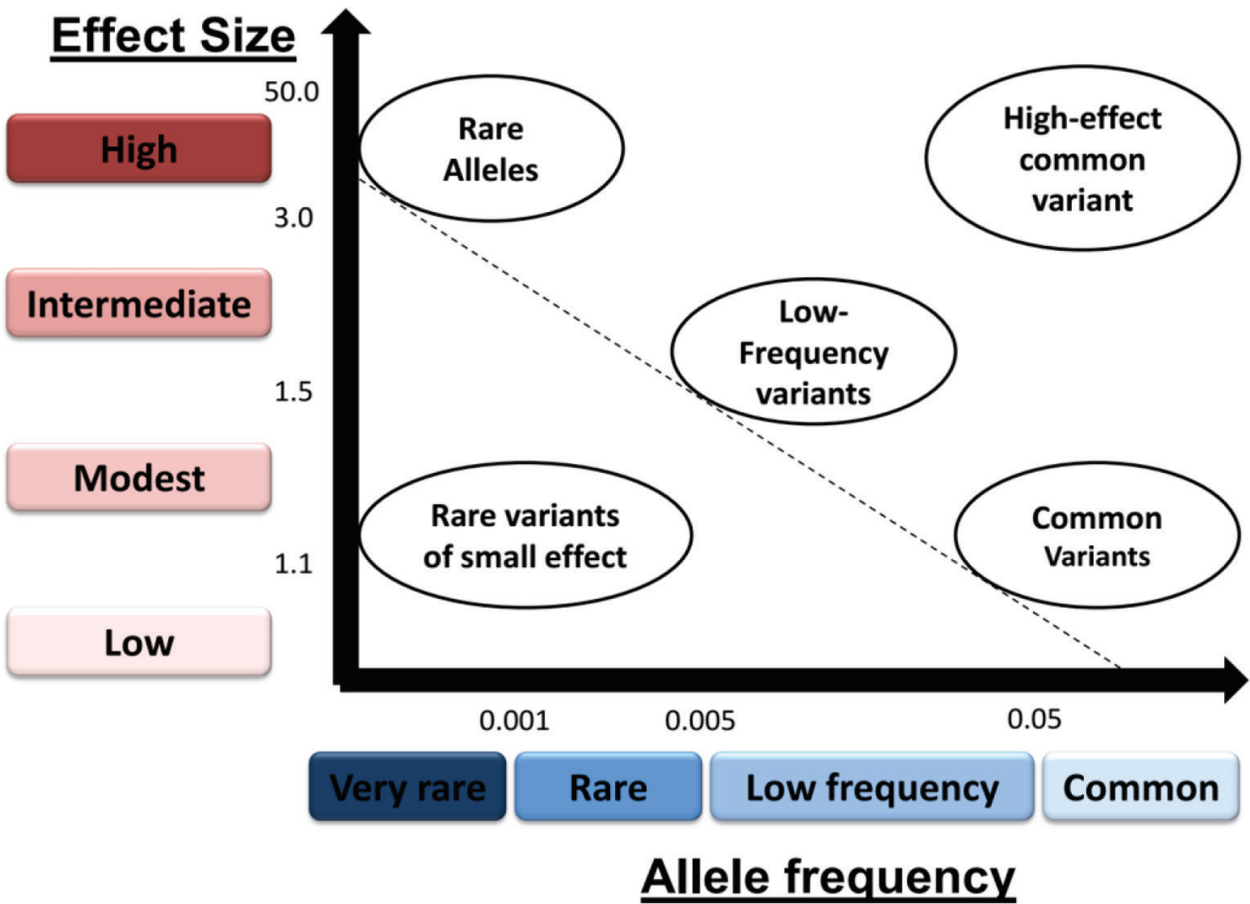


Figure 1. Features of genetic variants and correlation with disease severity. The panel shows the correlation between the frequency of alleles and the severity of the disease (odds ratio). Accordingly, Mendelian diseases (top left circle) have high effect on the individual, but the frequency of such mutations in the population is very rare. On the other hand, very rare variants with small effect (bottom left circle) are also found in the population, these features restrain the establishment of a reliable correlation between phenotype and genotype. GWAS have focused on identifying a massive number of genetic variants, which can be separated as (i) common variants associated with high effect size (top right circle) and (ii) abundant common alleles with apparent very low impact on human health (bottom right circle). Adapted from Manolio et al. [95].

A reflection of the urgency to unveil this research field is notable when looking through breast cancer numbers. Worldwide, the scenario is dramatic, with more than one million new cases of breast cancer diagnosed yearly (cancer genome atlas network 2012), and the fifth cause of death from cancer overall. In developing countries, breast cancer is the second cause of death from cancer and accounts for 15.4% of overall cancer-related deaths in women [5]. Moreover, it corresponds to the most common cancer-related death in women in the less developed regions (14.3%). In the United States, breast cancer is the second cause of cancer-related deaths among women, and it is estimated that one of eight American women will develop invasive breast cancer over the course of her lifetime. Accordingly, in the year of 2016, only in the United States, more than 240,000 new cases of the disease and 40,450 related deaths are expected [6].

Breast cancer comprises multiple diseases harbouring different genetic alterations; each subtype responds differently to treatments, and this feature leads to distinct clinical outcomes [7, 8]. Based on tumour histological biomarkers, breast cancer can be separated into three basic clinical types, such as HR positive (estrogen receptor and progesterone receptor), HER2+ (human epidermal growth factor receptor 2 positive), and triple-negative breast cancer, which are an essential part of the diagnostic workup of all breast cancer patients [9]. Approximately, 85% of all breast cancers are HR positive, about 20% are HER2+ and nearly 15% are triple-negative.

It is well understood that breast cancer is a complex and heterogeneous disease with a multifactorial etiology involving genetic, dietary, hormonal and reproductive factors. Among these, genetic is of particular importance. Epidemiological studies estimate that women with history of breast cancer in a first-degree relative show nearly twofold higher risk to develop breast cancer than women without a family history, indicating that the genetic factors are important determinants of disease risk [10]. At least 10–15% of all breast cancer cases may be due to the inheritance of a single gene mutation or multiple genetic variants [10, 11]. In the 1990s, two major susceptibility genes for breast cancer, breast cancer 1 (BRCA1) and breast cancer 2 (BRCA2), were the first ones to be identified on the long arm of chromosome 17 and the short arm of chromosome 13, respectively [12–14]. These genes are responsible for 20–30% of hereditary breast cancer cases worldwide. BRCA1 and BRCA2 are important on the maintenance of genome stability by playing a critical role in the regulation of different cellular processes, such as transcription, cell cycle, DNA repair, cell proliferation and differentiation, in response to DNA damage [15]. Indeed, woman carrying such pathogenic variants have an increased risk of 60–80% of breast cancer [16, 17]. Moreover, inherited BRCA1/2 gene mutations are associated with a 39–80% lifetime risk of female breast cancer [18–21]. It is also well established that BRCA1/2 carriers with breast cancer have a strong lifetime risk of developing contralateral breast cancer range from 10 to 40% and are 2–6 times higher than the risk for non-carriers [22–27].

The identification of mutations in BRCA, considered as a critical factor for the development of breast cancer in some women, has boosted the interest of scientists to discover more mutations that drive tumour development. In this context, advances in DNA sequencing technologies empowered massive parallel sequencing, and, as a consequence, it has led to a fantastic

discovery and assignment of other hereditary pre-disposition genes to high (TP53, PALB2, PTEN), moderated (CHEK2, ATM, NF1, NBN) and elevated, but imprecise, breast cancer risk (CDH1, STK11) [28–34]. Altogether, high and moderate penetrance breast cancer susceptibility mutations in these genes account for just over 30% of familial breast cancer cases, because linkage studies are not amenable to the identification of common alleles with small effects.

However, the major advance over the several years has led by genome-wide association studies (GWAS). This approach is based on genome-wide genotyping for thousands to millions of single-nucleotide polymorphisms (SNPs) in a large number of individuals and contrast between the groups with and without a specific phenotype. Therefore, this approach has successfully identified thousands of *loci* associated with hundreds of traits (National Human Genome Research Institute GWAS catalogue) [35]. Hence, this recent technology opens a new way to understand the underlying genetic causes of common diseases [1].

2. Genome-wide association studies

In the past, studying polymorphisms were limited by the technologies that only permitted analysis of one or a few *loci* at time, hence limiting the aims to particular genes or pathways. The selection for candidate genes or pathways to be studied were based on the potential relation with carcinogenesis, metabolism, cell cycle control and hormone synthesis. Therefore, the initial studies focused on single nucleotide polymorphisms which were associated with a crucial role in the cell functionality. With the advancement of techniques, sets of tagged SNPs included ‘known common variants’ across a gene. However, even though the number of candidates being analysed have increased, the number of well-validated association in those studies did not increase as expected. Association studies, involving direct testing of genetic polymorphisms in large series of cases versus controls, provide a powerful approach to identify lower penetrance alleles that cannot be detected by genetic linkage studies [36, 37]. However, additional susceptibility genes in which rare coding variants are associated with a moderate cancer risk have emerged through candidate gene re-sequencing [38].

GWAS have emerged as a powerful new approach that has the capacity of analysing the whole human genome in order to identify common variations in the population possibly associated with genetic factors of a specific disease. In other words, the intent of GWAS is to predict who is at the risk and develop new strategies for prevention and treatments of genetic diseases [39]. One of the initial successes of GWAS was the identification of the *complement factor H* gene as a major risk factor for age-related macular degeneration [40–42].

The GWAS technology is based on genotyping platforms (chip-based microarray technology) that can evaluate hundreds to thousands of SNPs simultaneously. The two primary platforms that have been used for most GWAS were developed by Illumina (San Diego, CA) and Affymetrix (Santa Clara, CA). These two competing technologies use different approaches to detect SNP variation. Accordingly, the Affymetrix platform prints short DNA sequences on a chip that recognizes a specific SNP allele. Alleles (i.e. nucleotides) are detected by a differential

DNA hybridization between the samples. Illumina, on the other hand, uses a bead-based technology with slightly longer DNA sequences to detect alleles. The Illumina technology is more expensive, but provides better specificity. Hence, it is possible to conduct association studies using sets of SNPs that tag most known common variants in the genome, and therefore, scan for the associations without prior knowledge of function or position [39, 43].

GWAS arrays have identified SNPs that are associated with many complex diseases or traits [44]; although they do not contain all mapped SNPs, rather they contain only index SNPs that represent SNPs in the same linkage disequilibrium (LD) block. The SNPs identified by GWAS are significantly correlated with a disease (or case) and are called as risk-associated SNPs, and the genomic regions containing the SNPs are called as risk *loci* for that particular disease [45, 46]. One common trend of the SNPs associated with the trait is that they are not frequently found in coding regions of the genome. Instead, most of them are located in non-coding regions of the genome and are equally distributed between intronic and intergenic compartments [47, 48]. This might initially reduce the potential of the index, SNPs being the causal, but it is important to keep in mind that all the SNPs in the same haplotype block with the index SNP could possibly play the role of a causal SNP. A commonly used approach to investigate SNPs other than the index SNPs present on the standard GWAS array has been to use an LD calculation [49–51] together with the 1000 Genomes Project reference panels from different populations [52, 53].

To move from the index SNP to a more refined list of putative causal SNPs located within the identified region, another approach called fine-mapping has also been used. Fine-mapping studies employ dense genotyping arrays that contain all common SNPs within the previously identified risk *loci*, which together with imputation [49–51] allow investigators to perform a more complete analysis of the risk regions. The most fine-mapping analyses have been done by international consortia with the shared interests for specific diseases or traits; examples include: the immunochip [54], the metabochip [55], the iCOGs array [56] and the Oncoarray [57].

3. GWAS in breast cancer

Over the past years, the results from GWAS have been published for breast cancer reporting well-validated novel associations. In total, these scans have identified approximately 100 common genetic susceptibility *loci* for breast cancer risk, and as additional scans are ongoing at some point, the number of cancer susceptibility *loci* is likely to change rapidly over the next years. This is only possible because there are many worldwide consortium groups, for example Asia Breast Cancer Consortium (ABCC) and Breast Cancer Association Consortium (BCAC), and it is through them that has been possible identifying the susceptibility of SNPs in large-scale and different populations.

The first GWAS for breast cancer was published in 2007 and identified novel susceptibility *loci* associated with this illness. Accordingly, as in reference [58], they studied 4398 breast cancer cases and 4316 controls, followed by a third stage in which 30 SNPs were tested for confirmation in 21860 cases and 22578 controls from 22 studies. In total 227876 SNPs were analyzed,

which represented a coverage of approximately 77% of known common SNPs in Europeans at $r^2 > 0.5$. As a result, they found five novel independent *loci* associated with the breast cancer ($P < 10^{-7}$ using a stratified Cochran-Armitage trend test). The genes found around four *loci* are plausible causative genes (FGFR2, TNRC9, MP3K1 and LSP1). The most strongly associated SNP was in the intron 2 of the FGFR2 gene, a receptor tyrosine kinase that is amplified and overexpressed in 5–10% of breast tumours [59]. The 16q *locus* contains the candidate genes TNRC9 and LOC643714. The function of TNRC9 genes is currently unknown; however, the presence of the HMG box motif suggests that it possibly acts as a transcription factor [60]. MAP3K1, located at the 5q *locus*, is a gene involved in signal transduction and has not been previously reported to be involved with cancer. LSP1 is located at 11p *locus* and is an F-actin bundling cytoskeletal protein expressed in hematopoietic and endothelial cells. Other evidence of association pointed to a SNP around the H19 gene, a maternally imprinted gene that encodes an untranslated mRNA closely involved in regulation of IGF2. The fifth *locus* is an interval of 110 kb lacking known genes and located in the genomic region 8q24. Despite the absence of genes in the segment of 110 kb, the region 8q24 contains *loci* associated with prostate and colorectal cancers. The second stage of this study identified 1792 SNPs with P -value < 0.05 , while the estimated by chance would be 1343. These observations have indicated that many additional common susceptibility alleles might be identifiable by this approach, but the detection of further susceptibility *loci* is associated with the increased coverage and use of larger number of cases and controls [58].

In the following years, nine articles using GWAS to identify genetic factors linked with breast cancer were published [61–69]. These works have not only increased the number of new markers associated with the illness, but also validated the genetic factors that were previously identified. Furthermore, the cancer genetic markers of susceptibility (CGEMS) group detected the association of FGFR2 in a second genome scan, genotyping 528,173 SNPs in 1145 cases of invasive breast cancer among postmenopausal white women and 1142 controls they detected a set of four SNPs in intron 2 of FGFR2 [62]. All the variants are related with FGFR2 expression in normal breast tissue, and interesting two of them are likely related to biological mechanism for interrupting active transcription factor-binding sites [70]. The deCODE group later on, using approximately 1000 unselected breast cancer cases and illumina 317k panel, found two additional *loci* at 2q and 5p [61, 63]. A further *locus* on 6q was identified by Gold et al. [64] studying 249 familial Ashkenazi Jewish breast cancer cases. This region contains two potential candidate genes, ECHDC1 and RNF146. The CGEMS group again added two novel *loci* with genome-wide significance: (i) one SNP, on the genomic region 1p11.2 neighbouring NOTCH2 and FCGR1B, is predominantly associated with estrogen receptor-positive breast cancer; (ii) the second SNP is located on chromosome 14q24.1, localizes to RAD51L1, a prior candidate pathway for breast cancer susceptibility [67]. Additional *loci* associated with the breast cancer were found with a more refined analysis of the first GWAS. Accordingly, Ahmed et al. [66] tested over 800 promising associations in the two stages involving 37,012 cases and 40,069 controls from 33 studies in the CGEMS and BCAC, finding strong evidence for additional susceptibility *loci* on 3p24 and 17q23.2; the causative genes include SLC4A7 and NEK10 on 3p and COX11 on 17q. Finally, Zheng et al. [65] conducted a GWAS among Chinese women and studied 607,728 SNPs in 1505 cases and 1522 controls; this analysis revealed 29

promising SNPs. The SNP at 6q25.1, located upstream of the estrogen receptor 1 gene (ESR1), exhibited consistent association with breast cancer across all the three stages performed, providing strong evidence of a susceptibility *locus* for breast cancer.

In 2010, a group conducted a new GWAS in which 582,886 SNPs were genotyped in 3659 cases with a family history of the disease and 4897 controls. They identified five new susceptibility *loci* on the chromosomes 9, 10 and 11, and found three SNPs in the 6q25.1, 8q24 and 11p15 regions with a higher correlation risk to develop cancer than the ones reported previously [69]. In the same year, other group, based on the fact that the germline BRCA1 mutations predispose to breast cancer, aimed to identify genetic modifiers of this risk in 1193 individuals with BRCA1 mutations who were diagnosed with invasive breast cancer under the age 40. This group was contrasted with 1190 BRCA1 carriers without breast cancer diagnosis over age 35. The first stage of this study had led to the identification of 96 SNPs; after the further stages of analysis, five SNPs on 19p13 were highly associated with breast cancer risk and also associated with triple-negative breast cancer in a separate study of 2301 triple-negative cases and 3949 controls [68].

Three studies were published in 2011 revealing new *loci* associated with the breast cancer. Haiman et al. [71] searching for common risk alleles for ER-negative breast cancer, combined GWAS data from women of African ancestry (1004 ER-negative cases and 2745 controls) and European ancestry (1718 ER-negative cases and 3670 controls). This study was further replicated with an additional 2292 ER-negative cases and 16,901 controls of European ancestry. Their conclusion pinpointed a common risk variant for ER-negative breast cancer at the TERT-CLPT1L *locus* on chromosome 5p15 in multiple populations. Furthermore, the same variant was also significantly associated with the triple-negative breast cancer, particularly in younger women (<50 years old). Cai et al. [72] published a four-stage GWAS including 17,153 cases and 16,943 controls among East-Asian women, after analysing 684,457 SNPs. The final result revealed one SNP at 10q21.2 (ZNF365) strongly implicated as a genetic risk variant for breast cancer among East-Asian women. Fletcher et al. [73] compared 296,114 tagging SNPs in 1694 cases of breast cancer and 2365 controls, with validation in three independent series totalling 11,880 cases and 12,487 controls, identifying a novel *locus* risk for breast cancer at 9q31.2 (the nearest genes around the SNP found are KLF4, RAD23B and ACTL7A), as well two variants mapping to 6q25.1, a *locus* previously reported. Although approximately 25 common genetic susceptibility *loci* have been identified to be independently associated with breast cancer risk, the genetic risk variants reported only explain a small fraction of the heritability of breast cancer.

Long et al. [74] aimed to discover novel genetic susceptibility *loci* for breast cancer, therefore they conducted a four-stage GWAS in 19,091 cases and 20,606 controls of East-Asian descent (Chinese, Korean and Japanese women were included). It was analysed 690,947 SNPs, from this group the final stage showed an SNP in chromosome 6q25.1, near to TGF- β activated kinase (TAB2), with consistent association with breast cancer risk across all four stages, reaching a *P*-value of 3.8×10^{-12} when the analysis was done with all samples combined. In addition, they identified two possible susceptibility SNPs, one located in the intron 5 of the ESR1 gene and the other at 11q24.3, with consistent association in each of the four

stages. Kim et al. [75] conducted a GWAS to evaluate previously identified *loci* in Korean women and to identify additional novel breast cancer susceptibility variants. Accordingly, they conducted a three-stage GWAS that included 6322 cases and 5897 controls. The results revealed one SNP in the epidermal growth factor receptor (ERB4) gene, located at chromosome 2q34, and showed that seven breast cancer susceptibility *loci* that were previously identified in European and/or Chinese population could be directly replicated in Korean women. Another GWAS study was conducted in Japanese patients with hormone receptor-positive, invasive breast cancer receiving adjuvant tamoxifen therapy. This study detected significant associations with recurrence-free survival of 15 SNPs on nine chromosomal *loci* 1p31, 1q41, 5q33, 7p11, 10q22, 12q13, 13q22, 18q12 and 19p13. Among them, the one in the C10orf11 gene in 10q22 was significantly associated with recurrence-free survival in breast cancer patients treated with tamoxifen [76]. Besides these articles, two more were published in the same year. Ghoussaini et al. [77] reported a follow up of 72 promising associations from two independent GWAS using approximately 70,000 cases and 68,000 controls from 41 case-control studies and nine breast cancer GWAS. Through this study, three new breast cancer risk *loci* on 12p11 (PTHLH gene), 12q24 and 21q21 (NRIP1 gene) were identified. An interesting fact was that two SNPs were associated only with ER-positive disease, whereas the SNP on 12p11 was associated with similar relative risks for both ER-negative and ER-positive breast cancer. Because the GWAS of breast cancer separated by immunohistochemical have revealed *loci* contributing to the susceptibility of ER-negative subtypes. Siddiq et al. [78] conducted a large meta-analysis of ER-negative disease, comprising 4754 ER-negative cases and 31,663 controls from three GWAS, to identify additional genetic variants for ER-negative breast cancer. They performed an *in silico* replication with 86 SNPs using a P -value $\leq 10^{-5}$ in an additional population of 11,209 cases of breast cancer, where 946 were with ER-negative disease, and 16,057 controls of Japanese, Latino and European ancestry. As result two novel *loci* were identified, one at 6q14 and other at 20q11. At the *locus* 6q14 the SNP was associated with breast cancer and both ER-positive and ER-negative disease. In contrast, the SNP at 20q11 was associated with ER-negative breast cancer, but showed weaker association with overall breast cancer and no association with ER-positive disease. This work also confirmed three known *loci* associated with both ER-negative and ER-positive breast cancer. These findings highlight the relevance of large-scale collaborative studies to identify novel breast cancer risk *loci*.

In order to obtain a more comprehensive knowledge on the genetic factors controlling breast cancer development, the project collaborative oncological gene-environment study (COGS) was created through collaboration among four consortia [56]. The project consisted of a meta-analysis of nine GWAS, involving 10,052 breast cancer cases and 12,575 controls of European ancestry. 29,807 SNPs were selected for further genotyping. The selected SNPs were genotyped in 41 studies in BCAC, using 45,290 cases and 41,880 controls in European ancestry population. Another important point of the study was the custom Illumina iSelect genotyping array (iCOGS) utilized that comprises more than 200,000 SNPs. The combined efforts identified SNPs at 41 new breast cancer susceptibility *loci* at genome-wide significance ($P < 5 \times 10^{-8}$). Two other studies were published in 2013. One aimed to identify further cancer risk-modifying *loci* using multi-stage GWAS of 11,705 BRCA1 carriers (5920 diagnosed with breast cancer and 1839 diagnosed with ovarian cancer); further replication was done with an additional sample

of 2646 BRCA1 carriers. Looking specifically at breast cancer factors they identified a novel risk modifier *locus* at 1q32 for BRCA1 carriers [79]. The other study was focused on identification of susceptibility *loci* specific to ER-negative disease, using a meta-analysis of 3 GWAS with 4193 ER-negative breast cancer cases and 35,194 controls with a series of 40 follow-up studies and also used the iCOGS to genotype. Their conclusion reported SNPs at four *loci*, 1q32.1 (MDM4 and LGR6), 2p24.1 and 16q12.2 associated with ER-negative but not ER-positive breast cancer. Once again providing further evidence for distinct etiological pathways associated with invasive ER-positive and ER-negative breast cancer [80].

GWAS have also been proven to be a powerful strategy to identify genetic factors associated with adverse reactions caused by drugs. The first GWAS for chemotherapy-induced alopecia was conducted in Japanese breast cancer patients, and identified SNPs significantly associated with drug-induced grade 2 alopecia. For instance, the rs3820706 (calcium channel voltage-dependent subunit beta) on 2q23 and its nearby SNP rs16830728 could be associated with significant molecular alterations in genes such as ion channel-related genes and genes related to the β -catenin signalling pathway [81].

The lack of concordance among some studies for breast cancer led a group to study 41 common non-synonymous SNP (nsSNP) for which evidence of association with breast cancer risk had been previously reported. This work combined 38 studies of white European women (46,450 cases and 42,600 controls), and showed strong association for one previously reported, 7q21; one novel susceptibility *locus*, 3p21 and the third *locus* is located in an established breast cancer susceptibility region, 3p24 [82]. Another study with 22,780 cases and 24,181 controls provided additional insights into the genetics and biology of breast cancer in East Asian women. It was identified that three genetic *loci* located at 1q32.1, 5q14.3 and 15q26.1 were recently associated with breast cancer risk [83]. Purrington et al. [84], interested on identify *loci* that influence triple-negative breast cancer risk, conducted a two-stage GWAS of triple-negative breast cancer with 1529 cases and 3399 controls in the first stage and 2148 cases and 1309 controls in the second. Variants at 19p13.1 and PTHLH *loci* showed significant association in both stages. Moreover, 25 SNPs already known as breast cancer susceptibility were associated with risk of triple-negative breast cancer ($P < 0.05$).

One particular article published in 2014 called attention for running a meta-analysis of GWAS of three mammographic density phenotype: dense area, non-dense and percent density in up to 7916 women in stage 1 and 10,379 women in the second stage. The results showed *loci* that reached genome-wide significance for all three phenotypes, for dense area (AREG, ESR1, ZNF365, LSP1, IGF1, TMEM184B and SGSM3), non-dense area (8p11.23) and percent density (PRDM6, 8p11.23 and TMEM184B). Interestingly, some regions are known as breast cancer susceptibility *loci* and the others regions were found, after a large meta-analysis, to be associated with breast cancer ($P < 0.05$). Based on the ability to identify known as well as putative novel breast cancer *loci* by studying mammographic density phenotypes, the authors demonstrated the power of using quantitative intermediate phenotypes to discover new disease *loci* [85].

In 2015, there were more than 90 established breast cancer risk *loci*, with 57 new ones, revealed through GWAS during 2013 and 2014. Nevertheless, new studies were published identifying new

susceptibility *loci*. A group performed a meta-analysis restricted to women of European ancestry. They worked with 11 GWAS comprising of 15,748 breast cancer cases and 18,084 controls, and 46,785 cases and 42,892 controls from 41 studies genotyped on iCOGS, and used imputation to estimate genotypes for more than 11 million SNPs, identifying 15 novel *loci* associated with breast cancer at $P < 5 \times 10^{-8}$ [85]. Palomba et al. [86], also following the assumption that analyses in genetically-homogeneous population could represent an additional approach to detect low penetrance alleles, conducted a GWAS study comparing 1431 Sardinian patients with non-familial, BRCA1/2-mutation-negative breast cancer to 2171 healthy Sardinian blood donors, where 2,067,645 SNPs were analysed. The study concludes the role of TOX3 and FGRF2 as breast cancer susceptibility genes in BRCA1/2-wild-type breast cancer patients from Sardinian population.

In 2016, three GWAS were published describing novel genetic susceptibility *loci*. It was a study including 14,224 cases and 14,829 controls of East Asian women, where two SNPs in two *loci* were found to be associated with breast cancer risk at the genome-wide significance level, one at 1p22.3 and other at 21q22.12 [87]. The identification of four previously unidentified *loci* including the ones at 13q22 (KLF5), 2p23.2 (WDR43) and 2q33 (PPIL3) with genome-wide significant association with ER-negative breast cancer, performing a meta-analysis of 11 GWAS consisting of 4939 ER-negative cases and 14,352 controls, combined with 7333 ER-negative cases and 42,468 controls and 15,252 BRCA1 mutation carriers genotyped on the iCOGS array [88]. GWAS also can be useful to identify SNPs associated with response to anthracycline-based neoadjuvant chemotherapy in breast cancer patients. A group identified two SNPs that were significantly associated with pathologic complete response after neoadjuvant chemotherapy. After the validation using 401 patients who received anthracycline-based neoadjuvant regimens the authors found that only one SNP, located in the WT1 gene, was associated with the pathologic complete response after anthracycline-based neoadjuvant therapy, suggesting that WT1 may be a potential target of anthracycline-based neoadjuvant therapy for breast cancer [89].

4. Conclusion

GWAS have been successful in identifying many genetic variants that are significantly associated with human diseases. However, a gap has emerged between the ability to detect these associations and the ability to meaningfully interpret their biological significance [90]. Currently, the challenges facing GWAS include the translation of associated *loci* into suitable biological hypotheses, missing heritability [91], and the understanding of how multiple modestly associated *loci* within genes interact to influence a phenotype [92]. Thus, the new trend for susceptibility *loci* identification has moved forward to describe precisely the functional effects and target genes. The post-GWAS include detailed genetic and epidemiological dissection, bioinformatics prediction of functionality and *in vitro* and *in vivo* experimental verification of the molecular mechanisms for the causal variants and their target genes [93, 94]. Although identification of common risk variants is an emerging field, it will create a routine screening method for earlier diagnosis and direct breast cancer treatment strategies.

Abbreviations

GWAS	Genome-wide association studies
BRCA 1	Breast cancer 1 gene
BRCA 2	Breast cancer 2 gene
SNP	Single-nucleotide polymorphisms
HER 2+	Human epidermal growth factor receptor 2 positive
HR	Progesterone receptor
LOD	Logarithm of odds
LD	Linkage disequilibrium
TP53	Tumour protein p53
PALB2	Partner and localizer of BRCA2
PTEN	Phosphatase and tensin homolog
CHEK2	Checkpoint kinase 2
ATM	Serine/threonine kinase
NF1	Nuclear factor 1
NBN	Nibrin
CDH1	Cadherin-1
STK11	Serine/threonine kinase 11
FGFR2	Fibroblast growth factor receptor 2
TNRC9	Trinucleotide-repeat-containing 9
LSP1	Lymphocyte-specific protein 1
IGF2	Insulin-like growth factor 2
CGEMS	Cancer genetic markers of susceptibility
RNF146	RING finger protein 146
RAD51L1	DNA repair protein RAD51 homolog 2
NOTCH2	Neurogenic <i>locus</i> notch homolog protein 2
FCGR1B	Cluster of differentiation 64
BCAC	Breast Cancer Association Consortium
SLC4A7	Solute carrier family 4, sodium bicarbonate cotransporter, member 7
NEK10	NIMA-related kinase 10
COX11	Cytochrome C oxidase copper chaperone

ER	Estrogen receptor
KLF4	Kruppel-like factor 4
RAD23B	UV excision repair protein RAD23 homolog B
ACTL7A	Actin-like protein 7A
ESR1	Estrogen receptor 1
ERB4	Epidermal growth factor receptor
PTH1H	Parathyroid hormone-related protein
NR1P1	Nuclear receptor-interacting protein 1
COGS	Collaborative oncological gene-environment study
LGR6	Leucine-rich repeat-containing G-protein coupled receptor 6
AREG	Amphiregulin
ZNF365	Zinc finger protein 365
IGF1	Insulin-like growth factor 1
TMEM184B	Transmembrane protein 184B
SGSM3	Small G protein signaling modulator 3
KLF5	Kruppel-like factor 5
WDR43	WD repeat domain 43
PP1L3	Peptidyl-prolyl cis-trans isomerase-like 3
WT1	Wilms tumour protein

Author details

Paulo C.M. Lyra-Junior¹, Nayara G. Tessarollo, Isabella S. Guimarães¹, Taciane B. Henriques¹, Diandra Z. dos Santos¹, Marcelle L.M. de Souza¹, Hugo M. Victor Marques², Laura F.R.L. de Oliveira², Krislayne V. Siqueira², Ian V. Silva³, Leticia B.A. Rangel^{1*} and Alan T. Branco²

*Address all correspondence to: lbarangel@yahoo.com

¹ Laboratory of Cellular and Molecular Biology of Human Cancer, Department of Pharmaceutical Sciences and Biotechnology Program, Federal University of Espírito Santo State, Vitória, Brazil

² Laboratory of Cellular and Molecular Biology of Human Cancer, Department of Pharmaceutical Sciences, Vitória, Brazil

³ Department of Morphology, Health Sciences Center, Federal University of Espírito Santo State, Vitória, Brazil

References

- [1] Monteiro ANA, Freedman ML. Lessons from postgenome-wide association studies: functional analysis of cancer predisposition loci. *Journal International Medicine*. 2013; **274**: 414–424. DOI: 10.1111/joim.12085
- [2] Venter JC, et al. The sequence of the human genome. *Science*. 2001; **291**: 1304–1351. DOI:10.1126/science.1058040
- [3] Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; **409**: 860–921. DOI: 10.1038/35057062
- [4] Altshuler D, et al. Haplotype map of the human genome. *Nature*. 2005; **437**: 1299–1320. DOI: 10.1038/nature04226
- [5] Ferlay J, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal Cancer*. 2015; **136**: 359–386. DOI: 10.1002/ijc.29210. Epub 2014 Oct 9.
- [6] American Cancer Society. *Cancer Facts & Figures 2016*. Atlanta: American Cancer Society; 2016.
- [7] Perou CM, et al. Molecular portraits of human breast tumours. *Nature*. 2000; **406**: 747–752. DOI: 10.1038/35021093
- [8] Sorlie T, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; **98**: 10869–10874. DOI: 10.1073/pnas.191367098
- [9] Rakha EA, Chan S. Metastatic triple-negative breast cancer. *Clinical Oncology*. 2011; **23**: 587–600. DOI: 10.1016/j.clon.2011.03.013
- [10] Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *The Lancet*. 2001; **358**: 1389–1399. DOI: 10.1016/S0140-6736(01)06524-2
- [11] Pharoah PD, et al. Polygenic susceptibility to breast cancer and implications for prevention. *Nature Genetics*. 2002; **31**: 33–36. DOI: 10.1038/ng853
- [12] Miki Y, et al. A strong candidate for the breast and ovarian-cancer susceptibility gene BRCA1. *Science*. 1994; **266**: 66–71. DOI: 10.1126/science.7545954
- [13] Friedman LS, et al. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nature Genetics*. 1994; **8**: 399–404. DOI: 10.1038/ng1294-399.
- [14] Wooster R, et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature*. 1995; **378**: 789–792. DOI: 10.1038/378789a0

- [15] Gudmundsdottir K, Ashworth A. The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. *Oncogene*. 2006; **25**: 5864–5874. DOI: 10.1038/sj.onc.1209874
- [16] Lakhani SR, et al. Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. *Clinical Cancer Research*. 2005; **11**: 5175–5180. DOI: 10.1158/1078-0432
- [17] Foulkes WD, Shuen AY. In brief: BRCA1 and BRCA2. *The Journal of pathology*. 2013; **230**: 347–349. DOI: 10.1002/path.4205
- [18] King M.C, Marks J.H, Mandell J.B. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*. 2003; **302**: 643–646. DOI: 10.1126/science.1088759
- [19] Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *Journal of Clinical Oncology*. 2007; **25**: 1329–1333. DOI: 10.1200/JCO.2006.09.1066
- [20] Mavaddat N, et al. Cancer risks for BRCA1 and BRCA2 mutation carriers: Results from prospective analysis of EMBRACE. *Journal of the National Cancer Institute*. 2013; **105**: 812–822. DOI: 10.1093/jnci/djt095
- [21] Gabai-Kapara E, et al. Population-based screening for breast and ovarian cancer risk due to BRCA1 and BRCA2. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; **111**: 14205–14210. DOI: 10.1073/pnas.1415979111
- [22] Brekelmans CT, et al. Survival and prognostic factors in BRCA1-associated breast cancer. *Annals of Oncology*. 2006; **17**: 391–400. DOI: 10.1093/annonc/mdj095
- [23] Graeser MK, et al. Contralateral breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Journal of Clinical Oncology*. 2009; **27**: 5887–5892. DOI:10.1200/JCO.2008.19.9430
- [24] Van Der Kolk DM, et al. Penetrance of breast cancer, ovarian cancer and contralateral breast cancer in BRCA1 and BRCA2 families: high cancer incidence at older age. *Breast Cancer Research and Treatment*. 2010; **124**: 643–651. DOI: 10.1007/s10549-010-0805-3
- [25] Bernstein JL, et al. Contralateral breast cancer after radiotherapy among BRCA1 and BRCA2 mutation carriers: A WECARE study report. *European Journal of Cancer*. 2013; **49**: 2979–2985. DOI: 10.1016/j.ejca.2013.04.028
- [26] Evans DG, et al. Contralateral mastectomy improves survival in women with BRCA1/2-associated breast cancer. *Breast Cancer Research and Treatment*. 2013; **140**: 135–142. DOI: 10.1007/s10549-013-2583-1
- [27] Van den Broek AJ, et al. Impact of age at primary breast cancer on contralateral breast cancer risk in BRCA1/2 mutation carriers. *Journal of Clinical Oncology*. 2015; **34**: 409–418. DOI: 10.1200/JCO.2015.62.3942
- [28] Giardiello FM, et al. Very high risk of cancer in familial Peutz-Jehgers syndrome. *Gastroenterology*. 2000; **119**: 1447–1453. DOI: 10.1053/gast.2000.20228

- [29] Antoniou A, 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. *The American Society of Human Genetics*. 2003; **72**: 1117–1130. DOI: 10.1086/375033
- [30] The CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: A collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from ten studies. *The American Society of Human Genetics*. 2004; **74**: 1175–1182. DOI: 10.1086/421251
- [31] Thompson D, et al. Cancer risks and mortality in heterozygous ATM mutation carriers. *Journal of the National Cancer Institute*. 2005; **97**: 813–822. DOI: 10.1093/jnci/dji141
- [32] Rahman N, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genetics*. 2006; **39**: 165–167. DOI: 10.1038/ng1959
- [33] Robson M. CHEK2, breast cancer, and the understanding of clinical utility. *Clinical Genetics*. 2010; **78**: 8–10. DOI: 10.1111/j.1399-0004.2010.01444.x
- [34] Meindl A, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nature Genetics*. 2010; **42**: 410–414. DOI: 10.1038/ng.569
- [35] Manolio TA. Genomewide association studies and assessment of the risk of disease. *The New England Journal of Medicine*. 2010; **363**: 166–176. DOI: 10.1056/NEJMra0905980
- [36] García-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. *The Lancet*. 2005; **366**: 649–59. DOI: 10.1016/S0140-6736(05)67137-1
- [37] Cox A, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nature Genetics*. 2007; **39**: 352–358. DOI: 10.1038/ng1981
- [38] Easton DF, Eeles RA. Genome-wide association studies in cancer. *Human Molecular Genetics*. 2008; **17**: R109–R115. DOI: 10.1093/hmg/ddn287
- [39] Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLoS Computation Biology*. 2012; **8**: e1002822. DOI:10.1371/journal.pcbi.1002822
- [40] Haines JL, et al. Complement factor H variant increases the risk of age-related macular degeneration. *Science*. 2005; **308**: 419–421. DOI: 10.1126/science.1110359
- [41] Edwards AO, et al. Complement factor H polymorphism and age-related macular degeneration. *Science*. 2005; **308**: 421–424. DOI: 10.1126/science.1110189
- [42] Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005; **308**: 385–389. DOI: 10.1126/science.1109557
- [43] The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; **447**: 661–678. DOI: 10.1038/nature05911

- [44] Welter D, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*. 2014; **42**: 1001–1006. DOI: 10.1093/nar/gkt1229
- [45] Li Y, et al. Genotype imputation. *Annual Review Genomics Human Genetics*. 2009; **10**: 387–406. DOI: 10.1146/annurev.genom.9.081307.164242
- [46] Kichaev G, Pasaniuc B. Leveraging functional-annotation data in transethnic fine-mapping studies. *American Journal Human Genetics*. 2015; **97**: 260–271. DOI: 10.1016/j.ajhg.2015.06.007
- [47] Freedman ML, Monteiro AN, Gayther SA et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genetics*. 2011; **43**: 513–8. DOI:10.1038/ng.840.
- [48] Blattler A, et al. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biology*. 2014; **15**: 469. DOI: 10.1186/s 13059-014-0469-0
- [49] Howie BN, Donnelly P, Marchini JA. Flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*. 2009; **5**: e1000529. DOI: 10.1371/journal.pgen.1000529
- [50] Li Y, et al. Mach: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*. 2010; **34**: 816–834. DOI: 10.1002/gepi.20533
- [51] Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*. 2007; **81**: 1084–1097. DOI: 10.1086/521987
- [52] Genomes Project C, et al. A global reference for human genetic variation. *Nature*. 2015; **526**: 68–74. DOI: 10.1038/nature15393
- [53] 1000 Genomes Project Consortium, et al. An integrated map of genetic variation from 1092 human genomes. *Nature*. 2012; **491**: 56–65. DOI: .1038/nature11632
- [54] Trynka G, Hunt KA, Bockett NA et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genetics*. 2011; **43**: 1193–201. DOI: 10.1038/ng.998
- [55] Voight BF, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genetics*. 2012; **8**: e1002793. DOI: 10.1371/journal.pgen.1002793
- [56] Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genetics*. 2013; **45**: 353–361. DOI: 10.1038/ng.2563
- [57] Available from: <http://epi.grants.cancer.gov/oncoarray/> [Accessed: 2016-07-14].
- [58] Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007; **447**: 1087–1093. DOI: 10.1038/nature05887.
- [59] Adnane J, et al. BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene*. 1991; **6**: 659–663.

- [60] Brantjes H, et al. All Tcf HMG box transcription factors interact with Groucho-related co-repressors. *Nucleic Acids Research*. 2001; **29**: 1410–1419. DOI: 10.1093/nar/29.7.1410
- [61] Stacey SN, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics*. 2007; **39**: 865–869. DOI: 10.1038/ng2064
- [62] Hunter DJ, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics*. 2007; **39**: 870–874. DOI: 10.1038/ng2075
- [63] Stacey SN, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics*. 2008; **40**: 703–706. DOI: 10.1038/ng.131
- [64] Gold B, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; **105**: 4340–4345. DOI: 10.1073/pnas.0800441105
- [65] Zheng W, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature Genetics*. 2009; **41**: 324–328. DOI: 10.1038/ng.318
- [66] Ahmed S, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nature Genetics*. 2009; **41**: 585–590. DOI: 10.1038/ng.354
- [67] Thomas G, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nature Genetics*. 2009; **41**: 579–584. DOI: 10.1038/ng.353
- [68] Antoniou AC, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nature Genetics*. 2010; **42**: 885–892. DOI: 10.1038/ng.669
- [69] Turnbull C, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature Genetics*. 2010; **42**: 504–507. DOI: 10.1038/ng.586
- [70] Meyer KB, et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biology*. 2008; **6**: e108. DOI: 10.1371/journal.pbio.0060108
- [71] Haiman CA, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nature Genetics*. 2011; **43**: 1210–1214. DOI: 10.1038/ng.985
- [72] Cai Q, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. *Human Molecular Genetics*. 2011; **20**: 4991–4999. DOI: 10.1093/hmg/ddr405
- [73] Fletcher O, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *Journal of the National Cancer Institute*. 2011; **103**: 425–435. DOI: 10.1093/jnci/djq563

- [74] Long J, et al. Genome-wide association study in East Asians identifies novel susceptibility loci for breast cancer. *PLoS Genetics*. 2012; **8**: e1002532. DOI: 10.1371/journal.pgen.1002532
- [75] Kim HC, et al. A genome-wide association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study. *Breast Cancer Research*. 2012; **14**: R56. DOI: 10.1186/bcr3158
- [76] Kiyotani K, et al. A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. *Human Molecular Genetics*. 2012; **21**: 1665–1672. DOI: 10.1093/hmg/ddr597
- [77] Ghoussaini M, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nature Genetics*. 2012; **44**: 312–318. DOI: 10.1038/ng.1049
- [78] Siddiq A, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Human Molecular Genetics*. 2012; **21**: 5373–5384. DOI: 10.1093/hmg/dds381
- [79] Couch FJ, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genetics*. 2013; **9**: e1003212. DOI: 10.1371/journal.pgen.1003212
- [80] Garcia-Closas M, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nature Genetics*. 2013; **45**: 392–398. DOI: 10.1038/ng.2561
- [81] Chung R, et al. Roles of Wnt/ β -catenin signalling pathway in the bony repair of injured growth plate cartilage in young rats. *Bone*. 2013; **52**: 651–658. DOI: 10.1016/j.bone.2012.10.035
- [82] Milne RL, et al. Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. *Human Molecular Genetics*. 2014; **23**: 6096–6111. DOI: 10.1093/hmg/ddu311
- [83] Cai Q, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nature Genetics*. 2014; **46**: 886–890. DOI: 10.1038/ng.3041
- [84] Purrington KS, et al. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis*. 2014; **35**: 1012–1019 DOI: 10.1093/carcin/bgt404
- [85] Michailidou K, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature Genetics*. 2015; **47**: 373–380. DOI: 10.1038/ng.3242
- [86] Palomba G, et al. Genome-wide association study of susceptibility loci for breast cancer in Sardinian population. *BMC cancer*. 2015; **15**: 383. DOI: 10.1186/s12885-015-1392-9

- [87] Han MR, et al. Genome-wide association study in East Asians identifies two novel breast cancer susceptibility loci. *Human Molecular Genetics*. 2016; **0**: 1–11. DOI: 10.1093/hmg/ddw164
- [88] Couch FJ, et al. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. *Nature Communication*. 2016; **7**: 11375. DOI: 10.1038/ncomms11375
- [89] Wu L, et al. A genome-wide association study identifies WT1 variant with better response to 5-fluorouracil, pirarubicin and cyclophosphamide neoadjuvant chemotherapy in breast cancer patients. *Oncotarget*. 2016; **7**: 5042–5052. DOI: 10.18632/oncotarget.5837
- [90] Qian DC, et al. Identification of shared and unique susceptibility pathways among cancers of the lung, breast, and prostate from genome-wide association studies and tissue-specific protein interactions. *Human Molecular Genetics*. 2015; **24**: 7406–7420. DOI: 10.1093/hmg/ddv440
- [91] Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *American Journal of Human Genetics*. 2010; **86**: 6–22. DOI: 10.1016/j.ajhg.2009.11.017
- [92] Peng G, et al. Gene and pathway-based second-wave analysis of genome-wide association studies. *European Journal of Human Genetics*. 2010; **8**: 111–117. DOI: 10.1038/ejhg.2009.115
- [93] Edwards SL, et al. Beyond GWASs: illuminating the dark road from association to function. *American Journal of Human Genetics*. 2013; **93**: 779–797. DOI: 10.1016/j.ajhg.2013.10.012
- [94] Fachal L, Dunning AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. *Current Opinion in Genetics & Development*. 2015; **30**: 32–41. DOI: 10.1016/j.gde.2015.01.004
- [95] Manolio TA, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; **461**: 747–53. DOI: 10.1038/nature08494

