Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

From candidate gene studies to GWAS and post-GWAS analyses in breast cancer

Laura Fachal[1,2] and Alison M Dunning[1]

There are now more than 90 established breast cancer risk loci, with 57 new ones, revealed through genome-wide-association studies (GWAS) during the last two years. Established high, moderate and low penetrance genetic variants currently explain ~49% of familial breast cancer risk. GWAS-discovered variants account for 14%, and it is estimated that another 1000 yet-to-be-discovered loci could contribute an additional ~14% of familial risk. Polygenic risk scores can already be used to stratify breast cancer risk in the female population and could improve the targeting of mammographic screening programmes, which are at present largely based on age-specific risks. Fine-scale mapping and functional analyses are revealing candidate causal variants and the molecular mechanisms by which GWAS-hits may act. Better-powered GWAS and genome-wide sequencing projects are likely to continue identifying new breast cancer causal variants.

Addresses
[1] Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK
[2] Genomic Medicine Group, CIBERER, University of Santiago de Compostela, 15706 Santiago de Compostela, Spain

Corresponding author: Dunning, Alison M (amd24@medschl.cam.ac.uk)

Current Opinion in Genetics & Development 2015, **30**:32–41

This review comes from a themed issue on **Cancer genomics**

Edited by **Christine A Iacobuzio-Donahue** and **Elaine A Ostrander**

For a complete overview see the Issue

Available online 27th February 2015

http://dx.doi.org/10.1016/j.gde.2015.01.004

Introduction

During the 1990s, the two major susceptibility genes for breast cancer, *BRCA1* [1] and *BRCA2* [2], were identified in addition to two other genes, *TP53* [3] and *STK11* [4], which cause breast cancer among other cancers. Mutations in these high penetrance genes account for approximately 30% of the familial risk of breast cancer [5]. Following these discoveries, a large number of candidate gene studies were conducted over the following decade, aimed at identifying moderate and low penetrance alleles believed to be responsible for the remaining familial risk. Selected common variants (mostly single nucleotide polymorphisms, SNPs) in genes involved in steroid hormone metabolism, carcinogen metabolism or DNA repair were evaluated by this approach. Several genes, implicated in DNA repair, such as *ATM* [6], *CHEK2* [7], *BRIP1* [8] and *PALB2* [9]; or apoptosis, for example, *CASP8* [10,11] were identified. However, the great majority of reported SNP associations in candidate genes could not be replicated [12]. The lack of success of these hypothesis-driven candidate gene studies made researchers question whether the relevant genes had been examined, and ultimately, lead to rapid adoption of empirical genome wide association study (GWAS) approaches, once new technological advances had made these possible. Since GWAS simply point to a genomic region (locus) where the causal variant is located, post-GWAS analyses are subsequently required to identify the causal variant, the mechanism by which this causal variant is acting, and the target gene.

Summary of GWAS findings

In 2007 one of the very first large GWAS reported five significant loci associated with breast cancer risk [13]. Since then 78 new loci have been identified through similar genome-wide approaches [14–30,31•,32••,33,34] (Table 1). The rate of discovery has continued apace with 57 of these new loci being identified in the last two years [26–30,31•,32••,33,34]. The international Collaborative Oncological Gene–environment Study (COGS) [35] revealed 47 new loci and validated 23 previously reported loci. The large statistical power afforded by combining more than 50 individual studies, comprising up to 55 342 cases and 54 455 controls of both European and Asian ancestry, has enabled loci with ever smaller magnitudes of effect to be identified and these new variants are associated with relative increases in breast cancer risk between 1.05 and 1.26 [30,31•,32••,34]. Recent studies have also been sufficiently powered to focus on specific breast tumour subtypes: to-date most loci have been found to be associated with estrogen receptor (ER) positive disease but eight are associated with predominantly ER-negative disease (see Table 1).

Contribution of new loci to familial risk

The 83 common breast cancer risk loci identified via GWAS (Table 1), explain approximately 14% of the inherited genetic component of breast cancer [32••]. However, Michailidou and colleagues [32••] estimate,

Table 1

Summary of the breast cancer susceptibility loci identified by a GWAS approach

Study	Year	SNP[a]	Locus	Gene[d]	Overall risk		Tumour subtype risk		
					OR[e]	*P* value	OR[e]	*P* value	Phenotype
Easton *et al*. [13]	2007	rs889312	5q11.2[b,c]	intergenic	1.13	7.00×10^{-20}			
		rs13281615	8q24.21[b,c]	*CASC21, CASC8*	1.08	5.00×10^{-12}			
		rs2981582	10q26.13[c]	*FGFR2*	1.26	2.00×10^{-76}			
		rs3817198	11p15.5[b,c]	*LSP1*	1.07	3.00×10^{-9}			
		rs3803662	16q12.1[c]	*CASC16*	1.20	1.00×10^{-36}			
Stacey *et al*. [15]	2007	rs13387042	2q35[b,c]	intergenic	1.20	1.30×10^{-13}			
Stacey *et al*. [16]	2008	rs10941679	5p12[c]	intergenic	1.19	2.90×10^{-11}	1.27	2.5×10^{-12}	ER+
Zheng *et al*. [17]	2009	rs2046210	6q25.1[b,c]	intergenic	1.29	2.00×10^{-15}			
Ahmed *et al*. [18]	2009	rs4973768	3p24.1[b,c]	*SLC4A7*	1.11	4.10×10^{-23}			
		rs6504950	17q22[c]	*STXBP4*	0.95	1.40×10^{-8}			
Thomas *et al*. [19]	2009	rs11249433	1p11.2[c]	*EMBP1*	1.16	$6.74 \times 10^{-1\circ}$			
		rs999737	14q24.1[b,c]	*RAD51B*	0.94	1.74×10^{-7}			
Turnbull *et al*. [20]	2010	rs3757318	6q25.1[b,c]	*CCDC170*	1.30	2.90×10^{-6}			
		rs1562430	8q24.21[b]	*CASC21, CASC8*	1.17	5.80×10^{-7}			
		rs1011970	9p21.3[c]	*CDKN2B*	1.09	2.50×10^{-8}			
		rs2380205	10p15.1	intergenic	0.94	4.60×10^{-7}			
		rs10995190	10q21.2[b,c]	*ZNF365*	0.86	5.10×10^{-15}			
		rs704010	10q22.3[c]	*ZMIZ1*	1.07	3.70×10^{-9}			
		rs909116	11p15.5[b]	*TNNT3*	1.17	7.30×10^{-7}			
		rs614367	11q13.3[c]	intergenic	1.15	3.20×10^{-15}			
Antoniou *et al*. [21]	2010	rs8170	19p13.11[b]	*BABAM1*	1.26[f]	2.30×10^{-9}	1.28	1.2×10^{-6}	TNBC
		rs2363956	19p13.11[b]	*ANKLE1*	0.84[f]	5.50×10^{-9}	0.80	1.1×10^{-7}	TNBC
Haiman *et al*. [23]	2011	rs10069690	5p15.33[c]	*TERT*	ne		1.18	$1.0 \times 10^{-1\circ}$	ER–
Fletcher *et al*. [24]	2011	rs9383938	6q25.1[b]	*ESR1*	1.18	1.41×10^{-7}			
		rs865686	9q31.2[b,c]	intergenic	0.89	1.75×10^{-10}			
Cai *et al*. [25]	2011	rs10822013	10q21.2[b]	*ZNF365*	1.12	5.87×10^{-9}			
Ghoussaini *et al*. [26]	2012	rs10771399	12p11.22[c]	intergenic	0.85	2.70×10^{-35}			
		rs1292011	12q24.21[c]	intergenic	0.92	4.30×10^{-19}	0.90	2.0×10^{-15}	ER+
		rs2823093	21q21.1[c]	intergenic	0.94	1.10×10^{-12}	0.93	4.6×10^{-8}	ER+
Siddiq *et al*. [27]	2012	rs17530068	6q14.1[c]	intergenic	1.12	1.10×10^{-9}			
		rs2284378	20q11.22	*RALY*	1.08	1.30×10^{-6}	1.16	1.1×10^{-8}	ER–
Long *et al*. [28]	2012	rs9485372	6q25.1[b]	*TAB2*	0.90	3.86×10^{-12}			
Kim *et al*. [29]	2012	rs13393577	2q34	*ERBB4*	1.53	8.80×10^{-14}			
Couch *et al*. [30]	2013	rs2290854	1q32.1[b]	*MDM4*	ne		1.16	1.26×10^{-7}	ER–
Garcia-Closas *et al*. [31]	2013	rs6678914	1q32.1[b]	*LGR6*	ne		1.10	1.4×10^{-8}	ER–
		rs12710696	2p24.1	intergenic	ne		1.10	4.6×10^{-8}	ER–
		rs11075995	16q12.2[b]	*FTO*	ne		1.11	4.0×10^{-8}	ER–
Michailidou *et al*. [32••]	2013	rs616488	1p36.22	*PEX14*	0.94	2.00×10^{-10}			
		rs11552449	1p13.2	*DCLRE1B*	1.07	1.80×10^{-8}			
		rs4849887	2q14.2	intergenic	0.91	3.70×10^{-11}			
		rs2016394	2q31.1[b]	intergenic	0.95	1.20×10^{-8}	0.94	1.1×10^{-8}	ER+
		rs1550623	2q31.1[b]	intergenic	0.94	3.00×10^{-8}			
		rs16857609	2q35[b]	*DIRC3*	1.08	1.10×10^{-15}			
		rs6762644	3p26.1	*ITPR1*	1.07	2.20×10^{-12}	1.07	1.4×10^{-8}	ER+
		rs12493607	3p24.1[b]	*TGFBR2*	1.06	2.30×10^{-8}	1.07	1.0×10^{-7}	ER+
		rs9790517	4q24	*TET2*	1.05	4.20×10^{-8}			
		rs6828523	4q34.1	*ADAM29*	0.90	3.50×10^{-16}	0.87	2.9×10^{-14}	ER+
		rs10472076	5q11.2[b]	intergenic	1.05	2.90×10^{-8}			
		rs1353747	5q11.2[b]	*PDE4D*	0.92	2.50×10^{-8}			
		rs1432679	5q33.3	*EBF1*	1.07	2.00×10^{-14}			
		rs11242675	6p25.3	intergenic	0.94	7.10×10^{-9}			
		rs204247	6p23	intergenic	1.05	8.30×10^{-9}	1.06	9.0×10^{-8}	ER+
		rs720475	7q35	*ARHGEF5*	0.94	7.00×10^{-11}	0.93	2.9×10^{-8}	ER+
		rs9693444	8p12	intergenic	1.07	9.20×10^{-14}			
		rs6472903	8q21.11[b]	*CASC9*	0.91	1.70×10^{-17}			
		rs2943559	8q21.11[b]	*HNF4G*	1.13	5.70×10^{-15}			
		rs11780156	8q24.21[b]	intergenic	1.07	3.40×10^{-11}			
		rs10759243	9q31.2[b]	intergenic	1.06	1.20×10^{-08}	1.08	6.0×10^{-10}	ER+

Table 1 (*Continued*)

Study	Year	SNP[a]	Locus	Gene[d]	Overall risk		Tumour subtype risk		
					OR[e]	*P* value	OR[e]	*P* value	Phenotype
		rs7072776	10p12.31[b]	intergenic	1.07	4.30×10^{-14}	1.09	2.5×10^{-11}	ER+
		rs11814448	10p12.31[b]	intergenic	1.26	9.30×10^{-16}			
		rs7904519	10q25.2	*TCF7L2*	1.06	3.10×10^{-8}			
		rs11199914	10q26.12	intergenic	0.95	1.90×10^{-8}	0.94	9.1 × 10-8	ER+
		rs3903072	11q13.1	intergenic	0.95	8.60×10^{-12}			
		rs11820646	11q24.3	intergenic	0.95	1.10×10^{-9}			
		rs12422552	12p13.1	intergenic	1.05	3.70×10^{-8}			
		rs17356907	12q22	intergenic	0.91	1.80×10^{-22}			
		rs11571833	13q13.1	*BRCA2*	1.26	4.90×10^{-8}			
		rs2236007	14q13.3	*PAX9*	0.93	1.70×10^{-13}	0.91	1.9×10^{-10}	ER+
		rs2588809	14q24.1[b]	*RAD51B*	1.08	1.40×10^{-10}	1.10	5.7×10^{-9}	ER+
		rs941764	14q32.11	*CCDC88C*	1.06	3.70×10^{-10}			
		rs17817449	16q12.2[b]	*FTO*	0.93	6.40×10^{-14}			
		rs13329835	16q23.2	*CDYL2*	1.08	2.10×10^{-16}	1.09	3.4×10^{-10}	ER+
		rs527616	18q11.2[b]	intergenic	0.95	1.60×10^{-10}			
		rs1436904	18q11.2[b]	*CHST9*	0.96	3.20×10^{-8}	0.93	7.3×10^{-8}	ER+
		rs4808801	19p13.11[b]	*ELL*	0.93	4.60×10^{-15}			
		rs3760982	19q13.31	intergenic	1.06	2.10×10^{-10}			
		rs132390	22q12.2	*EMID1*	1.12	3.10×10^{-9}			
		rs6001930	22q13.1	*MKL1*	1.12	8.80×10^{-19}			
Cai *et al*. [33]	2014	rs4951011	1q32.1[b]	*ZC3H11A*	1.09	8.82×10^{-9}			
		rs10474352	5q14.3	intergenic	1.09	1.67×10^{-9}			
		rs2290203	15q26.1	*PRC1*	1.08	4.25×10^{-8}			
Milne *et al*. [34]	2014	rs1053338	3p14.1	*ATXN7*	1.07	1.00×10^{-8}			
		rs6964587	7q21.2	*AKAP9*	1.05	2.00×10^{-10}			

ne: non evaluated; TNBC: triple negative (estrogen receptor, progesterone receptor and human epidermal growth factor receptor 2) breast cancer.
[a] Independent associated variant.
[b] Locus mapped by more than one independent variant.
[c] Locus replicated in Michailidou *et al*. [32••] at a GWAS significance level.
[d] Name of the gene where the variant lies (intronic, exonic or at 5′ or 3′UTR regions). Note that not always the gene where the variant is placed is the one affected by the causal variant within the associated locus.
[e] Per allele OR.
[f] Breast cancer risk for *BRCA1* mutation carriers.

from Quantile–Quantile (Q–Q) plots, that another 1000 common variants with smaller effects may still be undiscovered.[3] The combined effect of these with the known common variants could explain up to 28% of familial risk [32••], while high-penetrance alleles account for approximately 30%, and moderate penetrance ones explain a further 5% of familial breast cancer risk [5,36] (Figure 1 summarizes the known genetic architecture of breast cancer). The great majority of undiscovered common alleles are predicted to confer increases in risk between 1.02 and 1.05 and, given that the smallest GWAS-significant odds ratios found are currently ~1.05, even larger sample sizes will be needed to identify such variants.

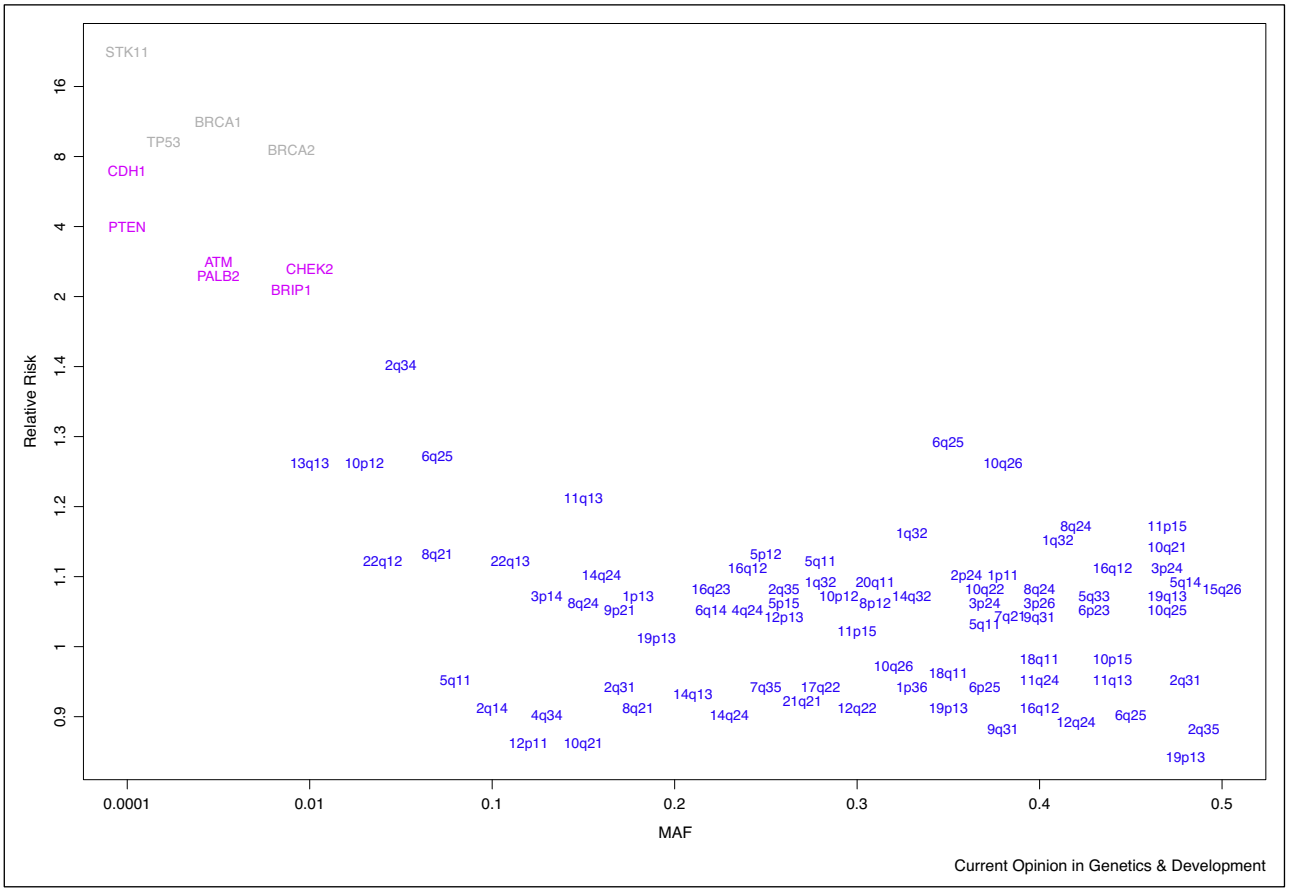
The OncoArray consortium (URL: http://epi.grants.cancer.gov/oncoarray/) has recently designed a custom genotyping chip, which includes ~530,000 rationally selected variants. The aim of this umbrella consortium is to gather new insights into the genetic architecture and mechanisms underlying breast, ovarian, prostate, colorectal, and lung cancers. Through this, an additional 100,000 new breast cancer cases and controls are being analysed. The combined sample size achieved (summed with the COGS breast cancer samples) should permit the detection of variants with smaller effect sizes than has been previously possible at GWAS-appropriate significance levels.

SNP profiles for risk prediction

The individual risks conferred by GWAS-discovered loci are low but their combined effects are useful for population-based risk stratification. For example, Mavaddat *et al*. [37••] estimated that, by stratifying the female population using the current set of known risk loci, the top 1% would have a ~3.36-fold higher risk than the population average. Eligibility for current breast screening programmes is based on population-risk at particular ages — a mammographic breast-screening program is offered to women aged 47–73 in UK (at which ages they have a 10-year absolute risk of developing

[3] Q-Q plots compare the expected probability distribution (like –log10(P value) or x^2 statistics) under the null hypothesis of no association with the observed values. Deviations of the largest observed values from the expected distribution suggest that these variants could represent a true association. Thus, Michailidou et al. [32] identified up to 1,168 independent variants associated with risk, although they did not reach genome-wide levels of significance.

Figure 1



Allelic architecture of breast cancer (updated from [52] and [53]). Dark blue, low penetrance alleles. Magenta, moderate penetrance alleles. Grey, high penetrance alleles.

breast cancer of ≥2.5%) [38], and 40–64 in USA (USA 10-year absolute risk ≥1.45[4]). Thus, risk scores, encompassing known common variants could allow younger women, at equivalent absolute risks, to benefit from screening programmes, whilst decreasing, by ~24%, the proportion of women in the current age-groups for whom breast cancer screening would be considered useful [39•]. Moreover, according to Mavaddat *et al.* [37••], polygenic risk scores based on the set of known loci alone could identify the 17% of women with the highest lifetime risks (this is the recommended enhanced surveillance according to UK NICE guidelines; URL: https://www.nice.org.uk/guidance).

Although genetic profiles are potentially very useful for risk stratification in prevention programmes, their discriminatory accuracy in predicting individual breast cancer risks is limited. However, SNP profiles could be added to existing individual risk models, which are currently based on established risk factors such as personal, clinical and family history information, reproductive information, and environmental or lifestyle factors. Dite and colleagues [40] have evaluated the effect of incorporating seven SNPs: rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs3817198 and rs4415084 [correlated with rs10941679 in Table 1] to the NCI's Breast Cancer Risk Prediction Tool (BCRAT, URL: http://www.cancer.gov/bcrisktool/). Their results indicate that inclusion of genetic data improves significantly the discriminative accuracy of BCRAT (P value <0.001), increasing the area under the receiver operating characteristic curve (AUC) from 0.58 to 0.61, which represents an increase in specificity and sensitivity of the model.

Another consideration is that current GWAS hits are very rarely directly functional — panels of SNPs on GWAS arrays are selected as tags for all other common variants and the probability of them being directly causal is remote. It is thus reasonable to expect that the identification of the

[4] Age-conditional absolute risk was derived using DevCan 6.7.2 (URL: http://surveillance.cancer.gov/devcan/), as in Pashayan *et al.* [38], using all races database (2000–2011) from National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) Program.

Table 2

Summary of candidates for being the functionally relevant gene mapped by GWAS hits

Pathway	GWAS hit	Study	Locus	Gene	Gene function, and proposed mechanism of action
DNA damage recognition and repair	rs999737 rs2588809	Thomas *et al*. [19] Michailidou *et al*. [32••]	14q24.1	*RAD51B*	*RAD51B* forms the BCDX2 complex (with *RAD51C*, *RAD51D* and *XRCC2*), which is involved in repair of DNA double-strand breaks through the homologous recombination mediated pathway, and in the Fanconi anemia (FA) pathway [54].
	rs8170	Antoniou *et al*. [21]	19p13.11	*BABAM1*	*BABAM1* interacts with *BRCA1* stabilizing the BRCA1A complex (alongside with *RAP80, ABRAXAS, BRCC36*, and *BRCC45*), which is involved in G2–M checkpoint control [55].
	rs11552449	Michailidou *et al*. [32••]	1p13.2	*DCLRE1B*	*DCLRE1B* interacts with the Mre11-Rad50-Nbs1 complex, as part of the homologous recombination pathway; and with FancD2, within the FA pathway [56].
	rs10759243	Michailidou *et al*. [32••]	9q31.2	*RAD23B*	*RAD23B* is involved in the nucleotide excision repair pathway. Defects in the global genome NER sub-pathway result in cancer predisposition [57].
	rs3903072	Michailidou *et al*. [32••]	11q13.1	*MUS81*	*MUS81* cooperates with *FANCC* in response to crosslink damage and chromosomal integrity [58].
Apoptosis	rs4245739	Garcia-Closas *et al*. [31]	1q32.1	*MDM4*	MDM4 is a specific inhibitor of p53 [59]. Thus the observed increase in risk associated with these two alleles could be mediated though the regulation of p53, which is involved in cell cycle arrest, after DNA damage, and triggers apoptosis.
	rs4808801	Michailidou *et al*. [32••]	19p13.11	*ELL*	*ELL* can reduce functional activity of p53 [60].
	rs1353747	Michailidou *et al*. [32••]	5q11.2	*PDE4D*	Depletion of endogenous *PDE4D* cause apoptosis and growth inhibition in multiple types of cancer cells, including breast cancer cells [61].
	rs2236007	Michailidou *et al*. [32••]	14q13.3	*PAX9*	*PAX9* gene encodes a transcription factor whose inhibition has been related with the induction of apoptosis [62].
	rs2290203	Cai *et al*. [33].	15q26.1	*PRC1*	A significant increase in endogenous Prc1 levels has been observed in breast cancer cell lines, whereas knockdown of endogenous *PRC1* caused dysfunction in the cytokinesis process in breast cancer cells and results in cell death [63]. *PRC1* function is regulated by p53 [64].
Estrogen receptor signalling	rs2981582	Easton *et al*. [13]	10q26.13	*FGFR2*	Fine-scale mapping of 10q26.13 locus revealed that the effect of one causal variant is mediated through the modification of a FOXA1 (Forkhead Box A1) finding site [43••]. Meyer and colleagues [43••] demonstrated that one of the 10q26 causal variants is affecting FOXA1, and it is able to recruit Estrogen Receptor alpha (ERα) to this site in an allele-specific manner, which is in agreement with the association between *FGFR2* and ER+ disease.
	rs13387042	Stacey *et al*. [15]	2q35	*IGFBP5*	Two recent studies identify that the SNP rs4442975 (r2 with rs13387042 = 0.93) affect a consensus binding site for FOXA1 [49•,51•]. This association could be mediated by ER binding, which would be consistent with the observed association between rs13387042 and ER+ breast cancer. Although no ER binding sites are located in the vicinity of the associated variant[51•], Hurtado *et al*. [65] suggest that FOXA1 could also act stabilizing ER binding from a distance, possibly through chromatin loops between distinct ER binding regions. Ghoussaini *et al*. [49•] demonstrated that the associated variant modifies the expression of *IGFBP5* gene, which is related with mammary development and increased apoptotic cell death.

Table 2 (*Continued*)

Pathway	GWAS hit	Study	Locus	Gene	Gene function, and proposed mechanism of action
	rs9383938	Fletcher *et al*. [24]	6q25.1	*ESR1*	*ESR1* codes for the ERα. Due to its location at 5′ UTR of *ESR1*, the authors proposed that the associated variant could be modifying *ESR1* levels of expression [24].
	rs2823093	Ghoussaini *et al*. [26]	21q21.1	*NRIP1*	*NRIP1* is an essential factor for normal mammary gland development. According to a recent study, Nrip1 acts with ERα as a co-regulator of several factors that influence key mitogenic pathways that regulate normal mammary gland development [66].
Tumour progression and metastatic disease	rs614367	Turnbull *et al*. [20]	11q13.3	*CCD1*	French et a. [46•] demonstrated that the variants associated with breast cancer risk at the 11q13 locus modify Cyclin D1 expression. A recent study observed that the repression of *CCD1* by the tumour suppressor miR-206 activates cell cycle arrest resulting in a decrease in cell proliferation [67].
	rs10771399	Ghoussaini *et al*. [26]	12p11.22	*PTHLH*	*PTHLH* plays a key role during the formation of the mammary glands, and it has been related with breast cancer bone metastasis [68].
	rs4808611 rs2943559	Antoniou *et al*. [21] Michailidou *et al*. [32••]	19p13.11 8q21.11	*NR2F6 HNF4G*	*HNF4G*, and *NR2F6* are both involved in cancer progression, *HNF4G* promotes both the proliferation and invasion of bladder cancer cells [70].
	rs12493607	Michailidou *et al*. [32••]	3p24.1	*TGFBR2*	Loss of Tgfbr2 has been related with breast cancer progression [71], and metastatic disease [72].
	rs720475	Michailidou *et al*. [32••]	7q35	*ARHGEF5*	*ARHGEF5* has been related with breast tumour progression and proliferative breast disease [73]
	rs6001930	Michailidou *et al*. [32••]	22q13.1	*MKL1*	*MKL1* has been related with breast tumour progression and proliferative breast disease [69]
Epigenetic changes	rs9790517 rs1432679	Michailidou *et al*. [32••]	4q24 5q33.3	*TET2* *EBF1*	A recent study describes that the transcription factor *EBF1* is an interaction partner of *TET2* gene, and suggest that both are involved in the regulation of DNA methylation [74]

directly causal variant(s) at each locus, via fine scale mapping studies, will further increase the accuracy of risk models.

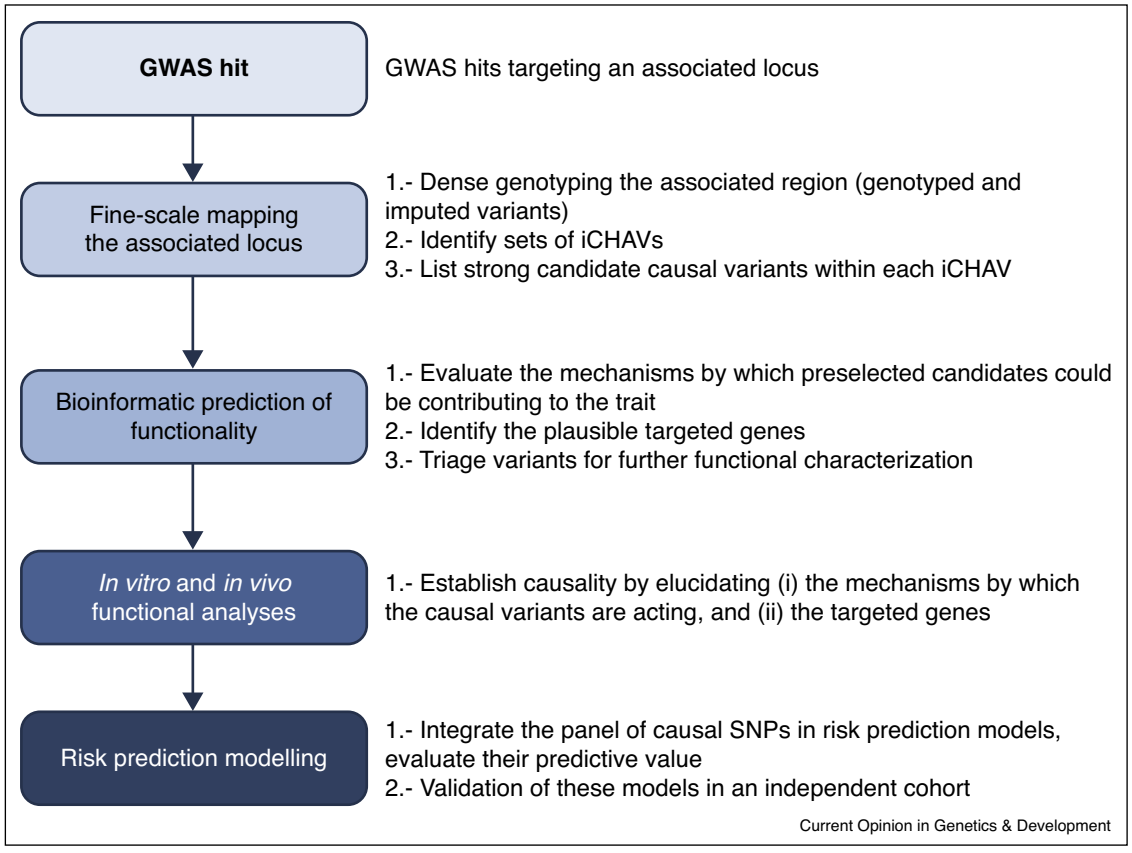
Future directions — fine scale mapping and assays to identify the target genes of GWAS-identified loci

It is always tempting to speculate on the genetic mechanisms and molecular pathways by which GWAS loci affect risk of breast cancer (see Table 2). Some loci are near genes with known functions, indicating these may be good candidates for being the functionally relevant gene. Others lie in 'gene deserts' with no nearby genes. Despite much speculation, the functional effects and target genes at most of the confirmed loci have yet to be elucidated, while those that have been examined in detail have often generated surprising results.

A battery of post-GWAS analyses (see Figure 2) is needed to properly identify the truly functional variants, directly responsible for the observed risk-differences, and to unravel the mechanisms underlying their effects. Post-GWAS analyses include detailed genetic epidemiological dissection of the associated locus, bio-informatic prediction of functionality, and *in vitro* and *in vivo* experimental verification of the molecular mechanisms for the causal variants and their target genes [41••]. The initial epidemiological studies require dense SNP genotyping in large sample sizes, first, to analyse the effects of less common candidate variants (minor frequency ~1–5%), and second, to separate neighbouring genetic variants that are often correlated (the difficulty is in recognising the truly causal variants from among many highly correlated SNPs). Udler *et al*. [42] calculated that fine mapping studies, which include all the common variation at the GWAS-discovered locus, require sample sizes one to four times larger than the original GWAS, which only includes tag SNPs, to identify the true causal variant. Thus, to date only eight confirmed loci have published fine-scale mapping studies [43••,44,45,46•,47•,48,49•]. Chakravarti *et al*. suggested that to formally establish causality of a given variant it is necessary to demonstrate in human cells/tissues or animal models, that recreating the risk variants would generate analogous phenotypes in the model system [50].

Detailed analysis of the eight loci examined so fare has revealed several strong candidate functional variants and their targeted genes. For instance, the fine scale mapping of the 2q35 locus indicated that candidate causal SNP, rs4442975, confers increased breast cancer risk through the regulation of the *IGFBP5* gene (encoding insulin-like growth factor binding protein 5), although this is not the nearest gene to the GWAS hit [49•,51•]. A study of

Figure 2



Post-GWAS analyses workflow. ICHAV, independent set of correlated, highly trait-associated variants.

the 5p15 locus reported multiple functional variants in the *TERT* gene with various effects on telomere length, breast, prostate and ovarian cancer risk. These variants lie in three different functional elements with effects on the *TERT* promoter activity, a silencer element, and the generation of an aberrant splice site, causing truncation of the translated telomere reverse transcriptase protein [47•]. Unexpectedly, given previous hypotheses about telomere length and cancer risk, the variant with the greatest effect on telomere length did not greatly affect hormonal cancer risks, while those with the biggest effects on risk clearly did not act by altering telomere length. Analysis of the 10q26 locus, revealed three independent causal variants [43••,44], each situated at a DNase hypersensitive site. Two of these variants (rs45631563, rs2981578) were found to alter transcription factor (TF) binding sites for E2F1 and FOXA1/Erα, respectively. Chromatin conformation capture demonstrated the target gene of these variants to be *FGFR2*, encoding the fibroblast growth factor receptor 2. In a similar manner, three functional variants have been identified at 11q13 [46•]. These three variants affect (rs554219, rs78540526) or create (rs75915166) TF binding sites (ELK4 and GATA3, respectively), and hence affect the regulation of target gene *CCDN1*, encoding Cyclin D1, as determined by chromatin conformation assays.

Conclusions

Recent GWAS and follow-up studies have identified 57 new risk loci and substantially increased understanding of the mechanisms underlying breast cancer. Now that the target genes of GWAS hits are being revealed we can see that old hypotheses about candidate breast cancer pathways were not entirely incorrect — candidate gene studies largely failed because the selected variants did not tag sufficient variability at the evaluated genes and that they were underpowered. Post-GWAS analyses are now demonstrating that the functional consequences of candidate causal variants are mainly on transcriptional regulation, rather than protein coding. Although individual SNP effects are small, there is a likely prospect that SNP-risk-profiling can improve the targeting of breast cancer screening programmes in the near future. The next steps in the genetic epidemiology of breast cancer will need to include the assessment of variants with lower frequencies and smaller effect sizes. These will necessitate even larger cohorts of breast cancer patients, as well as the development of new statistical methods, to

comprehensively evaluate combinations of variants conferring low to moderate increases in risk on an already complex disease.

Acknowledgements

LF is supported by a postdoc fellowship from Xunta de Galicia and European Social Fund (POS-A/2013/034). AMD holds a Cancer Research-UK program [c8197/A16565].

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W *et al.*: **A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1***. *Science* 1994, **266**:66-71.
2. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G *et al.*: **Identification of the breast cancer susceptibility gene *BRCA2***. *Nature* 1995, **378**:789-792.
3. Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA *et al.*: **Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms**. *Science* 1990, **250**:1233-1238.
4. Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, Bignell G, Warren W, Aminoff M, Hoglund P *et al.*: **A serine/threonine kinase gene defective in Peutz–Jeghers syndrome**. *Nature* 1998, **391**:184-187.
5. Melchor L, Benitez J: **The complex genetic landscape of familial breast cancer**. *Hum Genet* 2013, **132**:845-863.
6. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K *et al.*: **ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles**. *Nat Genet* 2006, **38**:873-875.
7. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, Hollestelle A, Houben M, Crepin E, van Veghel-Plandsoen M *et al.*: **Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations**. *Nat Genet* 2002, **31**:55-59.
8. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K *et al.*: **Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles**. *Nat Genet* 2006, **38**:1239-1241.
9. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T *et al.*: **PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene**. *Nat Genet* 2007, **39**:165-167.
10. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MWR, Pooley KA, Scollen S, Baynes C, Ponder BAJ, Chanock S *et al.*: **A common coding variant in CASP8 is associated with breast cancer risk**. *Nat Genet* 2007, **39**:352-358.
11. Lin W-Y, Camp NJ, Ghoussaini M, Beesley J, Michailidou K, Hopper JL, Apicella C, Southey MC, Stone J, Schmidt MK *et al.*: **Identification and characterisation of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk**. *Hum Mol Genet* 2015, **24**:285-298.
12. Dunning AM, Healey CS, Pharoah PDP, Teare MD, Ponder BAJ, Easton DF: **A systematic review of genetic polymorphisms and breast cancer risk**. *Cancer Epidemiol Biomarkers Prev* 1999, **8**:843-854.
13. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R *et al.*: **Genome-wide association study identifies novel breast cancer susceptibility loci**. *Nature* 2007, **447**:1087-1093.
14. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A *et al.*: **A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer**. *Nat Genet* 2007, **39**:870-874.
15. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A *et al.*: **Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer**. *Nat Genet* 2007, **39**:865-869.
16. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, Jakobsdottir M, Bergthorsson JT, Gudmundsson J, Aben KK *et al.*: **Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer**. *Nat Genet* 2008, **40**:703-706.
17. Zheng W, Long J, Gao Y-T, Li C, Zheng Y, Xiang Y-B, Wen W, Levy S, Deming SL, Haines JL *et al.*: **Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1**. *Nat Genet* 2009, **41**:324-328.
18. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R *et al.*: **Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2**. *Nat Genet* 2009, **41**:585-590.
19. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K *et al.*: **A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1)**. *Nat Genet* 2009, **41**:579-584.
20. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS *et al.*: **Genome-wide association study identifies five new breast cancer susceptibility loci**. *Nat Genet* 2010, **42**:504-507.
21. Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, Healey S, Morrison J, Kartsonaki C, Lesnick T *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**. *Nat Genet* 2010, **42**:885-892.
22. Long J, Cai Q, Shu X-O, Qu S, Li C, Zheng Y, Gu K, Wang W, Xiang Y-B, Cheng J *et al.*: **Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium**. *PLoS Genet* 2010, **6**:e1001002.
23. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, Wang X, Ademuyiwa F, Ahmed S, Ambrosone CB *et al.*: **A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer**. *Nat Genet* 2011, **43**:1210-1220.
24. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C *et al.*: **Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study**. *J Natl Cancer Inst* 2011, **103**:425-435.
25. Cai Q, Long J, Lu W, Qu S, Wen W, Kang D, Lee J-Y, Chen K, Shen H, Shen C-Y *et al.*: **Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium**. *Hum Mol Genet* 2011, **20**:4991-4999.
26. Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, Dennis J, Wang Q, Humphreys MK, Luccarini C *et al.*: **Genome-wide association analysis identifies three new breast cancer susceptibility loci**. *Nat Genet* 2012, **44**:312-318.
27. Siddiq A, Couch FJ, Chen GK, Lindström S, Eccles D, Millikan RC, Michailidou K, Stram DO, Beckmann L, Rhie SK *et al.*: **A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11**. *Hum Mol Genet* 2012, **21**:5373-5384.
28. Long J, Cai Q, Sung H, Shi J, Zhang B, Choi J-Y, Wen W, Delahanty RJ, Lu W, Gao Y-T *et al.*: **Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer**. *PLoS Genet* 2012, **8**:e1002532.
29. Kim H-C, Lee J-Y, Sung H, Choi J-Y, Park S, Lee K-M, Kim Y, Go M, Li L, Cho Y *et al.*: **A genome-wide association study**

identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study. *Breast Cancer Res* 2012, **14**:R56.

30. Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbaecker KB, Soucy P, Fredericksen Z, Barrowdale D, Dennis J *et al.*: **Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk**. *PLoS Genet* 2013, **9**:e1003212.

31. • Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhie SK, Riboli E, Feigelson HS *et al.*: **Genome-wide association studies identify four ER negative-specific breast cancer risk loci**. *Nat Genet* 2013, **45**:392-398.
This study reports three ER negative loci, and validates a fourth, which represents a half of the ER negative locus identified to date. These findings provide further evidence for distinct etiological pathways associated with ER-positive and ER-negative breast cancers.

32. •• Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK *et al.*: **Large-scale genotyping identifies 41 new loci associated with breast cancer risk**. *Nat Genet* 2013, **45**:353-361.
The largest GWAS on breast cancer published to date. It reports 41 new breast cancer-associated loci, and validates 23 previously reported risk loci. Moreover, the authors estimate that 1168 additional loci are associated with increased breast cancer risk.

33. Cai Q, Zhang B, Sung H, Low SK, Kweon SS, Lu W, Shi J, Long J, Wen W, Choi JY *et al.*: **Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1**. *Nat Genet* 2014, **46**:886-890.

34. Milne RL, Burwinkel B, Michailidou K, Arias-Perez JI, Zamora MP, Menendez-Rodriguez P, Hardisson D, Mendiola M, Gonzalez-Neira A, Pita G *et al.*: **Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium**. *Hum Mol Genet* 2014, **23**:6096-6111.

35. Bahcall OG: **iCOGS collection provides a collaborative model**. *Nat Genet* 2013, **45**:343.

36. Easton DF: **How many more breast cancer predisposition genes are there?** *Breast Cancer Res* 1999, **1**:14-17.

37. •• Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M *et al.*: **Prediction of breast cancer risk based on profiling with common genetic variants**. *J Natl Cancer Inst* 2014. (in press).
The authors investigate the value of using 77 breast cancer GWAS hits for risk stratification, and conclude that the observed level of risk discrimination could improve targeted screening and prevention strategies.

38. Pashayan N, Duffy SW, Chowdhury S, Dent T, Burton H, Neal DE, Easton DF, Eeles R, Pharoah P: **Polygenic susceptibility to prostate and breast cancer: implications for personalised screening**. *Br J Cancer* 2011, **104**:1656-1663.

39. • Burton H, Chowdhury S, Dent T, Hall A, Pashayan N, Pharoah P: **Public health implications from COGS and potential for risk stratification and screening**. *Nat Genet* 2013, **45**:349-351.
This paper analyses the public health implications from COGS results.

40. Dite G, Mahmoodi M, Bickerstaffe A, Hammet F, Macinnis R, Tsimiklis H, Dowty J, Apicella C, Phillips K-A, Giles G *et al.*: **Using SNP genotypes to improve the discrimination of a simple breast cancer risk prediction model**. *Breast Cancer Res Treat* 2013, **139**:887-896.

41. •• Edwards SL, Beesley J, French JD, Dunning AM: **Post-GWAS studies: illuminating the dark road from association to function**. *Am J Hum Genet* 2013, **93**:779-797.
Comprehensive review of the methodology of post GWAS analysis.

42. Udler MS, Tyrer J, Easton DF: **Evaluating the power to discriminate between highly correlated SNPs in genetic association studies**. *Genet Epidemiol* 2010, **34**:463-468.

43. •• Meyer Kerstin B, O'Reilly M, Michailidou K, Carlebur S, Edwards Stacey L, French Juliet D, Prathalingham R, Dennis J, Bolla MK, Wang Q *et al.*: **Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1**. *Am J Hum Genet* 2013, **93**:1046-1060.
Reports the causal variants underlying the most significant breast cancer locus identified to date. The authors also demonstrate that one of the independently associated risk alleles binds preferentially FOXA1, which is consistent with the strong association of this locus with ER-positive disease.

44. Udler MS, Meyer KB, Pooley KA, Karlins E, Struewing JP, Zhang J, Doody DR, MacArthur S, Tyrer J, Pharoah PD *et al.*: **FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation**. *Hum Mol Genet* 2009, **18**:1692-1703.

45. Udler MS, Ahmed S, Healey CS, Meyer K, Struewing J, Maranian M, Kwon EM, Zhang J, Tyrer J, Karlins E *et al.*: **Fine scale mapping of the breast cancer 16q12 locus**. *Hum Mol Genet* 2010, **19**:2507-2515.

46. • French Juliet D, Ghoussaini M, Edwards Stacey L, Meyer Kerstin B, Michailidou K, Ahmed S, Khan S, Maranian Mel J, O'Reilly M, Hillman Kristine M *et al.*: **Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers**. *Am J Hum Genet* 2013, **92**:489-503.
This study identifies three independent signals at the 11q13 locus, which are mediating the increase of breast cancer risk through the regulation of its likely target gene *CCND1*.

47. • Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, Edwards SL, Pickett HA, Shen HC, Smart CE *et al.*: **Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer**. *Nat Genet* 2013, **45**:371-384.
Reveals the complexity and pleiotropy of *TERT* locus, providing definitive evidence for genetic control of telomere length by common genetic variants in this locus.

48. Chen F, Chen GK, Millikan RC, John EM, Ambrosone CB, Bernstein L, Zheng W, Hu JJ, Ziegler RG, Deming SL *et al.*: **Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans**. *Hum Mol Genet* 2011, **20**:4491-4503.

49. • Ghoussaini M, Edwards SL, Michailidou K, Nord S, Cowper-Sal lari R, Desai K, Kar S, Hillman KM, Kaufmann S, Glubb DM *et al.*: **Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation**. *Nat Commun* 2014:4.
This study reveals that the 2q35 locus is associated with breast cancer risk through the regulation of *IGFBP5*, a gene with known roles in breast cell biology.

50. Chakravarti A, Clark AG, Mootha VK: **Distilling pathophysiology from complex disease genetics**. *Cell* 2013, **155**:21-26.

51. • Dryden NH, Broome LR, Dudbridge F, Johnson N, Orr N, Schoenfelder S, Nagano T, Andrews S, Wingett S, Kozarewa I *et al.*: **Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C**. *Genome Res* 2014, **24**:1854-1868.
The authors use CHi-C to investigate long-range interactions at three breast cancer gene deserts (2q35, 8q24.21, 9q31.2). They confirm that the 2q35 could be related with the regulation of IGFBP5.

52. Varghese JS, Easton DF: **Genome-wide association studies in common cancers—what have we learnt?** *Curr Opin Genet Dev* 2010, **20**:201-209.

53. Ghoussaini M, Pharoah PD, Easton DF: **Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning?** *Am J Pathol* 2013, **183**:1038-1051.

54. Skorski T: **Oncogenic tyrosine kinases and the dna-damage response**. *Nat Rev Cancer* 2002, **2**:351-360.

55. Huen MS, Sy SM, Chen J: **BRCA1 and its toolbox for the maintenance of genome integrity**. *Nat Rev Mol Cell Biol* 2010, **11**:138-148.

56. Bae JB, Mukhopadhyay SS, Liu L, Zhang N, Tan J, Akhter S, Liu X, Shen X, Li L, Legerski RJ: **Snm1B/Apollo mediates replication fork collapse and S Phase checkpoint activation in response to DNA interstrand cross-links**. *Oncogene* 2008, **27**:5045-5056.

57. Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JH: **Understanding nucleotide excision repair and its roles in cancer and ageing**. *Nat Rev Mol Cell Biol* 2014, **15**:465-481.

58. Larin M, Gallo D, Tamblyn L, Yang J, Liao H, Sabat N, Brown GW, McPherson JP: **Fanconi anemia signaling and Mus81 cooperate to safeguard development and crosslink repair**. *Nucleic Acids Res* 2014, **42**:9807-9820.

59. Toledo F, Wahl GM: **MDM2 and MDM4: p53 regulators as targets in anticancer therapy**. *Int J Biochem Cell Biol* 2007, **39**:1476-1482.

60. Shinobu N, Maeda T, Aso T, Ito T, Kondo T, Koike K, Hatakeyama M: **Physical interaction and functional antagonism between the RNA polymerase II elongation factor ELL and p53**. *J Biol Chem* 1999, **274**:17003-17010.

61. Lin DC, Xu L, Ding LW, Sharma A, Liu LZ, Yang H, Tan P, Vadgama J, Karlan BY, Lester J *et al.*: **Genomic and functional characterizations of phosphodiesterase subtype 4D in human cancers**. *Proc Natl Acad Sci U S A* 2013, **110**:6109-6114.

62. Lee J-C, Sharma M, Lee Y-H, Lee N-H, Kim S-Y, Yun J-S, Nam S-Y, Hwang P-H, Jhee E-C, Yi H-K: **Pax9 mediated cell survival in oral squamous carcinoma cell enhanced by c-myb**. *Cell Biochem Funct* 2008, **26**:892-899.

63. Shimo A, Nishidate T, Ohta T, Fukuda M, Nakamura Y, Katagiri T: **Elevated expression of protein regulator of cytokinesis 1, involved in the growth of breast cancer cells**. *Cancer Sci* 2007, **98**:174-181.

64. Li C, Lin M, Liu J: **Identification of PRC1 as the p53 target gene uncovers a novel function of p53 in the regulation of cytokinesis**. *Oncogene* 2004, **23**:9336-9347.

65. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS: **FOXA1 is a key determinant of estrogen receptor function and endocrine response**. *Nat Genet* 2011, **43**:27-33.

66. Nautiyal J, Steel JH, Mane MR, Oduwole O, Poliandri A, Alexi X, Wood N, Poutanen M, Zwart W, Stingl J *et al.*: **The transcriptional co-factor RIP140 regulates mammary gland development by promoting the generation of key mitogenic signals**. *Development* 2013, **140**:1079-1089.

67. Elliman SJ, Howley BV, Mehta DS, Fearnhead HO, Kemp DM, Barkley LR: **Selective repression of the oncogene cyclin D1 by the tumor suppressor miR-206 in cancers**. *Oncogenesis* 2014, **3**:e113.

68. Wang Y, Lei R, Zhuang X, Zhang N, Pan H, Li G, Hu J, Pan X, Tao Q, Fu D *et al.*: **DLC1-dependent parathyroid hormone-like hormone inhibition suppresses breast cancer bone metastasis**. *J Clin Invest* 2014, **124**:1646-1659.

69. Gurbuz I, Ferralli J, Roloff T, Chiquet-Ehrismann R, Asparuhova MB: **SAP domain-dependent Mkl1 signaling stimulates proliferation and cell migration by induction of a distinct gene set indicative of poor prognosis in breast cancer patients**. *Mol Cancer* 2014, **13**:1476-4598.

70. Okegawa T, Ushio K, Imai M, Morimoto M, Hara T: **Orphan nuclear receptor HNF4G promotes bladder cancer growth and invasion through the regulation of the hyaluronan synthase 2 gene**. *Oncogenesis* 2013, **2**:e58.

71. Novitskiy SV, Pickup MW, Gorska AE, Owens P, Chytil A, Aakre M, Wu H, Shyr Y, Moses HL: **TGF-beta receptor II loss promotes mammary carcinoma progression by Th17 dependent mechanisms**. *Cancer Discov* 2011, **1**:430-441.

72. Forrester E, Chytil A, Bierie B, Aakre M, Gorska AE, Sharif-Afshar AR, Muller WJ, Moses HL: **Effect of conditional knockout of the type II TGF-beta receptor gene in mammary epithelia on mammary gland development and polyomavirus middle T antigen induced tumor formation and metastasis**. *Cancer Res* 2005, **65**:2296-2302.

73. Debily MA, Camarca A, Ciullo M, Mayer C, El Marhomy S, Ba I, Jalil A, Anzisi A, Guardiola J, Piatier-Tonneau D: **Expression and molecular characterization of alternative transcripts of the ARHGEF5/TIM oncogene specific for human breast cancer**. *Hum Mol Genet* 2004, **13**:323-334.

74. Guilhamon P, Eskandarpour M, Halai D, Wilson GA, Feber A, Teschendorff AE, Gomez V, Hergovich A, Tirabosco R, Fernanda Amary M *et al.*: **Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2**. *Nat Commun* 2013, **4**.