

# Functional germline variants in driver genes of breast cancer

Stella Göhler<sup>1</sup> · Miguel Inacio Da Silva Filho<sup>1</sup> · Robert Johansson<sup>2</sup> ·  
Kerstin Enquist-Olsson<sup>3</sup> · Roger Henriksson<sup>2,4</sup> · Kari Hemminki<sup>1,5</sup> · Per Lenner<sup>2</sup> ·  
Asta Försti<sup>1,5</sup>

Received: 6 March 2015 / Accepted: 7 January 2017 / Published online: 25 February 2017  
© Springer International Publishing Switzerland 2017

## Abstract

**Purpose** Germline mutations in tumour suppressor genes cause various cancers. These genes are also somatically mutated in sporadic tumours. We hypothesized that there may also be cancer-related germline variants in the genes commonly mutated in sporadic breast tumours.

**Methods** After excluding the well-characterized breast cancer (BC) genes, we screened 15 novel genes consistently classified as BC driver genes in next-generation sequencing approaches for single nucleotide polymorphisms (SNPs). Altogether 40 SNPs located in the core promoter, 5'- and 3'-UTR or which were nonsynonymous SNPs were genotyped in 782 Swedish incident BC cases and 1,559 matched controls. After statistical analyses, further evaluations related to functional prediction and signatures of selection were performed.

**Results** *TBX3* was associated with BC risk (rs2242442: OR = 0.76, 95% CI 0.64–0.92, dominant model) and with less aggressive tumour characteristics. An association with BC survival and aggressive tumour characteristics was detected for the genes *ATR* (rs2227928: HR = 1.63; 95% CI 1.00–2.64, dominant model), *RUNXI* (rs17227210: HR = 3.50, 95% CI 1.42–8.61, recessive model) and *TTN* (rs2303838: HR = 2.36; 95% CI 1.04–5.39; rs2042996: HR = 2.28; 95% CI 1.19–4.37, recessive model). According to the experimental ENCODE data all these SNPs themselves or SNPs in high linkage disequilibrium with them ( $r^2 \geq 0.80$ ) were located in regulatory regions. *RUNXI* and *TTN* showed also several signatures of positive selection.

**Conclusion** The study gave evidence that germline variants in BC driver genes may have impact on BC risk and/or survival. Future studies could discover further germline variants in known or so far unknown driver genes which contribute to cancer development.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10552-017-0849-3) contains supplementary material, which is available to authorized users.

✉ Stella Göhler  
stella-goehler@gmx.de

<sup>1</sup> Department of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69121 Heidelberg, Germany

<sup>2</sup> Department of Radiation Science, Oncology, Umeå University, 90187 Umeå, Sweden

<sup>3</sup> Department of Public Health and Clinical Medicine/ Nutritional Research, Umeå University, 90187 Umeå, Sweden

<sup>4</sup> Cancer Center Stockholm Gotland, 10239 Stockholm, Sweden

<sup>5</sup> Center for Primary Health Care Research, Clinical Research Center, Lund University, 20502 Malmö, Sweden

**Keywords** Breast cancer · Driver genes · Germline variants · Case–control study · Single nucleotide polymorphism

## Introduction

Breast cancer (BC) is the most frequent cancer among women in both developing and developed regions of the world. In 2012, around 522,000 women died due to BC and another 1.67 million new cancer cases were diagnosed worldwide [1, 2]. About 5–10% of women with a diagnosis of BC do have a family history of BC, which is a known risk factor for this disease [3]. Breast tumours can be caused by germline variants in tumour suppressor genes like *TP53*, which may also be somatically mutated

in sporadic tumours [4]. To identify somatic mutations and to determine genes which are critical in the development of human cancer, the International Cancer Genome Consortium (ICGC) was launched. The first spectra of somatic mutations in human protein-coding genes in BC were published in 2006 and 2007 [5, 6]. Around 90% of BC tumours are caused by somatic mutations, so-called driver mutations, which initiate the carcinogenic process [7–9]. To identify possible driver genes in sporadic breast tumours a number of studies using next-generation sequencing were published in 2012 [10–13]. However, it was not investigated, whether the driver genes could also contain inherited variants, which influence the development of cancer. Thus, the aim of this study was to identify germline variations in potential driver genes which influence BC risk and/or survival.

## Materials and methods

### Study population

The present study was performed using a population-based Swedish cohort consisting of 782 prospectively collected cases and 1,559 age- and gender-matched controls from the Västerbotten intervention project (VIP), the mammary screening project (MSP) and from the Department of Oncology at the Norrlands University Hospital in Umeå. Controls were matched with cases by age at baseline ( $\pm 6$  months) and time of sampling ( $\pm 2$  months). Blood samples were collected from an ethnically homogeneous population living in Umeå (North Sweden) and its surroundings between January 1990 and January 2001 [14]. Prospective cases were identified from the cohorts by record linkage to the regional cancer registry. Date and the reason of death were collected until 30 January 2012 from the Swedish population register while clinical data were enquired from the registry managed by the Northern Sweden Breast Cancer Group (Table 1).

All participants gave informed consents to the use of their samples for research purpose. The study was approved by the ethical committees of the participating institutes.

### Gene/SNP selection

We focused on genes described to carry BC driver mutations in at least two of the following publications: Banerji et al. [10], Ellis et al. [11], Shah et al. [12], Stephens et al. [13]. We were mainly interested in genes not previously reported as BC driver genes, consequently well-known and intensively studied genes, such as BRCA1, BRCA2, TP53 and PTEN, were excluded from the study. SNP selection

**Table 1** Characteristics of breast tumours at the time of diagnosis

Characteristics	Samples
Cases/controls	782/1559
Age at diagnosis, mean $\pm$ SD (years)	58 $\pm$ 8
Range (years)	30–86
Estrogen receptor (ER)	No. (%)
Positive	402 (51.41)
Negative	133 (17.01)
Missing	247 (31.58)
Progesterone receptor (PR)	
Positive	283 (36.19)
Negative	214 (27.37)
Missing	285 (36.45)
Estrogen/progesterone receptor	
ER+/PR+	228 (29.16)
ER+/PR–	71 (9.08)
ER–/PR+	12 (1.53)
ER–/PR–	114 (14.59)
Missing	357 (45.65)
Tumour size	
$\leq 2$ cm	499 (63.81)
$> 2$ cm	228 (29.16)
Missing	055 (7.03)
Lymph node status	
Positive	221 (28.26)
Negative	469 (59.97)
Missing	92 (11.76)
Distant metastases	
Positive	13 (1.66)
Negative	759 (97.06)
Missing	10 (1.28)
Stage at diagnosis	
0	2 (0.26)
I	405 (51.79)
II	325 (41.56)
III	26 (3.32)
IV	15 (1.92)
Missing	9 (1.15)
Histological grade	
1	159 (20.33)
2	357 (45.65)
3	225 (28.77)
Missing	41 (5.24)

was done by Ensembl browser release 69 (<http://www.ensembl.org/index.html>). We emphasized regions with known functions like core promoter, 5'- and 3'-untranslated regions (UTRs) and nonsynonymous SNPs which were described in well-verified transcripts marked in Ensembl as HAVANA manually curated gold transcripts and consensus coding sequences (CCDSs). Haploview was used

to select SNPs on the basis of linkage disequilibrium (LD) ( $r^2 \geq 0.80$ ) to minimize the number of SNPs to be genotyped. An additional inclusion criterion was defined by a minor allele frequency (MAF)  $>10\%$ , with the exception of *ARID2* (rs7570492), *MAP3K13* (rs13091808), *MLL2* (rs11168827) and *MLL3* (rs1323116, rs3735156). CFBF was excluded as it did not have any functional SNPs fulfilling our selection criteria. Reported driver genes with SNPs fulfilling our selection criteria are listed in Online Resource 1.

### Genotyping

Either the KASPar SNP Genotyping System (KBioscience, Hoddesdon, Great Britain) or the TaqMan SNP Genotyping Assay (Life Technologies, Darmstadt, Germany) was used. Both are based on an allele-specific PCR. Master Mix for the KASPar assays was prepared according to the KBioscience's conditions and products, whereas 5x HOT FIREPol Probe qPCR Mix Plus from Solis BioDyne (Tartu, Estonia) was used for the TaqMan assays. In case an assay could not be designed, an assay for a highly linked SNP ( $r^2 \geq 0.80$ ) was ordered instead. PCRs were performed in a 384-well plate format using a total reaction volume of 4  $\mu$ l per well. Endpoint genotype detection was performed using the ViiA7 Real-Time PCR System (Applied Biosystems, Weiterstadt, Germany).

### Statistical analysis

The  $\chi^2$  test was used to test the observed genotype frequencies in the controls for Hardy–Weinberg equilibrium (HWE). To estimate the associations between genotypes and BC risk, odds ratios (ORs) and 95% confidence intervals (CIs) were calculated by logistic regression (PROC LOGISTIC, SAS Version 9.2; SAS Institute, Cary, NC). Relative risk of death was estimated as hazard ratio (HR) (PROC PHREG, SAS Version 9.2, SAS Institute, Cary, NC) via Cox regression. Polymorphisms that showed significant differences in BC-specific survival in the unadjusted model were analysed further by adjusting the data for size of tumour, lymph node metastases, histological grade and estrogen (ER) and progesterone receptor (PR) status (PROC PHREG, SAS Version 9.2, SAS Institute, Cary, NC).  $p$  values  $\leq 0.05$  were considered statistically significant. Kaplan–Meier method (PROC LIFETEST, SAS Version 9.2; SAS Institute) was used to generate survival curves. To measure the differences between the survival functions among the different genotypes log-rank test (PROC LIFETEST, SAS Version 9.2; SAS Institute, Cary, NC) was used. To account for multiple testing, empirical  $p$  values were generated using the permutation option in Plink [15]. Number of permutations used was equal to 10,000.

In this prospective study, we followed the REMARK recommendations for reporting of tumour marker prognostic studies [16].

### In silico functional analyses

To increase our knowledge about the consequences of the SNPs on protein-binding sites, chromatin structure and promoter and enhancer strength, HaploReg (<http://www.broadinstitute.org/mammals/haploreg/haploreg.php>) was used; RegulomeDB (<http://regulome.stanford.edu/>) was utilized to gain detailed information of possible effects on histone modification. All effects were proofed for data in MCF7 (Michigan Cancer Foundation-7 breast cancer cell line), T-47D (epithelial cell line derived from mammary ductal carcinoma), human mammary epithelial cells (HMEC) or MCF10A-ER-SRc (breast epithelial cell line-estrogen receptor-src) cell lines. Effects on transcription factor-binding sites (TFBSs) were calculated through position weight matrices (PWM). PolyPhen/SIFT prediction was used to evaluate a possible impact of an amino acid change on protein structure and function (Ensembl release 75, <http://www.ensembl.org/index.html>). Influence of the 3'-UTR SNPs on micro-RNA binding was studied using microSNiPer (<http://epicenter.ie-freiburg.mpg.de/services/microsniper/>). All linked SNPs mentioned in HaploReg with an  $r^2 \geq 0.80$  among the European population were studied for their influence on promoter, enhancer or chromatin structure.

### Signatures of selection

To gain information about the functional consequences of the SNPs, likelihood of mutations, conservation and recombination rate (RR) of the genes and the SNPs associated with BC risk, survival or tumour characteristics were evaluated. Recombination rate of a specific region was assessed in comparison to the whole chromosome. As an estimate for conservation, phylogenetic  $p$  value (phyloP) was used. A region is conserved if the value is  $\geq .3$  and has a positive prefix. Both variables were analysed with the UCSC Browser (<http://genome.ucsc.edu>). Indication of selective pressure was assessed by analysing first the Fixation Index ( $F_{ST}$ ). It compares allele frequencies between two populations. Values  $\geq 0.25$  indicate strong genetic differentiation and values  $>0.05$  moderate genetic differentiation. Second, integrated haplotype score (iHs), which determines the length of haplotypes around a SNP, was assessed. Scores with  $IiHsI > 2$  are a proof of selection and  $IiHsI > 1.5$  is an indication of selection; a negative score refers to a longer haplotype for derived alleles and a positive score for ancestral alleles. The third value, Fay Wu's  $H$ , distinguishes between a DNA sequence evolving randomly

and the one evolving under positive selection. Strong negative values starting at  $-40$  are considered as a signature of a selective sweep [17, 18]. To identify these values, haplotter was used (<http://haplotter.uchicago.edu>). In case the genotyped SNP was not listed, values of a highly linked SNP ( $r^2 \geq 0.80$ ) were used instead.

## Results

Altogether 20 SNPs in 14 potential driver genes were associated with BC risk, survival and/or clinical and pathological tumour characteristics at  $p \leq 0.05$  level (Tables 2, 3; Online Resource 2 and 3).

### SNPs associated with risk

Five genes were associated with BC risk (Table 2; Online Resource 2). The genotype distribution of rs2242442 (*TBX3*) and rs10497520 (*TTN*) was significantly different between the cases and the controls (overall  $p=0.01$  and  $p=0.03$ , respectively). The most significant association was observed for rs2242442 (*TBX3*): both heterozygous and homozygous carriers of the minor allele were at a decreased risk of BC (OR 0.76, 95% CI 0.64–0.92; dominant model). Also minor allele carriers of another *TBX3* SNP, rs12366395, had a decreased risk (OR 0.83, 95% CI 0.69–1.00; dominant model). LD between these SNPs determined by 1000 genomes was  $r^2=0.01$ . Interestingly, three of the four genotyped SNPs in *TBX3* were associated with less aggressive tumour features: rs2242442 with small tumour size and rs8853 and rs1061651 with low histological grade (Table 2, Online Resource 3).

Among *TTN* rs10497520 minor allele carriers, only the homozygous ones were at increased risk (OR 1.96, 95% CI 1.18–3.26). Four additional SNPs in *TTN* showed associations with less favourable tumour characteristics, large tumour size, high-grade and/or negative hormone receptor status (Online Resource 3).

An increased risk was observed for homozygous carriers of two SNPs ( $r^2=0.25$ ) in *MAP3K1* (rs702688 OR 1.33, 95% CI 0.99–1.76; rs72758040 with OR 1.36, 95% CI 1.01–1.83). However, no association with clinical tumour characteristics was observed for any of the eight genotyped SNPs.

One SNP in *MLL2*, rs11168827, was associated both with risk (OR 1.31, 95% CI 1.00–1.72 for homozygotes), positive hormone receptor status and low grade. A decreased risk was observed for homozygous minor allele carriers of *SF3B1* rs4685 (OR 0.73, 95% CI 0.54–0.97). The SNP was also associated with negative lymph node metastasis and hormone receptor status. By applying the

permutation test, none of these associations remained statistically significant.

### SNPs associated with BC-specific survival

SNPs in *ARID1B*, *ATR*, *RUNX1* and *TTN* showed association with BC-specific survival (Table 3; Online Resource 2). Poor survival was observed for carriers of the minor allele of the SNPs rs73013281 (*ARID1B*) and rs2227928 (*ATR*) (HR 1.58, 95% CI 1.02–2.45 and HR 1.63, 95% CI 1.00–2.64, dominant model), whereas for SNPs rs17227210 (*RUNX1*), rs2303838 (*TTN*) and rs2042996 (*TTN*) poor survival was observed only among homozygous carriers. These results are supported by the Kaplan–Meier plots (Fig. 1). Tumours of the *ARID1B* and *ATR* SNP carriers were also diagnosed at high stage. The association between rs2227928 (*ATR*) and high stage remained statistically significant after applying the permutation test ( $p=0.023$ ). Consequently, the associations of *ARID1B* and *ATR* with survival did not stay significant after adjustment for clinical tumour characteristics (Table 4). The strength of the associations between the *RUNX1* and *TTN* SNPs and survival remained at the same level after adjustment with tumour size, lymph node metastasis status and grade, they even became stronger after further adjustment with hormone receptor status. However, these results should be taken with caution because of the small numbers of the minor homozygote genotypes and because of incomplete hormone receptor status data (Table 1). Nevertheless, as two additional *RUNX1* SNPs, rs8130963 and rs7276777, were associated with stage and four SNPs in *TTN* with less favourable tumour characteristics, a true direct or indirect association of the genetic variation with survival cannot be ruled out (Online Resource 3).

### Functional characterization of the associated SNPs

We used HaploReg to search for SNPs in high LD ( $r^2 \geq 0.80$ ) with SNPs associated with either BC risk, tumour characteristics or survival and used the experimental data obtained in any mammary epithelial (tumour) cell line (MCF7, T-47D, HMEC, MCF10A-ER-SRC) to assess their possible functional role in breast tumorigenesis (Online Resource 4; summarized in Table 5). All promoter and 5'-UTR SNPs, except rs12465459 (*TTN*), were located in an active promoter. According to the ENCODE data, all these SNPs can affect chromatin structure, histone modification, regulatory protein and/or transcription factor binding. Of the eight 3'-UTR SNPs covered by our study, five were predicted to change the binding site for one or more miRNAs and they all had an impact on histone modification and transcription factor-binding sites. A possible or

**Table 2** Genes with SNPs associated with BC risk and their associations with clinical characteristics ( $p \leq 0.05$ )

Gene/SNP	Genotype	Association with risk			Association with clinical parameters			p value	
		Cases (%)	Controls (%)	OR (95% CI)	Clinical characteristics	No. of patients Group 1 (%)	No. of patients Group 2 (%)		OR (95% CI)
<i>MAP3K1</i>									
rs72758040	GG	499 (69.60)	993 (71.90)	1.00					
	AG	125 (17.43)	251 (18.18)	1.00 (0.78–1.28)					
	AA	93 (12.97)	137 (9.92)	<b>1.36 (1.01–1.83)</b>					
	AG+AA	218 (30.40)	388 (28.10)	1.12 (0.92–1.38)					
	Overall								0.12
<i>MAP3K1</i>									
rs702688	GG	300 (39.58)	665 (43.66)	1.00					
	AG	354 (46.70)	682 (44.78)	1.15 (0.96–1.40)					0.14
	AA	104 (13.72)	176 (11.56)	<b>1.33 (0.99–1.77)</b>					<b>0.05</b>
	AG+AA	458 (60.42)	858 (56.34)	1.19 (0.99–1.42)					0.06
	Overall								0.11
<i>MLL2</i>									
rs11168827	GG	291 (38.24)	601 (40.94)	1.00					
	CG	350 (45.99)	669 (45.57)	1.11 (0.91–1.35)					0.29
	CC	120 (15.77)	198 (13.49)	<b>1.31 (1.00–1.72)</b>					<b>0.05</b>
	CG+CC	470 (61.76)	867 (59.60)	1.15 (0.96–1.39)					0.13
	Overall								0.15
<i>SF3B1</i>									
rs4685	GG	319 (42.36)	596 (39.58)	1.00					
	AG	348 (46.22)	688 (45.68)	0.94 (0.78–1.13)					0.50
	AA	86 (11.42)	222 (14.74)	<b>0.73 (0.54–0.97)</b>					<b>0.03</b>
	AG+AA	434 (57.64)	910 (60.42)	0.89 (0.74–1.06)					0.19
	Overall								0.10
Progesterone receptor									
					PR+	PR–			
	GG				94 (33.81)	91 (44.17)	1.00		
	CG				139 (50.00)	89 (43.17)	<b>0.66 (0.45–0.98)</b>		<b>0.04</b>
	CC				45 (16.19)	26 (12.62)	0.60 (0.34–1.05)		0.07
	CG+CC				184 (66.19)	115 (55.83)	<b>0.65 (0.45–0.94)</b>		<b>0.02</b>
Histological grade									
					I+II	III			
	GG				180 (35.79)	102 (46.58)	1.00		
	CG				242 (48.11)	87 (39.73)	<b>0.63 (0.45–0.90)</b>		<b>0.01</b>
	CC				81 (16.10)	30 (13.70)	0.65 (0.40–1.06)		0.09
	CG+CC				323 (64.21)	117 (53.42)	<b>0.64 (0.46–0.88)</b>		<b>0.007</b>
Estrogen receptor									
					ER+	ER–			
	GG				171 (44.30)	51 (39.53)	1.00		
	AG				184 (47.67)	57 (44.19)	1.04 (0.67–1.60)		0.86
	AA				31 (8.03)	21 (16.28)	<b>2.27 (1.20–4.29)</b>		<b>0.01</b>
	AG+AA				215 (55.70)	78 (60.47)	1.22 (0.81–1.83)		0.34
Estrogen receptor/progesterone receptor									
					ER+/PR+	ER–/PR–			
	GG				92 (41.44)	41 (36.61)	1.00		
	AG				111 (50.00)	51 (45.54)	1.03 (0.63–1.69)		0.90
	AA				19 (8.56)	20 (17.86)	<b>2.36 (1.14–4.89)</b>		<b>0.02</b>
	AG+AA				130 (58.56)	71 (63.39)	1.23 (0.77–1.96)		0.39

Table 2 (continued)

Gene/SNP	Genotype	Association with risk			Association with clinical parameters				p value
		Cases (%)	Controls (%)	OR (95% CI)	Clinical characteristics	No. of patients Group 1 (%)	No. of patients Group 2 (%)	OR (95% CI)	
<i>TBX3</i>	GG	430 (57.18)	767 (50.66)	1.00	Regional lymph node metastasis	Negative 189 (41.63)	Positive 97 (45.97)	1.00	
	AG	274 (36.44)	622 (41.08)	<b>0.78 (0.65–0.95)</b>		208 (45.81)	98 (46.45)	0.92 (0.65–1.29)	0.62
	AA	48 (6.38)	125 (8.26)	<b>0.66 (0.46–0.95)</b>		57 (12.56)	16 (7.58)	<b>0.55 (0.30–1.00)</b>	<b>0.05</b>
	AG+AA	322 (42.82)	747 (49.34)	<b>0.76 (0.64–0.92)</b>	Tumour size	265 (58.37)	114 (54.03)	0.84 (0.60–1.16)	0.29
	Overall			<b>0.01</b>		≤2 cm 266 (55.19)	>2 cm 137 (63.13)	1.00	
<i>TBX3</i>	AA	469 (62.12)	873 (58.01)	1.00		181 (37.55)	71 (21.72)	0.76 (0.54–1.07)	0.12
	AG	249 (32.98)	541 (35.95)	0.86 (0.71–1.04)		35 (7.26)	9 (4.15)	0.50 (0.23–1.07)	0.07
	GG	37 (4.90)	91 (6.05)	0.70 (0.47–1.04)		216 (44.81)	80 (36.87)	<b>0.72 (0.52–1.00)</b>	<b>0.05</b>
	AG+GG	286 (37.88)	632 (41.99)	<b>0.83 (0.69–1.00)</b>					
	Overall			<b>0.05</b>					
<i>TTN</i>	CC	525 (69.63)	1091 (72.01)	1.00					
	CT	198 (26.26)	391 (25.81)	1.08 (0.88–1.32)					
	TT	31 (4.11)	33 (2.18)	<b>1.96 (1.18–3.26)</b>					
	CT+TT	229 (30.37)	424 (27.99)	1.15 (0.95–1.40)					
	Overall			<b>0.03</b>					

Significant values are given in bold  
OR odds ratio, CI confidence interval

**Table 3** Genes with SNPs associated with BC-specific survival and their associations with clinical characteristics ( $p \leq 0.05$ )

Gene/SNP	Genotype	Association with survival				Association with clinical parameters				<i>p</i> value
		No. of cases	No. of cases died (%)	HR (95% CI)	<i>p</i> value	Clinical characteristics	No. of patients Group 1(%)	No. of patients Group 2(%)	OR (95% CI)	
<i>ARID1B</i> rs73013281	TT	629	83 (13.20)	1.00		Stage	0+I	II-IV	1.00	
	CT	123	26 (21.14)	<b>1.64 (1.05–2.54)</b>	<b>0.03</b>		339 (85.82)	282 (80.11)	1.00	
	CC	4	0 (0.00)	0 (0.00)	–		53 (13.42)	69 (19.60)	<b>1.57 (1.06–2.31)</b>	
	CT+CC	127	26 (20.47)	<b>1.58 (1.02–2.45)</b>	<b>0.04</b>	Tumour size	3 (0.76)	1 (0.28)	0.40 (0.04–3.88)	
							56 (14.18)	70 (19.89)	<b>1.50 (1.02–2.20)</b>	
<i>ATR</i> rs2227928	GG	189	20 (10.58)	1.00			≤2 cm	>2 cm	1.00	
	AG	395	65 (16.46)	1.62 (0.98–2.67)	0.06		137 (28.84)	42 (18.92)	1.00	
	AA	166	27 (16.27)	1.65 (0.92–2.94)	0.09		238 (50.11)	124 (55.86)	<b>1.70 (1.13–2.56)</b>	
	AG+AA	561	92 (16.40)	<b>1.63 (1.00–2.64)</b>	<b>0.05</b>	Regional lymph node metastasis	100 (21.05)	56 (25.23)	<b>1.83 (1.14–2.94)</b>	
							338 (71.16)	180 (81.08)	<b>1.74 (1.18–2.57)</b>	
<i>RUNX1</i> rs17227210	GG					Negative		Positive		
	AG	555	75 (13.51)	1.00		119 (26.62)	43 (20.28)	1.00		
	TT	184	30 (16.30)	1.18 (0.77–1.81)	0.44		238 (53.24)	111 (52.36)	1.29 (0.85–1.96)	
	*TT vs. CC+CT	14	5 (35.71)	<b>3.66 (1.48–9.08)</b>	<b>0.01</b>		90 (20.13)	58 (27.36)	<b>1.78 (1.10–2.88)</b>	
		14	5 (35.71)	<b>3.50 (1.42–8.61)</b>	<b>0.01</b>	Stage	328 (73.38)	169 (79.72)	1.43 (0.96–2.12)	
<i>TTN</i> rs2303838	CC	528	73 (13.83)	1.00		0+I	II-IV	1.00		
	CT	195	28 (14.36)	1.09 (0.71–1.69)	0.68		118 (30.41)	70 (19.77)	1.00	
	TT	20	6 (30.00)	<b>2.42 (1.05–5.57)</b>	<b>0.04</b>		199 (51.29)	190 (53.67)	<b>1.61 (1.13–2.30)</b>	
	*TT vs. CC+CT	20	6 (30.00)	<b>2.36 (1.04–5.39)</b>	<b>0.04</b>	Estrogen receptor	71 (18.30)	94 (26.55)	<b>2.23 (1.46–3.42)</b>	
		20	6 (30.00)	<b>2.36 (1.04–5.39)</b>	<b>0.04</b>		270 (69.59)	284 (80.23)	<b>1.77 (1.26–2.49)</b>	



Table 3 (continued)

Gene/SNP	Genotype	Association with survival			Association with clinical parameters			p value
		No. of cases	No. of cases died (%)	HR (95% CI)	Clinical characteristics	No. of patients Group 1 (%)	No. of patients Group 2 (%)	
rs2042996	GG	468	64 (13.68)	1.00		247 (63.50)	67 (52.76)	1.00
	AG	252	35 (13.89)	1.04 (0.69–1.58)	0.84	123 (31.62)	55 (43.31)	<b>1.65 (1.09–2.50)</b>
	AA	36	10 (27.78)	<b>3.66 (1.19–4.51)</b>	<b>0.01</b>	19 (4.88)	5 (3.94)	0.97 (0.35–2.69)
	*AA vs. GG+AG	36	10 (27.78)	<b>2.28 (1.19–4.37)</b>	<b>0.01</b>	19 (4.88)	5 (3.94)	0.80 (0.29–2.18)
	GG					I+II	III	
	AG					324 (64.80)	119 (55.35)	1.00
	AA					152 (30.40)	84 (39.07)	<b>1.50 (1.07–2.11)</b>
	*AA vs. GG+AG					24 (4.80)	12 (5.58)	1.36 (0.66–2.81)
						24 (4.80)	12 (5.58)	1.17 (0.58–2.39)

Significant values are given in bold

\*Recessive model

HR hazard ratio, CI confidence interval, OR odds ratio

probably damaging functional effect was predicted by PolyPhen for two of the six missense SNPs.

### Signatures of selection

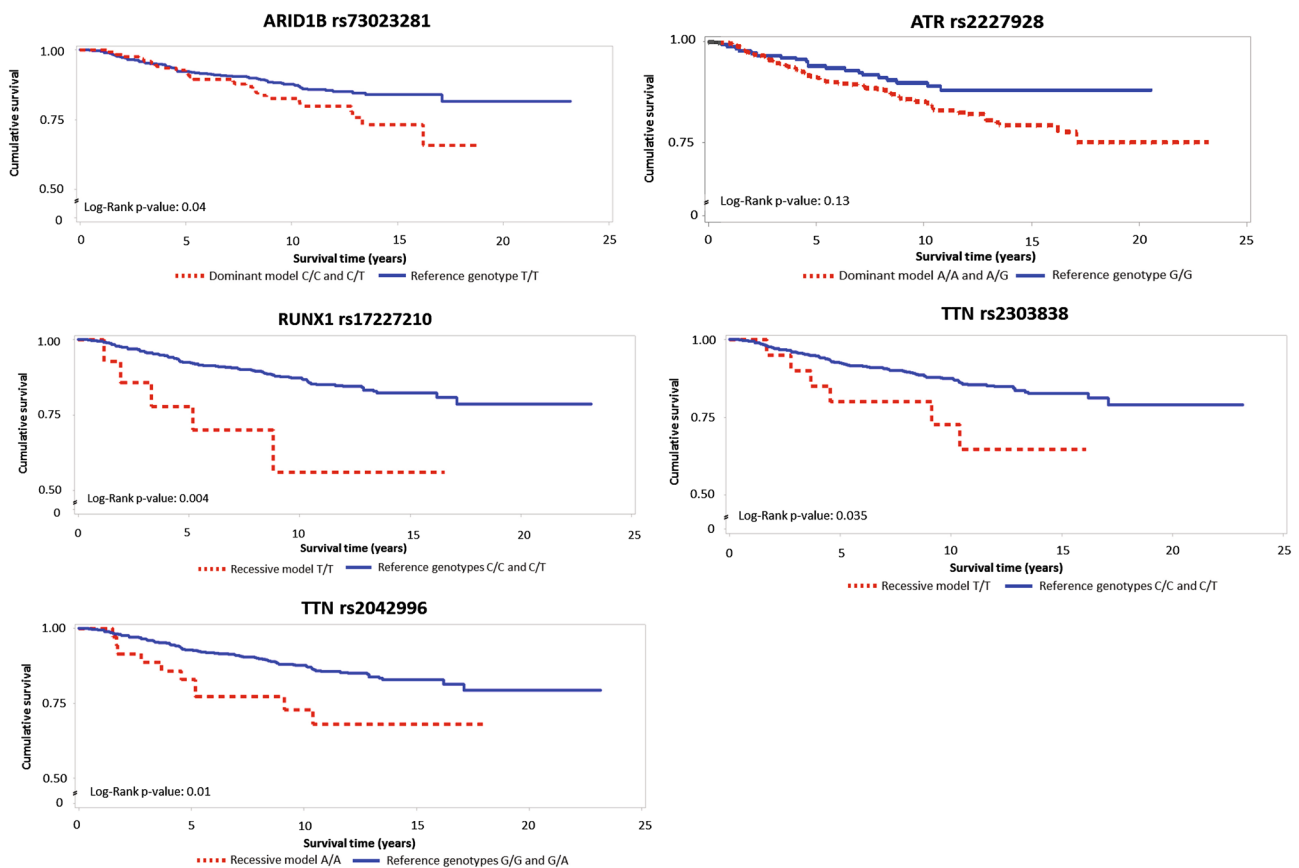
*TTN* was the most noticeable gene in this study. Four out of six SNPs are conserved and all six SNPs show either signatures of a strong ( $F_{ST} \geq 0.25$ ) or moderate ( $F_{ST} \geq 0.05$ ) genetic drift and an evolution under positive selection (Fay Wu's  $H \leq -40$ ) (Online Resource 5). In addition, the iHs score ( $-1.5$ ) for s10497520 is an indication for a selection based on the derived allele. All these values lead to a strong indication of a functionality of the genetic variation in *TTN*. Furthermore, rs8130963 (*RUNXI*) is significant among three values. Fay Wu's  $H = -56$  indicates an evolution under positive selection which is supported by  $F_{ST} = 0.346$  (CEU vs. YRI) and  $F_{ST} = 0.054$  (CEU vs. ASN).

### Discussion

Our aim was to identify cancer-related germline variants in novel genes classified as potential BC driver genes in four studies published in 2012. After genotyping and statistical analysis *ATR*, *RUNXI*, *TBX3* and *TTN* became the focus of the study. These potential driver genes carry germline variants with a statistical and functional impact and could, therefore, be potential predisposing genes with an influence on the development of BC. Whereas SNPs in *TBX3* were associated with less aggressive tumour markers, SNPs in *ATR*, *RUNXI* and *TTN* showed an opposite association. All associated or highly linked SNPs ( $r^2 \geq 0.80$ ) affected gene regulation according to the experimental ENCODE data. *RUNXI* and *TTN* showed also signatures of positive selection.

Ataxia Telangiectasia mutated and Rad3-related (*ATR*), together with Ataxia Telangiectasia mutated (*ATM*), plays an important role in cell cycle regulation by transducing DNA damaging signals [19]. Thus, *ATR* has been studied as a target for cancer therapy [20]. In the present study, two missense SNPs rs2227928 and rs2229032 were genotyped. These SNPs were also identified in Finnish and French breast/ovarian cancer families [21, 22], however, at a similar frequency as in the healthy control populations. Both SNPs are predicted to be benign/tolerated according to PolyPhen/SIFT. However, rs2227928 captures rs6768093 ( $r^2 = 0.99$ ), which is located at the active promoter of *ATR*. Rs2227928 was associated with high stage, large tumour size and positive regional lymph node metastases, consequently poor survival was observed. Several SNPs which are linked to rs2227928 with an  $r^2$  between 0.85 and 0.97 are located in the PLS1 (Plastin 1) gene. The encoded actin-binding protein has been found at high levels specifically





**Fig. 1** Kaplan–Meier plots of SNPs associated with breast cancer-specific survival

in small intestine [23]. An association with BC is so far unknown. The significant  $F_{ST}$  value (0.076 EUR vs. YRI) for rs2227928 could be an indication of genetic hitchhiking, as other values were not noticeable.

Runt-related transcription factor 1 (*RUNX1*) is a tumour suppressor which is highly expressed in breast epithelial cells [24]. Downregulation of *RUNX1* is part of a 17-gene signature that has been suggested to predict BC metastasis [25]. *RUNX1* may stimulate E-cadherin and inhibit epithelial-to-mesenchymal transition [26]. However, how *RUNX1* promotes BC development has to be clarified. In the present study, three out of the four genotyped SNPs were associated with high stage (rs8130963, rs7276777) or poor survival (rs17227210). All these SNPs are located in introns and were genotyped instead of potentially functional SNPs, rs72813661 in the active promoter, rs13051066 in the 3'-UTR and rs56045941 in the 5'-UTR, respectively. Rs8130963 shows a strong negative Fay Wu's  $H$  value of  $-56$  and a strong genetic differentiation between the European and African population ( $F_{ST}=0.346$ ), which is an indication for positive selection. Rs17227210 was associated with poor survival before (HR 3.66, 95% CI 1.48–9.08) and after adjustment for ER, PR, T, N and

grade (HR 5.07, 95% CI 1.15–22.47). All linked SNPs are located at a position of either a strong or a weak enhancer. One of these, rs17227231 ( $r^2=0.92$  to rs17227210), affects *GATA3* (*GATA* binding protein 3) binding. As *GATA3* has already been classified as a high confident cancer driver gene and as a possible marker for metastatic breast carcinomas [27, 28], the change of the *GATA*-binding site could explain the poor survival associated with rs17227210.

The T-box transcription factor 3 (*TBX3*) is expressed in mammary tissues and plays a context-dependent role in mammary gland development and tumorigenesis [29]. *TBX3* interacts with several major oncogenic pathways and is overexpressed in many tumours but most commonly in BC [30]. Recently, somatic mutations in *TBX3* have been classified as BC driver mutations [10–13, 31, 32]. Our study suggests an additional layer to the involvement of *TBX3* in the development of BC by showing that SNPs are associated with protective tumour traits. The two risk-associated SNPs rs2242442 and rs12366395 are located at the active promoter and have an effect on the transcription factor-binding site of *STAT* (signal transducer and activator of transcription). Mutations in *STAT* proteins, which lead to unregulated cell proliferation, have been found in

**Table 4** Associations of SNPs with BC-specific survival without and with adjustment for established clinical markers

Gene/SNP	Genotype	Without adjustment			Adjustment for T, N, grade			Adjustment for T, N, grade, ER, PR					
		No.	No. died (%)	HR (95% CI)	p value	No.	No. died (%)	HR (95% CI)	p value	No.	No. died (%)	HR (95% CI)	p value
<i>ARD1B</i> rs73013281	TT	629	83 (13.20)	1.00		509	70 (13.75)	1.00		299	36 (12.04)	1.00	
	CT	123	26 (21.14)	<b>1.64 (1.05–2.54)</b>	<b>0.03</b>	106	25 (23.58)	1.53 (0.96–2.41)	0.07	67	12 (17.91)	1.41 (0.73–2.73)	0.31
	CC	4	0 (0.00)	0 (0.00)	–	4	0 (0.00)	0 (0.00)	–	2	0 (0.00)	0 (0.00)	–
<i>ATR</i> rs2227928	CT+CC	127	26 (20.47)	<b>1.58 (1.02–2.45)</b>	<b>0.04</b>	110	25 (22.73)	1.48 (0.94–2.34)	0.09	69	12 (17.39)	1.37 (0.71–2.66)	0.35
	GG	189	20 (10.58)	1.00		150	17 (11.33)	1.00		90	7 (7.78)	1.00	
	AG	395	65 (16.46)	1.62 (0.98–2.67)	0.06	326	56 (17.18)	1.38 (0.80–2.37)	0.25	198	28 (14.14)	1.52 (0.66–3.49)	0.33
<i>RUNX1</i> rs17227210	AA	166	27 (16.27)	1.65 (0.92–2.94)	0.09	138	24 (17.39)	1.34 (0.72–2.50)	0.36	84	15 (17.86)	2.12 (0.85–5.27)	0.11
	AG+AA	561	92 (16.40)	<b>1.63 (1.00–2.64)</b>	<b>0.05</b>	464	80 (17.24)	1.36 (0.81–2.31)	0.25	282	43 (15.25)	1.68 (0.75–3.75)	0.21
	CC	555	75 (13.51)	1.00		452	63 (13.94)	1.00		273	33 (12.09)	1.00	
<i>TTN</i> rs2303838	CT	184	30 (16.30)	1.18 (0.77–1.81)	0.44	154	29 (18.83)	1.18 (0.76–1.84)	0.47	87	14 (16.09)	1.32 (0.68–2.53)	0.41
	TT	14	5 (35.71)	<b>3.66 (1.48–9.08)</b>	<b>0.01</b>	12	4 (33.33)	<b>3.36 (1.22–9.25)</b>	<b>0.02</b>	8	2 (25.00)	<b>5.07 (1.15–22.46)</b>	<b>0.03</b>
	*TT vs. CC+CT	14	5 (35.71)	<b>3.50 (1.42–8.61)</b>	<b>0.01</b>	12	4 (33.33)	<b>3.21 (1.17–8.77)</b>	<b>0.02</b>	8	2 (25.00)	<b>2.83 (1.15–6.92)</b>	<b>0.02</b>
<i>TTN</i> rs2042996	CC	528	73 (13.83)	1.00		438	62 (14.16)	1.00		255	26 (10.20)	1.00	
	CT	195	28 (14.36)	1.09 (0.71–1.69)	0.68	157	25 (15.92)	1.42 (0.89–2.28)	0.14	100	17 (17.00)	<b>2.79 (1.44–5.40)</b>	<b>0.002</b>
	TT	20	6 (30.00)	<b>2.42 (1.05–5.57)</b>	<b>0.04</b>	17	6 (35.29)	<b>2.93 (1.25–6.83)</b>	<b>0.01</b>	10	5 (50.00)	<b>9.11 (3.27–25.40)</b>	<b>2 × 10<sup>-5</sup></b>
<i>TTN</i> rs2042996	*TT vs. CC+CT	20	6 (30.00)	<b>2.36 (1.04–5.39)</b>	<b>0.04</b>	174	6 (35.29)	<b>2.66 (1.15–6.15)</b>	<b>0.02</b>	110	5 (50.00)	<b>6.17 (2.31–16.45)</b>	<b>2.7 × 10<sup>-4</sup></b>
	GG	468	64 (13.68)	1.00		390	56 (14.36)	1.00		225	22 (9.78)	1.00	
	AG	252	35 (13.89)	1.04 (0.69–1.58)	0.84	202	29 (14.36)	1.09 (0.69–1.71)	0.72	123	17 (13.82)	1.78 (0.92–3.44)	0.08
<i>TTN</i> rs2042996	AA	36	10 (27.78)	<b>2.31 (1.19–4.51)</b>	<b>0.01</b>	28	10 (35.71)	<b>2.86 (1.45–5.65)</b>	<b>0.002</b>	18	9 (50.00)	<b>10.52 (4.44–24.92)</b>	<b>9 × 10<sup>-8</sup></b>
	*AA vs. GG+AG	36	10 (27.78)	<b>2.28 (1.19–4.37)</b>	<b>0.01</b>	28	10 (35.71)	<b>2.78 (1.43–5.39)</b>	<b>0.002</b>	18	9 (50.00)	<b>8.10 (3.66–17.93)</b>	<b>2.5 × 10<sup>-7</sup></b>

Significant values are given in bold

Polymorphisms that showed significant differences in BC-specific survival in the unadjusted model were analysed further by adjusting the data for tumour size (T), regional lymph node metastases (N), histological grade and estrogen (ER) and progesterone receptor (PR) status; HR hazard ratio; CI confidence interval

\*Recessive model

**Table 5** Summary of the possible functional effects of genes and SNPs associated with BC risk, survival or clinical markers

Gene	Genotyped SNP	Association	Linked functional SNP ( $r^2$ )	Region/function	Possible functional effect	PolyPhen/SIFT	Protein binding	TFBS <sup>a</sup>	Chromatin structure	Histone modification	miRNA
ARID1B	rs73013281	Survival, stage		Promoter region	-		YES	YES	YES	YES	-
	rs2227928	Survival, T, N, stage	rs6768093 (0.99)	Missense	Benign/tolerated		NO	YES	NO	NO	-
MAP3K1	rs72758040	Risk	rs62355943 (0.99)	Promoter region	-		NO	YES	NO	YES	-
	rs702688	Risk	rs72757038 (0.97)	Promoter region	-		YES	NO	YES	YES	-
	rs11168827	Risk, PR, grade		Promoter region	-		NO	YES	YES	YES	-
MLL2	rs8130963	Stage		Promoter region	-		YES	YES	NO	N/A	-
	rs7276777	Stage	rs7281361 (0.95)	3'-UTR	-		NO	YES	NO	YES	NO
RUNX1	rs17227210	Risk, PR, grade		Intron	-		NO	YES	NO	N/A	-
	rs4685	Stage	rs13051066 (0.81)	Intron	-		NO	YES	NO	NO	-
SF3B1	rs12456284	Survival	rs56045941 (0.80)	Promoter region	-		YES	YES	YES	YES	-
	rs8853	Risk, ER, ER/PR, N	rs788023 (0.99)	Intron	-		NO	YES	NO	YES	-
SMAD4	rs1061651	Grade		3'-UTR	-		NO	NO	NO	YES	-
	rs2242442	Grade		3'-UTR	-		NO	YES	NO	YES	-
TBX3	rs12366395	Grade		3'-UTR	-		NO	YES	NO	YES	-
	rs12463674	Grade	rs1061657 (0.95)	3'-UTR	-		NO	YES	NO	YES	-
TTN	rs2303838	risk, T	rs3741698 (0.91)	3'-UTR	-		NO	YES	NO	YES	-
	rs2042996	risk		3'-UTR	-		NO	YES	NO	YES	-
TTN	rs2244492	PR		3'-UTR	-		NO	YES	NO	YES	-
	rs10497520	PR, ER/PR		3'-UTR	-		NO	YES	NO	YES	-
TTN	rs12465459	Risk		3'-UTR	-		NO	YES	NO	YES	-
	rs12465459	T		3'-UTR	-		NO	YES	NO	YES	-
TTN	rs2303838	Survival		Missense	Probably damaging/unknown		NO	YES	NO	N/A	-
	rs2042996	Survival, ER grade		Missense	Benign/unknown		NO	YES	NO	YES	-
TTN	rs2244492	PR, ER/PR		Missense	Benign/unknown		NO	YES	NO	N/A	-
	rs10497520	Risk		Missense	Unknown/tolerated		NO	YES	NO	N/A	-
TTN	rs12465459	Risk		Missense	possibly damaging/tolerated		NO	YES	NO	N/A	-
	rs12465459	T		Promoter region	-		NO	YES	NO	YES	-

Further knowledge about the consequences of SNPs on protein structure and function due to an amino acid change, their possible effects on protein-binding sites, transcription factor-binding sites (TFBSs), chromatin structure and histone modification and the influence of 3'-UTR (untranslated region) SNPs on micro-RNA binding, PolyPhen/SIFT, HaploReg, RegulomeDB, and microSNIPer were used for the predictions. All effects were proofed for data in mammary epithelial (tumour) cell lines (MCF7, T-47D, HMEC, MCF10A-ER-SRC)

T tumour size, N regional lymph node metastases, grade histological grade, ER estrogen receptor status, PR progesterone receptor status

<sup>a</sup>Effects on TFBS (transcription factor-binding site) were calculated through PWM (position weight matrices); N/A not available. No eQTLs were identified. Further information is available at Online Resource 4

myeloproliferative disorders. However, these aspects are not well studied in BC [33]. Gene expression of *TBX3* could be influenced by the SNPs rs8853, rs1061651 and two other 3'UTR SNPs linked to them due to their impact on miRNA-binding sites. An association of miR-1290, which binding is affected by rs3741698, a SNP linked to rs1061651 with  $r^2=0.91$ , and estrogen receptor-positive BC has been described [34]. Furthermore, *TBX3* overexpression has been observed in primary breast tumours and BC cell lines with higher expression in estrogen receptor-positive tumour cells [29]. However, other publications have described that estrogen-induced *TBX3* overexpression results in a pool of estrogen receptor-negative cancer stem-like cells [30].

*TTN* (Titin) has been intensively studied as a component of the muscle contractile machinery [35, 36]. However, *TTN* seems to play a role in non-muscle cells during chromosome condensation and chromosome segregation [37, 38]. Furthermore, the disease rhabdomyosarcoma is associated with Titin. Rhabdomyosarcoma can also affect the breast, although rarely [39]. Thus, *TTN* may play a role in oncogenesis, although the biological mechanisms need to be evaluated. In contrast, *TTN* is also described as a false-positive driver gene due to mutational heterogeneity which dominates over true driver events [35, 40]. In our study, six out of nine genotyped SNPs were associated with increased risk, aggressive tumour characteristics and/or poor survival. Three of these SNPs showed an association with negative hormone receptor status. Two of five missense SNPs (rs12463674 and rs10497520) are predicted to cause a probably or possibly damaging amino acid change by PolyPhen. Thus, these SNPs could have impact on the structure and function of the protein. Beside rs12463674 and rs10497520, *TTN* has a large number of other missense mutations, though phyloP estimates for *TTN* indicate high conservation. For all genotyped SNPs, the values of  $F_{ST}$ , Fay Wu's  $H$  and  $iHS$  indicated strong positive selection.

Although our study provides new knowledge about genes and mutations influencing BC risk, tumour characteristics and survival certain limitations have to be considered: first, a relatively small sample size and especially missing information on hormone receptor status decreased the power to detect associations with genotypes; second, the associations did not stay statistically significant after correction for multiple testing; third, the results need to be replicated in another population. The strengths of our study included a population-based design, with prospectively collected blood samples, long follow-up, and detailed clinical data. As several SNPs within one gene were associated with BC risk, tumour characteristics and/or survival, all of them with functional consequences and some even with signatures of positive selection, a true direct or indirect

association of the studied SNPs with BC development cannot be ruled out.

## Conclusion

Our study suggests that germline variants in driver genes of sporadic tumours can have an impact on BC risk, tumour characteristics and/or survival. Several SNPs in *ATR*, *RUNX1*, *TBX3* and *TTN* showed associations with BC development and progression. In silico analyses provided evidence of possible functional consequences for the associated SNPs. For *TTN* and *RUNX1*, strong signatures of positive selection gave further insights on the functionality of the SNPs. However, to verify the results on BC survival and the influence of *TTN*, further investigations are necessary.

**Acknowledgments** We thank Åsa Ågren (Department of Public Health and Clinical Medicine/Nutritional Research, Umeå University, Sweden) for her commitment in keeping track of the samples and data and the Northern Sweden Breast Cancer group for providing the clinical data.

**Compliance with ethical standards**

**Conflict of interest** None.

## References

1. Bray F, Ren JS, Masuyer E et al (2013) Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer* 1325:1133–1145
2. Ferlay JSI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2012) GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer-Base No. 11. Internet; Lyon, France: International Agency for Research on Cancer; 2013. <http://globocan.iarc.fr>. Accessed 05 Feb 2014
3. Frank C, Fallah M, Ji J et al (2014) The population impact of familial cancer, a major cause of cancer. *Int J Cancer* 1348:1899–1906
4. Olivier M, Hollstein MHainaut P (2010) TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol* 21:a001008
5. Sjoblom T, Jones S, Wood LD et al (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 3145797:268–274
6. Wood LD, Parsons DW, Jones S et al (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 3185853:1108–1113
7. Teschendorff A, ECaldas C (2009) The breast cancer somatic 'muta-ome': tackling the complexity. *Breast Cancer Res* 112:301
8. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature* 4587239:719–724
9. Tomasetti C, Marchionni L, Nowak MA et al (2015) Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc Natl Acad Sci USA* 1121:118–123

10. Banerji S, Cibulskis K, Rangel-Escareno C et al (2012) Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 4867403:405–409
11. Ellis MJ, Ding L, Shen D et al (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 4867403:353–360
12. Shah SP, Roth A, Goya R et al (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 4867403:395–399
13. Stephens PJ, Tarpey PS, Davies H et al (2012) The landscape of cancer genes and mutational processes in breast cancer. *Nature* 4867403:400–404
14. Kaaks R, Lundin E, Rinaldi S et al (2002) Prospective study of IGF-I, IGF-binding proteins, and breast cancer risk, in northern and southern Sweden. *Cancer Causes Control* 134:307–316
15. Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 813:559–575
16. McShane LM, Altman DG, Sauerbrei W et al (2005) Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 9716:1180–1184
17. Voight BF, Kudaravalli S, Wen X et al (2006) A map of recent positive selection in the human genome. *PLoS Biol* 43:e72
18. Oleksyk TK, Smith M, WO'Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond B Biol Sci* 3651537:185–205
19. Dai Y, Grant S (2010) New insights into checkpoint kinase 1 in the DNA damage response signaling network. *Clin Cancer Res* 162:376–383
20. Peasland A, Wang LZ, Rowling E et al (2011) Identification and evaluation of a potent novel ATR inhibitor, NU6027, in breast and ovarian cancer cell lines. *Br J Cancer* 1053:372–381
21. Heikkinen K, Mansikka V, Karppinen SM et al (2005) Mutation analysis of the ATR gene in breast and ovarian cancer families. *Breast Cancer Res* 74:R495–R501
22. Durocher F, Labrie Y, Soucy P et al (2006) Mutation analysis and characterization of ATR sequence variants in breast cancer cases from high-risk French Canadian breast/ovarian cancer families. *BMC Cancer* 6:230
23. Grimm-Gunter E M, Revenu C, Ramos S, et al., (2009) Plastin 1 binds to keratin and is required for terminal web assembly in the intestinal epithelium. *Mol Biol Cell*. 2010:2549–2562
24. Wang L, Brugge J, Janes KA (2011) Intersection of FOXO- and RUNX1-mediated gene expression programs in single breast epithelial cells during morphogenesis and tumor progression. *Proc Natl Acad Sci USA* 10840:E803–E812
25. Ramaswamy S, Ross KN, Lander ES et al (2003) A molecular signature of metastasis in primary solid tumors. *Nat Genet* 331:49–54
26. Ching N, OFrenkel B (2013) The RUNX family in breast cancer: relationships with estrogen signaling. *Oncogene* 3217:2121–2130
27. Shield PW, Papadimos D, Walsh MD, (2014) GATA3: a promising marker for metastatic breast carcinoma in serous effusion specimens. *Cancer Cytopathol*
28. Tamborero D, Gonzalez-Perez A, Perez-Llomas C et al (2013) Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep* 3:2650
29. Douglas N, CPapaoiannou VE (2013) The T-box transcription factors TBX2 and TBX3 in mammary gland development and breast cancer. *J Mammary Gland Biol Neoplasia* 182:143–147
30. Washkowitz AJ, Gavrilov S, Begum S et al (2012) Diverse functional networks of Tbx3 in development and disease. *Wiley Interdiscip Rev Syst Biol Med* 43:273–283
31. Kandoth C, McLellan MD, Vandin F et al (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 5027471:333–339
32. Lawrence MS, Stojanov P, Mermel CH et al (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 5057484:495–501
33. Hosford S, RMiller T W, (2014) Clinical potential of novel therapeutic targets in breast cancer: CDK4/6, Src, JAK/STAT, PARP, HDAC, and PI3K/AKT/mTOR pathways. *Pharmgenomics Pers Med* 7:203–215
34. Endo Y, Toyama T, Takahashi S et al (2013) miR-1290 and its potential targets are associated with characteristics of estrogen receptor alpha-positive breast cancer. *Endocr Relat Cancer* 201:91–102
35. Greenman C, Stephens P, Smith R et al (2007) Patterns of somatic mutation in human cancer genomes. *Nature* 4467132:153–158
36. Dowling JJ (2013) Titin and centronuclear myopathy: the tip of the iceberg for TTN-ic mutations? *Neurology* 8114:1189–1190
37. Mayans O, van der Ven PF, Wilm M et al (1998) Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* 3956705:863–869
38. Machado C, Andrew DJ (2000) Titin as a chromosomal protein. *Adv Exp Med Biol* 481:221–232 (**discussion 232–236**)
39. Bhosale SJ, Kshirsagar AY, Sulhyan SR et al (2013) Rhabdomyosarcoma of the breast—a rare malignancy. *Am J Case Rep* 14:250–252
40. Lawrence MS, Stojanov P, Polak P et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 4997457:214–218