Alternative splicing regulation by GWAS risk loci for breast cancer

Juliana Machado^{1,2,3}, Ramiro Magno^{1,2}, Joana M Xavier^{1,2,3}, and Ana-Teresa Maia^{1,2,3,*}

¹Centre for Biomedical Research (CBMR), University of Algarve, Campus of Gambelas, 8055-139 Faro, Portugal

²Algarve Biomedical Center, Campus Gambelas, Edifício 2. Ala Norte. 8005-139, Faro, Portugal

³Department of Biomedical Sciences and Medicine (DCBM), University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

*atmaia@ualg.pt

ABSTRACT

Recent genome-wide association studies (GWAS) have revealed the association of hundreds of single nucleotide polymorphisms (SNPs) with breast cancer (BC) risk, which mostly locate in non-coding regions, suggesting regulatory roles to the causal variants. Functional characterisation of GWAS loci has been biased towards the effect of regulatory SNPs on transcription factor binding. Here we set out to determine the extent of the contribution of breast cancer risk-associated SNPs to alternative splicing (AS).

We screened genome-wide significant ($P \leq 5 \times 10^{-8}$) BC risk SNPs for association with AS, using expression and genotype data from normal breast samples, from the GTEx project. We identified four splicing quantitative trait loci (sQTL). In locus 6p22.1, rs6456883 is a significant cis-sQTL for the expression of *ZNF311* gene isoforms. Three SNPs in locus 8p23.3, rs6682326, rs3008282 and rs2906324, were also identified as significant cis-sQTLs/svQTLs for the expression of *RPL23AP53* gene isoforms. In-silico functional analysis revealed that these variants can potentially alter enhancer splicing elements within the target genes.

Our work shows that BC risk-associated variants at two loci are associated with AS isoforms in normal breast tissue, thus demonstrating that AS plays an important role in breast cancer susceptibility. Furthermore, it supports that all cis-regulatory mechanisms should be considered in the functional characterisation of risk loci.

Introduction

Recent genome-wide association studies (GWAS) have revealed the association of hundreds of single nucleotide polymorphisms (SNPs) with breast cancer (BC) risk, increasing the total knowledge of the risk's genetic component to approximately 50%^{1,2}. However, GWASes do not identify directly the causative variants and are unable to pinpoint the underlying biological mechanisms. Interestingly, most risk-associated SNPs are located in non-coding regions, suggesting a regulatory role to these risk loci. Several studies have elucidated the biological mechanism of a small portion of risk loci (e.g.³⁻⁵), but have mainly focused on the role variants can have in regulating transcription via modulation of transcription factor binding to cis-regulatory elements (CREs). However, cis-regulation of gene expression can also be achieved by genetic variants altering binding of splicing factors⁶⁻⁸, including those associated with breast cancer risk^{9,10}, as well as miRNA binding^{11,12}. Our aim was to determine whether breast cancer risk-associated SNPs contribute to alternative splicing (AS) in normal breast tissue, the primary tissue of disease risk, and what the importance of this mechanism is to breast cancer risk.

Results

Splicing QTL analysis in breast cancer risk loci

Four hundred and thirty two BC GWAS SNPs with association p-value $\leq 5 \times 10^{-8}$, corresponding to 151 unique loci, were evaluated for their potential to modulate alternative splicing (Supplementary Table S1 indicates trait search criteria used). Upon identification of proxy SNPs, filtering of datasets and overlap of information, we obtained genotype data on 4,223 genetic variants and expression data on 120,812 transcripts, corresponding to 58,336 gene annotations. In total, sQTLseeker tested 50 gene-SNP pairs (Supplementary Table S2), and four significant sQTL associations were identified (FDR < 0.05) (Table 1 and Figure 1A and B).

rs6456883 was identified as a significant sQTL (q-value = 3.89×10^{-2}) for *ZNF311* (gene zinc finger protein 311 gene), localised in chromosome 6p22, and it localises to its third intron (Figure 1C). This variant is a proxy for the GWAS SNP rs9257408 (LD - $r^2 = 0.93$) associated with breast cancer risk (C risk allele, p-val = 5×10^{-8} , OR = 1.05 and 95%CI = [1.03-1.08])¹ (Supplementary Table S3). In our study, the T allele of rs6456883, is associated with a predominance of the isoform *ZNF311-203*, the only protein producing transcript of *ZNF311* (Figure 1A). Conversely, the C allele is associated with

predominance of the isoform *ZNF311-204*, an untranslated transcript characterised by the retention of an intron, resulting in a larger fourth exon. The same SNP is also an eQTL for the expression of *ZNF311* in testis, the tissue with higher expression of this gene, but not in breast tissue (data from¹³). We also found evidence of DAE for *ZNF311* in normal breast tissue (Supplementary Figure 1B), in two variants in LD with the sQTL variant (rs3129792 $r^2 = 0.74$ and rs3129795 $r^2 = 0.28$), further supporting the presence of cis-regulatory variants at this locus.

The other three significant sQTLs/svQTLs are rs66823261, rs3008282 and rs2906324 (q-values = 1.28×10^{-2} , 3.94×10^{-3} and 3.94×10^{-3} , respectively), associated with alternative splicing of the pseudogene *RPL23AP53* (ribosomal protein L23a pseudogene 53) (Figure 1B), located in chromosome 8p23, and widely expressed across tissues (data from¹³). These three SNPs are all located in the third intron of *RPL23AP53* (Figure 1D), with rs66823261 being the variant associated with risk to breast cancer in this locus (p-val = 3×10^{-9} for BC and 6×10^{-9} for ERnegBC, OR = 1.09 and 95%CI = [1.06-1.12])¹⁴ (Supplementary Table S3). The most predominant transcript *RPL23AP53-203*, and a shorter AS isoform containing only the first three exons (*RPL23AP53-202*), are both more expressed in the context of the minor allele of all three svQTLs. A minor AS isoform containing only a small portion of the 3'UTR of *RPL23AP53*, *RPL23AP53-201*, is the least expressed form, but more expressed in the context of the common allele. These three associations were significant for both sQTL and svQTL, meaning that there is heterogeneous variation between the levels of expression of all isoforms (heterogeneity of variation between the genotype groups). Due to this, we carried further analysis and found that rs3008282 and rs2906324 were significant eQTLs for each of the AS isoforms (Supplementary Figure 1A). We also found strong evidence of DAE for *RPL23AP53*, particularly at rs3008282 for which only one in 23 heterozygotes did not display a significant allelic expression difference (Supplementary Figure 1B). This result is a strong indicator that not only the expression of *RPL23AP53-203* is affected by cis-regulatory variants, but also that rs3008282 is a strong candidate causal variant of the DAE, as described previously¹⁵.

In silico Predictive Functional Analysis

In order to unveil the possible functional impact of the significant sQTLs we investigated modification of donor/acceptor splice sites and ESE sites (exon splicing enhancer). We found a putative stronger intronic ESE site for *ZNF311* in the presence of the T allele of rs6456883, as predicted by two algorithms reported by the Human Splicing Finder tool (Supplementary Table S4). This is corroborated by the prediction using the NetGene2 website, which uses artificial neural networks applied to the prediction of splice site locations in human pre-mRNA (Supplementary Table S5). The allelic variation for the prediction is small (-0.86% for HSF matrix score and 33% for NetGene2 confidence), but it could explain the also small variation in the ratio of isoform expression. Nevertheless, we did not find evidence supporting that the rs6456883 alleles alter the binding of RNA-binding proteins.

For 8p23 locus we found predictions for all three sQTLs/svQTLs to affect both the binding of RBPs, as well as, potential ESE sites. Specifically, all three sQTL variants are predicted to potentially affect the binding of RBPs, with rs66823261 having the strongest supporting evidence (iCLIP and eQTL data)¹⁶ (Supplementary Table S6). The allelic differential predictions of RBP binding for these sQTLs, which could explain the association with different isoforms, are presented in Supplementary Tables S4, S7 and S8. rs66823261 and rs2906324 have the strongest predictions for the creation of intronic ESE sites, corroborated by the prediction of creation of new binding sites for several RBPs. Although rs3008282 is not predicted to modify AS by the Human Splicing Finder tool, RBPs are predicted to bind differentially to its alleles, with preferential binding to the A allele.

Discussion

After more than a decade of GWAS efforts, the possible mechanisms by which cis-acting genetic variants regulate gene expression and contribute to disease susceptibility remain poorly understood. Most attempts at functionally characterising risk variants have focused on their potential to alter transcription factor binding. However, mRNA splicing is equally likely to be affected by genetic variants, whether by introducing changes in levels of expression and sequence of mRNA splicing-related genes or by modifying of RBP target sequences. The former has been assessed in liquid malignancies¹⁷, in lung cancer¹⁸, and even in breast cancer¹⁰. However, we believe that, in order to understand the biology of risk variants and how they increase risk during the lifetime of an individual, functional studies should be focused on the primary healthy tissue where disease develops. Thus, we carried an evaluation of the potential of GWAS risk variants to regulate AS in normal breast tissue.

Several computational tools are available that can identify cis-variants affecting alternative splicing¹⁹. In our study we used sQTLseeker²⁰, an R package which performs sQTL analysis by using the information of all alternatively spliced isoforms of a gene, and using their relative abundances as a vector, which is then used in a distance-based approach to test for association with genotype groups. Advantages of this package include the non-limitation of number of alternative splice isoforms that can be tested and the power to detect both cis- and trans- effects, as well as simple and complex alternative splicing events.

In our study, we identified four significant associations between BC risk-associated variants and differential transcript isoform expression levels and/or isoform relative amounts, from a total of 50 SNP-isoform pairs that were filtered for analysis

by sQTLseeker. Overall, this represents 8% of all 50 testable SNP-isoform pairs identified by sQTLseeker, which suggest that AS is an important contributor to risk for BC. Additionally, none of these loci have previously been functionally characterised.

The significant association at 6p22 indicated that the intronic variant rs6456883 is an sQTL for *ZNF311*, a gene encoding for a zinc-finger protein, predicted to be a transcription factor²¹, with broad expression across many tissues, including breast. This sQTL is in complete linkage disequilibrium with rs9257408, which is associated with risk to BC¹, and also with lung cancer and tuberculosis^{22,23}. Also, in the GTEx¹³ database, rs6456883 is indicated as an eQTL for *ZNF311*, albeit only in testis, the tissue where it is most expressed.

In our study we found that the minor allele of rs6456883, is associated with decreasing relative expression of isoform *ZNF311-203*, a protein coding isoform, and increasing relative expression of isoform *ZNF311-204*, an untranslated isoform with a retained intron. Intron retention is a known regulatory mechanism^{24,25}, which works in the way of decreasing the major isoform's expression, by using the transcription machinery to produce a transcript which will not be translated. The in silico analysis suggests that rs6456883 contributes to a higher intron retention of *ZNF311* by weakening an intronic ESE. Altogether, our data suggests that lower expression of the main coding isoform might be the mechanism underlying risk to BC associated with this locus, but as the function of *ZNF311* remains unknown, we cannot predict more about its involvement in BC risk.

In locus 8p23, we detected significant svQTLs for three SNPs (rs66823261, rs3008282, rs2906324) associated with varying relative amounts of *RPL23AP53* AS isoforms. The *RPL23AP53* gene, is a processed pseudogene with unknown function, which is moderately expressed across all tissues (GTEx data¹³), and has three AS isoforms, with a predominant larger isoform (*RPL23AP53-203*). The locus has been associated with risk to breast cancer, in particular with risk to Estrogen Receptor Negative BC¹⁴. In fact, one of the identified svQTL signals corresponds to the GWAS SNP rs66823261. The other two SNPs, rs3008282 and rs2906324, are in almost complete LD with each other ($r^2 = 0.99$), and in strong LD with the GWAS SNP ($r^2 = 0.82$ and 0.83 , respectively). All three svQTLs are located in the second intron of *RPL23AP53*, were characterised by increasing relative abundance of the main isoform associated with the minor alleles and have strong evidence of affecting the binding of RBPs. No sQTL is indicated in the GTEx database for *RPL23AP53*, but all three svQTLs we found have been identified as eQTLs for *RPL23AP53* in other tissues, but not in breast. Here, we found evidence for rs66823261 being an eQTL for *RPL23AP53* all three isoforms, and DAE data provides further supports the functional role of rs3008282 as a cis-regulatory SNP. Taken together, this data suggests that the risk for BC in this locus may be due to higher expression of the *RPL23AP53-203* isoform, whose expression is inversely correlated with that of *RPL23AP53-201*, the smallest isoform, consisting of a single exon. The higher expression of the pseudogene *RPL23AP53* could be working as a miRNA decoy protecting the parent gene Ribosomal Protein L23a from silencing, via competition for miRNA binding²⁶, but the link to breast cancer increased risk needs further investigation. To our knowledge, this is the first study to systematically mine GWAS-significant SNPs for breast cancer to identify sQTL signals in normal breast tissue. We found that two risk loci, 6p22 and 8p23, may harbour causal variants that affect alternative splicing of two genes *ZNF311* and *RPL23AP53*, respectively, by altering the binding of RBPs and ESEs. This is the first functional analysis of these risk loci, with evidence supporting that the identified variants may be the causal of risk. Overall, our results demonstrate that alternative splicing is a mechanism by which risk-associated genetic variants, identified via GWAS, may be contributing to risk to breast cancer. Moreover, our work adds support to unbiased study approaches for complete unveiling of the biological mechanism underlying risk to breast cancer in particular, and complex diseases in general.

Methods

Data Sources

Breast cancer risk variants were retrieved, on 2018-06-12, from the GWAS Catalog website²⁷, using a p-value cut-off significance of 5×10^{-8} , and a list of related disease/trait with BC (Supplementary Table S1). Gene transcripts' expression and genotyping data was obtained from GTEx project (dbGaP accession number: phs000424.v7.p2)¹³ on 2018-03-14. Data was retrieved for a total of 103 female breast samples. All data throughout the analysis was converted to the genome assembly GRCh38.

Linkage disequilibrium (LD) analysis and Annotation

Variants in LD with the breast cancer GWAS SNPs were retrieved with the R package proxysnps (<https://github.com/slowkow/proxysnps>) using CEU population genotype data from the third phase of the 1000 Genomes Project²⁸. Threshold parameters were set as $r^2 \geq 0.8$ and distance limit of ± 500 kb on either side of the queried SNP.

Splicing QTL (sQTL) Analysis

Splicing QTL (sQTL) analysis was carried with the R package sQTLseeker²⁰, which filters data for a threshold of isoforms' expression level of 0.01, threshold of gene expression of 0.01 and threshold of dispersion of 0.1. The authors differentiate between an sQTL, a genetic variant associated with changes in the splicing isoform ratios of a gene, and a splicing variation QTL (svQTL), a genetic variant associated with changes in the level of expression of a gene's splicing isoforms, but not

their relative amounts. Analysis was carried for both sQTLs and svQTLs, and p-values were corrected for multiple testing (FDR<5%).

Expression QTL (eQTL) Analysis

The same data was used to perform eQTL analyses for the identified sQTL variants in the locus 8p23.3, as it was the only locus with evidence of eQTL in any tissue at the GTEx website. Analysis of variance (ANOVA) test was fitted, and p-values were corrected for multiple testing by using the Bonferroni method.

Differential Allelic Expression (DAE) Analysis

Processed allelic expression and genotype data from normal breast tissue, described previously²⁹, was used to determine differential allelic expression ratios (DAE) as $\log_2((\text{RNA allele A}/\text{RNA allele B})/(\text{DNA allele A}/\text{DNA allele B}))$. Analysis was carried in R using several Bioconductor packages³⁰ and custom functions³¹. DAE was inferred when the absolute value of DAE was ≥ 0.58 (1.5 fold or greater difference between alleles).

In silico Predictive Functional Analysis

In silico prediction of the effect of significant sQTL variants on the binding of RNA-binding proteins was carried using the RBP-Var database (RBP-VAR, <http://www.rbp-var.biols.ac.cn/>)¹⁶ and the RBPmap (<http://rbpmap.technion.ac.il/>)³² online tools. For splice variant and exonic splice enhancer analysis we used Human Splicing Finder (<http://www.umd.be/HSF3/index.html>)³³ and NetGene2 Server (<http://www.cbs.dtu.dk/services/NetGene2/>)³⁴ tools. In particular, we queried up to 200bp of sequence, centred on the sQTL variants, and searched for differential allelic predictions.

References

1. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 1–24 (2017).
2. Ghoussaini, M., Pharoah, P. D. P. & Easton, D. F. Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? *The Am. J. Pathol.* **183**, 1038–1051 (2013).
3. Meyer, K. B. *et al.* Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol.* **6**, e108 (2008).
4. Glubb, D. M. *et al.* Fine-Scale Mapping of the 5q11.2 Breast Cancer Locus Reveals at Least Three Independent Risk Variants Regulating MAP3K1. *Am. J. Hum. Genet.* **96**, 5–20 (2015).
5. French, J. D. *et al.* Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am. J. Hum. Genet.* **92**, 489–503 (2013).
6. Kwan, T. *et al.* Tissue Effect on Genetic Control of Transcript Isoform Variation. *PLoS Genet.* **5**, e1000608 (2009).
7. Zhang, X. *et al.* Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nat. Genet.* **47**, 345–352 (2015).
8. Matsumoto, N., Takata, A. & Kato, T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* **8**, 1–11 (2017).
9. Dunning, A. M. *et al.* Association of ESR1 gene tagging SNPs with breast cancer risk. *Hum. Mol. Genet.* **18**, 1131–1139 (2009).
10. Caswell, J. L. *et al.* Multiple breast cancer risk variants are associated with differential transcript isoform expression in tumors. *Hum. Mol. Genet.* **24**, 7421–7431 (2015).
11. Nicoloso, M. S. *et al.* Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res.* **70**, 2789–2798 (2010).
12. Brewster, B. L. *et al.* Identification of fifteen novel germline variants in the BRCA1 3'UTR reveals a variant in a breast cancer case that introduces a functional miR-103target site. *Hum. Mutat.* **33**, 1665–1675 (2012).
13. The GTEx Consortium *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
14. Milne, R. L. *et al.* Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat. Genet.* **49**, 1–15 (2017).
15. Xiao, R. & Scott, L. J. Detection of cis-acting regulatory SNPs using allelic expression data. *Genet. Epidemiol.* **35**, 515–525 (2011).

16. Mao, F. *et al.* RBP-Var: a database of functional variants involved in regulation mediated by RNA-binding proteins. *Nucleic Acids Res.* **44**, D154–D163 (2016).
17. Je, E. M., Yoo, N. J., Kim, Y. J., Kim, M. S. & Lee, S. H. Mutational analysis of splicing machinery genes SF3B1, U2AF1 and SRSF2 in myelodysplasia and other common tumors. *Int. J. Cancer* **133**, 260–265 (2013).
18. Pan, Y. *et al.* Associations between genetic variants in mRNA splicing-related genes and risk of lung cancer: a pathway-based analysis from published GWASs. *Sci. Reports* **7**, 1–8 (2017).
19. Park, E., Pan, Z., Zhang, Z., Lin, L. & Xing, Y. The Expanding Landscape of Alternative Splicing Variation in Human Populations. *The Am. J. Hum. Genet.* **102**, 11–26 (2018).
20. Monlong, J., Calvo, M., Ferreira, P. G. & Guigo, R. Identification of genetic variants associated with alternative splicing using sQTLseeker. *Nat. Commun.* **5**, 4698–12 (2014).
21. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
22. McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* **49**, 1126 EP – (2017).
23. Tian, C. *et al.* Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat. Commun.* **8**, 599 (2017).
24. Jacob, A. G. & Smith, C. W. J. Intron retention as a component of regulated gene expression programs. *Hum. Genet.* **136**, 1043–1057 (2017).
25. Schmitz, U. *et al.* Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol.* **18**, 216 (2017).
26. Hu, X., Yang, L. & Mo, Y.-Y. Role of Pseudogenes in Tumorigenesis. *Cancers* **10**, 256–14 (2018).
27. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
28. Consortium, T. . G. P. *et al.* A global reference for human genetic variation. *Nature* **526**, 68 EP – (2015).
29. Liu, R. *et al.* Allele-specific expression analysis methods for high-density SNP microarray data. *Bioinformatics* **28**, 1102–1108 (2012).
30. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
31. Xavier, J. *et al.* Integrative differential allelic expression analysis efficiently reveals the biology underlying risk to breast cancer. *Mol. Cancer Res.* **14**, A31–A31 (2016).
32. Paz, I., Kosti, I., Ares Jr, M., Cline, M. & Mandel-Gutfreund, Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* **42**, W361–W367 (2014).
33. Desmet, F. O., Hamroun, D., acids, M. L. N. & 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37** (2009).
34. Brunak, S., Engelbrecht, J. & Knudsen, S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J. Mol. Biol.* **220**, 49–65 (1991).

Acknowledgements

Firstly, the authors would like to acknowledge the contributions made by all specimen donors who allow databases to be created and used to the good of medical care and science. This work was supported by national Portuguese funding through FCT-Fundação para a Ciência e a Tecnologia, UID/BIM/04773/2013-CBMR, SFRH/BPD/99502/2014 (JMX), and ALG-01-0145-FEDER-31477, which was co-funded by CRESC ALGARVE 2020. The authors would also like to thank the members of the Functional Genomics of Cancer group at CBMR for helpful discussions, Mr V. Morais at UAIC for administrative support, and Serv. de Informática, both at UAlg.

Author contributions statement

J.M., R.M. and J.M.X. conducted bioinformatic analysis, A.T.M. conducted functional analysis and drafted the manuscript, R.M. and A.T.M. conceived and supervised the work. All authors reviewed the manuscript.

Additional information

Accession codes Data used is available from GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) and GTEx Project (<https://gtexportal.org/home/>) databases (dbGaP accession number: phs000424.v7.p2).

Competing interests The authors declare no competing interests.

The corresponding author is [Ana-Teresa Maia](#).

SNP	sQTL		svQTL		qvalue	Most varying transcripts
	F	pvalue	F	pvalue		
rs6456883	6.13	2.68×10^{-3}	0.94	0.39	3.89×10^{-2}	ZNF311-204 ZNF311-203
rs66823261	6.00	6.66×10^{-4}	4.91	8.61×10^{-3}	1.28×10^{-2}	
rs3008282	7.32	1.36×10^{-4}	4.63	1.10×10^{-2}	3.94×10^{-3}	RPL23AP53-201 RPL23AP53-203
rs2906324	7.32	1.36×10^{-4}	4.63	1.10×10^{-2}	3.94×10^{-3}	

Table 1. Significant sQTL associations for BC risk loci. Legend: F values and p-values, for both sQTL and svQTL, are indicated as obtained from sQTLseeker, and qvalues as obtained after multiple testing correction (5%FDR).

Figure 1. Breast cancer risk loci associated with alternative splicing. **A** and **B** - sQTLseeker analysis for the risk variants in loci 6p22 (**A**) and 8p23 (**B**). Splicing isoforms for *ZNF311* and *RPL23AP53* are coloured differently as indicated below the graphs; y axis indicates the relative expression of the isoforms, i.e. the isoform specific RNA-seq read counts divided by the total read counts for the gene. Each panel represents a genotype group, as indicated in the grey bars above (unk - samples with unknown genotype). Each box-and-whiskers plot describes the distribution of expression values. The q-value of association is shown for each set. **C** and **D** - Schematic representation of breast cancer risk loci associated with alternative splicing (**C** - locus 6p22 and **D** - locus 8p23) with GWAS Catalog reported risk-associated variants displayed in green and identified sQTLs from current study displayed in black. Genes are displayed according to GENCODE 29 or 27, respectively, and colour-coded as follows: red for Ensemble protein coding genes; yellow for merged Ensembl/Havana protein coding genes; grey for non coding pseudogenes; and blue for non-coding processed transcripts.