



Kazmi, N., Robinson, T., Zheng, J., Kar, S., Martin, R. M., & Ridley, A. J. (2022). Rho GTPase gene expression and breast cancer risk: A Mendelian randomization analysis. *Scientific Reports*, 12(1), [1463 ]. <https://doi.org/10.1038/s41598-022-05549-5>

Publisher's PDF, also known as Version of record

Link to published version (if available):  
[10.1038/s41598-022-05549-5](https://doi.org/10.1038/s41598-022-05549-5)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Nature Research at <https://doi.org/10.1038/s41598-022-05549-5>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



OPEN

## Rho GTPase gene expression and breast cancer risk: a Mendelian randomization analysis

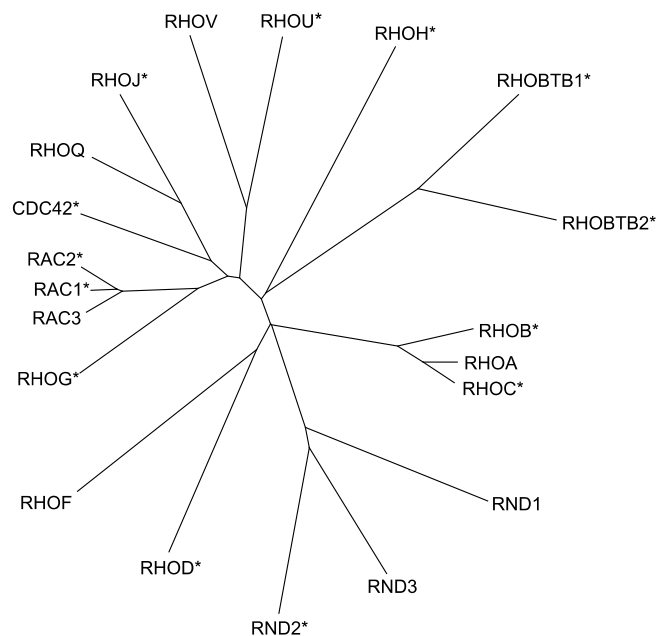
Nabila Kazmi<sup>1,2,5</sup>✉, Tim Robinson<sup>2,5</sup>✉, Jie Zheng<sup>1,2</sup>, Siddhartha Kar<sup>1,2</sup>, Richard M. Martin<sup>1,2,3</sup> & Anne J. Ridley<sup>4</sup>

The Rho GTPase family consists of 20 genes encoding intracellular signalling proteins that influence cytoskeletal dynamics, cell migration and cell cycle progression. They are implicated in breast cancer progression but their role in breast cancer aetiology is unknown. As aberrant Rho GTPase activity could be associated with breast cancer, we aimed to determine the potential for a causal role of Rho GTPase gene expression in breast cancer risk, using two-sample Mendelian randomization (MR). MR was undertaken in 122,977 breast cancer cases and 105,974 controls, including 69,501 estrogen receptor positive (ER+) cases and 105,974 controls, and 21,468 ER negative (ER-) cases and 105,974 controls. Single nucleotide polymorphisms (SNPs) underlying expression quantitative trait loci (eQTLs) obtained from normal breast tissue, breast cancer tissue and blood were used as genetic instruments for Rho GTPase expression. As a sensitivity analysis, we undertook co-localisation to examine whether findings reflected shared causal variants or genomic confounding. We identified genetic instruments for 14 of the 20 human Rho GTPases. Using eQTLs obtained from normal breast tissue and normal blood, we identified evidence of a causal role of *RHOD* in overall and ER+ breast cancers (overall breast cancer: odds ratio (OR) per standard deviation (SD) increase in expression level 1.06; (95% confidence interval (CI) 1.03, 1.09;  $P = 5.65 \times 10^{-5}$ ) and OR 1.22 (95% CI 1.11, 1.35;  $P = 5.22 \times 10^{-5}$ ) in normal breast tissue and blood respectively). There was a consistent direction of association for ER- breast cancer, although the effect-estimate was imprecisely estimated. Using eQTLs from breast cancer tissue and normal blood there was some evidence that *CDC42* was negatively associated with overall and ER+ breast cancer risk. The evidence from colocalization analyses strongly supported our MR results particularly for *RHOD*. Our study suggests a potential causal role of increased *RHOD* gene expression, and, although the evidence is weaker, a potential protective role for *CDC42* gene expression, in overall and ER+ breast cancers. These findings warrant validation in independent samples and further biological investigation to assess whether they may be suitable targets for drug targeting.

The Rho family of GTPases are key molecular regulators of actin and microtubule cytoskeletal dynamics that influence oncogenic processes such as cellular adhesion, migration, survival and cell cycle progression<sup>1</sup>. There are 20 Rho family genes in humans<sup>2</sup> (Fig. 1) and several Rho GTPases, and their associated signalling pathways, have been implicated in biological processes involved in cancer initiation and progression<sup>3,4</sup>. For example, increased expression of RhoA results in malignant transformation of mouse fibroblasts, stimulating carcinogenesis in a mouse model, and loss of function of RhoB increases chemically-induced carcinogenesis in an in vivo skin cancer model<sup>5,6</sup>. In breast cancer, multiple in vitro and in vivo experimental studies, across all breast cancer subtypes, have implicated a role for aberrant Rho GTPase activity in aggressive biological phenotypes<sup>1,3,4,7</sup>. However, a causal role of aberrant Rho GTPase signalling in the risk of developing breast cancer in humans is less clear.

Mendelian randomization (MR) uses germline genetic variants as instruments (“proxies”) to generate evidence for an association of potentially modifiable extrinsic risk factors and intrinsic metabolic processes on disease outcomes<sup>8,9</sup>. The aim of using germline genetic instruments is to minimise confounding and reverse

<sup>1</sup>MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK. <sup>2</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK. <sup>3</sup>National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals NHS Trust and University of Bristol, Bristol, UK. <sup>4</sup>School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK. <sup>5</sup>These authors contributed equally: Nabila Kazmi and Tim Robinson. ✉email: nabila.kazmi@bristol.ac.uk; tim.robinson@bristol.ac.uk



**Figure 1.** Twenty human Rho GTPase family members. A ClustlW alignment using the amino acid sequences of the 20 human Rho GTPases was used to generate the phylogenetic tree. \*, 14 genes able to be analysed by MR (see Table 1).

causation, as germline genotype is assigned at random at conception and fixed thereafter. These properties can make MR-derived effect-estimates independent of confounding by future lifestyle or environmental factors, and less likely to be affected by reverse causation, provided the following three assumptions are met: (i) robust association of the genetic instrument with the exposure of interest; (ii) no confounding of the instrument-outcome relationship; and (iii) lack of an alternative pathway through which an instrument influences the outcome except through the exposure<sup>8</sup>. The feasibility, precision and statistical power of MR analysis can be increased by using a “two-sample MR” framework in which summary genetic association data from independent samples representing firstly, genetic variant-exposure and secondly, genetic variant-outcome associations are combined in order to assess causal effects<sup>10</sup>.

The aim of our study was to assess whether a potential association exists between the expression of genes encoding Rho GTPases and risk of overall, estrogen receptor-positive (ER+) and estrogen receptor-negative (ER-) breast cancer.

## Methods

**Study population.** Summary data were obtained from a genome-wide association study (GWAS) of 122,977 overall breast cancer cases (including 69,501 ER+ and 21,468 ER- breast cancer cases) and 105,974 controls of European ancestry from the Breast Cancer Association Consortium (BCAC)<sup>11</sup>. All participating studies in BCAC were approved by their appropriate ethics review board and all subjects provided informed consent and all the methods were performed in accordance with the relevant guidelines and regulations. Full details of the 78 studies that contributed data to this BCAC analysis and their ethical approval committees are listed in<sup>11</sup>.

In studies that participated in BCAC, the ER status provided by each individual study was usually from either; (1) clinical records, (2) immunohistochemistry (IHC) whole sections, (3) IHC stain of Tissue Micro Arrays (TMAs), (4) derived from Ariol data of TMAs (a high throughput automated imaging system) (5) derived from semi-quantitative IHC stain data (not defined whether from whole sections or TMAs), and (6) other (not defined). Sometimes, the source is unknown.

A diagrammatic representation of the included datasets and the methodology is shown as a flow diagram (Fig. 2).

**Instrument construction.** To generate genetic instruments to proxy for Rho GTPase gene expression, we performed a multi-step approach. First, single nucleotide polymorphisms (SNPs) marking expression quantitative trait loci (eQTLs) underlying expression of the genes encoding 20 Rho GTPases were obtained from normal breast tissue, breast cancer tissue and blood from patients without breast cancer. We obtained normal breast tissue specific eQTLs by searching for the expression of each gene in the Genotype-Tissue Expression (GTEx) project (v8) (<https://gtexportal.org/home/>)<sup>12</sup> and selected the top SNP associated with the gene expression (defined by the smallest *P*-value). Second, we extracted eQTLs (the top SNP; smallest *P*-value) from the eQTLGen consortium (<https://www.eqtlgen.org/>) which has performed cis- and trans-eQTL analysis in blood from 31,684 individuals<sup>13</sup> of largely European ancestry. The consortium defined the cis-eQTLs as every SNP-gene combination with a distance < 1 Mb from the center of the gene and were tested in at least 2 cohorts. Third,

we obtained breast cancer tissue eQTLs from The Cancer Genome Atlas (TCGA) ([https://albertlab.shinyapps.io/tcga\\_eqtl/](https://albertlab.shinyapps.io/tcga_eqtl/)) which has systematically performed eQTL analyses across 24 human cancer types<sup>14</sup>. Finally, we selected cis-SNPs (within 1 Mb of the target gene) that were associated with gene expression level at a  $P$ -value threshold of  $P < 5 \times 10^{-08}$  based on summary data available on the three platforms. To retain only independent SNPs, we used linkage disequilibrium [LD] clumping with a threshold of  $r^2 \leq 0.01$  based on the 1000 Genomes European ancestry reference panel data<sup>15</sup>.

In sensitivity analyses, *TP53* gene expression was used as a control in this study: the genetic instrument to proxy this (rs78378222; a germline variant in *TP53* with ability to influence p53 activity)<sup>16,17</sup> was only available in the eQTLGen consortium dataset.

**Two-sample MR analysis.** We extracted the following information for each selected eQTL—effect allele, other allele, beta coefficient and standard error—and calculated the proportion of variance in gene expression explained by the SNP. We calculated  $R^2$  and  $F$ -statistics to assess for both the strength of the genetic instruments and for weak instrument bias using previously reported methods<sup>18</sup>. Exposure and outcome data were harmonised such that the effect of each SNP on the outcome and exposure was relative to the same allele<sup>19</sup>.

For our primary analyses using the top eQTL, we used the Wald ratio method, equivalent to  $\beta_{YG}/\beta_{XG}$  (where  $Y$  = outcome [overall, ER+, and ER– breast cancer],  $G$  = germline genetic variant, and  $X$  = Rho GTPase gene expression). In secondary analyses, when the genetic instrument consisted of multiple SNPs (‘a multi-allelic instrument’), we used the inverse-variance weighted (IVW) method, which performs an inverse variance weighted meta-analysis of each Wald ratio for each SNP<sup>20</sup>. By default our inverse-variance weighted (IVW) Mendelian randomisation analyses were based on random effects models. If under dispersion in these IVW models was detected, i.e., if the residual standard error of the IVW model was less than 1, the standard errors of the regression coefficients were corrected by dividing by 1 rather than by the residual standard error<sup>21</sup>.

We used summary genetic association data (beta coefficients and standard errors) and conducted colocalisation analysis<sup>22</sup> to investigate the probability that the genetic associations with both gene expression level and risk of breast cancer shared the same underlying causal variants. Here, we used the SNPs that were located within 1 Mb windows of the eQTL studied in the MR analysis.

For multi-allelic instruments, it was possible to perform sensitivity analyses to assess whether there was any evidence of violations of the MR assumption of no pleiotropic effects. For an instrument made up of  $\geq 2$  independent SNPs, Cochran’s  $Q$  was computed to quantify heterogeneity across the individual causal effects of SNPs and for an instrument with  $\geq 3$  independent SNPs, weighted median<sup>23</sup>, weighted mode<sup>24</sup> and MR-Egger regression<sup>25</sup> methods were applied. Violations of the MR ‘no horizontal pleiotropy’ assumption (where a locus influences the outcome through independent pathways separate to the exposure)<sup>26</sup> were assessed by visual inspection of plots constituting a multi-allelic instrument (funnel<sup>27</sup>, forest, scatter and leave-one-out<sup>19</sup>).

To account for multiple testing, Bonferroni corrections were used to establish  $P$ -value thresholds for strong evidence ( $P < 0.004$ ; alpha of 0.05/14 genes) and suggestive evidence ( $0.004 < P < 0.05$ ) of a causal effect.

All analyses were carried out using the TwoSampleMR and MRInstruments R packages, curated by the latest version of MR-Base (0.5.4)<sup>19</sup>, <http://www.mrbase.org>. The study was in accordance with relevant guidelines and regulations.

## Results

Of the 20 genes of the Rho GTPase family, 14 could be analysed using MR as they had at least one strongly associated SNP in at least one of the three databases that were searched (Fig. 1). Of these 14 genes, two showed evidence of an association of an eQTL with breast cancer risk (Table 2 and Fig. 3).

Using eQTLs obtained from both normal breast tissue and blood, increased expression of *ras homolog family member D (RHOD)* was positively associated with overall and ER+ breast cancer risk (Overall breast cancer: Odds ratio (OR) per standard deviation (SD) increase in expression level 1.06; (95% confidence interval (CI) 1.03, 1.09;  $P = 5.65 \times 10^{-5}$ ) and OR 1.22 (95% CI 1.11, 1.35;  $P = 5.22 \times 10^{-5}$ ) in normal breast tissue and blood respectively. ER+ breast cancer: OR per SD increase in expression level was 1.08 (95% CI 1.05, 1.12;  $P = 2.29 \times 10^{-5}$ ) and 1.32 (95% CI 1.18, 1.49;  $P = 2.74 \times 10^{-6}$ ) in normal breast tissue and blood respectively).

The direction of association was consistent for ER– breast cancer but the evidence of association was suggestive (OR per SD increase in expression level was 1.06 (95% CI 1.01, 1.12;  $P = 0.03$ ) and OR per SD increase in expression level 1.19 (95% CI 0.99, 1.42;  $P = 0.06$ ) in normal breast tissue and blood respectively). As we obtained only one SNP for the eQTL for *RHOD* from both breast tissue and blood, and none from breast cancer tissue, we were unable to perform multiple SNP sensitivity analyses.

There was some suggestive evidence that increased expression of *cell division cycle 42 (CDC42)* was inversely associated with overall and ER+ breast cancer risk. (Overall breast cancer: observed OR per SD increase in expression level 0.91 (95% CI 0.84, 0.98;  $P = 0.02$ ) and OR per SD increase in expression level 0.96 (95% CI 0.93, 1.00;  $P = 0.03$ ) using eQTLs obtained from breast cancer tissue and blood in general population respectively. ER+ breast cancer: OR per SD increase in expression level 0.90 (95% CI 0.82, 0.99;  $P = 0.03$ ) and 0.95 (95% CI 0.91, 0.98;  $P = 0.005$ ) using eQTLs obtained from breast cancer tissue and blood, respectively). The direction of association was consistent for ER– breast cancer using eQTL from breast cancer but the evidence of association was weak (OR per SD increase in expression level 0.88 (95% CI 0.76, 1.02;  $P = 0.09$ )). Using an eQTL from blood, the evidence of association was inconclusive (OR per SD increase in expression level 1.01 (95% CI 0.95, 1.07;  $P = 0.78$ )).

For expression level of each gene, we calculated the power a priori, to detect an odds ratio (OR) of 1.2 (or conversely a protective OR of 0.80) given an alpha level of 0.05, the variance explained in the expression level of the gene by the SNP instrument and the sample size of the outcome dataset against overall, ER+ and ER– breast cancer risk, as described previously<sup>18</sup>. The power to detect the odds ratio of 1.2 (or equally 0.80) was  $> 99\%$  for

Gene	Resource	Tissue	R <sup>2</sup>	F-statistic	Power (overall BC)	Power (ER+ BC)	Power (ER– BC)
<i>RHOD</i>	GTE <sub>x</sub>	Breast mammary	27.28	93.42	>99	>99	>99
<i>RHOD</i>	eQTLGen	Blood	0.35	107.03	>99	59.63	30.06
<i>CDC42</i>	eQTLGen	Blood	3.41	1118.52	>99	>99	>99
<i>CDC42</i>	TCGA	Breast cancer	0.31	14.97	67.57	54.56	27.19
<i>RHOBTB1</i>	GTE <sub>x</sub>	Breast mammary	20.76	65.24	>99	>99	>99
<i>RHOBTB1</i>	eQTLGen	Blood	1.85	582.89	>99	>99	91.20
<i>RHOBTB2</i>	eQTLGen	Blood	0.49	154.52	86.01	74.25	39.85
<i>RHOU</i>	eQTLGen	Blood	2.95	803.91	>99	>99	98.70
<i>RHOU</i>	TCGA	Breast cancer	0.34	16.00	70.45	57.32	28.72
<i>RHOQ</i>	eQTLGen	Blood	1.07	257.98	>99	97.16	71.24
<i>RHOQ</i>	TCGA	Breast cancer	0.44	21.29	82.09	69.57	36.40
<i>RHOH</i>	eQTLGen	Blood	0.31	97.50	67.34	54.34	27.07
<i>RHOG</i>	eQTLGen	Blood	0.65	206.04	93.78	85.14	49.93
<i>RHOG</i>	TCGA	Breast cancer	0.35	16.90	72.78	59.63	30.06
<i>RHOC</i>	eQTLGen	Blood	1.90	609.56	>99	>99	91.90
<i>RHOB</i>	eQTLGen	Blood	0.59	188.24	91.67	81.88	46.50
<i>RHOB</i>	TCGA	Breast cancer	0.27	12.94	61.28	48.77	24.13
<i>RHOJ</i>	TCGA	Breast cancer	0.54	26.11	89.01	78.13	43.04
<i>RAC1</i>	eQTLGen	Blood	2.03	546.86	>99	>99	93.46
<i>RAC1</i>	TCGA	Breast cancer	0.28	13.67	63.65	50.91	25.24
<i>RAC2</i>	eQTLGen	Blood	2.65	863.39	>99	>99	97.77
<i>RAC2</i>	TCGA	Breast cancer	0.34	16.44	71.62	58.47	29.38
<i>RND2</i>	TCGA	Breast cancer	0.78	38.02	97.00	90.96	57.58

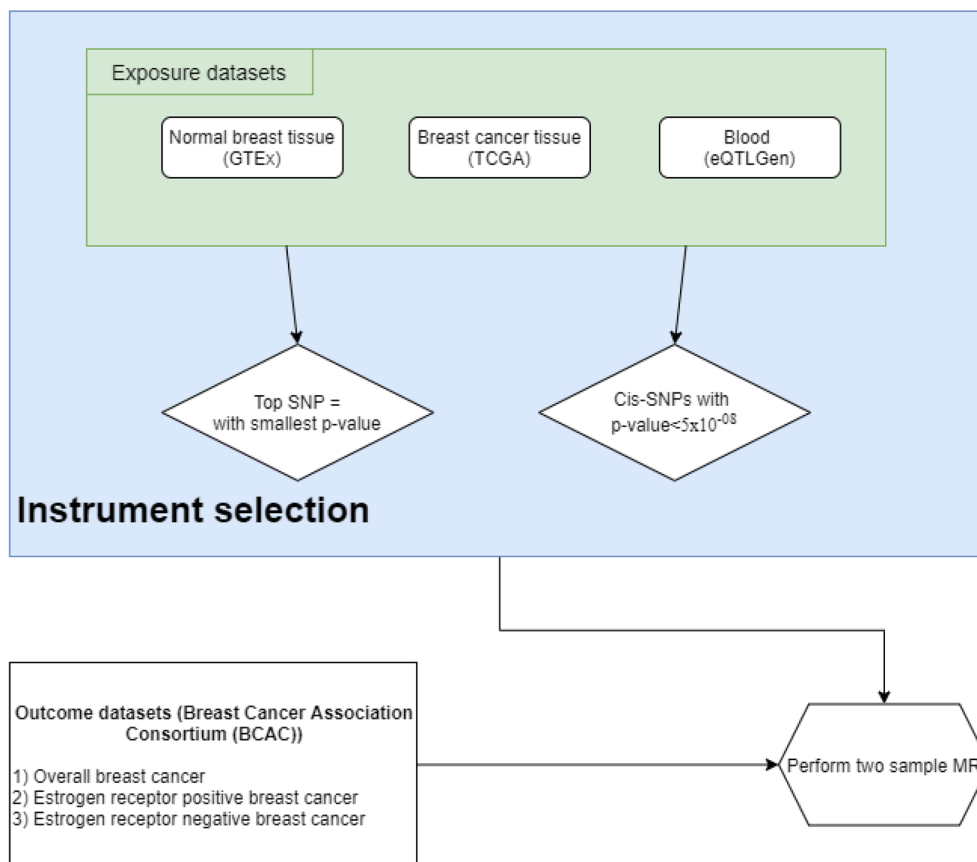
**Table 1.** Results of power calculation for the main analysis (using top eQTL). BC=breast cancer; R<sup>2</sup> represents the variance explained in the expression level of the gene by the instrument; F-statistic indicates strength of the instrument used for each gene expression (a strong instrument is sometimes defined as an F-statistic > 10); and the Power represents the power to detect an odds ratio of 1.2 (or 0.80) for an association of the expression of gene expression with breast cancer at an alpha-level of 0.05, given the values in the R<sup>2</sup> column and the number of breast cancer cases and controls.

*RHOD* in overall, ER+ and ER– breast cancer using eQTL obtained from breast tissue and ≥ 30% using eQTL obtained from blood (Table 1). The power was >99% for *CDC42* in overall, ER+ and ER– breast cancer using eQTL obtained from blood and ≥ 27.19 using eQTL obtained from breast cancer. The results for R<sup>2</sup>, F-statistic and power calculations are provided in Table 1.

In sensitivity analyses, using multiple SNPs obtained from breast cancer tissue, the direction of association for ER+ and ER– together was consistent with single SNP analyses but the evidence of association was weak (overall: OR per SD increase in expression level: 0.97; 95% CI 0.91, 1.04; P = 0.41, ER+ : OR per SD increase in expression level: 0.96; 95% CI 0.90, 1.03; P = 0.26 and ER–: OR per SD increase in expression level: 0.98; 95% CI 0.89, 1.08; P = 0.68). Using multiple SNPs obtained from blood, there was evidence of protective association with overall (OR per SD increase in expression level: 0.96; 95% CI 0.93, 0.99; P = 0.01) and ER+ breast cancer (OR per SD increase in expression level: 0.94; 95% CI 0.91, 0.98; P = 0.001) consistent to single SNP analyses. The evidence of association was inconclusive for ER– breast cancer risk (OR per SD increase in expression level: 1.01; 95% CI 0.96, 1.06; P = 0.66). For multiple SNP sensitivity analyses, the results for the effect of *CDC42* on overall and ER+ breast cancer were consistent across the various sensitivity analyses (Fig. 4 and 5). We did not find evidence of heterogeneity and pleiotropy across the individual causal effects (Supplementary Table 1).

In colocalization analyses for *RHOD* (using summary data from the GTE<sub>x</sub> platform) the posterior probability of colocalization (i.e. exposure and outcome are associated and share the same causal variant) was 84% for overall breast cancer risk and 98% for ER+ breast cancer risk suggesting that breast cancer risk and *RHOD* expression are both associated and share a single causal variant (Table 3). However, the evidence of colocalization was weak for ER– breast cancer risk and *RHOD* eQTLs (9%). There was weaker evidence of colocalization of the *CDC42* expression (using summary data from the eQTLGen platform) and breast cancer risk signals (Table 3).

We then assessed the prognostic role of increased *RHOD* expression using KM Plotter<sup>28</sup>. Using the mean expression of the 2 *RHOD* probes (209885\_at and 31846\_at) within KM Plotter and splitting patients by an auto-select best cut off (and adjusted our P-values accordingly for multiple comparisons induced by the auto-selection procedure), we found that increased *RHOD* expression was associated with an improved Recurrence Free Survival (RFS) in overall breast cancer (Hazard Ratio (HR) 0.86, (95% CI 0.78, 0.95), P = 0.0042) but a worsening trend towards overall survival (HR 1.19, (95% CI 0.98, 1.45), P = 0.072) (Fig. 6). We repeated the same analyses above in the oestrogen receptor positive (ER+) subgroup, as our Mendelian Randomization results suggested a stronger relationship between *RHOD* and ER+ breast cancer. This repeat analysis showed a concordant result for RFS and OS, with higher *RHOD* expression showing a trend towards worsening PFS (HR 1.16 (95%

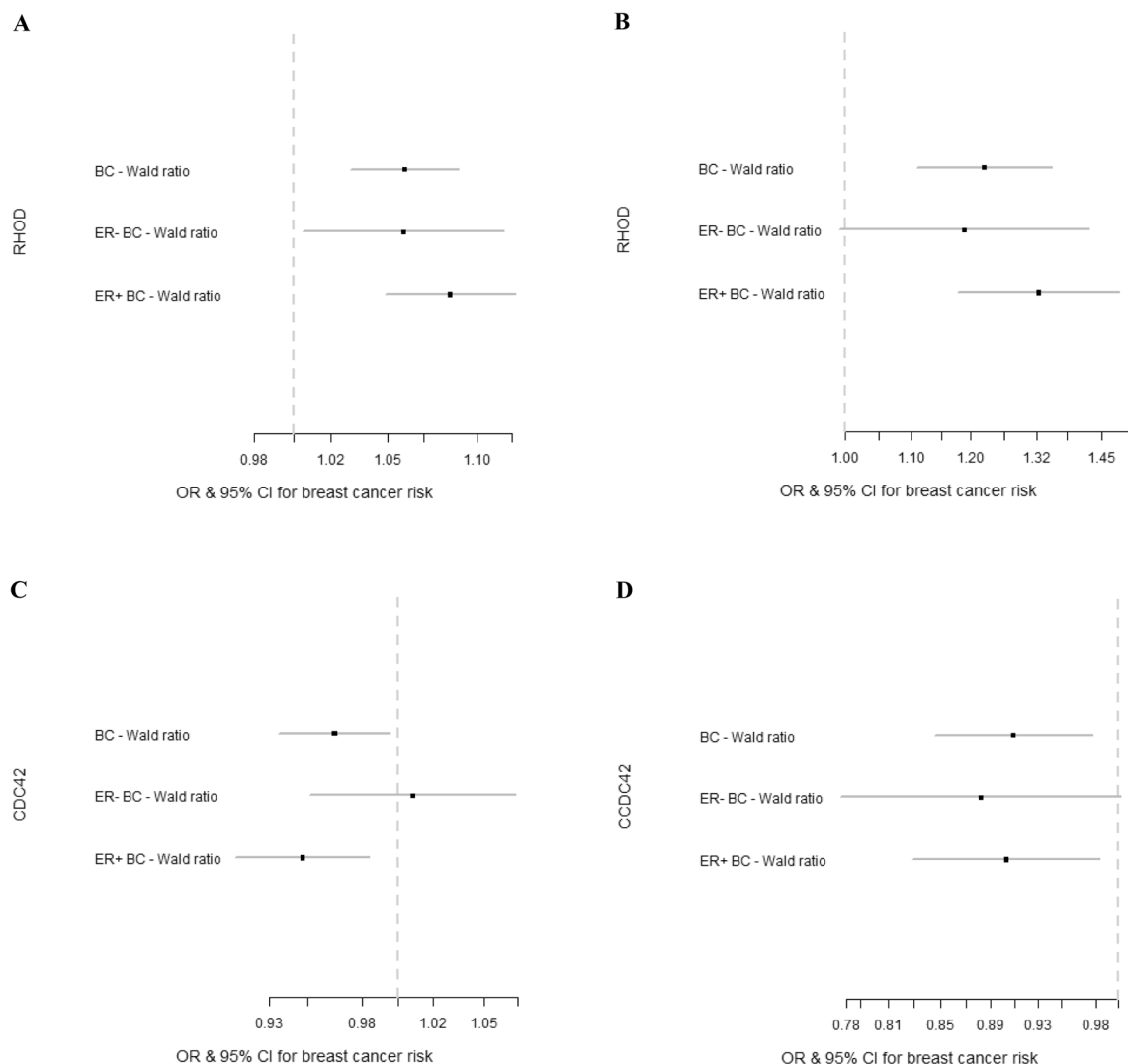


**Figure 2.** A diagrammatic representation of the datasets included and the methodology is shown as a flow diagram.

Outcome	Exposure (gene expression level)	Tissue (data source)	MR method	nsnp	OR	LCI	UCI	P-value
Overall breast cancer	<i>RHOD</i>	Breast mammary (GTEx)	Wald ratio	1	1.06	1.03	1.09	$5.65 \times 10^{-5}$
ER+ breast cancer	<i>RHOD</i>	Breast mammary (GTEx)	Wald ratio	1	1.08	1.05	1.12	$2.29 \times 10^{-5}$
ER- breast cancer	<i>RHOD</i>	Breast mammary (GTEx)	Wald ratio	1	1.06	1.01	1.12	0.03
Overall breast cancer	<i>RHOD</i>	Blood (eqlGen)	Wald ratio	1	1.22	1.11	1.35	$5.22 \times 10^{-5}$
ER+ breast cancer	<i>RHOD</i>	Blood (eqlGen)	Wald ratio	1	1.32	1.18	1.49	$2.74 \times 10^{-6}$
ER- breast cancer	<i>RHOD</i>	Blood (eqlGen)	Wald ratio	1	1.19	0.99	1.42	0.06
Overall breast cancer	<i>CDC42</i>	Breast cancer (TCGA)	Wald ratio	1	0.91	0.84	0.98	0.02
ER+ breast cancer	<i>CDC42</i>	Breast cancer (TCGA)	Wald ratio	1	0.90	0.82	0.99	0.03
ER- breast cancer	<i>CDC42</i>	Breast cancer (TCGA)	Wald ratio	1	0.88	0.76	1.02	0.09
Overall breast cancer	<i>CDC42</i>	Blood (eqlGen)	Wald ratio	1	0.96	0.93	1.00	0.03
ER+ breast cancer	<i>CDC42</i>	Blood (eqlGen)	Wald ratio	1	0.95	0.91	0.98	0.005
ER- breast cancer	<i>CDC42</i>	Blood (eqlGen)	Wald ratio	1	1.01	0.95	1.07	0.78

**Table 2.** Mendelian randomisation analyses of the association of *RHOD* and *CDC42* with overall, ER+ and ER- breast cancer risk. *nsnp* number of SNPs used in the analysis, *OR* odds ratio, *LCI* 95% lower confidence interval, *UCI* 95% upper confidence interval, *P-value* P-value for association.

CI 0.98–1.38),  $p=0.08$ ) and worse OS (HR 1.44 (95% CI 1.05–1.98)  $p=0.025$ ). We then assessed the relationship between PAM50 subtype, *RHOD* expression and RFS and OS, demonstrating conflicting results between OS and RFS. For increased *RHOD* expression the results were: Luminal A breast cancer, a protective effect on both OS (OS 0.52 (95% CI 0.33–0.81,  $p=0.0034$ ) and RFS 0.8, 95% CI 0.65–0.99,  $p=0.039$ ); Luminal B a worse OS (HR 1.75 (95% CI 1.21–2.53,  $p=0.0028$ ) but a trend towards improved RFS (HR 0.85 (95% CI, 0.69–1.04,  $p=0.11$ )); Her2 positive, a worse OS (HR 1.59 (95% CI, 1.06–2.37,  $p=0.024$ ) but improved PFS (0.86 (95% CI

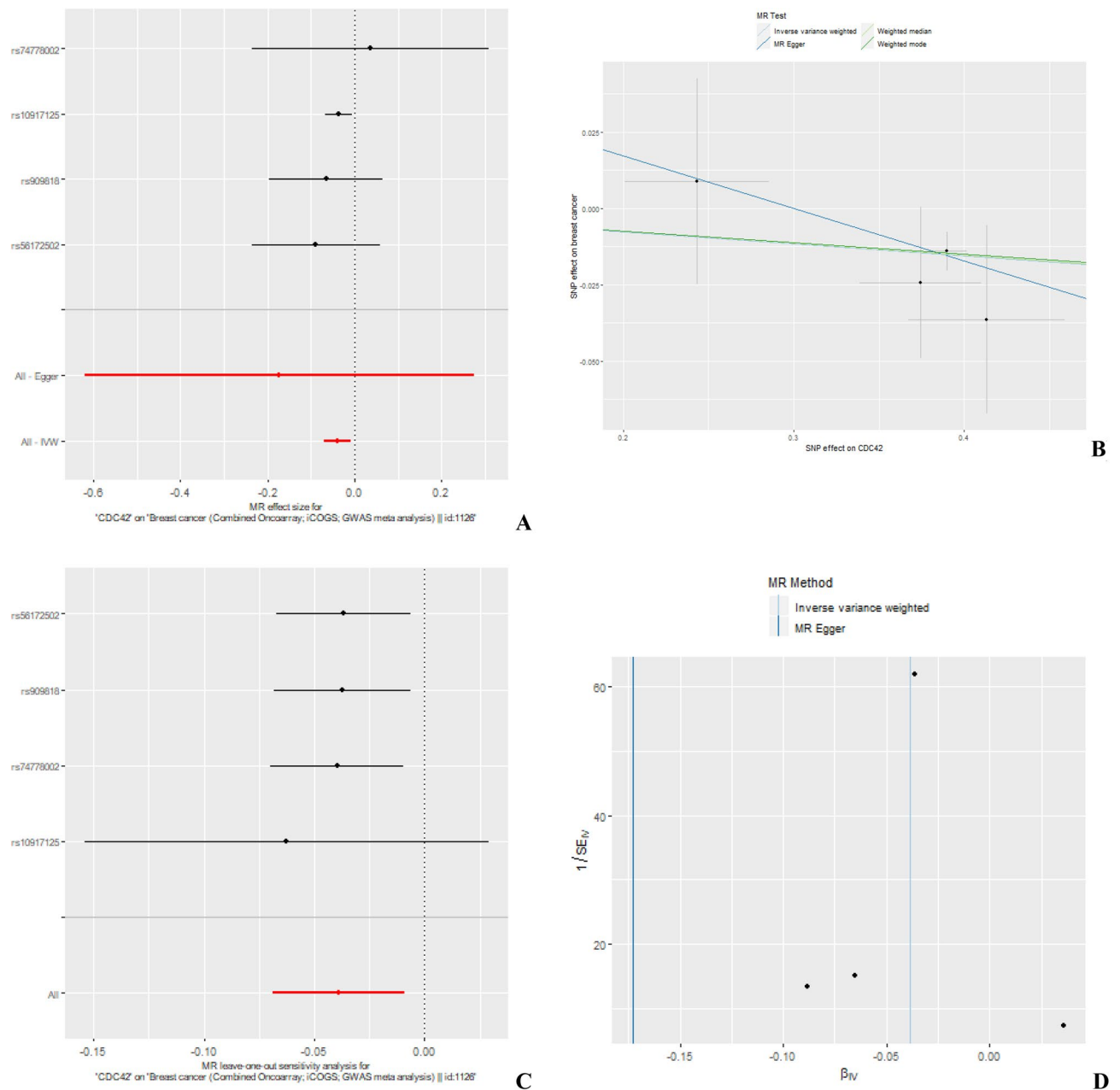


**Figure 3.** Results of MR analyses performed for overall, ER+ and ER– breast cancer risk for *RHOD* and *CDC42*. The genetic instruments for *RHOD* were obtained from breast mammary tissue (GTEx) (A) and from blood (eQTLGen) (B) and for *CDC42* were obtained from blood (eQTLGen) (C) and from breast cancer tissue (TCGA) (D). BC breast cancer, OR odds ratio, CI confidence interval.

0.68–1.1,  $p = 0.023$ ); and lastly Basal where there was a worse OS (HR 1.85 (95% CI, 1.27–2.71,  $p = 0.0013$ ) and a trend towards worsening PFS (HR: 1.15 (95% CI 0.93–1.42,  $p = 0.21$ )). These results highlight that further work is needed in larger datasets or with functional laboratory experiments to investigate the underlying biology by manipulating *RHOD* in representative cell lines.

We then extracted the rates of amplifications and copy number alterations (CNA) in *RHOD* using cBioPortal<sup>29</sup> and The Cancer Genome Atlas dataset<sup>30</sup> and found a low percentage of amplification events across this dataset (47/993 patients, 4.7%). This varied by molecular subtype of breast cancer, ranging from 2.3% for basal-like breast cancer to 8.7% in the Luminal B (ER+ with aggressive features) subgroup (Table 4). For CNA, 26.7% of patients (266/933) had amplifications in *RHOD* (Table 5).

To compare our novel results on *RHOD* expression with a known breast cancer-associated gene, we chose *TP53*, a well-known tumor suppressor<sup>31</sup>. Somatic mutations affecting the *TP53* gene (tumor protein p53) have been reported to be altered in breast cancer in approximately 20–40% of all cases, although this depends on the size of the tumor, molecular subtype and disease stage<sup>32</sup>. MR results showed evidence of a negative association between increased expression of *TP53* and ER+ breast cancer risk (OR per SD increase in expression level: 0.77; 95% CI 0.66, 0.90;  $P = 0.001$ ) using eQTLs obtained from blood. The direction of effect was consistent with overall breast cancer risk, but the evidence was weak (OR per SD increase in expression level: 0.92; 95% CI 0.81, 1.05;  $P = 0.22$ ). However, a positive association between *TP53* expression and ER– breast cancer was observed (OR per SD increase in expression level: 1.69; 95% CI 1.28, 2.24;  $P = 0.0002$ ) consistent with recent literature (see “Discussion”)<sup>17</sup>.



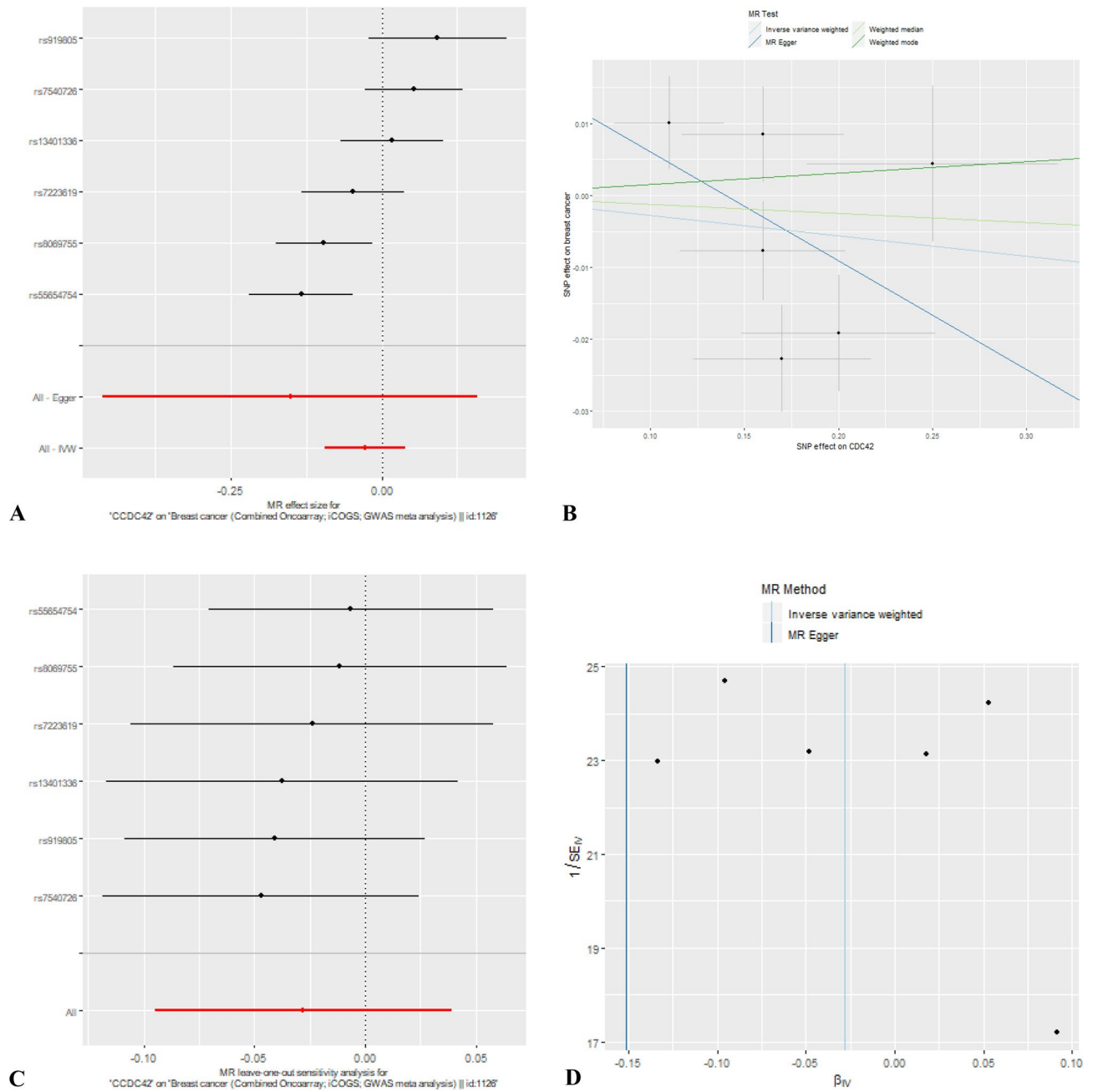
**Figure 4.** MR results for association between *CDC42* and overall breast cancer risk using eQTLs from blood. (A) Forest plot of single SNP analysis; (B) comparison of results using different MR methods; (C) leave-one-out sensitivity analysis; (D) funnel plot of IVW and MR-Egger regression.

## Discussion

Here we investigated the potential causal role of gene expression of Rho GTPases in breast cancer using a two-sample MR approach. Of the fourteen genes for which we found suitable germline genetic instruments, two—*RHOD* and *CDC42*—showed evidence of a link with breast cancer risk with the evidence being stronger for *RHOD*. The expression level of *RHOD* and *CDC42* demonstrated positive and inverse effects on overall breast cancer, respectively. Given that most cases in the overall breast cancer analysis were ER+, these associations were primarily driven by the ER+ subtype. There was a trend towards the same direction of effects in ER- breast cancer but this did not reach Bonferroni-corrected statistical significance, reflecting the lower numbers of ER- breast cancer cases in BCAC<sup>11</sup>.

To our knowledge, *RHOD* expression has not previously been implicated in breast cancer aetiology or indeed in any other cancer types, and thus our results are the first to link *RHOD* to cancer risk. There are very few studies on *RHOD* function compared to highly characterized Rho GTPases such as *CDC42*. *RHOD* was initially discovered as a regulator of membrane trafficking (reviewed in<sup>33</sup>) and contributes to the trafficking of tyrosine kinases such as SRC family and the PDGFR $\beta$  that are implicated in breast cancer progression<sup>34–37</sup>. It has also been reported to regulate several other cellular responses that are important for cellular transformation and





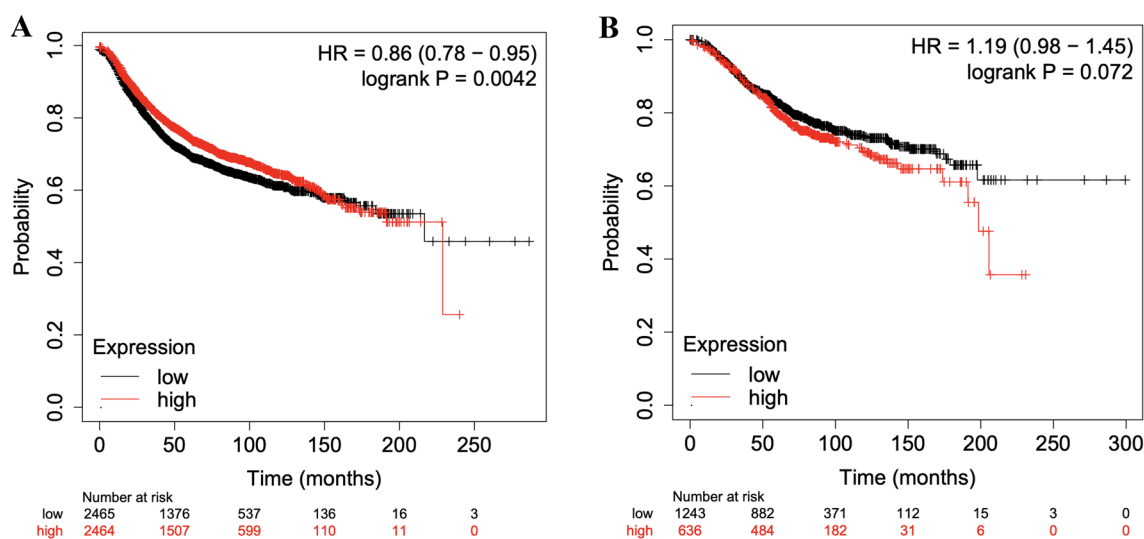
**Figure 5.** MR results for association between *CDC42* and overall breast cancer risk using eQTLs from breast cancer. **(A)** forest plot of single SNP analysis; **(B)** comparison of results using different MR methods; **(C)** leave-one-out sensitivity analysis; **(D)** funnel plot of IVW and MR-Egger regression.

early cancer development, including cell cycle progression, cytoskeletal dynamics, cell motility and centrosome duplication<sup>38,39</sup>.

Our results using publicly available datasets show a relatively low frequency of amplifications and copy number alterations within TCGA for *RHOD* (Tables 4 and 5) and a prognostic role of *RHOD* in both RFS and OS of oestrogen receptor positive breast cancer (Fig. 7). These data suggest that *RHOD* gene amplification could contribute to breast cancer progression in some women. Further analyses based on PAM50 subtype of breast cancer suggest that higher *RHOD* expression led to worse OS in more aggressive tumour subtypes (Luminal B, Her2 positive and Basal), but some results for RFS were conflicting. There are numerous caveats and considerations that need to be taken into account when examining RFS and OS in this context—including the lack of detailed information as to the treatment that the patients included in these datasets received, the longer relapse times that we typically see in oestrogen receptor positive breast cancer (meaning that relationships often take many years to observe) and the fact that overall survival data assesses all causes of death including those unrelated to breast cancer. Further work with larger numbers of women will be needed to assess whether the relationship between *RHOD* and RFS and OS is truly in opposite directions, with a focus on more aggressive breast cancer subtypes.

	Overall breast cancer	ER+ breast cancer	ER- breast cancer
<b><i>RHOD</i> (GTEX)</b>			
PP Hypothesis0	$4.68 \times 10^{-10}$	$2.29 \times 10^{-11}$	$9.26 \times 10^{-9}$
PP Hypothesis1	0.04	0.002	0.85
PP Hypothesis2	$9.63 \times 10^{-10}$	$1.98 \times 10^{-10}$	$6.77 \times 10^{-10}$
PP Hypothesis3	0.09	0.02	0.06
PP Hypothesis4	0.87	0.98	0.09
<b><i>CDC42</i> (eQTLGen)</b>			
PP Hypothesis0	$9.36 \times 10^{-306}$	$1.78 \times 10^{-305}$	$2.54 \times 10^{-305}$
PP Hypothesis1	0.33	0.62	0.89
PP Hypothesis2	$3.42 \times 10^{-306}$	$3.15 \times 10^{-306}$	$3.05 \times 10^{-306}$
PP Hypothesis3	0.12	0.11	0.11
PP Hypothesis4	0.55	0.27	0.007

**Table 3.** Results of colocalisation analyses for *RHOD* and *CDC42* genes. *n*<sub>snp</sub> number of SNPs used in the analysis, *PP* posterior probability.



**Figure 6.** Kaplan–Meier plots of (A) recurrence-free survival in patients and (B) overall survival with ‘high’ and ‘low’ *RHOD* expression in overall breast cancer.

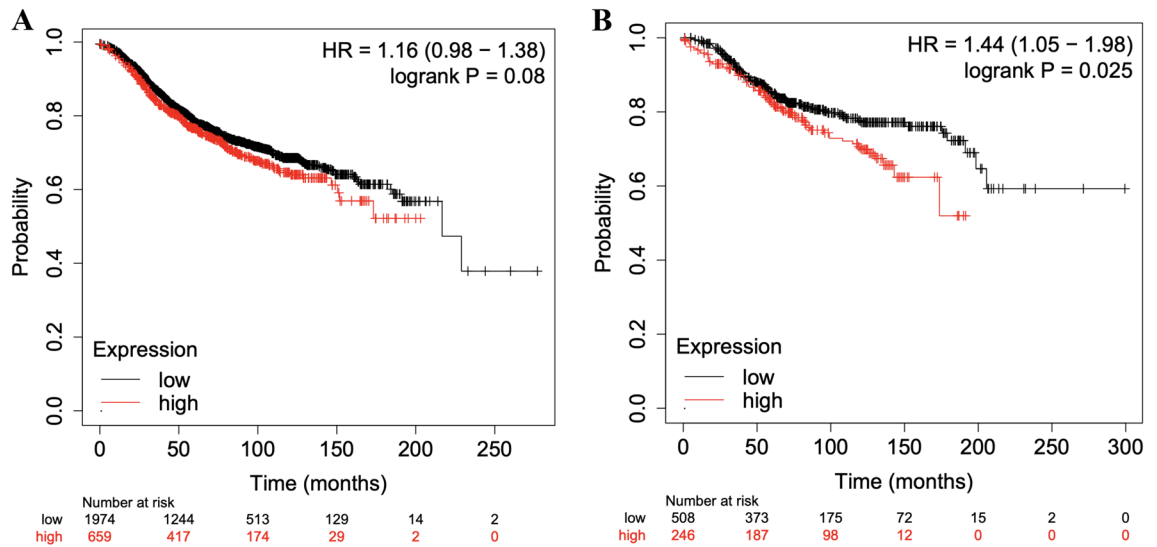
Molecular subtype	Number of patients	Number (percentage) with amplifications
Basal	130	3 (2.3)
Her2 positive	67	6 (9.0)
Luminal A	395	14 (3.5)
Luminal B	183	16 (8.7)
Normal	22	1 (4.5)
Unclassified	196	7 (3.6)
Overall	993	47 (4.7)

**Table 4.** Amplifications in *RHOD* in The Cancer Genome Atlas according to molecular subtype.

Our results demonstrated a protective effect of increased *TP53* expression on ER+ breast cancer risk, consistent with the known tumor suppressor role for *TP53*<sup>31</sup>. By contrast, the opposite effect was observed for ER- breast cancer. The relationship between germline and somatic changes affecting intra-tumoral *TP53* activity is complex and is altered dependent upon the rate of somatic mutations in *TP53* (approximately 80% in ER- and 20% in ER+ breast cancer). A discussion around this relationship and the effects of *TP53* on breast cancer risk has been recently published<sup>17</sup> but interestingly the strength of effect on breast cancer risk we report with *RHOD* is even stronger than that of a well-established target (rs78378222 at *TP53*) on breast cancer risk.

Molecular subtype	Number of patients	Number of copy number alterations (percentage)			
		-1 Loss	0 Neutral	1 Gain	2 Amplification
Basal	130	27 (20.8)	67 (51.5)	33 (25.4)	3 (2.3)
Her2 positive	67	11 (16.4)	30 (44.8)	20 (29.9)	6 (9.0)
Luminal A	395	44 (11.1)	264 (66.8)	73 (18.5)	14 (3.5)
Luminal B	183	37 (20.2)	76 (41.5)	54 (29.5)	16 (8.7)
Normal	22	1 (4.5)	17 (77.3)	3 (13.6)	1 (4.5)
Unclassified	196	27 (13.8)	127 (64.8)	36 (18.4)	7 (3.6)
Overall	993	147 (14.8)	581 (58.5)	219 (22.1)	47 (4.7)

**Table 5.** Copy number alterations in *RHOD* in The Cancer Genome Atlas according to molecular subtype.



**Figure 7.** Kaplan–Meier plots of recurrence free survival (A) and overall survival (B) according to expression of *RHOD* in oestrogen receptor positive breast cancer by IHC.

Together, our results suggest that *RHOD* may play an important role in breast cancer aetiology and the mechanisms through which *RHOD* could drive breast cancer formation merit further investigation.

*CDC42* is a key regulator of both cell motility and cell cycle progression<sup>40</sup> as well as in establishing normal epithelial cell polarity and promoting migratory polarity<sup>7</sup>. Compared to other Rho GTPases, *CDC42* is relatively less studied in cancer<sup>41</sup>. In single-cell in vitro experiments it promotes a mesenchymal phenotype with cellular invasion<sup>42</sup> but its relationship with breast cancer phenotypes is complex. Previous reports have demonstrated overexpression in 42–57% of human breast cancers on a protein level associated with more aggressive clinical behaviour, ductal carcinomas and a poorer prognosis with increased cytoplasmic expression but the opposite with nuclear expression<sup>43</sup>. There are also previous reports suggesting that *CDC42* activation leads to a decreased migratory phenotype in breast cancer cell lines<sup>44,45</sup>. A genomic and transcriptomic analysis by the Molecular Taxonomy of the Breast Cancer International Consortium (METABRIC) demonstrated that less aggressive, ER+ tumours were enriched for altered expression of genes in the *CDC42* pathway<sup>4</sup>. Despite this, mutations in *CDC42*-related genes are very low at between 0.1 and 1.7% and the elevated *CDC42* expression in breast cancer is thought to be due to activation of oncogenes or cell surface receptors (for example; epidermal growth factor receptor (EGFR)) that lead to *CDC42* upregulation<sup>46</sup>. Differences in tissue-specific *CDC42* expression biasing the result are unlikely as the effects of SNPs for *CDC42* were derived from both breast cancer tissue and from blood and were in concordance. The more plausible explanation for the apparent protective effect of *CDC42* is that *CDC42* maintains epithelial polarity<sup>47,48</sup> and hence protects against cancer initiation. At later stages of breast cancer development, increased *CDC42* expression could promote cancer progression via its effects on cell cycle progression and invasion<sup>3</sup>. The activity of most Rho GTPases, including *RHOD* and *CDC42*, is also controlled by over 70 guanine nucleotide exchange factors (GEFs), 60 GTPase-activating proteins (GAPs) and 3 Rho GDI proteins (guanine–nucleotide-dissociation inhibitors) that switch the Rho GTPases between active and inactive forms<sup>39</sup>. Interestingly, several GAPs are upregulated in basal-like breast cancers and contribute to breast cancer cell growth<sup>49</sup>.

We performed sensitivity analyses to disentangle the causal effects of gene expression from associations driven by horizontal pleiotropy, reverse causation, and genetic confounding through LD. We found robust evidence of colocalization for *RHOD*, suggesting that our MR findings could not be driven by genetic confounding through

LD between eQTLs and other disease-causal SNPs strengthening the evidence of causality. Evidence of colocalization thus served as a complementary approach to reinforce the MR finding for *RHOD*.

We have tested the effects of Rho GTPases in human samples rather than cancer-derived cell lines or animal models that often give variable and inconsistent results. In contrast to conventional observational studies, MR is less susceptible to problems of confounding, reverse causation and measurement error. The use of two-sample MR enabled us to use the largest GWAS of breast cancer to date. Ideally, we would have liked to examine protein-level expression of the Rho GTPases as well as gene expression, but there are very few small studies on protein levels compared to the large datasets available for normal breast, breast cancer and blood-based transcriptomic profiles, and none to our knowledge for *RHOD*. Although we could not detect strong evidence of effects for Rho GTPases for the ER- breast cancer subtype, this could be due to a lack of power to detect smaller effect sizes in this cohort. Larger studies with matched germline genotype and tissue-specific normal and tumour gene expression will be required to improve the statistical power of similar MR analyses in the future and would help to causally link other members of the Rho GTPase pathway with breast cancer risk.

## Conclusion

We found robust evidence that *RHOD* expression may be causally and positively related to breast cancer risk, and that *CDC42* may potentially be causally and negatively related to breast cancer risk. Given that the activity of *RHOD* and *CDC42* proteins is regulated by a variety of other proteins, it will be interesting to determine whether any of the genes encoding these regulators are also associated with breast cancer risk. The role of *RHOD* warrants further biological investigation to assess its role in breast carcinogenesis.

## Data availability

All data analysed during this study were previously generated. Data availability repository links are given below: (1) BCAC GWAS: <http://bcac.ccg.med.schl.cam.ac.uk/bcacdata/oncoarray/oncoarray-and-combined-summary-result/gwas-summary-results-breast-cancer-risk-2017/>. (2) TCGA eQTLs: [https://albertlab.shinyapps.io/tcga\\_eqtl/](https://albertlab.shinyapps.io/tcga_eqtl/). (3) GTEx: <https://www.gtexportal.org/home/datasets>.

Received: 12 February 2021; Accepted: 12 January 2022

Published online: 27 January 2022

## References

- Ridley, A. J. Rho proteins and cancer. *Breast Cancer Res. Treat.* **84**(1), 13–19 (2004).
- Hodge, R. G. & Ridley, A. J. Regulating Rho GTPases and their regulators. *Nat. Rev. Mol. Cell Biol.* **17**(8), 496–510 (2016).
- Orgaz, J. L., Herraiz, C. & Sanz-Moreno, V. Rho GTPases modulate malignant transformation of tumor cells. *Small GTPases*. **5**, e983867 (2014).
- Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**(7403), 346–352 (2012).
- Liu, A.-X., Rane, N., Liu, J.-P. & Prendergast, G. C. RhoB Is dispensable for mouse development, but it modifies susceptibility to tumor formation as well as cell adhesion and growth factor signaling in transformed cells. *Mol. Cell Biol.* **21**(20), 6906–6912 (2001).
- Avraham, R. & Weinberg, R. A. Characterization and expression of the human rhoH12 gene product. *Mol. Cell Biol.* **9**(5), 2058–2066 (1989).
- De, P. *et al.* RAC1 GTP-ase signals Wnt-beta-catenin pathway mediated integrin-directed metastasis-associated tumor cell phenotypes in triple negative breast cancers. *Oncotarget* **8**(2), 3072–3103 (2017).
- Yarmolinsky, J. *et al.* Causal inference in cancer epidemiology: What is the role of Mendelian randomization?. *Cancer Epidemiol. Biomarkers Prev.* **27**(9), 995–1010 (2018).
- Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?. *Int. J. Epidemiol.* **32**(1), 1–22 (2003).
- Pierce, B. L. & Burgess, S. Efficient design for Mendelian randomization studies: Subsample and 2-sample instrumental variable estimators. *Am. J. Epidemiol.* **178**(7), 1177–1184 (2013).
- Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**(7678), 92–94 (2017).
- The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**(6509), 1318–1330 (2020).
- Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv. 447367 (2018).
- Lim, Y. W. *et al.* Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. *PNAS* **115**(50), E11701–E11710 (2018).
- Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**(7571), 68–74 (2015).
- Stacey, S. N. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.* **43**(11), 1098–1103 (2011).
- Zhang, P. *et al.* Germline and somatic genetic variants in the p53 pathway interact to affect cancer risk, progression, and drug response. *Cancer Res.* **81**(7), 1667–1680 (2021).
- Burgess, S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int. J. Epidemiol.* **43**(3), 922–929 (2014).
- Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
- Johnson, T. *Efficient Calculation for Multi-SNP Genetic Risk Scores* 1.
- Thompson, S. G. & Sharp, S. J. Explaining heterogeneity in meta-analysis: A comparison of methods. *Stat. Med.* **18**(20), 2693–2708 (1999).
- Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genet.* **10**(5), e1004383 (2014).
- Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**(4), 304–314 (2016).
- Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**(6), 1985–1998 (2017).
- Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**(2), 512–525 (2015).

26. Angrist, J. D., Imbens, G. W. & Rubin, D. B. Identification of causal effects using instrumental variables: Rejoinder. *J. Am. Stat. Assoc.* **91**(434), 468 (1996).
27. Sterne, J. A. C. *et al.* Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* **343**, d4002–d4002 (2011).
28. Nagy, Á., Munkácsy, G. & Györffy, B. Pancancer survival analysis of cancer hallmark genes. *Sci. Rep.* **11**(1), 6047 (2021).
29. Cerami, E. *et al.* The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**(5), 401–404 (2012).
30. Network, C. G. A. Comprehensive molecular portraits of human breast tumours. *Nature* **490**(7418), 61–70 (2012).
31. Kruiswijk, F., Labuschagne, C. F. & Vousden, K. H. p53 in survival, death and metabolic health: A lifeguard with a licence to kill. *Nat. Rev. Mol. Cell Biol.* **16**(7), 393–405 (2015).
32. Børresen-Dale, A.-L. TP53 and breast cancer. *Hum. Mutat.* **21**(3), 292–300 (2003).
33. Phuyal, S. & Farhan, H. Multifaceted rho GTPase signaling at the endomembranes. *Front. Cell Dev. Biol.* **7**, 127 (2019).
34. Sandilands, E. *et al.* Src kinase modulates the activation, transport and signalling dynamics of fibroblast growth factor receptors. *EMBO Rep.* **8**(12), 1162–1169 (2007).
35. Nehru, V., Voytyuk, O., Lennartsson, J. & Aspenström, P. RhoD binds the Rab5 effector rabankyrin-5 and has a role in trafficking of the platelet-derived growth factor receptor. *Traffic* **14**(12), 1242–1254 (2013).
36. Espada, J. & Martín-Pérez, J. An update on Src family of nonreceptor tyrosine kinases biology. *Int. Rev. Cell Mol. Biol.* **1**(331), 83–122 (2017).
37. Jitariu, A.-A., Raica, M., Cimpean, A. M. & Suci, S. C. The role of PDGF-B/PDGFR-BETA axis in the normal development and carcinogenesis of the breast. *Crit. Rev. Oncol. Hematol.* **131**, 46–52 (2018).
38. Kyrkou, A. *et al.* RhoD participates in the regulation of cell-cycle progression and centrosome duplication. *Oncogene* **32**(14), 1831–1842 (2013).
39. Vega, F. M. & Ridley, A. J. Rho GTPases in cancer cell biology. *FEBS Lett.* **582**(14), 2093–2101 (2008).
40. Fritz, G., Just, I. & Kaina, B. Rho GTPases are over-expressed in human tumors. *Int. J. Cancer* **81**(5), 682–687 (1999).
41. Humphries, B., Wang, Z. & Yang, C. Rho GTPases: Big players in breast cancer initiation, metastasis and therapeutic responses. *Cells* **9**(10), E2167 (2020).
42. Gaggioli, C. *et al.* Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nat. Cell Biol.* **9**(12), 1392–1400 (2007).
43. Chrysanthou, E. *et al.* Phenotypic characterisation of breast cancer: The role of CDC42. *Breast Cancer Res. Treat.* **164**(2), 317–325 (2017).
44. Zuo, Y., Wu, Y. & Chakraborty, C. Cdc42 negatively regulates intrinsic migration of highly aggressive breast cancer cells. *J. Cell Physiol.* **227**(4), 1399–1407 (2012).
45. Prasad, C. P., Chaurasiya, S. K., Axelsson, L. & Andersson, T. WNT-5A triggers Cdc42 activation leading to an ERK1/2 dependent decrease in MMP9 activity and invasive migration of breast cancer cells. *Mol. Oncol.* **7**(5), 870–883 (2013).
46. Zhang, Y. *et al.* Focus on Cdc42 in breast cancer: New insights, target therapy development and non-coding RNAs. *Cells* **8**(2), 146 (2019).
47. Pichaud, F., Walther, R. F. & Nunes de Almeida, F. Regulation of Cdc42 and its effectors in epithelial morphogenesis. *J. Cell Sci.* **132**(10), jcs217869 (2019).
48. Mack, N. A. & Georgiou, M. The interdependence of the Rho GTPases and apicobasal cell polarity. *Small GTPases.* **5**(2), 10 (2014).
49. Lawson, C. D. *et al.* Rho GTPase transcriptome analysis reveals oncogenic roles for Rho GTPase-activating proteins in basal-like breast cancers. *Cancer Res.* **76**(13), 3826–3837 (2016).

## Acknowledgements

The authors would like to thank the participants of the individual studies contributing to the BCAC for their participation in these studies along with the principal investigators of BCAC for generating the data utilised for this analysis and for making these data available in the public domain.

## Author contributions

T.R., R.M. and A.R. conceived the study. N.K., T.R., J.Z. and S.K. designed and conducted the analyses. T.R. and N.K. wrote the manuscript. N.K., T.R., J.Z., S.K., R.M. and A.R. drafted, revised the manuscript and have approved the final version for publication.

## Funding

This work was supported by a Cancer Research UK (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme). TR is supported by an NIHR Academic Clinical Lectureship at the University of Bristol. RMM was supported by the NIHR Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. JZ is funded by a Vice-Chancellor's Fellowship from the University of Bristol. This research was also funded by the UK Medical Research Council Integrative Epidemiology Unit (MC\_UU\_00011/4and MC\_UU\_00011/5).

## Competing interests

TR has received educational grants from Daiichi-Sankyo and Amgen to attend educational workshops. This has not had any influence on this work. RMM has received a grant from Cancer Research UK, which is paid to the institution. All other authors have no competing interests to declare.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-05549-5>.

**Correspondence** and requests for materials should be addressed to N.K. or T.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022