BMC Cancer

RESEARCH

Open Access

# Novel insight of critical genes involved in breast cancer brain metastasis: evidence from a cross-tissue transcriptome association study and validation through external clinical cohorts

Jinsong Liu[1,2†], Xiao Guan[3†], Songlin Gao[1,2†], Liuliu Quan[1,2†], Min Dou[4], Jian Yue[2], Mengwu Shi[2] and Peng Yuan[2*]

## Abstract

**Background** Breast cancer represents the most prevalent form of tumors among females and is characterized by a significant genetic component. The brain is a frequent site of metastasis for breast cancer. Although numerous loci associated with breast cancer brain metastasis (BCBM) have been identified, the critical regulatory genes underlying BCBM remain largely unclear.

**Methods** The FinnGen R11 dataset was combined with Genotype-Tissue Expression Project (GTEx) for Transcriptome-wide Association Study (TWAS). The Unified Test for Molecular Signatures (UTMOST), Multimarker Analysis of Genomic Annotation (MAGMA), and Functional Summary-based Imputation (FUSION) were used to identify candidate genes. Summary-data-based mendelian randomization (SMR) and co-localization were performed further to elucidate the association between key genes and BCBM. Finally, multiple external cohorts were obtained to validate the findings.

**Result** In our study, 12 new genes associated with breast cancer were identified with TWAS. Subsequently, both SMR and co-localization have shown that CAPS8 was only expressed in brain tissues including frontal cortex and cerebellar hemispheres associated with breast cancer. Potential regulation of CASP8 could occur in BCBM. Finally, the findings were ultimately validated by external clinical cohorts.

**Conclusion** Our study identified key gene CASP8, which was associated with BCBM, providing new insights into the occurrence of BCBM.

**Keywords** Breast cancer brain metastasis, TWAS, UTMOST, FUSION, SMR, Co-localization

†Jinsong Liu, Xiao Guan, Songlin Gao and Liuliu Quan contributed equally to this work.

*Correspondence:
Peng Yuan
yuanpengyp01@163.com
[1]Department of Medical Oncology, National Clinical Research Center for Cancer/Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China
[2]Department of VIP Medical Services, National Clinical Research Center for Cancer/Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China
[3]State Key Lab of Molecular Oncology, Department of Pancreatic and Gastric Surgery, National Clinical Research Center for Cancer/Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China
[4]BengBu Medical University, BengBu 233030, China

## Introduction

Breast cancer is a highly prevalent tumor worldwide [1]. According to the global cancer report, BC ranks the second in incidence and the fourth in mortality rates of all tumors, which poses a serious challenge to the lives and health of people [2]. Advanced metastasis is the leading cause of death and brain metastasis is the most common distant metastasis of breast cancer. Certain subtypes such as HER2-positive breast cancer are prone to brain metastasis, which occurs in 50% of patients, with poorer survival outcomes [3]. Although there are standard treatments available for BCBM, including surgery, radiation, chemotherapy, targeted therapies, and new antibody-coupled drugs, mortality rates from BCBM remain high. Therefore, identifying novel key targets for BCBM is especially urgent for breast cancer patients.

Previous studies showed that the occurrence of breast cancer was mainly attributed to the polygenic trait of the disease. Recently, research on identifying genetic loci of diseases with GWAS was increasingly being conducted. Joana M Xavier identified candidate pathogenic variants at 41 breast cancer risk loci with GWAS [4]. Guochong Jia Identified genetic risk loci and blood risk biomarkers of breast cancer by integrating genomics and proteomics data [5]. However, many trait loci of the disease were located in non-coding regions, posing a challenge in identifying their functional value [6]. In addition, complex linkage disequilibrium (LD) could mask the screening of causal genes driving the disease [7].

TWAS is an approach for identifying genes associated with complex diseases through the transcriptome. Researchers can identify genes that may affect diseases through altered expression, providing insights for subsequent functional validation [8]. UTMOST is a common TWAS tool that integrates multiple datasets across tissues to identify important genes that may be missed in traditional single-tissue analysis [9]. Moreover, FUSION identifies hub genes that potentially influence phenotype by utilizing gene expression and GWAS data, which is critical to understanding the biology of disease and developing new therapeutic strategies [10]. MAGMA serves as a tool for association analysis of genes or sets of genes and can identify genes as potential biomarkers [11]. In recent years, cross-tissue TWAS has been widely used to screen vital genes for complex diseases, including cardiovascular or endocrine diseases, and various cancers.

In this work, we integrated eQTL data from the GTEx project and GWAS of European populations with breast cancer from the Finnish R11 database to conduct cross-tissue TWAS. Next, we performed three TWAS approaches including FUSION, UTMOST, and MAGMA to identify candidate genes. Subsequently, SMR and co-localization analysis were used to further elucidate the relationships between hub genes and BCBM. Finally, multiple external cohorts with BCBM were obtained to validate our findings.

## Materials and methods

### The data source of breast cancer

The FinnGen study was a large-scale genomics initiative that analyzed over 500,000 samples. GWAS data for breast cancer were obtained from cohort II Neoplasms, from cancer register (ICD-O-3) of the FinnGen R11 dataset, which included 20,586 cases and 201,494 controls of European ancestry [12]. The GTEx v8 project, which included 49 tissues from 838 post-mortem donors, aimed to discover the genetic effects of diseases and characterize the underlying molecular mechanisms [13]. Primary model weights files for gene expression were obtained from GTEx v8. We acquired BCBM datasets from Gene Expression Omnibus [14]. GSE46928 and GSE52604 datasets contained the breast cancer patients who were examined with brain metastasis and some patients with no brain metastasis.

### TWAS analysis in cross-tissue and single tissues

EQTL Integration was an effective method to connect SNPs, genes and complex traits in genetic association analysis. TWAS has provided new insights into the genetic basis of many diseases and traits [15]. Common TWAS analysis included FUSION and UTMOST.

Based on the tissue-dependent trait of transcriptional regulation, we used UTMOST to quantify gene-trait association. This method identified more genes in tissue with enriched trait heritability and enhanced inference accuracy. Subsequently, we used the generalized Berk-Jones test to integrate the relationship of gene and traits [16]. After applying a false discovery rate (FDR) correction, FDR < 0.05 was considered statistically significant.

FUSION was designed to identify relationships between GWAS and phenotypes by constructing a predictive model and predicting and testing the above association [17]. We integrated eQTL data with breast cancer to estimate the associations between each gene and with disease. First, the LD between the prediction model and the SNP at each locus of GWAS was estimated using 1,000 Genomes European samples. Subsequently, FUSION integrated several prediction models to assess the impact of SNPs on gene expression weights. Model representations utilized the highest predictive performance to determine gene weights [18]. Finally, we combined the genetic effect of breast cancer with these gene weights for TWAS analysis.

### Conditional and joint analysis

In genomic association studies, a single locus might contain multiple features (multiple SNPs or genes) that were associated with phenotype [19]. One of the roles

of COJO (the post-process module in FUSION) was to identify which of these features were statistically conditionally independent, which meant that their effects on the phenotype were not accounted for the effects of other features [17]. COJO ensured a more comprehensive understanding of the variation by interpreting LDs between genes [20]. Following testing, genes representing independent associations were termed significant, while those no longer showing significance were considered marginally significant.

#### Analysis of genetic effects of genomes

MAGMA was used for gene set association analysis, which could directly find functional genes or functional modules (gene regulatory pathways) related to the trait of interest, and was also conducive to the discovery of genes associated with multiple micro-effect SNPs [21]. By integrating gene expression data, MAGMA could assess gene expression patterns in different tissues and determine their correlation to specific phenotypes [22].

#### Summary data based Mendelian randomization and bayesian colocalization

The integrative analytical framework of Summary-based Mendelian Randomization (SMR) was originally developed by Zhu et al. (2016) to investigate genotype-phenotype associations [23]. In the current investigation, transcriptome-wide association signals for breast cancer were evaluated through systematic integration of genome-wide association statistics with cis-expression quantitative trait loci (cis-eQTL) profiles encompassing 49 anatomically distinct tissues from the Genotype-Tissue Expression project (GTEx) version 8. Population-specific linkage disequilibrium (LD) patterns were quantified utilizing phased genotype data obtained from the European ancestry cohort of the 1000 Genomes Project Phase III. All analytical procedures adhered to the protocol-specified parameter configurations embedded within the SMR software (https://yanglab.westlake.edu.cn/software/smr/#Overview). A strict Bonferroni-corrected SMR threshold of $p < 0.05$ was performed. Heterogeneity in the dependent instrument (HEIDI) test was used to distinguish pleiotropy from linkage, where $P$-HEIDI $< 0.01$ was considered likely due to pleiotropy and was therefore discarded from the analysis [24].

Co-localization analysis could be used to genetically co-locate two potentially related phenotypes to determine whether they shared common genetic causal variants in a given region [25]. Co-localization was based on sample sizes and converted correlation statistics into effect sizes. The method estimated the posterior probability (PP) that two outcomes within a locus were correlated and driven by a common causal variable [26]. The hypothesis has five PP statistics: H0: no features were genetically associated with SNPs in the region; H1: only feature 1 was genetically associated with SNPs; H2: only feature 2 was genetically associated with an SNP; H3: both traits were associated with SNPs, but using different dependent variables; H4: both traits were associated with the SNP and shared a causal variable (focus on the probability of H4, PPH4, satisfying PPH4 > 0.7, which indicated that there could be co-localization between the two traits) [27].

#### Validation with mRNA expression levels of key gene in multiple clinical cohorts

We then obtained breast cancer tissues from 10 breast cancer and 10 breast cancer patients with brain metastases (BM) for transcriptome sequencing from Cancer Hospital, Chinese Academy of Medical Sciences (CHCAMS). All approaches were carried out according to the relevant guidelines. The detailed method has been described previously [28]. The above experiments were approved by the Ethical Review Committee of CHCAMS and the patient has given informed consent.

#### Statistical analysis

GraphPad Prism 9.0 and R software (version 4.3.1) were used in our analysis. Wilcoxon test was used to compare clinicopathological characteristics between groups. The differences between breast cancer tissues from the No-BM and BM patients were compared using unpaired t-tests.

## Results

#### TWAS analysis in different tissues

The detailed workflow of this study was shown in Fig. 1. In the process of cross-tissue association analysis using UTMOST analysis, we first performed 49 single-tissue analysis, then cross-tissue analysis was conducted using GBJ and p-value adjustment was applied to the results. Finally, the results were obtained as shown in Table 1 and we obtained 67 genes in the UTMOST analysis for the subsequent study.

Next, we performed FUSION analysis with 49 tissue-sourced gene expression weight files, which were first conducted according to individual tissues, calculated adjusted p-value, combined the results from multiple tissues, and finally filtered the results that met the criteria. The detailed results were shown in Table S1, where we obtained 599 target genes during FUSION analysis. A total of 24 candidate genes including CASP8, TNS1, STRADB, EFR3B, SLC35E4, HSCB, MTMR3, DNAJC27, POMC, AAMP, ADCY3, CENPO, FLACC1, FZD7, DUSP18, AQP4, APOBEC3B, CBX6, CHEK2, PTRHD1, MAFF, RPL37A, DNAJC27-AS1, LINC00570 met the strict thresholds in both the cross-tissue and single-tissue analyses.

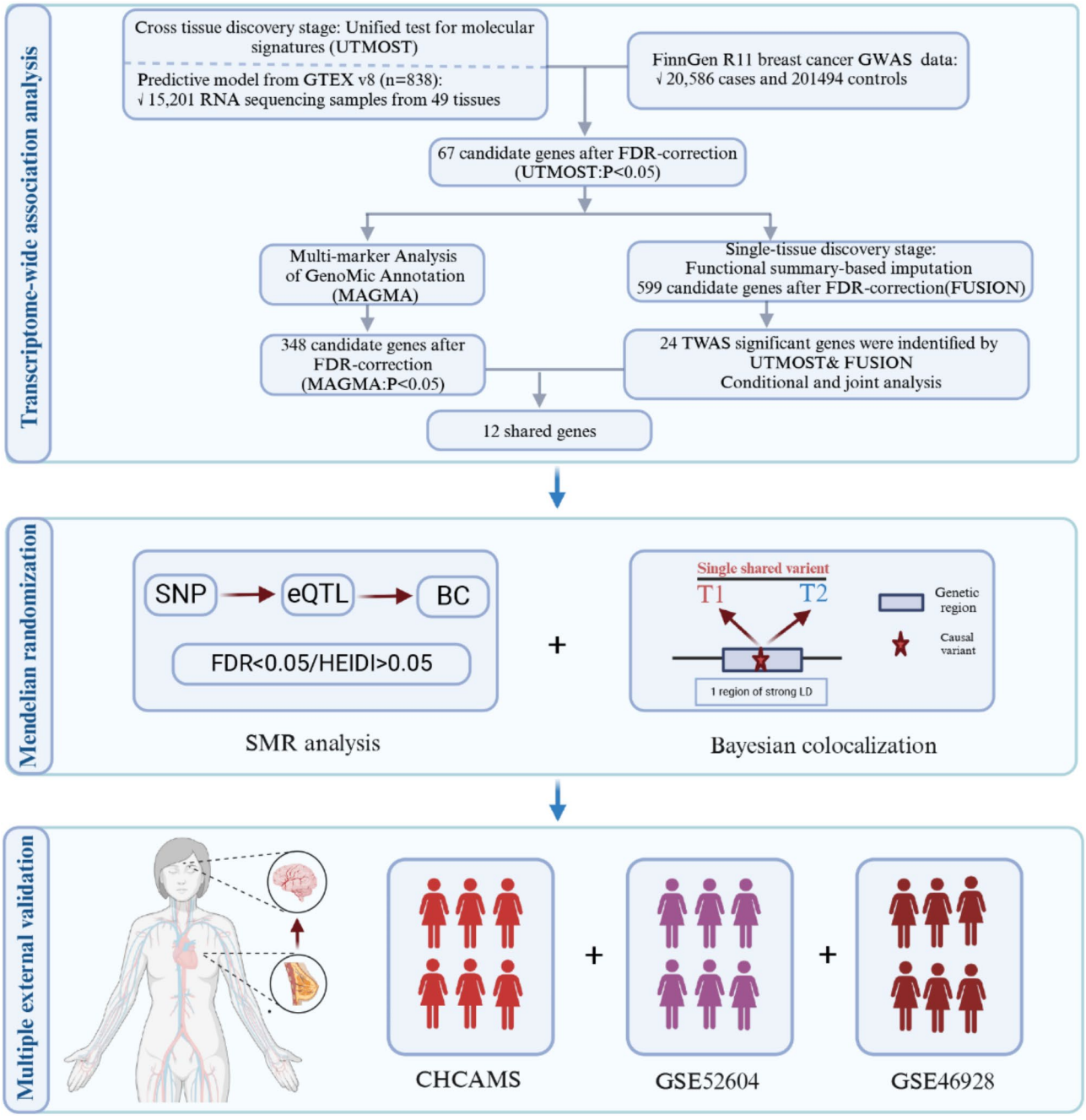


**Fig. 1** The flowchart of our study. GWAS, genome-wide association study; GTEx, Genotype-Tissues Expression Project; TWAS, transcriptome-wide association studies; BC, breast cancer; CHCAMS, Cancer Hospital, Chinese Academy of Medical Sciences

### The results of COJO analysis

To ensure whether genes were independently linked to phenotypes, we performed COJO analysis. The results of COJO analysis were shown in Table S2 and we found that CASP8 and FLACC1 could play a key role in various brain regions, including the amygdala, anterior cingulate cortex, caudate basal ganglia, cortex, frontal cortex, nucleus accumbens basal ganglia, hippocampus, and cerebellar hemispheres in Fig. 2. It has been noted that certain genes, which were genetically regulated, could be the driving force by other genes. Figure 2 demonstrated that in most brain tissues, FLACC1, labeled in blue, was the relatively less significant gene, whereas CASP8, labeled in green, was the most significant gene within the same chromosomal region. This suggested that the effect of CASP8 on the risk of breast cancer was statistically independent and its effect on phenotype was not affected by other genes. In contrast, the effect of FLACC1 on the

**Table 1** The significant genes for breast cancer risk in cross-tissue UTMOST analysis

| Gene symbole | Ensemeble ID | Chromosomes | Test_score | Start | End | Width | P_value | P_fdr |
|---|---|---|---|---|---|---|---|---|
| CASP8 | ENSG00000064012 | 2 | 11.867936 | 2.01E+08 | 2.01E+08 | 128,394 | 5.01E-06 | 0.00089 |
| TNS1 | ENSG00000079308 | 2 | 10.887393 | 2.18E+08 | 2.18E+08 | 234,395 | 1.78E-05 | 0.002388 |
| STRADB | ENSG00000082146 | 2 | 10.079696 | 2.01E+08 | 2.01E+08 | 92,989 | 9.26E-05 | 0.009918 |
| EFR3B | ENSG00000084710 | 2 | 17.177232 | 25,042,076 | 25,159,135 | 117,060 | 3.06E-08 | 1.91E-05 |
| KIF3C | ENSG00000084731 | 2 | 15.747699 | 25,926,598 | 25,982,749 | 56,152 | 9.06E-08 | 4.25E-05 |
| OSM | ENSG00000099985 | 22 | 9.1663835 | 30,262,829 | 30,266,851 | 4023 | 8.00E-05 | 0.008823 |
| SLC35E4 | ENSG00000100036 | 22 | 6.9291161 | 30,635,781 | 30,669,016 | 33,236 | 0.000564 | 0.037875 |
| SLC16A8 | ENSG00000100156 | 22 | 8.9948523 | 38,078,134 | 38,084,184 | 6051 | 7.80E-05 | 0.008823 |
| HSCB | ENSG00000100209 | 22 | 12.262935 | 28,742,039 | 28,757,515 | 15,477 | 1.66E-06 | 0.000366 |
| TOMM22 | ENSG00000100216 | 22 | 7.4151031 | 38,681,957 | 38,685,421 | 3465 | 0.000606 | 0.037875 |
| XBP1 | ENSG00000100219 | 22 | 12.894936 | 28,794,555 | 28,800,597 | 6043 | 1.05E-06 | 0.000262 |
| C22orf31 | ENSG00000100249 | 22 | 20.531281 | 29,058,672 | 29,061,831 | 3160 | 3.49E-10 | 3.27E-07 |
| AP1B1 | ENSG00000100280 | 22 | 13.373215 | 29,327,680 | 29,388,583 | 60,904 | 1.24E-06 | 0.00029 |
| CABP7 | ENSG00000100314 | 22 | 14.65697 | 29,720,003 | 29,731,833 | 11,831 | 3.90E-07 | 0.000126 |
| ASCC2 | ENSG00000100325 | 22 | 13.545938 | 29,788,609 | 29,838,304 | 49,696 | 4.05E-07 | 0.000126 |
| MTMR3 | ENSG00000100330 | 22 | 8.820537 | 29,883,169 | 30,030,868 | 147,700 | 3.66E-05 | 0.004737 |
| DNAJC27 | ENSG00000115137 | 2 | 13.000825 | 24,943,636 | 24,972,094 | 28,459 | 8.67E-07 | 0.000232 |
| POMC | ENSG00000115138 | 2 | 15.792371 | 25,160,853 | 25,168,903 | 8051 | 2.61E-08 | 1.91E-05 |
| FANCL | ENSG00000115392 | 2 | 6.7418857 | 58,159,243 | 58,241,410 | 82,168 | 0.000745 | 0.042986 |
| PRKAG3 | ENSG00000115592 | 2 | 8.0884554 | 2.19E+08 | 2.19E+08 | 9779 | 0.000382 | 0.028107 |
| SUMO1 | ENSG00000116030 | 2 | 8.5338296 | 2.02E+08 | 2.02E+08 | 32,418 | 0.000286 | 0.023316 |
| TNP1 | ENSG00000118245 | 2 | 91.755697 | 2.17E+08 | 2.17E+08 | 607 | 5.95E-11 | 7.44E-08 |
| KLHL29 | ENSG00000119771 | 2 | 6.8728296 | 23,385,179 | 23,708,611 | 323,433 | 0.000895 | 0.048646 |
| TMEM175 | ENSG00000127419 | 4 | 7.0481212 | 932,387 | 958,656 | 26,270 | 0.000726 | 0.042986 |
| TUBA4A | ENSG00000127824 | 2 | 7.4578938 | 2.19E+08 | 2.19E+08 | 28,193 | 0.000604 | 0.037875 |
| VIL1 | ENSG00000127831 | 2 | 7.0790771 | 2.18E+08 | 2.18E+08 | 34,175 | 0.000861 | 0.047497 |
| AAMP | ENSG00000127837 | 2 | 7.1350539 | 2.18E+08 | 2.18E+08 | 6050 | 0.000469 | 0.033811 |
| ADCY3 | ENSG00000138031 | 2 | 246.51319 | 24,819,169 | 24,920,237 | 101,069 | 1.00E-05 | 0.001568 |
| CENPO | ENSG00000138092 | 2 | 7.789083 | 24,793,136 | 24,822,376 | 29,241 | 0.000253 | 0.021101 |
| DTNB | ENSG00000138101 | 2 | 7.9807731 | 25,377,198 | 25,673,647 | 296,450 | 0.000248 | 0.021101 |
| FAM13A | ENSG00000138640 | 4 | 7.125542 | 88,725,955 | 89,111,398 | 385,444 | 0.000597 | 0.037875 |
| DLX1 | ENSG00000144355 | 2 | 8.5125908 | 1.72E+08 | 1.72E+08 | 4938 | 0.000243 | 0.021101 |
| CHST9 | ENSG00000154080 | 18 | 8.8972789 | 26,906,481 | 27,185,308 | 278,828 | 0.000211 | 0.019304 |
| FLACC1 | ENSG00000155749 | 2 | 11.105026 | 2.01E+08 | 2.01E+08 | 69,128 | 3.12E-06 | 0.000651 |
| TMEM237 | ENSG00000155755 | 2 | 7.2748577 | 2.02E+08 | 2.02E+08 | 23,387 | 0.0006 | 0.037875 |
| FZD7 | ENSG00000155760 | 2 | 8.1289412 | 2.02E+08 | 2.02E+08 | 4587 | 0.00014 | 0.013436 |
| LRATD1 | ENSG00000162981 | 2 | 6.9916824 | 14,632,700 | 14,650,814 | 18,115 | 0.000723 | 0.042986 |
| WDCP | ENSG00000163026 | 2 | 11.289658 | 24,029,347 | 24,049,575 | 20,229 | 1.18E-05 | 0.001698 |
| DUSP18 | ENSG00000167065 | 22 | 6.3804357 | 30,652,051 | 30,667,887 | 15,837 | 0.000929 | 0.04977 |
| KCNJ4 | ENSG00000168135 | 22 | 7.956443 | 38,426,327 | 38,455,199 | 28,873 | 0.000306 | 0.02387 |
| CPLX1 | ENSG00000168993 | 4 | 10.874508 | 784,957 | 826,129 | 41,173 | 7.63E-06 | 0.001243 |
| ANTXR1 | ENSG00000169604 | 2 | 9.3671672 | 69,013,176 | 69,249,327 | 236,152 | 0.000108 | 0.01098 |
| AQP4 | ENSG00000171885 | 18 | 8.4754139 | 26,852,043 | 26,865,771 | 13,729 | 0.00022 | 0.019612 |
| HORMAD2 | ENSG00000176635 | 22 | 7.5859602 | 30,080,464 | 30,177,075 | 96,612 | 0.0003 | 0.02387 |
| APOBEC3B | ENSG00000179750 | 22 | 23.866751 | 38,982,347 | 38,992,804 | 10,458 | 9.85E-12 | 3.69E-08 |
| ZBTB12BP | ENSG00000180610 | 4 | 8.6051568 | 39,770,081 | 39,771,371 | 1291 | 0.000205 | 0.019263 |
| PRR34 | ENSG00000182257 | 22 | 7.8980532 | 46,049,478 | 46,054,144 | 4667 | 0.000348 | 0.026065 |
| EWSR1 | ENSG00000182944 | 22 | 12.764608 | 29,268,009 | 29,300,525 | 32,517 | 5.22E-06 | 0.00089 |
| CBX6 | ENSG00000183741 | 22 | 8.8658415 | 38,861,422 | 38,872,249 | 10,828 | 0.000104 | 0.010881 |
| KREMEN1 | ENSG00000183762 | 22 | 14.1233 | 29,073,035 | 29,168,333 | 95,299 | 3.32E-07 | 0.000125 |
| CHEK2 | ENSG00000183765 | 22 | 88.464757 | 28,687,743 | 28,742,422 | 54,680 | 2.91E-11 | 5.46E-08 |
| SMTN | ENSG00000183963 | 22 | 10.546485 | 31,064,105 | 31,104,757 | 40,653 | 4.22E-05 | 0.005273 |
| PLA2G6 | ENSG00000184381 | 22 | 5.8397291 | 38,111,495 | 38,214,778 | 103,284 | 0.000577 | 0.037875 |

**Table 1** (continued)

| Gene symbole | Ensemeble ID | Chromosomes | Test_score | Start | End | Width | P_value | P_fdr |
|---|---|---|---|---|---|---|---|---|
| PTRHD1 | ENSG00000184924 | 2 | 11.517625 | 24,789,728 | 24,793,391 | 3664 | 1.09E-05 | 0.001635 |
| MAFF | ENSG00000185022 | 22 | 7.1764991 | 38,200,767 | 38,216,507 | 15,741 | 0.00067 | 0.041189 |
| PCGF3 | ENSG00000185619 | 4 | 7.5289471 | 705,748 | 770,089 | 64,342 | 0.000136 | 0.013436 |
| CCDC157 | ENSG00000187860 | 22 | 6.7352357 | 30,356,635 | 30,378,673 | 22,039 | 0.000782 | 0.04379 |
| RPL37A | ENSG00000197756 | 2 | 16.001079 | 2.16E+08 | 2.17E+08 | 80,356 | 4.86E-08 | 2.61E-05 |
| MYH9-DT | ENSG00000223695 | 22 | 6.904549 | 36,388,626 | 36,396,517 | 7892 | 0.00053 | 0.037532 |
| LAPTM4A-DT | ENSG00000223734 | 2 | 7.6079342 | 20,052,114 | 20,054,501 | 2388 | 0.000739 | 0.042986 |
| DNAJC27-AS1 | ENSG00000224165 | 2 | 12.732983 | 24,971,390 | 25,039,716 | 68,327 | 8.44E-07 | 0.000232 |
| LINC00570 | ENSG00000224177 | 2 | 9.6856409 | 11,372,612 | 11,403,175 | 30,564 | 6.57E-05 | 0.007701 |
| MTND4P23 | ENSG00000225796 | 2 | 8.1665249 | 2.01E+08 | 2.01E+08 | 1363 | 0.000327 | 0.024995 |
| LEF1-AS1 | ENSG00000232021 | 4 | 9.6815871 | 1.08E+08 | 1.08E+08 | 90,513 | 6.28E-05 | 0.007598 |
| LINC02261 | ENSG00000249699 | 4 | 3.3631447 | 27,217,479 | 27,282,225 | 64,747 | 0.000771 | 0.043784 |
| AQP4-AS1 | ENSG00000260372 | 18 | 11.927423 | 26,655,737 | 27,190,698 | 534,962 | 4.28E-06 | 0.000844 |
| PCAT18 | ENSG00000265369 | 18 | 15.025169 | 26,687,621 | 26,703,638 | 16,018 | 2.59E-07 | 0.000108 |

risk of breast cancer was influenced by CASP8. FLACC1 would also be excluded in subsequent analysis.

### Gene analysis of MAGMA

MAGMA analysis identified 348 significant genes associated with breast cancer (FDR < 0.05) (Table S3). To increase the robustness of the findings, we integrated the results from FUSION and UTMOST analyses to obtain 12 significant shared genes. The detailed results were shown in Fig. 3.

### The results of SMR and colocalization

To verify whether there was a causal relationship between 12 genes and breast cancer, we performed SMR analysis with GTEx v8 data. And we found that the expression of CASP8 was positively correlated with breast cancer only in a wide range of brain tissues including the Anterior cingulate cortex, Cerebellar Hemisphere, Cerebellum, Cortex and Frontal Cortex (bSMR > 0; pSMR < 0.05), and there was no significant heterogeneity in this process (pHEIDI > 0.05) (Table 2). CASP8 was located on chromosome 2 and co-localization analysis showed that most PP.H4 values were greater than 0.9 (Table 3). Among them, rs10197246 was the most significant co-localized locus with breast cancer in the above brain tissues (Fig. 4).

### External clinical cohort validation analyses

We obtained RNA sequencing data from brain metastasis tissues as well as external datasets of brain metastasis. We found that CASP8 had higher expression levels in breast cancer patients with brain metastasis than no brain metastasis in the CHCAMS cohort (Fig. 5A). The same validation results occur in the GSE52604 and GSE46928 cohorts (Fig. 5B, C). In conclusion, CASP8 was differentially expressed in patients with brain metastasis and could play a significant regulatory role in BCBM.

## Discussion

Breast cancer is the most common tumor in women and also has one of the highest mortality rates [29]. A key factor affecting 5-year survival from breast cancer is the occurrence of brain metastasis [30]. The regulatory network of brain metastasis is especially complex. It is particularly important to find key targets of BCBM and explore their potential regulatory mechanisms.

Multi-omics association studies were currently being used to identify hub genes. Yi-Xiao Chen et al. identified key regulators of breast cancer and determined that RNASEH2A was identified as a new candidate gene for breast cancer based on an integrated multi-omics analysis [31]. Meanwhile, Jonathan Beesley used data from the Breast Cancer Society Consortium and eQTL to identify shared genetic relationships, with 17 genes, identified as potential mediators of breast cancer [32]. Moreover, Zhihao Zhang identified genes associated with breast cancer by integrating GWAS, eQTL, and mQTL, and ultimately found that the association of ATG10 and RCCD1 with breast cancer was significant [33]. Variations in results might be attributed to differences in sample size and sources, as well as differences between the methods used. However, this emphasized the significance of performing multiple ways of identifying potential genes contributing to breast cancer.

In our study, we integrated eQTL data from the GTEx project and GWAS from the Finnish R11 database to conduct cross-tissue TWAS. Meanwhile, unlike previous studies, we performed multiple TWAS approaches including FUSION, UTMOST, and MAGMA to identify candidate genes, which ensured the stability of our results. Subsequently, the results of both SMR and co-localization analysis have shown that across multiple tissues, CAPS8 was only expressed in multiple brain tissues associated with breast cancer. The above multiple genetic evidence demonstrated that CASP8 had a strong link

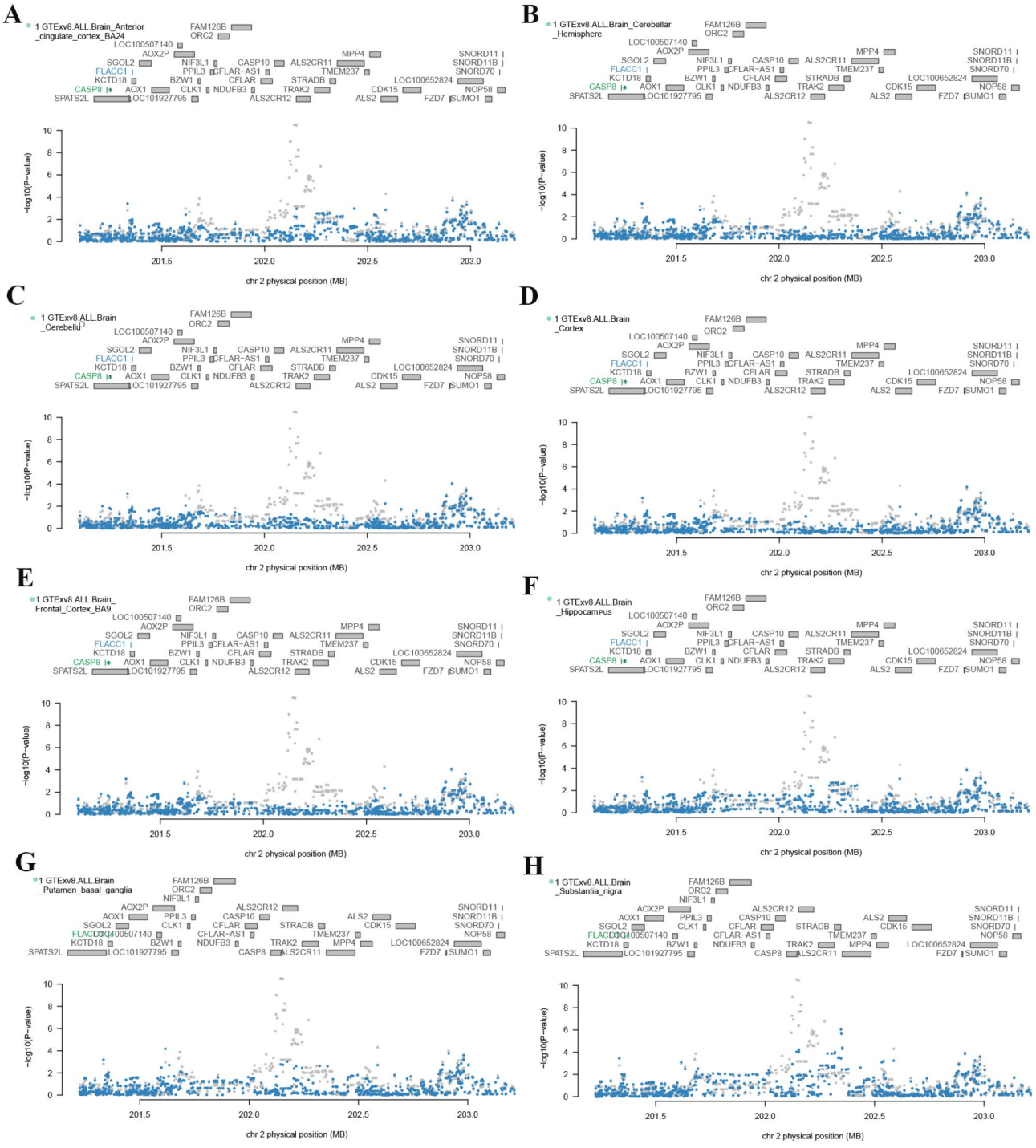


**Fig. 2** Transcriptome-wide conditional and joint analysis. Genes independently associated with breast cancer are highlighted (green) and genes not independently associated with breast cancer are highlighted (blue) in the top of Fig. 2. SNPs associated with breast cancer before conditional analysis are highlighted (gray), and secondary SNPs associated with breast cancer after conditional analysis are highlighted (blue) in the bottom of Fig. 2

between breast cancer and the nervous system. Previous studies have also shown that brain metastasis occurred in approximately 50% of HER2-positive breast cancer and common site of metastasis for breast cancer was the brain [34]. Sites of metastasis from breast cancer with brain metastasis include parenchymal metastasis, meningeal metastasis, and both (3.6% incidence, but worst prognosis) [35]. Breast cancer and the nervous system were strongly linked [36]. With the above finding, we have sufficiently inferred that potential genetic regulation

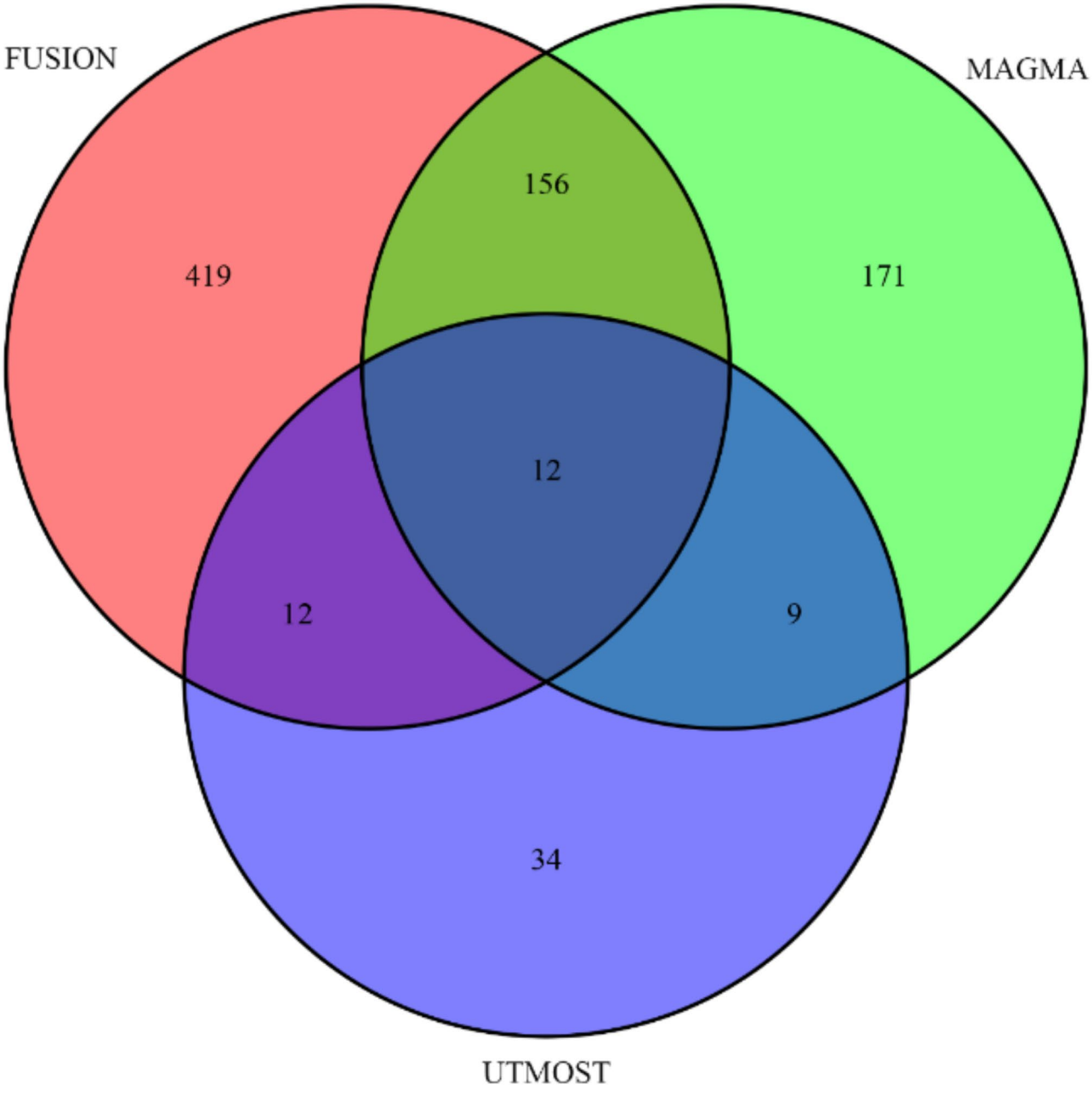


**Fig. 3** Venn diagram of shared genes identified by FUSION, UTMOST, and MAGMA

**Table 2** Outcomes of summary data-level Mendelian randomization analysis

| Tissue | Gene | b_SMR | se_SMR | p_SMR_multi | p_HEIDI | nsnp_HEIDI |
|---|---|---|---|---|---|---|
| Anterior_cingulate_cortex | CASP8 | 0.0847715 | 0.0165362 | 2.95E-07 | 3.82E-01 | 20 |
| Cerebellar_Hemisphere | CASP8 | 0.0783516 | 0.0138049 | 8.18E-08 | 7.03E-01 | 20 |
| Cerebellum | CASP8 | 0.0731566 | 0.0128932 | 1.39E-08 | 7.44E-01 | 20 |
| Cortex | CASP8 | 0.088889 | 0.0158294 | 1.96E-08 | 8.23E-01 | 20 |
| Frontal_Cortex | CASP8 | 0.0809234 | 0.0140482 | 8.39E-09 | 7.45E-01 | 20 |

**Table 3** The results of bayesian colocalization analysis

| Gene symbole | Panel | Nsnps | PP.H0 | PP.H1 | PP.H2 | PP.H3 | PP.H4 |
|---|---|---|---|---|---|---|---|
| CASP8 | Anterior_cingulate_cortex | 4887 | 4.86E-16 | 2.07E-07 | 2.06E-10 | 0.086817826 | 0.91318197 |
| | Cerebellar_Hemisphere | 4865 | 3.40E-23 | 5.49E-08 | 1.44E-17 | 0.022270628 | 0.97772932 |
| | Cerebellum | 4776 | 7.55E-23 | 5.18E-08 | 3.20E-17 | 0.020959575 | 0.97904037 |
| | Cortex | 4885 | 2.05E-22 | 5.24E-08 | 8.70E-17 | 0.021244036 | 0.97875591 |
| | Frontal_Cortex | 4891 | 1.91E-25 | 5.20E-08 | 8.10E-20 | 0.021073049 | 0.9789269 |
| | Hippocampus | 4820 | 5.03E-08 | 9.10E-08 | 0.02130536 | 0.037605456 | 0.94108904 |
| | Hypothalamus | 4876 | 8.26E-08 | 9.53E-08 | 0.03498323 | 0.039444224 | 0.92557237 |
| | Nucleus_accumbens_basal_ganglia | 4882 | 3.32E-07 | 3.09E-07 | 0.14054559 | 0.130195884 | 0.72925789 |

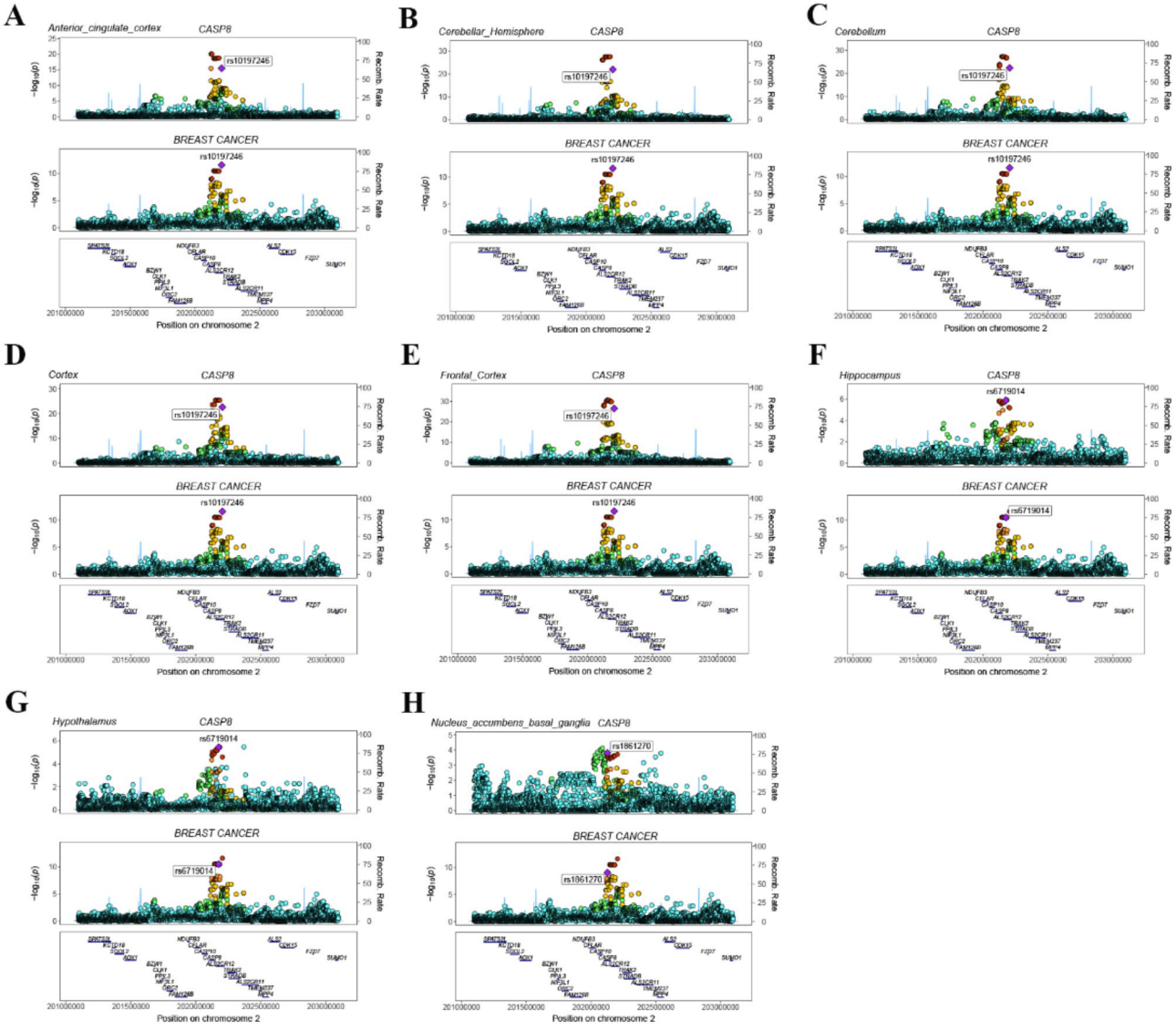


**Fig. 4** Co-localization results between CASP8 and breast cancer in various brain tissues

of CASP8 could occur in breast cancer brain metastasis. Ultimately, we used the BCBM dataset as well as our self-assessment dataset for external validation and found that CASP8 was more highly expressed in BCBM, and CASP8 might play an important role in BCBM.

CASP8 encoded a member of the cysteine-aspartate protease family [37]. The activation of cyst-aspartase played a pivotal role in the execution phase of apoptosis. CASP8 was involved in programmed cell death induced by Fas and various apoptotic stimuli [38]. CASP8 was

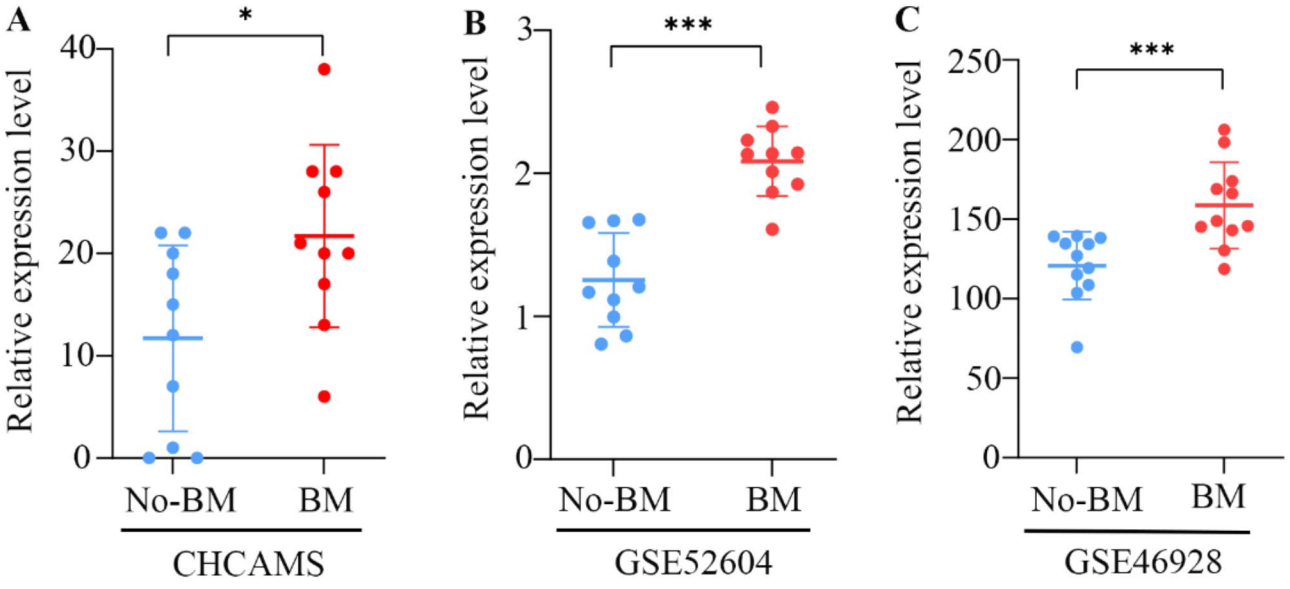


**Fig. 5** The expression level of CASP8 in No-BM and BM tissues of the CHCAMS cohort, GSE52604 and GSE46928. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

detected in insoluble fractions of affected brain regions in Huntington's disease patients, suggesting a role in neurodegenerative disease. Previous studies have identified the co-occurrence of regulatory and dysfunctional T cells in a tumor microenvironment that proliferated cells and was enriched for BRCA1 and CASP8 mutations and predicted poor outcomes in breast cancer [39]. In addition, a study conducted by Irene Catucci et al. found that the rs3834129 polymorphism in the promoter region of the CASP8 gene was associated with breast cancer [40]. However, the association of the rs10197246 polymorphism in the CASP8 gene region with breast cancer has been poorly reported. Although there have been some reports finding that CASP8 might be associated with the occurrence of breast cancer, there was a lack of research on BCBM, which was remedied by our study.

Finally, our study identified novel genes associated with breast cancer with various TWAS analysis and determined the correlation between CASP8 and BCBM by SMR and co-localization, and subsequently validated this finding by multiple cohort validation with BCBM, which showed the rigor of our findings and contributed to the stability of our conclusions. However, our study had several limitations. Firstly, due to the criteria of significant cis-genetic power genes, not all genes could be captured and those SNP that affect breast cancer but were independent of cis-expression would be ignored. Secondly, although the finding was validated by genetic analysis and multiple external cohorts, the validation was at the expression level and a series of biological experiments would be required to be performed to confirm our conclusion. In conclusion, this study provided new perspectives and genetic insights in the effect of CASP8 on the occurrence and progression of BCBM.

## Conclusion

The aim of this study was to identify causal genes driving brain metastasis in breast cancer by integrating multi-tissue TWAS and mendelian randomization. And the above results indicated that the expression of CASP8 was associated with BCBM, which provided new insights into the genetic architecture of BCBM. This provided a target for subsequent treatment of BCBM.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14095-y.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

**Acknowledgements**

The authors are grateful for the help of all colleagues and the public database.

**Author contributions**

Jinsong Liu: design of the study, data analysis, and manuscript writing. Songlin Gao and Xiao Guan: data collection and formal analysis. Liuliu Quan, Min Dou and Jian Yue: data curation, funding acquisition and figures plotted. Peng Yuan and Mengwu Shi: project administration, supervision, and validation. The final manuscript received the approval of all the authors.

**Funding**

This study was supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2023-JKCS-23 to P.Y.) and the Special Research Fund for Central Universities, Peking Union Medical College. [No.2022-I2M-C&T-A-014, CAMS Innovation Fund for Medical Sciences(CIFMS) to P.Y.].

**Data availability**
The original information supporting the assay's conclusions can be obtained through getting in touch with the author.

## Declarations

**Ethical approval**
The study was approved by the Ethical Review Committee, Cancer Hospital of the Chinese Academy of Medical Sciences and the patient has given informed consent. Our research adheres to the declaration of Helsinki and is in line with the fundamental principles of scientific research. Consent for publication is not applicable.

**Competing interests**
The authors declare no competing interests.

Received: 18 November 2024 / Accepted: 7 April 2025
Published online: 16 April 2025

**References**

1. Bi Y, et al. An ultra-sensitive and rapid immunosensor for the onsite detection of Circulating tumor DNA in breast cancer. Front Bioeng Biotechnol. 2024;12:1412598.
2. Al-Alem U et al. Understanding the role of Toll-Like receptors 9 in breast cancer. Cancers (Basel), 2024. 16(15).
3. Sun X, et al. CILP, a putative gene associated with immune infiltration in breast cancer brain metastases. Front Genet. 2022;13:862264.
4. Xavier JM, et al. Identification of candidate causal variants and target genes at 41 breast cancer risk loci through differential allelic expression analysis. Sci Rep. 2024;14(1):22526.
5. Jia G, et al. Identification of target proteins for breast cancer genetic risk loci and blood risk biomarkers in a large study by integrating genomic and proteomic data. Int J Cancer. 2023;152(11):2314–20.
6. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190–5.
7. Tam V, et al. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467–84.
8. Gamazon ER, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47(9):1091–8.
9. Hu Y, et al. A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet. 2019;51(3):568–76.
10. Gui J, et al. Identification of novel proteins for sleep apnea by integrating genome-wide association data and human brain proteomes. Sleep Med. 2024;114:92–9.
11. de Leeuw CA, et al. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219.
12. Kurki MI, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508–18.
13. Consortium GT. The Genotype-Tissue expression (GTEx) project. Nat Genet. 2013;45(6):580–5.
14. Barrett T, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41(Database issue):D991–5.
15. Mai J, et al. Transcriptome-wide association studies: recent advances in methods, applications and available databases. Commun Biol. 2023;6(1):899.
16. Sun R, et al. Powerful gene set analysis in GWAS with the generalized Berk-Jones statistic. PLoS Genet. 2019;15(3):e1007530.
17. Gusev A, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245–52.
18. Li SJ, et al. Identifying causal genes for migraine by integrating the proteome and transcriptome. J Headache Pain. 2023;24(1):111.
19. Ren S, et al. Gaining new insights into the etiology of ulcerative colitis through a cross-tissue transcriptome-wide association study. Front Genet. 2024;15:1425370.
20. Liao C, et al. Transcriptome-wide association study of attention deficit hyperactivity disorder identifies associated genes and phenotypes. Nat Commun. 2019;10(1):4450.
21. de Leeuw CA, et al. Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. Nat Commun. 2018;9(1):3768.
22. Yuan F, et al. Fine mapping-based multi-omics analysis interprets the gut-lung axis function of SGLT2 inhibitors. Front Cell Infect Microbiol. 2024;14:1447327.
23. Zhu Z, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48(5):481–7.
24. Porcu E, et al. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat Commun. 2019;10(1):3300.
25. Giambartolomei C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383.
26. Yang Y, et al. Epigenetic and integrative cross-omics analyses of cerebral white matter hyperintensities on MRI. Brain. 2023;146(2):492–506.
27. Tao T, Mo X, Zhao L. Identifying novel potential drug targets for endometriosis via plasma proteome screening. Front Endocrinol (Lausanne). 2024;15:1416978.
28. Liu J, et al. A novel m6A/m5C/m1A score signature to evaluate prognosis and its immunotherapy value in colon cancer patients. J Cancer Res Clin Oncol. 2023;149(13):11995–2012.
29. Liu S, et al. TP53AIP1 induce autophagy via the AKT/mTOR signaling pathway in the breast cancer cells. Cancer Biol Ther. 2024;25(1):2398297.
30. Zhao YY et al. Density of tertiary lymphoid structures predicts clinical outcome in breast cancer brain metastasis. J Immunother Cancer, 2024. 12(7).
31. Chen YX, et al. An integrative multi-omics network-based approach identifies key regulators for breast cancer. Comput Struct Biotechnol J. 2020;18:2826–35.
32. Ferreira MA, et al. Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. Nat Commun. 2019;10(1):1741.
33. Zhang Z, et al. Multi-omics Mendelian randomization integrating GWAS, eQTL, and mQTL data identified genes associated with breast cancer. Am J Cancer Res. 2024;14(3):1433–45.
34. Santos L et al. ENPP1 induces blood-brain barrier dysfunction and promotes brain metastasis formation in HER2-positive breast cancer. Neuro Oncol, 2024.
35. Karthik J, et al. Navigating breast cancer brain metastasis: risk factors, prognostic indicators, and treatment perspectives. World J Clin Oncol. 2024;15(5):594–8.
36. Mo H, Zhang X, Ren L. Analysis of neuroglia and immune cells in the tumor microenvironment of breast cancer brain metastasis. Cancer Biol Ther. 2024;25(1):2398285.
37. Mandal R, et al. Caspase-8: the double-edged sword. Biochim Biophys Acta Rev Cancer. 2020;1873(2):188357.
38. Sena-Dos-Santos C, et al. Association of apoptosis-related variants to malaria infection and parasite density in individuals from the Brazilian Amazon. Malar J. 2023;22(1):295.
39. Danenberg E, et al. Breast tumor microenvironment structures are associated with genomic features and clinical outcome. Nat Genet. 2022;54(5):660–9.
40. Catucci I, et al. The CASP8 rs3834129 polymorphism and breast cancer risk in BRCA1 mutation carriers. Breast Cancer Res Treat. 2011;125(3):855–60.

## Publisher's note