

University of Nottingham
UK | CHINA | MALAYSIA

Germline determinants of the immune response in Asian breast cancer

Beverley Yap Ling Shan, BSc

20474573

Thesis submitted to the University of Nottingham Malaysia for the Master of Philosophy in Biosciences

December 2023

Acknowledgements

I would like to first thank God for His favour and strength. On days where I grow tired, to the brightest of days, and every day in between, His grace has been more than sufficient for me.

I could not have completed my Master's journey without the support from my incredible supervisors. To Dr Tai Mei Chee and Dr Pan Jia Wern, thank you for every weekly catch-up, for your patient guidance as I picked up all things genomics and bioinformatics for the first time, and for every word of encouragement. The two of you are not just first-class cancer researchers, but also amazing leaders. To Dr Ho Weang Kee, you are kind, thoughtful, and an all-round brilliant statistician. Finally, Dr Teo Soo Hwang, thank you for the space you have given me to explore, make mistakes and discover who I am as a scientist and person at Cancer Research Malaysia. I have said this for the last 4 years, and I will say it again: you are my inspiration. To all supervisors, thank you for believing in me.

This study is a testament to the invaluable contributions made by my colleagues at Cancer Research Malaysia. To Dr Yip Cheng Har, Prof Aishah Mohd Taib, Kartini Rahmat, Yook Sook Yee, Tiara Hassan, Eldarina Azfar Wijaya, Siti Norhidayu Hassan, Patsy Ng Pei Sze, and Joanna Lim for their work in the MyBrCa Study. To Mamduh Zabidi and Stephanie Meng who collected and analysed the breast tumour data, and Haslina Makmur for generating the immune scores analysed in this study. I thank the Caldas Lab and the Core Genomics Facility at the CRUK Cambridge Institute for their sequencing work. I acknowledge and thank the Malaysian Ministry of Higher Education for funding the MyBrCa Study.

To my parents and family, thank you for your endless love, advice, and support. To Gary and my wonderful friends, thank you for cheering me on and celebrating every milestone. I wouldn't be who I am today if it were not for you.

Abstract

Breast cancer is a complex disease with heterogeneity in tumour characteristics and treatment response. Recently, Asian breast cancer patients were reported to have elevated immune scores in comparison to Caucasian breast cancer patients, however, germline genetic determinants associated with immune scores in Asian breast cancer patients remain undercharacterised. This study aims to identify the germline genetic variants associated with heritable immune scores in Asian breast tumours and determine the cellular pathways that drive immune scores within this population.

By evaluating 112 immune scores that have previously been described in literature, we observed significantly higher heritability in Asian breast tumours compared to Caucasians. The major histocompatibility complex class-I (MHC-I) immune score had high heritability in both cohorts. By conducting a genome-wide association study (GWAS), the findings revealed, as predicted, the *HLA* locus on chromosome (chr) 6 in both Asian and Caucasian cohorts.

Regulatory T (Treg) cells, Immune non-small cell lung cancer (NSCLC) score, Triggering receptor expressed on myeloid cells 1 (TREM1) score and T follicular helper (Tfh) cells were the most heritable immune scores in Asian breast tumours. A total of 37 candidate SNPs were identified through GWAS. Expression quantitative trait loci (eQTL) analysis and a review of current literature revealed variants in *TNFRSF13B*, *PRKCQ*, *IL2RA*, and *MIR149* as candidate *cis*-eQTL loci with potential roles in immune function and immune scores. Taken together, these findings suggest that there may be a population-specific genetic influence on heritable immune traits. This study identifies candidate variants that may underlie these observed differences.

Table of Contents

Acknowledgements 1

Abstract 2

Table of Contents 3

List of Tables 5

Abbreviations 8

Chapter 1 : Introduction 10

1.1 Breast cancer incidence 10

1.2 Immunotherapeutic strategies 11

1.3 Immune scores as a cancer prognostics tool 12

1.4 Germline genomic landscape of breast tumours 14

1.5 Research Aims & Objectives 15

Chapter 2 : Methodology 17

2.1 Sample selection 17

2.2 Genotyping procedures 17

2.3 Germline genotype data 18

2.4 RNA Extraction, sequencing, alignment, and QC 18

2.5 Immune scores 19

2.6 Immune trait normalisation 20

2.7 Genome-wide heritability analysis 20

2.8 SNP-association test 21

2.9 Genome-wide association study 21

2.10 Expression quantitative trait loci (eQTL) analysis 22

2.11 Statistical Analysis 23

Chapter 3 : Results 24

3.1 Genome-wide heritability of 112 immune traits 24

3.2 Genome-wide association analysis of highly heritable immune traits 28

3.3 Genome-wide association analysis of highly heritable immune traits in Asian breast tumours 34

3.4 Expression quantitative trait loci analysis 44

Chapter 4 : Discussion 58

4.1 Summary of main findings 58

4.2 Genome-wide heritability of immune scores in breast tumours 58

4.3 Shared genomic loci in highly heritable immune traits in Asian and Caucasian breast tumours 59

4.4 Most heritable immune traits in Asian breast tumours 60

4.5 Strengths 64

4.6 Limitations 64

Chapter 5 : Conclusion 65

Chapter 6 : References 66

List of Figures

Figure 1. Overview of study approach 26

Figure 2. Genome-wide heritability of immune traits 27

Figure 3. Genome-wide association analysis of MHC.I_19272155 in MyBrCa and TCGA populations. 31

Figure 4. Genome-wide association analysis of MHCI_21978456 in MyBrCa and TCGA populations. 32

Figure 5.1. Manhattan plot of Treg cells in the MyBrCa population (n = 960). 35

Figure 5.2. Regional association plots for candidate SNPs of Treg cells in the MyBrCa population. 36

Figure 5.3. Manhattan plot of Immune NSCLC Score in the MyBrCa population (n = 960). ... 37

Figure 5.4. Regional association plots for candidate SNPs of Immune NSCLC Score in the MyBrCa population. 38

Figure 5.5. Manhattan plot of TREM1 score in the MyBrCa population (n = 960). 39

Figure 5.6. Regional association plots for candidate SNPs of TREM1 score in the MyBrCa population. 40

Figure 5.7. Manhattan plot of Tfh cells in the MyBrCa population (n = 960). 41

Figure 5.8. Regional association plots for candidate SNPs of Tfh cells in the MyBrCa population. 42

Figure 6.1. eQTpLot of the colocalization of eQTLs for gene *TNFRSF13B* and GWAS signals for Treg cells on candidate SNP chr17:16764104 in the MyBrCa population. 45

Figure 6.2. eQTpLot of the colocalization of eQTLs for gene *PRKCQ* and GWAS signals for Treg cells on candidate SNP chr10:6499101 in the MyBrCa population. 46

Figure 6.3. eQTpLot of the colocalization of eQTLs for gene *IL2RA* and GWAS signals for Treg cells on candidate SNP chr10:6499101 in the MyBrCa population. 47

Figure 6.4. eQTpLot of the colocalization of eQTLs for gene MIR149 and GWAS signals for TREM1 score on candidate SNP chr2:241431311 in the MyBrCa population. 48

Figure 6.5. eQTpLot of the colocalization of eQTLs for gene LHX4 and GWAS signals for Treg cells on candidate SNP chr1:180194037 in the MyBrCa population. 49

Figure 6.6. eQTpLot of the colocalization of eQTLs for gene SNX25 and GWAS signals for Treg cells on candidate SNP chr4:185964047 in the MyBrCa population. 50

Figure 6.7. eQTpLot of the colocalization of eQTLs for gene ZNF624 and GWAS signals for Treg cells on candidate SNP chr17:16764104 in the MyBrCa population. 51

Figure 6.8. eQTpLot of the colocalization of eQTLs for gene EFEMP1 and GWAS signals for Immune NSCLC score on candidate SNP chr2:55917604 in the MyBrCa population. 52

Figure 6.9. eQTpLot of the colocalization of eQTLs for gene LINC01813 and GWAS signals for Immune NSCLC score on candidate SNP chr2:55917604 in the MyBrCa population. 53

Figure 6.10. eQTpLot of the colocalization of eQTLs for gene SMC2-AS1 and GWAS signals for Immune NSCLC score on candidate SNP chr9:107044828 in the MyBrCa population. 54

Figure 6.11. eQTpLot of the colocalization of eQTLs for gene C10orf76 and GWAS signals for Tfh cells on candidate SNP chr10:103933363 in the MyBrCa population. 55

Figure 6.12. eQTpLot of the colocalization of eQTLs for gene RPS26P5 and GWAS signals for Tfh cells on candidate SNP chr21:17948888 in the MyBrCa population. 56

List of Tables

Table 1. Gene sets for selected immune traits**.** 30

Table 2. Validated SNPs for MHC.I_19272155 and MHC1_21978456. 33

Table 3. Candidate SNPs for Treg cells, Immune NSCLC, TREM1 and Tfh cells. 43

Table 4. GWAS summary statistics of variants associated with TNFRSF13B, IL2RA,PRKCQ and MIR149. 57

Abbreviations

AMR	American
APRIL	A proliferation-inducing ligand
BAFF	B-cell activating factor
CAML	Calcium-modulating cyclophilin ligand
CCDC134	Coiled-coil domain containing 134 protein
eQTL	Expression quantitative trait loci
EAS	East Asian
EMT	Epithelial to mesenchymal transition
ESTIMATE	Estimation of Stromal and Immune Cells in Malignant Tumours
EUR	European
GATK	Genome analysis toolkit
GCTA	Genome-wide complex trait analysis
GEP	Gene expression profile
GREML	Genome-based restricted maximum likelihood statistical approach
GSVA	Gene set variation analysis
GWAS	Genome-wide association study
HLA	Human leukocyte antigen
ICB	Immune checkpoint blockade
IFN	Interferon
IMPRES	Immune-predictive score
LD	Linkage disequilibrium
M/I	Mortality-to-incidence ratio
MAF	Minor allele frequency
MAPK	Mitogen-activated protein kinase
Mb	Megabase
MHC	Major histocompatibility complex
MyBrCa	The Malaysian Breast Cancer Genetic Study
NES	Normalized effect size
NSCLC	Non-small cell lung cancer
OS	Overall survival

PKC	Protein Kinase C
RNA-seq	RNA sequencing
SJMC	Subang Jaya Medical Centre
SNP	Single-nucleotide polymorphism
TCGA	The Cancer Genome Atlas
Tfh	Follicular helper T cells
TIL	Tumour-infiltrating lymphocyte
TNBC	Triple-negative breast cancer
TPM	Transcripts-per-million
Treg	Regulatory T cells
TREM1	Triggering receptor expressed on myeloid cells-1
UMMC	University Malaya Medical Centre
V(G)/V(p)	Proportion of genotypic variance to phenotypic variance

Chapter 1 : Introduction

1.1 Breast cancer incidence

Cancer is widely recognized as a global health concern due to its rapid increase in incidence and mortality. In 2020, cancer was reported as the first or second leading cause of death in 112 of 183 countries for patients ages 70 and below ([1]). From this global statistic, breast cancer was reported as the most diagnosed cancer with a new record high of 2.3 million new cases (11.7%), and the fifth (6.9%) leading cause of cancer-related deaths in 2020 ([1]). While global incidence continues to surge, the burden of breast cancer varies by region and country.

In Asia, breast cancer is a growing concern as women residing in this region accounted for 45.4% of all breast cancer diagnoses worldwide in 2020 ([1, 2]) and has experienced the greatest increase in incidence over the last two decades ([3]). Worryingly, the mortality-to-incidence ratio (M/I) of breast cancer in Asia was higher than the world's average M/I ratio, and ranked second globally ([2]). Mortality rates have been shown to be especially higher in low-to-middle income countries as compared to high-income countries ([1, 2]) due to disparities in national cancer control strategies. Low-to-middle income countries are more likely to lack access to early detection screening programs, treatment options, and have inadequate medical facilities, thus resulting in a higher prevalence of delayed diagnoses that are much harder and more costly to treat ([4-7]). In Malaysia, breast cancer has been the most commonly diagnosed cancer since 2007 ([8]) and accounted for 32.9% of all reported cancer cases among females in 2020. Furthermore, it was also reported as the second leading cause of cancer-related deaths (11.9%) in Malaysia in the same year.

The rising incidence of mortality and incidence rates of breast cancer in Asia, coupled with the lack of effective treatment options, underscores the critical need to explore and implement innovative therapies and treatment strategies to curb its prevalence.

1.2 Immunotherapeutic strategies

The host immune system has been shown to play a significant role in the anti-tumour immune response and the prevention of tumour formation and progression ([9]). Thus, there has been great interest in developing effective immunotherapeutic strategies to treat and prevent breast cancer recurrence in patients. Examples of such therapies include CTLA-4 and PD-1/PD-L1, which are immune checkpoint blockade (ICB) therapies that stimulate the host immune system to recognize and prevent immune evasion of cancer cells by killing them ([10, 11]). PD-1/PD-L1 blockades, such as nivolumab, atezolizumab, durvalumab and avelumab, have shown particularly strong therapeutic promise when used in combination with conventional cancer therapies such as chemotherapy and radiotherapy. In the randomized phase 3 clinical trial IMpassion130, the trial reported that patients with PD-L1 immune cell-positive disease receiving a combined use of atezolizumab with nab-paclitaxel (a drug used in chemotherapy) had significant improvements in overall survival (OS) compared to those in the placebo group receiving only nab-paclitaxel ([12-14]). In the KEYNOTE-355 trial where 1,372 triple-negative breast cancer (TNBC) patients were randomized to receive either a pembrolizumab-chemotherapy treatment or a placebo-chemotherapy treatment, patients in the former group had a significant improvement in progression-free survival in less than 2 years ([15]). Findings from these clinical trials have shed light to a possible role for the addition of immunotherapy to the standard care of treatment for cancer patients. These results are especially meaningful

for TNBC patients, where responses to chemotherapy are often short-lived with a median OS of less than 18 months ([16, 17]).

ICB therapy has most certainly revolutionised the field of cancer immunotherapy. However, despite its promising results, only less than 15% of patients respond to treatment ([18]). In the context of breast cancer, the efficacy of immunotherapy is limited to only a small subset of patients as breast cancer is typically immunologically quiescent compared to other cancer types ([19, 20]). There has been strong evidence to show that the success of immunotherapy varies patient-to-patient and is largely influenced by intrinsic differences across host immune systems, gut microbiota, immune status and also by tumour-related characteristics such as its mutational load, specific somatic genetic mutations, location, and more ([21-26]). Tumour resistance to immunotherapy could, in turn, result in lower quality of life, reduced OS and also an increased economic burden on the healthcare system ([27]). Therefore, understanding the underlying cellular mechanisms influencing therapeutic response and patient outcome is critical in the development of effective, targeted immunotherapy for breast cancer patients and in patient stratification.

1.3 Immune scores as a cancer prognostics tool

A cancer patient's disease progression and outcome have conventionally been predicted through the assessment of tumour histopathological features, but this has shown limited prognostic accuracy as cancer outcomes can vary significantly among patients ([28]). In more recent times, molecular profiling has revealed underlying gene expression, copy number number and mutational landscape that can be used for more accurate prognostication ([29, 30]). Through the characterization of the tumour microenvironment and its relationship to the immune system, molecular-based prognostics tools, such as immune scores, have also been

developed to predict the likelihood of cancer recurrence, progression, and patient prognosis ([31]). An immune score is a statistical estimate that provides insight into the activity and status of the immune system in the presence of a tumour ([32]). It is typically derived from gene expression data, but the selection of genes used would vary from score to score depending on the aspect of the immune system being studied ([33-37]). Typically, an elevated immune score corresponds to higher levels of immune activity and response, lower tumour heterogeneity and thus, better overall survival ([38]).

The immune scoring system as a prognostic marker was first established in colorectal cancer, wherein T-lymphocyte populations were quantified in tumours and given a score, called the Immunoscore, that predicts patient survival ([39-41]). The Immunoscore has set the stage for the emergence of new molecular prognostics tools incorporating multiple classes of immune cells and traits across multiple cancers. A study by Auslander et al., (2018) showed that through the use of an immune-predictive score (IMPRES), they were able to predict patient response to ICB therapies in melanoma ([36]) while Yoshihara et al., (2013) developed a method known as the Estimation of Stromal and Immune Cells in Malignant Tumours using Expression data (ESTIMATE) score to measure tumour purity by inferring the level of infiltrating stromal and immune cells in tumour tissues ([35]). Additionally, T-cell-inflamed gene expression profiles (GEPs), also known as IFN-γ-related mRNA profiles have been reported to predict clinical response to PD-1 blockade ([37]). The implementation of immune scores could contribute to precision medicine and facilitate the development of personalised treatment plans for breast cancer patients.

1.4 Germline genomic landscape of breast tumours

Studies thus far have extensively investigated the influence of germline genetic variant-associated immune traits in host immune response to diseases ([42-45]). In addition, a few studies have also characterised the influence of germline genetic variants on the tumour immune microenvironment. Genome-wide association studies (GWAS) thus far have identified host germline genetic factors that are associated with shaping the cancer immune microenvironment ([25, 46-48]) and an expression quantitative trait loci (eQTL) discovery analysis by Lim et al., (2018) defined the germline genetic determinants associated with tumour-infiltrating lymphocytes (TILs) across 24 cancer types ([49]). A pan-cancer GWAS carried out by Shahamatdar et al., (2020) investigated 17 immune infiltration phenotypes and identified a variant in *SIK1* that was associated with the amount of follicular helper T (Tfh) cells found in bulk tumour, 23 candidate genes associated with cytokine-mediated signalling pathways and networks containing genes that impact DNA repair and transcription elongation pathways ([46]). Most recently, Sayaman et al., (2021) uncovered the common and rare genetic determinants that affect the tumour immune microenvironment, and the mechanisms and modulatory genes involved ([47]). Common variants of *IFIH1*, *TMEM173* (*STING1*), and *TMEM108* were found to be associated with differential interferon (IFN) signalling and variants of RBL1 were linked to differential T cell infiltration of tumour tissue. Rare cancer predisposition variants reportedly affected different immunomodulatory pathways that regulate anti-tumour immunity ([47, 50]).

Breast cancer is a complex and heterogenous disease that poses significant challenges for diagnosis and treatment. While significant progress has been made in characterising the molecular landscape of breast tumours in Caucasian populations, much less is known about

the disease in Asian patients. There is therefore a significant need for advancing breast tumour cell characterisation and developing immunotherapeutic options within this region, as this will lead to improved survival and an increased overall quality of life amongst Asian breast cancer patients. We recently reported that compared to Caucasians, Asian breast tumours exhibited an elevated immune score (Pan et al., (2020), ([48])), which was associated with better overall survival and prognosis, however its germline genetic drivers remain to be uncovered. This study evaluated the association between germline genetic variants and its effect on immune scores and determines the plausible cellular pathways underlying different immune scores in a population of Asian breast cancer patients as a follow-up study from Pan et al., (2020).

1.5 Research Aims & Objectives

1.5.1 Aim

The aim of this research study is to evaluate the role and contribution of germline genetic variants in shaping the immune landscape of Asian breast tumours.

1.5.2 Objectives

a. Identify the germline genetic determinants associated with elevated immune scores in a population of Asian breast cancer patients

In a recent study by Pan et al., (2020) at Cancer Research Malaysia, we reported that compared to Caucasians, Asian breast tumours exhibited an enriched immune score, which was associated with better overall prognosis and survival ([48]). Here in this analysis, RNA sequencing (RNA-seq) data and gene expression profiles of Malaysian breast cancer patients from the Malaysian Breast Cancer Genetic study (MyBrCa) cohort were used to determine the germline genetic variants that were associated with enriched immune scores in Asian breast tumours. This objective was accomplished by first calculating the genome-wide heritability for immune

traits from common genetic variants, followed by a GWAS and eQTL analysis to identify the loci with the strongest effect on heritable immune scores in Asian breast tumours.

b. Determine the cellular pathways underlying the variation in immune scores among Asian breast cancer patients

This study aims to explore the molecular mechanisms, such as functional and genetic pathways, that are associated with genome-wide significant and suggestive loci found among heritable immune traits in this analysis within the MyBrCa cohort. In doing so, we hope to better understand the genetic influences that modulate the immune response in Asian breast tumour immune profiles and shed light to potential strategies to treat the disease in the future.

Chapter 2 : Methodology

2.1 Sample selection

The Malaysian Breast Cancer Genetic (MyBrCa) Study is a case-control study of breast cancer risk factors among Asian female breast cancer patients receiving treatment at Subang Jaya Medical Centre (SJMC) and University Malaya Medical Centre (UMMC), Malaysia. For SJMC, women were recruited from 2012 to 2017, and for UMMC, women were recruited from 2002 to 2019 all women provided written informed consent prior to participating in the study. Matching blood samples were obtained from patients prior to surgery to extract for germline genomic DNA, while breast tumours were sectioned during surgery and immediately frozen and stored in liquid nitrogen for a proportion of breast cancer patients to extract tumour RNA ([48, 51]). A total of 1023 participants were included in this study. After filtering (refer to Methods section 2.2 – 2.3), the final sample size was 960 participants. The MyBrCa study was approved by the Sime Darby Medical Centre Independent Ethics Committee (reference no: 201109.4 and 201208.1) and the University Malaya Ethics Committee (reference no: 842.9). The Cancer Genome Atlas Breast (TCGA) dataset comprises of tumour and matched blood normal samples from 1102 Caucasian breast cancer patients ([52]). There was a total of 976 samples after filtering.

2.2 Genotyping procedures

Germline DNA samples from the MyBrCa study were genotyped using -533,000-SNP OncoArray ([53]). Genotype calling, quality control procedures and imputation was described previously ([54-56]). In brief, individuals were excluded if samples had a low call rate (<95%), were not in Hardy-Weinberg equilibrium ($p<1x10^{-12}$), had extreme heterozygosity ($p<1x10^{-5}$) or were found to have non-female genotypes (non-XX). Additionally, samples were also excluded

if they had discordant genotypes or were cryptic duplicates despite phenotype data showing the individuals were unrelated. For cryptic duplicates, the sample with a lower call rate was excluded. For first-degree relative pairs, the controls were excluded, otherwise the sample with a lower call rate was excluded. All remaining samples were imputed using a two-stage imputation method with IMPUTE2 ([57]), with the 1000Genomes Project (Phase3) data as the reference panel ([58]) and SHAPEIT2 ([59]) for phasing. Post-imputation quality was determined by the imputation accuracy score INFOSCORE from IMPUTE2 ([57]), with higher values (ranging from 0 – 1) indicating higher imputation quality. A clean dataset of genotyped and imputed SNPs was made available for this study.

2.3 Germline genotype data

In addition to the MyBrCa dataset described above, the TCGA Affymetrix SNP 6.0 genotyping data and clinical data were downloaded from the legacy archive of GDC (https://portal.gdc.cancer.gov/legacy-archive). All data was converted into a CEL file through Birdseed ([60]), imputed using a two-stage imputation approach with IMPUTE2 ([57]) and SHAPEIT2 ([59]). Further quality control (QC) steps were taken by utilising PLINK ([61]).

2.4 RNA Extraction, sequencing, alignment, and QC

The RNA from tumour samples was extracted using the QIAGEN miRNeasy Mini Kit with a QIA cube, according to standard protocol. Total RNA was quantified with the Nanadrop 2000 Spectrophotometer and RNA integrity was confirmed using Agilent 2100 Bioanalyzer. Only RNA samples with a concentration of >20.0ng/uL and an RIN number of 7 and above were sequenced. RNA libraries were prepared from 50 ng of genomic DNA using the TruSeq Stranded Total RNA HT kit with Ribo-Zero Gold (Illumina, San Diego, USA) following the manufacturer's instruction, The libraries were then subjected to paired-end 75 sequencing on

a HiSEQ4000 or NovaSeq platform (Illumina, San Diego, USA). RNA-seq reads were mapped to the hs37d5 human genome and ENSEMBL GRCh37 release 87 human transcriptome using the STAR aligner (v. 2.5.3a) ([62]). Samples with less than 15 million mapped fragments were excluded from further analysis. Gene-level aligned fragment counts were generated using featureCounts (v. 1.5.3), and gene-level expression in transcripts-per-million (TPM) was calculated using RSEM (v1.2.31)([63]). Genes with an average count of less than 10 or an average TPM score of <0.1 were filtered out and excluded from further analysis. To confirm sample IDs and for downstream analysis, variant calling from RNA-seq data was conducted using the Genome Analysis Toolkit (GATK) Best Practices workflow for RNA-seq. Lastly, variants were called using HaplotypeCaller.

These methods were previously completed and described by Pan et al., (2020) ([48]) and a clean dataset of RNA-seq data was made available for this study.

2.5 Immune scores

Using the RNA-seq data, individual immune scores were derived for a total of 112 immune traits including (1) 81 immune traits previously characterized in an immune analysis using gene set variation analysis (GSVA) by Thorsson et al., (2018) ([25]), (2) 29 GSVAs from the combined immune cell gene sets of immune traits from Bindea et al., (2013) ([34]), (3) IMPRES scores, a score that predicts response to immune checkpoint blockade (ICB) therapy ([36]), (4) ESTIMATE scores, a score that assesses immune cell infiltration in bulk tumour samples ([35]) and (5) GSVA with Expanded IFN-γ gene sets from Ayers et al., (2017) ([37]). The immune scores were categorised into six "modules" previously defined by Thorsson et al., (2018) ([25]). These six modules include (1) IFN response, (2) Leukocyte infiltration, (3) Macrophage/Monocyte,

(4) T-cell/Cytotoxic, (5) TGF-β response and (6) Wound healing. Out of the 112 immune scores, 25 were unassigned to any immune modules.

The relative abundance of immune cell populations ("immune score") was calculated from RNA-seq TPM scores with CIBERSORT ([64]) and TIMER ([65]) web tools. Normalisation was completed to reduce any potential errors caused by batch effects or differences in RNA-seq methods. After excluding samples with no immune score data, the MyBrCa dataset had a total of 960 samples, while the TCGA dataset had a total of 976 samples.

2.6 Immune trait normalisation

Fourteen housekeeping genes were included to normalize the gene expression data used in this study and account for any potential variations in the preparation and quality of the data. These genes were *ANKRD17*, *CCSER2*, *CIAO1*, *DAD1*, *GAPDH*, *HSP90AB1*, *IPO8*, *PFN1*, *PUM1*, *SF3A1*, *SRSF4*, *SYMPK*, *UBC*, and *YWHAZ* due to previous reports of stable expression levels in cancer studies ([66-70]). The geometric mean of the expression levels of these genes were used for normalisation. This was accomplished by plotting comparative Beeswarm boxplots, paired with a Student's T-test to determine the distribution of gene expression levels between the MyBrCa and TCGA cohorts. This step was accomplished by using the RStudio statistical environment software (version 1.3.1093).

2.7 Genome-wide heritability analysis

All 112 immune scores were estimated for its genome-wide heritability in the MyBrCa and TCGA population using the Genome-wide Complex Trait Analysis (GCTA) genome-based restricted maximum likelihood (GREML) statistical approach ([71, 72]). The GCTA-GREML statistical approach first calculates pairwise genetic relationships between all single-nucleotide polymorphisms (SNPs) by generating a Genetic Relatedness Matrix (GRM) for all

SNPs with a minor allele frequency (MAF) of > 0.01. This is followed by the GREML method that estimates the proportion of variance in each immune scores that could be attributed to common genetic variants. The proportion of phenotypic variance that can be explained by genotypic variance is calculated as V(Genotype)/V(Phenotype) ([71, 72]). A summary statistics table was generated to present all immune traits and its corresponding V(G)/V(p) values in the MyBrCa and TCGA cohort to select immune scores for downstream association analyses.

2.8 SNP-association test

Genotype data from the MyBrCa and TCGA cohorts was analysed for SNP associations with the selected immune scores as phenotypes using SNPTEST v2.5.6. The frequentist method was selected to calculate association statistics, and the expected test was used to test for associations between each SNP and each chosen immune trait. For the association analysis with the TCGA cohort, ethnicity was included as a covariate while genetic ancestry from a principal component analysis (PC1-4), batch number, RNA-SEQ and sample collection center were covariates for the MyBrCa cohort. MAF was calculated for all SNPs, and variants with an MAF of <0.01 were excluded. SNPs with an IMPUTE information score of <0.3 were also excluded.

2.9 Genome-wide association study

Manhattan plots were generated for each population and selected trait to identify overlapping genomic regions with the highest association signals. Standard cut-off values of $p = 5x10^{-8}$ for genome-wide significant SNPs, and $p = 1x10^{-5}$ for suggestive loci were used. To strengthen the statistical power of the analysis and reduce false-positive findings, a two-stage discovery and validation analysis was used to identify candidate SNPs associated with immune scores that had high heritability in both study cohorts. In the discovery stage, candidate SNPs with p-

values that exceeded the suggestive loci threshold ($p < 1 \times 10^{-5}$) from the MyBrCa cohort were shortlisted and subsequently validated using the TCGA cohort by comparing p-values and effect sizes (beta) of the same effect alleles. This process was repeated with candidate SNPs from the TCGA cohort, validated with the MyBrCa cohort. Validated candidate SNPs were confirmed based on a significance threshold of $p < 0.05$ in the validation cohort, as well as the criteria of having effect sizes in the same direction. For immune scores that were highly heritable only in the MyBrCa cohort, $p_{GWAS} < 1 \times 10^{-5}$ was applied to GWAS summary statistics and the SNP with the lowest p_{GWAS} in a genomic region was selected as a candidate SNP. All genome-wide significant and suggestive candidate SNPs that were found either through visualisation on Manhattan plots or summary statistics of the discovery and validation analysis were further examined and plotted using LocusZoom ([73]) to visualize association with the immune scores, linkage disequilibrium (LD) estimates and recombination rates (r^2 colour map) around the chosen SNPs.

MAF information associated with European (EUR), American (AMR) and East Asian (EAS) populations were extracted from 1000 Genomes and gnomAD Gnomes from SNPnexus (www.snp-nexus.org) ([74]).

2.10 Expression quantitative trait loci (eQTL) analysis

An eQTL analysis was performed using the R package Matrix eQTL (version 2.3) ([75]) to identify *cis*-eQTLs located within a genomic window of 1 megabase (Mb) for each candidate SNP identified at the GWAS step. A linear model was employed to assess the relationship between SNPs and gene expression levels. At this stage, a conservative p-value threshold of $p = 1$ was applied to capture all SNP-eQTL information for every gene-SNP pair within a genomic region. Following this, eQTpLot R package (version 0.0.0.9000) ([76]) was used to visualise the

colocalization of eQTLs and GWAS signals, while also examining the correlation of their p-values. eQTLs were considered significant if $p_{eQTL} < 0.01$.

2.11 Statistical Analysis

Standard descriptive statistics were used to describe differences across the study populations. Wilcoxon signed-rank test was used to test for differences in heritability of all immune scores between the MyBrCa and TCGA cohort, and a p-value of < 0.05 was considered statistically significant. The analysis was done using the RStudio statistical environment software (version 1.3.1093).

Chapter 3 : Results

3.1 Genome-wide heritability of 112 immune traits

A genome-wide heritability analysis was conducted for 112 immune traits (Fig 1) previously characterised by Thorsson et al., (2018) and Sayaman et al., (2021) using the GCTA-GREML statistical approach to estimate proportion of phenotypic variance that can be explained by common genetic variants across the MyBrCa and TCGA cohorts. Correlated immune traits were categorized within immune modules previously characterised by Thorsson et al., (2018) and ranked in order of their genetic variance to phenotypic variance (V(G)/V(p)) values in the MyBrCa cohort (Fig 2A). Most immune traits were categorised in the "Lymphocyte Infiltration" module (n=28), followed by the "T-cell/Cytotoxic" module (n=21), "Macrophage/Monocyte" module (n=16), "IFN Response" module (n=11), 7 in the "TGF-b Response" module and lastly 4 within the "Wound Healing" immune module. A subset of 25 immune traits did not fall within a module and were labelled as "Unassigned".

The results from the heritability analysis showed that the top 3 most heritable traits overall were MHC.1_19272155, MHC1_21978456 and IGG_Cluster_21214954, while the least heritable traits overall were IFIT3, LYMPHS_PCA_16704732 and DAP12_data. Within the Asian population, regulatory T (Treg) cells were most heritable (V(G)/V(p) = 0.667), while Bcell_21978456 was most heritable (V(G)/V(p) = 0.193) in the Caucasian population. In general, immune traits were more highly heritable in the Asian population (0.402 ± 0.016), compared to the Caucasian population (0.041 ± 0.005) as the former had higher median V(G)/V(p) values comparatively (Fig 2B; $p < 2.2 \times 10^{-16}$). Particularly, 103 (92%) scores had greater heritability in the Asian population, while only 9 (8%) were heritable in the Caucasian

population. Differences between the MyBrCa and TCGA cohort does not appear to be limited to any specific immune module.

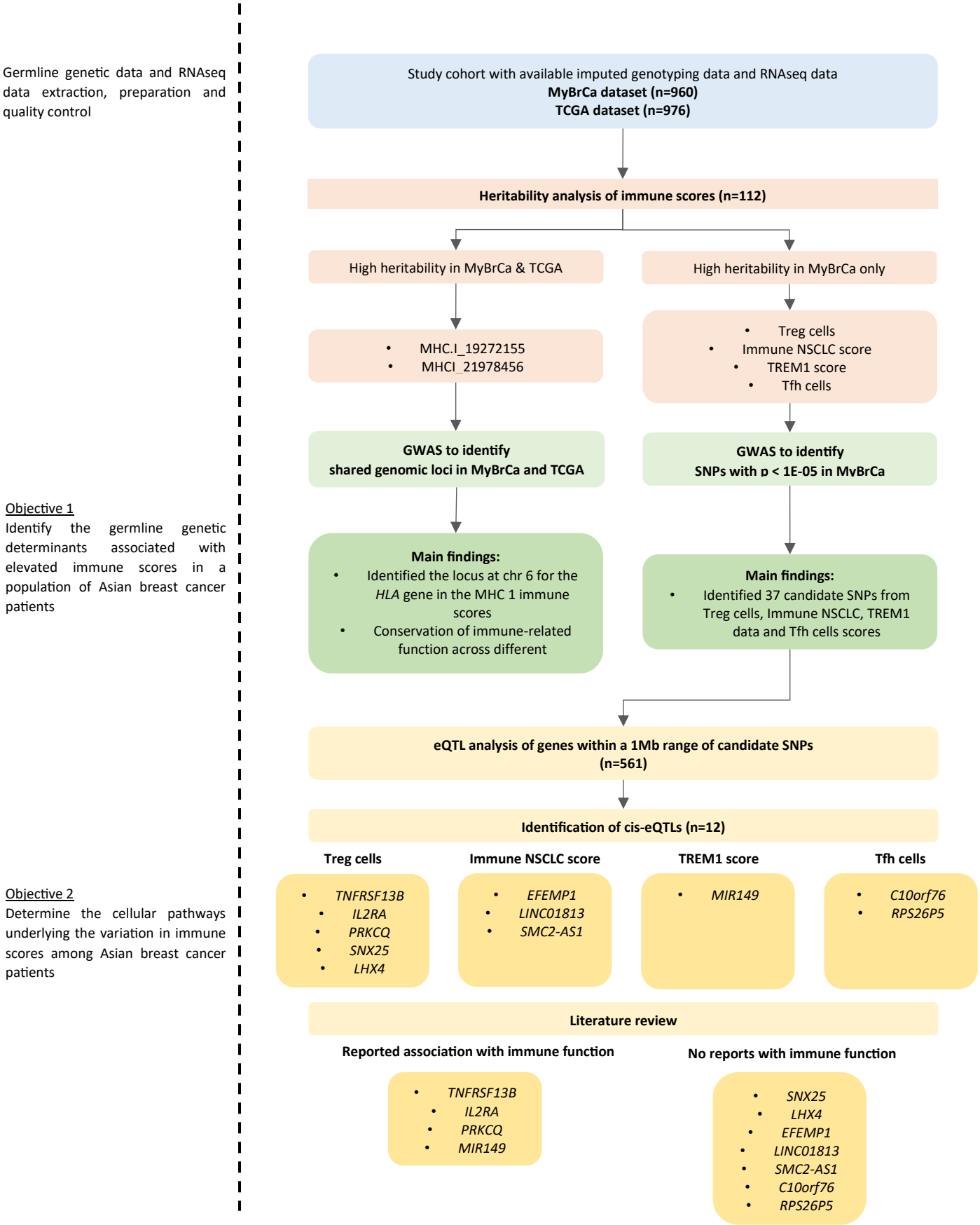


Figure 1. Overview of study approach

Flowchart depicting the objectives, methods, and main results of the study.

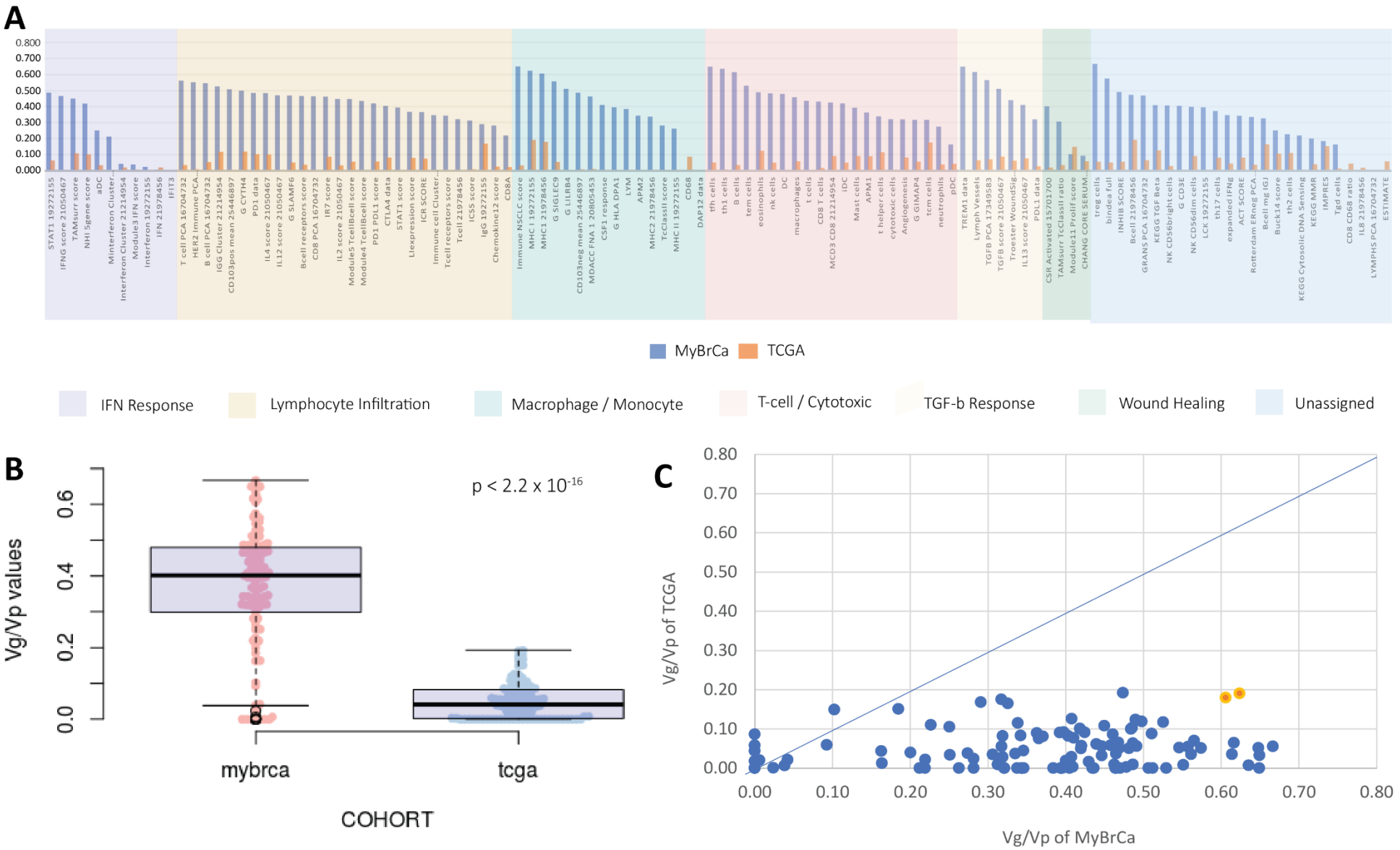


Figure 2. Genome-wide heritability of immune traits

A Genome-wide heritability estimates, V(G)/V(p), of 112 immune scores categorised by immune module in the MyBrCa and TCGA populations. **B** Overall comparison of genome-wide heritability betweeb MyBrCa and TCGA. **C** Genome-wide heritability of immune scores in MyBrCa and TCGA plotted against each other. Immune traits with high heritability in both populations are highlighted in orange.

3.2 Genome-wide association analysis of highly heritable immune traits

This study hypothesized that immune traits with high genome-wide heritability in both the MyBrCa and TCGA cohorts contribute to host immune function through the same molecular mechanisms and pathways regardless of ethnicity, implying a strong likelihood of having the same genomic regions of interest. To visualise this, V(G)/V(p) values for both cohorts were plotted in Figure 2C and two immune traits were chosen for further analysis (highlighted in orange). These immune traits are two variants of the major histocompatibility complex class-I (MHC.I_19272155, MHC1_21978456). MHC.I_19272155 had a V(G)/V(p) value of 0.623 in the MyBrCa cohort and 0.191 in the TCGA cohort, while MHC1_21978456 had a V(G)/V(p) value of 0.606 in MyBrCa and 0.180 in TCGA. The gene sets for both immune scores are presented in Table 1.

Manhattan plots were generated for MHC.I_19272155 (Fig 3A) and MHC1_21978456 (Fig 4A) to identify associated loci that surpassed the suggestive p-value threshold of 1×10^{-5} in both populations. This hypothesis was further investigated through a discovery and validation analysis. In total, for MHC.I_19272155, 333/1196 (27.8%) candidate SNPS from the MyBrCa cohort were identified and subsequently validated in the TCGA population, while 248/1857 (13.4%) candidate SNPs from the TCGA population were validated in the MyBrCa population. For MHC1_21978456, a total of 540/1458 (37%) candidate SNPs from MyBrCa were validated in TCGA, and 349/2026 (17.2%) vice versa. It is worth noting that despite the substantial number of validated candidate SNPs for both immune traits in each population, most were found to be SNPs clustered around the same genomic regions. Therefore, Table 2 highlights the representative candidate SNP for each genomic region that achieved the lowest p-value in the validating cohort.

Among all validated candidate SNPs in the two immune traits, the genomic region at chr 6 was of greatest interest as it had many SNPs with a low p value (Fig 3A; Fig 4A) and was also most strongly validated in both populations (Tab 2). Therefore, the most significant candidate SNP in this genomic region of the MyBrCa population was further investigated through regional association plots (Fig 3B; Fig 4B). For MHC.I_19272155, this SNP was chr6:229911228 ($p = 1.53 \times {}^{10-}7$), and chr6:29923554 ($p = 6.49 \times 10^{-6}$) for MHC1_21978456. Genes found nearby this region include multiple different classes of the Human Leukocyte Antigen gene (*HLA-H, HLA-G, HLA-A, HLA-F, HLA-F-AS1, HLA-J* and *HLA-L*).

Treg cells (n=1)	Immune NSCLC (n=72)	TREM1 (n=1)	Tfh cells (n=32)	MHC1_19272155 (n=5)	MHC1_21978456 (n=7)
FOXP3	C3orf41	TREM1	B3GAT1	HLA-G	HLA-E
	C1orf24		BLR1	HLA-A	HLA-G
	PLEK		C18orf1	HLA-B	HLA-B
	RHOH		CDK5R1	HLA-C	HLA-C
	PSCDBP		CHGB	HLA-F	HLA-J
	TCEA2		CHI3L2		HLA-F
	FLJ21963		CXCL13		HLA-A
	CLEC4E		HEY1		
	USP51		HIST1H4K		
	WFDC10B		ICA1		
	IGH@		KCNK5		
	SLC4A3		KIAA1324		
	CD53		MAF		
	LOC401431		MAGEH1		
	SCFV		MKL2		
	THRAP2		MYO6		
	PRDM13		MYO7A		
	OBSL1		PASK		
	C7orf40		PDCD1		
	TAGAP		POMT1		
	MGC11271		PTPN13		
	IGLV6-57		PVALB		
	CD38		SH3TC1		
	FKBP9		SIRPG		
	ADAMTSL2		SLC7A10		
	CD48		SMAD1		
	GNPTAB		ST8SIA1		
	DHRS8		STK39		
	LOC388886		THADA		
	CNIH3		TOX		
	PSMA6		TSHR		
	CCRK		ZNF764		
	SHROOM1				
	GPSM1				
	TRO				
	GSTT2				
	NQO2				
	EAF2				
	MUM1L1				
	MUC4				
	C13orf21				
	PABPC1				
	PLA2G7				
	PARK2				
	AOAH				
	IGL@				
	LOC642480				
	TMSB4X				
	LOC390712				
	ACOT8				
	GIMAP7				
	LOC375010				
	ASAH1				
	TRIM45				
	C2orf30				
	EXT2				
	IFI6				
	KCNE3				
	CTSF				
	SULT1C1				
	RASL11B				
	LOC148898				
	HMGCL				
	IGHA1				
	C1QTNF3				
	C20orf46				
	IL11RA				
	ADRA2C				
	IGKC				
	CEACAM5				
	PURB				
	TPD52				

Table 1. Gene sets for selected immune traits**.**

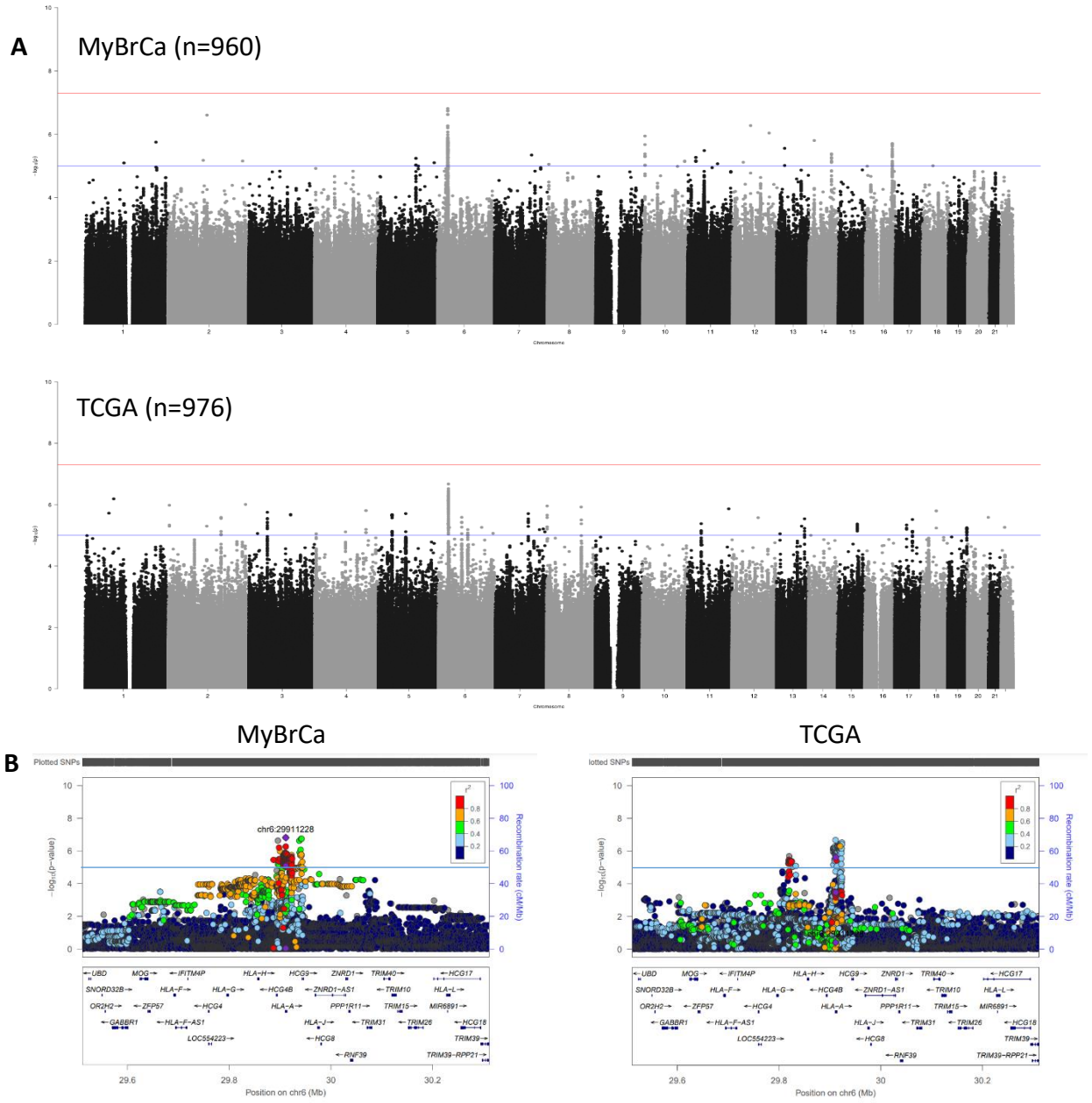


Figure 3. Genome-wide association analysis of MHC.I_19272155 in MyBrCa and TCGA populations.

A Manhattan plots of MHC.I_19272155 score in MyBrCa (n=960) and TCGA (n=976). The MHC.1_19272155 score, calculated using GSVA, includes genes *HLA-G, HLA-A, HLA-B, HLA-C,* and *HLA-F.* The plot shows -log10 transformed p-values for all SNPs analysed. Horizontal blue line represents suggestive loci threshold ($p = 1 \times 10^{-5}$), and the red line represents genome-wide significant threshold of $p = 5 \times 10^{-8}$. **B** Regional association plots around the strongest associated candidate SNP, chr6:29911228 (purple diamond), from MyBrCa and in TCGA. Surrounding SNPs are colour-coded to by LD strength (pairwise r^2-values). Nearby genes annotated by the UCSC Genome Browser are below the association plots. Plots were generated with LocusZoom.

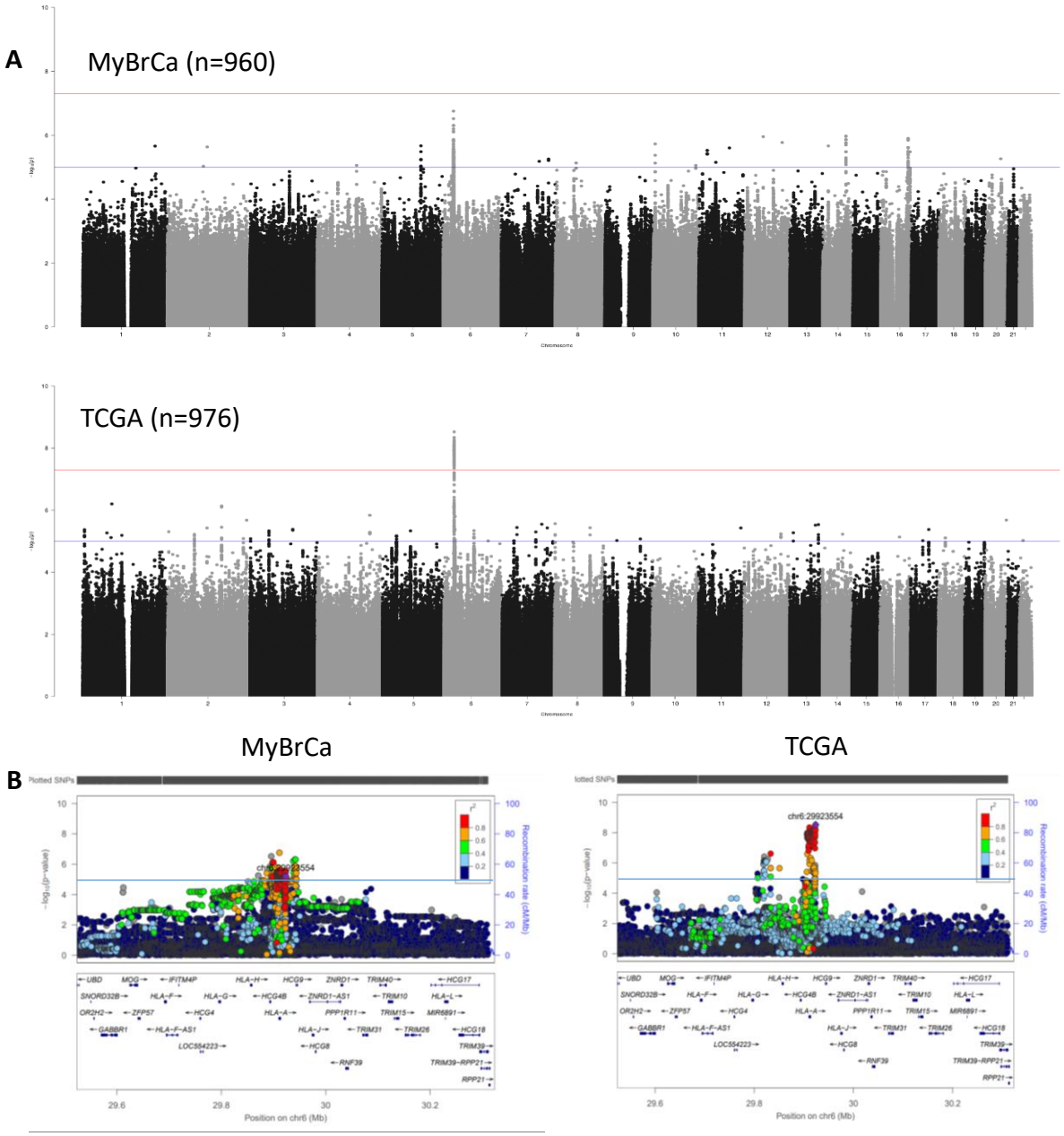


Figure 4. Genome-wide association analysis of MHCI_21978456 in MyBrCa and TCGA populations.

A Manhattan plots of MHC.I_19272155 score in MyBrCa (n=960) and TCGA (n=976). The MHCI_21978456 score, calculated using GSVA, includes genes *HLA-E, HLA-G, HLA-B, HLA-C, HLA-J, HLA-A* and *HLA-F.* The plot shows -log10 transformed p-values for all SNPs analysed. Horizontal blue line represents suggestive loci threshold ($p = 1 \times 10^{-5}$), and the red line represents genome-wide significant threshold of $p = 5 \times 10^{-8}$. **B** Regional association plots around the strongest associated candidate SNP, chr6:29923554 (purple diamond), from TCGA and in MyBrCa. Surrounding SNPs are colour-coded to by LD strength (pairwise r^2-values). Nearby genes annotated by the UCSC Genome Browser are below the association plots. Plots were generated with LocusZoom.

SNP Details				*Discovery: MyBrCa – MHC_19272155 (n = 5)*			*Validation - TCGA*		
chr	**pos**	**Allele A**	**Allele B**	**P-value**	**Effect size**	**MAF**	**P-value**	**Effect size**	**MAF**
6	29923554	T	C	2.67E-06	0.215	0.370	2.99E-09	0.300	0.345
3	116422008	T	G	7.64E-05	0.210	0.253	2.73E-03	-0.210	0.157
19	46119884	C	T	3.60E-05	-0.568	0.031	3.07E-02	1.096	0.005
10	90022647	C	T	8.76E-05	-0.177	0.415	3.14E-02	0.127	0.238
5	114725793	G	T	5.24E-05	0.350	0.075	3.86E-02	0.140	0.156
				Discovery: TCGA – MHC_19272155 (n = 18)				**Validation - MyBrCa**	
6	29909593	T	G	8.19E-07	0.249	0.351	1.30E-06	0.225	0.356
6	164181088	A	G	8.59E-06	-0.656	0.033	6.36E-03	0.288	0.058
6	24436760	G	A	8.47E-05	0.234	0.198	2.11E-02	0.102	0.442
6	32635605	G	A	2.97E-05	-0.224	0.323	4.82E-02	-0.096	0.308
6	159374559	C	T	9.72E-05	0.302	0.102	4.97E-02	-0.096	0.359
1	172381157	C	G	8.52E-05	-0.781	0.019	1.13E-02	1.187	0.004
1	105795244	C	T	8.95E-05	-0.200	0.303	3.73E-02	0.118	0.193
2	7566811	G	A	6.25E-05	-0.648	0.028	2.71E-02	-0.987	0.003
3	193862478	A	G	2.49E-05	-0.385	0.087	2.10E-03	0.195	0.133
3	55071234	C	G	2.63E-05	0.260	0.174	4.63E-02	-0.098	0.264
4	160919694	GTA	G	1.48E-05	-0.252	0.293	4.02E-02	-0.144	0.182
5	159011137	A	T	7.67E-05	0.351	0.084	3.15E-02	-0.433	0.013
5	41207011	T	C	7.33E-06	0.238	0.278	3.87E-02	0.093	0.424
7	154003120	C	T	2.17E-05	-0.374	0.098	3.71E-02	0.481	0.015
10	133149831	A	T	5.13E-05	0.217	0.325	8.83E-03	-0.153	0.187
14	99550624	T	C	1.48E-05	-0.444	0.059	2.56E-02	0.198	0.114
16	86071714	C	A	1.80E-05	1.049	0.013	1.09E-02	-0.190	0.157
17	15120704	T	C	7.05E-05	-0.213	0.273	4.79E-02	0.119	0.174
				Discovery: MyBrCa - MHC1_21978456 (n= 12/540)				**Validation - TCGA**	
6	29923554	T	C	6.49E-06	0.210	0.370	2.99E-09	0.300	0.345
1	154907082	C	T	7.84E-05	-0.673	0.021	2.52E-02	-1.917	0.002
3	116422008	T	G	1.96E-05	0.231	0.253	2.73E-03	-0.21	0.157
3	142836128	C	CA	6.90E-05	-0.181	0.451	9.75E-03	-0.134	0.345
3	20643480	A	G	4.41E-05	0.196	0.468	3.57E-02	0.11	0.338
5	114725793	G	T	2.29E-05	0.373	0.075	3.86E-02	0.14	0.156
7	13032859	A	C	8.43E-05	0.310	0.101	7.43E-03	0.669	0.013
8	54393521	C	T	9.97E-05	0.305	0.093	4.49E-02	0.24	0.043
10	44122491	C	TT	6.94E-05	0.183	0.478	2.19E-02	0.137	0.237
10	90022647	C	T	3.79E-05	-0.189	0.415	3.14E-02	0.127	0.238
16	86112755	TAAAAAA AAAAAAA AA	T	7.53E-05	0.232	0.286	2.05E-02	-0.129	0.458
19	46119884	C	T	2.63E-05	-0.589	0.031	3.07E-02	1.096	0.005
				Discovery: TCGA – MHC_21978456 (n = 15)				**Validation - MyBrCa**	
6	29911228	A	T	6.23E-07	0.248	0.427	1.76E-07	0.239	0.474
6	24436760	G	A	7.60E-05	0.237	0.198	9.88E-03	0.116	0.442
6	32614387	A	G	2.20E-05	-0.459	0.072	4.50E-02	0.166	0.102
6	164181088	A	G	1.21E-05	-0.648	0.033	7.81E-03	0.286	0.058
1	4818197	T	C	8.26E-05	0.187	0.437	3.97E-02	0.123	0.220
3	193862478	A	G	6.75E-05	-0.366	0.087	7.00E-03	0.174	0.133
4	127656784	A	G	3.39E-05	-0.212	0.351	2.57E-02	-0.100	0.450
4	160919694	GTA	G	1.87E-05	-0.25	0.293	3.77E-02	-0.148	0.182
5	80223613	G	C	9.08E-05	0.253	0.174	2.27E-03	0.159	0.287
5	41223861	C	A	2.43E-05	0.223	0.287	3.86E-02	0.094	0.421
7	154003120	C	T	1.03E-05	-0.390	0.098	1.73E-02	0.559	0.015
9	73758049	G	A	6.86E-05	-0.276	0.270	3.95E-02	0.136	0.236
10	133151585	T	TAGTC	7.03E-05	0.209	0.360	2.95E-02	-0.130	0.187
12	69255455	C	T	5.05E-05	0.258	0.188	1.88E-02	0.246	0.072
16	86071714	C	A	6.00E-05	0.986	0.013	8.36E-03	-0.201	0.157

Table 2. Validated SNPs for MHC.I_19272155 and MHC1_21978456.

chr: chromosome number; pos: SNP position number; MAF: minor allele frequency

3.3 Genome-wide association analysis of highly heritable immune traits in Asian breast tumours

Next, a GWAS was conducted on immune traits that were most heritable in the MyBrCa population (Fig 1). These immune traits were Treg cells (from the "Unassigned" module), immune non-small cell lung cancer (NSCLC) score belonging to the Macrophage/Monocyte module, triggering receptor expressed on myeloid cells-1 (TREM1) from the TGF-β Response module and T follicular helper (Tfh) cells from the T-cell/Cytotoxic module (Fig 2A). The gene sets for all 4 immune scores are presented in Table 1.

Manhattan plots were generated for all 4 immune scores to identify genomic regions that exceeded the suggestive loci threshold (Fig 5.1, Fig 5.3, Fig 5.5, Fig 5.7), and candidate SNPs were systematically shortlisted through the use of GWAS summary statistics. These candidate SNPs have been reported in Table 3. A total of 12 candidate SNPs were shortlisted for Treg cells, 7 for Immune NSCLC score, 8 for TREM1 and 10 for Tfh cells. Using regional association plots, it was confirmed that all shortlisted candidate SNPs were in strong linkage disequilibrium (LD) with neighbouring alleles that also shared relatively low p-values (Fig 5.2, Fig 5.4, Fig 5.6, Fig 5.8).

Regulatory T (Treg) Cells

MyBrCa (n=960)

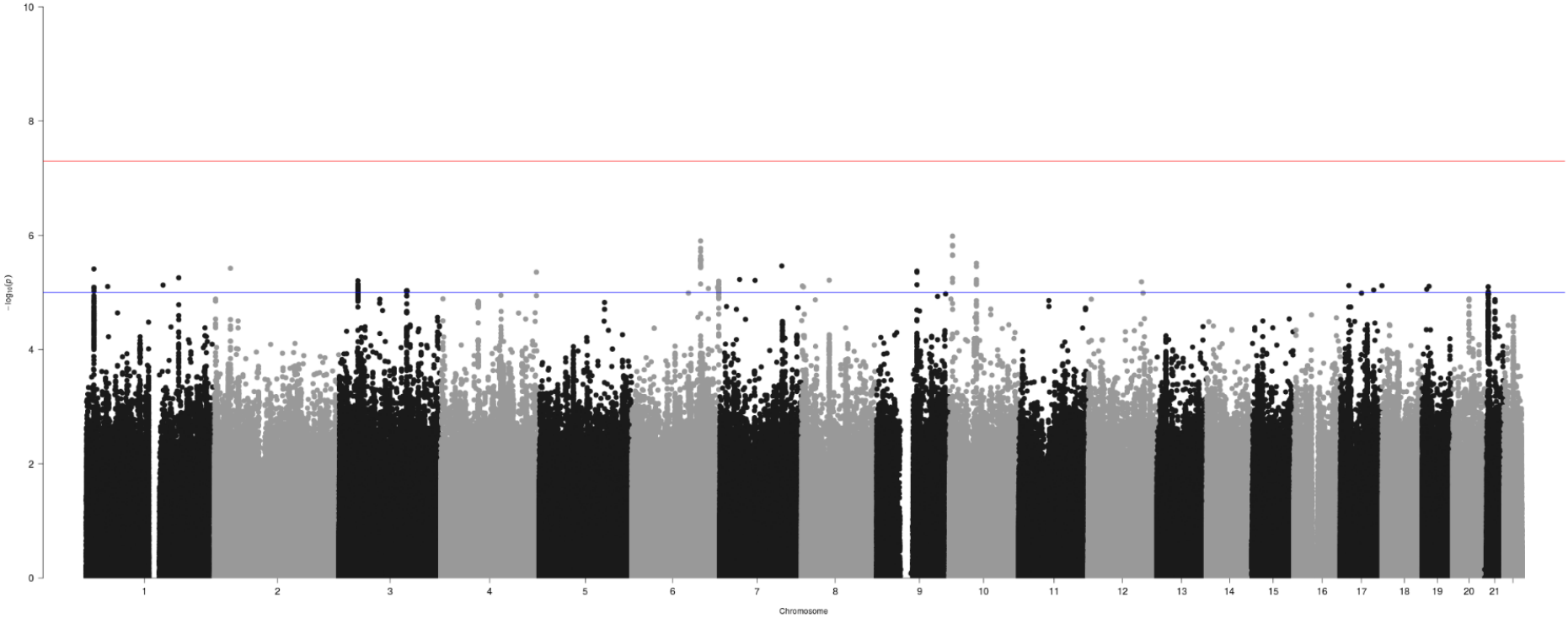


Figure 5.1. Manhattan plot of Treg cells in the MyBrCa population (n = 960).

Horizontal blue line represents suggestive loci threshold of $p = 1 \times 10^{-5}$, and the red line represents genome-wide significant threshold of $p = 5 \times 10^{-8}$.

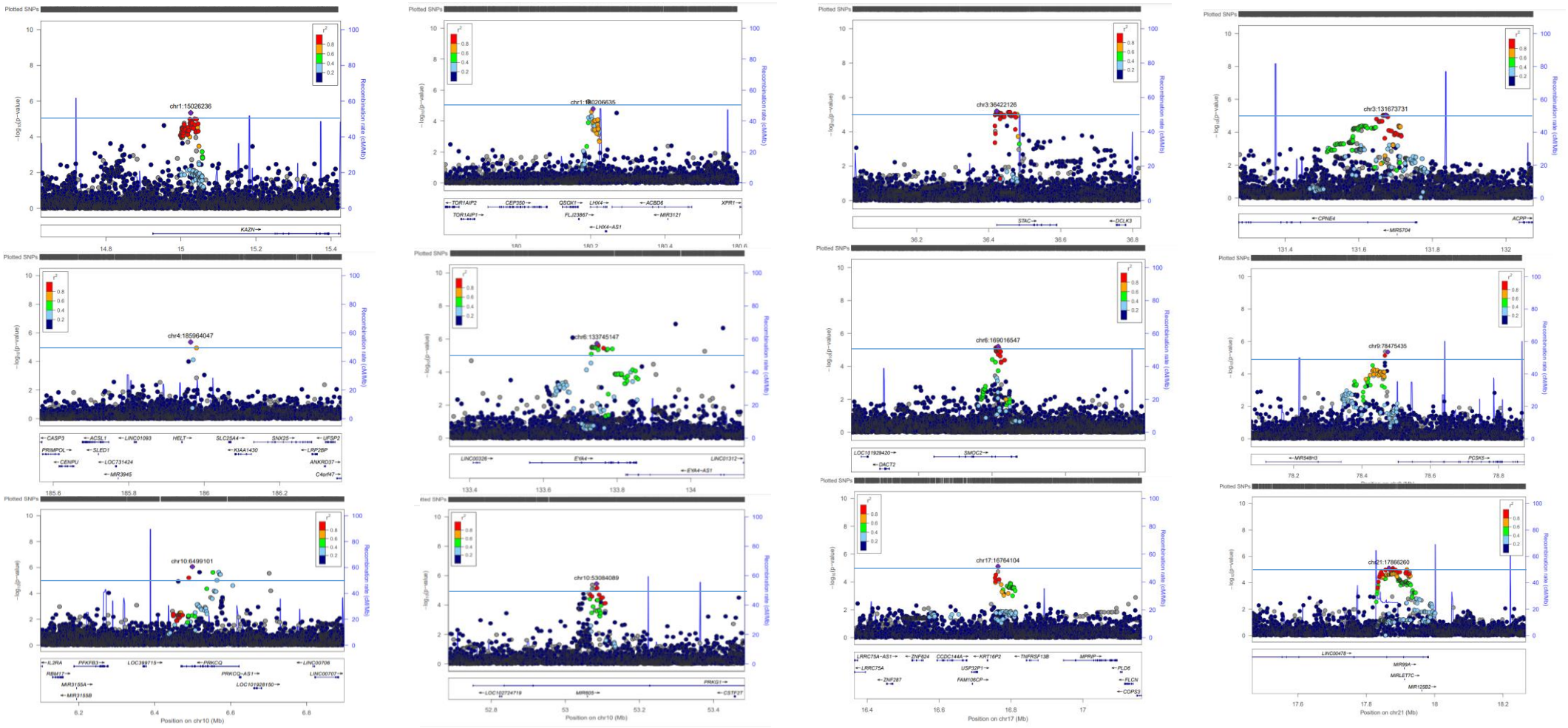


Figure 5.2. Regional association plots for candidate SNPs of Treg cells in the MyBrCa population.

Horizontal blue line represents suggestive loci threshold of $p = 1 \times 10^{-5}$.

Immune Non-Small Cell Lung Cancer (NSCLC) Score

MyBrCa (n=960)

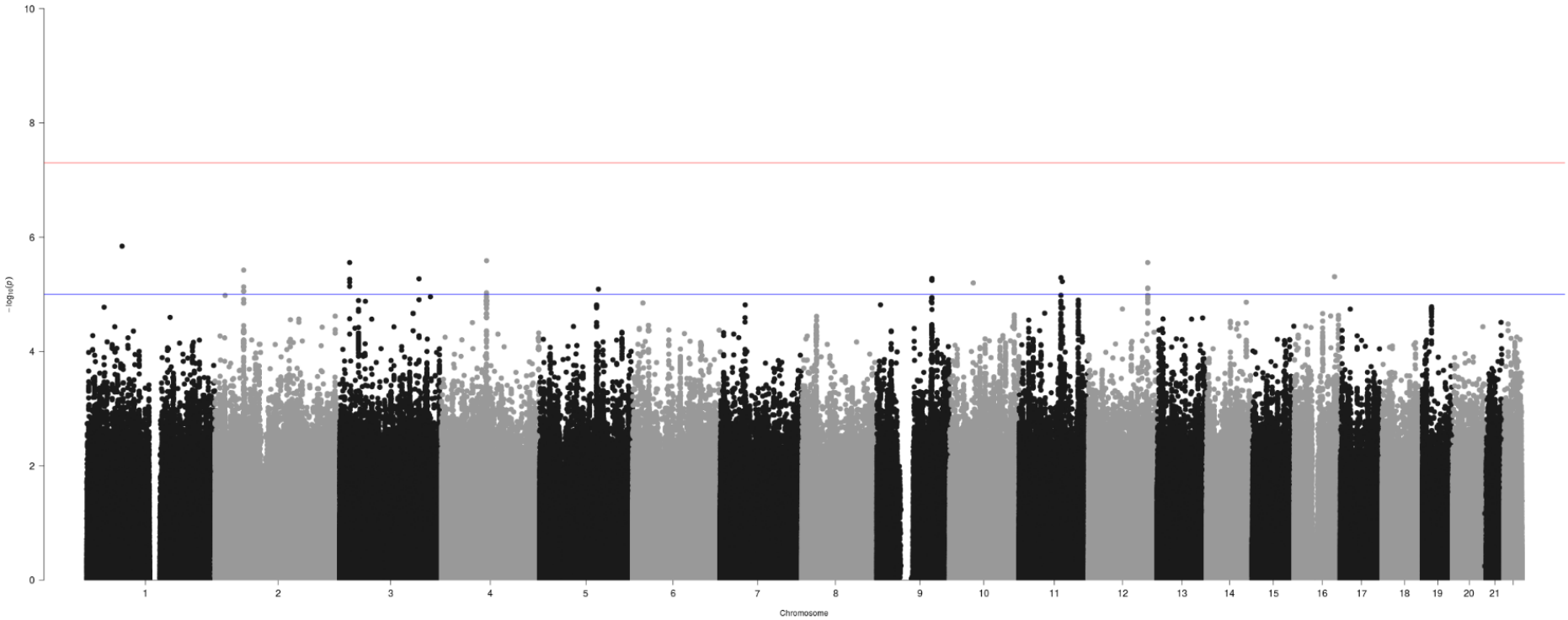


Figure 5.3. Manhattan plot of Immune NSCLC Score in the MyBrCa population (n = 960).

Horizontal blue line represents suggestive loci threshold of $p = 1 \times 10^{-5}$, and the red line represents genome-wide significant threshold of $p = 5 \times 10^{-8}$.

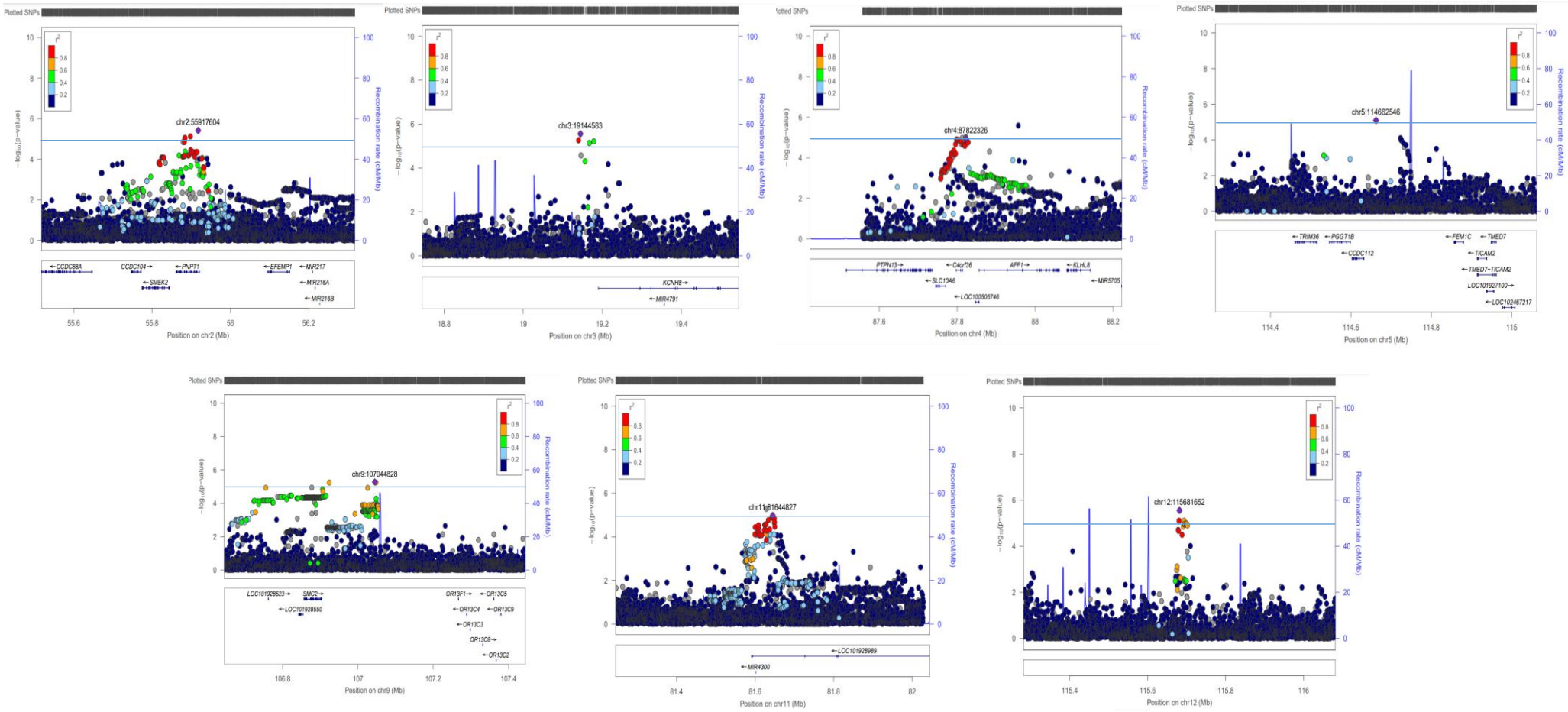


Figure 5.4. Regional association plots for candidate SNPs of Immune NSCLC Score in the MyBrCa population.

Horizontal blue line represents suggestive loci threshold of $p = 1 \times 10^{-5}$.

Triggering Receptor Expressed On Myeloid Cells 1 (TREM1)

MyBrCa (n=960)

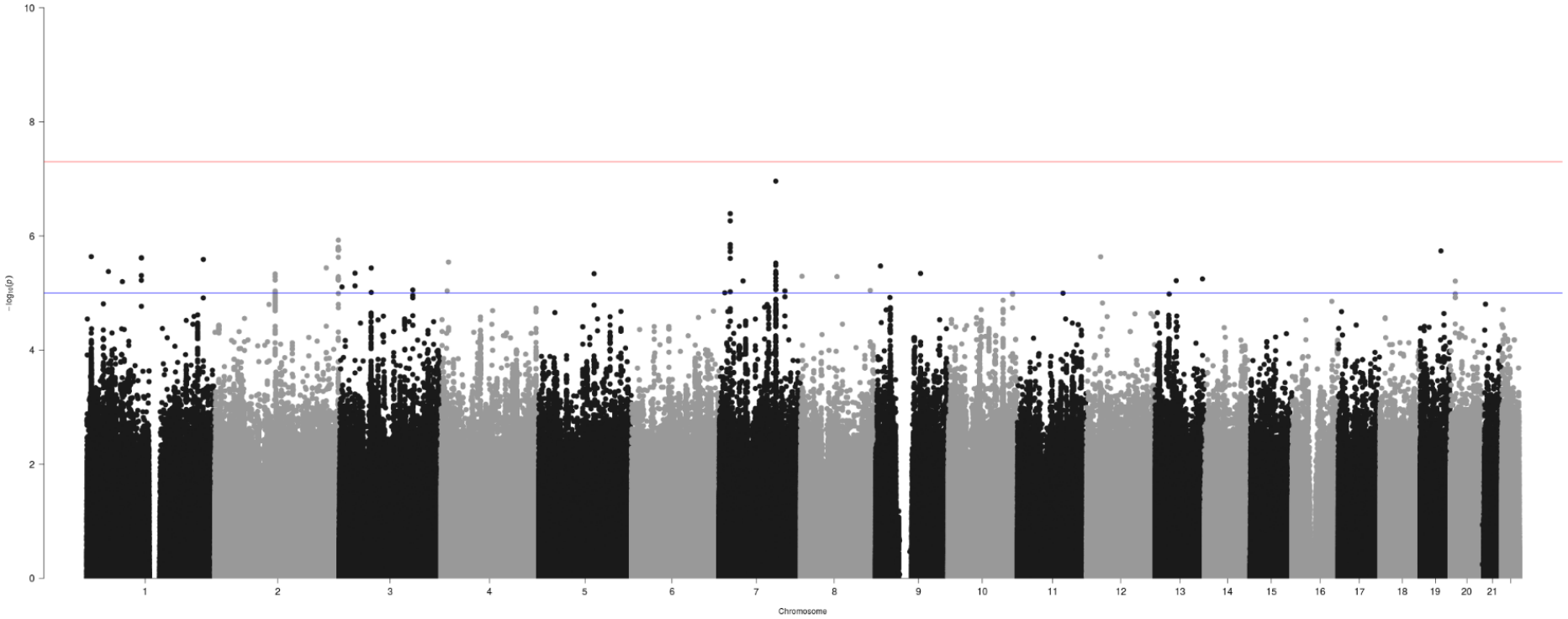


Figure 5.5. Manhattan plot of TREM1 score in the MyBrCa population (n = 960).

Horizontal blue line represents suggestive loci threshold of $p = 1 \times 10^{-5}$, and the red line represents genome-wide significant threshold of $p = 5 \times 10^{-8}$.

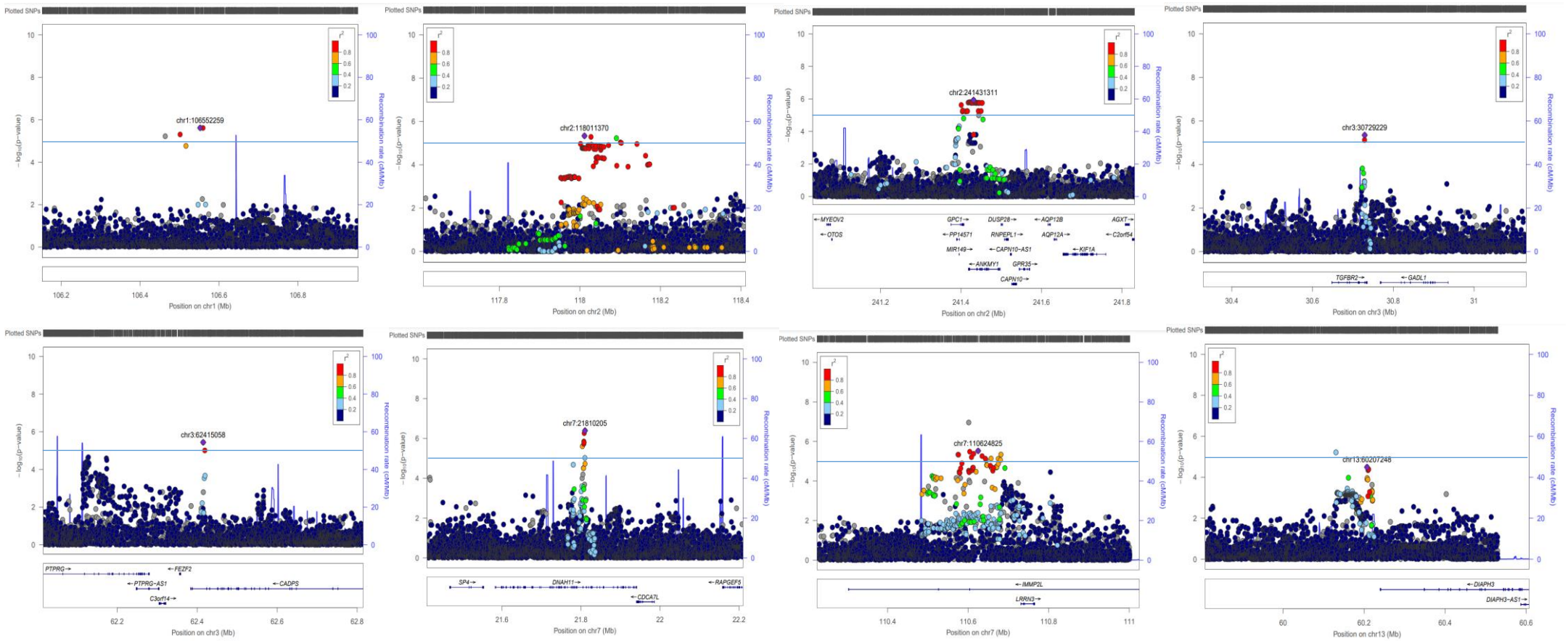


Figure 5.6. Regional association plots for candidate SNPs of TREM1 score in the MyBrCa population.

Horizontal blue line represents suggestive loci threshold of $p = 1 \times 10^{-5}$.

T Follicular Helper (Tfh) Cells

MyBrCa (n=960)

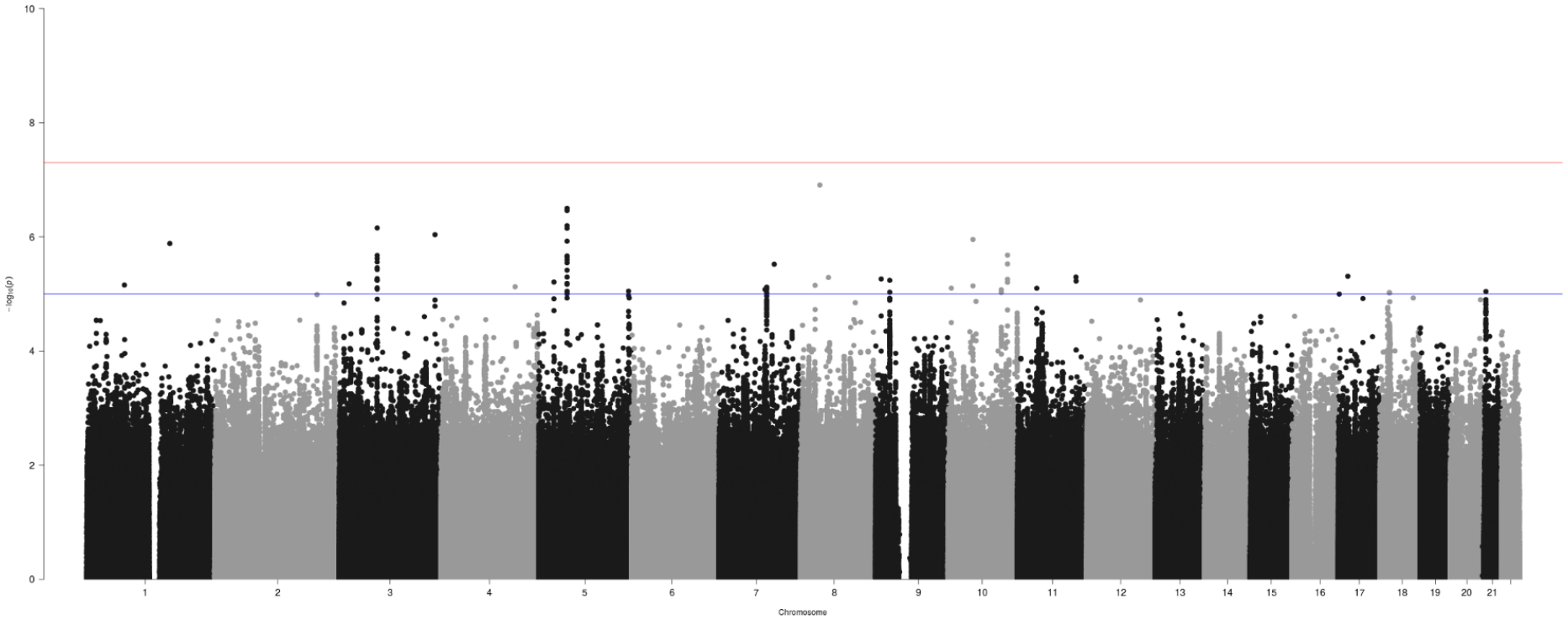


Figure 5.7. Manhattan plot of Tfh cells in the MyBrCa population (n = 960).

Horizontal blue line represents suggestive loci threshold of $p = 1 \times 10^{-5}$, and the red line represents genome-wide significant threshold of $p = 5 \times 10^{-8}$.

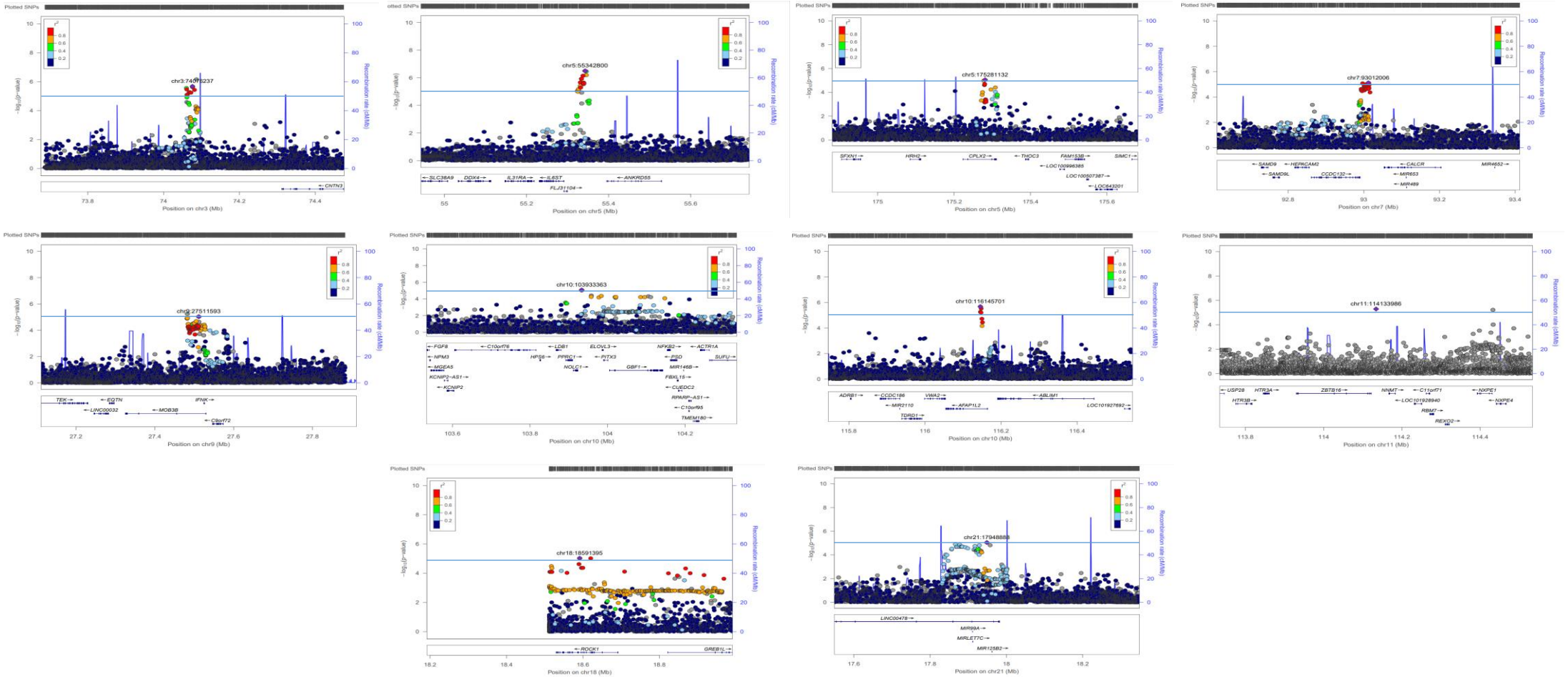


Figure 5.8. Regional association plots for candidate SNPs of Tfh cells in the MyBrCa population.

Horizontal blue line represents suggestive loci threshold of $p = 1 \times 10^{-5}$.

SNP Details				GWAS Summary Statistics		
chr	pos	Allele A	Allele B	MAF	P-value	Effect size
			Treg cells			
1	15026236	A	C	0.469	3.87E-06	-0.203
1	180194037	T	TG	0.323	5.53E-06	0.225
3	36422126	C	T	0.170	6.24E-06	0.258
3	131673731	G	A	0.118	9.30E-06	-0.284
4	185964047	A	G	0.013	4.41E-06	-1.050
6	133745147	A	G	0.222	1.25E-06	0.257
6	169016547	G	A	0.259	6.36E-06	0.218
9	78467855	G	GC	0.214	4.22E-06	-0.248
10	6499101	T	A	0.028	1.03E-06	0.671
10	53084089	G	A	0.368	3.08E-06	0.215
17	16764104	C	T	0.226	7.56E-06	0.241
21	17866260	A	G	0.072	7.95E-06	0.367
			Immune NSCLC			
2	55917604	A	G	0.172	3.78E-06	0.262
3	19144583	C	T	0.248	2.78E-06	0.289
4	87956639	G	C	0.016	2.58E-06	-0.832
5	114662546	A	T	0.013	8.12E-06	1.023
9	107044828	A	G	0.015	5.28E-06	0.779
11	81628400	T	TTATA	0.433	5.13E-06	-0.192
12	115681652	A	G	0.039	2.79E-06	0.511
			TREM1			
1	106552259	T	C	0.050	2.41E-06	-0.635
2	118011370	G	T	0.016	4.64E-06	-0.815
2	241431311	A	G	0.019	1.19E-06	0.821
3	30729229	G	C	0.454	4.49E-06	0.215
3	62415058	A	C	0.377	3.65E-06	-0.280
7	110601488	GT	G	0.386	1.10E-07	-0.247
7	21810205	C	G	0.495	4.07E-07	0.230
13	60130495	T	C	0.410	6.08E-06	-0.234
			Tfh cells			
3	74086643	CAAAAAAAAAA	C	0.451	7.00E-07	0.213
5	55342800	C	T	0.107	3.17E-07	0.348
5	175281132	A	C	0.431	8.95E-06	-0.176
7	93012006	C	T	0.222	7.65E-06	-0.217
9	27482648	G	GC	0.271	5.78E-06	-0.209
10	103933363	T	C	0.136	8.47E-06	0.260
10	116145701	C	G	0.465	2.10E-06	0.211
11	114133986	CT	C	0.014	5.07E-06	1.057
18	18591395	T	C	0.052	9.48E-06	-0.405
21	17948888	T	A	0.099	9.07E-06	0.331

Table 3. Candidate SNPs for Treg cells, Immune NSCLC, TREM1 and Tfh cells.

chr: chromosome number; pos: SNP position number; MAF: minor allele frequency

3.4 Expression quantitative trait loci analysis

In this study, it was hypothesized that some of the genetic variants associated with immune traits in Asian breast tumours may exert its effect by influencing gene expression. All genes within a 1Mb genomic window of the 37 candidate SNPs were extracted (n=561, data not shown) and after testing for significant gene-SNP associations with Matrix eQTL and eQTpLot, 382 genes had at least one significant *cis*-eQTL ($p_{eQTL} < 0.05$). A more stringent significance level ($p_{eQTL} < 0.01$) was applied to select *cis*-eQTLs for further downstream analyses, resulting in 12 significant *cis*-eQTLs. These *cis*-eQTLs were *LHX4*, *SNX25*, *IL2RA*, *PRKCQ* and *TNFRSF13B* from the Treg cells immune score, *EFEMP1*, LINC01813, and *SMC2-AS1* from the Immune NSCLC score, *MIR149* from TREM1 score and lastly, *C10orf76* and *RPS26P5* from the Tfh cells immune score. Of the 12 significant *cis*-eQTL genes, 4 have previously been reported as being associated with immune scores. These 4 genes, presented in Figures 6.1 to 6.4, were *TNFRSF13B*, *PRKCQ*, *IL2RA*, and *MIR149* ([77-86]). All remaining 8 eQTpLots can be found in Figures 6.5 to 6.12. In addition to their roles in immune scores, the 2 variants specifically associated with Treg cells immune score were found to have distinct population frequency. The variant chr10:6499101 had higher MAF values in cohorts of Asian descent (MyBrCa: 0.028, EAS: 0.031) compared to those of European descent (TCGA: 0.003). The same was observed for variant chr17:16764104 (MyBrCa: 0.226, EAS: 0.281 vs. TCGA: 0.034, AMR: 0.029, EUR: 0.024) (Tab 4).

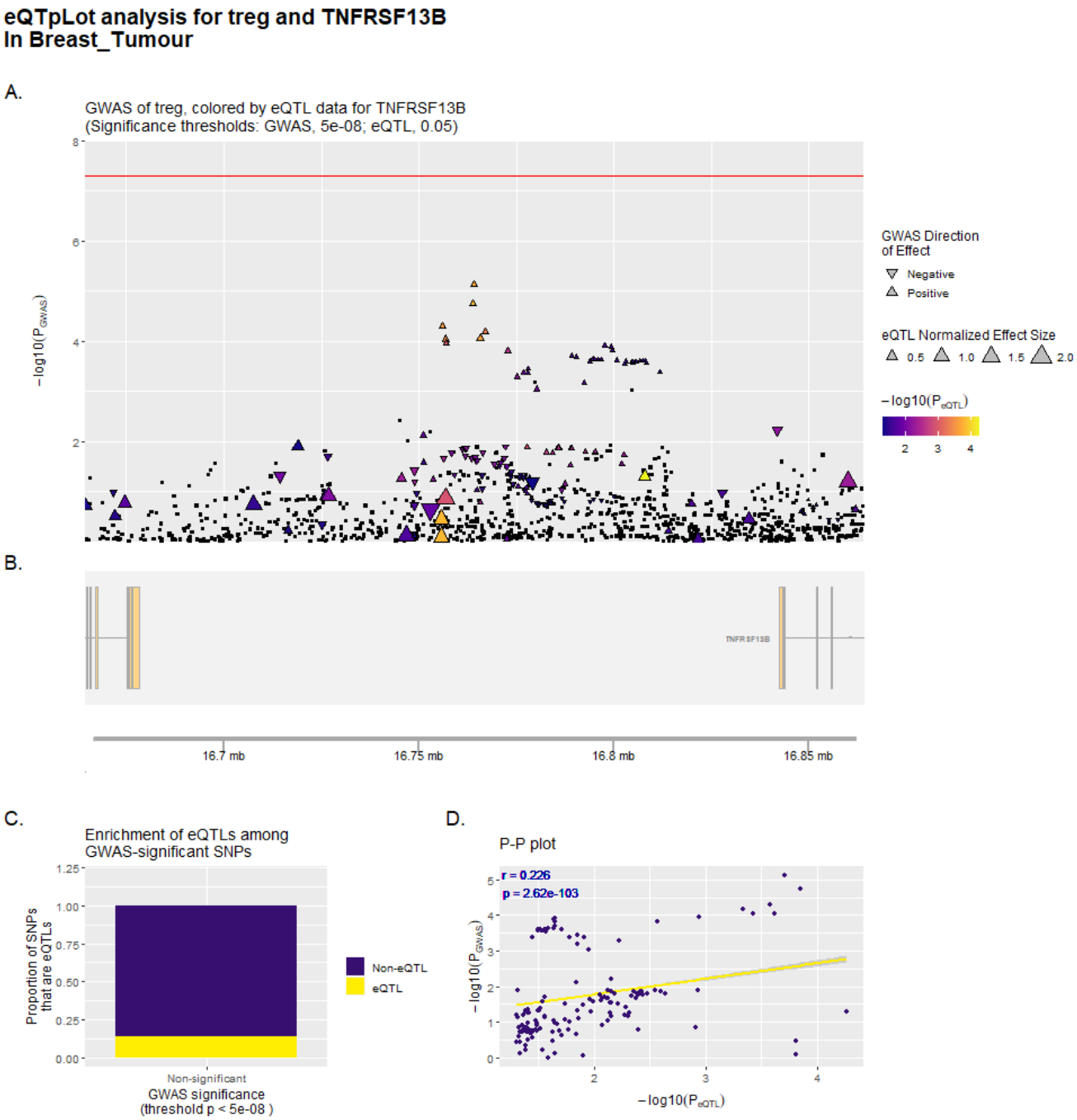


Figure 6.1. eQTpLot of the colocalization of eQTLs for gene *TNFRSF13B* and GWAS signals for Treg cells on candidate SNP chr17:16764104 in the MyBrCa population.

A Genomic region of interest around chr17:16764104. Colour scale corresponds to the magnitude of association between each SNP and *TNFRSF13B* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of p = 5 x 10^{-8}. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *TNFRSF13B* and Treg cells with Pearson correlation coefficient (r) and p-value (p).

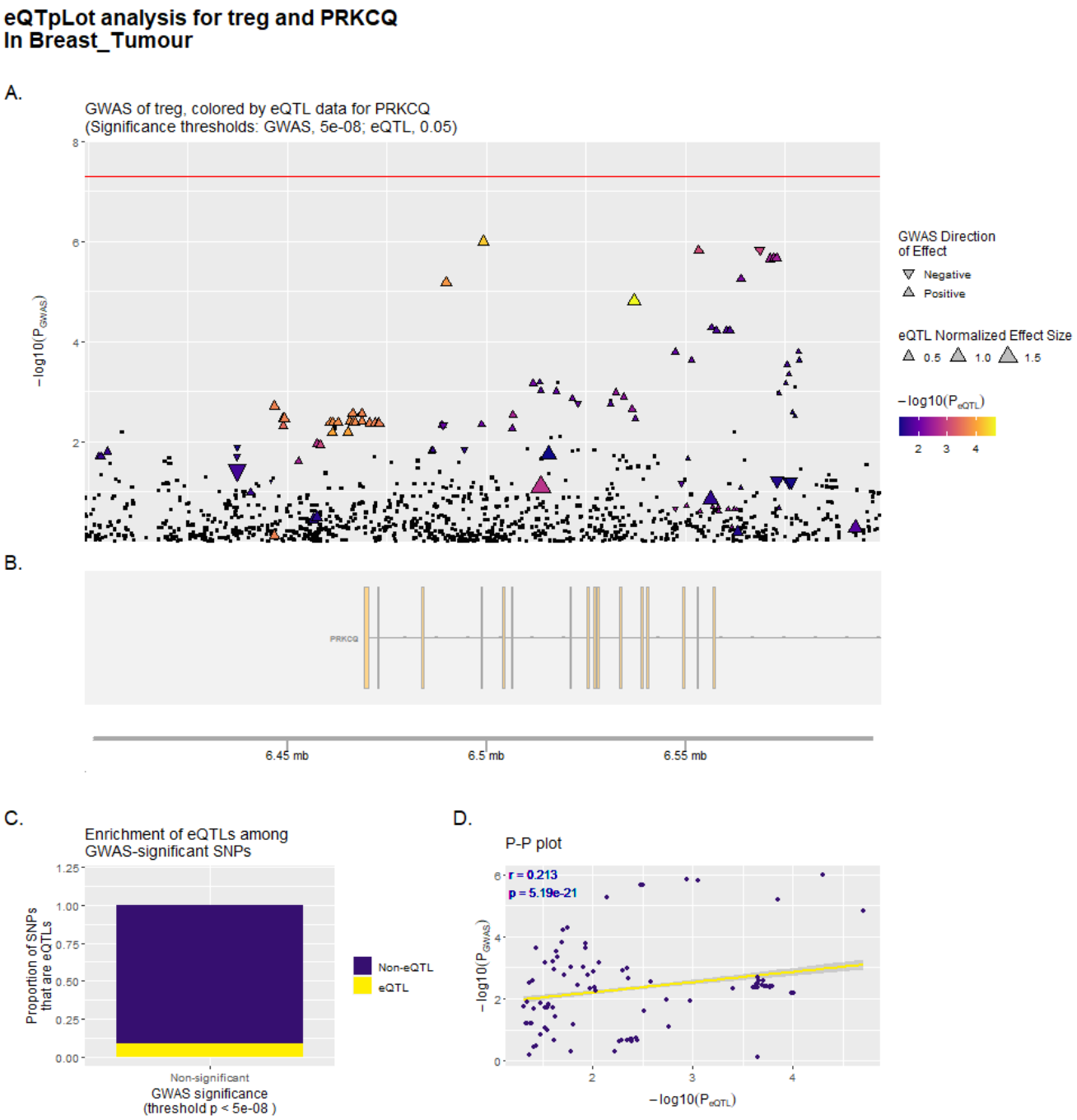


Figure 6.2. eQTpLot of the colocalization of eQTLs for gene *PRKCQ* and GWAS signals for Treg cells on candidate SNP chr10:6499101 in the MyBrCa population.

A Genomic region of interest around chr10:6499101. Colour scale corresponds to the magnitude of association between each SNP and *PRKCQ* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of $p = 5 \times 10^{-8}$. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *PRKCQ* and Treg cells with Pearson correlation coefficient (r) and p-value (p).

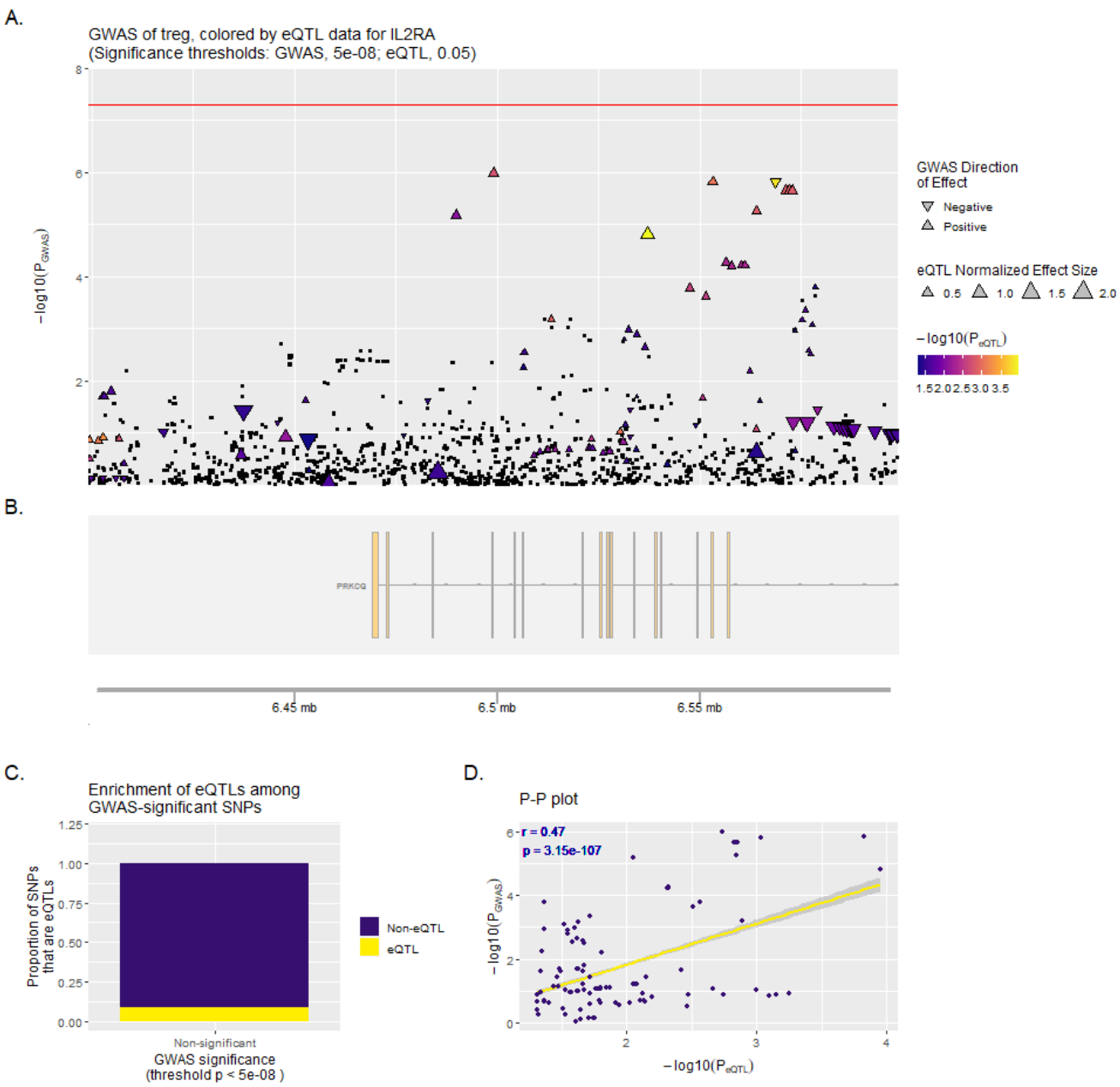


Figure 6.3. eQTpLot of the colocalization of eQTLs for gene *IL2RA* and GWAS signals for Treg cells on candidate SNP chr10:6499101 in the MyBrCa population.

A Genomic region of interest around chr10:6499101. Colour scale corresponds to the magnitude of association between each SNP and *IL2RA* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of $p = 5 \times 10^{-8}$. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *IL2RA* and Treg cells with Pearson correlation coefficient (r) and p-value (p).

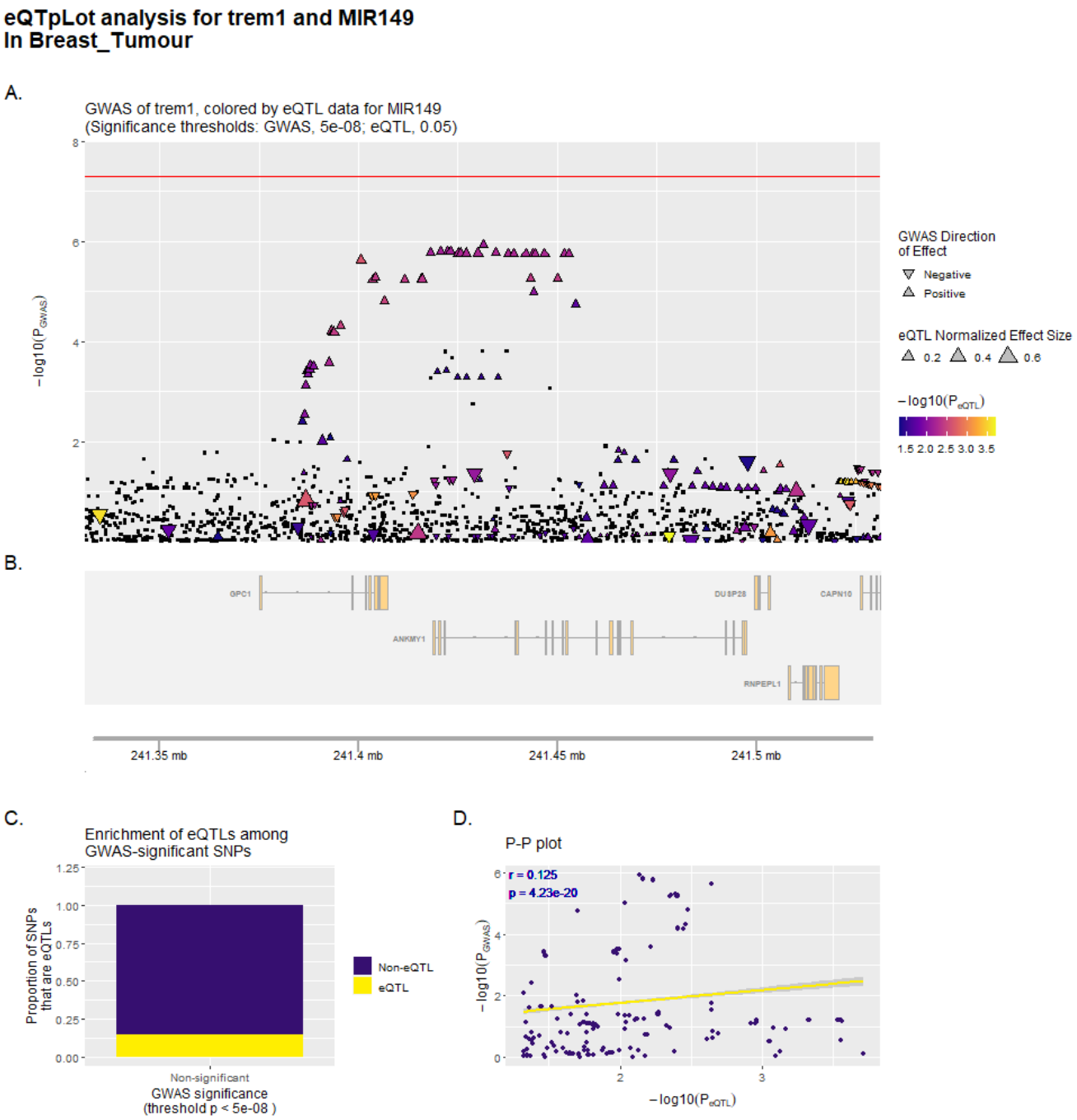


Figure 6.4. eQTpLot of the colocalization of eQTLs for gene *MIR149* and GWAS signals for TREM1 score on candidate SNP chr2:241431311 in the MyBrCa population.

A Genomic region of interest around chr2:241431311. Colour scale corresponds to the magnitude of association between each SNP and *MIR149* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of p = 5 x 10^{-8}. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *MIR149* and TREM1 with Pearson correlation coefficient (r) and p-value (p).

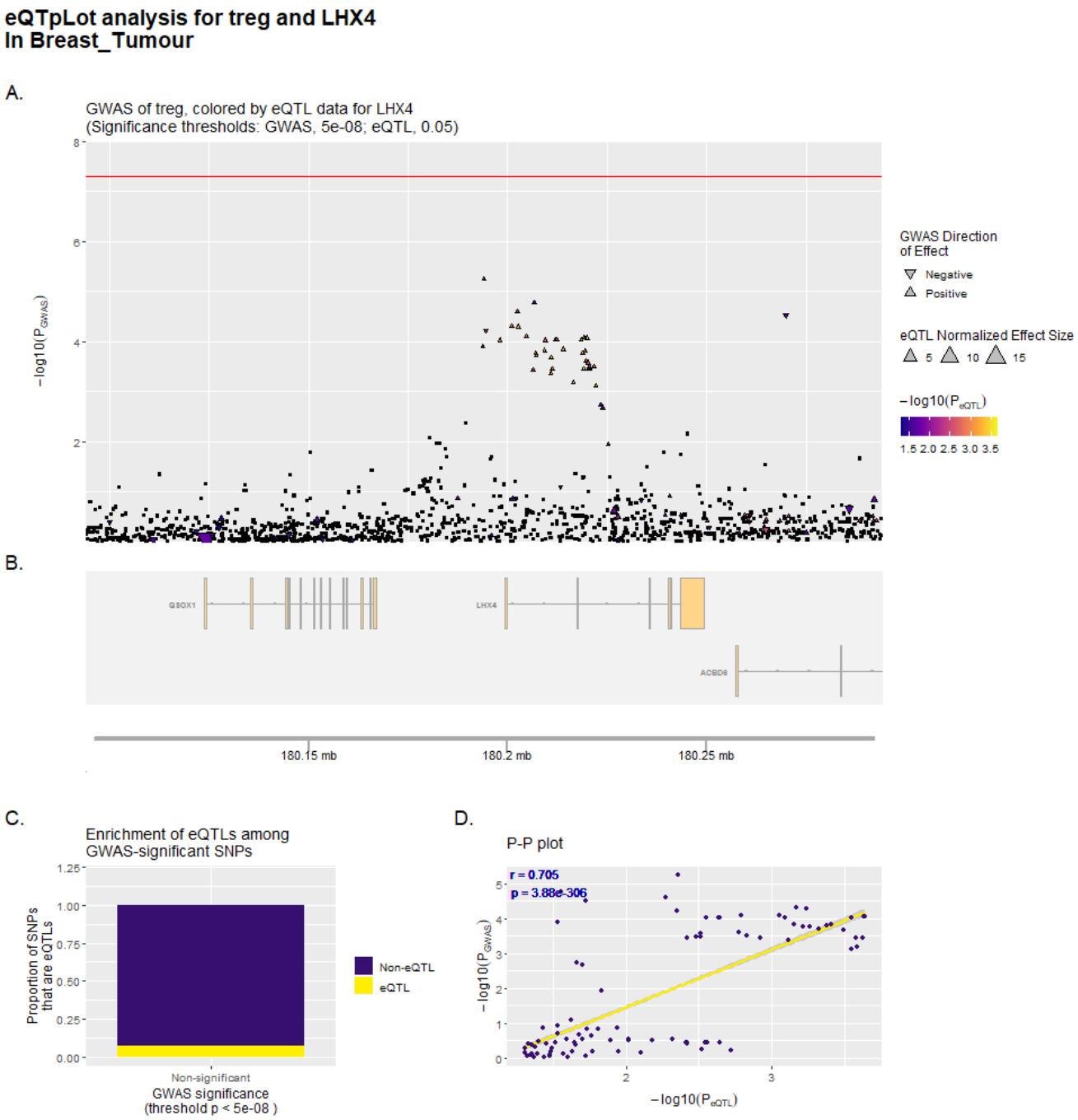


Figure 6.5. eQTpLot of the colocalization of eQTLs for gene *LHX4* and GWAS signals for Treg cells on candidate SNP chr1:180194037 in the MyBrCa population.

A Genomic region of interest around chr1:180194037. Colour scale corresponds to the magnitude of association between each SNP and *LHX4* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of $p = 5 \times 10^{-8}$. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *LHX4* and Treg cells with Pearson correlation coefficient (r) and p-value (p).

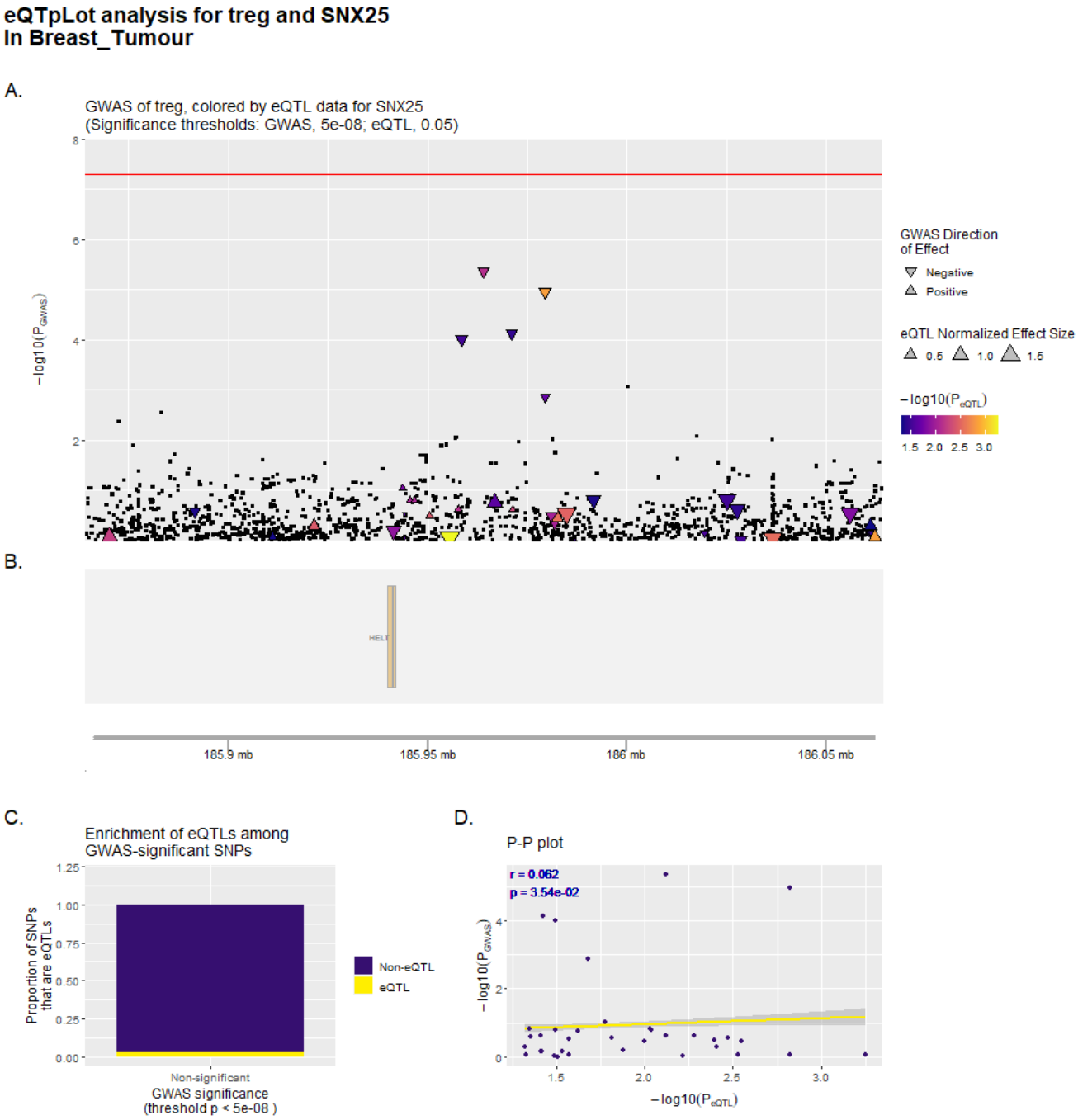


Figure 6.6. eQTpLot of the colocalization of eQTLs for gene *SNX25* and GWAS signals for Treg cells on candidate SNP chr4:185964047 in the MyBrCa population.

A Genomic region of interest around chr4:185964047. Colour scale corresponds to the magnitude of association between each SNP and *SNX25* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of $p = 5 \times 10^{-8}$. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *SNX25* and Treg cells with Pearson correlation coefficient (r) and p-value (p).

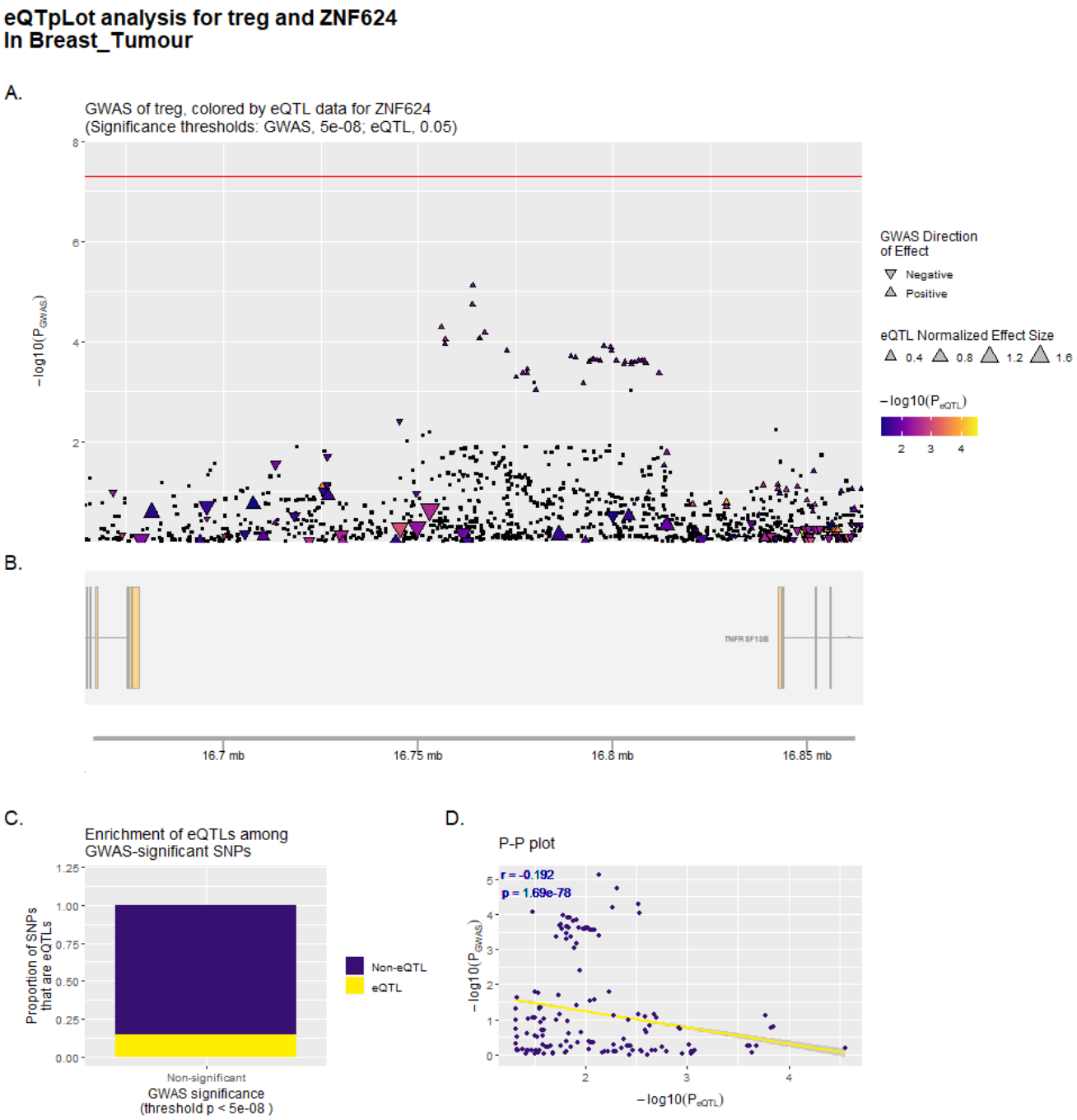


Figure 6.7. eQTpLot of the colocalization of eQTLs for gene *ZNF624* and GWAS signals for Treg cells on candidate SNP chr17:16764104 in the MyBrCa population.

A Genomic region of interest around chr17:16764104. Colour scale corresponds to the magnitude of association between each SNP and *ZNF624* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of $p = 5 \times 10^{-8}$. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *ZNF624* and Treg cells with Pearson correlation coefficient (r) and p-value (p).

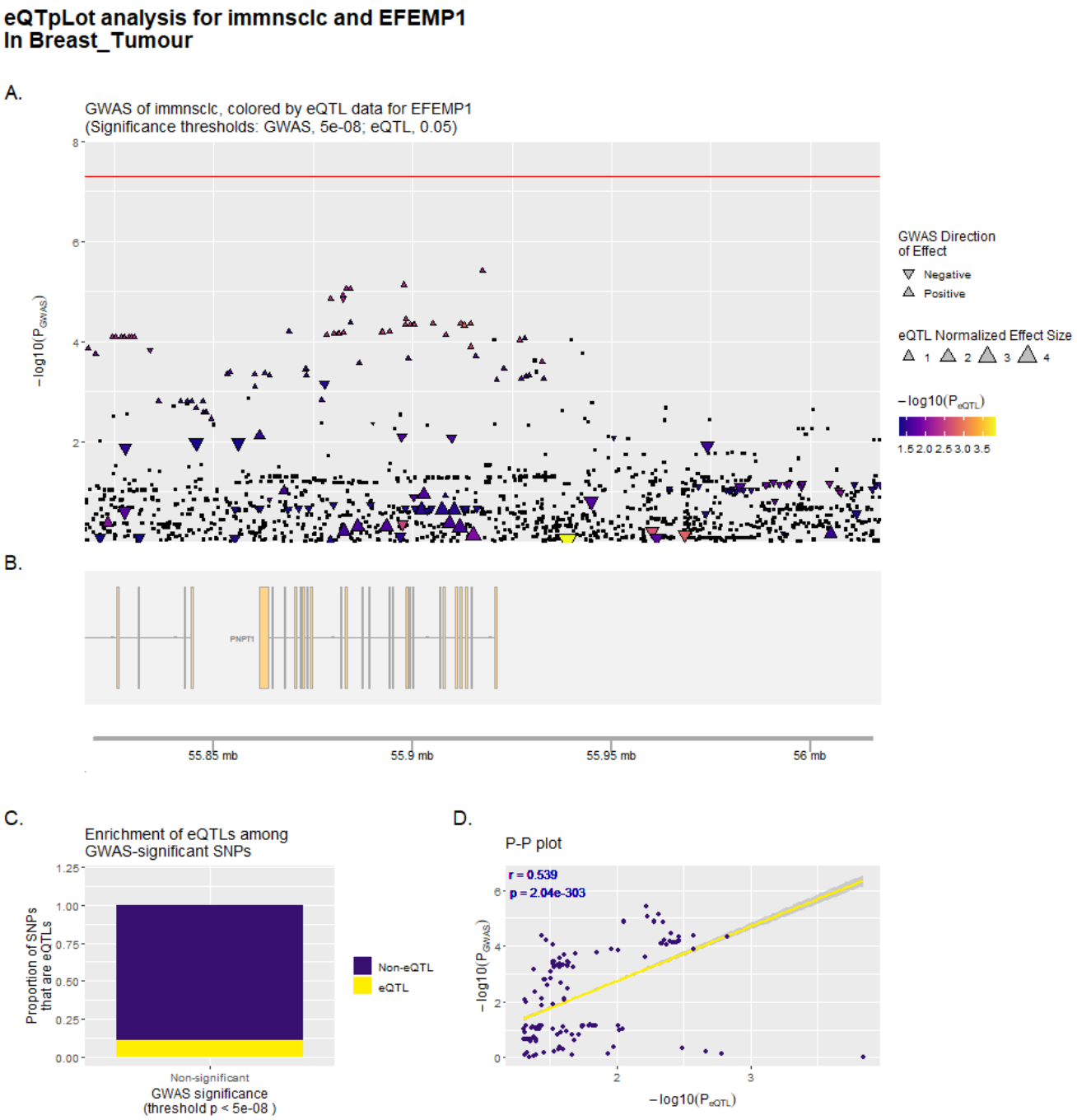


Figure 6.8. eQTpLot of the colocalization of eQTLs for gene *EFEMP1* and GWAS signals for Immune NSCLC score on candidate SNP chr2:55917604 in the MyBrCa population.

A Genomic region of interest around chr2:55917604. Colour scale corresponds to the magnitude of association between each SNP and *EFEMP1* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of $p = 5 \times 10^{-8}$. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *EFEMP1* and Immune NSCLC score with Pearson correlation coefficient (r) and p-value (p).

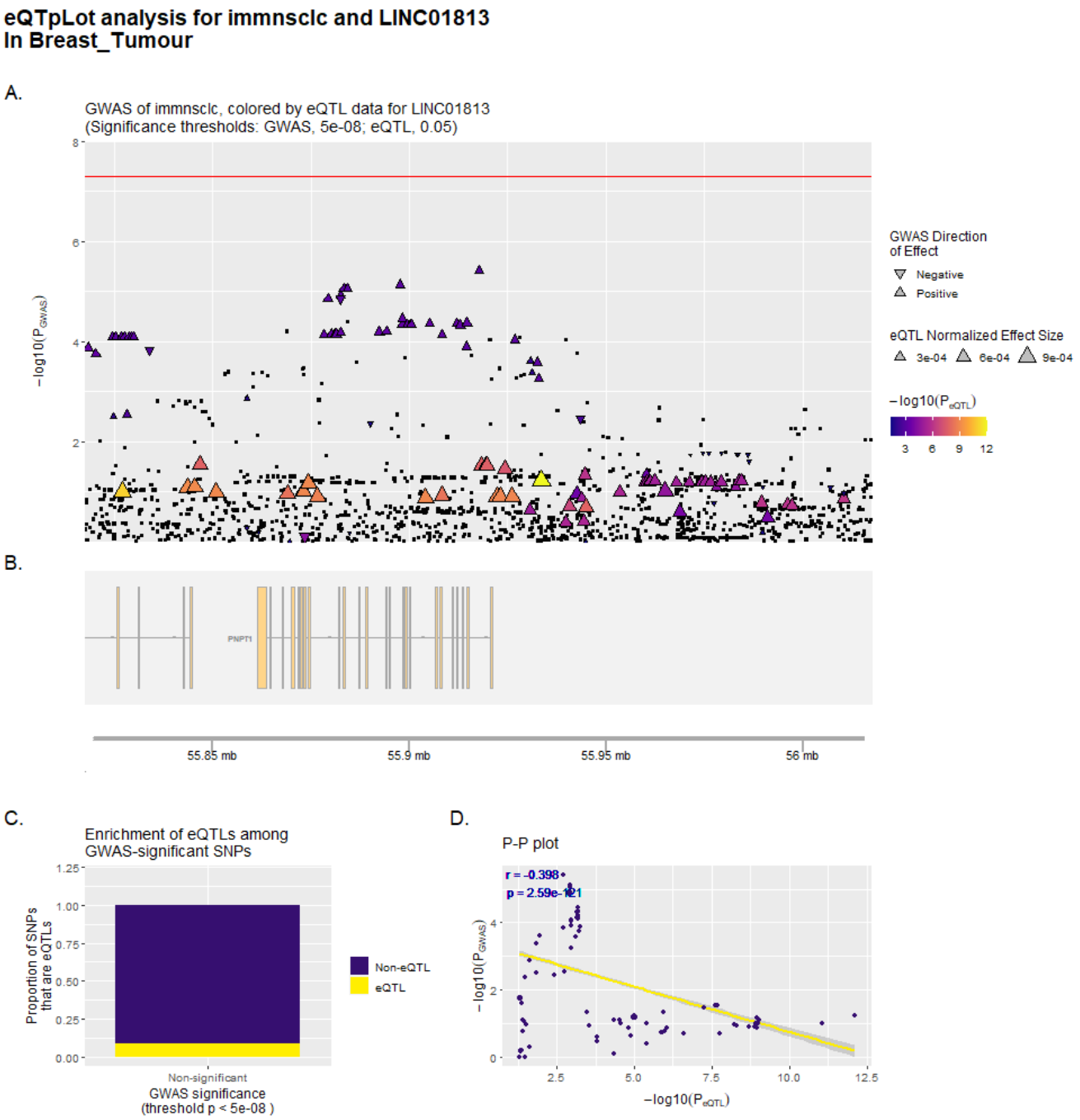


Figure 6.9. eQTpLot of the colocalization of eQTLs for gene *LINC01813* and GWAS signals for Immune NSCLC score on candidate SNP chr2:55917604 in the MyBrCa population.

A Genomic region of interest around chr2:55917604. Colour scale corresponds to the magnitude of association between each SNP and *LINC01813* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of p = 5 x 10^{-8}. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *LINC01813* and Immune NSCLC score with Pearson correlation coefficient (r) and p-value (p).

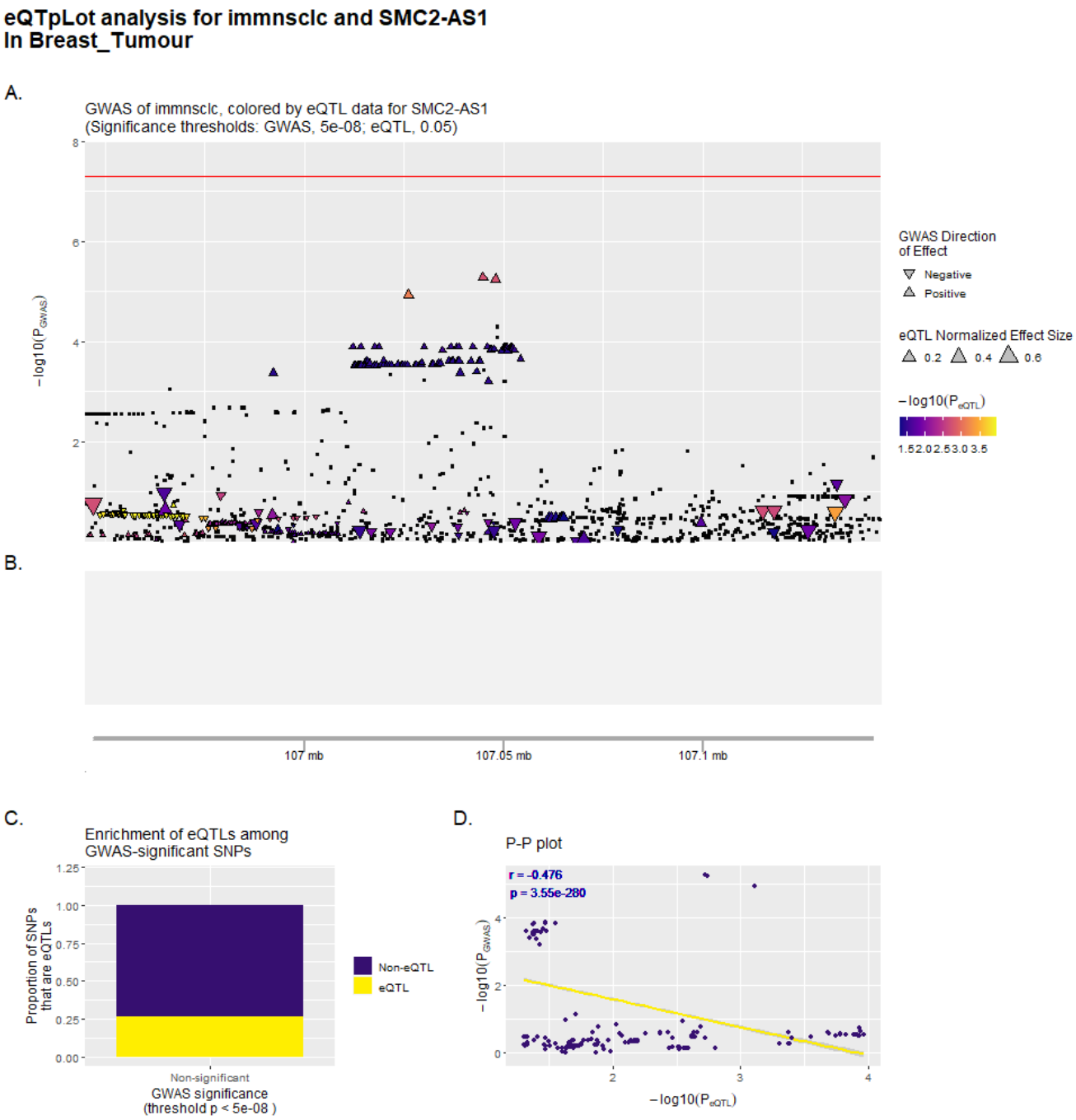


Figure 6.10. eQTpLot of the colocalization of eQTLs for gene *SMC2-AS1* and GWAS signals for Immune NSCLC score on candidate SNP chr9:107044828 in the MyBrCa population.

A Genomic region of interest around chr9:107044828. Colour scale corresponds to the magnitude of association between each SNP and *SMC2-AS1* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of p = 5 x 10^{-8}. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *SMC2-AS1* and Immune NSCLC score with Pearson correlation coefficient (r) and p-value (p).

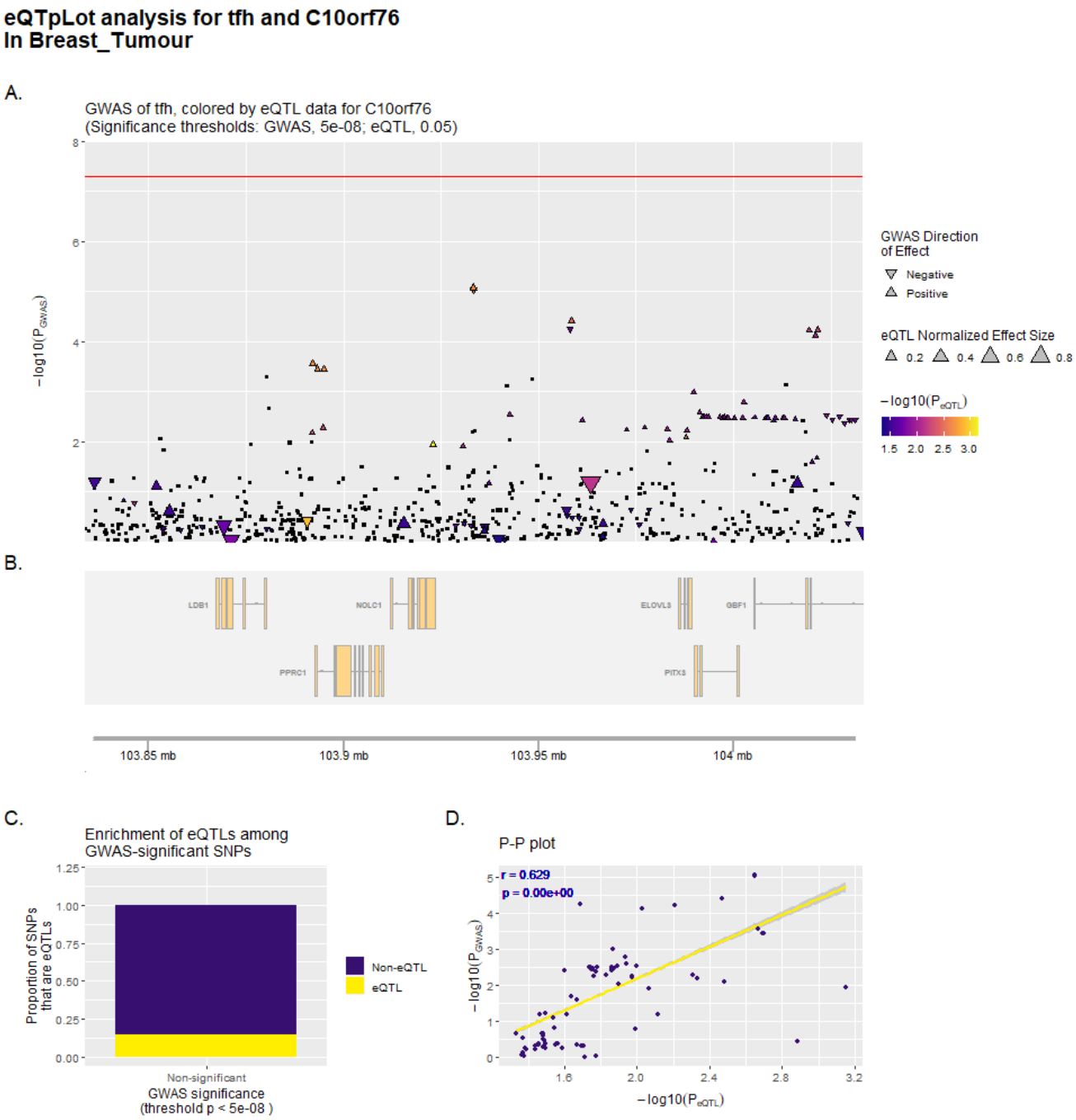


Figure 6.11. eQTpLot of the colocalization of eQTLs for gene *C10orf76* and GWAS signals for Tfh cells on candidate SNP chr10:103933363 in the MyBrCa population.

A Genomic region of interest around chr10:103933363. Colour scale corresponds to the magnitude of association between each SNP and *C10orf76* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of p = 5 x 10^{-8}. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *C10orf76* and Tfh cells with Pearson correlation coefficient (r) and p-value (p).

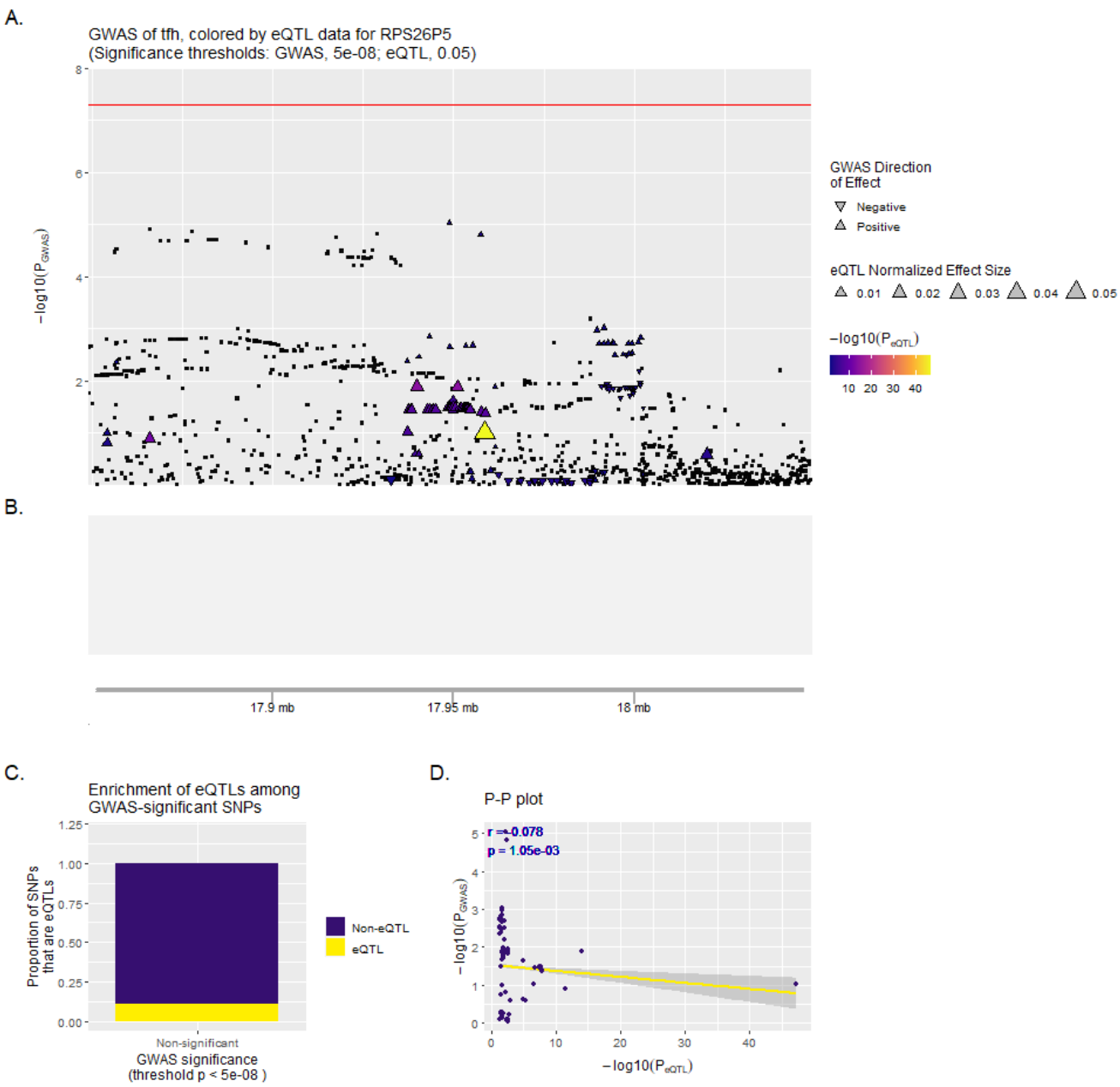


Figure 6.12. eQTpLot of the colocalization of eQTLs for gene *RPS26P5* and GWAS signals for Tfh cells on candidate SNP chr21:17948888 in the MyBrCa population.

A Genomic region of interest around chr21:17948888. Colour scale corresponds to the magnitude of association between each SNP and *RPS26P5* expression. Directionality of triangle represents GWAS direction of effect, size of triangle is the normalised effect size (NES) for the eQTL data. Red horizontal line represents genome wide significance threshold of p = 5 x 10^{-8}. **B** Gene positions. **C** Enrichment of eQTLs among GWAS-significant SNPs. **D** Correlation P-P plot between p_{GWAS} and p_{eQTL} for *RPS26P5* and Tfh cells with Pearson correlation coefficient (r) and p-value (p).

Immune Score	SNP Details						GWAS Summary Statistics				
	chr	pos	Allele A	Allele B	Effect size	P-value	MAF MyBrCa	MAF TCGA	MAF AMR	MAF EAS	MAF EUR
Treg cells	10	6499101	T	A	0.671	1.03E-06	0.028	0.003	NA	0.031	NA
Treg cells	17	16764104	C	T	0.241	7.56E-06	0.226	0.034	0.029	0.281	0.024
TREM1 score	2	241431311	A	G	0.821	1.19E-06	0.019	0.036	0.019	0.011	0.039

Table 4. GWAS summary statistics of variants associated with *TNFRSF13B*, *IL2RA*,*PRKCQ* and *MIR149*.

chr: chromosome; pos: position; MAF: minor allele frequency; AMR: American; EAS: East Asian; EUR: European; NA: Not available

Chapter 4 : Discussion

4.1 Summary of main findings

This study explored the germline genetic variants associated with immune traits in Asian and Caucasian breast tumours. Here, we assessed the genome-wide heritability of 112 immune traits and found that in general, heritability was significantly higher in Asian breast tumours, compared to Caucasian breast tumours. By implementing GWAS analyses, immune traits that exhibited high heritability in both populations were found to have overlapping genomic loci, suggesting a shared genetic influence and its underlying mechanisms, irrespective of population ethnicity. Immune scores with high heritability exclusively in Asian breast tumours were Treg cells, Immune NSCLC score, TREM1 score and Tfh cells. Across the 4 scores, 37 associated candidate SNPs were identified with a total of 561 nearby genes. Twelve significant eQTLs were shortlisted, from which *TNFRSF13B*, *PRKCQ*, *IL2RA*, and *MIR149* have been extensively characterised as key modulators of the immune landscape. Variants associated with *TNFRSF13B*, *PRKCQ* and *IL2RA* had distinct population frequencies, potentially contributing to the observed variation in heritable immune scores between the 2 study populations.

4.2 Genome-wide heritability of immune scores in breast tumours

One of the most notable findings from this analysis is that immune traits were, in general, significantly more heritable in Asian breast tumours compared to its Caucasian counterparts, thus suggesting that there may be a population-specific genetic influence on the immune landscape of breast tumours. These findings were consistent with a comprehensive GWAS study that characterised the immune landscape of 33 different cancer types in TCGA. They report that the top 10 immune traits among cancer patients of European ancestry had

heritability estimates ranging from ~15 to 20% while patients in the Asian ancestry group had estimates ranging from ~77% to 131% (although maximum heritability is 100%) ([47]). The observed differences in immune trait genome-wide heritability between the two ethnic groups could be due to population-specific differences in environmental factors, genetic variances, and the complex interplay between the two factor types ([87-92]). Further studies are warranted to elucidate the specific drivers of variation in the heritability of immune traits across diverse populations.

4.3 Shared genomic loci in highly heritable immune traits in Asian and Caucasian breast tumours

This study investigates whether immune traits with high heritability in both populations share similar genetic architecture that persists across populations, thereby influencing downstream functional mechanisms. This hypothesis was supported by findings in the GWAS analysis. The distinct locus at chr 6 for MHC was surrounded by a cluster of *HLA* variants in both Asian and Caucasian populations. This well-known locus has been reported to be the most polymorphic gene cluster of the entire human genome, among which *HLA-B* is the most polymorphic gene with 511 reported alleles ([93]). The MHC system is essential in T Cell-mediated adaptive immunity as it facilitates antigen presentation, while each variant of the *HLA* gene performs a distinct role to support its overall function within the immune system ([94]). The findings of this study suggest that while there are immune traits with population-specific genetic influences, there are also immune-related genetic determinants that have been conserved across different ethnic populations. These shared determinants consistently contribute to the modulation of the immune landscape of breast tumours.

4.4 Most heritable immune traits in Asian breast tumours

4.4.1 Summary of findings

In this study, Treg cells, Immune NSCLC score, TREM1 score and Tfh cells were immune traits with the highest heritability in the Asian population of breast tumours. Despite not achieving genome-wide significance, 37 candidate SNPs were found to have suggestive associations with the 4 immune traits. Twelve genes were found to be *cis*-eQTLs, out of which 4 have been reported to be associated with immune scores. These genes were *TNFRSF13B*, *PRKCQ*, *IL2RA* of the Treg cells immune score, and *MIR149* of the TREM1 score. Noteworthily, differences in the allelic frequencies for variants associated with Treg cells may explain the difference in heritable immune scores between Asian and Caucasian breast cancer patients.

4.4.2 TNFRSF13B in Treg cells

The *TNFRSF13B* gene encodes the tumour necrosis factor superfamily member 13B, or otherwise also known as TACI, transmembrane activator and CAML (calcium-modulating cyclophilin ligand) interactor ([95]). *TNFRSF13B* plays an essential role in immunoglobulin recombination ([96-99]) and B cell memory, differentiation, proliferation, and survival ([85, 99, 100]). Its dysregulation has been associated with several autoimmune disorders and cancer ([86, 96-98, 101]). While most studies have investigated its role in B cells, one study reported heightened expressions in Treg cells compared to T cells. This elevated expression of *TNFRSF13B* was found to promote immunosuppression in multiple myeloma and was associated with improved overall patient outcomes ([83]). Similar results were also reported for patients with endometroid epithelial ovarian cancer, chronic lymphocytic leukemia, and head and neck squamous cell carcinoma ([102-104]). Conversely, elevated expression of the gene was associated with poor prognosis in patients of melanoma, lymphoma and prostate cancer ([105-107]). In

breast cancer, increased expressions of *TNFRSF13B* and its associated ligands, B-cell Activating Factor (BAFF) and A Proliferation-inducing Ligand (APRIL), have been associated with tumour initiation, progression, and aggressiveness ([108, 109]). Few studies have reported *TNFRSF13B* as a potential therapeutic target and biomarker in breast cancer with some promising results, however, the role of the gene in breast cancer still warrants further characterization ([108-111]). Further investigation into the relationship between *TNFRSF13B* and Treg cells may also be an avenue of interest, as there may be population-specific germline genetic differences contributing to its mechanism in breast cancer.

4.4.3 PRKCQ and IL2RA in Treg cells

Two well-characterised eQTLs on chr 10 in the analysis for Treg cells as an immune score were *PRKCQ* and *IL2RA*. Both genes play unique and vital roles in the T cell signalling pathway involving Treg cells. *IL2RA* is highly expressed in Treg cells to regulate immune responses ([80]) while *PRKCQ* is vital in the coordination of transcriptional networks involved in T cell activation ([82]). The two have been implicated as potential biomarkers for the tumour immune landscape due to their prognostic value.

PRKCQ has been extensively characterized for its prognostic value in breast cancer. The inhibition of *PRKCQ* expression was shown to suppress the proliferation, migration and survival of breast cancer cells. This occurs through the regulation of various intermediary pathways and gene expressions such as the mitogen-activated protein kinase (MAPK) pathway, Fra-1 protein, the epithelial to mesenchymal transition (EMT), miR-203-mediated cell migration, and miR-361/PIK3C3-mediated autophagy ([112-117]). Additionally, the inhibition of *PRKCQ* reportedly enhances chemosensitivity among TNBC patients, potentially improving the effectiveness of chemotherapy among this subgroup ([118]). In a pan-cancer study of the

Protein Kinase C (PKC) Family, *PRKCQ* strongly correlated with the infiltration of tumour immune cells and an elevation of its levels was associated with improved response to HLA-dependent or PD-L1 blocking-dependent immunotherapies ([81]).

Elevated levels of *IL2RA* has been linked to worse infiltration and survival outcomes in several cancers including pancreatic ductal carcinoma and acute myeloid leukaemia, ([79, 119-121]), but favourable prognosis in thyroid cancer patients ([122]). In breast cancer, the modulation of *IL2RA* by the secretory protein coiled-coil domain containing 134 protein (CCDC134) and mir-490 has been linked to immune evasion and disease progression ([123, 124]).

Taken together, *PRKCQ* and *IL2RA* have been extensively characterized for their role in cancer growth and immune regulation. Understanding the interplay between these genes and its associated targeted pathways, especially in a population of Asian breast cancer patients, could provide insights into novel therapeutic strategies, potentially improving treatment outcomes in patients.

4.4.4 MIR149 in TREM1 score

The *MIR149* gene was identified as an eQTL for the candidate SNP in the TREM1 score. MicroRNAs (miRNAs) have garnered significant attention due to their promising therapeutic potential in cancer immunotherapy ([125-131]). While *MIR149* is not known to directly influence TREM1 score, it is known to modulate the AKT1/MTOR pathway as a tumour promoter in gastric cancer ([132]). In TREM1, the inhibition of the AKT1/MTOR pathway has been linked to TREM1-induced metabolic reprogramming in acute lung injury. The MTOR pathway has also been reported to modulate TREM1-mediated mitochondrial fission, mitophagy and necroptosis ([133]). Further studies are warranted to investigate the association between *MIR149* and the TREM1 immune score. *MIR149* also acts as a tumour suppressor gene in

multiple cancers including breast cancer. It does so, at least in part, by downregulating the expression of T-cell inhibitor receptors that are associated with T-cell exhaustion, PD-1, TIM-3, BTLA and FOXP1, to strengthen immune surveillance in CD8+ T-cells ([77]). Additionally, increased *MIR149* expression decreases G-protein-coupled receptor kinase-interacting protein 1 (GIT1), which in turn, supresses breast cancer invasion and metastasis. Its potential to reverse resistance to anticancer drugs, such as paclitaxel and trastuzumab, has made it a promising candidate for therapeutic treatment ([134, 135]).

4.4.5 Future implications

Taken together, this study identified 4 genes that may be associated with the most heritable immune traits in Asian breast tumours. These genes may play a significant role in modulating the immune landscape in different capacities and has been postulated in other studies as potential tumour immune biomarkers and targets for cancer treatment ([77-85, 95, 100, 125-127, 131]). However, further research is warranted still to assess their suitability in a population of Asian breast cancer patients. This includes validating the results of this study in an additional Asian cohort, exploring associations between genetic variants with patient clinical subtypes, conducting molecular studies to elucidate the mechanisms of action, and performing clinical trials to evaluate their effectiveness as biomarkers or therapeutic interventions. Additionally, the relationship between immune scores and breast cancer risk can be examined by exploring associations with established genetic variants such as *BRCA1* and *BRCA2.* Given the distinct genetic architecture of Asian populations and the heightened heritability observed for these specific immune traits, these genes could hold even greater promise in the context of Asian breast tumours.

4.5 Strengths

To the best of my knowledge, this is the first study to investigate the germline genetic determinants driving immune scores in Asian breast tumours, and to explore whether there may be population-specific differences. By comparing the differences in heritability and genetic distribution between Asians and Caucasians populations, this enhanced the validity of the study conclusions and provided insight into any cross-population differences. Additionally, the implementation of a heritability analysis revealed immune traits that were most likely influenced by genetic factors of one population, thus not only setting the stage for all downstream analyses, but potentially for future studies to develop immune biomarkers or treatment strategies catered to that specific population. Through quality control and applying stringent filtering criteria, this ensured robust and reliable germline genotyping and RNAseq data. The use of eQTpLot was useful in visualising the colocalization of GWAS and eQTL signals.

4.6 Limitations

The main limitation of this study would be the small sample size, and therefore studies with larger sample sizes are warranted to strengthen the current conclusions of this study. This study investigated the influence of candidate SNPs on heritable immune scores by exploring changes in their gene expression levels. It is worth noting, however, that gene expression only comprises up to ~15% of heritability ([136]). Hence, while these results may be informative, it is only one piece of the complex puzzle that shapes these immune traits. Other factors such as environmental influences, epigenetics, and post-translational modifications also play a large role in heritability ([137]), and should therefore be investigated in future studies.

Chapter 5 : Conclusion

In conclusion, the data presented in this thesis demonstrate that immune scores are more heritable in Asian breast tumours, as compared to Caucasian breast tumours. Immune traits that were highly heritable in both populations had similar genetic architecture, highlighting shared modulatory genes and mechanisms involved in their function. The results of this study suggest that the tumour immune microenvironment may be modulated by germline genetic variants. Some variants, such as those near immune-related genes like *TNFRSF13B*, *PRKCQ*, *IL2RA*, and *MIR149*, are likely to affect the tumour immune microenvironment by modulating gene expression levels. These genes have been extensively implicated in modulating the immune landscape, and this study particularly highlights their significant contributions to Asian breast tumours, therefore making them promising targets for further research and therapeutic avenues for the treatment of breast cancer in the Asian population.

Chapter 6 : References

1. Sung H, Ferlay J, Siegel RL*, et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. CA Cancer J Clin*. 2021; **71**(3): 209-49.

2. Lim YX, Lim ZL, Ho PJ*, et al.* Breast Cancer in Asia: Incidence, Mortality, Early Detection, Mammography Programs, and Risk-Based Screening Initiatives*. Cancers (Basel)*. 2022; **14**(17).

3. Heer E, Harper A, Escandor N*, et al.* Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study*. The Lancet Global Health*. 2020; **8**(8): e1027-e37.

4. Rivera-Franco MM, Leon-Rodriguez E. Delays in Breast Cancer Detection and Treatment in Developing Countries*. Breast Cancer (Auckl)*. 2018; **12**: 1178223417752677.

5. da Costa Vieira RA, Biller G, Uemura G*, et al.* Breast cancer screening in developing countries*. Clinics (Sao Paulo)*. 2017; **72**(4): 244-53.

6. Kashyap D, Pal D, Sharma R*, et al.* Global Increase in Breast Cancer Incidence: Risk Factors and Preventive Measures*. Biomed Res Int*. 2022; **2022**: 9605439.

7. Fidler MM, Soerjomataram I, Bray F. A global view on cancer incidence and national levels of the human development index*. Int J Cancer*. 2016; **139**(11): 2436-46.

8. Malaysia MoH. Malaysia National Cancer Registry Report 2012 - 2016. Ministry of Health Malaysia; 2019.

9. Hanahan D, Robert. Hallmarks of Cancer: The Next Generation*. Cell*. 2011; **144**(5): 646-74.

10. Fife BT, Pauken KE, Eagar TN*, et al.* Interactions between PD-1 and PD-L1 promote tolerance by blocking the TCR–induced stop signal*. Nature Immunology*. 2009; **10**(11): 1185-92.

11. Tivol EA, Borriello F, Schweitzer AN*, et al.* Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4*. Immunity*. 1995; **3**(5): 541-7.

12. Schmid P, Rugo HS, Adams S*, et al.* Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial*. Lancet Oncol*. 2020; **21**(1): 44-59.

13. Schmid P. ESMO 2018 presidential symposium-IMpassion130: atezolizumab+nab-paclitaxel in triple-negative breast cancer*. ESMO Open*. 2018; **3**(6): e000453.

14. Emens LA, Molinero L, Loi S*, et al.* Atezolizumab and nab-Paclitaxel in Advanced Triple-Negative Breast Cancer: Biomarker Evaluation of the IMpassion130 Study*. J Natl Cancer Inst*. 2021; **113**(8): 1005-16.

15. Cortes J, Cescon DW, Rugo HS*, et al.* Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial*. Lancet*. 2020; **396**(10265): 1817-28.

16. Garrido-Castro AC, Lin NU, Polyak K. Insights into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment*. Cancer Discov*. 2019; **9**(2): 176-98.

17. Cardoso F, Senkus E, Costa A*, et al.* 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4)dagger*. Ann Oncol*. 2018; **29**(8): 1634-57.

18. Haslam A, Prasad V. Estimation of the Percentage of US Patients With Cancer Who Are Eligible for and Respond to Checkpoint Inhibitor Immunotherapy Drugs. *JAMA Network Open*. 2019; **2**(5): e192535.

19. Adams S, Gatti-Mays ME, Kalinsky K*, et al.* Current Landscape of Immunotherapy in Breast Cancer: A Review. *JAMA Oncol*. 2019; **5**(8): 1205-14.

20. Bates JP, Derakhshandeh R, Jones L*, et al.* Mechanisms of immune evasion in breast cancer. *BMC Cancer*. 2018; **18**(1).

21. Barzaman K, Moradi-Kalbolandi S, Hosseinzadeh A*, et al.* Breast cancer immunotherapy: Current and novel approaches. *Int Immunopharmacol*. 2021; **98**: 107886.

22. Snyder A, Makarov V, Merghoub T*, et al.* Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *New England Journal of Medicine*. 2014; **371**(23): 2189-99.

23. Bedognetti D, Maccalli C, Al Bader SBJ*, et al.* Checkpoint Inhibitors and Their Application in Breast Cancer. *Breast Care*. 2016; **11**(2): 108-15.

24. Kalbasi A, Ribas A. Tumour-intrinsic resistance to immune checkpoint blockade. *Nat Rev Immunol*. 2020; **20**(1): 25-39.

25. Thorsson V, Gibbs DL, Brown SD*, et al.* The Immune Landscape of Cancer. *Immunity*. 2018; **48**(4): 812-30.e14.

26. Gajic ZZ, Deshpande A, Legut M*, et al.* Recurrent somatic mutations as predictors of immunotherapy response. *Nat Commun*. 2022; **13**(1): 3938.

27. Lu S, Stein JE, Rimm DL*, et al.* Comparison of Biomarker Modalities for Predicting Response to PD-1/PD-L1 Checkpoint Blockade. *JAMA Oncology*. 2019; **5**(8): 1195.

28. Mlecnik B, Bindea G, Pages F*, et al.* Tumor immunosurveillance in human cancers. *Cancer Metastasis Rev*. 2011; **30**(1): 5-12.

29. Rueda OM, Sammut SJ, Seoane JA*, et al.* Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups*. Nature*. 2019; **567**(7748): 399-404.

30. Perou CM, Sorlie T, Eisen MB*, et al.* Molecular portraits of human breast tumours*. Nature*. 2000; **406**(6797): 747-52.

31. Galon J, Pages F, Marincola FM*, et al.* The immune score as a new possible approach for the classification of cancer*. J Transl Med*. 2012; **10**: 1.

32. Mlecnik B, Tosolini M, Kirilovsky A*, et al.* Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *J Clin Oncol*. 2011; **29**(6): 610-8.

33. Mezheyeuski A, Backman M, Mattsson J*, et al.* An immune score reflecting pro- and anti-tumoural balance of tumour microenvironment has major prognostic impact and predicts immunotherapy response in solid cancers*. EBioMedicine*. 2023; **88**: 104452.

34. Bindea G, Mlecnik B, Tosolini M*, et al.* Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer*. Immunity*. 2013; **39**(4): 782-95.

35. Yoshihara K, Shahmoradgoli M, Martínez E*, et al.* Inferring tumour purity and stromal and immune cell admixture from expression data*. Nature Communications*. 2013; **4**(1).

36. Auslander N, Zhang G, Lee JS*, et al.* Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma*. Nature Medicine*. 2018; **24**(10): 1545-9.

37. Ayers M, Lunceford J, Nebozhyn M*, et al.* IFN-γ–related mRNA profile predicts clinical response to PD-1 blockade*. Journal of Clinical Investigation*. 2017; **127**(8): 2930-40.

38. McGranahan N, Furness AJ, Rosenthal R*, et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade*. Science*. 2016; **351**(6280): 1463-9.

39. Fridman WH, Pages F, Sautes-Fridman C*, et al.* The immune contexture in human tumours: impact on clinical outcome*. Nat Rev Cancer*. 2012; **12**(4): 298-306.

40. Angell H, Galon J. From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer*. Curr Opin Immunol*. 2013; **25**(2): 261-7.

41. Galon J, Fridman WH, Pages F. The adaptive immunologic microenvironment in colorectal cancer: a novel perspective*. Cancer Res*. 2007; **67**(5): 1883-6.

42. Marty R, Kaabinejadian S, Rossell D*, et al.* MHC-I Genotype Restricts the Oncogenic Mutational Landscape*. Cell*. 2017; **171**(6): 1272-83.e15.

43. Marty Pyke R, Thompson WK, Salem RM*, et al.* Evolutionary Pressure against MHC Class II Binding Cancer Mutations*. Cell*. 2018; **175**(2): 416-28.e13.

44. Roederer M, Quaye L, Mangino M*, et al.* The Genetic Architecture of the Human Immune System: A Bioresource for Autoimmunity and Disease Pathogenesis*. Cell*. 2015; **161**(2): 387-403.

45. Orrù V, Steri M, Sole G*, et al.* Genetic Variants Regulating Immune Cell Levels in Health and Disease*. Cell*. 2013; **155**(1): 242-56.

46. Shahamatdar S, He MX, Reyna MA*, et al.* Germline Features Associated with Immune Infiltration in Solid Tumors*. Cell Rep*. 2020; **30**(9): 2900-8 e4.

47. Sayaman RW, Saad M, Thorsson V*, et al.* Germline genetic contribution to the immune landscape of cancer*. Immunity*. 2021; **54**(2): 367-86 e8.

48. Pan J-W, Zabidi MMA, Ng P-S*, et al.* The molecular landscape of Asian breast cancers reveals clinically relevant population-specific differences*. Nature Communications*. 2020; **11**(1).

49. Lim YW, Chen-Harris H, Mayba O*, et al.* Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration*. Proc Natl Acad Sci U S A*. 2018; **115**(50): E11701-E10.

50. Kraya AA, Maxwell KN, Wubbenhorst B*, et al.* Genomic Signatures Predict the Immunogenicity of BRCA-Deficient Breast Cancer*. Clinical Cancer Research*. 2019; **25**(14): 4363-74.

51. Tan M-M, Ho W-K, Yoon S-Y*, et al.* A case-control study of breast cancer risk factors in 7,663 women in Malaysia*. PLOS ONE*. 2018; **13**(9): e0203469.

52. Comprehensive molecular portraits of human breast tumours*. Nature*. 2012; **490**(7418): 61-70.

53. Amos CI, Dennis J, Wang Z*, et al.* The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers*. Cancer Epidemiol Biomarkers Prev*. 2017; **26**(1): 126-35.

54. Michailidou K, Lindstrom S, Dennis J*, et al.* Association analysis identifies 65 new breast cancer risk loci*. Nature*. 2017; **551**(7678): 92-4.

55. Michailidou K, Hall P, Gonzalez-Neira A*, et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk*. Nat Genet*. 2013; **45**(4): 353-61, 61e1-2.

56. Ho WK, Tan MM, Mavaddat N*, et al.* European polygenic risk score for prediction of breast cancer shows similar performance in Asian women*. Nat Commun*. 2020; **11**(1): 3833.

57. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009; **5**(6): e1000529.

58. International Genome Sample Resource Phase 3: International Genome Sample Resource; [Available from: https://www.internationalgenome.org/category/phase-3/.

59. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011; **9**(2): 179-81.

60. Korn JM, Kuruvilla FG, McCarroll SA*, et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet*. 2008; **40**(10): 1253-60.

61. Chang CL, Villalun M, Geib SM*, et al.* Pupal X-ray irradiation influences protein expression in adults of the oriental fruit fly, Bactrocera dorsalis. *J Insect Physiol*. 2015; **76**: 7-16.

62. Dobin A, Davis CA, Schlesinger F*, et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; **29**(1): 15-21.

63. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011; **12**: 323.

64. Newman AM, Liu CL, Green MR*, et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015; **12**(5): 453-7.

65. Li B, Severson E, Pignon JC*, et al.* Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016; **17**(1): 174.

66. Eisenberg E, Levanon EY. Human housekeeping genes are compact. *Trends Genet*. 2003; **19**(7): 362-5.

67. Iyer G, Wang AR, Brennan SR*, et al.* Identification of stable housekeeping genes in response to ionizing radiation in cancer research*. Sci Rep*. 2017; **7**: 43763.

68. Janssens N, Janicot M, Perera T*, et al.* Housekeeping genes as internal standards in cancer research*. Mol Diagn*. 2004; **8**(2): 107-13.

69. Chua SL, See Too WC, Khoo BY*, et al.* UBC and YWHAZ as suitable reference genes for accurate normalisation of gene expression using MCF7, HCT116 and HepG2 cell lines*. Cytotechnology*. 2011; **63**(6): 645-54.

70. Tilli TM, Castro Cda S, Tuszynski JA*, et al.* A strategy to identify housekeeping genes suitable for analysis in breast cancer diseases*. BMC Genomics*. 2016; **17**(1): 639.

71. Yang J, Lee SH, Goddard ME*, et al.* GCTA: A Tool for Genome-wide Complex Trait Analysis*. The American Journal of Human Genetics*. 2011; **88**(1): 76-82.

72. Yang J, Benyamin B, McEvoy BP*, et al.* Common SNPs explain a large proportion of the heritability for human height*. Nature Genetics*. 2010; **42**(7): 565-9.

73. Pruim RJ, Welch RP, Sanna S*, et al.* LocusZoom: regional visualization of genome-wide association scan results*. Bioinformatics*. 2010; **26**(18): 2336-7.

74. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms*. Bioinformatics*. 2009; **25**(5): 655-61.

75. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations*. Bioinformatics*. 2012; **28**(10): 1353-8.

76. Drivas TG, Lucas A, Ritchie MD. eQTpLot: a user-friendly R package for the visualization of colocalization between eQTL and GWAS signals*. BioData Min*. 2021; **14**(1): 32.

77. Zhang M, Gao D, Shi Y*, et al.* miR-149-3p reverses CD8(+) T-cell exhaustion by reducing inhibitory receptors and promoting cytokine secretion in breast cancer cells*. Open Biol*. 2019; **9**(10): 190061.

78. Ren FJ, Yao Y, Cai XY*, et al.* MiR-149-5p: An Important miRNA Regulated by Competing Endogenous RNAs in Diverse Human Cancers*. Front Oncol*. 2021; **11**: 743077.

79. Fan L, Wang X, Chang Q*, et al.* IL2RA is a prognostic indicator and correlated with immune characteristics of pancreatic ductal adenocarcinoma*. Medicine (Baltimore)*. 2022; **101**(42): e30966.

80. O'Gorman WE, Dooms H, Thorne SH*, et al.* The initial phase of an immune response functions to activate regulatory T cells*. J Immunol*. 2009; **183**(1): 332-9.

81. Abdelatty A, Sun Q, Hu J*, et al.* Pan-Cancer Study on Protein Kinase C Family as a Potential Biomarker for the Tumors Immune Landscape and the Response to Immunotherapy*. Front Cell Dev Biol*. 2021; **9**: 798319.

82. Brezar V, Tu WJ, Seddiki N. PKC-Theta in Regulatory and Effector T-cell Functions*. Front Immunol*. 2015; **6**: 530.

83. Tai YT, Lin L, Xing L*, et al.* APRIL signaling via TACI mediates immunosuppression by T regulatory cells in multiple myeloma: therapeutic implications*. Leukemia*. 2019; **33**(2): 426-38.

84. von Bulow GU, van Deursen JM, Bram RJ. Regulation of the T-independent humoral response by TACI*. Immunity*. 2001; **14**(5): 573-82.

85. de Mattos Barbosa MG, Lefferts AR, Huynh D*, et al.* TNFRSF13B genotypes control immune-mediated pathology by regulating the functions of innate B cells*. JCI Insight*. 2021; **6**(17).

86. Salzer U, Chapel HM, Webster AD*, et al.* Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans*. Nat Genet*. 2005; **37**(8): 820-8.

87. Visscher PM, Medland SE, Ferreira MA*, et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings*. PLoS Genet*. 2006; **2**(3): e41.

88. Mangino M, Roederer M, Beddall MH*, et al.* Innate and adaptive immune traits are differentially affected by genetic and environmental factors*. Nat Commun*. 2017; **8**: 13850.

89. Visscher PM, Gordon S, Neale MC. Power of the classical twin design revisited: II detection of common environmental variance*. Twin Res Hum Genet*. 2008; **11**(1): 48-54.

90. Docherty AR, Kremen WS, Panizzon MS*, et al.* Comparison of Twin and Extended Pedigree Designs for Obtaining Heritability Estimates*. Behav Genet*. 2015; **45**(4): 461-6.

91. Roederer M, Quaye L, Mangino M*, et al.* The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis*. Cell*. 2015; **161**(2): 387-403.

92. Gielen M, Lindsey PJ, Derom C*, et al.* Modeling genetic and environmental factors to increase heritability and ease the identification of candidate genes for birth weight: a twin study*. Behav Genet*. 2008; **38**(1): 44-54.

93. Mungall AJ, Palmer SA, Sims SK*, et al.* The DNA sequence and analysis of human chromosome 6*. Nature*. 2003; **425**(6960): 805-11.

94. Charles A Janeway J, Paul Travers, Mark Walport, and Mark J Shlomchik. Immunobiology: The Immune System in Health and Disease. 5th ed. New York: Garland Science; 2001.

95. von Bulow GU, Bram RJ. NF-AT activation induced by a CAML-interacting member of the tumor necrosis factor receptor superfamily. *Science*. 1997; **278**(5335): 138-41.

96. Castigli E, Wilson SA, Garibyan L*, et al.* TACI is mutant in common variable immunodeficiency and IgA deficiency. *Nat Genet*. 2005; **37**(8): 829-34.

97. Zhang J, van Oostrom D, Li J*, et al.* Innate Mechanisms in Selective IgA Deficiency. *Front Immunol*. 2021; **12**: 649112.

98. Yang C, Jie W, Yanlong Y*, et al.* Genome-wide association study identifies TNFSF13 as a susceptibility gene for IgA in a South Chinese population in smokers. *Immunogenetics*. 2012; **64**(10): 747-53.

99. Tsuji S, Cortesao C, Bram RJ*, et al.* TACI deficiency impairs sustained Blimp-1 expression in B cells decreasing long-lived plasma cells in the bone marrow. *Blood*. 2011; **118**(22): 5832-9.

100. Romberg N, Virdee M, Chamberlain N*, et al.* TNF receptor superfamily member 13b (TNFRSF13B) hemizygosity reveals transmembrane activator and CAML interactor haploinsufficiency at later stages of B-cell development. *J Allergy Clin Immunol*. 2015; **136**(5): 1315-25.

101. Cumbo C, Orsini P, Tarantini F*, et al.* TNFRSF13B gene mutation in familial acute myeloid leukemia: A new piece in the complex scenario of hereditary predisposition? *Hematol Oncol*. 2023.

102. Wang Y, Xu Y, Hua Q*, et al.* Novel Prognostic Model Based on Immune Signature for Head and Neck Squamous Cell Carcinoma. *Biomed Res Int*. 2020; **2020**: 4725314.

103. Mamara A, Germenis AE, Kompoti M*, et al.* TACI expression and signaling in chronic lymphocytic leukemia. *J Immunol Res*. 2015; **2015**: 478753.

104. Block MS, Charbonneau B, Vierkant RA*, et al.* Variation in NF-kappaB signaling pathways and survival in invasive epithelial ovarian cancer*. Cancer Epidemiol Biomarkers Prev*. 2014; **23**(7): 1421-7.

105. Moreaux J, Veyrune JL, De Vos J*, et al.* APRIL is overexpressed in cancer: link with tumor progression*. BMC Cancer*. 2009; **9**: 83.

106. Li CY, Huang SP, Chen YT*, et al.* TNFRSF13B is a potential contributor to prostate cancer*. Cancer Cell Int*. 2022; **22**(1): 180.

107. Zhang C, Dang D, Cong L*, et al.* Pivotal factors associated with the immunosuppressive tumor microenvironment and melanoma metastasis*. Cancer Med*. 2021; **10**(14): 4710-20.

108. Garcia-Castro A, Zonca M, Florindo-Pinheiro D*, et al.* APRIL promotes breast tumor growth and metastasis and is associated with aggressive basal breast cancer*. Carcinogenesis*. 2015; **36**(5): 574-84.

109. Kampa M, Notas G, Stathopoulos EN*, et al.* The TNFSF Members APRIL and BAFF and Their Receptors TACI, BCMA, and BAFFR in Oncology, With a Special Focus in Breast Cancer*. Front Oncol*. 2020; **10**: 827.

110. Hinterleitner C, Zhou Y, Tandler C*, et al.* Platelet-Expressed TNFRSF13B (TACI) Predicts Breast Cancer Progression*. Front Oncol*. 2021; **11**: 642170.

111. Abo-Elfadl MT, Gamal-Eldeen AM, Ismail MF*, et al.* Silencing of the cytokine receptor TNFRSF13B: A new therapeutic target for triple-negative breast cancer*. Cytokine*. 2020; **125**: 154790.

112. Byerly J, Halstead-Nussloch G, Ito K*, et al.* PRKCQ promotes oncogenic growth and anoikis resistance of a subset of triple-negative breast cancer cells*. Breast Cancer Res*. 2016; **18**(1): 95.

113. Belguise K, Milord S, Galtier F*, et al.* The PKCtheta pathway participates in the aberrant accumulation of Fra-1 protein in invasive ER-negative breast cancer cells*. Oncogene*. 2012; **31**(47): 4889-97.

114. Belguise K, Cherradi S, Sarr A*, et al.* PKCtheta-induced phosphorylations control the ability of Fra-1 to stimulate gene expression and cancer cell migration*. Cancer Lett*. 2017; **385**: 97-107.

115. Salem S, Mosaad R. Crosstalk between miR-203 and PKCtheta regulates breast cancer stem cell markers*. Ann Hum Genet*. 2021; **85**(3-4): 105-14.

116. Zheng S, Fu W, Huang Q*, et al.* LncRNA PRKCQ-AS1 regulates paclitaxel resistance in triple-negative breast cancer cells through miR-361-5p/PIK3C3 mediated autophagy*. Clin Exp Pharmacol Physiol*. 2023; **50**(6): 431-42.

117. Zafar A, Wu F, Hardy K*, et al.* Chromatinized protein kinase C-theta directly regulates inducible genes in epithelial to mesenchymal transition and breast cancer stem cells*. Mol Cell Biol*. 2014; **34**(16): 2961-80.

118. Byerly JH, Port ER, Irie HY. PRKCQ inhibition enhances chemosensitivity of triple-negative breast cancer by regulating Bim*. Breast Cancer Res*. 2020; **22**(1): 72.

119. Kuhn DJ, Dou QP. The role of interleukin-2 receptor alpha in cancer*. Front Biosci*. 2005; **10**: 1462-74.

120. Du W, He J, Zhou W*, et al.* High IL2RA mRNA expression is an independent adverse prognostic biomarker in core binding factor and intermediate-risk acute myeloid leukemia*. J Transl Med*. 2019; **17**(1): 191.

121. Nguyen CH, Schlerka A, Grandits AM*, et al.* IL2RA Promotes Aggressiveness and Stem Cell-Related Properties of Acute Myeloid Leukemia*. Cancer Res*. 2020; **80**(20): 4527-39.

122. Pan Z, Bao L, Lu X*, et al.* IL2RA(+)VSIG4(+) tumor-associated macrophage is a key subpopulation of the immunosuppressive microenvironment in anaplastic thyroid cancer*. Biochim Biophys Acta Mol Basis Dis*. 2023; **1869**(1): 166591.

123. Seif F, Vaseghi H, Ariana M*, et al.* Overexpression of miR-490-5p/miR-490-3p Potentially Induces IL-17-Producing T Cells in Patients With Breast Cancer*. Eur J Breast Health*. 2022; **18**(2): 141-7.

124. Huang Z, Zhen S, Jin L*, et al.* miRNA-1260b Promotes Breast Cancer Cell Migration and Invasion by Downregulating CCDC134*. Curr Gene Ther*. 2023; **23**(1): 60-71.

125. Omar HA, El-Serafi AT, Hersi F*, et al.* Immunomodulatory MicroRNAs in cancer: targeting immune checkpoints and the tumor microenvironment*. FEBS J*. 2019; **286**(18): 3540-57.

126. Kipkeeva F, Muzaffarova T, Korotaeva A*, et al.* The Features of Immune Checkpoint Gene Regulation by microRNA in Cancer*. Int J Mol Sci*. 2022; **23**(16).

127. Zhou H, Jia W, Lu L*, et al.* MicroRNAs with Multiple Targets of Immune Checkpoints, as a Potential Sensitizer for Immune Checkpoint Inhibitors in Breast Cancer Treatment*. Cancers (Basel)*. 2023; **15**(3).

128. O'Brien J, Hayder H, Zayed Y*, et al.* Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation*. Front Endocrinol (Lausanne)*. 2018; **9**: 402.

129. Han R, Zhao J, Lu L. MicroRNA-34a expression affects breast cancer invasion in vitro and patient survival via downregulation of E2F1 and E2F3 expression*. Oncol Rep*. 2020; **43**(6): 2062-72.

130. Han R, Chen X, Li Y*, et al.* MicroRNA-34a suppresses aggressiveness of hepatocellular carcinoma by modulating E2F1, E2F3, and Caspase-3*. Cancer Manag Res*. 2019; **11**: 2963-76.

131. Zhang Y, Tanno T, Kanellopoulou C. Cancer therapeutic implications of microRNAs in the regulation of immune checkpoint blockade. *ExRNA*. 2019; **1**(1).

132. Zhang X, Wang S, Wang H*, et al.* Circular RNA circNRIP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway. *Mol Cancer*. 2019; **18**(1): 20.

133. Zhong WJ, Zhang J, Duan JX*, et al.* TREM-1 triggers necroptosis of macrophages through mTOR-dependent mitochondrial fission during acute lung injury. *J Transl Med*. 2023; **21**(1): 179.

134. Tian D, Tian M, Ma ZM*, et al.* Anesthetic propofol epigenetically regulates breast cancer trastuzumab resistance through IL-6/miR-149-5p axis. *Sci Rep*. 2020; **10**(1): 8858.

135. Xiang F, Fan Y, Ni Z*, et al.* Ursolic Acid Reverses the Chemoresistance of Breast Cancer Cells to Paclitaxel by Targeting MiRNA-149-5p/MyD88. *Front Oncol*. 2019; **9**: 501.

136. Yao DW, O'Connor LJ, Price AL*, et al.* Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet*. 2020; **52**(6): 626-33.

137. Mayhew AJ, Meyre D. Assessing the Heritability of Complex Traits in Humans: Methodological Challenges and Opportunities. *Curr Genomics*. 2017; **18**(4): 332-40.