博士論文

Identification of genetic variants associated with hormonal receptor positive breast cancer in the Japanese population

(日本人集団におけるホルモン受容体陽性乳がんに関わる遺伝子多型の同定)

エルガザール　シーハム　アブデル　ハルク

Elgazzar Seham Abdel khalk

CONTENTS

Abbreviations

MAF	minor allele frequency
GWAS	genome-wide association study
SNP	single nucleotide polymorphism
BBJ	the Biobank Japan
SIAH2	seven in absentia homolog 2
bp	base-pair
DNA	deoxyribonucleic acid
LD	linkage disequilibrium
OR	odds ratio
QC	quality control
FGFR2	fibroblast growth factor receptor 2
pGL3A_Major/Minor	pGL3 basic plasmid having Construct of region A (-2,151 and -571) with major or minor allele of rs2018246.
pGL3B_Major/Minor	pGL3 promoter plasmid having Construct of region B(+10,063 and +11,152) with major or minor allele of rs16862837.
pGL3C_Major/Minor	pGL3 promoter plasmid having Construct of region C(+13,956 and 15,864) with major or minor allele of rs72008905
pGL3B290_Major/Minor	pGL3 promoter plasmid having 290 bp Construct of region B (+10,267 and +10,556) with major or minor allele of rs16862837.
pGL3B60_Major/Minor	pGL3 promoter plasmid having 60 bp Construct of region B (+10,413and +10,470) with major or minor allele of rs16862837.
pGL3B290_Major/Minor_mutant	pGL3 promoter plasmid having 290 bp Construct of region B (+10,267 and +10,556) with mutant major or minor allele of rs1533711.

Abstract

To identify genetic variants associated with the disease susceptibility, I performed a genome-wide association study (GWAS) using a total of 1086 Japanese female patients with hormonal receptor-positive (HRP) breast cancer and 1816 female controls. I selected 33 single-nucleotide polymorphisms (SNPs) with suggestive associations in GWAS (P-value less than 1×10^{-4}) as well as 4 SNPs that were previously implicated their association with breast cancer for further replication by an independent set of 1653 cases and 2797 controls. I identified significant association of the disease with a SNP rs6788895 ($P_{combined} = 9.43 \times 10^{-8}$ with odds ratio (OR) = 1.22) in the *SIAH*2 (seven in absentia homolog 2) gene on chromosome 3q25.1. In addition, a SNP rs3750817 in intron 2 of the *FGFR2* (fibroblast growth factor receptor 2) gene, which was reported to be associated with breast cancer susceptibility, was significantly replicated ($P_{combined} = 8.47 \times 10^{-8}$ with OR =1.22). To identify functional variations of the *SIAH2* gene, among the SNPs showing strong LD with rs6788895. I focused on three SNPs, SNP X, SNP Y, and SNP Z that are located in the transcription factor binding sites. I prepared six kinds of reporter plasmids carrying either major or minor allele of each SNP. Subsequent reporter gene assay disclosed that the minor non-susceptible T allele of SNP Y enhanced the reporter activity of *SIAH2*, suggesting that this variant might affect the expression or induction of *SIAH2*.

Chapter1 Introduction / Background

Identification of genetic loci/genes associated with common diseases

Variations in the human genome are associated with phenotypes and characteristics of individuals including color of skin and hair, height, blood type, the sensitivity of diseases, and response to drugs. Therefore, these variations have been used for the identification of responsible genes of hereditary diseases and genes associated with common diseases and adverse effect of drugs. Among the variations, single nucleotide polymorphism or SNP is a DNA sequence variation occurring when a single nucleotide- A, T, C, or G - in the genome differs between the individual. Within a population, almost all SNPs have only two alleles; the less common variant is termed a minor allele, while the more common variant is termed a major allele. Minor allele frequency (MAF) is the frequency carrying a minor allele in the population, and the MAF of each SNP is different among ethnic groups. SNPs are localized within coding or non-coding regions in exons, introns, 5`- or 3`-flanking regions of genes, or intergenic regions. Among the SNPs allocated incoding regions, those that do not change the amino acid sequence of the predicted protein are termed synonymous SNPs, while those that change the sequence are termed non-synonymous SNPs. In addition, SNPs located out of protein coding regions may affect gene splicing, transcription factor binding, or the sequence of non-coding RNA.

Genome-wide association studies (GWAS) have been applied for the identification of loci/genes involved in common diseases and response to treatment, and loci/genes associated with adverse effect of drugs. In this method, the frequencies of SNPs across the genome are compared between individuals with a particular phenotype and those without the phenotype. The advancement of typing technologies enabled to investigate thousands of SNPs at the same time. This approach has successfully identified SNPs related to several

complex conditions[1] such as diabetes[2-4], heart abnormalities[5-7], Parkinson disease[8, 9], and Crohn's disease[10-12]. Additionally, GWA studies have identified more than 200 susceptibility loci associated with different types of human cancer[13]. The SNPs identified by GWAS may be applicable for the prediction and prevention of human diseases, and used for the identification of biomarkers of drug efficacy and adverse effect.

The Biobank Japan Project

The Biobank Japan (BBJ) project started as a leading project of Ministry of Education, Culture, Sports, Science and Technology, Japan in June 2003. This project was conducted by the collaboration of the Institute of Medical Sciences, the University of Tokyo, RIKEN, and 66 collaborative hospitals. BBJ has successfully collected genomic DNA, serum, and clinical information of about 200,000 patients with different diseases. The collection covers 47 human diseases including 10 cancers. The aim of this project include (1) discovery genes that susceptible to diseases and or drug response (pharmacogenomics), (2) identification of molecular targets for evidence based development of drugs or diagnostic tools, (3) identification of important genetic information that implemented in the establishment of "Personalized Medicine", and (4) studies on gene-environment interaction for prevention of diseases[14].

Breast cancer

Breast cancer originates from the epithelial cells in mammary gland.The gland is made up of different tissues including lobules with tiny tube-like structures generating milk, ducts carrying the milk from the lobes to the nipple, blood and lymphatic vessels; blood and lymphatic vessels serving for the delivery of oxygen and nutrients to the cells and drainage of bicarbonates and wastes, and the lymphatic vessels connecting lymph nodes for immune systems and coping with infection[15]. Breast cancer comprises of two main types, non-invasive types and invasive types, and rare types. Non-invasive types include ductal

carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS), while invasive types include invasive lobular and invasive ductal carcinoma. In addition, rare types include inflammatory breast cancer, medullary breast cancer, mucinous breast cancer, metaplastic breast cancer and papillary breast cancer. The treatment of patients with breast cancer is decided by multiple factors including stage and grade of the disease, and the presence or absence of receptor expression such as estrogen receptor, progesterone receptor and HER2/new receptor in cancer cells. “Triple negative” cancer indicates a tumor that do not express any of these three receptors[16].

Hormones including estrogen and progesterone have key roles both in the development and progression of the disease[17-18]. Nearly 70% of breast cancersare known to express estrogen and progesterone receptors and their growth is dependent of the hormones. The exposures to higher level and/or for longer period of estrogen such as early menarche, late menopause, late age at first pregnancy, nulliparity, postmenopausal obesity and high serum estrogen level in postmenopausal women is considered to be risk factors for breast cancer[19-21]. Furthermore, progestin, synthetic progesterone, was shown to markedly increase the risk of breast cancer in postmenopausal women when this hormonal therapy was provided for 10 years[22]. In addition, genetic factors play an important role in the development of breast cancer. Although germ-line mutations in *BRCA1* and *BRCA2*, and rare variants in *CHEK2* and *ATM* have been shown to be implicated in high and moderate risks to breast cancer, respectively, several as yet unidentified common susceptibility variants are associated with low to moderate risk of the tumor[23-25].

Clarification of the link between breast carcinogenesis and SNPs

GWAS issued for the identification of loci or genes that are statistically associated with specific phenotypes such as the sensitivities to drugs and the risks to disease through

genotyping thousands of SNPs[26-27]. Although GWAS enables to find a list of SNPs that are statistically associated with a phenotype of interest, it will not provide any information about the biological processes that link the SNPs to the phenotype. One of the major challenges to post GWAS research is to find the functional or causative SNPs associated with the phenotype and the mechanism(s) underlying the process. To resolve the link between the marker SNP and cancer, I searched additional SNPs that are in perfect linkage disequilibrium with the marker SNP. However, the clarification of functional link between the SNPs and phenotype is a hard task, because all SNPs with complete LD are located in the introns or intergenic regions. I hypothesized that if a SNP alters the ability of binding to the transcription factor, the variation should result in the different expression of the gene[28]**.** To identify the loci/genes associated with hormone receptor positive (HRP) breast cancer, I carried out a GWAS and a replication study, and successfully identified SNPs that are significantly associated with the disease. Among the SNPs, the most significantly associated SNP was located in the intron 1 of *SIAH2*. To resolve the link between the marker SNP and cancer, I searched additional SNPs that are in perfect linkage disequilibrium with the marker SNP. Among these SNPs, I focused on SNPs within or close to the transcription factor binding elements, and analyzed the effect of variants on transcriptional activity using reporter assay.

Chapter 2 Aims of my study

In Japan, breast cancer is the most common cancer among women and its incidence has been doubled in both pre- and post-menopausal women in the last 20 years. The marked increase is caused by the increase of estrogen receptor-positive subgroup[29]. Surgical removal of breast cancer is the main therapeutic modality, but patients at advanced stages or those with recurrent diseases need chemotherapy and/or radiotherapy. Although patients at early stages are cured by operation with or without adjuvant chemotherapy, more than ten thousands of patients die of this disease. Endocrine therapy for HRP tumors is effective, but cancer cells often become resistant to the treatment. Additionally, only one third of HRP breast cancer patients with metastatic disease respond to radiotherapy[30]. Importantly, nearly half of patients with HRP tumor at an advanced stage suffer from recurrence[31-33]. Therefore, novel therapeutic drugs and/or strategies are needed. The aim of this study is to identify the genetic factors susceptible to HRP breast cancer in the Japanese population and facilitate the development of novel approaches to prevent and/or treat this type of cancer.

Chapter 3 Materials and Methods

Case and control subjects

BBJ project and the GWAS analysis were approved by the Human Genome Research Ethics Committee in the Institute of Medical Science, The University of Tokyo (IMSUT-IRB, 15-3-0627) and Ethical Committeein the Center for Genomic Medicine of RIKEN. All participating subjects provided written informed consent to the study. Characteristics of study subjects are shown in Table 1. Most of the breast cancer cases and all the controls in this study were registered in BBJ and stored at the Institute of Medical Science, The University of Tokyo. All cases were diagnosed to have a HRP breast cancer by immunohistochemistry with estrogen receptor and progesterone receptor-specific antibodies using breast tissue obtained in clinical examination (biopsy or cytology).

In the initial GWAS, 1086 subjects with HRP breast cancer had been selected as cases; 846 samples were obtained from BBJ and the remaining 240 samples were collected from collaborative hospitals; Tokushima Breast Care Clinic, Yamakawa Breast Clinic, Shikoku Cancer Center, and Itoh Surgery and Breast Clinic, Kansai Rosai Hospital, Sapporo Breast Surgical Clinic and Sapporo Medical University Hospital.Controls were 1816 female subjects including 231 healthy volunteers from the Midosuji Rotary Club, Osaka, Japan,and 1585 subjects registered in BBJ as subjects with different eight non-cancerous diseases (hepatitis B, keloid, drug eruption, pulmonary tuberculosis, peripheral artery disease, arrhythmias, stroke, myocardial infarction) because they will not affect hormonal disturbance. In the replication study, we analyzed a total of 1652 cases including 1547 from BBJ and 105 from the collaborative hospitals. A total of 2797 female subjects registered in BBJ for other diseases (rheumatoid arthritis, amyotrophic lateral sclerosis, and liver cirrhosis) were used as controls.

Table 1. Characterization of subjects used in HRP breast cancer

		GWAS	Replication
Case	Number of subjects	1,086	1,653
	Mean age(±SD)	66.7 (18.5)	60.7 (9.3)
	Mean age of menarche	12.4	12.2
	Mean age of menopause	48.3	47.9
	cases with DCIS*	52	207
	cases with Invasion	1,034	1,446
	Body weight index		
	Platform	Illumina HumanHap 610K	invader assay
	Source	BioBank Japan collaborative hospitals[2]	BioBank Japan collaborative hospitals[1]
Control	Number of subjects	1,816	2,797
	Mean age(±SD)	61.3 (12.6)	65.9 (13.2)
	platform	Illumina HumanHap 610K	Illumina HumanHap 610K
	Source	BioBank Japan	BioBank Japan
	Diseases in control[2]	MRC healthy volunteer Hepatitis B keloid Drug eruption Pulmonary tuberculosis Peripheral artery disease Arrhythmias Stroke Myocardial infarction	Rheumatoid arthritis Amyotrophic lateral sclerosis Liver cirrhosis

[1] Tokushima Breast Care Clinic, Yamakawa Breast Clinic, Shikoku Cancer Center, and Itoh Surgery and Breast Clinic, Kansai Rosai Hospital, Sapporo Breast Surgical Clinic and Sapporo Medical University Hospital.

[2] The control groups from BioBank Japan consisted of female individuals without cancer also without any disease related to breast cancer.

*DCIS: Ductal carcinoma in situ

2266 cases with HRP breast cancer and 497 female controls with fourdifferent diseases (hepatitis B, keloid, drug eruption or pulmonary tuberculosis) from BBJ as well as 231 healthy female volunteers from the Midosuji Rotary Club, Osaka, Japan were used for re-sequencing analysis.

Sample quality control

In order to detect population stratification in this GWAS, Eigenstrat analysis was performed, which is based on principal component analysis (PCA) that detect the ancestry differences between cases and controls along continuous axes of variation[34].

SNP genotyping

Illumina HumanHap 610 Genotyping Bead Chip (Illumina, San Diego, CA, USA) was used for the first stage genotyping. As a SNP quality control standard, p-value of Hardy-Weinberg equilibrium test should be $<1.0 \times 10^{-6}$ in controls, and call rate of ≥ 0.99 in both cases and controls was applied. SNPs with minor allele frequency (MAF) of <0.01 in both case and control samples were excluded further analysis. A total of 453,627 SNPs on autosomal chromosomes and 10,525 SNPs on X chromosome passed the quality control filters and were further analyzed. Illumina HumanHap 610 Bead Chip was also used for control samples in the replication study. All cluster plots were checked by visual inspection by trained personnel, and SNPs with ambiguous calls were excluded.

I used the multiplex PCR-based Invader assay (Third Wave Technologies, Madison, WI, USA) for the genotyping of HRP breast cancer cases in the replication study. I designed all the primers within the Tm ranged from 55^{o}C to 60^{o}C as shown in Table2. The multiplex-PCR amplification was carried out using 15 ng of genomic DNA. Each total reaction volume is 50μl, consisting of PCR buffer, dNTP, 50 pmol of each primer and 10 unit of HotStart.

Table 2. Primer sequences and probes used in invader assay for the SNPs identified by GWAS.

SNP	Sequence ID	Sequence
rs708766	rs708766_F	GGGTGGGTTAAAGCCTTCTC
	rs708766_R	GCCCAGAGTCCAGTGGTATC
	rs708766_INV	CCCTGGAACCCACATCCACCCTGAT
	rs708766_VIC	CGCGCCGAGGATATTGCTGGAGGTACTC
	rs708766_FAM	ATGACGTGGCAGACCTATTGCTGGAGGTACT
rs53915	rs53915_F	TACGTTTACAGGCCTCTGCT
	rs53915_R	CTCATTACAGCCATCGAACC
	rs53915_INV	CCTTTAGAGAGGTTACAAAGGGAGGGAGACTGGTA
	rs53915_VIC	CGCGCCGAGGTTCAAATCACTGGAATCCAGAG
	rs53915_FAM	ATGACGTGGCAGACCTCAAATCACTGGAATCCAGA
rs1012357	rs1012357_F	TTTTACATTTGGCTCAGTGCTC
	rs1012357_R	TGAAAAATCATCTGCTTTTGAGTC
	rs1012357_INV	GAAATAGTAAAAAAACAAAAAACTA
	rs1012357_VIC	CGCGCCGAGGGAAGTATCATATCTCCA
	rs1012357_FAM	ATGACGTGGCAGACAAAGTATCATATCTCCAT
rs560737	rs560737_F	GGGCGCAATAACTATTTCTGA
	rs560737_R	CTAAATTCCCTGGGTCCTTTG
	rs560737_INV	CCTTTAGAGAGGTTACAAAGGGAGGGAGACTGGTA
	rs560737_VIC	CGCGCCGAGGTTCAAATCACTGGAATCCAGAG
	rs560737_FAM	ATGACGTGGCAGACCTCAAATCACTGGAATCCAGA
rs6788895	rs6788895_F	TGGTCCAAGGTAAACTCCAC

	rs6788895_R	TCATTCAGTGGTCTTTGTGG
	rs6788895_INV	TTGGGAGGTGAACCTGGAAGTGGGAAGAAGAGATATAA
	rs6788895_VIC	CGCGCCGAGGTGAGCTCAACTGGAAGG
	rs6788895_FAM	ATGACGTGGCAGACGGAGCTCAACTGGAAG
rs9835984	rs9835984_F	AAAGTGGGTTCATCCATTGC
	rs9835984_R	ACACCTGGCCCTGAGTTCTA
	rs9835984_INV	CTCACTGTACTTACTGCTCAATTTTTCTGTAAACCTAAAGTGCTCTAAAAT
	rs9835984_VIC	CGCGCCGAGGATAAAGACTATTAACAAACAGCAAC
	rs9835984_FAM	ATGACGTGGCAGACGTAAAGACTATTAACAAACAGCAA
rs10017875	rs10017875_F	TTGTCTTCCTGAATATCTCCTCCT
	rs10017875_R	CCAGAATGTTGGCTATGTATCAC
	rs10017875_INV	CCTCCTTCCACACACACTTCTTTCATCG
	rs10017875_VIC	CGCGCCGAGGACTGATGATTTTTCTTTCA
	rs10017875_FAM	ATGACGTGGCAGACGCTGATGATTTTTCTTTC
rs6844819	rs6844819_F	CATGTCTTGATGTTCGTTTGG
	rs6844819_R	TCTCTCAACAGTCGTCATCTTCA
	rs6844819_INV	CAATTTGTTACTTTTGTCACTATTTCAG
	rs6844819_VIC	CGCGCCGAGGATGTTTACAGGACTTCGT
	rs6844819_FAM	ATGACGTGGCAGACGTGTTTACAGGACTTCG
rs4921443	rs4921443_F	AACCTGCTTGAAATTCCCCTA
	rs4921443_R	AGGTCTCCAATGTGCATACTCTT
	rs4921443_INV	GTCTTTGCAATAATTCAAAGGCAGTATA
	rs4921443_VIC	CGCGCCGAGGGCTTTAGTGGTTAAGAGT
	rs4921443_FAM	ATGACGTGGCAGACACTTTAGTGGTTAAGAGTG

rs10039851	rs10039851_F	GGGAATACTTCTGAACTCAC
	rs10039851_R	TTTATAAGTTTTTACAGATT
	rs10039851_INV	GACAAAGGCATTATAAGAAAATAAAAT
	rs10039851_VIC	CGCGCCGAGGCGGCAGACCAATATCTCT
	rs10039851_FAM	ATGACGTGGCAGACTGGCAGACCAATATCTCTC
rs2278386	rs2278386_F	CCAACCTCTCATGTGCCAAT
	rs2278386_R	TCCCTATTGTACCAGGCATTC
	rs2278386_INV	GTTGTTCTGGTTTAATGATAATCAAATGAGGTCCACAGACCCT
	rs2278386_VIC	CGCGCCGAGGATGCCAGTTAGCAAACTG
	rs2278386_FAM	ATGACGTGGCAGACGTGCCAGTTAGCAAACT
rs9456909	rs9456909_F	ATTTGACATTGGTGTCCTTGG
	rs9456909_R	CTCAAATGCCTGCACACAATA
	rs9456909_INV	AACAATATGTCCCAATTAATTCTG
	rs9456909_VIC	CGCGCCGAGGATATTAATTGGGGGGGG
	rs9456909_FAM	ATGACGTGGCAGACGTATTAATTGGGGGGG
rs13199020	rs13199020_F	AGAAGGAAAAAAGTCACCTC
	rs13199020_R	GGCATGAGGATTGCTTGAAC
	rs13199020_INV	ATAGTGGACTTGCCCTTAAAACCAGT
	rs13199020_VIC	CGCGCCGAGGCATAAGCTGGAAGAGAATA
	rs13199020_FAM	ATGACGTGGCAGACTATAAGCTGGAAGAGAAT
rs6453815	rs6453815_F	TCAAAGCATAGCACAGTATTTGG
	rs6453815_R	ATGTGGAATGAGGTCCAAGA
	rs6453815_INV	CAAGCTTCTTTCTCAAGCCCTATTTTGCCAACATAGTGTTTTGCT

	rs6453815_VIC	CGCGCCGAGGATATAGTAGGTGTCCAACTG
	rs6453815_FAM	ATGACGTGGCAGACGTATAGTAGGTGTCCAACT
rs10456973	rs10456973_F	AGGCACCGTCCTCCAGTATT
	rs10456973_R	TCAGAGTGAGGCAGATGGAA
	rs10456973_INV	CTGGCTGGGGGTCCTCAGTATTTCCA
	rs10456973_VIC	CGCGCCGAGGTGGAACACCTAACCCAC
	rs10456973_FAM	ATGACGTGGCAGACCGGAACACCTAACCCA
rs10250216	rs10250216_F	CCAAAAATGGGGAATCACAG
	rs10250216_R	GAGAGCACCTACCCCGTTCT
	rs10250216_inv	GGAAAGGAGAGATTTGCCTACAAATTAGGCTGGAAACCT
	rs10250216_VIC	CGCGCCGAGGATAGGGAAGAAATTGTTTGCG
	rs10250216_FAM	ATGACGTGGCAGACGTAGGGAAGAAATTGTTTGC
rs7814396	rs7814396_F	CCATACCCTCCATTCTGTGTG
	rs7814396_R	AGCCTTGTGGGGAATTAGAGA
	rs7814396_inv	GAATCAGGCCTTATCATATCTCCCTGTA
	rs7814396_VIC	CGCGCCGAGGAGTGTAGATACATTTACAT
	rs7814396_FAM	ATGACGTGGCAGACACTGTAGATACATTTACA
rs822304	rs822304_F	TTCCAAACAAGCAAGAACTCAA
	rs822304_R	TCATCTGCCTTCTTTCTTTCTCA
	rs822304_inv	TATGGAAAGGAAGAGGCTATA
	rs822304_VIC	CGCGCCGAGGGAGGCTACTCTTCAATCAT
	rs822304_FAM	ATGACGTGGCAGACGAGGCTACTCTTCAATCA
rs4741635	rs4741635_F	CCAAAACACATTCAATGGAAGA

	rs4741635_R	TGCCAAGCTCCTTAGCATAAC
	rs4741635_inv	TTCTGTTTTCCTTTTTGAAGCACTCTGTGTG
	rs4741635_VIC	CGCGCCGAGGATTGTGTGGGAGAGTTATGCTA
	rs4741635_FAM	ATGACGTGGCAGACGTTGTGTGGGAGAGTTATGCT
rs6602097	rs6602097_F	GCCTAAGGCTTCGGAGTTAAA
	rs6602097_R	CCTTAGCCAAGGGGAGATGT
	rs6602097_inv	TTCTTCATCTGTAAGAAAGGTGACTAAC
	rs6602097_VIC	CGCGCCGAGGTACCTCATTAGAGTGTTG
	rs6602097_FAM	ATGACGTGGCAGACCACCTCATTAGAGTGTT
rs1028722	rs1028722_F	AAAAGCCTAACCGAGTGCAG
	rs1028722_R	CAAACATATGTGGGCCTGCT
	rs1028722_INV	CGTATGGTCACTGCTAGAAATGTTTGCCACCCA
	rs1028722_VIC	CGCGCCGAGGTGGTCTAACTACACGAGTAG
	rs1028722_FAM	ATGACGTGGCAGACCGGTCTAACTACACGAGTA
rs12591707	rs12591707_F	CCAGACCGAAGACAAGAAGTTG
	rs12591707_R	CCCAAATTTATTTATGCCAGTCA
	rs12591707_INV	TTTAATTTTCAGAAACAACTGTATTTAC
	rs12591707_VIC	CGCGCCGAGGTTTTCTTATTTGTATTTC
	rs12591707_FAM	ATGACGTGGCAGACCTTTCTTATTTGTATTT
rs11076442	rs11076442_F	AAGTAATGACCGCCAACAAAA
	rs11076442_R	TTCTGCAAAGGGTAGTCTCCA
	rs11076442_INV	AACTGATTTTAGGTATAATGTGGGTCC
	rs11076442_VIC	CGCGCCGAGGTGCTGCTTGTACTAATCGGT
	rs11076442_FAM	ATGACGTGGCAGACCGCTGCTTGTACTAATCGG

rs4793427	rs4793427_F	AGATTCTCCCACACAAATGGA
	rs4793427_R	AGTATCATCGGGGGATAGAGC
	rs4793427_INV	AAGCAGAAGAGGCCACTTGACATAC
	rs4793427_VIC	CGCGCCGAGGTGTTTACTAAACACCCGCTCT
	rs4793427_FAM	ATGACGTGGCAGACCGTTTACTAAACACCCGCTC
rs7219201	rs7219201_F	GTGCTGGGATTACAGGCATAA
	rs7219201_R	AAAACTGTGTCGCAAAGGAGA
	rs7219201_INV	TCTGGATGCTGCCCCTCATCAGATAC
	rs7219201_VIC	CGCGCCGAGGTACAGTTTGCAGATATTT
	rs7219201_FAM	ATGACGTGGCAGACCACAGTTTGCAGATATT
rs871739	rs871739_F	AGCAGTGAGGTTAGGGAGTCA
	rs871739_R	AGACATCGAGGACAGATGGAG
	rs871739_INV	TCTCATCCGCCCAGCAGGGAGTAGGCAC
	rs871739_VIC	CGCGCCGAGGTGGGATCGCGGCCTGGGCA
	rs871739_FAM	ATGACGTGGCAGACCGGGATCGCGGCCTGGGC
rs6040836	rs6040836_F	TTTACCCATCCCTCAGCTACC
	rs6040836_R	TACCTAATGTCAGGGGCTGAA
	rs6040836_INV	GGGCAGATAGGGAGGGAGGTAAGTC
	rs6040836_VIC	CGCGCCGAGGAGGTAAGCAGGGTATTTAT
	rs6040836_FAM	ATGACGTGGCAGACCGGTAAGCAGGGTATTTA
rs6036227	rs6036227_F	AAGTCAGTAACCAGCCCCTTG
	rs6036227_R	CTCCAGAAGCCAGCAGAATTA
	rs6036227_INV	AAGGCTGGTGATGCTTTACACCTT

	rs6036227_VIC	CGCGCCGAGGCGGGGTCCTTCAGAGAAAT
	rs6036227_FAM	ATGACGTGGCAGACTGGGGTCCTTCAGAGAAA
rs2853218	rs2853218_F	GGACTGCTGGCAGAATCTAA
	rs2853218_R	CCATTAGCCCTTCCATTTCT
	rs2853218_inv	GTACAACAGAGCTTTTGCCAGCCAGTGTCCTGAAAAATA
	rs2853218_VIC	CGCGCCGAGGTAGAGGGGCTGGCAAATTTA
	rs2853218_FAM	ATGACGTGGCAGACCAGAGGGGCTGGCA
rs8129581	rs8129581_F	ATCAAATGTTCCCTTCACCA
	rs8129581_R	CAAGGTGCTATGTGGCTTCT
	rs8129581_inv	CGTGAAAACGGAGACATCATTTAAAATCATCACTGATAATTAAAATGGGAACAGAT
	rs8129581_VIC	CGCGCCGAGGATTTAGAGATTTGGAGAATTCTC
	rs8129581_FAM	ATGACGTGGCAGACGTTTAGAGATTTGGAGAATTCT
rs17820958	rs17820958_F	AGCAAGTTGTCCTGTGCTATGA
	rs17820958_R	TGGTTCTGCTTTGTTTTAGGG
	rs17820958_inv	TGAGGAGACAAGGCTAATGGGAAAG
	rs17820958_VIC	CGCGCCGAGGAAGGCTGTTAGCCCTGCCC
	rs17820958_FAM	ATGACGTGGCAGACGAGGCTGTTAGCCCTGCC
rs5997921	rs5997921_F	TCTCTCCACATCTCCACAACC
	rs5997921_R	ATTTCTCTGAGCCATCCAGGT
	rs5997921_inv	TCTGCCTAGAATGCCCCCTTACTCTG
	rs5997921_VIC	CGCGCCGAGGTTTCACTTAACTCCTGCT
	rs5997921_FAM	ATGACGTGGCAGACGTTCACTTAACTCCTGC
rs1034587	rs1034587_F	ACCCAGCCTAATCCCTGTTT

	rs1034587_R	ACCCTTGACTGCTCTTCCAG
	rs1034587_inv	GTAAAGGAGGTTGCTGAGGAAAGCAAGGCGT
	rs1034587_VIC	CGCGCCGAGGAGGAAGTTAAACTCAAAAGCC
	rs1034587_FAM	ATGACGTGGCAGACCGGAAGTTAAACTCAAAAGC

F: Forward primer.
R: Reverse primer.
INV: Invader.
VIC-sequence: CGCGCCGAGG
FAM-sequence: ATGACGTGGCAGAC

DNA polymerase Promega, Madison, WI. The multiplex PCR condition was for initial denaturation 94^{o}C for 2 min, followed by 35 cycles of denaturation at 94°C for 15 s, annealing at 60°C for 45s, and extension at 72°C for 2 min. The PCR products were diluted in 1:10 and transferred into plates (384 well). The products were air-dried overnight before Invader assay. Invader assay combines a universal fluorescent resonance energy transfer (FRET) system and structure-specific cleavage enzymes. FRET probes were labeled with FAM or VIC that corresponded to alleles. Signal intensity was calculated as the ratio of FAM or VIC to ROX, an internal reference. The dried PCR products were resolved in the reaction mixture of 3µl, containing 0.5µl of distilled water, 0.5µl of signal buffer, 0.5µl of structure-specific cleavage enzyme, 1µl of allele-specific probe mix and 0.5µl of FRET probes. The reaction was carried out by the incubation of samples at 95°C for 5 minfollowed by 5min at 63°C, and the genotype was analyzed by Applied Biosystems 7900 HT Fast Real-time PCR system. Additional incubation was done when the allele discrimination plot did not reveal distinct cluster of risk homozygote, heterozygote and non-risk homozygote.

For the genotyping of additional 22 variations discovered by re-sequencing analysis, the multiplex PCR-based Invader assay was also applied (Table 3)

Statistical analysis

I tested the associations of SNPs by employing the Cochran–Armitage trend test for both the initial GWAS and replication analyses. For the combined study, the simple combined method was applied. In the replication analyses, significance level was applied to be P-value of $<1.35 \times 10^{-3}$ (calculated as 0.05/37). Odds ratios (ORs) and confidence intervals were calculated using the non-susceptible allele as a reference. Heterogeneity between the initial GWAS and replication subjects was examined using the Breslow– Day test. The genomic inflation factor (lGC) was calculated from the median of the Cochran–Armitage trend test.

Table 3. Primer sequences and probes of the SNPsdiscovered by re-sequencing analysis

SNP	Sequence ID	Sequence
rs6788895	rs6788895_F	TGGTCCAAGGTAAACTCCAC
	rs6788895_R	TCATTCAGTGGTCTTTGTGG
	rs6788895_INV	TTGGGAGGTGAACCTGGAAGTGGGAAGAAGAGATATAA
	rs6788895_VIC	CGCGCCGAGGTGAGCTCAACTGGAAGG
	rs6788895_FAM	ATGACGTGGCAGACGGAGCTCAACTGGAAG
rs476663	rs476663_F	GGAGGATGCTAGAAAACTACCTCA
	rs476663_R	TTCACTTGTGAGGCATTACAGAA
	rs476663_INV	GTAAAATGTACCTTACGGTAAACC
	rs476663_VIC	CGCGCCGAGAAATAGTTGAAAAGAGGA
	rs476663_FAM	ATGACGTGGCAGACCAATAGTTGAAAAGAGG
rs7615292	rs7615292_F	CCACTGCTGTTTTTGCTTCTC
	rs7615292_R	GAGGCTGTCATCCTTGCTTC
	rs7615292_INV	TATCACTAGCAACAAGCAGATG
	rs7615292_VIC	CGCGCCGAGGAGTATGCACTATGTGACT
	rs7615292_FAM	ATGACGTGGCAGACGGTATGCACTATGTGAC
rs7636866	rs7636866_F	CAAAAGACCACACCAAGAGGA
	rs7636866_R	GGGGTCTAAGGCTGACACCTA
	rs7636866_INV	TGTTAATTCCTAGATACCAGATGGAGA
	rs7636866_VIC	CGCGCCGAGGTTCAGCCCACAAGTAGACACC
	rs7636866_FAM	ATGACGTGGCAGACATCAGCCCACAAGTAGACAC
rs2116684	rs2116684_F	CTGCATATACCCTTGGGCTCT

	rs2116684_R	AATGGCCACATATTGCAGTTC
	rs2116684_INV	AAAAAATCTCAGTAAGAAAAG
	rs2116684_VIC	CGCGCCGAGGCAGAATTTAAATGACCAC
	rs2116684_FAM	ATGACGTGGCAGACGAGAATTTAAATGACCA
rs57359526	rs57359526_F	GCAAGCTGTACAGGAAGCATA
	rs57359526_R	TCCTCTTGGTGCTGTTCTCAT
	rs57359526_INV	AGAAAGCTTCCAATCATGGTGGAAGG
	rs57359526_VIC	CGCGCCGAGGTCAAAGGGGGAGTGAGGCAT
	rs57359526_FAM	ATGACGTGGCAGACGCAAAGGGGGAGTGAGGCA
rs6789306	rs6789306_F	ATTACCCCACTGCATCCTTTT
	rs6789306_R	GGGTCAGCAAGCTTTGAGAG
	rs6789306_INV	CTGGTCAGTCTGGTTCTGGCCAGTA
	rs6789306_VIC	CGCGCCGAGGGGAGGTTAACATGAGA
	rs6789306_FAM	ATGACGTGGCAGACAGAGGTTAACATGAG
rs1148369	rs1148369_F	GAATCAGATCAGAGGCTGCAT
	rs1148369_R	CCAAAGTGCTGGGATTACAAG
	rs1148369_INV	CGCTTTGTCTGCAAGTGGTGGT
	rs1148369_VIC	CGCGCCGAGGCTATACATGTGAAGTCA
	rs1148369_FAM	ATGACGTGGCAGACTTATACATGTGAAGTC
rs76182650	rs76182650_F	CTCTACCACAGAACGGGAGAG
	rs76182650_R	TCCACACATAACTGGCCAAAG
	rs76182650_INV	CTTTCACTGATAACTATTTACTCT
	rs76182650_VIC	CGCGCCGAGGCAGTTCCAGGTCATGGT
	rs76182650_FAM	ATGACGTGGCAGACTAGTTCCAGGTCATGG

rs1833811	rs1833811_F	TGGAGAACCAGTTTTGTGAGG
	rs1833811_R	TGAGGGGAATAGAGCTGTTTG
	rs1833811_INV	ATGCCAGGTTCCCACCTGCACAACCACA
	rs1833811_VIC	CGCGCCGAGGGTAACTGCCCTGACAGTTC
	rs1833811_FAM	ATGACGTGGCAGACATAACTGCCCTGACAGTT
rs61628582	rs61628582_F	AGGAAGCATAGTGGCTTCTGC
	rs61628582_R	TCCTCTTGGTGCTGTTCTCAT
	rs61628582_INV	GCAAGGGTGGGAGAGGTGCTTCACACTTC
	rs61628582_VIC	CGCGCCGAGGTTTAAATAACCAGATCTCATA
	rs61628582_FAM	ATGACGTGGCAGACCTTAAATAACCAGATCTCAT

F: Forward primer.
R: Reverse primer.
INV: Invader.
VIC-sequence: CGCGCCGAGG
FAM-sequence: ATGACGTGGCAGAC

The quantile–quantile plot of the logarithms of the genome-wide P-values was generated by the "snpMatrix" package in R program v2.10.0, and the Manhattan plot was generated using Haploview v4.1. Haplotype analysis was performed by the use of Haploview v4.1 by considering genotyped SNPs located within 500 kb upstream or downstream of the marker SNP. (Haploview software was used to analyze linkage disequilibrium (LD) values, visualize haplotype.)

Genotype imputation

The SNPs with low genotyping rate (<99%), showing deviations from Hardy-Weinberg equilibrium (>1.0 x 10^{-6}), or MAF (<0.01) were excluded from the analysis. MACH version 1.0 was used to estimate haplotypes, map crossover and error rates using 50 iterations of the Markov chain Monte Carlo algorithm. Imputation analysis was carried out by referring to the genotype data of Japanese (JPT) individuals as deposited in the Phase II HapMap database using MACH v1.0 Genotypes of SNPs that are located in the genomic region within 500 kb upstream or downstream of the marker SNP (the SNP that showed the strongest association with HRP breast cancer) were imputed. The imputed SNPs with an imputation quality score of $r^2 < 0.3$ were excluded from the subsequent analysis.

Re-sequencing analysis

I explored additional SNPs in the genomic region of the *SIAH2* (seven in absentia homolog 2) gene using DNA samples of 96 cases with HRP breast cancer. A total of 98 sets of primers were utilized (Table 4) to amplify the 22,353 bps (two exons, one intron, 5 UTR and 3UTR) of the genomic region corresponding to the *SIAH2* gene. The sequence information was obtained from UCSC Genome Bioinformatics database (NM_005067). PCRs were performed using GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, USA). The PCR

Table 4. Primer sequences for re-sequencing of *SIAH2*

	PrimerSequences	position	product size
F1	CCTGAAGTTGCCTTTCTCGT	Exon 1	362
R1	CACCCGCTTCTGGAACAG		
F2	CAGCGGCTGTTCCAGAAG	Exon 1	366
R2	CTGCTTGCTGCAGGGTTTAT		
F3	AATAAACCCTGCAGCAAGCA	Exon 1	223
R3	AGTCAAAGCAGACCGGACAC		
F4	TCAGGAACCTGGCTATGGAG	Intron	420
R4	CTCCTCAATCCCCACACAGT		
F5	TGCAGACGTGTATTCGGGTA	Intron	520
R5	GGAGGGGAGAGATCTTGGTC		
F6	AAAATGCAAGCCACCAAGAC	Intron	646
R6	GGGCAGCAGACCAAATTAAA		
F7	GGCCATTTAAGCTGTCTTGC	Intron	439
R7	AGAAATCCTATGGGGCATCC		
F8	TGGTTGGAAAAGGTTTGGTG	Intron	417
R8	AAGGCTGAGAACCCCTGATT		
F9	CCCAAGAGAAAGGGAAGGAA	Intron	403
R9	AGCTGCCCTGCCAATAGTTT		
F10	GGCAGCTCAGCCAGTTAGTC	Intron	476
R10	AGCTGACCCCTGCAAAGTAA		
F11	TGCAGGGGTCAGCTTTTACT	Intron	425
R11	GCAGTCCAGAGCTGTGATGA		
F12	AAACCGTCTGGTTGGATGAG	Intron	479
R12	TTACAGACGTGAGCCACTGC		
F13	GGGAACCGGATGCATTATTA	Intron	489
R13	CCACCACACCTGGCTATTTT		
F14	AAAGTTAGCTGGGCATGGTG	Intron	443
R14	GGGCACAGGATGTCAAAAGT		
F15	GCAGTGAGCTGAGATTGTGC	Intron	434
R15	TCCAGCTGAGGCTTTCAAAT		
F16	GTGCCCAGACTTAAACTCACC	Intron	556
R16	GAGTAAAGGCTGCCTCTTGG		
F17	TCACCTCTGGCTGATTGGT	Intron	528
R17	GTTCCTGATGGCCAGAAAAT		
F18	TCCCAAGGTTCTTTCTTACCC	Intron	497
R18	AGCCCCGGAGTTAATGTATG		
F19	ATTAACTCCGGGGCTCAAG	Intron	536
R19	AACTCGCCAGCAATACACAG		
F20	TGGCCACATATTGCAGTTCT	Intron	381
R20	ATTAGCCAAGCCTGATGGTG		
F21	CTACTTGGGAGGCTGAGAGG	Intron	545
R21	CAACCTTGAGGCTTGGAGAG		
F22	ATGCAAAGTGCCTGACACAG	Intron	547

R22	GGGACATAGACCCAAACCAT		
F23	TGAGATGGTGATATGGTTTGG	Intron	552
R23	CTGCAATATGTGGCCATTAAA		
F24	TGGCCACATATTGCAGTTCT	Intron	593
R24	TGGGAAGAAATGGACATTCA		
F25	CAGTTGAATGTCCATTTCTTCC	Intron	579
R25	AGAGGTTGTAGTGAAAAAGACACTG		
F26	AACCTCTTGCTTTTTCTGTGC	Intron	403
R26	AGCCTGAAATGCCAGCTAAT		
F27	GGGGCCGGAGAAGAATAA	Intron	462
R27	AGCTCTATTCCCCTCAGCAG		
F28	ACTGCTGAGGGGAATAGAGC	Intron	628
R28	CCAAAAATCTGTCCTTTTTGC		
F29	GCAAAAAGGACAGATTTTTGG	Intron	545
R29	GGTAAATTCCTGAACTCTGCTG		
F30	AGCAGAGTTCAGGAATTTACCA	Intron	526
R30	GAAATGACGGGGTCTCACTT		
F31	AAGTGAGACCCCGTCATTTC	Intron	522
R31	AATGGGAGTTGGCAGCTAAG		
F32	AATTCATCCTTCCAGTTGAGC	Intron	599
R32	GTGAGCCGTGATTGTACCAC		
F33	GTGGTACAATCACGGCTCAC	Intron	533
R33	TCTGCTGTGTGTCAAATTTCC		
F34	GGAAATTTGACACACAGCAGA	Intron	547
R34	AGTGAAGGTCATGGATGCTG		
F35	CAGCATCCATGACCTTCACT	Intron	357
R35	ACCTGTAATCCCAGCTACCC		
F36	GATTACAGGTGCCCACCAC	Intron	618
R36	AAAAACAGCCCTGAAAGTGC		
F37	GGCTGTTTTTCTCTCCTTGC	Intron	611
R37	CAGAAATGTCTCCAGCTCTCC		
F38	CCTGGAGAGCTGGAGACATT	Intron	570
R38	GGAGGAAACAGTGGGGATAA		
F39	GTTTCCTCCACTCACCCTGT	Intron	612
R39	CCTTGGGCTTGTGTTTTCTT		
F40	ACTCCAGCTTGGGTGACAG	Intron	454
R40	AGCCTTAGGTGCTTACTCTGCT		
F41	GGATGTAGCAAGATCCTTATGAAA	Intron	482
R41	TGTGAGTGATGGGTGAGACA		
F42	TGTCTCACCCATCACTCACA	Intron	479
R42	TCACTAGCAACAAGCAGATGG		
F43	TGCATACCATCTGCTTGTTG	Intron	450
R43	CCACACCAAGAGGAACTGAA		
F44	ACTGTCATCCTTGTGAGAGTCTG	Intron	478
R44	CCTGGAGACCCCTTCATACA		
F45	GGGTTAGGCCCTTCTCATGT	Exon 2	522

R45	CGCTGTCAAAGGGATCAGA		
F46	CCAGCTTTTATAGGGCTTGC	Exon 2	568
R46	CATCATCACCCAGTCGACAG		
F47	CTGTCGACTGGGTGATGATG	Exon 2	546
R47	GAACAGCATCCAAATCACCA		
F48	TTGAACAGCAGTGTCCCATC	Exon 2	663
R48	GGCCATATGACTTGCTTTCC		
F49	GTGGGAAAGGGTCCAGATTT	Exon 2	502
R49	GGTTCATTGCACTTGCACAC		

F=Forward primer, R=Reverse primer

Products were sequenced with forward or reverse primer and BigDye® Terminators by ABI PRISM 3730 Genetic Analyzer (Applied Biosystems). SNPs were detected by Sequencher software v4.8 (Gene Codes, Ann Arbor, MI, USA).

Cell lines

Three breast cancer cell lines,MCF7, T-47D, and BT-474, were obtained from the American Type Culture Collection (ATCC, Manassas, VA). T-47D and BT-474 cells were grown in RPMI 1640 medium, and MCF7 cells were in Eagle's minimal essential medium (EMEM) supplemented with 10% fetal bovine serum (FBS, Life Technologies, Carlsbad, CA) and 1% antibiotic/antimycotic solution (Sigma, St. Louis, MO). were maintained at 37^{o}C in an atmosphere of humidified air with 5% CO2.

Preparation of reporter plasmids

The genomic regions encompassing the three SNPs, SNP X, SNP Y and SNP Z, were amplified by PCR using different sets of primers (Table 5) with genomic DNA from the subjects as templates. The reaction was performed in a reaction mixture of 50μl containing KOD buffer, 5mM of dNTP, 50 pmol of each primer, 30 ng of genomic DNA, and KOD Taq polymerase (TOYOBO, Tokyo, Japan). All of the reaction involved initial denaturation at 94°C for 2 min, followed by 35 cycles at 94°C for 15 s, 58°C for 30s, and 72°C for 2-5 min on a GeneAmp PCR System 9700 (Applied Biosystems, USA). PCR products were purified with QIAquick PCR Purification Kit (Quiagen, Valencia, CA), digested with appropriate restriction enzymes, and cloned into the appropriate enzyme sites of pGL3-Promoter vector or pGL3-basic vector (Promega, Madison, WI).The reporter plasmids, pGL3A_Major and pGL3A_Minor, included a genomic region A (between -2,151 and -571) in the 5'-flanking region of *SIAH2*encompassing SNP X upstream of the luciferase gene. The reporter

Table 5. Primer sequencesfor reporter plasmids.

Plasmid' name	Forward primer	Reverse primer	Restriction enzyme_1	Restriction enzyme_2	Length of product (bp)
pGL3A	5'CTCGGTATCCTATCCAAAACTCAGGTTCAGC 3'	5'TCGAGATCTCGTCGAAGTCCTCTGG 3'	*Kpn* l	*Bgl*ll	1,580
pGL3PB	5'TGGTACCAGCAGGTCCTCTTA 3'	5'CTTGGGAAGACTTTGGCATAA3'	*Nhe* l	*Bgl*ll	1.089
pGL3PC	5'TGCCCCAGTAATGAATGTCCT 3'	5'TGTGGTCATTTGTGGCTTCA 3'	*Mlu* l	*Xho* l	1.905
pGL3PB290	5'AAGGGAGGGTGTGTTTAGCC 3'	5'ACCTGGCATGGAAGAGTGAG 3'	*Mlu* l	*Xho* l	290
pGL3PB60	5'TCATGGAGAACCAGTTTTGTG 3'	5'TTGCTTGAACATGCACTTTG 3'	*Mlu* l	*Xho* l	60

pGL3A: construct of region A using pGL3-Basic.
pGL3PB: construct of region B using pGL3-promoter.
pGL3PC: construct of region C using pGL3-promoter.
pGL3PB290: construct of region B 290bp using pGL3-promoter.
pGL3PB60: construct of region B 60pb using pGL3-promoter.

plasmids, pGL3PB_Major and pGL3PB_Minor, included a genomic region B (between +10,063 and +11,152) in the intron 1 of *SIAH2* encompassing SNP Y and plasmids, pGL3PC_Major and pGL3PC_Minor, included a genomic region C (between +13,956 and 15,864) in the intron 1 encompassing SNP Z upstream of SV40 promoter. Plasmids DNA was purified from bacterial cells using NucleoBond® Xtra plasmid purification kit according to the supplier's protocol.

Site directed mutagenesis

Site direct mutagenesis was performed with Quick Change II XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA) according to the manufacture's instruction. Primer sequences used for the mutagenesis were shown in Table 6. I changed A to G in the SNP Y1 (A/G) using forward primer 5`-CACCTGCACAACCACGTAACTGCCCTGACAG-3` and reverse primer 5`-CTGTCAGGGCAGTTACGTGGTTGTGCAGGTG-3` and G into T in the SNP using forward primer 5`-CACCTGCACAACCACTTAACTGCCCTGACAG-3` and reverse primer 5`- CTGTCAGGGCAGTTAAGTGGTTGTGCAGGTG -3`.

Table 6. Primer sequences for site directed mutagenesis

B209_major_mutant_F	CACCTGCACAACCACGTAACTGCCCTGACAG
B209_major_mutant_R	CTGTCAGGGCAGTTACGTGGTTGTGCAGGTG
B209_minor_mutant_F	CACCTGCACAACCACTTAACTGCCCTGACAG
B209_minor_mutant_R	CTGTCAGGGCAGTTAAGTGGTTGTGCAGGTG

B209_major_mutant_F: Forward primer for region B 290 bp has mutant major allele of Y1
B209_major_mutant_R: Reverse primer for region B 290 bp has mutant major allele of Y1
B209_minor_mutant_F: Forward primer for region B 290 bp has mutant minor allele of Y1
B209_minor_mutant_R: Reverse primer for region B 290 bp has mutant minor allele of Y1
Under line indicates mutated nucleotide of Y1 in the primers.

Dual-Luciferase reporter assay

BT-474 (2× 10^5cells /well) or T-47D (1.5× 10^5cells /well) cells seeded on six-well plates were transfected with 1μg of reporter plasmids and 0.1 μg of pRL-TK plasmids (Promega) by FuGENE®6 reagent (Roche, Indianapolis, USA). The cells were harvested after 24 h of incubation, and luciferase activities were measured using dual luciferase assay system (TOYO B-Net, Tokyo, Japan).

17 β estradiol treatment

MCF7 (2× 10^5cells /well) and T-47D (1.5× 10^5cells /well) cells were grownin phenol red free-medium supplemented with 5% FBS treated with dextran coated charcoal to remove the steroid hormone agonist and 1% streptomycin for 24 hrs before the transfection. The cells were co-transected with reporter plasmids and pRL-TK *Renilla* luciferase plasmid using FuGENE® 6 Transfection Reagent according to the supplier's protocol. Twenty-four hrs after transfection, the cells were treated with 17β-estradiol (Sigma) freshly resolved in ethanol at the concentration of 10 nM, or vehicle (ethanol). Reporter activities were measured after 24 hrs of the treatment.

Chapter 4 Results

Identification of SNPs associated with HRP breast cancer

To identify genetic variants susceptible to HRP breast cancer in Japanese population, I performed an initial case-control study using 1086 female patients and 1816 female controls by GWAS with Illumina HumanHap 610k Bead Chip. A SNP quality control standard was applied for the study, and SNPs showing p-value of Hardy-Weinberg equilibrium more than 1.0×10^{-6} in controls and SNPs with call rate lower than 0.99 in both cases and controls were excluded. After this quality check of SNP genotyping data, a total of 453,627 SNPs were selected for further analysis. Principal component analysis revealed that all the subjects participating in this study were clustered in the HapMap Asian population (Figure 1). The quantile-quantile plot (Q-Q plot), a plot of the percentiles (or quintiles) of a standard normal distribution against the corresponding percentiles of the observed data, showed a highly positive association between the expected and observed p-values. The observations follow approximately a normal distribution because the resulting plot is roughly a straight line with a positive slope as shown in Figure 2. The genomic inflation factor (lGC) of the test statistic in this study was 1.053 indicating a very low possibility of false-positive associations resulted from the population stratification.

Although no SNP achieved genome-wide significance level P-value $<5\times10^{-8}$, 46 SNPs in various chromosomes showed suggestive association (P-value $<1\times10^{-4}$) as illustrated in Manhattan plot in Figure 3. Among these 46 SNPs, I excluded 13 SNPs possessing strong LD ($r^2>0.8$) with the others, and selected 33 SNPs for replication study. I included four additional SNPs (rs1292011, rs3803662, rs2981579 and rs3750817) that were previously reported their association with breast cancer and showed P-value of $<1.0\times10^{-2}$ in our GWAS. A total of 37 SNPs (Table 7) were analyzed using an independent set of 1653 female patients and 2797 female controls. In this replication study, a SNP, rs6788895, was successfully replicated with

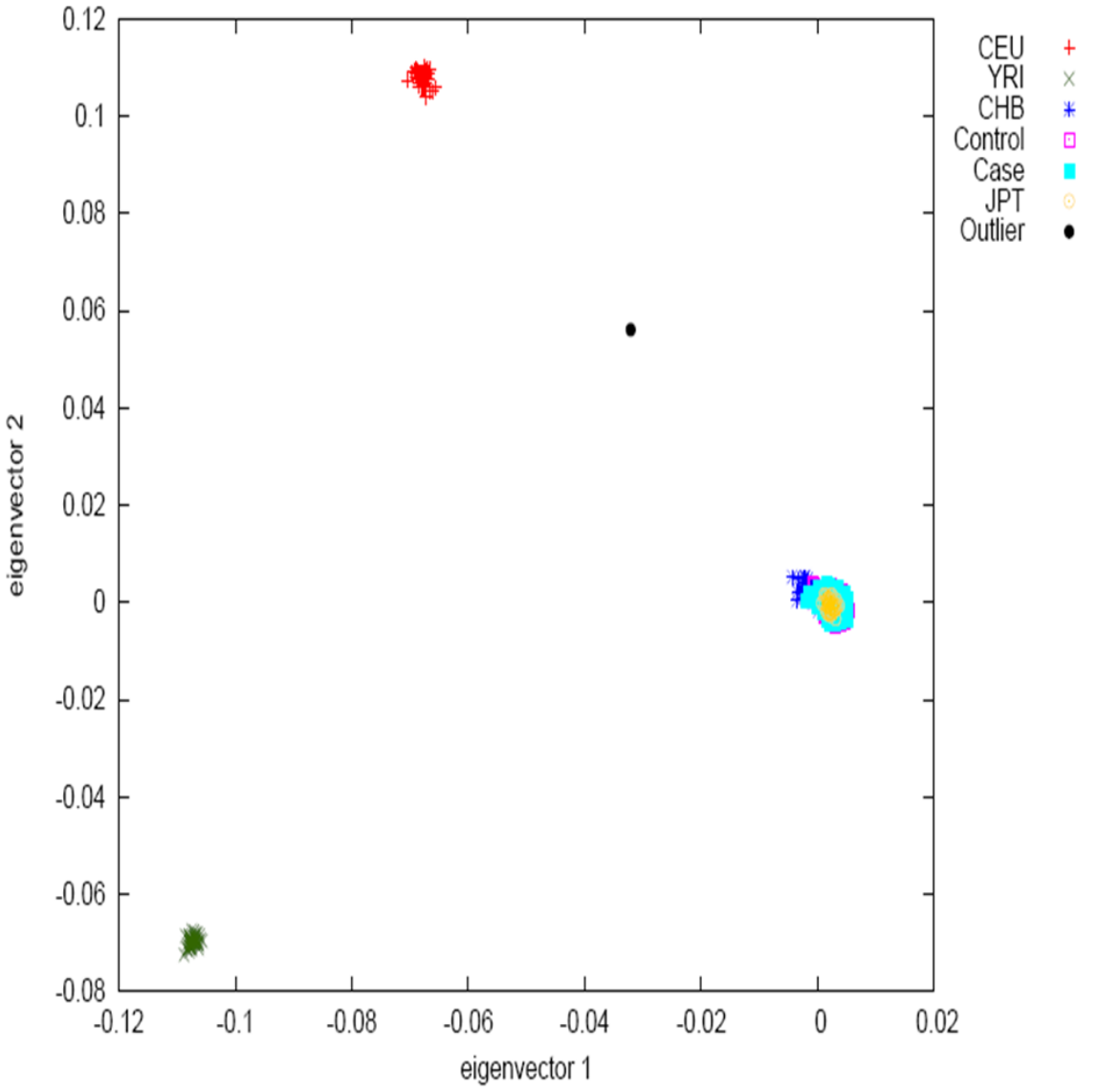


Figure1. Principal component analysis (PCA) plot of samples enrolled in the genome wide association study (GWAS) of hormonal receptor positive breast cancer. All cases and controls who participated in the GWAS were plotted based on eigenvectors1 and 2 obtained from the PCA, along with the European(CEU), African(YRI), and Japanese(JPT) and Chinese(CHB) individuals obtained from the PhaseII HapMap database.

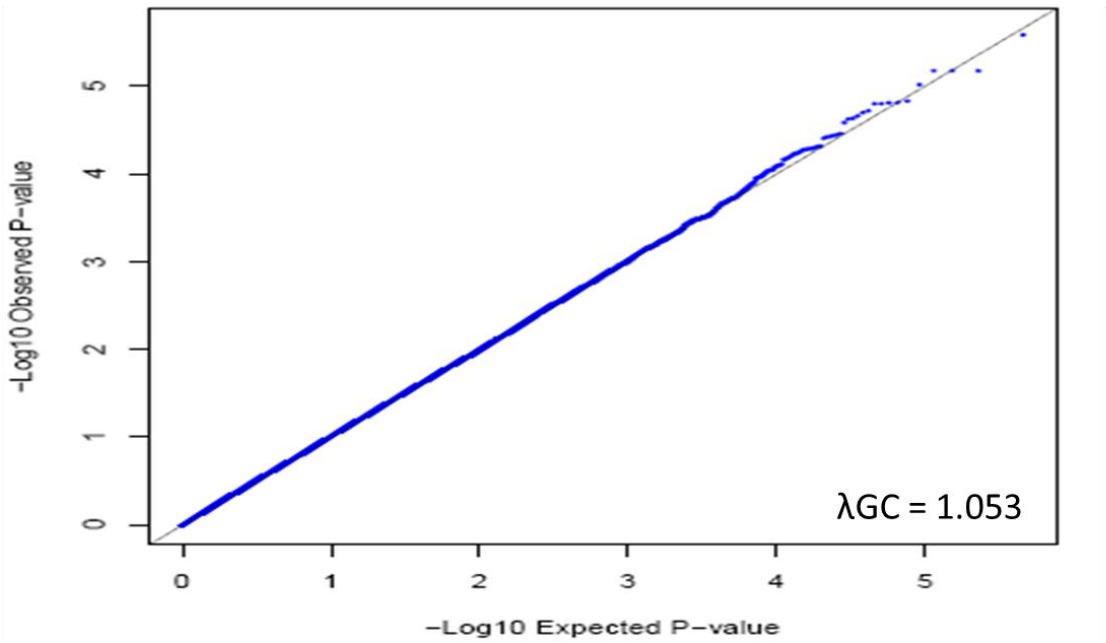


Figure2.A quantile-quantile (Q.Q) plot for GWAS of hormonal receptor positive breast cancer. Horizontal and vertical lines represent expected *P* values under a null distribution and observed P values, respectively. Fall the SNPs were not associated with the disease, all plots would lie on the line y=x.

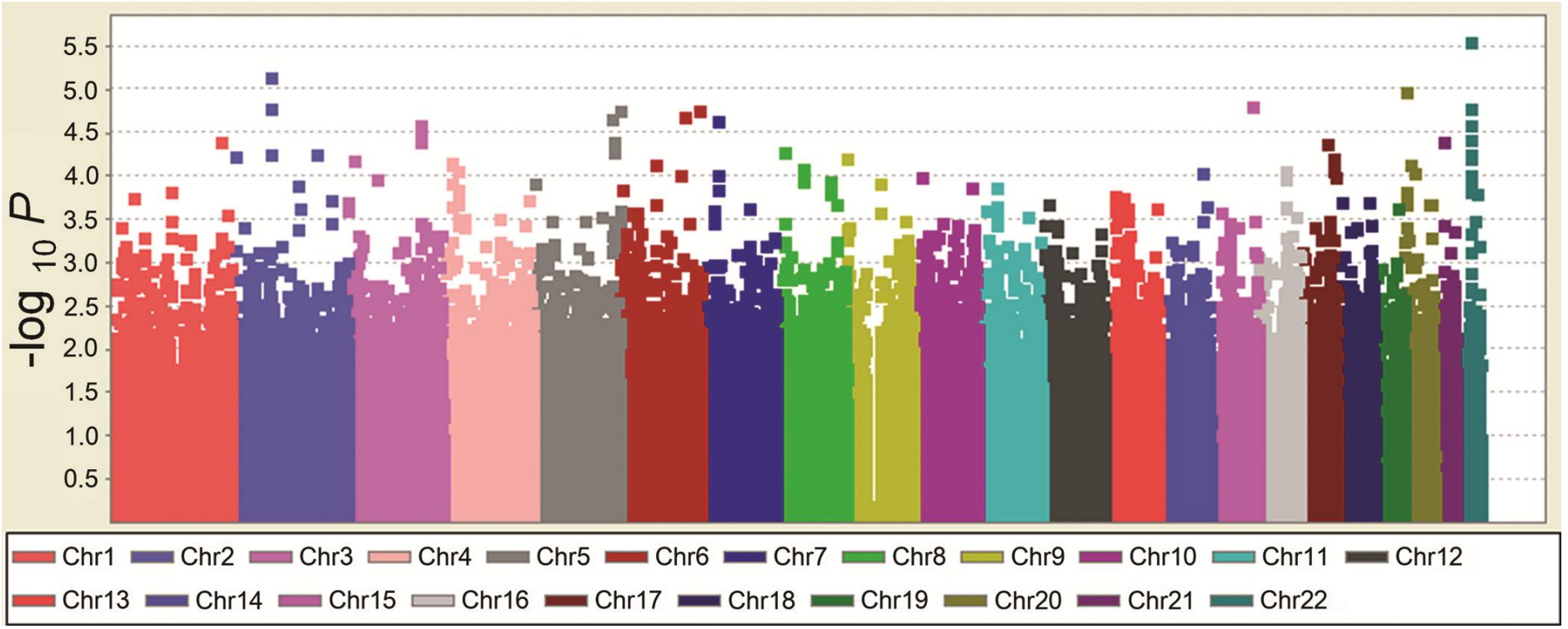


Figure 3. Manhattan plot for GWAS of HRP breast cancer indicating _log10P of the Cochran–Armitage trend test for 453627 single-nucleotide polymorphisms (SNPs) plotted against their respective positions on each chromosome.

Table 7. Associations of the 33 SNPs identifiedin GWAS with HRP breast cancer.

		case				control									
SNP	Stage	11	12	22	MAF	11	12	22	MAF	P_{assoc}[a]	OR		Phet[b]	chr	Gene
rs708766	GWAS	691	345	50	0.2	1273	498	45	0.16	3.65E-05	1.33	(1.16-1.53)	8.99E-03	1	*ITPKB*
	Rep	1114	463	74	0.19	1909	785	103	0.18	3.61E-01	1.05	(0.94-1.18)			
	Combined	1805	808	124	0.19	3182	1283	148	0.17	1.07E-03	1.16	(1.06-1.26)			
rs53915	GWAS	567	434	85	0.28	1081	650	85	0.23	6.74E-06	1.32	(1.17-1.49)	1.26E-05	2	*CTNNA2*
	Rep	967	589	94	0.24	1586	1029	182	0.25	1.53E-01	1.08	(0.97-1.19)			
	Combined	1534	1023	179	0.25	2667	1679	267	0.24	8.88E-02	1.07	(0.99-1.16)			
rs1012357	GWAS	648	388	50	0.22	966	710	140	0.27	5.19E-05	1.29	(1.14-1.46)	4.14E-05	2	*KIAA1715*
	Rep	890	640	122	0.27	1566	1045	186	0.25	1.40E-01	1.08	(0.98-1.19)			
	Combined	1538	1028	172	0.25	2532	1755	326	0.26	1.67E-01	1.06	(0.98-1.14)			
rs560737	GWAS	11	216	859	0.11	10	265	1541	0.08	5.65E-05	1.45	(1.21-1.73)	4.00E-02	2	*SOX11*
	Rep	43	262	1346	0.11	28	470	2299	0.09	9.29E-02	1.14	(0.98-1.31)			
	Combined	54	478	2205	0.11	38	735	3840	0.09	1.76E-04	1.24	(1.11-1.39)			
rs6788895	GWAS	106	456	524	0.31	242	832	742	0.36	2.34E-05	1.28	(1.14-1.43)	2.33E-01	3	*SIAH2*
	Rep	164	694	786	0.31	337	1265	1195	0.35	5.77E-04	1.18	(1.07-1.29)			
	Combined	270	1150	1310	0.31	579	2097	1937	0.35	9.43E-08	1.22	(1.13-1.31)			
rs9835984	GWAS	80	459	547	0.28	214	791	811	0.34	6.24E-05	1.27	(1.13-1.42)	6.83E-05	3	*NUP210*
	Rep	185	721	746	0.33	295	1179	1322	0.32	1.81E-01	1.07	(0.97-1.17)			
	Combined	233	1148	1259	0.31	509	1970	2133	0.32	2.36E-02	1.09	(1.01-1.17)			
rs10017875	GWAS	87	431	568	0.28	101	640	1074	0.23	8.08E-05	1.28	(1.13-1.44)	1.01E-03	4	*LOC727819*
	Rep	116	580	956	0.25	168	1057	1571	0.25	7.26E-01	1.02	(0.92-1.12)			
	Combined	203	1011	1524	0.26	269	1697	2645	0.24	2.70E-02	1.09	(1.01-1.18)			
rs6844819	GWAS	322	522	242	0.46	632	882	302	0.41	6.37E-05	1.25	(1.12-1.39)	1.80E-02	4	*LOC645174*
	Rep	514	782	356	0.45	892	1348	551	0.44	2.30E-01	1.06	(0.97-1.15)			
	Combined	836	1304	598	0.46	1524	2230	853	0.43	6.05E-04	1.13	(1.05-1.2)			
rs4921443	GWAS	10	157	919	0.08	3	187	1626	0.05	2.01E-05	1.58	(1.28-1.95)	2.78E-03	5	*GABRG2*
	Rep	7	201	1444	0.07	11	329	2457	0.06	6.64E-01	1.04	(0.87-1.24)			

	Combined	17	358	2363	0.07	14	516	4083	0.06	2.48E-03	1.23	(1.08-1.41)			
rs10039851	GWAS	625	396	65	0.24	1181	562	71	0.19	1.60E-05	1.33	(1.17-1.51)	1.38E-02	5	*GFPT2*
	Rep	1002	534	116	0.23	1713	946	138	0.22	1.51E-01	1.08	(0.97-1.2)			
	Combined	1627	930	181	0.24	2894	1508	209	0.21	1.55E-04	1.17	(1.08-1.27)			
rs2278386	GWAS	146	525	415	0.38	344	875	597	0.43	4.99E-05	1.25	(1.12-1.4)	6.44E-04	5	*SLIT3*
	Rep	283	818	550	0.42	484	1353	960	0.41	6.95E-01	1.02	(0.93-1.11)			
	Combined	429	1343	965	0.4	828	2228	1557	0.42	2.44E-02	1.08	(1.01-1.16)			
rs9456909	GWAS	16	260	809	0.13	65	516	1234	0.18	1.59E-05	1.39	(1.2-1.62)	6.50E-04	6	*LOC728275*
	Rep	54	461	1137	0.17	93	780	1923	0.17	9.49E-01	1.00	(0.9-1.12)			
	Combined	70	721	1946	0.18	158	1296	3157	0.17	6.59E-03	1.13	(1.04-1.24)			
rs13199020	GWAS	9	172	905	0.09	24	401	1391	0.12	1.92E-05	1.47	(1.23-1.76)	6.16E-03	6	*FAM8A6P*
	Rep	18	322	1312	0.11	38	570	2189	0.12	3.04E-01	1.07	(0.94-1.23)			
	Combined	27	494	2217	0.1	62	971	3580	0.12	5.21E-04	1.21	(1.09-1.35)			
rs6453815	GWAS	93	459	534	0.3	208	845	763	0.35	6.87E-05	1.26	(1.12-1.41)	8.09E-04	6	*COL12A1*
	Rep	190	691	770	0.32	289	1211	1297	0.32	6.61E-01	1.02	(0.93-1.12)			
	Combined	283	1150	1304	0.31	497	2056	2060	0.33	3.26E-02	1.08	(1.01-1.16)			
rs10456973	GWAS	697	357	32	0.19	1290	490	36	0.15	8.98E-05	1.31	(1.14-1.51)	9.34E-05	6	*PTPRK*
	Rep	1170	433	48	0.16	1915	799	83	0.17	1.36E-01	1.09	(0.97-1.23)			
	Combined	1867	790	80	0.17	3205	1289	119	0.17	2.06E-01	1.06	(0.97-1.16)			
rs10250216	GWAS	16	301	769	0.15	21	377	1418	0.16	2.21E-05	1.39	(1.19-1.62)	6.09E-03	7	*BBS9*
	Rep	37	377	1238	0.14	51	630	2116	0.13	4.53E-01	1.05	(0.93-1.19)			
	Combined	53	678	2007	0.14	72	1007	3534	0.12	1.39E-03	1.17	(1.06-1.29)			
rs7814396	GWAS	680	362	44	0.21	1022	669	125	0.25	7.72E-05	1.30	(1.14-1.47)	6.56E-04	8	*SOX17*
	Rep	945	600	106	0.25	1605	1032	160	0.24	6.54E-01	1.02	(0.93-1.13)			
	Combined	1625	962	150	0.23	2627	1701	285	0.25	3.29E-02	1.09	(1.01-1.18)			
rs822304	GWAS	477	490	119	0.34	685	851	280	0.39	4.91E-05	1.26	(1.13-1.41)	2.37E-03	8	*MRPL49P2*
	Rep	652	767	232	0.37	1088	1319	390	0.38	8.20E-01	1.01	(0.92-1.1)			
	Combined	1129	1257	351	0.36	1773	2170	670	0.38	6.25E-03	1.10	(1.03-1.18)			
rs4741635	GWAS	174	509	401	0.40	199	847	768	0.34	5.84E-05	1.25	(1.12-1.4)	1.30E-02	9	*C9orf66*
	Rep	238	750	663	0.37	365	1287	1142	0.36	3.31E-01	1.05	(0.96-1.14)			

	Combined	412	1259	1064	0.38	564	2134	1910	0.35	1.09E-03	1.12	(1.05-1.2)			
rs6602097	GWAS	414	543	129	0.37	842	788	186	0.32	9.63E-05	1.25	(1.11-1.39)	1.84E-02	10	*PTER*
	Rep	684	796	171	0.34	1247	1230	320	0.33	3.14E-01	1.05	(0.96-1.15)			
	Combined	1098	1339	300	0.35	2089	2018	506	0.33	1.24E-03	1.12	(1.05-1.2)			
rs1028722	GWAS	51	366	669	0.22	58	514	1244	0.17	8.42E-05	1.31	(1.15-1.5)	1.78E-04	14	*CHORDC2P*
	Rep	62	470	1120	0.18	107	845	1845	0.19	2.70E-01	1.07	(0.95-1.19)			
	Combined	113	836	1789	0.19	165	1359	3089	0.18	1.06E-01	1.07	(0.99-1.17)			
rs12591707	GWAS	35	332	719	0.19	80	684	1052	0.23	1.49E-05	1.33	(1.17-1.52)	3.16E-04	15	*LOC100128104*
	Rep	80	559	1011	0.22	121	930	1689	0.21	6.57E-01	1.02	(0.92-1.14)			
	Combined	115	891	1730	0.2	201	1614	2741	0.22	1.87E-02	1.10	(1.02-1.2)			
rs11076442	GWAS	114	472	493	0.32	266	832	715	0.38	8.08E-05	1.26	(1.12-1.41)	4.80E-03	16	*LOC643358*
	Rep	215	782	654	0.37	395	1280	1107	0.37	6.39E-01	1.02	(0.93-1.12)			
	Combined	329	1254	1147	0.35	661	2112	1822	0.37	4.47E-03	1.11	(1.03-1.19)			
rs4793427	GWAS	961	123	2	0.06	1517	279	20	0.09	5.90E-05	1.55	(1.25-1.92)	5.82E-03	17	*LOC727973*
	Rep	1427	214	11	0.07	2395	380	21	0.08	4.88E-01	1.06	(0.9-1.25)			
	Combined	2388	337	13	0.07	3912	659	41	0.08	2.03E-03	1.23	(1.08-1.4)			
rs7219201	GWAS	830	240	16	0.13	1496	308	12	0.09	3.96E-05	1.42	(1.2-1.69)	9.94E-03	17	*BCAS3*
	Rep	1305	315	32	0.11	2229	531	37	0.11	3.49E-01	1.07	(0.93-1.22)			
	Combined	2135	555	48	0.12	3725	839	49	0.10	1.17E-03	1.19	(1.07-1.33)			
rs871739	GWAS	36	340	710	0.19	104	640	1072	0.23	9.48E-05	1.30	(1.14-1.48)	1.07E-02	17	*LOC100133332*
	Rep	78	541	1033	0.21	126	969	1701	0.22	4.12E-01	1.04	(0.94-1.16)			
	Combined	114	881	1743	0.2	230	1609	2773	0.22	1.92E-03	1.14	(1.05-1.24)			
rs6040836	GWAS	500	472	114	0.32	714	845	257	0.37	6.86E-05	1.26	(1.12-1.41)	1.93E-05	20	*LOC728573*
	Rep	604	746	202	0.37	1188	1257	352	0.35	6.28E-02	1.09	(1-1.19)			
	Combined	1104	1218	316	0.35	1902	2102	609	0.36	2.66E-01	1.04	(0.97-1.12)			
rs6036227	GWAS	43	328	715	0.19	43	461	1312	0.15	8.28E-05	1.33	(1.15-1.53)	2.20E-02	20	*CYB5P4*
	Rep	42	424	1186	0.15	58	696	2043	0.15	2.71E-01	1.07	(0.95-1.21)			

	Combined	85	752	1901	0.17	101	1157	3355	0.15	6.85E-04	1.17	(1.07-1.28)			
rs2853218	GWAS	24	291	771	0.16	30	359	1427	0.12	9.67E-06	1.42	(1.22-1.65)	5.95E-04	20	*ATRN*
	Rep	37	356	1259	0.13	43	643	2111	0.13	9.81E-01	1.00	(0.88-1.14)			
	Combined	61	647	2030	0.14	73	1002	3538	0.12	5.60E-03	1.15	(1.04-1.27)			
rs8129581	GWAS	99	474	513	0.26	128	685	1003	0.26	3.63E-05	1.28	(1.14-1.44)	3.90E-03	21	*PPIAL3*
	Rep	142	640	870	0.27	220	1096	1479	0.27	6.23E-01	1.02	(0.93-1.13)			
	Combined	241	1114	1383	0.27	348	1781	2482	0.27	2.99E-03	1.12	(1.04-1.21)			
rs17820958	GWAS	171	549	366	0.41	383	932	501	0.47	1.56E-05	1.26	(1.13-1.41)	4.13E-03	22	*PATZ1*
	Rep	328	807	517	0.44	584	1350	861	0.45	4.87E-01	1.03	(0.95-1.12)			
	Combined	499	1356	883	0.43	967	2282	1362	0.46	1.28E-03	1.12	(1.04-1.19)			
rs5997921	GWAS	194	550	342	0.43	415	934	467	0.49	5.48E-05	1.24	(1.12-1.38)	5.83E-03	22	*LIMK2*
	Rep	350	830	471	0.46	630	1365	802	0.47	5.92E-01	1.02	(0.94-1.12)			
	Combined	544	1380	813	0.45	1045	2299	1269	0.48	3.40E-03	1.11	(1.03-1.18)			
rs1034587	GWAS	186	556	344	0.37	246	858	711	0.37	2.39E-05	1.26	(1.13-1.4)	5.79E-04	22	*PIK3IP1*
	Rep	243	792	617	0.39	447	1289	1061	0.39	7.50E-01	1.01	(0.93-1.11)			
	Combined	429	1348	961	0.38	693	2147	1772	0.38	1.73E-02	1.09	(1.01-1.16)			

MAF, Minor allele frequency; OR, Odds ratio (Calculated based on non susceptible allele); CI, Confidence intervals

[a] *P*-assoc: *P*-value for GWAS and replication study obtained from the Cochran-Armitage trend test; *P*-value for the combined study obtained from simple combined analysis test

[b] *P*-het: *P*-value for heterogeneity test obtained from the Breslow-Day test

the P-value of $<1.35\times10^{-3}$ even after the Bonferroni correction (0.05/37) as shown in Table 8. Combined analysis of the results of the GWAS and the replication study suggested strong association of the locus of the *SIAH2* gene on chromosome 3q25.1 (rs6788895, $P_{combined}$ of 9.43×10^{-8} with OR of 1.22, 95% confidence interval 1.13–1.31) without any significant heterogeneity between the two studies ($P_{heterogeneity}= 2.33\times10^{-01}$). I further performed subgroup analysis to examine the association of rs6788895 with the subgroups of breast cancer; an invasive papilloductal breast cancer group and a HER2-negative breast cancer group, and found significant associations with them ($P_{combined}=3.61\times10^{-07}$, 6.78×10^{-06}, OR=1.23, 1.21 respectively) although they did not reach to the genome-wide significant level (Table 9).

Imputation analysis of this locus identified nine additional SNPs in strong LD ($r^2 > 0.8$) with rs6788895, and they showed similar levels of association with HRP tumors (Figure 4). I performed re-sequencing of 22,353 bps corresponding to the *SIAH2* gene and identified 10 novel genetic variations in addition to 37 variations reported previously. After the exclusion of 25 SNPs showing strong LD with rs6788895, I further genotyped 22 of the 47 variants. As a result, no genetic variant showed significant association with HRP breast cancer (Table 10 and Table 11). Using our sample sets, I additionally investigated the 22 SNPs (Table 12) that were reported to be associatedwith HRP breast cancer[25, 34-41]. Moderate association of four genetic variants, rs1292011, rs3803662, rs2981579 and rs3750817, was observed in our GWAS sample set ($P_{GWAS}= 5.89\times10^{-02}$, 6.95×10^{-03}, 8.68×10^{-04} and 5.03×10^{-04}, respectively). Further analysis using the replication set found significant association in two of the four SNPs, rs3750817 ($P_{replication}=5.39\times10^{-5}$, OR =1.22) and rs2981579 ($P_{replication}=1.21\times10^{-3}$, OR=1.20). Both SNPs are located within intron 2 of the fibroblast growth factor receptor 2 (*FGFR2*) gene. The combined analysis of the GWAS and replication phases of rs3750817 revealed strong association with $P_{combined}=8.47\times10^{-08}$ (OR=1.22) and that of rs2981579 was

Table 8. Association of SNP rs6788895 with HRP breast cancer

					Case				Control							
Chr.	Chrloc.	SNP	RA	Stage	TT	TG	GG	RAF	TT	TG	GG	RAF	P_{assoc}[1]	OR	(95% CI)	*P*het[2]
3	151950498	rs6788895	G	GWAS	106	456	524	0.69	242	832	742	0.64	2.34E-05	1.28	(1.14-1.43)	
				Rep	164	694	786	0.69	337	1,265	1,195	0.65	5.77E-04	1.18	(1.07-1.29)	2.33E-01
				Combined	270	1,150	1,310	0.69	579	2,097	1,937	0.65	9.43E-08	1.22	(1.13-1.31)	

Chr., chromosome; chrloc., chromosomal location (bp); RA, risk allele; 11, homozygous non risk genotype; 12, heterozygous genotype; 22, homozygous risk genotype; RAF, risk allele frequency; OR, odds ratio (calculated based on the risk allele); CI, confidence intervals; Rep, replication.

[1]Passoc, P value for the GWAS and replication study obtained from the Cochran-Armitage trend test and P value for the combined study obtained from the simple combined test.

[2]Phet, P value for heterogeneity test obtained from the Breslow-Day test.

Table 9. Associations of rs6788895 in the subgroups

		Case				Control								
	Stage	TT	TG	GG	RAF	TT	TG	GG	RAF	risk allele	P_{assoc}[a]	OR	(95% CI)	*P*het[b]
Invasive papilloductal	1st	56	316	346	0.70	242	832	742	0.64	G	1.37E-05	1.34	(1.17-1.53)	1.00E-01
	Rep	134	578	640	0.69	337	1265	1195	0.65	G	2.32E-03	1.17	(1.06-1.29)	
	Total	190	894	986	0.69	579	2097	1937	0.65	G	3.61E-07	1.23	(1.13-1.33)	
Her-2 receptor negative	1st	75	314	346	0.68	242	832	742	0.64	G	1.63E-03	1.23	(1.08-1.4)	7.33E-01
	Rep	129	411	549	0.69	337	1265	1195	0.65	G	1.13E-03	1.20	(1.08-1.33)	
	Total	204	725	895	0.69	579	2097	1937	0.65	G	6.78E-06	1.21	(1.11-1.31)	
Hormonal receptor negative	1st	44	176	178	0.33	242	832	742	0.36	G	1.03E-01	1.15	(1.07-1.13)	
Her-2 receptor positive	1st	23	72	89	0.32	242	832	742	0.36	G	1.15E-01	1.2	(1. 07-1.23)	

11, homozygous non risk genotype; 12, heterozygous genotype; 22, homozygous risk genotype; RAF, risk allele frequency; OR, odds ratio (calculated based on the risk allele); CI, confidence intervals; Rep, replication.

[a]Passoc, P value for the GWAS and replication study obtained from the Cochran-Armitage trend test and P value for the combined study obtained from the simple combined test.

[b]Phet, P value for heterogeneity test obtained from the Breslow-Day test.

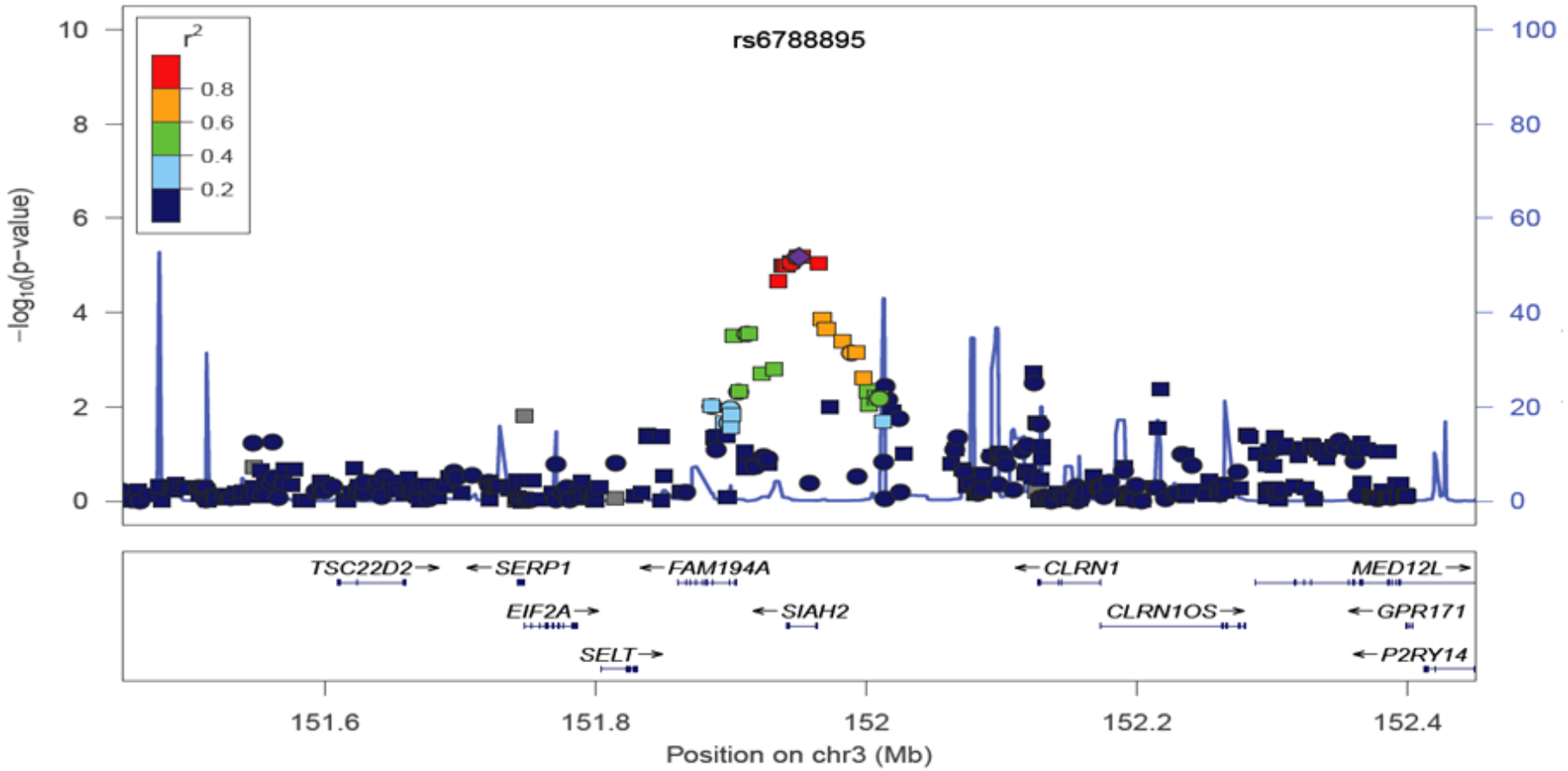


Figure4. Regional association plots of the locus associated with HRP breast cancer on chromosomes 3q25.1 (intron of seven in absentia homolog 2 (*SIAH2*) For each plot, _log10P of the Cochran–Armitage trend test of SNPs in GWAS was plotted against relative chromosomal locations. The square and rounded signs represent imputed and genotyped SNPs, respectively. All SNPs are color coded as red (r^2=0.8–1.0), orange (r^2=0.6–0.8), green (r^2=0.4–0.6), light blue (r^2=0.4–0.6), and dark blue (r^2=0.2) according to their pair wise r^2 to the marker SNP. The marker SNP is represented in purple color. SNP positions followed NCBI build 36 coordinates. Estimated recombination rates (cM/Mb) are plotted as a blue line.

Table 10. Association study of reported SNPs resulted from re-sequence analysis in HRP breast cancer

SNP	Case			Control			MAF		P_{assoc} [a]	OR	95% CI
	11	12	22	11	12	22	case	Ctrl			
rs6788895	221	947	1098	100	332	296	0.31	0.37	3.36E-05	1.30	(1.15-1.47)
rs476663	959	988	316	305	336	87	0.36	0.35	6.02E-01	1.03	(0.91-1.17)
rs7615292	1014	1060	192	315	336	77	0.32	0.34	1.88E-01	1.08	(0.96-1.23)
rs7636866	321	974	970	93	316	319	0.36	0.34	4.20E-01	1.05	(0.93-1.19)
rs2116684	333	1033	900	126	312	288	0.37	0.39	3.65E-01	1.06	(0.94-1.2)
rs19862837	253	1020	993	62	349	316	0.34	0.33	4.16E-01	1.05	(0.93-1.19)
rs57359526	274	1009	983	112	296	318	0.34	0.36	3.19E-01	1.07	(0.94-1.21)
rs6789306	287	991	988	102	313	313	0.35	0.36	5.05E-01	1.04	(0.92-1.18)
rs1148369	38	109	2118	0	68	660	0.04	0.05	3.90E-01	1.15	(0.87-1.53)
rs76182650	909	1100	257	307	335	86	0.36	0.35	5.73E-01	1.04	(0.91-1.17)
rs1733711	261	974	1031	90	301	335	0.33	0.33	9.35E-01	1.01	(0.89-1.14)
rs61628582	930	1036	300	301	345	80	0.36	0.35	3.60E-01	1.06	(0.94-1.2)

MAF, minor allele frequency, OR, odds ratio; CI, confidence intervals.
[a]Passoc, P value obtained from the Cochran-Armitage trend test.

Table 11. Association study of unreported SNPs resulted from re-sequence analysis in HRP breast cancer

SNP position		Case			Control			MAF		P_{assoc} [a]	OR	95% CI
		11	12	22	11	12	22	case	Ctrl			
150481266	5' UTR	10	179	2074	3	68	657	0.04	0.05	2.88E-01	1.16	(0.89-1.53)
150471735	Intron	201	1036	1029	76	332	320	0.32	0.33	2.71E-01	1.07	(0.94-1.21)
150470898	Intron	2128	105	33	685	33	10	0.04	0.04	8.42E-01	1.04	(0.76-1.42)
150467858	Intron	2226	37	3	718	8	2	0.01	0.01	6.89E-01	1.15	(0.61-2.19)
150466004	Intron	2138	106	22	686	40	2	0.03	0.03	6.25E-01	1.10	(0.78-1.55)
150465616	Intron	2148	112	6	700	26	2	0.03	0.02	1.73E-01	1.34	(0.89-2)
150465581	Intron	2123	110	33	690	28	10	0.04	0.03	3.79E-01	1.19	(0.86-1.64)
150461443	Intron	2186	76	4	710	17	1	0.02	0.01	1.78E-01	1.43	(0.87-2.36)
150460001	exon 2	2138	120	8	686	41	1	0.03	0.03	9.29E-01	1.02	(0.72-1.44)
150459679	3` UTR	0	100	2166	0	36	692	0.022	0.02	5.49E-01	1.12	(0.76-1.65)

MAF, minor allele frequency, OR, odds ratio; CI, confidence intervals.
[a]Passoc, P value obtained from the Cochran-Armitage trend test.

Table 12: GWAS of HRP breast cancer reported in other population

SNP	Case			Control			P_{assoc} [a]	OR	(95% CI)	Ref
	11	12	22	11	12	22				
rs3750817 (FGFR2)	256	545	285	519	901	396	5.03E-04	1.21	(1.09-1.34)	34
rs2981579 (FGFR2)	253	538	295	346	889	580	8.68E-04	0.83	(0.75-0.93)	42
rs1219648 (FGFR2)	393	524	169	741	844	231	4.34E-03	1.17	(1.05-1.31)	43
rs1292011 (12q24)	582	422	82	914	738	163	5.89E-02	0.89	(0.79-1)	44
rs11249433 (1p11.2)	1013	71	2	1717	98	1	1.32E-01	1.26	(0.93-1.71)	45
rs7752591 (C6orf97)	171	530	384	276	835	705	1.33E-01	0.92	(0.82-1.03)	46
rs12500093 (MTNR1A)	76	411	599	133	722	961	2.86E-01	1.07	(0.95-1.2)	38
rs2380205 (TRNAV)	5	155	926	9	231	1576	2.88E-01	0.90	(0.73-1.1)	38
rs7703618 (MRPS30)	127	488	471	226	838	752	2.94E-01	1.06	(0.95-1.19)	47
rs1011970 (ANRIL)	3	119	964	11	212	1593	2.97E-01	1.13	(0.9-1.41)	38
rs13387042 (TNP1)	11	223	852	23	395	1398	3.20E-01	1.09	(0.92-1.28)	48
rs6929137(C6orf97)	76	429	580	128	680	1008	4.02E-01	0.95	(0.84-1.07)	46
rs490361(NCOA7)	693	349	44	1186	561	69	4.18E-01	1.06	(0.92-1.21)	36
rs3734805 (C6orf97)	598	422	66	1038	664	114	4.20E-01	1.05	(0.93-1.19)	46
rs9383932 (C6orf97)	454	498	133	786	818	212	4.33E-01	1.05	(0.94-1.17)	46
rs4866929 (HCN1)	392	508	186	651	890	275	5.13E-01	1.04	(0.93-1.16)	47
rs9375411(NCOA7)	212	544	330	357	884	575	6.73E-01	0.98	(0.88-1.09)	36
rs10454853 (5p13.3)	789	268	29	1305	463	48	6.99E-01	0.97	(0.84-1.13)	49
rs852003 (C6orf97)	124	486	476	217	778	821	7.44E-01	0.98	(0.88-1.1)	46
rs2823093(NRIP1)	0	27	1059	1	46	1769	7.99E-01	1.06	(0.66-1.71)	44
rs4973768 (SLC4A7)	38	334	714	64	564	1188	8.68E-01	1.01	(0.88-1.16)	48
rs549438 (NCOA7)	615	408	63	1028	677	111	8.86E-01	0.99	(0.88-1.12)	36

OR, odds ratio; CI, confidence intervals. [a]Passoc, P value for the GWAS obtained from the Cochran-Armitage trend test.

1.77×10^{-06} (OR=1.20) (Table 13). Imputation analysis of this locus identified three additional SNPs (rs9420318, rs11199914 and rs10736303) that showed similar levels of association with rs3750817 (Figure 5).

To identify functional variants of the *SIAH2* gene, I focused on the 25 SNPs showing strong LD with rs6788895 that had a significant association with HRP-breast cancer. The *SIAH2* gene is composed of two exons, and one of the 25 SNPs are located in the 5'-flanking region and 24 are in intron 1. None of the 25 SNPs are localized in the exons. I selected three SNPs SNP X, SNP Y and SNP Z which are located in the transcription factor binding sites; SNP X in the 5'-flanking region, and SNP Y and SNP Z in intron 1. To study the effect of the three SNPs on the *SIAH2* expression, I prepared six reporter plasmids containing three regions; one in the 5' flanking region (region A: between -2,151 and -571, Figure 6) and two in the intron 1 (region B: between +10,063 and +11,152, Figure 7, and region C: between +13,956 and 15,864, Figure 8). These constructs contained either major or minor allele of the three SNPs. Subsequently, a reporter gene assay was performed using these constructs (pGL3A_Major and pGL3A_Minor for SNP X, pGL3PB_Major and pGL3PB_Minor for SNP Y, and pGL3PC_Major and pGL3PC_Minor for SNP Z) in T-47D and BT-474 breast cancer cells. Mock vectors (pGL3-Basic or pGL3-Promoter) were used for their controls.

As a result, I did not find any significant difference of reporter activity between pGL3A_Major (containing major susceptible G allele) and pGL3A_Minor (containing minor non susceptible T allele) in T-47D or BT-474 cells, although both constructs revealed markedly higher activity compared to mock vector (Figure9). This result suggested that the 5' flanking region between (-2,151 and -571) contains the promoter activity and that the SNPs in this region may not affect the activity at least in the condition.

Table13. Association of rs2981579 and rs3750817 in ***FGFR2*** in different population

SNPs	Minor/Major allele	MAF	OR	*P*-trend	Population
rs3750817	T/C	0.49	1.22	8.47×10^{-08}	Japanese
rs3750817	T/C	0.37	0.78	8.20×10^{-08}	American[34]
rs2981579	A/G	0.47	1.20	1.77×10^{-06}	Japanese
rs2981579	A/G	0.42	1.43	3.60×10^{-31}	UK[37]
rs2981579	A/G	0.44	1.31	2.60×10^{-09}	American[34]

MAF, minor allele frequency; OR, odds ratio (calculated based on the non susceptible allele) except rs3750817 in American population OR, calculated based on the susceptible allele).

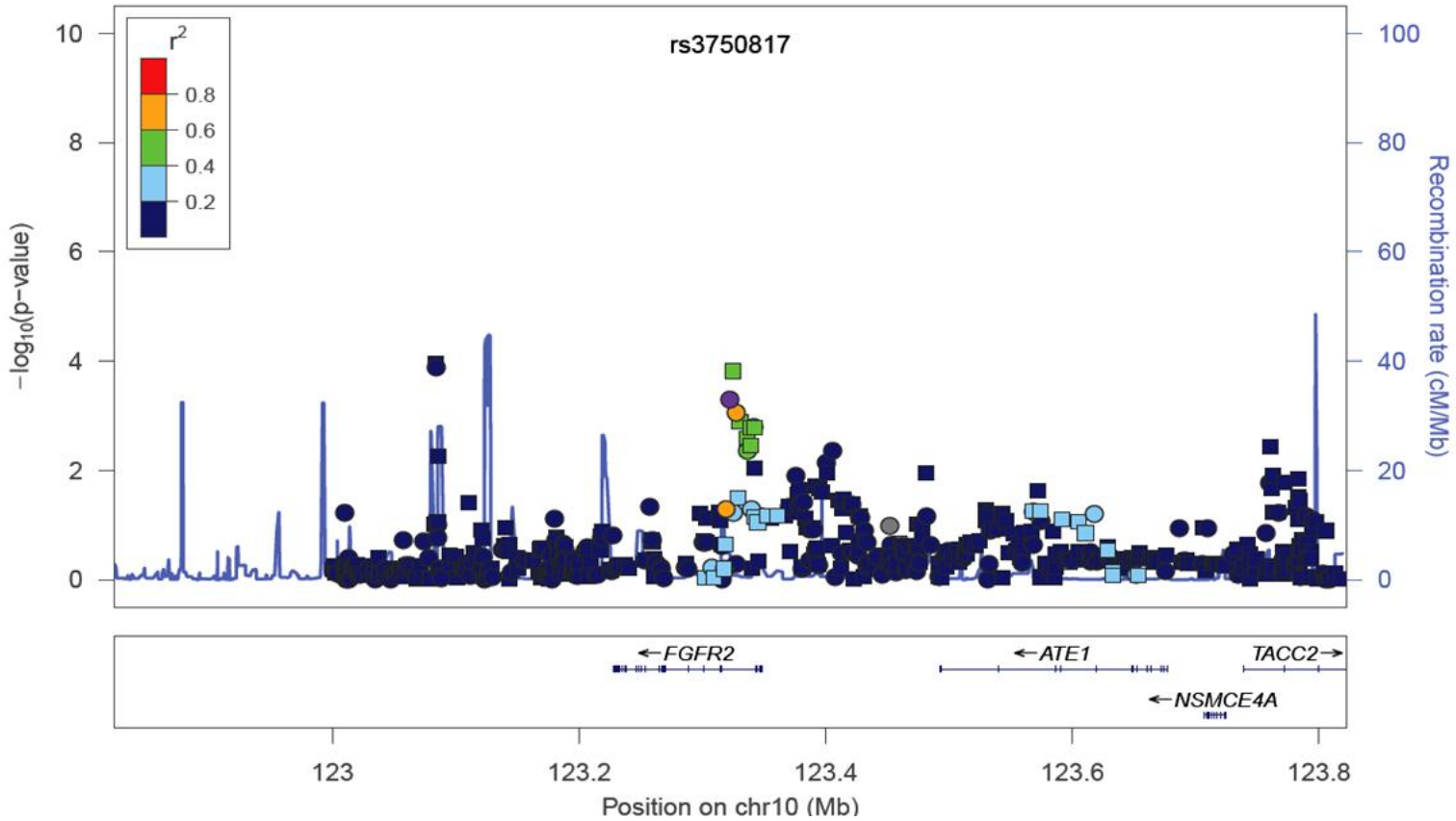


Figure5. Regional association plots of the locus associated with HRP 10q26 (fibroblast growth factor receptor 2 (*FGFR2*)). For each plot, _log10P of the Cochran–Armitage trend test of single-nucleotide polymorphisms (SNPs) in the genome-wide association study (GWAS) was plotted against relative chromosomal locations. The square and rounded signs represent imputed and genotyped SNPs, respectively. All SNPs are color coded as red (r^2=0.8–1.0), orange (r^2=0.6–0.8), green (r^2=0.4–0.6), light blue (r^2=0.4–0.6), and dark blue (r^2=0.2) according to their pair wise r^2 to the marker SNP. The marker SNP is represented in purple color. SNP positions followed NCBI build 36 coordinates. Estimated recombination rates (cM/Mb) are plotted as a blue line.

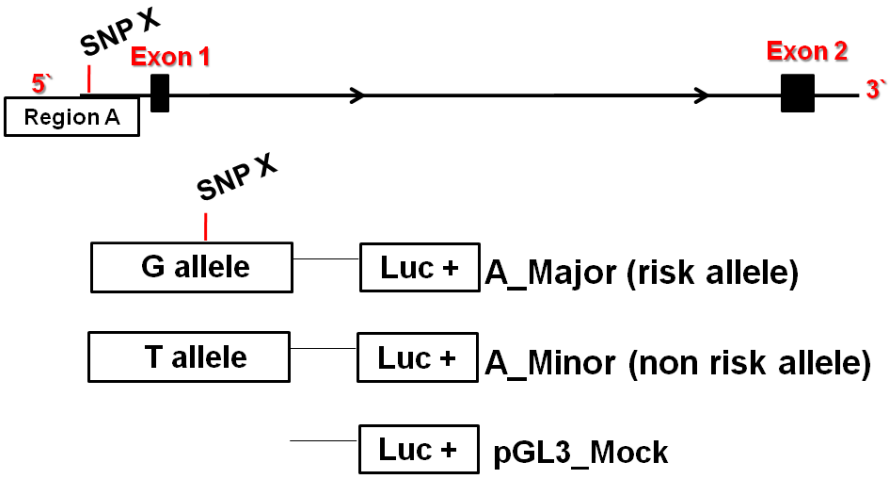


Figure 6: pGL3A_Major and pGL3A_Minor

Reporter plasmid contains region A (-2.151Kb to -571 kb) of *SIAH2*. Plasmid pGL3A_Major contains major susceptible G allele and pGL3A_Minor contains minor non susceptible T allele of SNP X. pGL3_Mock is pGL3 basic empty vector.

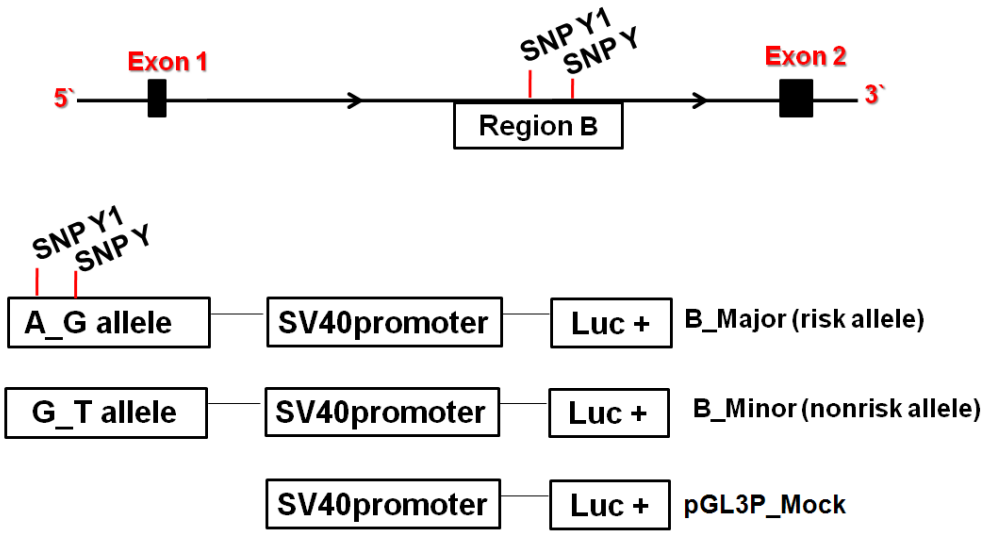


Figure 7:pGL3PB_Major and pGL3PB_Minor

Reporter plasmid contains region B(+10.063 to+ 11.152)of *SIAH2*. Plasmid pGL3PB_Major contains major susceptible allele and pGL3PB_Minor contains minor non susceptible allele of SNP Y and SNP Y1.pGL3P_Mock is pGL3 promoter empty vector.

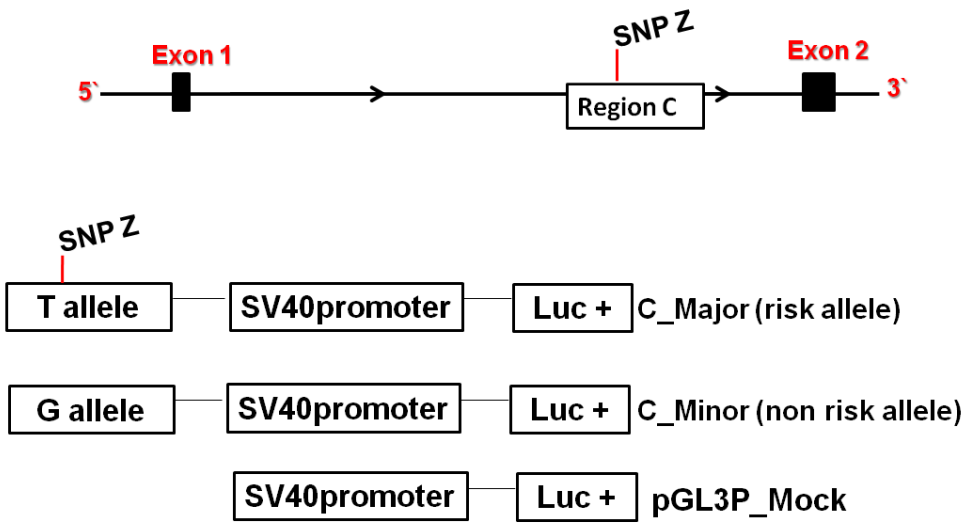
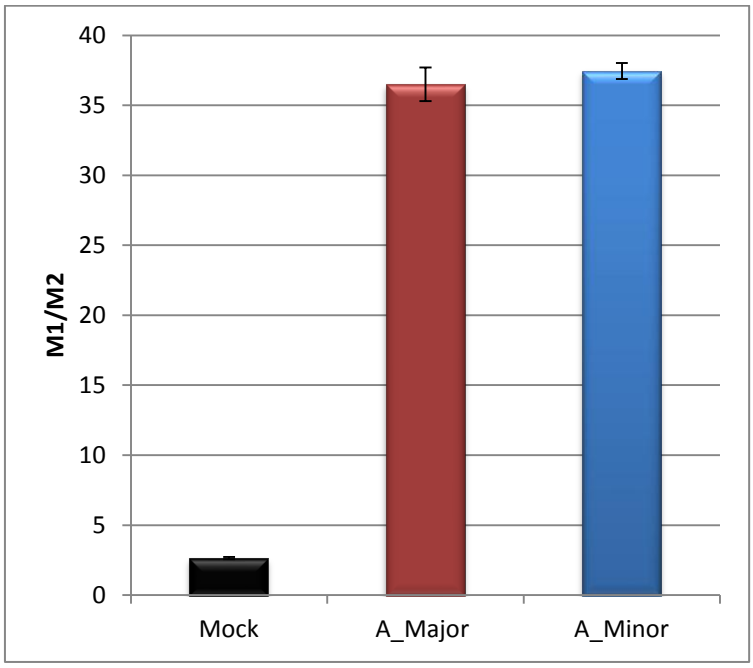


Figure 8:pGL3PC_Major and pGL3PC_Minor.

Reporter plasmid contains region C (+13.956 to +15.864) of *SIAH2*. Plasmid pGL3PC_Major contains major susceptible T allele and pGL3PB_Minor contains minor non susceptible Callele of SNP Z. pGL3P_Mock is pGL3 promoter empty vector.

A)



B)

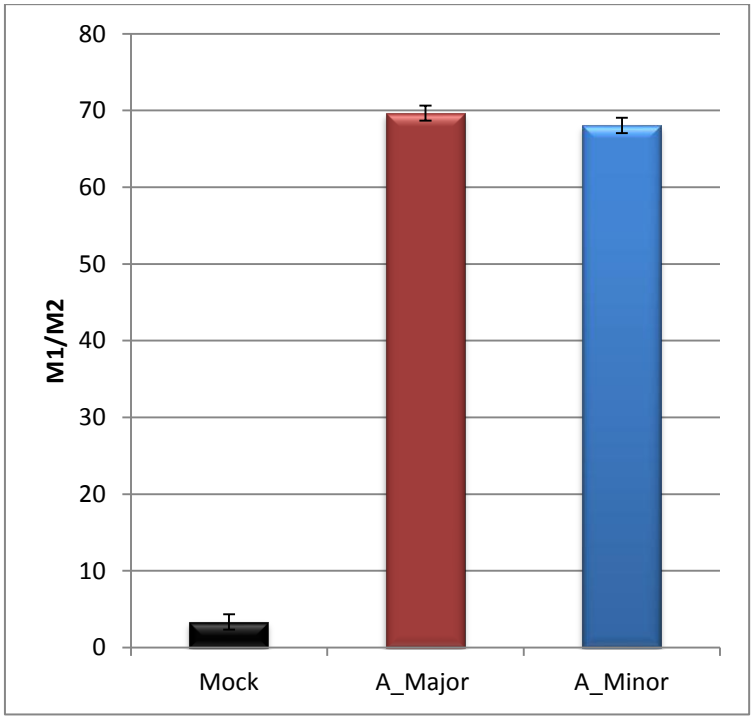


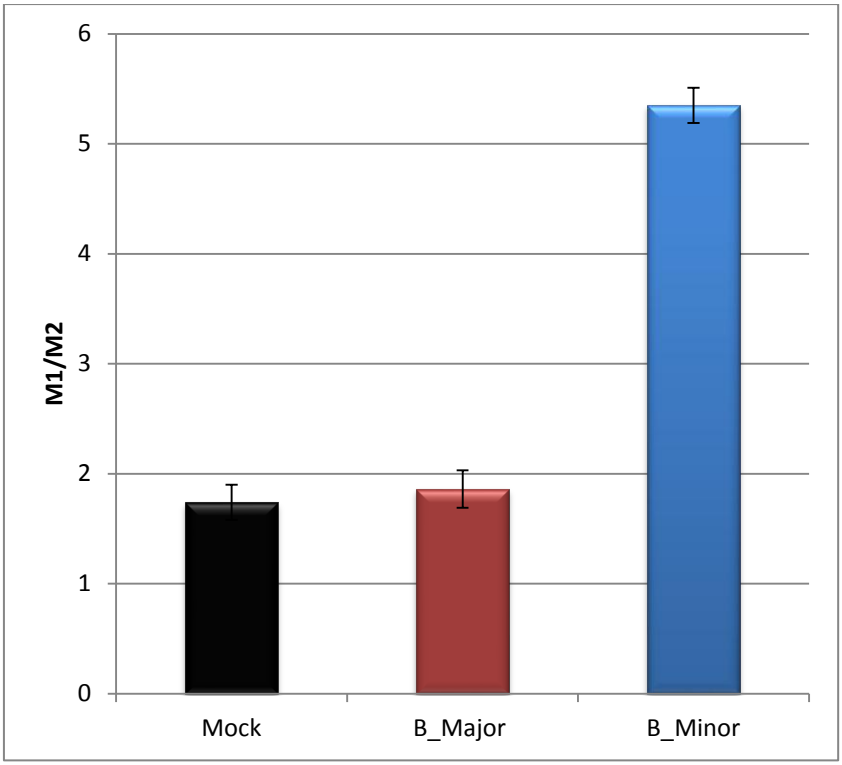
Figure9. Reporter gene assay of region A contains major or minor allele of SNP X in:

A) T47D and B) BT474 cell lines.
Mock is pGL3 basic.

Interestingly, we found significant difference of reporter activity between pGL3PB_Major and pGL3PB_Minor in both T-47D and BT-474 cells. The activity of pGL3PB_Minor (containing the minor allele) was 2.9 fold higher in T-47D cells and 3.2 fold higher in BT-474 cells than that of pGL3PB_Major (containing major allele) (Figure 10). On the other hand, inconsistent reporter activity was detected between pGL3PC_Major and pGL3PC_Minor in T-47D and BT-474 cells. The activity of pGL3PC_Minor was slightly higher p value=5×10^{-2} than pGL3PC_Major in T-47D cells, but no significant difference was found between the two reporter plasmids in BT-474 cells p value = 3×10^{-1} (Figure 11). These data suggested that region B may include an enhancer region(s) for the transcriptional activation of *SIAH2*, and that minor allele of SNP Y may be associated with augmented expression of *SIAH2* compared to the major allele.

Therefore, I further analyzed functional roles of SNPs within region B. I identified 5 SNPs in region B, and found that a SNP, rs6788895, showed complete LD with SNP Y (r^2=1) among the 5 SNPs. I additionally prepared four reporter plasmids containing SNP Y; two containing 290 bp and two containing 60 bp in region B (Figure 12 and 13). All four constructs harbored either susceptible minor allele G or non-susceptible major allele T of SNP Y. The activities of these four plasmids were analyzed by the reporter gene assay in T47D and BT474 cells (Figure 14). Consistently, the reporter plasmids pGL3PB290_Minor containing non susceptible allele showed approximately 2.3 and 3.6 fold higher reporter activities compared to pGL3PB290_Major containing susceptible allele in T-47D and BT-474 cells, respectively (Figure 14). This result suggested that this region carries a crucial element(s) for the transactivation. Although the increase was lower compared to the 290 bp-region, the reporter plasmids pGL3PB60_Minor showed increased activity compared to pGL3PB60_Major. Since the 290 bp region but not the 60 bp region included another SNPY1, located 138 bp at a 5' side of SNP Y. I further investigated whether this SNP also associate with the enhancer

A)



B)

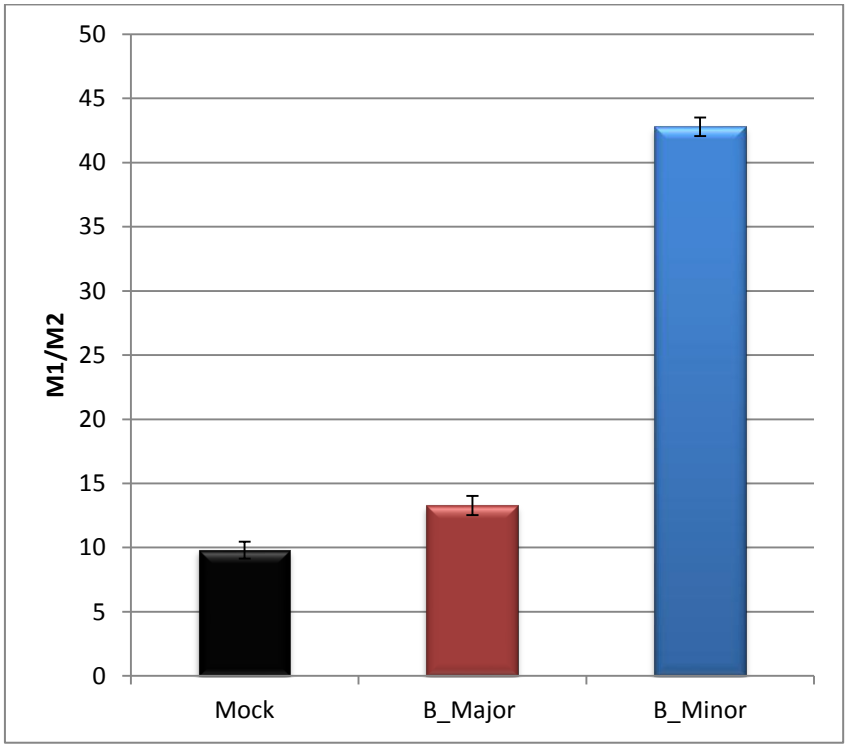
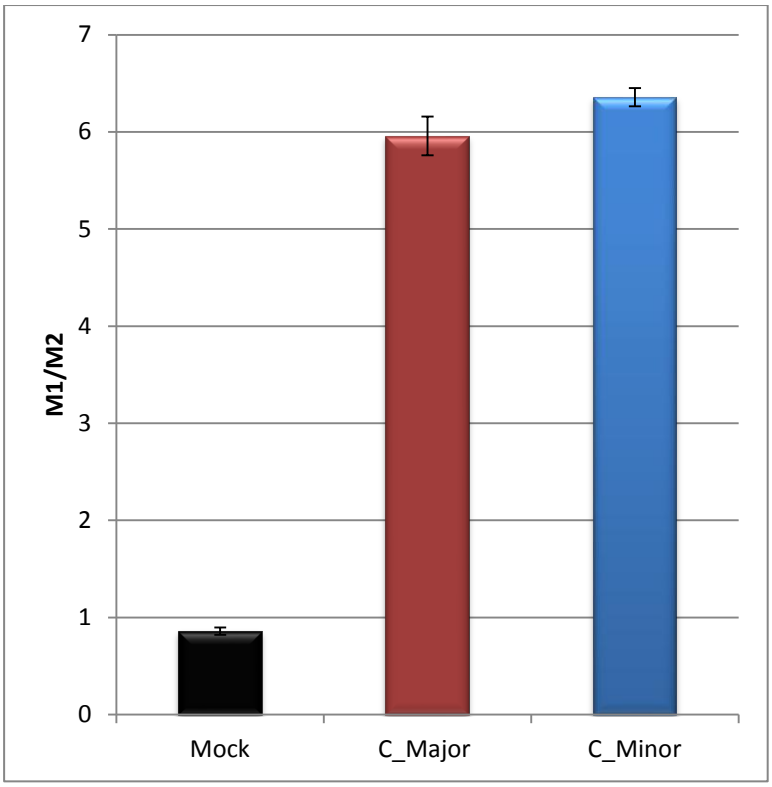


Figure10. Reporter gene assay of region B containing major or minor allele of SNP Y and SNP Y1 in

A) T47D and B) BT474 cell lines. Mock is pGL3promoter.

A)



B)

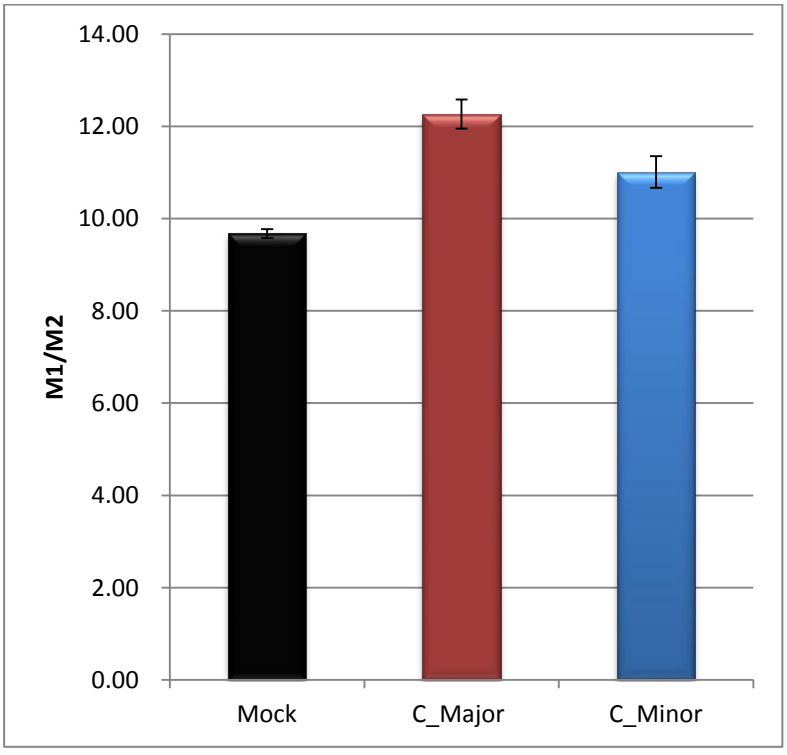


Figure11. Reporter gene assay of region C containing major or minor allele of SNP Z in A) T47D and B) BT474 cell lines. Mock is pGL3promoter

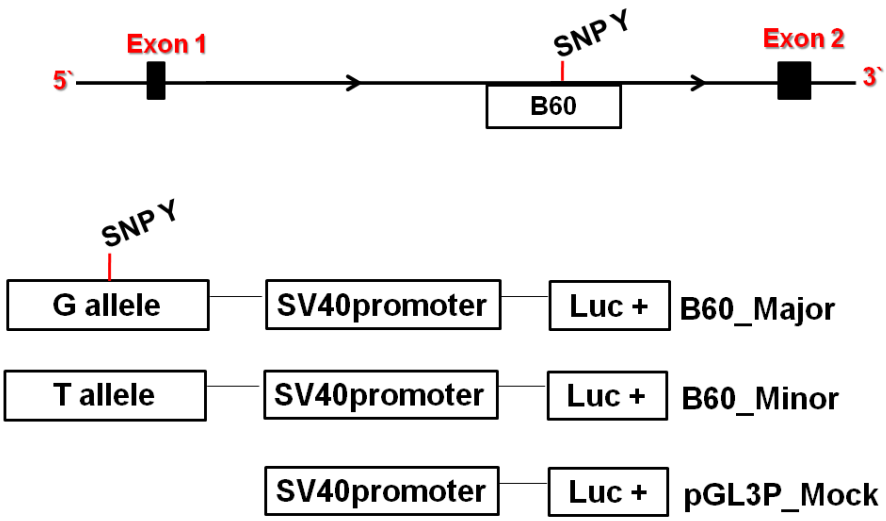


Figure12. B60_Major and B60_Minor

Reporter plasmid contains region B (+10.413 to+ 10.470) of *SIAH2*. Plasmid B60_Major contains major susceptible G allele and B60_Minor contains minor non susceptible T allele of SNP Y. pGL3P_Mock is pGL3 promoter empty vector.

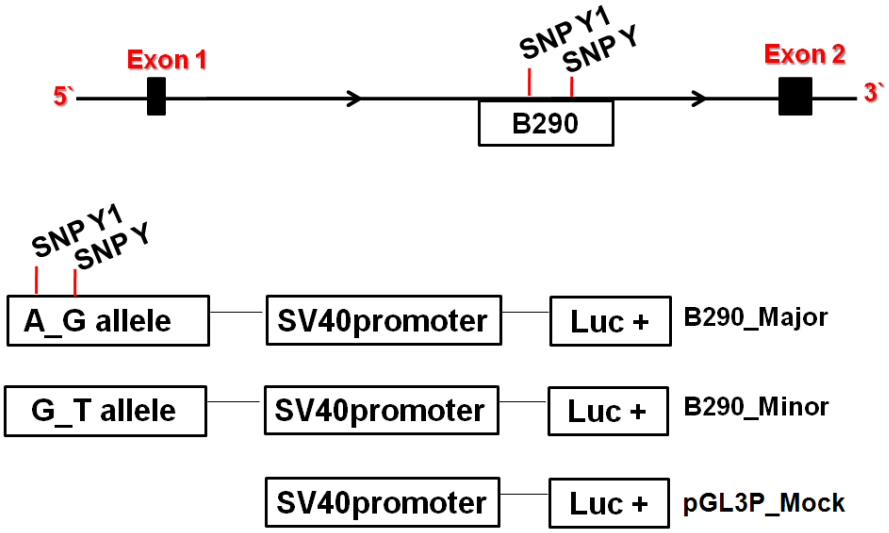
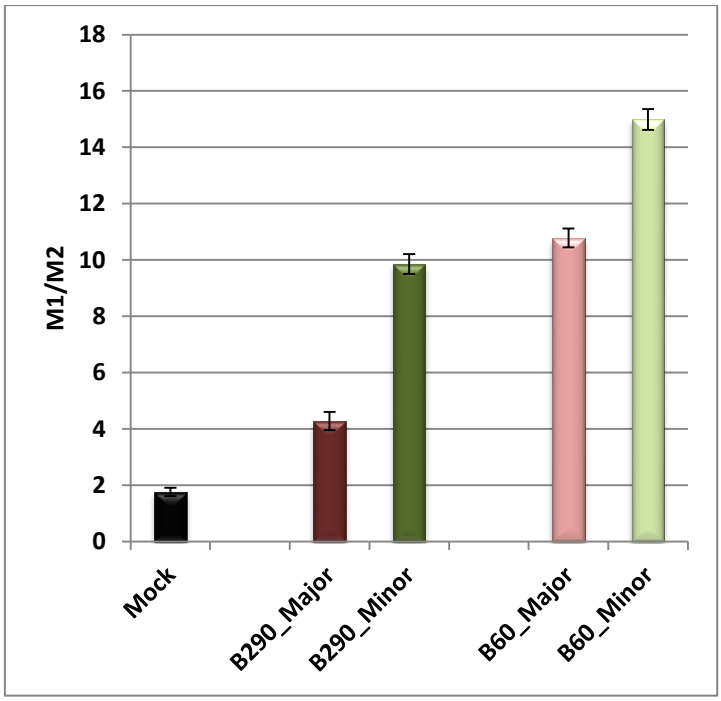


Figure13. B290_Major and B290_Minor

Reporter plasmid contains region B (+10.267to + 10.556) of *SIAH2*. Plasmid B290_Major contains major susceptible alleles and B290_Minor contains minor non susceptible alleles of SNP Y and SNP Y1. pGL3P_Mock is pGL3 promoter empty vector.

A)



B)

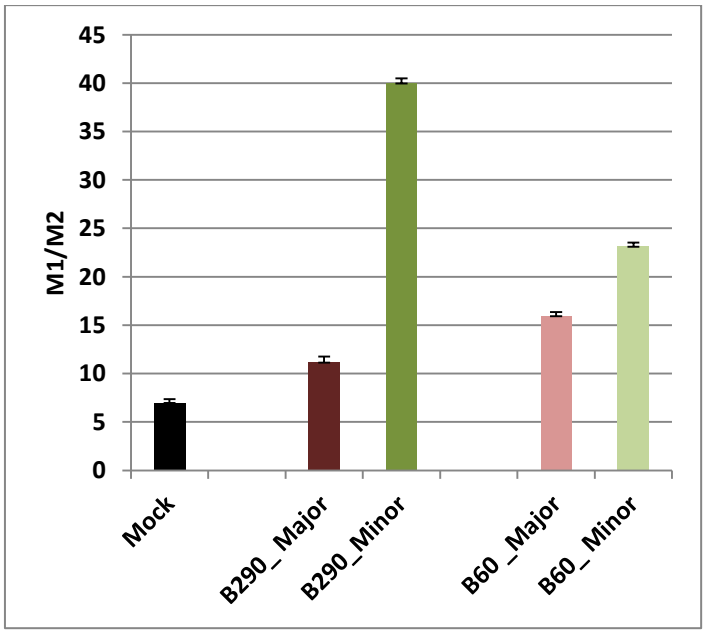


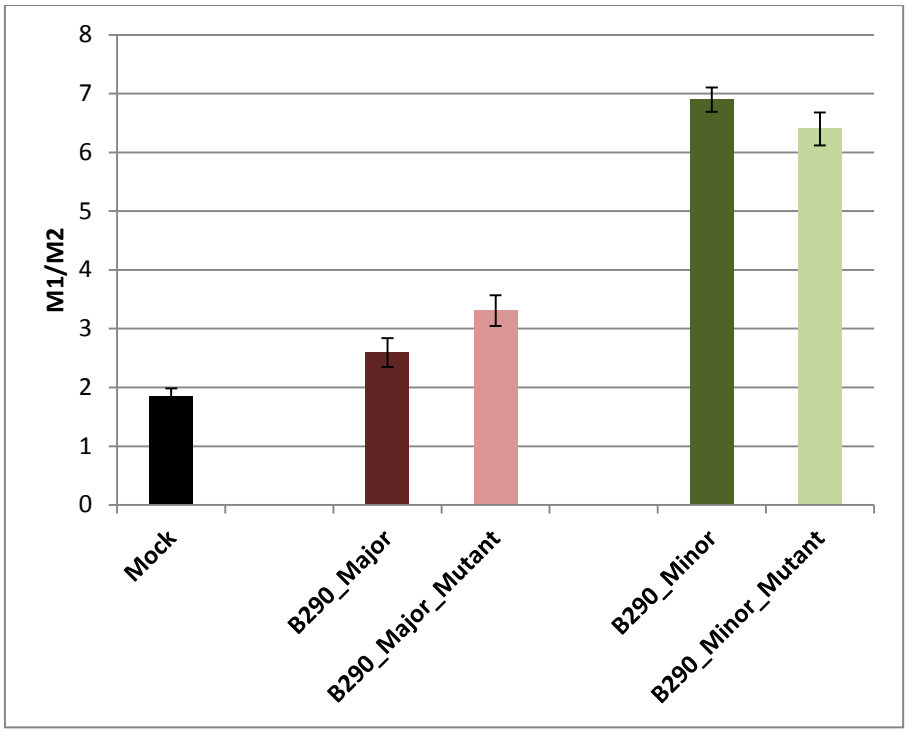
Figure14. Reporter gene assay of plasmids B290 and B60 containing major or minor allele of SNP Y in

A) T47D and B) BT474 cell lines. Mock is pGL3promoter.

activity of *SIAH2*. This SNP, Y1, is completely linkage to SNP Y (r^2=1); G in SNP Y and A in Y1, and T in SNP Y and G in SNP Y1. I prepared additional reporter plasmids, pGL3PB290_Major_Mutant containing G in SNP Y and mutated allele (G) in SNP Y1, and pGL3PB290_Minor_Mutant containing T in SNP Y and mutated allele (T) in SNPY1 by site directed mutagenesis, and performed a reporter assay. As shown in Figure 15, compared to the variations of SNP Y, limited levels of difference in the reporter activity were observed by the mutation at SNP Y1. Therefore I concluded that the difference in transcriptional activity of the 290 bp region should be caused by the variation of SNP Y.

To examine the effect of estrogen on *SIAH2* expression, I measured the reporter activities of pGL3PB_Major and pGL3PB_Minor in the presence or absence of estradiol. T-47D and MCF7 cells transfected with pGL3PB_Major or pGL3PB_Minor were cultured in the presence or absence of 17β estradiol. Although slightly higher activity was detected with pGL3PB_Minor than pGL3PB_Major, no significant increase was observedby 17β estradiol (Figure 16).

A)



B)

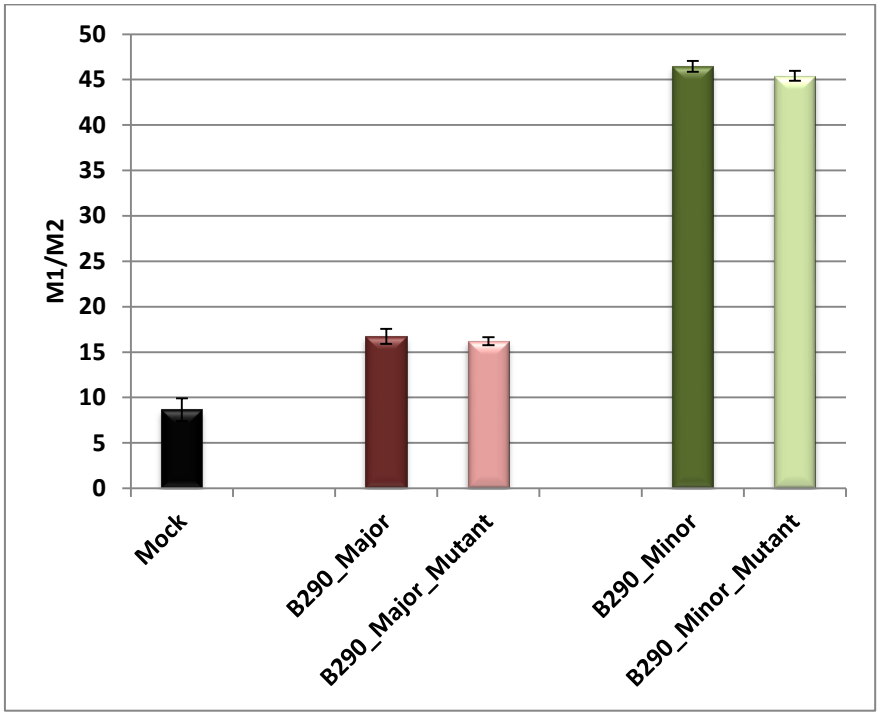
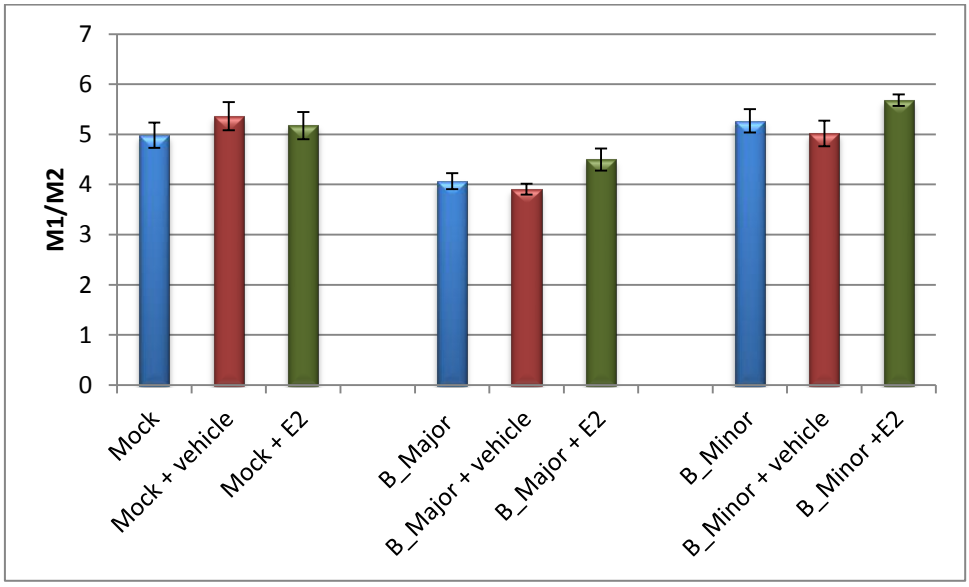


Figure15. Reporter gene assay of plasmids B290 contains major or minor allele of Y1 and mutant major or mutant minor allele of in A) T47D and B) BT474 cell lines. Mock is pGL3 promoter.

A)



B)

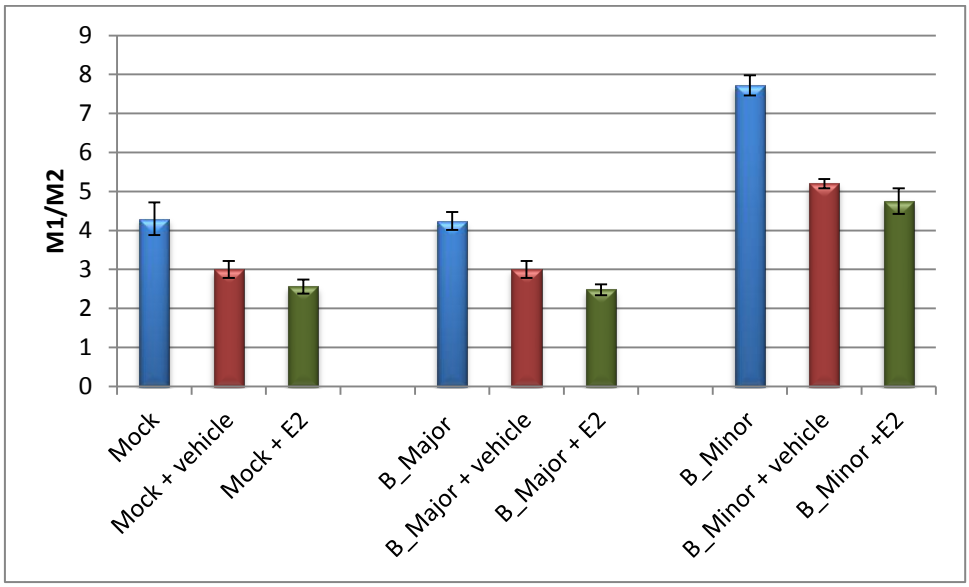


Figure16. Reporter gene assay of region B containing major or minor allele of SNP Y with 17β estradiol treatment in A) MCF7 and B) T47D cell lines. Mock is pGL3 promoter, E2 is 17β estradiol and vehicle is ethanol.

Chapter 5: Discussion

In this study, I successfully identified a new genetic locus, in *SIAH2*, that is associated with HRP breast cancer in Japanese population. In addition, I validated the significant association of two SNPs, rs3750817 and rs2981579 in intron 2 of *FGFR2* with Japanese breast cancer, which had been reported in various populations including Caucasians, Ashkenazi, Chinese and Siberians[38, 50, 51] (Rinella et al., 2013;Boyarskikh et al., 2009;Fanale et al., 2012; Chen et al., 2012;Raskin et al., 2008). Although other SNPs; rs2981582, rs2420946, rs11200014, located in intron 2 of *FGFR2* were reported to be associated with breast cancer[50], I could not find any association of these SNPs with HRP breast cancer in the Japanese population. Moreover these SNPs showed low LD and low r^2 with rs3750817 and rs2981579 in Japanese (Figure 17) indicating that these SNPs may not play a role in Japanese HRP breast cancer although they are found to be associated with breast cancer in other population. The minor allele of rs3750817, a SNP in intron 2 of *FGFR2*, is a protective allele for both Japanese and American populations with comparable ORs (Table 13). The minor allele of rs2981579 is a risk allele for UK and American populations as well as Japanese. In addition, the association of the two SNPs in *FGFR2* intron 2 with estrogen receptor positive cancer is consistent with earlier report[50]. Therefore FGFR2-mediated signaling should play a crucial role in HRP-breast carcinogenesis. The intron 2 is highly conserved in various species[37]and possesses putative binding sites for transcription factors including FOXA1, USF1, http://genome.ucsc.edu/ and ER[52]. Since we and others proved that intron 2 of *FGFR2* is associated with estrogen receptor positive breast cancer, it is tempting to speculate that the variant(s) in the ER-binding element is a causative SNP. Nevertheless, FOXA1 and USF1 are also reported tobe implicated in the estrogen receptor signaling pathway[52]. FOXA1 has the ability to bind to the same enhancer regions of ER as a pioneer factor that facilitates the

association of ER with compacted DNA. In addition USF1 is considered component of a multi-protein complex of transcription factors that interact with ER alpha in breast cancer[53]. Therefore future investigation of the variants in these binding elements may clarify the roles of these transcription factors and reveal the precise mechanisms in HRP breast cancer mediated by FGFR2.

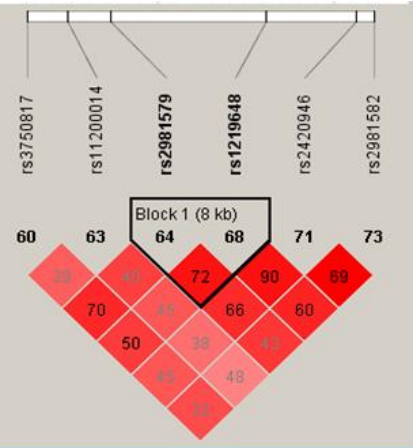


Figure17. r^2 between rs3750817, rs2981579 and rs2981582, rs28981579, rs2420946, rs11200014, rs1219648.

In the study of *SIAH2*, I focused on three SNPs; SNP X, SNP Y, and SNP Z Figure 18 A&B that are located in transcription factors binding sites. Among the three, the variation of SNP Y was significantly associated with the transcriptional activity, suggesting that the variation may play a role as a transcriptional enhancer. Although SNP Y, showing perfect LD with the marker SNP rs6788895, was significantly associated with HRP tumors, it was not associated with hormonal receptor negative breast cancer (Table 9). This may suggest that SIAH2 is involved in the pathogenesis of HRP-breast cancer but not in hormone-negative breast cancer. Notably, recent genetic studies showed that the *SIAH2* region on chromosome 3q25.1 might be involved in estrogen-dependent diseases such as development of peritoneal leiomyomatosis and uterine fibrosis[54,55]. In addition, SIAH2 expression is up-regulated by estrogenand down regulated by selective estrogen receptor modulator (SERM) tamoxifen[56, 57]. Therefore, *SIAH2* may play a crucial in the regulation of cell growth and/or cell fate in response to estrogen.

SIAH2 encodes a 324 amino acid protein that shares 68% identity with *Drosophila*sina and 77% identity with SIAH1[58]. SIAH proteins are suggested to play an important role in the specification of cell fate. It was reported that SIAH2 acts as an E3-ubiquitin ligase as a part of ubiquitin ligase-complexes and that the complexes are involved in protein degradation through ubiquitin-proteosome system. SIAH2 was known to induce ubiquitin-mediated degradation of many substrates, including proteins involved in transcriptional regulation (POU2AF1, PML and NCOR1), a cell surface receptor (DCC) and an anti-apoptotic protein (BAG1). These proteins were reported to have implication to breast cancer by different mechanisms[59-63]. In addition, SIAH2 protein was indicated to have an essential role in the hypoxic response by regulating the hypoxia-inducible factor-α[64]. The role of *SIAH*2 in human carcinogenesis remains controversial. It is reported that *SIAH2* has tumor suppressor effect by elimination of oncogenic proteins through proteasomal and kinase signaling pathways in leukemia[65]. On the contrary, increased *SIAH2* expression and its oncogenic role were shown in human neuroendocrine (NE) prostate tumor samples[66]. Additionally, nuclear accumulation of SIAH2 was detected in human hepatocellular carcinoma and its expression was associated with enhanced motility and proliferation of liver cancer cell[67]. Although previous reports implicated possible roles of *SIAH2* in breast carcinogenesis[32,64,68,69], the role of *SIAH*2 in breast carcinogenesis is also controversial. It was reported that SIAH2 was under-expressed in breast cancer and that the patients with low level of expression of SIAH2 had higher risk of recurrence and low chance of disease free survival with increased the aggressiveness of the breast cancer[70]. These data support the possible role of SIAH2 as a tumor suppressor for breast cancer. This view is in good agreement with the reduced reporter activity of the risk allele of SNP Y. Meanwhile, *SIAH2* was reported to have oncogenic effect in aggressive tumors associated with hypoxia such as basal-like phenotype breast cancer[71]. The expression of SIAH2 is increased from in-situ to invasive carcinomas due to the hypoxia resulting from

increase of the metabolic demand as well as the demand of the oxygen supply. It was reported that low levels of SIAH2 were associated with resistance of breast cancer cells to tamoxifen, an inhibitor of estrogen receptor[32]. Considering that *SIAH2* expression is regulated by ER, insufficient under-regulation of *SIAH2* by tamoxifen may render cancer cells to escape from apoptosis or growth suppression. Further investigation of the SNPs in SIAH2 may help the personalized tamoxifen treatment for patients with HRP breast cancer.

The SNP Y was predicted to present within the transcription factor binding site of I, II and III transcription factor using Figure 18 A. In addition to rs73008905 was predicted to be present in within IV and vicinity of I transcription factors binding sites (http://genome.ucsc.edu/). Surprisingly, the transcription factors ER alpha, USF1 and FOXA1 are also candidate regulators of *FGFR2* expression, and variants in intron 2 of *FGFR2* are implicated in HRP breast cancer. Since ER alpha, FOXA1 and USF1 are involved in estrogen receptor signaling as described previously, variations of SNP Y and SNP Z may affect the estrogen-mediated *SIAH2* expression. Future studies on *SIAH2* may disclose the role of *SIAH2* in HRP breast carcinogenesis.

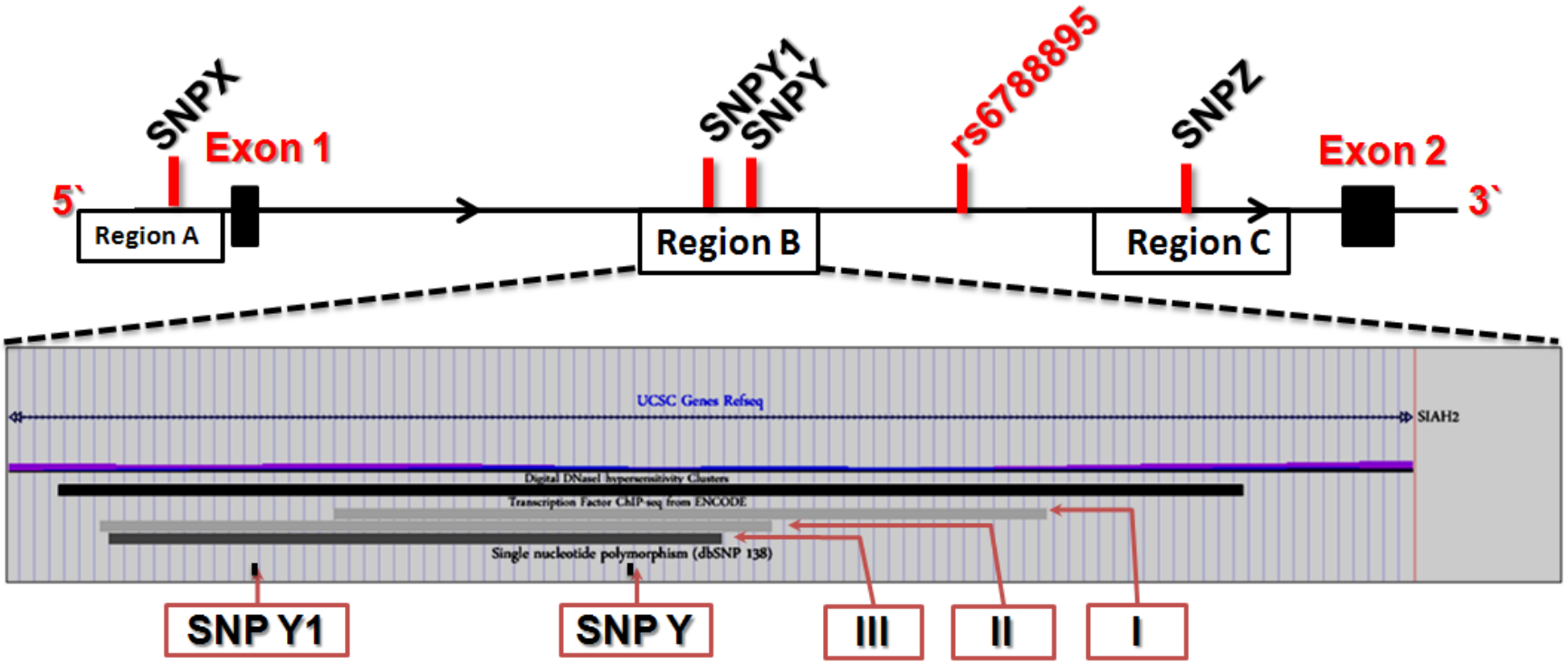


Figure 18A. Position of region B include SNP Y & Y1 and transcription factors I & II & III

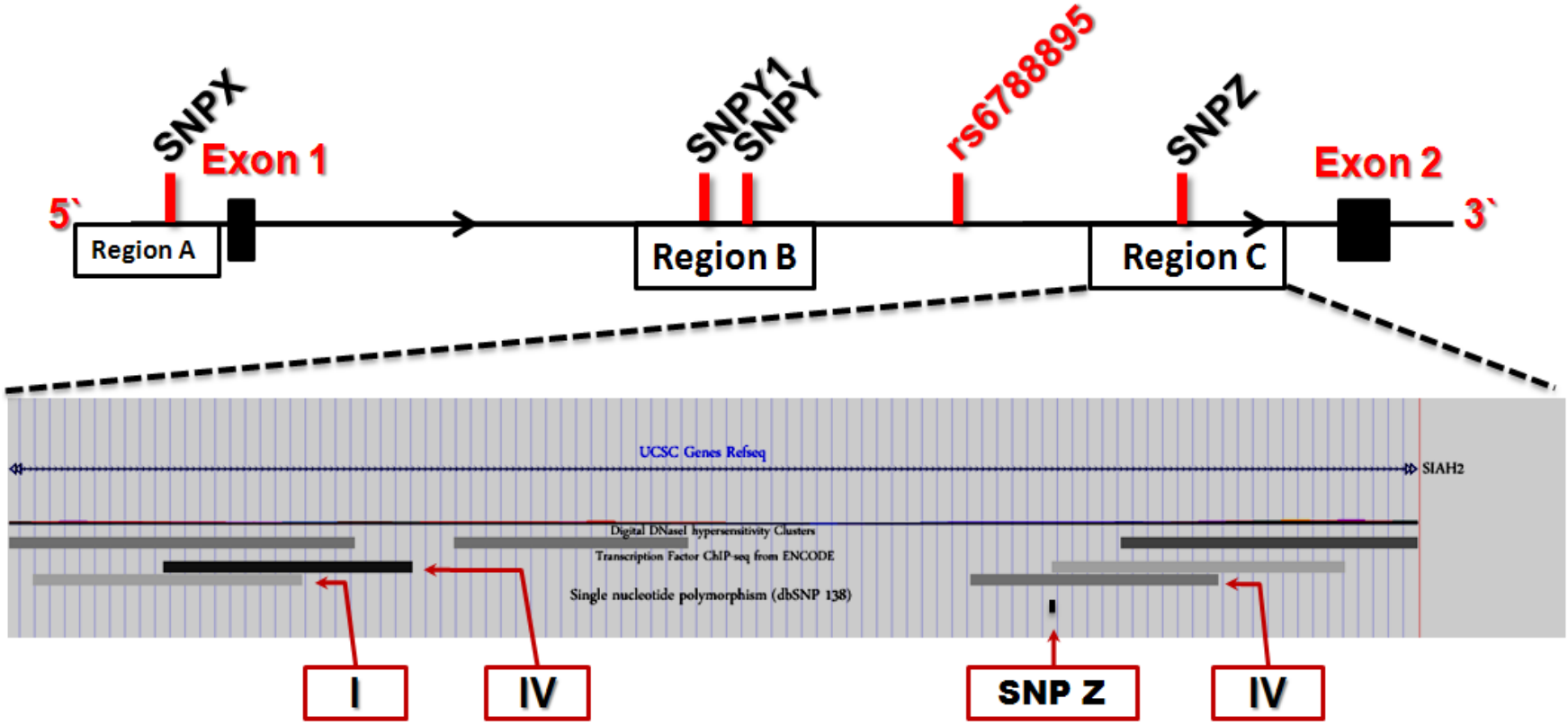


Figure 18 B. Position of region C include SNP Z and transcription factors I & IV

Chapter 6 Conclusion

GWAS is a useful method to identify genetic loci or genes associated with phenotype. The information will be used for various health care services. In my study, I discovered a link between SIAH2 and HRP breast cancer. Additionally, I successfully verified the association between FGFR2 and HRP breast cancer in Japanese population. Since SNP Y revealing significant association with the disease is located within I, II and III binding site these factors may also be involved in breast carcinogenesis. Since SIAH2 is involved in protein degradation via ubiquitin-proteosome system, altered expression of SIAH2 may affect the protein stability of tumor suppressor genes or oncogenes associated with HRP breast carcinogenesis. Therefore, further analysis of SIAH2 function in human carcinogenesis may be useful for the development of new therapeutic or preventive approaches to breast cancer.

Chapter 7 References

1. Stranger, B.E., Stahl, E.A. & Raj, T. Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. *Genetics***187**, 367-383 (2011).

2. Frayling, T.M. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet***8**, 657-662 (2007).

3. Billings, L.K. & Florez, J.C. The genetics of type 2 diabetes: what have we learned from GWAS? *Annals of the New York Academy of Sciences***1212**, 59-77 (2010).

4. McCarthy, M. & Zeggini, E. Genome-wide association studies in type 2 diabetes. *Curr Diab Rep***9**, 164-171 (2009).

5. Cordell, H.J. et al. Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat Genet***45**, 822-824 (2013).

6. Lettre, G. et al. Genome-Wide Association Study of Coronary Heart Disease and Its Risk Factors in 8,090 African Americans: The NHLBI CARe Project. *PLoS Genet***7**, e1001300 (2011).

7. Davies, R.W. et al. A Genome-Wide Association Study for Coronary Artery Disease Identifies a Novel Susceptibility Locus in the Major Histocompatibility Complex. *Circulation: Cardiovascular Genetics***5**, 217-225 (2012).

8. Simon-Sanchez, J. et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet***41**, 1308-1312 (2009).

9. Do, C.B. et al. Web-Based Genome-Wide Association Study Identifies Two Novel Loci and a Substantial Genetic Component for Parkinson's Disease. *PLoS Genet***7**, e1002141 (2011).

10. Franke, A. et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet***42**, 1118-1125 (2010).

11. Rivas, M.A. et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet***43**, 1066-1073 (2011).

12. Dinu, I. et al. SNP-SNP Interactions Discovered by Logic Regression Explain Crohn's Disease Genetics. *PLoS ONE***7**, e43035 (2012).

13. Freedman, M.L. et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet***43**, 513-518 (2011).

14. Nakamura, Y. The BioBank Japan Project. *Clin Adv Hematol Oncol***5**, 696-697 (2007).

15. Cooper, A.P. On the anatomy of the breast 9-15 (2011).

16. Peacock, J. Breast Cancer: Perspectives on disease and illness. 4-12 (2001).

17. Massarweh, S. & Schiff, R. Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk. *Endocrine-Related Cancer***13**, S15-S24 (2006).

18. Xu, W.-H. et al. Relation of FGFR2 Genetic Polymorphisms to the Association Between Oral Contraceptive Use and the Risk of Breast Cancer in Chinese Women. *American Journal of Epidemiology* (2011).

19. Hu, Y.-H., Nagata, C., Shimizu, H., Kaneda, N. & Kashiki, Y. Association of body mass index, physical activity, and reproductive histories with breast cancer: a case-control study in Gifu, Japan. *Breast Cancer Research and Treatment***43**, 65-72 (1997).

20. Tung, H.T. et al. Risk Factors for Breast Cancer in Japan, with Special Attention to Anthropometric Measurements and Reproductive History. *Japanese Journal of Clinical Oncology***29**, 137-146 (1999).

21. Miyoshi, Y., Tanji, Y., Taguchi, T., Tamaki, Y. & Noguchi, S. Association of Serum Estrone Levels with Estrogen Receptor-positive Breast Cancer Risk in Postmenopausal Japanese Women. *Clinical Cancer Research***9**, 2229-2233 (2003).

22. Colditz, G.A. & Rosner, B. Cumulative Risk of Breast Cancer to Age 70 Years According to Risk Factor Status: Data from the Nurses' Health Study. *American Journal of Epidemiology***152**, 950-964 (2000).

23. Ford, D. et al. Genetic Heterogeneity and Penetrance Analysis of the BRCA1 and BRCA2 Genes in Breast Cancer Families. *American journal of human genetics***62**, 676-689 (1998).

24. Bertucci, F. et al. Gene Expression Profiling Shows Medullary Breast Cancer Is a Subgroup of Basal Breast Cancers. *Cancer Research***66**, 4636-4644 (2006).

25. Walsh, T. et al. Spectrum of Mutations in BRCA1, BRCA2, CHEK2, and TP53 in Families at High Risk of Breast Cancer. *JAMA: The Journal of the American Medical Association***295**, 1379-1388 (2006).

26. Hindorff, L.A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences***106**, 9362-9367 (2009).

27. Manolio, T.A. Genomewide Association Studies and Assessment of the Risk of Disease. *New England Journal of Medicine***363**, 166-176 (2010).

28. Kasowski, M. et al. Variation in Transcription Factor Binding Among Humans. *Science***328**, 232-235 (2010).

29. Yamashita, H. et al. Estrogen receptor-positive breast cancer in Japanese women: trends in incidence, characteristics, and prognosis. *Annals of Oncology***22**, 1318-1325 (2011).

30. Prat, A. & Baselga, J. The role of hormonal therapy in the management of hormonal-receptor-positive breast cancer with co-expression of HER2. *Nat Clin Prac Oncol***5**, 531-542 (2008).

31. Ao, A. et al. Response of Estrogen Receptor-Positive Breast Cancer Tumorspheres to Antiestrogen Treatments. *PLoS ONE***6**, e18810 (2011).

32. Jansen, M. et al. Downregulation of SIAH2, an ubiquitin E3 ligase, is associated with resistance to endocrine therapy in breast cancer. *Breast Cancer Research and Treatment***116**, 263-271 (2009).

33. Kiyotani, K. et al. A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. *Human Molecular Genetics* (2011).

34. Turnbull, C. et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet***42**, 504-507 (2010).

35. Sehrawat, B. et al. Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Human Genetics***130**, 529-537 (2011).

36. Higginbotham, K.S.P. et al. A Multistage Association Study Identifies a Breast Cancer Genetic Locus at NCOA7. *Cancer Research***71**, 3881-3888 (2011).

37. Prentice, R.L. et al. Variation in the FGFR2 Gene and the Effects of Postmenopausal Hormone Therapy on Invasive Breast Cancer. *Cancer Epidemiology Biomarkers & Prevention***18**, 3079-3085 (2009).

38. Chen, J. et al. Genetic variants at chromosome 9p21, 10p15 and 10q22 and breast cancer susceptibility in a Chinese population. *Breast Cancer Research and Treatment***132**, 741-746 (2012).

39. Rebbeck, T.R. et al. Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. *Carcinogenesis***30**, 269-274 (2009).

40. Yoshimoto, N. et al. Genetic and environmental predictors, endogenous hormones and growth factors, and risk of estrogen receptor-positive breast cancer in Japanese women. *Cancer Science***102**, 2065-2072 (2011).

41. Hamaguchi, M. et al. Possible Difference in Frequencies of Genetic Polymorphisms of Estrogen Receptor α, Estrogen Metabolism and P53 Genes Between Estrogen Receptor-positive and -negative Breast Cancers. *Japanese Journal of Clinical Oncology***38**, 734-742 (2008).

42. Sueta, A. et al. A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. *Breast Cancer Research and Treatment***132**, 711-721 (2012).

43. Zheng, W. et al. Genetic and Clinical Predictors for Breast Cancer Risk Assessment and Stratification Among Chinese Women. *Journal of the National Cancer Institute***102**, 972-981 (2010).

44. Ghoussaini, M. et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet***44**, 312-318 (2012).

45. Antoniou, A.C. et al. Common alleles at 6q25.1 and 1p11.2 are associated with breast cancer risk for BRCA1 and BRCA2 mutation carriers. *Human Molecular Genetics***20**, 3304-3321 (2011).

46. Stacey, S.N. et al. Ancestry-Shift Refinement Mapping of the <italic>C6orf97-ESR1</italic> Breast Cancer Susceptibility Locus. *PLoS Genet***6**, e1001029 (2010).

47. Stacey, S.N. et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet***40**, 703-706 (2008).

48. Broeks, A. et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Human Molecular Genetics***20**, 3289-3303 (2011).

49. Garcia-Closas, M. & Chanock, S. Genetic Susceptibility Loci for Breast Cancer by Estrogen Receptor Status. *Clinical Cancer Research***14**, 8000-8009 (2008).

50. Fanale, D. et al. Breast cancer genome-wide association studies: there is strength in numbers. *Oncogene* (2011).

51. Raskin, L. et al. FGFR2 Is a Breast Cancer Susceptibility Gene in Jewish and Arab Israeli Populations. *Cancer Epidemiology Biomarkers & Prevention***17**, 1060-1065 (2008).

52. Carroll, J.S. et al. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet***38**, 1289-1297 (2006).

53. deGraffenried, L., Hopp, T., Valente, A., Clark, R. & Fuqua, S.W. Regulation of the Estrogen Receptor α Minimal Promoter by Sp1, USF-1 and ERα. *Breast Cancer Research and Treatment***85**, 111-120 (2004).

54. Ordulu, Z. et al. Disseminated peritoneal leiomyomatosis after laparoscopic supracervical hysterectomy with characteristic molecular cytogenetic findings of uterine leiomyoma. *Genes, Chromosomes and Cancer***49**, 1152-1160 (2010).

55. Cha, P.-C. et al. A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids. *Nat Genet***43**, 447-450 (2011).

56. Frasor, J., Danes, J.M., Funk, C.C. & Katzenellenbogen, B.S. Estrogen down-regulation of the corepressor N-CoR: Mechanism and implications for estrogen derepression of N-CoR-regulated genes. *Proceedings of the National Academy of Sciences of the United States of America***102**, 13153-13157 (2005).

57. Frasor, J. et al. Selective Estrogen Receptor Modulators: Discrimination of Agonistic versus Antagonistic Activities by Gene Expression Profiling in Breast Cancer Cells. *Cancer Research***64**, 1522-1533 (2004).

58. Hu, G. et al. Mammalian homologs of seven in absentia regulate DCC via the ubiquitin–proteasome pathway. *Genes & Development***11**, 2701-2714 (1997).

59. Koren, R. et al. The Expression of DCC Protein in Female Breast Cancer. *Breast Cancer Research and Treatment***80**, 215-220 (2003).

60. Zhen-huan, Z. et al. Nuclear corepressor 1 expression predicts response to first-line endocrine therapy for breast cancer patients on relapse. *Chinese Medical Journal***122**, 1764-1768 (2009).

61. PLEVOVA, P. et al. PML protein expression in hereditary and sporadic breast cancer. *Neoplasma***54**, 263-268 (2007).

62. Yang, X., Hao, Y., Ding, Z., Pater, A. & Tang, S.-C. Differential Expression of Antiapoptotic Gene BAG-1 in Human Breast Normal and Cancer Cell Lines and Tissues. *Clinical Cancer Research***5**, 1816-1822 (1999).

63. Baysal, B.E. A high-resolution integrated map spanning the SDHD gene at 11q23: a 1.1-Mb BAC contig, a partial transcript map and 15 new repeat polymorphisms in a tumour-suppressor region. *European journal of human genetics***9**, 121-129 (2001).

64. van 't Veer, L.J. et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature***415**, 530-536 (2002).

65. Pietschmann, K. et al. Differential regulation of PML–RARα stability by the ubiquitin ligases SIAH1/SIAH2 and TRIAD1. *The International Journal of Biochemistry & Cell Biology***44**, 132-138 (2012).

66. Qi, J., Pellecchia, M. & Ronai, Z.e.A. The Siah2-HIF-FoxA2 axis in prostate cancer – new markers and therapeutic opportunities. *Oncotarget; Vol 1, No 5: September 2010* (2010).

67. Malz, M. et al. Nuclear accumulation of seven in absentia homologue-2 supports motility and proliferation of liver cancer cells. *International Journal of Cancer***131**, 2016-2026 (2012).

68. Perou, C.M. et al. Molecular portraits of human breast tumours. *Nature***406**, 747-752 (2000).

69. Roh, M.S. et al. Gene Expression Profiling of Breast Cancers with Emphasis of β-Catenin Regulation. *J Korean Med Sci***19**, 275-282 (2004).

70. Confalonieri, S. et al. Alterations of ubiquitin ligases in human cancer and their association with the natural history of the tumor. *Oncogene***28**, 2959-2968 (2009).

71. Voss, M. et al. Luminal and basal-like breast cancer cells show increased migration induced by hypoxia, mediated by an autocrine mechanism. *BMC Cancer***11**, 1-12 (2011).

Acknowledgement

This is perhaps the hardest section I have written in my thesis. It will be easy to name all the people that helped to get this done, but it will be tough to thank them enough. I will nonetheless try…

First, I would like to thank Professor. Nakamura Yusuke my supervisor in GWAS project for his constant support and guidance throughout my GWAS project. The discussions, encouragement and critiques made by him throughout the 1st 2 years of my work, were of essence to the progress of this work and I am very grateful to him for letting me have this opportunity to work with him.

I would like to thank my supervisors, Professor Yoichi Furukawa his support and guidance. I am blessed to have such a wonderful mentor who provided me such challenging work that I never seem to get enough of it. I never forget how Prof. Furukawa supported me throughout my thesis with his patience and knowledge whilst allowing me the laboratories to work in my own way. I attribute the level of my Doctors degree to his encouragement and effort and without him this thesis, too, would not have been completed or written. One simply could not wish for a better or friendlier supervisor.

My deep gratitude toward Dr. Tsuneo Ikenoue, Dr. Hitoshi Zembutsu and Dr. Kiyoshi Yamaguchi, for their guidance, encouragement and invaluable discussion that make my research of great achievement.

My sincere appreciation to Dr. Michiaki Kubo, Dr. Atsushi Takahashi, from Center of Genomic Medicine, RIKEN, for the statistical analysis, management of genome-wide analysis.Also, I would like thank Dr. Fuminori Aki, Dr. Koichi Hirata, Dr. YuichiTakatsuka, Dr. Minoru Okazaki, Dr. Shozo Ohsumi, Dr. Takashi Yamakawa, Dr. Mitsunori Sasa, Dr. Toyomasa Katagiri, Dr. Yoshio Mikifor collecting samples and providing constructive

suggestions for breast cancer GWAS project, and all the staff and participants of Biobank Japan Project the Leading Project for Personalized Medicine.

I wish to extend my warmest thanks to all my colleagues in past and present Laboratory of Clinical Genome Research and Laboratory of Molecular Medicine. Special thanks to my friends who provided the scientific and familial environment for me during my work toward my degree, Norihiko Takahashi, Rei Noguchi, TerakadoYumi, Zhu cough, Osugi Tomoyuki, Hatakeyama Seira, and Yamaguchi Yuko.

Finally, especial thanks and apologize for my lovely kids because of busyness of my study.

Last but not least, I am indebted to my parents for their love, patience, moral support and encouragement in seeing me through my study.

My husband, can never thank you enough for whatever you have done for me.