Fig. 17d), reveals a significant preponderance of small *P* values $<10^{-3}$, suggesting many of these nominally significant SNPs likely represent true signals[12]. To further examine the shared genetic risk across our European and Japanese populations we performed polygenic prediction analysis[13] to evaluate whether the aggregate effects of many variants of small effect in the BBJ GWA cohort, could predict affection status in the European GWA cohorts. The BBJ-derived risk scores significantly predicted affection status in the QIMRHCS (R^2 = 0.0064; $P = 6.9 \times 10^{-7}$), OX (R^2 = 0.0057; $P = 9.6 \times 10^{-6}$) and combined QIMRHCS+OX all endometriosis case-control sets (R^2 = 0.0054; $P = 8.8 \times 10^{-11}$). For the individual and combined QIMRHCS and OX case-control sets, the variance explained peaked in the SNP sets with BBJ GWA $P < 0.1$, using all GWA meta-analysis SNPs (Fig. 4a) and after excluding all SNPs within ±2500 kb of the seven implicated SNPs listed in Table 1 (Fig. 4b). Analogously, performing the prediction in reverse, the QIMRHCS+OX-derived risk scores significantly predicted affection status in the BBJ case-control set (R^2 = 0.0106; $P = 3.3 \times 10^{-6}$) (Supplementary Fig. 18 and Supplementary Note).

A gene-based GWA analysis using VEGAS[14], which accounts for gene size and LD between SNPs, revealed 1,184 genes with a combined $P \leq 0.05$ and the top three ranked genes associated with endometriosis to be *WNT4* on 1p36.12 ($P = 5.0 \times 10^{-9}$), *VEZT* on 12q22 ($P = 5.7 \times 10^{-7}$) and *GREB1* on 2p25.1 ($P = 2.5 \times 10^{-5}$) (Supplementary Table 3). In addition to having genome-wide significant SNPs near these three genes, the *WNT4* and *VEZT* genes easily surpass our conservative gene-based significant association threshold of $P \leq 2.85 \times 10^{-6}$ (calculated as $P = 0.05 / 17{,}538$ independent genes). *WNT4* encodes for wingless-type MMTV integration site family, member 4 and is important for the development of the female reproductive tract[15] and steroidogenesis[16]. *VEZT* encodes vezatin, an adherens junction transmembrane protein that is down regulated in gastric cancer[17]. *GREB1* encodes growth regulation by estrogen in breast cancer 1, an early response gene in the estrogen regulation pathway involved in hormone dependent breast cancer cell growth[18]. For the four remaining implicated regions on 2p14, 6p22.3, 7p15.2 and 9p21.3, no genes were significant ($P \leq 1.3 \times 10^{-3}$) after adjusting VEGAS results for testing 37 genes across all seven regions, see Table 2, Supplementary Figs. 3-9 and Supplementary Table 4.

In conclusion, given their high gene-based ranking, proximity to genome-wide significant SNPs, known pathophysiology and reported gene expression (Supplementary Note and Supplementary Fig. 19), the *WNT4*, *VEZT* and *GREB1* genes are strong targets for further studies aimed at understanding the molecular pathogenesis of endometriosis. Our results also suggest that a considerable number of SNPs nominally implicated (e.g. $P < 0.1$) in the European and Japanese GWA cohorts represent true endometriosis risk loci. Moreover, the significant overlap in common polygenic risk for endometriosis indicates genetic risk prediction and future targeted disease therapy may be transferred across these populations.

ONLINE METHODS

GWA samples and phenotyping

Initially, 2,351 surgically-confirmed endometriosis cases were drawn from women recruited by The Queensland Institute of Medical Research (QIMR) study[19] and a further 1,030 cases were obtained from women recruited by the Oxford Endometriosis Gene (OXEGENE) study. Australian controls consisted of 1,870 individuals recruited by QIMR[2] and 1,244 individuals recruited by the Hunter Community Study (HCS)[7]. UK controls encompassed 6,000 individuals provided by the Wellcome Trust Case Control Consortium 2 (WTCCC2). Approval for the studies was obtained from the QIMR Human Ethics Research Committee, the University of Newcastle and Hunter New England Population Health Human Research

Europe PMC Funders Author Manuscripts

Ethics Committees, and the Oxford regional multi-centre and local research ethics committees. Informed consent was obtained from all participants prior to testing[2].

All Japanese GWA case and control samples were obtained from the BioBank Japan (BBJ) at the Institute of Medical Science, the University of Tokyo. A total of 1,423 cases were diagnosed with endometriosis by the following one or more examinations: multiple clinical symptoms, physical examinations, and laparoscopy or imaging tests. We utilized 1,318 female control samples consisting of healthy volunteers from Osaka-Midosuji Rotary Club, Osaka, Japan and women in the Biobank Japan who were registered to have no history of endometriosis. All participants provided written informed consent to this study. This study was approved by the ethical committees at the Institute of Medical Science, the University of Tokyo and Center for Genomic Medicine, RIKEN Yokohama Institute.

GWA genotyping and quality control (QC)

QIMR and OX cases, and QIMR controls were genotyped at deCODE Genetics on Illumina 670-Quad (cases) and 610-Quad (controls) BeadChips (Illumina Inc), respectively. HCS controls were genotyped at the University of Newcastle on 610-Quad BeadChips (Illumina Inc). The WTCCC2 controls were genotyped at the Wellcome Trust Sanger Institute using Illumina HumanHap1M BeadChips. Genotypes for QIMR cases and controls were called with the Illumina BeadStudio software. Standard quality control procedures were applied as outlined previously[20]. Briefly, individuals with call rates <0.95 then SNPs with a mean BeadStudio GenCall score < 0.7, call rates < 0.95, Hardy-Weinberg equilibrium $P < 10^{-6}$, and minor allele frequency (MAF) < 0.01 were excluded. Cryptic relatedness between individuals was identified through a full identity-by-state matrix. Ancestry outliers were identified using data from 11 populations of the HapMap 3 and five Northern European populations genotyped by the GenomeEUtwin consortium, using EIGENSOFT[21,22]. To increase the power of the Australian GWA dataset we ancestrally matched the existing QIMR cases and controls[2] to individuals from the Hunter Community Study (HCS)[7] genotyped on Illumina 610 chips. After stringent quality control, the resulting QIMRHCS GWA cohort consists of 2,262 endometriosis cases and 2,924 controls, increasing the Australian effective sample size by 24%.[2]

Quality control procedures for the OX genotype data resulted in the removal of SNPs with a genotype call rate < 0.99 and/or heterozygosity < 0.31 or > 0.33. Genome-wide IBS was estimated for each pair of individuals and one individual from each duplicate or related pair (IBS > 0.82) was removed. Genotype data were combined with CEU, CHB&JPT and YRI genotype data from HapMap 3 and individuals of non Northern European ancestry were identified using EIGENSOFT and subsequently removed. SNPs with a genotype call rate < 0.95 were removed, and this threshold was increased to 0.99 for SNPs with MAF < 0.05. In addition, SNPs showing a significant a) deviation from HWE ($P < 1 \times 10^{-6}$); b) difference in call rate between 58BC and NBS control groups ($P < 1 \times 10^{-4}$); c) difference in allele/genotype frequency between control groups ($P < 1 \times 10^{-4}$); d) difference in call rate between cases and controls ($P < 1 \times 10^{-4}$) and e) a MAF < 0.01 were removed.[2]

The BBJ cases and controls were genotyped using the Illumina HumanHap550v3 Genotyping BeadChip. Quality control included sample call rate ≥ 0.98, identity-by-state to exclude close relatedness samples and principal component analysis to exclude non-Asian samples. We also performed SNP quality control (call rate of ≥ 0.99 in both cases and controls and Hardy-Weinberg equilibrium test $P \geq 1.0 \times 10^{-6}$ in controls); 460,945 SNPs on all chromosomes passed the quality control filters and were further analyzed.[1]

GWA meta-analysis

For SNPs passing QC, tests of allelic association (--assoc) were performed using PLINK[23] in the separate QIMRHCS, OX and BBJ GWA datasets. The primary meta-analysis of all endometriosis cases versus controls in the QIMRHCS, OX and BBJ GWA data was performed using a fixed-effect (inverse variance-weighted) model, where the effect size estimates, or β-coefficients, are weighted by their estimated standard errors, utilizing the GWAMA software[24].

The threshold of 7.2×10^{-8} for GWA studies of dense SNPs and resequence data[25] proposed by Dudbridge and Gusnanto[26] was utilized to indicate genome-wide *significant* association, while SNPs with $P \leq 10^{-5}$ were considered to show a *suggestive* association [as used in the online 'Catalog of Published Genome-Wide Association Studies'].

Also, given the substantially greater genetic loading of moderate to severe (stage B) endometriosis (rAFS stage III or IV disease) compared to minimal (stage A) endometriosis (rAFS stage I or II disease)[2], a secondary analysis was performed for suggestive SNPs ($P \leq 10^{-5}$); where the association results from QIMRHCS and OX stage B cases versus controls, were meta-analyzed with the BBJ association results. As previously demonstrated[2], the exclusion of minimal endometriosis cases has the potential to enrich true genetic risk effects, even taking into account the reduced sample size.

Consistency of allelic effects across studies was examined utilizing the *Cochran's Q* test[27]. Between-study (effect) heterogeneity was indicated by *Q statistic P* values < 0.1[28]. Meta-analysis of SNPs associated with fixed-effect $P \leq 10^{-5}$ and showing evidence of effect heterogeneity were also analyzed using the recently developed Han and Eskin's random effects model (RE2) implemented in the Metasoft software[29]. In contrast to the conventional DerSimonian-Laird random effects (RE) model[30], the RE2 model *increases* power under heterogeneity[29].

Genotype imputation analysis

In order to assess the impact of variants not present on the Illumina BeadChips and better define the associated regions, we imputed genotypes within ±2500 kb of the most significant genotyped SNP using the full reference panel from the 1000 Genomes project Interim Phase I Haplotypes (2010-11 data freeze, 2011-06 haplotypes). Imputation was performed separately for the QIMRHCS, OX and BBJ GWA datasets with only the overlapping genotyped SNPs within ±2500kb of the most significant genotyped SNP, using the MaCH and minimac programs[31,32] and following the two-step approach outlined in the online 'Minimac: 1000 Genomes Imputation Cookbook'. Analysis of imputed genotype dosage scores was performed using mach2dat[31,32] and PLINK. The quality of imputation was assessed by means of the Rsq statistic. Results for poorly imputed SNPs, defined as having an Rsq < 0.3, were subsequently removed. The results from association analysis of imputed data in the QIMRHCS, OX and BBJ datasets were then combined via meta-analysis of the β-coefficients weighted by their estimated standard errors using GWAMA.

Replication samples and genotyping

Independent of the BBJ GWA case-control cohort, a total of 1,044 cases and 4,017 controls were obtained from the BioBank Japan and utilized for replication. We note that 653 of these 1,044 cases were also utilized in a small GWA study (Adachi et al. 2010) of 696 cases and 825 controls[9]. To utilize all available association data for rs13394619 maximally, given there is incomplete overlap between the Adachi and our replication cases and zero overlap between the controls, we worked with the published results for rs13394619 in Adachi et al

(2010) and the results from comparing our non-overlapping 391 replication cases to our 4,017 replication controls.

The seven SNPs (rs7521902, rs13394619, rs4141819, rs7739264, rs12700667, rs1537377 and rs10859871) reaching genome-wide significance in the GWA meta-analysis were genotyped in the independent Japanese replication cohort using the multiplex PCR-based Invader assay (Third Wave Technologies), as previously described[1].

Replication and total association analyses

Tests of allelic association were performed using PLINK in the independent Japanese replication cohort. Because only the associations in the same direction would be considered as replicated, one-sided *P* values were obtained by halving the standard (two-sided) PLINK *P* values. To determine the total evidence for association, the one-sided replication *P* values were meta-analyzed with the QIMRHCS, OX, BBJ [and Adachi[9] 500K (290 cases and 262 controls) and 6.0 (406 cases and 563 controls) for rs13394619] GWA *P* values using METAL[33]. The *P* values observed in each case-control cohort were converted into a signed Z-score. Z-scores for each allele were combined across samples in a weighted sum, with weights proportional to the square-root of the sample size for each cohort[34]. Given that our cohorts have unequal numbers of cases and controls, we utilized the effective sample size, where $N_{eff} = 4 / (1 / N_{cases} + 1 / N_{controls})$[33]. We also performed meta-analysis of the β-coefficients weighted by their estimated standard errors using GWAMA to estimate the overall odds ratio and 95% CI for the genome-wide significant SNPs.

Polygenic prediction

The aim of the prediction analysis was to evaluate the aggregate effects of many variants of small effect. We summarized variation across nominally associated loci into quantitative scores and related the scores to disease state in independent samples. Although variants of small effect (e.g., genotype relative risk of 1.05) are unlikely to achieve even nominal significance, increasing proportions of "true" effects will be detected at increasingly liberal *P* value thresholds, e.g. $P < 0.1$ (i.e., ~10% of all SNPs), $P < 0.2$, etc. Using such thresholds, we defined large sets of "allele specific scores" in the "discovery" sample of the Japanese BioBank (BBJ) endometriosis case-control set (1,423 cases, 1,318 controls) to generate risk scores for individuals in the "target" sample of the QIMRHCS (2,262 cases, 2,924 controls), OX (919 cases, 5,151 controls) and combined European (QIMRHCS+OX) endometriosis case-control sets (3,181 cases, 8,075 controls). The term risk score is used instead of risk, as it is impossible to differentiate the minority of true risk alleles from the non-associated variants. In the discovery sample, we selected sets of allele specific scores for SNPs with the following levels of significance; $P < 0.01$, $P < 0.05$, $P < 0.1$, $P < 0.2$, $P < 0.3$, $P < 0.4$, $P < 0.5$, $P < 0.6$, $P < 0.7$, $P < 0.8$, $P < 0.9$, $P < 1.0$. For each individual in the target sample, we calculated the number of score alleles that they possessed, each weighted by the log odds ratio from the discovery sample. To assess whether the aggregate scores reflect endometriosis risk, we tested for a higher mean score in cases compared to controls. Logistic regression was used to assess the relationship between target sample disease status and aggregate risk score. Nagelkerke's pseudo R^2 was used to assess the variance explained. Prediction was performed using all 407,632 SNPs overlapping the QIMRHCS, OX and BBJ GWA datasets, and after excluding the 6,163 SNPs within ±2500 kb of the seven implicated SNPs listed in Table 1. We also performed the predictions in reverse, using QIMRHCS +OX-derived risk scores to predict affection status in the BBJ case-control set.

Gene-based association analysis

Gene-based approaches can be more powerful than traditional individual-SNP-based approaches in the presence of allelic heterogeneity. Therefore, utilizing the QIMRHCS, OX

and BBJ GWA data, we performed a genome-wide gene-based association study using VEGAS[14]. Briefly, for the 407,632 overlapping SNPs, the *P* values from the European GWA study (i.e., FE meta-analysis of QIMRHCS and OX GWA data) and the *P* values from the Japanese (BBJ) GWA study were analyzed separately using VEGAS. The VEGAS test incorporates evidence for association from all SNPs across a gene and accounts for gene size (number of SNPs) and LD between SNPs by using simulations from the multivariate normal distribution. The resulting European and Japanese gene-based *P* values were meta-analyzed using Stouffer's Z-score combined p-value method[34]. A total of 17,538 genes (including 50 kb 5′ and 3′ of their transcription start and end site, respectively[14]) contained association results for ≥1 SNP, so a Bonferroni adjusted significance threshold of $P \leq 2.85 \times 10^{-6}$ (0.05 / 17,538) was utilized to indicate genome-wide gene-based *significant* association.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge with appreciation all the women who participated in the QIMR, OX and BBJ studies. We thank Endometriosis Associations for supporting the study recruitment. We also thank the many hospital directors and staff, gynecologists, general practitioners and pathology services in Australia, the UK and the United States who provided assistance with confirmation of diagnoses. We thank Sullivan and Nicolaides Pathology and the Queensland Medical Laboratory Pathology for pro bono collection and delivery of blood samples and other pathology services for assistance with blood collection. The Hunter Community Study team thanks the men and women of the Hunter region who participated in the study.

The QIMR Study was supported by grants from the National Health and Medical Research Council (NHMRC) of Australia (241944, 339462, 389927, 389875, 389891, 389892, 389938, 443036, 442915, 442981, 496610, 496739, 552485 and 552498), the Cooperative Research Centre for Discovery of Genes for Common Human Diseases (CRC), Cerylid Biosciences (Melbourne) and donations from N. Hawkins and S. Hawkins. D.R.N. was supported by the NHMRC Fellowship (339462 and 613674) and the ARC Future Fellowship (FT0991022) schemes. S.M. was supported by NHMRC Career Development Awards (496674, 613705). E.G.H. (631096) and G.W.M. (339446, 619667) were supported by the NHMRC Fellowships Scheme. We thank B. Haddon, D. Smyth, H. Beeby, O. Zheng, B. Chapman and S. Medland for project and database management, sample processing, genotyping and imputation. We thank Brisbane gynecologist D.T. O'Connor for his important role in initiating the early stages of the project and for confirmation of diagnosis and staging of disease from clinical records of many cases, including 251 in these analyses. We are grateful to the many research assistants and interviewers for assistance with the studies contributing to the QIMR collection. The Hunter Community Study was funded by the University of Newcastle, the Gladys M Brawn Fellowship scheme and the Vincent Fairfax Family Foundation in Australia.

The work presented here was supported by a grant from the Wellcome Trust (WT084766/Z/08/Z) and makes use of WTCCC2 control data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of these data is available from http://www.wtccc.org.uk. Funding for the WTCCC project was provided by the Wellcome Trust under awards 076113 and 085475. C.A.A. was supported by a grant from the Wellcome Trust (098051). A.P.M. was supported by a Wellcome Trust Senior Research Fellowship. S.H.K. is supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the Department of Health NIHR Biomedical Research Centres funding scheme. K.T.Z. is supported by a Wellcome Trust Research Career Development Fellowship (WT085235/Z/08/Z). We thank L. Cotton, L. Pope, G. Chalk and G. Farmer (University of Oxford). We also thank P. Koninckx (Leuven, Belgium), M. Sillem (Heidelberg, Germany), C. O'Herlihy and M. Wingfield (Dublin, Ireland), M. Moen (Trondheim, Norway), L. Adamyan (Moscow, Russia), E. McVeigh (Oxford, UK), C. Sutton (Guildford, UK), D. Adamson (Palo Alto, California, USA) and R. Batt (Buffalo, New York, USA) for providing diagnostic confirmation.

We thank the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting our study. This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government.

References

1. Uno S, et al. A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. Nat Genet. 2010; 42:707–10. [PubMed: 20601957]

Europe PMC Funders Author Manuscripts

2. Painter JN, et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. Nat Genet. 2011; 43:51–4. [PubMed: 21151130]
3. Treloar SA, O'Connor DT, O'Connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. Fertil Steril. 1999; 71:701–710. [PubMed: 10202882]
4. Montgomery GW, et al. The search for genes contributing to endometriosis risk. Hum Reprod Update. 2008; 14:447–57. [PubMed: 18535005]
5. Gao X, et al. Economic burden of endometriosis. Fertil Steril. 2006; 86:1561–72. [PubMed: 17056043]
6. Visscher PM, Brown MA, McCarthy MI, Yang J. Five Years of GWAS Discovery. Am J Hum Genet. 2012; 90:7–24. [PubMed: 22243964]
7. McEvoy M, et al. Cohort profile: The Hunter Community Study. Int J Epidemiol. 2010; 39:1452–63. [PubMed: 20056765]
8. American Society for Reproductive Medicine. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. Fertil Steril. 1997; 67:817–21. [PubMed: 9130884]
9. Adachi S, et al. Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population. J Hum Genet. 2010; 55:816–21. [PubMed: 20844546]
10. Goumenou AG, Arvanitis DA, Matalliotakis IM, Koumantakis EE, Spandidos DA. Loss of heterozygosity in adenomyosis on hMSH2, hMLH1, p16Ink4 and GALT loci. Int J Mol Med. 2000; 6:667–71. [PubMed: 11078826]
11. Martini M, et al. Possible involvement of hMLH1, p16(INK4a) and PTEN in the malignant transformation of endometriosis. Int J Cancer. 2002; 102:398–406. [PubMed: 12402310]
12. Yang J, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 2011; 19:807–12. [PubMed: 21407268]
13. Purcell SM, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–52. [PubMed: 19571811]
14. Liu JZ, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010; 87:139–45. [PubMed: 20598278]
15. Vainio S, Heikkila M, Kispert A, Chin N, McMahon AP. Female development in mammals is regulated by Wnt-4 signalling. Nature. 1999; 397:405–9. [PubMed: 9989404]
16. Guo X, et al. Down-regulation of VEZT gene expression in human gastric cancer involves promoter methylation and miR-43c. Biochem Biophys Res Commun. 2011; 404:622–7. [PubMed: 21156161]
17. Boyer A, et al. WNT4 is required for normal ovarian follicle development and female fertility. Faseb J. 2010; 24:3010–25. [PubMed: 20371632]
18. Rae JM, et al. GREB 1 is a critical regulator of hormone dependent breast cancer growth. Breast Cancer Res Treat. 2005; 92:141–9. [PubMed: 15986123]
19. Treloar SA, et al. Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. Am J Hum Genet. 2005; 77:365–376. [PubMed: 16080113]
20. Medland SE, et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet. 2009; 85:750–5. [PubMed: 19896111]
21. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]
22. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
23. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
24. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC bioinformatics. 2010; 11:288. [PubMed: 20509871]
25. Bajpai AK, et al. MGEx-Udb: a mammalian uterus database for expression-based cataloguing of genes across conditions, including endometriosis and cervical cancer. PLoS One. 2012; 7:e36776. [PubMed: 22606288]

Europe PMC Funders Author Manuscripts

26. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol. 2008; 32:227–34. [PubMed: 18300295]
27. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954; 10:101–129.
28. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. PLoS One. 2007; 2:e841. [PubMed: 17786212]
29. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet. 2011; 88:586–98. [PubMed: 21565292]
30. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; 7:177–88. [PubMed: 3802833]
31. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009; 10:387–406. [PubMed: 19715440]
32. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–34. [PubMed: 21058334]
33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]
34. Stouffer, SA.; Suchman, EA.; DeVinney, LC.; Star, SA.; Williams, RMJ. Adjustment During Army Life. Princeton University Press; Princeton, NJ: 1949.

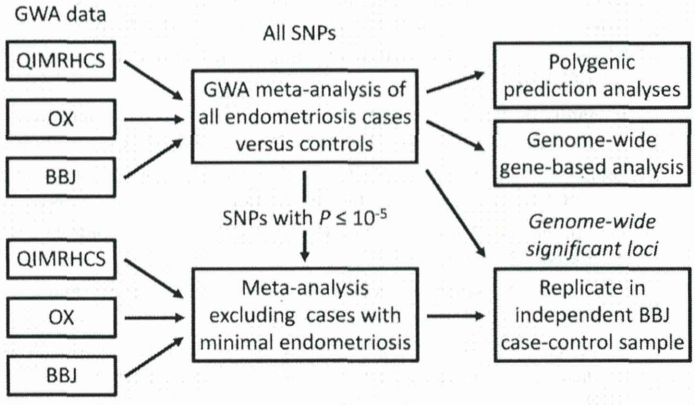

Figure 1.
Study design.

Europe PMC Funders Author Manuscripts

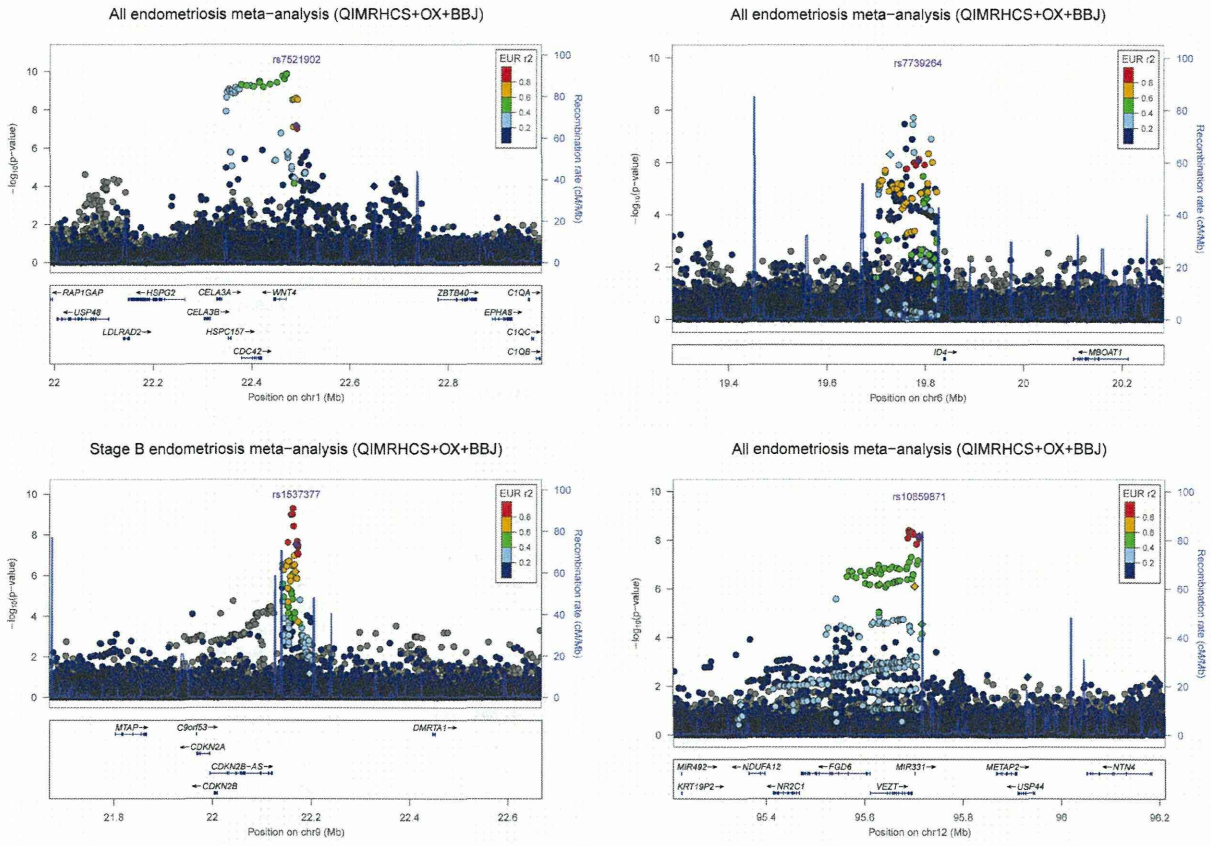

Figure 2.

Evidence for association with endometriosis from the QIMRHCS+OX+BBJ GWA meta-analysis across the 1p36.12 (**a**), 6p22.3 (**b**), 9p21.3 (**c**) and 12q22 (**d**) regions following imputation using the 1000 Genomes Project reference panel. Diamond and circle symbols represent genotyped and imputed SNPs, respectively. The most significant genotyped SNP is represented by a purple diamond. All other SNPs are color coded according to the strength of LD with the top genotyped SNP (as measured by r^2 in the European 1000 Genomes data).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

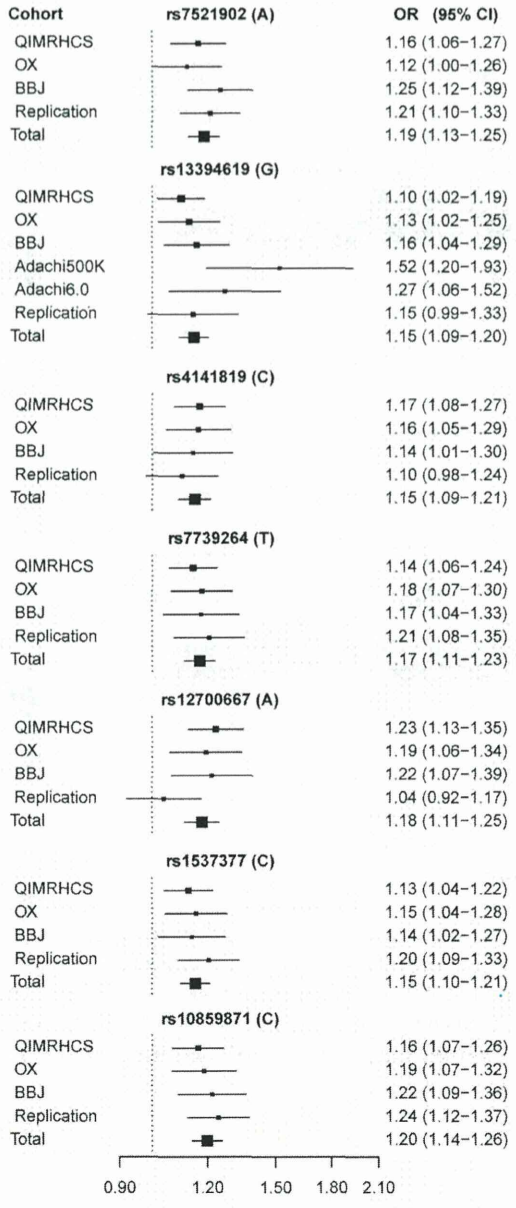

Figure 3.

Forest plots of risk allele effects for the seven genome-wide significant SNP loci in the individual and total endometriosis case-control cohorts.

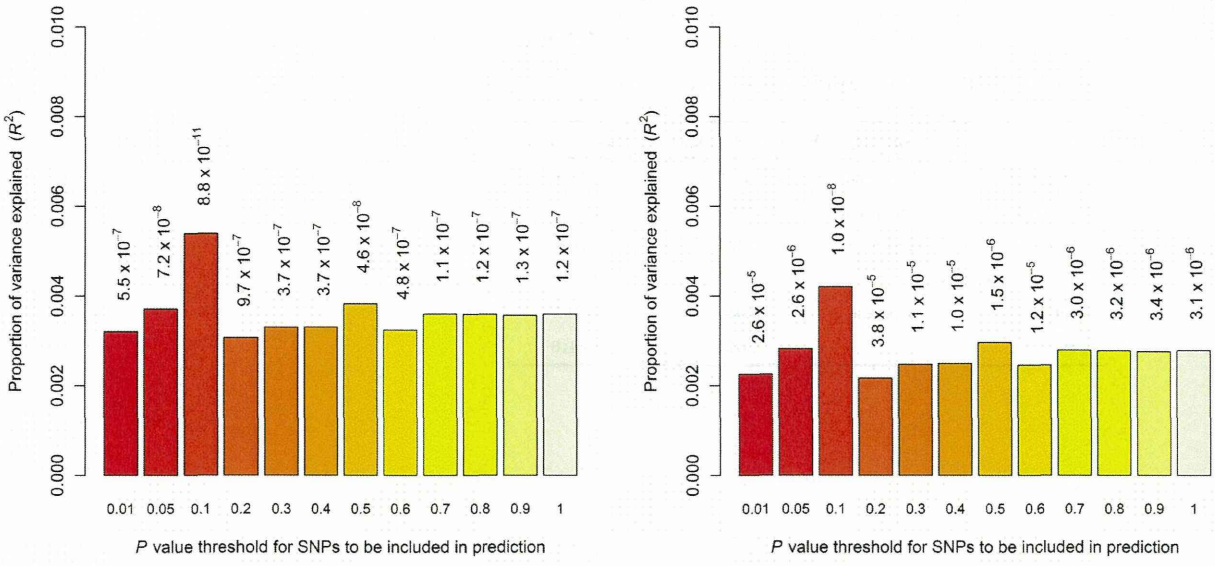

Figure 4.

Allele-specific score prediction for endometriosis, using the BBJ population as the discovery dataset and the QIMRHCS+OX population as the target dataset. The variance explained in the target dataset on the basis of allele-specific scores derived in the discovery dataset for twelve significance thresholds ($P < 0.01$, $P < 0.05$, $P < 0.1$, $P < 0.2$, $P < 0.3$, $P < 0.4$, $P < 0.5$, $P < 0.6$, $P < 0.7$, $P < 0.8$, $P < 0.9$, $P < 1.0$, plotted left to right). The y-axis indicates Nagelkerke's pseudo R^2 representing the proportion of variance explained. The number above each bar is the Pvalue for the target dataset prediction analysis (i.e. R^2 significance). Prediction was performed using all GWA meta-analysis SNPs (**a**) and after excluding all SNPs within ±2500 kb of the seven implicated SNPs listed in Table 1 (**b**). These figures show that the results were not driven by a few highly associated regions, indicating a substantial number of common variants underlie endometriosis risk.

Europe PMC Funders Author Manuscripts

Table 1

Summary of the endometriosis case-control cohorts

Cohort	Ancestry	No. of cases (stage B)	No. of controls
QIMRHCS GWA	European	2,262 (905)	2,924
OX GWA	European	919 (452)	5,151
BBJ GWA	Japanese	1,423	1,318
GWA meta-analysis		4,604	9,393
Replication	Japanese	1,044	4,017
Total		5,648	13,410

Europe PMC Funders Author Manuscripts

Table 2

Summary of the GWA and replication study results for the seven genome-wide significant loci

					GWA												
					QIMRHCS		OX		BBJ		Meta-analysis		Replication			Total	
Chr	SNP	Position	RA	OA	RAF_{case}	$RAF_{control}$	RAF_{case}	$RAF_{control}$	RAF_{case}	$RAF_{control}$	P_{all}	P_{stageB}	RAF_{case}	$RAF_{control}$	P	P_{all}	P_{stageB}
1	rs7521902	22490724	A	C	0.265	0.236	0.259	0.238	0.570	0.514	4.6×10^{-8}	2.3×10^{-9}	0.568	0.521	6.5×10^{-5}	3.2×10^{-11}	7.6×10^{-13}
2	rs13394619 *	11727507	G	A	0.538	0.514	0.551	0.521	0.485	0.449	6.1×10^{-8}	7.0×10^{-8}	0.489	0.455	3.5×10^{-2}	6.1×10^{-9}	6.7×10^{-9}
2	rs4141819	67864675	C	T	0.331	0.298	0.343	0.309	0.226	0.203	4.0×10^{-7}	6.5×10^{-8}	0.220	0.203	5.1×10^{-2}	8.5×10^{-8}	4.1×10^{-8}
6	rs7739264	19785588	T	C	0.545	0.512	0.556	0.515	0.772	0.742	1.3×10^{-7}	5.8×10^{-8}	0.778	0.744	6.9×10^{-4}	3.6×10^{-10}	2.1×10^{-10}
7	rs12700667	25901639	A	G	0.769	0.730	0.776	0.744	0.221	0.189	9.3×10^{-10}	3.8×10^{-11}	0.197	0.191	2.6×10^{-1}	3.6×10^{-9}	1.1×10^{-9}
9	rs1537377	22169700	C	T	0.424	0.395	0.436	0.401	0.410	0.379	2.5×10^{-6}	1.0×10^{-8}	0.402	0.359	1.3×10^{-4}	2.4×10^{-9}	5.8×10^{-12}
12	rs10859871	95711876	C	T	0.332	0.299	0.332	0.295	0.373	0.328	5.5×10^{-9}	3.7×10^{-7}	0.377	0.328	1.1×10^{-5}	5.1×10^{-13}	2.6×10^{-11}

Chr = Chromosome, Position = GRCh37 (hg19) bp position, RA = risk allele, OA = other allele, RAF = risk allele frequency.

* = GWA meta-analysis and total *P* values for rs13394619 include results published in Adachi et al. (2010), consisting of $P = 6.1 \times 10^{-4}$ ($RAF_{case} = 0.517$, $RAF_{control} = 0.414$) and $P = 1.0 \times 10\text{-}2$ ($RAF_{case} = 0.488$, $RAF_{control} = 0.429$) obtained in their 500K and 6.0 case-control cohorts, respectively. P_{all} includes all available endometriosis cases. P_{stageB} excludes unknown and minimal (rAFS I-II) endometriosis stage cases where detailed stage data was available.

Journal of Human Genetics (2012), 1–6

www.nature.com/jhg

ORIGINAL ARTICLE

A genome-wide association study identifies a genetic variant in the *SIAH2* locus associated with hormonal receptor-positive breast cancer in Japanese

Seham Elgazzar[1], Hitoshi Zembutsu[1], Atsushi Takahashi[2], Michiaki Kubo[3], Fuminori Aki[4], Koichi Hirata[5], Yuichi Takatsuka[6], Minoru Okazaki[7], Shozo Ohsumi[8], Takashi Yamakawa[9], Mitsunori Sasa[10], Toyomasa Katagiri[11], Yoshio Miki[12] and Yusuke Nakamura[1]

In Japan, breast cancer is the most common cancer among women and the second leading cause of cancer death among women worldwide. To identify genetic variants associated with the disease susceptibility, we 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. We selected 33 single-nucleotide polymorphisms (SNPs) with suggestive associations in GWAS (*P*-value of $<1\times10^{-4}$) 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. We identified significant association of the disease with a SNP rs6788895 ($P_{combined}$ of 9.43×10^{-8} with odds ratio (OR) of 1.22) in the *SIAH2* (*intron of seven in absentia homolog 2*) gene on chromosome 3q25.1 where the involvement in estrogen-dependent diseases was suggested. In addition, rs3750817 in intron 2 of the *fibroblast growth factor receptor 2* gene, which was reported to be associated with breast cancer susceptibility, was significantly replicated with $P_{combined}$ of 8.47×10^{-8} with OR = 1.22. Our results suggest a novel susceptibility locus on chromosome 3q25.1 for a HRP breast cancer.
Journal of Human Genetics advance online publication, 6 September 2012; doi:10.1038/jhg.2012.108

Keywords: breast cancer in the Japanese population; *FGFR2* gene; GWAS; hormonal receptor-positive breast cancer; *SIAH2* gene; 3q25.1 locus; 10q26 locus

INTRODUCTION

Nearly 70% of breast cancer is known to be hormone dependent, as estrogen and progesterone have key roles both in the development and progression of the disease.[1,2] 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.[3–5] 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.[6] In Japan, breast cancer is the most common cancer among women and its incidence has been doubled in both pre- and postmenopausal women in the last 20 years, mainly as an estrogen receptor-positive subgroup.[7] Although hormone therapy and radiotherapy are effective, cancer cells often become resistant to these treatments; nearly half of estrogen receptor-positive breast cancer patients at an advanced stage suffer from recurrence[8–10] and only one-third of hormonal receptor-positive (HRP) patients with metastatic disease respond to radiotherapy.[11] Therefore, new therapeutic options for the disease are eagerly awaited.

The aim of this study is to identify the genetic factors susceptible to HRP breast cancer in the Japanese population and should facilitate the development of novel approaches to prevent and/or treat breast cancer.

MATERIALS AND METHODS

Samples

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 the BioBank Japan, which begun in 2003 with the goal of collecting DNA and serum

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; [2]Laboratory for Statistical Analysis, RIKEN Center for Genomic Medicine, Yokohama, Japan; [3]Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine, Yokohama, Japan; [4]Department of Surgery, Itoh Surgery and Breast Clinic, Kochi, Japan; [5]First Department of Surgery, Sapporo Medical University, Sapporo, Japan; [6]Department of Breast Surgery, Kansai Rosai Hospital, Hyogo, Japan; [7]Department of Surgery, Sapporo Breast Surgical Clinic, Sapporo, Japan; [8]Department of Breast Oncology, Shikoku Cancer Center, Ehime, Japan; [9]Department of Surgery, Yamakawa Breast Clinic, Kochi, Japan; [10]Department of Surgery, Tokushima Breast Care Clinic, Tokushima, Japan; [11]Division of Genome Medicine, Institute for Genome Research, The University of Tokushima, Tokushima, Japan and [12]Genome Center, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan
Correspondence: Dr Y Nakamura, Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.
E-mail: yusuke@ims.u-tokyo.ac.jp
Received 3 April 2012; revised 5 August 2012; accepted 8 August 2012

Table 1 Characterization of samples used in hormonal receptor-positive breast cancer

	GWAS	*Replication*
Case		
Number of subjects	1086	1653
Mean age at interview (±s.d.)	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	1034	1446
Body mass index prime	1.08	1.03
Platform	Illumina HumanHap 610K	Invader assay
Source	BioBank Japan Collaborative hospitals[a]	BioBank Japan Collaborative hospitals[a]
Control		
Number of subjects	1816	2797
Mean age at interview (±s.d.)	61.3 (12.6)	65.9 (13.2)
Body mass index prime	1.06	1.02
Platform	Illumina HumanHap 610K	Illumina HumanHap 610K
Source	BioBank Japan	BioBank Japan
Diseases in control[b]	MRC healthy volunteer Hepatitis B Keloid Drug eruption Pulmonary tuberculosis Peripheral artery disease Arrhythmias Stroke Myocardial infarction	Rheumatoid arthritis Amyotrophic lateral sclerosis Liver cirrhosis

Abbreviations: DCIS, ductal carcinoma in situ; GWAS, genome-wide association study.
[a]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.
[b]The control groups from BioBank Japan consisted of female individuals without cancer also without any disease related to breast cancer.

samples, along with clinical information from 300 000 individuals who were diagnosed to have any of 47 different diseases from a collaborative network of 66 hospitals in Japan. All cases were diagnosed to have a HRP breast cancer by the following examinations: examination of breast tissue (biopsy or cytology), estrogen receptor and progesterone receptor positivities were evaluated by immunohistochemistry. For the genome-wide association study (GWAS) study, 1086 subjects with HRP breast cancer had been selected as cases (Table 1); 846 samples were collected from the BioBank Japan and the remaining 240 samples were collected from collaborative hospitals. Controls for the GWAS consisted of 1816 females including 231 healthy volunteers from the Midosuji Rotary Club, Osaka, Japan. In addition, we also used genome-wide screening data of 1585 female samples for 8 diseases registered in the BioBank Japan (Table 1). In the replication stage, 1547 cases were obtained from BioBank Japan and 105 cases from the collaborative hospitals. In all, 2797 female controls were registered in BioBank Japan and were genotyped in GWAS for other diseases (Table 1).

For re-sequencing analysis, we selected 2266 cases with HRP breast cancer from the BioBank Japan. We used 497 female controls with 4 diseases (hepatitis B, keloid, drug eruption and pulmonary tuberculosis) from the BioBank Japan as well as 231 healthy volunteers from the Midosuji Rotary Club, Osaka, Japan. All participating subjects provided written informed consent to participate in the study in accordance with the process approved by Ethical Committee at each of the Institute of Medical Science of the University of Tokyo and the Center for Genomic Medicine of RIKEN.

SNP genotyping

For the first stage, we genotyped 1086 female individuals with HRP breast cancer and 1816 female controls using the Illumina HumanHap 610 Genotyping BeadChip (Illumina, San Diego, CA, USA). We applied our single-nucleotide polymorphism (SNP) quality control standard (call rate of $\geqslant 0.99$ in both cases and controls, and Hardy–Weinberg equilibrium test of $P < 1.0 \times 10^{-6}$ in controls). 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. All control samples for the replication stage were genotyped using the Illumina HumanHap 610 BeadChip (female samples of three diseases as controls). All cluster plots were checked by visual inspection by trained personnel, and SNPs with ambiguous calls were excluded. For cases in the replication study, we used the multiplex PCR-based Invader assay (Third Wave Technologies).[12] In addition, 22 variations resulted from re-sequencing analysis were selected and genotyped in 2266 cases and 728 female controls also using the multiplex PCR-based Invader assay (Third Wave Technologies, Madison, WI, USA).

Statistical analysis

Associations of SNPs were tested by employing the Cochran–Armitage trend test in both the GWA and replication stages. 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) by Bonferroni correction. Odds ratios (ORs) and confidence intervals were calculated using the non-susceptible allele as a reference. Heterogeneity between the GWAS and replication sets was examined using the Breslow–Day test. The genomic inflation factor (λGC) was calculated from the median of the Cochran–Armitage trend test statistics. 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 (see URLs), and the Manhattan plot was generated using Haploview v4.1 (see URLs). 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. *In silico* prediction of functional consequences of SNP was done by the use of the SNP info web server (see URLs). (Haploview software was used to analyze linkage disequilibrium (LD) values, visualize haplotype.)

Imputation

Imputation was performed by referring to the genotype data of Japanese (JPT) individuals as deposited in the Phase II HapMap database using MACH v1.0 (see URLs). 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. In the process of imputation, 50 Markov chain iterations were implemented. Imputed SNPs with an imputation quality score of $r^2 < 0.3$ were excluded from the subsequent analysis.

Re-sequencing analysis

Initially, we carried out SNP discovery by using DNA samples of 96 cases with HRP breast cancer. We designed 98 sets of primers (Supplementary Table 1) using the genomic sequence information from UCSC Genome Bioinformatics data base (NM_005067) to amplify the 22 353 bps (two exons, one intron, 5′-UTR and 3′-UTR) of the genomic region corresponding to the *SIAH2* (*intron of seven in absentia homolog 2*) gene. For each of the 96 DNA samples, PCRs were performed by using GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, USA). We performed direct sequencing of the PCR products with the 96-capillary 3730 × l DNA Analyzer (Applied Biosystems) with Big Dye Terminators (Applied Biosystems) according to standard protocols. All amplified fragments were sequenced by two pairs of sequencing primers. Then SNPs were detected by Sequecher software v4.8 (Gene Codes, Ann Arbor, MI, USA).

RESULTS

To identify genetic variants susceptible to HRP breast cancer in the Japanese population, we performed a GWAS using 1086 female patients and 1816 female controls with Illumina HumanHap 610k BeadChip (Table 1). After the 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 (Supplementary Figure 1S). A quantile–quantile plot for this GWAS is shown in Supplementary Figure 2S. The genomic inflation factor (λGC) 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, 46 SNPs in various chromosomes showed suggestive association (P-values $< 1 \times 10^{-4}$) as illustrated in Figure 1.

Among these 46 SNPs, we excluded SNPs possessing strong LD ($r^2 > 0.8$) and selected 33 SNPs for replication analysis as well as 4 additional SNPs that were previously reported their association with breast cancer and showed P-value of $< 1.0 \times 10^{-2}$ in GWAS analysis, using an independent set of 1653 female patients and 2797 female controls. Among 37 SNPs analyzed in the replication study, an SNP rs6788895 was successfully replicated with the P-value of $< 1.35 \times 10^{-3}$ even after the Bonferroni correction (0.05/37) as shown in Table 2 and Supplementary Table 2S. 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 ($P_{heterogeneity} = 2.33 \times 10^{-01}$).

The SNP rs6788895 was further examined its association 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 ($P_{combined} = 3.61 \times 10^{-07}$, 6.78×10^{-06}, OR = 1.23, 1.21, respectively) although they did not reach to the genome-wide significant level (Supplementary Table 3S). Imputation analysis of this locus identified nine additional SNPs in strong LD (r^2 of > 0.8) that showed similar levels of association with rs6788895 (Figure 2a). The subsequent logistic regression analysis revealed no significant association of these nine SNPs when we accounted the effect of SNP rs6788895. The haplotype analysis found no haplotype revealing stronger association than the single SNP (Supplementary Table 4S). Although *in silico* prediction of the functional effect of rs6788895 identified no possible biological effect, one SNP rs2018246 showing strong LD with rs6788895 ($r^2 = 0.94$), which was located about 0.7 kb upstream from the transcription initiation site of *SIAH2*, was indicated to be present within the binding site of multiple transcription factors such as STAT1, LEF1, PAX2, which were reported to have some implication to breast cancer.[13–16] The re-sequencing of 22 353 bps corresponding to the *SIAH2* gene identified 10 novel genetic variations in addition to 37 genetic variations reported previously. We further genotyped 22 of the 47 variations after the exclusion of SNPs showing strong LD with the marker SNP (r^2 of > 0.8). As a result, we identified no genetic variant showing significant association in HRP breast cancer (Supplementary Table 5S and Supplementary Table 6S)

Furthermore, we examined the association of 37 previously reported SNPs with the HRP breast cancer[17–26] using our sample sets (Supplementary Table 7S) and found very moderate association of four genetic variants, rs1292011, rs3803662, rs2981579 and rs3750817, with HRP breast cancer in the GWAS phase ($P_{GWAS} = 5.89 \times 10^{-02}$, 6.95×10^{-03}, 8.68×10^{-04} and 5.03×10^{-04}, respectively). Further analysis of these four SNPs identified significant

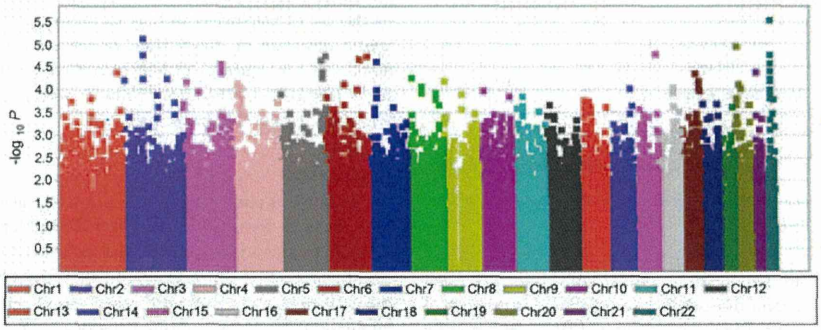

Figure 1 Manhattan plot for the genome-wide association study (GWAS) of hormonal receptor-positive breast cancer indicating −log10*P* of the Cochran–Armitage trend test for 453 627 single-nucleotide polymorphisms (SNPs) plotted against their respective positions on each chromosome.

Table 2 Association of SNP rs6788895 on chromosome 3q25.1 with hormonal receptor-positive breast cancer

					Case				*Control*							
Chr.	*Chrloc.*	*SNP*	*RA*	*Stage*	*11*	*12*	*22*	*RAF*	*11*	*12*	*22*	*RAF*	P_{assoc}[a]	*OR*	*(95% CI)*	*Phet*[b]
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	1265	1195	0.65	5.77E-04	1.18	(1.07–1.29)	2.33E-01
				Combined	270	1150	1310	0.69	579	2097	1937	0.65	9.43E-08	1.22	(1.13–1.31)	

Abbreviations: Chr., chromosome; chrloc., chromosomal location (bp); CI, confidence interval; GWAS, genome-wide association study; OR, odds ratio (calculated based on the risk allele); RA, risk allele; RAF, risk allele frequency; Rep, replication; SNP, single-nucleotide polymorphism; 11, homozygous non-risk genotype; 12, heterozygous genotype; 22, homozygous risk genotype.
[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.

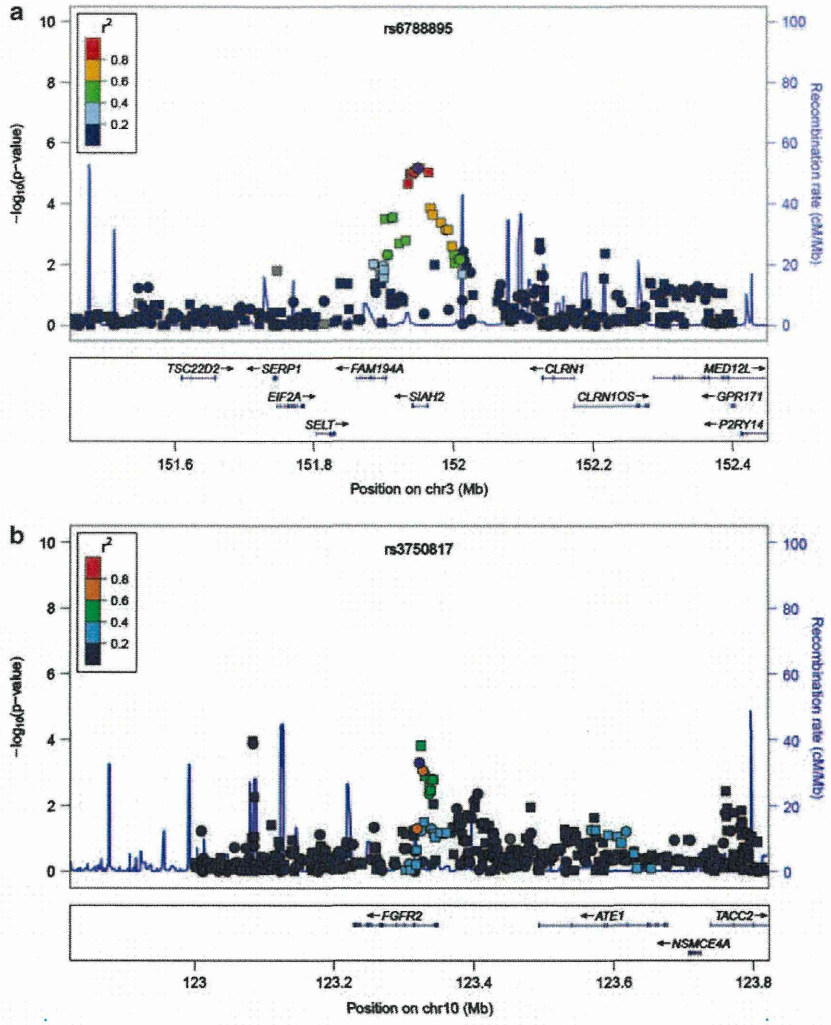

Figure 2 (**a**) Regional association plots of the locus associated with hormonal receptor-positive breast cancer on chromosomes 3q25.1 (*intron of seven in absentia homolog 2* (*SIAH2*)). (**b**) Regional association plots of the locus associated with hormonal receptor-positive breast cancer on chromosomes 10q26 (*fibroblast growth factor receptor 2* (*FGFR2*)). For each plot, $-\log_{10}P$ 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.

replication of two SNPs, rs3750817 ($P_{\text{replication}} = 5.39 \times 10^{-5}$, OR $= 1.22$) and rs2981579 ($P_{\text{replication}} = 1.21 \times 10^{-3}$, OR $= 1.20$). Both SNPs are located within intron 2 of the fibroblast growth factor receptor 2 *(FGFR2)* genes. The combined analysis of the GWAS and replication phases of rs3750817 revealed strong association with $P_{\text{combined}} = 8.47 \times 10^{-08}$ (OR $= 1.22$) and that of rs2981579 was 1.77×10^{-06} (OR $= 1.20$) (Table 3). Imputation analysis of this locus identified three additional SNPs, rs9420318, rs11199914 and rs10736303 that showed similar levels of association with rs3750817 (Figure 2b).

DISCUSSION

We reported here GWA and replication studies using a total of 2730 female breast cancer cases and 4613 female controls in the Japanese population to identify common genetic variants susceptible to the HRP breast cancer. The SNP rs6788895 located in the intronic region of the *SIAH2* gene on chromosome 3q25.1 revealed a significant association with the HRP breast cancer (P_{combined} of 9.43×10^{-08} with OR of 1.22, 95% confidence interval of 1.13–1.31). We further examined the association of rs6788895 with the subgroups of breast cancer. The analysis of two histological subgroups, an invasive papilloductal breast cancer group and a HER2-negative breast cancer group, indicated suggestive associations with P_{combined} of 3.61×10^{-07} (OR $= 1.24$) and with P_{combined} of 6.78×10^{-06} (OR $= 1.21$), respectively (Supplementary Table 3S). However, rs6788895 showed no association in the GWAS with the hormonal receptor-negative group (P_{trend} of 1.03×10^{-01}) or with the HER2-positive breast cancer group (P_{trend} of 1.15×10^{-01}).

For further characterization of the chromosome 3q25.1 locus, we imputed genotypes of SNPs that were not genotyped in the GWAS

Table 3 rs2981579 and rs3750817 in different population

SNPs	Minor/major allele	MAF	OR	P-trend	Population
rs2981579 (FGFR2)	A/G	0.42	1.43	3.60×10^{-31}	UK[20]
rs2981579	A/G	0.44	1.31	2.60×10^{-09}	American[25]
rs2981579	A/G	0.47	1.20	1.77×0^{-06}	Japanese
rs3750817 (FGFR2)	T/C	0.49	1.22	8.47×10^{-08}	Japanese
rs3750817	T/C	0.37	0.78	8.20×10^{-08}	American[25]

Abbreviations: FGFR2, fibroblast growth factor receptor 2; 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); SNP, single-nucleotide polymorphism.

and then examined their associations with HRP breast cancer, but found no SNP showing stronger association than the marker SNP rs6788895 although several SNPs having strong LD with rs6788895 ($r^2 > 0.8$) showed similar levels of associations (Figure 2a). Previous reports implicated possible roles of *SIAH2* in breast carcinogenesis and described that SIAH2 expression was highly associated with estrogen receptor levels.[9,27–29] In addition, SIAH2 protein was indicated to have an essential role in the hypoxic response by regulating the hypoxia-inducible factor-α.[30]

Moreover, *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 some relations to breast cancer by different mechanisms.[31–35] Recent genetic studies showed that the chromosome 3q25.1 region might have a critical role in some estrogen-dependent diseases such as development of peritoneal leioyomatosis.[36,37]

We also examined the association of previously reported loci with the breast cancer[17–26] using our sample sets and found very moderate association of four genetic variants in our GWAS. Further analysis of these four SNPs identified significant replication of two SNPs, rs3750817 and rs2981579 ($P_{combined} = 8.47 \times 10^{-8}$ and 1.77×10^{-06} with OR = 1.22 and OR = 1.20, respectively). A T allele for rs3750817 is a protective allele for both Japanese and American populations with comparable ORs (Table 3).

For characterization of the chromosome 10q26 locus, we imputed genotypes of SNPs that were not genotyped in the GWAS, and examined the associations of these SNPs with HRP breast cancer. As a result, three additional SNPs, rs9420318, rs11199914 and rs10736303 were found to have similar levels of association with rs3750817 (Figure 2b). The most strongly associated SNPs are located in intron 2 of the *FGFR* gene. The intron 2 region contains a highly conserved region and possess the transcription factor binding sites possibly related to the estrogen receptor signaling pathway.[38] FGFR2 encodes a receptor tyrosine kinase and has an important role in human mammary epithelial-cell transformation,[39,40] suggesting that FGFR2 is a good candidate for breast cancer susceptibility. Subsequent functional analyses are thus essential to pinpoint the causal variants and genes associated with HRP breast cancer. In addition, because breast cancer is multi factorial disease, we could not exclude the possibility that some subjects with undiagnosed early stage of cancers or undiagnosed hormonal-dependent diseases or subject have diseases related to breast cancer might have been included as controls. Hence, this study might not have sufficient power to detect SNPs having very modest effects on susceptibility to HRP breast cancer. In conclusion, our findings, the verification of the association of the FGFR2 to the risk of breast cancer in the Japanese population and the novel identification of significant association of genetic variations in the *SIAH2* gene, should contribute to the better understanding of the susceptibility to HRP breast cancer.

URLS

The Leading Project for Personalized Medicine, http://biobankjp.org/;
EIGENSTRATsoftwarev2.0, http://genepath.med.harvard.edu/~reich/Software.htm;
R project v2.10.0, http://www.r-project.org/;
Haploview v4.1, http://www.broadinstitute.org/haploview/haploview;
MACH v1.0, http://www.sph.umich.edu/csg/yli/mach/index.html;
PLINK statistical software v1.06, http://pngu.mgh.harvard.edu/~purcell/plink/;
SNP info web server, http://manticore.niehs.nih.gov/index.html.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We extend our heartfelt gratitude to all individuals who kindly participated in the current study. We also thank the BioBank Japan for providing DNA samples and clinical information and the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting this study. We extend our appreciation to Dr Siew-Kee Low for her useful discussion in the statistical analysis and Dr Osman Wael Mohammed for his insightful advices. Finally, we would like to express our gratefulness to technical staff of the Laboratory for Genotyping Development in RIKEN, Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, the University of Tokyo and technical staff of collaborative hospitals for their assistances.

1 Massarweh, S. & Schiff, R. Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk. *Endocr. Relat. Cancer* **13,** S15–S24 (2006).
2 Xu, W. -H., Shu, X. O., Long, J., Lu, W., Cai, Q., Zheng, Y. *et al.* Relation of FGFR2 genetic polymorphisms to the association between oral contraceptive use and the risk of breast cancer in Chinese women. *Am. J. Epidemiol.* **173,** 923–931 (2011).
3 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 Res. Treat.* **43,** 65–72 (1997).
4 Tung, H. T., Tsukuma, H., Tanaka, H., Kinoshita, N., Koyama, Y., Ajiki, W. *et al.* Risk factors for breast cancer in Japan, with special attention to anthropometric measurements and reproductive history. *Jpn. J. Clin. Oncol.* **29,** 137–146 (1999).
5 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. *Clin. Cancer Res.* **9,** 2229–2233 (2003).
6 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. *Am. J. Epidemiol.* **152,** 950–964 (2000).
7 Yamashita, H., Iwase, H., Toyama, T., Takahashi, S., Sugiura, H., Yoshimoto, N. *et al.* Estrogen receptor-positive breast cancer in Japanese women: trends in incidence, characteristics, and prognosis. *Ann. Oncol.* **22,** 1318–1325 (2011).
8 Ao, A., Morrison, B. J., Wang, H., López, J. A., Reynolds, B. A. & Lu, J. Response of estrogen receptor-positive breast cancer tumorspheres to antiestrogen treatments. *PLoS ONE* **6,** e18810 (2011).
9 Jansen, M., Ruigrok-Ritstier, K., Dorssers, L. C., van Staveren, I. L., Look, M. P., Meijer-van Gelder, M. E. *et al.* Downregulation of SIAH2, an ubiquitin E3 ligase, is associated with resistance to endocrine therapy in breast cancer. *Breast Cancer Res. Treat.* **116,** 263–271 (2009).
10 Kiyotani, K., Mushiroda, T., Tsunoda, T., Morizono, T., Hosono, N., Kubo, M. *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. *Hum. Mol. Genet.* **21,** 1665–1672 (2012).
11 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).
12 Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. & Nakamura, Y. A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46,** 471–477 (2001).

13 Chan, S., Vermi, W., Luo, J., Lucini, L., Rickert, C., Fowler, A. *et al.* STAT1-deficient mice spontaneously develop estrogen receptor alpha-positive luminal mammary carcinomas. *Breast Cancer. Res.* **14,** R16 (2012).
14 Holmes, K. A., Song, J. S., Liu, X. S., Brown, M. & Carroll, J. S. Nkx3-1 and LEF-1 function as transcriptional inhibitors of estrogen receptor activity. *Cancer Res.* **68,** 7380–7385 (2008).
15 Huang, F. -I., Chen, Y. -L., Chang, C. -L., Yuan, R. -H. & Jeng, Y. -M. Hepatocyte growth factor activates Wnt pathway by transcriptional activation of LEF1 to facilitate tumor invasion. *Carcinogenesis* **33,** 1142–1148 (2012).
16 Beauchemin, D., Lacombe, C. & Van Themsche, C. PAX2 is activated by estradiol in breast cancer cells of the luminal subgroup selectively, to confer a low invasive phenotype. *Mol. Cancer* **10,** 148 (2011).
17 Walsh, T., Casadei, S., Coats, K. H., Swisher, E., Stray, S. M., Higgins, J. *et al.* Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA* 295, 1379–1388, 2006).
18 Sehrawat, B., Sridharan, M., Ghosh, S., Robson, P., Cass, C. E., Mackey, J. R. *et al.* Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Hum. Genet.* **130,** 529–537 (2011).
19 Higginbotham, K. S. P., Breyer, J. P., Bradley, K. M., Schuyler, P. A., Plummer, Jr W. D., Freudenthal, M. E. *et al.* A multistage association study identifies a breast cancer genetic locus at NCOA7. *Cancer Res.* **71,** 3881–3888 (2011).
20 Prentice, R. L., Huang, Y., Hinds, D. A., Peters, U., Pettinger, M., Cox, D. R. *et al.* Variation in the FGFR2 gene and the effects of postmenopausal hormone therapy on invasive breast cancer. *Cancer Epidem. Biomarker. Prev.* **18,** 3079–3085 (2009).
21 Ghoussaini, M., Fletcher, O., Michailidou, K., Turnbull, C., Schmidt, M. K., Dicks, E. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat. Genet.* **44,** 312–318 (2012).
22 Rebbeck, T. R., DeMichele, A., Tran, T. V., Panossian, S., Bunin, G. R., Troxel, A. B. *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).
23 Yoshimoto, N., Nishiyama, T., Toyama, T., Takahashi, S., Shiraki, N., Sugiura, H. *et al.* Genetic and environmental predictors, endogenous hormones and growth factors, and risk of estrogen receptor-positive breast cancer in Japanese women. *Cancer Sci.* **102,** 2065–2072 (2011).
24 Fanale, D., Amodeo, V., Corsini, L. R., Rizzo, S., Bazan, V. & Russo, A. Breast cancer genome-wide association studies: there is strength in numbers. *Oncogene* **31,** 2121–2128 (2012).
25 Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42,** 504–507 (2010).
26 Hamaguchi, M., Nishio, M., Toyama, T., Sugiura, H., Kondo, N., Fujii, Y. *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. *Jpn. J. Clin. Oncol.* **38,** 734–742 (2008).
27 Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).
28 van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).
29 Roh, M. S., Hong, S. H., Jeong, J. S., Kwon, H. C., Kim, M. C., Cho, S. H. *et al.* Gene expression profiling of breast cancers with emphasis of β-catenin regulation. *J. Korean Med. Sci.* **19,** 275–282 (2004).
30 Chan, P., Möller, A., Liu, M. C., Sceneay, J. E., Wong, C. S., Waddell, N. *et al.* The expression of the ubiquitin ligase SIAH2 (seven in absentia homolog 2) is mediated through gene copy number in breast cancer and is associated with a basal-like phenotype and p53 expression. *Breast Cancer Res.* **13,** R19 (2011).
31 Koren, R., Dekel, Y., Sherman, E., Weissman, Y., Dreznik, Z., Klein, B. *et al.* The expression of DCC protein in female breast cancer. *Breast Cancer Res. Treat.* **80,** 215–220 (2003).
32 Zhen-huan, Z., Yamashita, H., Toyama, T., Yamamoto, Y., Kawasoe, T., Ibusuki, M. *et al.* Nuclear corepressor 1 expression predicts response to first-line endocrine therapy for breast cancer patients on relapse. *Chin. Med. J.* **122,** 1764–1768 (2009).
33 Plevova, P., Bouchal, J., Fiurásková, M., Foretová, L., Navrátilová, M., Zapletalová, J. *et al.* PML protein expression in hereditary and sporadic breast cancer. *Neoplasma* **54,** 263–268 (2007).
34 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. *Clin. Cancer Res.* **5,** 1816–1822 (1999).
35 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. *Eur. J. Human Genet.* **9,** 121–129 (2001).
36 Ordulu, Z., Dal C., P., Chong, W. W., Choy, K. W., Lee, C., Muto, M. G. *et al.* Disseminated peritoneal leiomyomatosis after laparoscopic supracervical hysterectomy with characteristic molecular cytogenetic findings of uterine leiomyoma. *Genes Chromosom. Cancer* **49,** 1152–1160 (2010).
37 Cha, P. -C., Takahashi, A., Hosono, N., Low, S. K., Kamatani, N., Kubo, M. *et al.* A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids. *Nat. Genet.* **43,** 447–450 (2011).
38 Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38,** 1289–1297 (2006).
39 Moffa, A. B. & Ethier, S. P. Differential signal transduction of alternatively spliced FGFR2 variants expressed in human mammary epithelial cells. *J. Cell. Physiol.* **210,** 720–731 (2007).
40 Koziczak, M., Holbro, T. & Hynes, N. E. Blocking of FGFR signaling inhibits breast cancer cell proliferation through downregulation of D-type cyclins. *Oncogene* **23,** 3501–3508 (2004).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

Clinical Cancer Research

A Genome-Wide Association Study Identifies Novel Loci for Paclitaxel-Induced Sensory Peripheral Neuropathy in CALGB 40101

R. Michael Baldwin, Kouros Owzar, Hitoshi Zembutsu, et al.

Clin Cancer Res 2012;18:5099-5109. Published OnlineFirst July 27, 2012.