Published OnlineFirst November 30, 2012; DOI:10.1158/1078-0432.CCR-12-2618

Neurite-outgrowth assays

Twenty-four hours following siRNA transfection, transfection media was removed from the NS-1 cells and 0, 12.5, or 100 nmol/L paclitaxel in NS-1 media (supplemented with 20 ng/mL NGF) was added to either the *siRfx2* or nontargeting control cells. After 24 hours in the presence of paclitaxel, phase-contrast images (×10) of the cells were taken using an Axiovert 200M inverted widefield fluorescence microscope (Zeiss). At least 500 cells per treatment in 6 randomly chosen fields were imaged and the longest neurite per cell was measured using ImageJ (34) software. The entire experiment was carried out in duplicate and mean neurite lengths were normalized relative to the 0 nmol/L drug treatment for each siRNA. Because tracing neurite lengths is somewhat qualitative, 2 scientists independently measured neurite lengths and the second scientist was blinded to siRNA/drug treatment. The mean of each set of measurements between the 2 scientists was assessed for significance by 2-way ANOVA (factors: siRNA treatment and drug treatment) to determine if the *siRfx2* affected neurite length upon paclitaxel treatment.

NS-1 cytotoxicity assays

Twenty-four hours after siRNA transfection, transfection media was removed from the NS-1 cells and 0, 6.25, 12.5, 25, 50, or 100 nmol/L paclitaxel in NS-1 media (supplemented with 20 ng/mL NGF) in triplicate was added to either the *siRfx2* or nontargeting control cells. After 72 hours of paclitaxel treatment, ATP levels were measured using the CellTiter-Glo assay (Promega) and percentage survival curves were generated. The entire experiment was done in duplicate and 2-way ANOVA was used to determine if the *siRfx2* significantly affected overall cytotoxicity upon paclitaxel treatment.

Results

Enrichment of LCL cytotoxicity SNPs in patient sensory peripheral neuropathy SNPs

We conducted a genome-wide meta-analysis (see Materials and Methods) to test common SNPs for association with paclitaxel-induced cytotoxicity in LCLs. We compared the results from this analysis with those from clinical trial CALGB 40101, a GWAS of paclitaxel-induced sensory peripheral neuropathy in patients with breast cancer (3). Neither study produced genome-wide significant results ($\alpha < 0.05$) nor did the very top SNPs match between the 2 studies (Fig. 1). However, through a permutation resampling analysis of the CALGB patient data, we found that the top sensory peripheral neuropathy-associated SNPs ($P < 0.05$) are significantly enriched for SNPs associated with paclitaxel-induced cytotoxicity in LCLs ($P < 0.001$) with consistent allelic directions of effect (Fig. 2; empirical $P = 0.007$). The observed enrichment of 24 SNPs between the LCL and patient studies is likely paclitaxel-specific, due to the sensory peripheral neuropathy SNPs not being enriched for either capecitabine- or carboplatin-induced cytotoxicity SNPs, which were tested as negative controls (Fig. 2). Positional information and effect sizes of all 24 overlap SNPs in the LCL and patient data can be found in Supplementary Table S1. When the inclusion thresholds for overlap SNPs were relaxed and when the LCL SNPs were tested for enrichment of patient SNPs, the significant overlap was present at a range of *P* value thresholds from 0.001 to 0.1, showing the robustness of our findings (Supplementary Table S2).

Enrichment of eQTLs in LCL/patient overlap SNPs

We tested the top paclitaxel-induced LCL cytotoxicity SNPs ($P < 0.001$) and the top paclitaxel-induced patient sensory peripheral neuropathy SNPs ($P < 0.05$) for eQTL

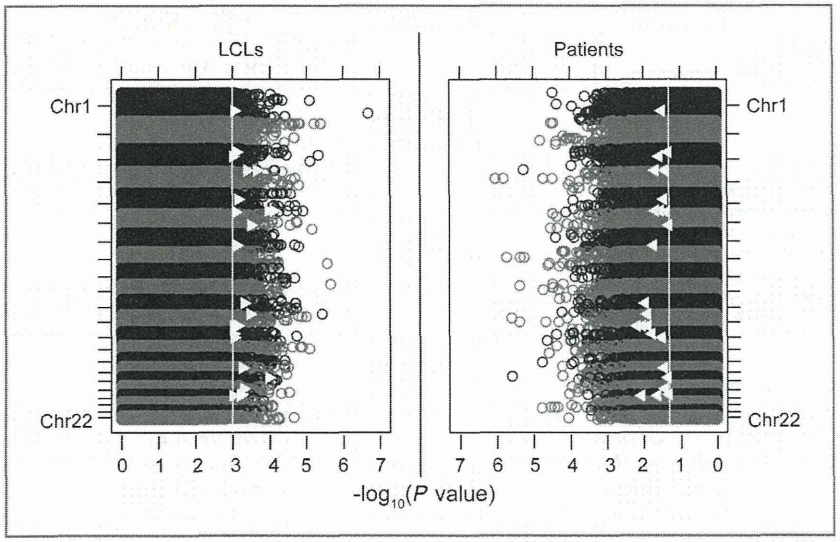

Figure 1. Comparison of individual GWAS results. Left, paclitaxel-induced cytotoxicity in LCLs. Right, paclitaxel-induced sensory peripheral neuropathy in patients. White lines represent the overlap thresholds used in the primary enrichment analysis ($P < 0.001$ for LCLs and $P < 0.05$ for patients) and white triangles represent the 24 overlap SNPs at these thresholds.

Published OnlineFirst November 30, 2012; DOI:10.1158/1078-0432.CCR-12-2618

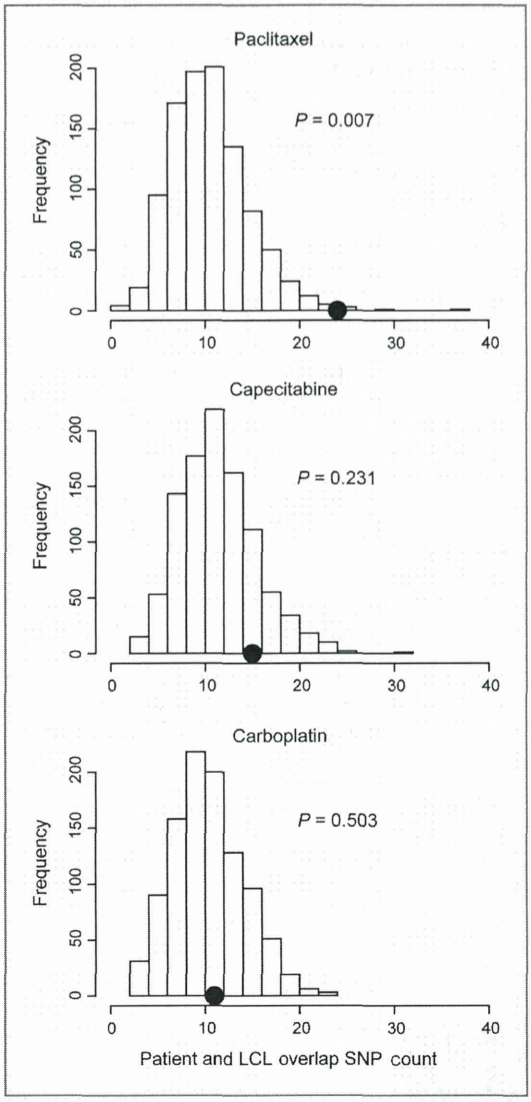

Figure 2. Patient paclitaxel-induced sensory peripheral neuropathy SNPs are enriched for SNPs associated with paclitaxel-induced cytotoxicity in LCLs. Distribution of chemotherapeutic-induced cytotoxicity SNP ($P < 0.001$) count in 1,000 permutations of neuropathy phenotype–genotype connections ($P < 0.05$). The dot is the observed SNP overlap at these thresholds. Of the 3 drug studies tested (paclitaxel, capecitabine, and carboplatin), only paclitaxel-induced cytotoxicity SNPs were significantly enriched in the patient GWAS (empirical $P = 0.007$).

enrichment because these were the thresholds used in the primary overlap analysis. We compared the observed number of eQTLs at these thresholds to the number of eQTLs in 10,000 randomly selected MAF-matched SNP sets (for details, see Materials and Methods). Neither cytotoxicity-associated SNPs nor neuropathy-associated SNPs alone were enriched for eQTLs (Fig. 3). However, we found that the 24 paclitaxel LCL/patient overlap SNPs at these thresholds are enriched for eQTLs when compared with MAF-matched SNP sets (empirical $P = 0.0447$), potentially revealing an important role for this functional class in paclitaxel toxicity.

Prioritization of LCL/patient overlap SNPs for functional analysis

First, we determined that 11 of 24 overlap SNPs from the enrichment analysis were located in or near (within 2 kb) gene transcripts (Table 1). The relationship of these 11 SNPs with paclitaxel-induced sensory peripheral neuropathy in patients and LCL cytotoxicity is shown in Supplementary Fig. S2. Next, we determined which of these 11 SNPs within genes were also eQTLs (31). Of the 8 eQTLs, we determined which had the most potential target genes at an arbitrary threshold of $P < 10^{-4}$. The SNP in *RFX2* had 18 target genes, more than any other of the 8 eQTLs. In addition, we tested the expression of the target genes for association with paclitaxel-induced cytotoxicity adjusted for growth rate (see Materials and Methods). We found that expression of 3 of the *RFX2* target genes associated with paclitaxel-induced cytotoxicity (Table 1 and Supplementary Table S3); therefore, we pursued evaluating *RFX2* in a model of neuropathy.

Functional validation of RFX2 in a paclitaxel-induced peripheral neuropathy model

We used neuroscreen (NS-1) cells, a subclone of the rat pheochromocytoma cell line PC-12 that has previously been used as a research model for chemotherapy-induced neuropathy (35, 36), to test *Rfx2*, the rat ortholog of *RFX2*, for functional involvement in paclitaxel response. Using siRNA, we decreased expression of *Rfx2* resulting in increased sensitivity of NS-1 cells to paclitaxel, as measured by reduced neurite outgrowth and increased cytotoxicity (Fig. 4). The 3 *RFX2* SNP target genes whose expression associated with paclitaxel-induced cytotoxicity in LCLs are *CYP51A1*, *BACH1*, and *CBARA1* (Table 1; Fig. 5A–C; $P < 0.05$). We measured the expression of these 3 potential *Rfx2* target genes upon knockdown of *Rfx2* in NS-1 cells and found that the expression of 1 of 3 genes, *Cyp51* (rat ortholog of *CYP51A1*), significantly decreased 24 hours posttransfection ($P < 0.05$), which is the expected direction of effect based on the LCL expression versus cytotoxicity data (Fig. 5D).

Discussion

We conducted a GWAS of paclitaxel-induced cytotoxicity in LCLs and showed significant enrichment of the top cytotoxicity-associated SNPs in a clinical GWAS of paclitaxel-induced sensory peripheral neuropathy in patients with breast cancer. This robust enrichment shows that susceptibilities to increased cytotoxicity in LCLs and sensory peripheral neuropathy in patients with breast cancer likely have some genetic mechanisms in common and supports the role of LCLs as a preclinical model for paclitaxel toxicity studies. Furthermore, the top SNPs that overlap between the 2 studies were enriched for eQTLs. This eQTL enrichment

Published OnlineFirst November 30, 2012; DOI:10.1158/1078-0432.CCR-12-2618

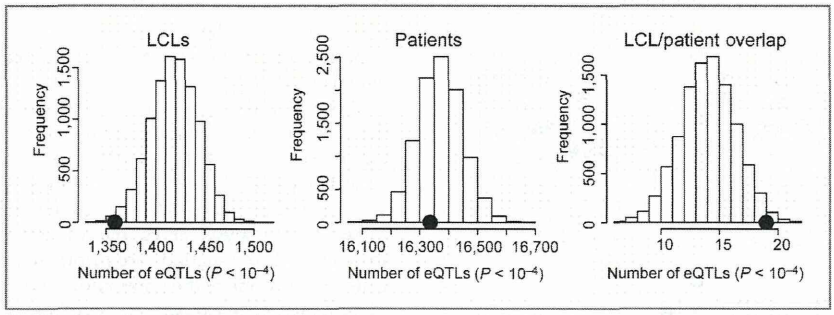

Figure 3. SNPs associated with both patient paclitaxel-induced sensory peripheral neuropathy and LCL paclitaxel-induced cytotoxicity are enriched for eQTLs. Distribution of eQTL ($P < 10^{-4}$) count in 10,000 simulations, each matching the MAF distribution of either LCL paclitaxel SNPs ($P < 0.001$), patient paclitaxel SNPs ($P < 0.05$), or the set of 24 LCL/patient overlap SNPs at these P value thresholds. Neither the LCL paclitaxel SNPs nor the patient paclitaxel SNPs alone were enriched for eQTLs, but the overlap SNP set between the 2 GWAS was enriched for eQTLs (empirical $P = 0.0447$).

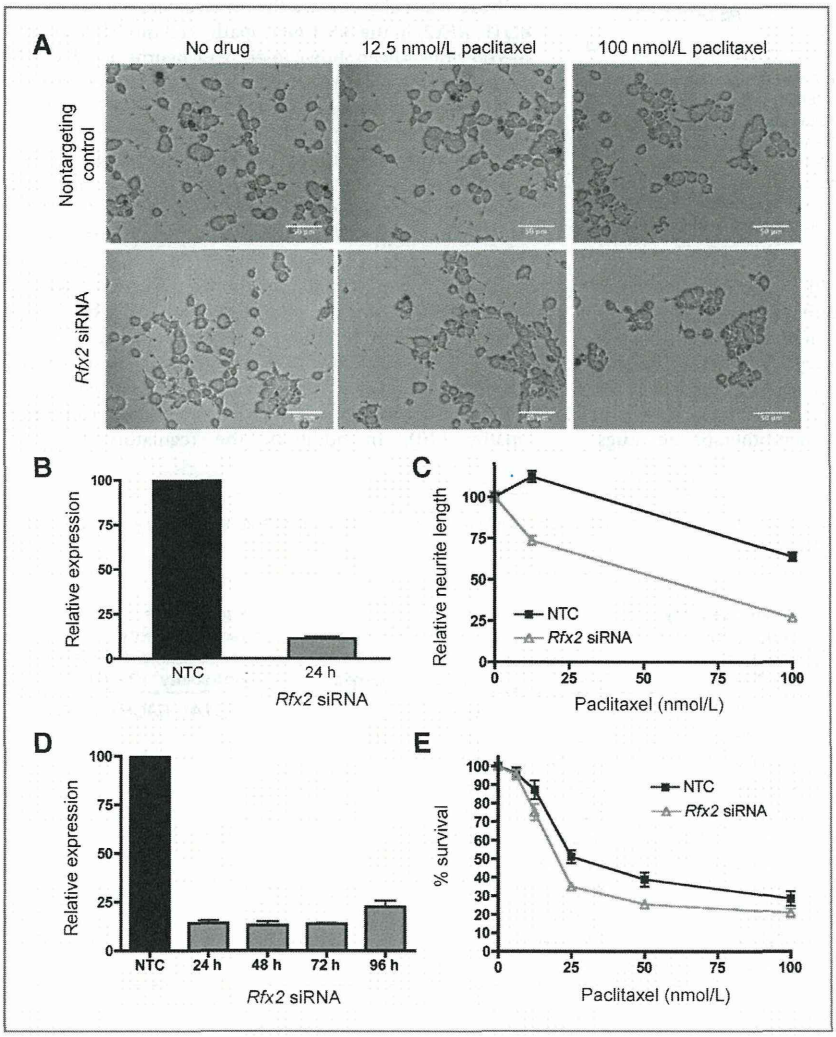

Figure 4. Functional validation of *RFX2* in paclitaxel response using a peripheral neuropathy cell model. A, representative micrographs comparing neurite lengths of NS-1 cells upon siRNA knockdown of *Rfx2* and treatment with paclitaxel (×10 phase-contrast). B, relative gene expression 24 hours posttransfection in the 2 neurite length experiments. NTC, nontargeting control. C, decreased expression of *Rfx2* causes decreased neurite length of differentiating NS-1 cells ($P < 10^{-4}$) 24 hours post-paclitaxel treatment (48 hours posttransfection). Error bars represent the SEM of the longest relative neurite length of at least 500 cells in each of 2 independent experiments. D, relative gene expression 24 to 96 hours posttransfection in the 2 cytotoxicity experiments. E, decreased expression of *Rfx2* causes decreased survival (increased cytotoxicity, $P < 10^{-4}$) of differentiating NS-1 cells measured by CellTiter-Glo 72 hours post-paclitaxel treatment (96 hours posttransfection). Error bars represent the SEM of survival in 2 independent experiments with 3 replicates each.

Published OnlineFirst November 30, 2012; DOI:10.1158/1078-0432.CCR-12-2618

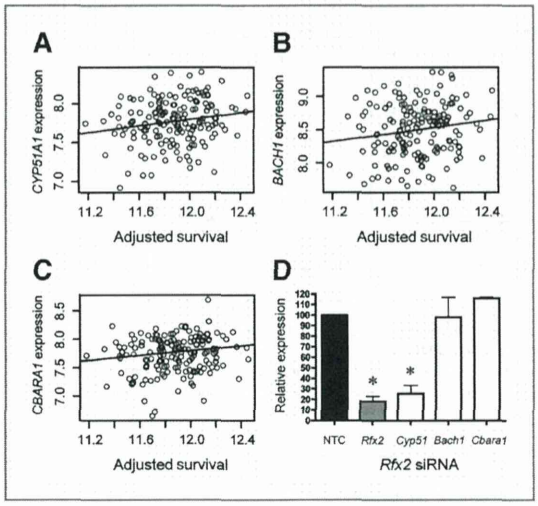

Figure 5. Target genes of *RFX2* eQTL rs7254081 in paclitaxel response. Increased baseline expression of the rs7254081 target genes (A) *CYP51A1*, (B) *BACH1*, and (C) *CBARA1* associate with increased cellular survival (adjusted for growth rate) of LCLs treated with 12.5 nmol/L paclitaxel ($P < 0.05$). D, *Rfx2* siRNA in NS-1 cells decreases the expression of *Rfx2* and *Cyp51*, but not *Bach1* and *Cbara1*, compared with the nontargeting control (NTC) 24 hours posttransfection. *, $P < 0.05$. Error bars represent the SEM of relative gene expression in 2 independent experiments with 3 replicates each.

indicates that SNPs associated with paclitaxel-induced toxicity phenotypes may be functioning through gene regulatory mechanisms. Interestingly, neither GWAS alone was enriched for eQTLs. Thus, our integration method may be reducing noise and revealing important functional SNPs. An enrichment of eQTLs has previously been shown in SNPs associated with 6 other chemotherapeutic drugs, which indicates that susceptibility to these drugs may depend on subtle gene expression differences across individuals (31).

The enrichment analyses were likely affected by the different linkage disequilibrium patterns among the populations studied. The LCL GWAS was a meta-analysis of African, African American, and European populations, whereas the patient GWAS was conducted in Europeans. In the meta-analysis, SNPs that are associated with cytotoxicity in all populations are prioritized over those associated in only one of the populations. We may have missed identifying European-specific overlap alleles. However, because the population linkage disequilibrium patterns differ and because African populations have shorter linkage disequilibrium blocks, overlap SNPs are more likely to be functional SNPs rather than SNPs that simply tag a functional locus (37).

We functionally assessed the involvement of one overlap eQTL, *RFX2*, in the NS-1 neuropathy cell model. Paclitaxel has previously been shown to decrease neurite outgrowth in the parent clone of the NS-1 cell line (36). Here, we showed that decreased expression of *Rfx2* sensitizes NS-1 cells to paclitaxel by reducing neurite outgrowth and survival. This result validates our approach by showing that patient neuropathy and LCL cytotoxicity overlap analyses can reveal genes mechanistically involved in paclitaxel response. Although most previous work on *RFX2* in mammalian cells describes its role in spermatogenesis (38, 39), several studies point to a potential role for the protein in sensory neurons. RFX2 and the related protein RFX1 have been shown to directly bind and regulate the transcription of *ALMS1* (40). Mutations in *ALMS1* cause the rare genetic disorder Alström syndrome, which is characterized by neurosensory degeneration, metabolic defects, and cardiomyopathy (40). In addition, the regulatory factor X

Table 1. Paclitaxel-induced LCL cytotoxicity ($P < 0.001$) and paclitaxel-induced patient sensory peripheral neuropathy ($P < 0.05$) overlap SNPs located in genes

SNP	LCL cytotoxicity *P* value	Patient sensory peripheral neuropathy *P* value	Gene	eQTL	Number of target genes	Target genes associated with LCL paclitaxel-induced cytotoxicity[a] ($P < 0.05$)
rs7254081	5.9E-04	4.8E-02	*RFX2*	yes	18	*CYP51A1*, *BACH1*, *CBARA1*
rs7642318	2.2E-04	3.9E-02	*TMEM44*	yes	6	
rs10933663	4.0E-04	2.1E-02	*TMEM44*	yes	4	
rs8002545	9.2E-04	3.1E-02	*DIS3*	yes	3	
rs4782010	5.5E-04	3.5E-02	*XYLT1*	yes	2	
rs11111539	7.9E-04	6.7E-03	*C12orf42*	yes	1	
rs7306825	7.2E-04	9.3E-03	*C12orf42*	yes	1	
rs8069856	1.1E-04	4.5E-02	*RICH2*	yes	1	
rs4868011	8.2E-04	4.2E-02	*KCNIP1*			
rs10778237	9.3E-04	1.3E-02	*C12orf42*			
rs323285	5.1E-04	3.7E-02	*KIAA1328*			

[a]Adjusted for growth rate.

Published OnlineFirst November 30, 2012; DOI:10.1158/1078-0432.CCR-12-2618

transcription factors present in *Caenorhabditis elegans* and *Drosophila*, which are called DAF-19 and RFX, respectively, regulate ciliated sensory neuron differentiation (41, 42).

Upon knockdown of *Rfx2* in NS-1 cells, the potential target gene *Cyp51* also decreased expression, which was the expected direction of effect based on the preliminary gene expression analysis in LCLs. However, *CYP51A1* does not contain an X-box RFX-binding domain (43) in the promoter region (2 kb upstream of the transcription start site), which means it is unlikely a direct target of *RFX2* and may instead be further downstream in the pathway. Alternatively, *RFX2* could be regulating an enhancer of *CYP51A1* that is further outside the gene region. CYP51A1 is a member of the cytochrome P450 superfamily of enzymes, which catalyze many reactions involved in the metabolism of drugs and endogenous compounds. Specifically, CYP51A1 is known to participate in the synthesis of cholesterol (44). CYP51A1 has not been previously implicated in paclitaxel metabolism (45).

In the CALGB GWAS, one of the top SNPs that associated with patient paclitaxel-induced sensory peripheral neuropathy (rs10771973, $P = 2.6 \times 10^{-6}$) is located in *FGD4* (3). Mutations in *FGD4* can cause the congenital peripheral neuropathy Charcot–Marie–Tooth disease type 4H, and thus the gene is a plausible candidate for involvement in variation in peripheral neuropathy induced by paclitaxel. This SNP association was replicated in a second cohort of self-reported White patients with breast cancer ($n = 154$; $P = 0.013$) and in a cohort of self-reported African American patients with breast cancer ($n = 117$; $P = 6.7 \times 10^{-3}$; ref. 3). However, this SNP was not associated with paclitaxel-induced cytotoxicity in LCLs ($P = 0.65$). *FGD4* is not expressed in LCLs (21), and thus the SNP is not expected to function in this model system. While our integrative approach can reveal variants and genes acting in paclitaxel response in both patients and LCLs, it does not identify genes potentially acting in patients that are not expressed in LCLs.

Effectively, LCLs have been used as an additional cohort to study the pharmacogenomics of various chemotherapeutics (16–19) because limited resources and *in vivo* confounders make obtaining large, homogeneous patient cohorts difficult. Here, we saw greater SNP overlap than expected by chance between SNPs associated with paclitaxel-induced cytotoxicity in LCLs and SNPs associated with paclitaxel-induced sensory peripheral neuropathy in patients at multiple *P* value thresholds, which confirms a role for the LCL model in the analysis of at least a subset of genes involved in patient neurotoxicity. This significant enrichment among a relatively large number of top SNPs is consistent with an underlying polygenic architecture for paclitaxel-induced toxicity. Functional siRNA studies in the NS-1 neuropathy model validated the involvement of RFX2 in paclitaxel toxicity, supporting our multi-gene hypothesis. Our novel integrative enrichment approach that combines clinical and LCL GWAS results can be used to expand patient cohort sizes for any drug phenotype of interest, including other toxicities, such as neutropenia, to find genes of potential impact that can be studied in cellular models.

Disclosure of Potential Conflicts of Interest

E.P. Winer has a Commercial Research Grant from Roche. No potential conflicts of interest were disclosed by the other authors. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute (NCI).

Authors' Contributions

Conception and design: H.E. Wheeler, E.R. Gamazon, K. Owzar, M. Kubo, C. Hudis, L.N. Shulman, Y. Nakamura, M.J. Ratain, N.J. Cox, M.E. Dolan
Development of methodology: H.E. Wheeler, E.R. Gamazon, C. Wing, K. Owzar, N.J. Cox
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H.E. Wheeler, C. Wing, U.O. Njiaju, C. Njoku, R. M. Baldwin, M. Kubo, H. Zembutsu, C. Hudis, L.N. Shulman, Y. Nakamura, D.L. Kroetz, M.E. Dolan
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): H.E. Wheeler, E.R. Gamazon, C. Wing, R. M. Baldwin, K. Owzar, D. Watson, I. Shterev, C. Hudis, L.N. Shulman, M.J. Ratain, D.L. Kroetz, N.J. Cox, M.E. Dolan
Writing, review, and/or revision of the manuscript: H.E. Wheeler, E.R. Gamazon, C. Wing, R.M. Baldwin, K. Owzar, C. Jiang, H. Zembutsu, E. Winer, C. Hudis, L.N. Shulman, M.J. Ratain, D.L. Kroetz, N.J. Cox, M.E. Dolan
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C. Wing, C. Hudis
Study supervision: H.E. Wheeler, C. Hudis, L.N. Shulman, M.E. Dolan

Acknowledgments

The authors thank the Pharmacogenomics of Anticancer Agents Research LCL core for maintaining and distributing LCLs.

Grant Support

This study is supported by NIH/NCI Cancer Biology Training grant T32CA009594 and National Research Service Award F32CA165823 (to H. E. Wheeler), NIH/National Institute of General Medical Sciences (NIGMS) Pharmacogenomics of Anticancer Agents grant U01GM61393 (to M.J. Ratain, N.J. Cox, M.E. Dolan, and K. Owzar), NIH/NIGMS Pharmacogenetics of Membrane Transporters grant U01GM61390 (to D.L. Kroetz), University of Chicago SPORE Grant NCI P50CA125183 (to M.E. Dolan), NIH/NCI Medical Oncology Training grant T32CA009566 (to U.O. Njiaju), and the Biobank Japan Project funded by the Japanese Ministry of Education, Culture, Sports, Science and Technology. This work is part of the NIH Pharamcogenomics Research Network-RIKEN Center for Genomic Medicine Global Alliance.

The research for CALGB 60202 and 40101 was supported, in part, by grants from the NCI (CA31946) to the CALGB (Monica M. Bertagnolli) and to the CALGB Statistical Center (Daniel J. Sargent, CA33601).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 7, 2012; revised November 1, 2012; accepted November 8, 2012; published OnlineFirst November 30, 2012.

References

1. Shulman LN, Cirrincione CT, Berry DA, Becker HP, Perez EA, O'Regan R, et al. Six cycles of doxorubicin and cyclophosphamide or paclitaxel are not superior to four cycles as adjuvant chemotherapy for breast cancer in women with zero to three positive axillary nodes: cancer and leukemia group B 40101. J Clin Oncol 2012;30:4071–6.
2. Pachman DR, Barton DL, Watson JC, Loprinzi CL. Chemotherapy-induced peripheral neuropathy: prevention and treatment. Clin Pharmacol Ther 2011;90:377–87.
3. Baldwin RM, Owzar K, Zembutsu H, Chhibber A, Kubo M, Jiang C, et al. A genome-wide association study identifies novel loci for paclitaxel-

Published OnlineFirst November 30, 2012; DOI:10.1158/1078-0432.CCR-12-2618

induced sensory peripheral neuropathy in CALGB 40101. Clin Cancer Res 2012;18:5099–109.
4. Green H, Soderkvist P, Rosenberg P, Horvath G, Peterson C. mdr-1 single nucleotide polymorphisms in ovarian cancer tissue: G2677T/A correlates with response to paclitaxel chemotherapy. Clin Cancer Res 2006;12:854–9.
5. Hertz DL, Motsinger-Reif AA, Drobish A, Winham SJ, McLeod HL, Carey LA, et al. CYP2C8*3 predicts benefit/risk profile in breast cancer patients receiving neoadjuvant paclitaxel. Breast Cancer Res Treat 2012;134:401–10.
6. Leandro-Garcia LJ, Leskela S, Jara C, Green H, Avall-Lundqvist E, Wheeler HE, et al. Regulatory polymorphisms in beta-tubulin IIa are associated with paclitaxel-induced peripheral neuropathy. Clin Cancer Res 2012;18:4441–8.
7. Leskela S, Leandro-Garcia LJ, Mendiola M, Barriuso J, Inglada-Perez L, Munoz I, et al. The miR-200 family controls beta-tubulin III expression and is associated with paclitaxel-based treatment response and progression-free survival in ovarian cancer patients. Endocr Relat Cancer 2011;18:85–95.
8. Sissung TM, Mross K, Steinberg SM, Behringer D, Figg WD, Sparreboom A, et al. Association of ABCB1 genotypes with paclitaxel-mediated peripheral neuropathy and neutropenia. Eur J Cancer 2006;42:2893–6.
9. Huang RS, Duan S, Bleibel WK, Kistner EO, Zhang W, Clark TA, et al. A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. Proc Natl Acad Sci U S A 2007;104:9758–63.
10. Li L, Fridley BL, Kalari K, Jenkins G, Batzler A, Weinshilboum RM, et al. Gemcitabine and arabinosylcytosin pharmacogenomics: genome-wide association and drug response biomarkers. PLoS ONE 2009;4:e7765.
11. Watters JW, Kraja A, Meucci MA, Province MA, McLeod HL. Genome-wide discovery of loci influencing chemotherapy cytotoxicity. Proc Natl Acad Sci U S A 2004;101:11809–14.
12. Wheeler HE, Dolan ME. Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation. Pharmacogenomics 2012;13:55–70.
13. Wheeler HE, Gamazon ER, Stark AL, O'Donnell PH, Gorsic LK, Huang RS, et al. Genome-wide meta-analysis identifies variants associated with platinating agent susceptibility across populations. Pharmacogenomics J. 2011 Aug 16. [Epub ahead of print].
14. Ingle JN, Schaid DJ, Goss PE, Liu M, Mushiroda T, Chapman JA, et al. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. J Clin Oncol 2010;28:4674–82.
15. Shukla SJ, Duan S, Wu X, Badner JA, Kasza K, Dolan ME. Whole-genome approach implicates CD44 in cellular resistance to carboplatin. Hum Genomics 2009;3:128–42.
16. Huang RS, Johnatty SE, Gamazon ER, Im HK, Ziliak D, Duan S, et al. Platinum sensitivity-related germline polymorphism discovered via a cell-based approach and analysis of its association with outcome in ovarian cancer patients. Clin Cancer Res 2011;17:5490–500.
17. Mitra AK, Crews K, Pounds S, Cao X, Downing JR, Raimondi S, et al. Impact of genetic variation in FKBP5 on clinical response in pediatric acute myeloid leukemia patients: a pilot study. Leukemia 2011;25:1354–6.
18. Tan XL, Moyer AM, Fridley BL, Schaid DJ, Niu N, Batzler AJ, et al. Genetic variation predicting Cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based chemotherapy. Clin Cancer Res 2011;17:5801–11.
19. Ziliak D, O'Donnell PH, Im HK, Gamazon ER, Chen P, Delaney S, et al. Germline polymorphisms discovered via a cell-based, genome-wide approach predict platinum response in head and neck cancers. Transl Res 2011;157:265–72.
20. Cox NJ, Gamazon ER, Wheeler HE, Dolan ME. Clinical translation of cell-based pharmacogenomic discovery. Clin Pharmacol Ther 2012;92:425–7.
21. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, et al. SCAN: SNP and copy number annotation. Bioinformatics 2010;26:259–62.
22. Njiaju UO, Gamazon ER, Gorsic LK, Delaney SM, Wheeler HE, Im HK, et al. Whole-genome studies identify solute carrier transporters in cellular susceptibility to paclitaxel. Pharmacogenet Genomics 2012;22:498–507.
23. Abecasis GR, Cookson WO, Cardon LR. Pedigree tests of transmission disequilibrium. Eur J Hum Genet 2000;8:545–51.
24. Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet 2009;5:e1000519.
25. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009;84:210–23.
26. Wheeler HE, Gorsic LK, Welsh M, Stark AL, Gamazon ER, Cox NJ, et al. Genome-wide local ancestry approach identifies genes and variants associated with chemotherapeutic susceptibility in African Americans. PLoS ONE 2011;6:e21920.
27. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997–1004.
28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–1.
29. Shterev ID, Jung SH, George SL, Owzar K. permGPU: using graphics processing units in RNA microarray association studies. BMC Bioinformatics 2010;11:329.
30. O'Donnell PH, Stark AL, Gamazon ER, Wheeler HE, McIlwee BE, Gorsic L, et al. Identification of novel germline polymorphisms governing capecitabine sensitivity. Cancer 2012;118:4063–73.
31. Gamazon ER, Huang RS, Cox NJ, Dolan ME. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. Proc Natl Acad Sci U S A 2010;107:9287–92.
32. Duan S, Huang RS, Zhang W, Bleibel WK, Roe CA, Clark TA, et al. Genetic architecture of transcript-level variation in humans. Am J Hum Genet 2008;82:1101–13.
33. Stark AL, Zhang W, Mi S, Duan S, O'Donnell PH, Huang RS, et al. Heritable and non-genetic factors as variables of pharmacologic phenotypes in lymphoblastoid cell lines. Pharmacogenomics J 2010;10:505–12.
34. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012;9:671–5.
35. Geldof AA. Nerve-growth-factor-dependent neurite outgrowth assay; a research model for chemotherapy-induced neuropathy. J Cancer Res Clin Oncol 1995;121:657–60.
36. Verstappen CC, Postma TJ, Geldof AA, Heimans JJ. Amifostine protects against chemotherapy-induced neurotoxicity: an *in vitro* investigation. Anticancer Res 2004;24:2337–41.
37. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet 2010;11:149–60.
38. Horvath GC, Kistler WS, Kistler MK. RFX2 is a potential transcriptional regulatory factor for histone H1t and other genes expressed during the meiotic phase of spermatogenesis. Biol Reprod 2004;71:1551–9.
39. Wolfe SA, Wilkerson DC, Prado S, Grimes SR. Regulatory factor X2 (RFX2) binds to the H1t/TE1 promoter element and activates transcription of the testis-specific histone H1t gene. J Cell Biochem 2004;91:375–83.
40. Purvis TL, Hearn T, Spalluto C, Knorz VJ, Hanley KP, Sanchez-Elsner T, et al. Transcriptional regulation of the Alstrom syndrome gene ALMS1 by members of the RFX family and Sp1. Gene 2010;460:20–9.
41. Swoboda P, Adler HT, Thomas JH. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. Mol Cell 2000;5:411–21.
42. Dubruille R, Laurencon A, Vandaele C, Shishido E, Coulon-Bublex M, Swoboda P, et al. *Drosophila* regulatory factor X is necessary for ciliated sensory neuron differentiation. Development 2002;129:5487–98.
43. Gajiwala KS, Chen H, Cornille F, Roques BP, Reith W, Mach B, et al. Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature 2000;403:916–21.
44. Halder SK, Fink M, Waterman MR, Rozman D. A cAMP-responsive element binding site is essential for sterol regulation of the human lanosterol 14alpha-demethylase gene (CYP51). Mol Endocrinol 2002;16:1853–63.
45. Oshiro C, Marsh S, McLeod H, Carrillo M, Klein T, Altman R. Taxane Pathway. Pharmacogenet Genomics 2009;19:979–83.

Published in final edited form as:
Nat Genet. 2012 October 28; 44(12): 1355–1359. doi:10.1038/ng.2445.

Genome-wide association meta-analysis identifies new endometriosis risk loci

Dale R. Nyholt[1,16], **Siew-Kee Low**[2,16], **Carl A. Anderson**[3], **Jodie N. Painter**[1], **Satoko Uno**[2,4], **Andrew P. Morris**[5], **Stuart MacGregor**[1], **Scott D. Gordon**[1], **Anjali K. Henders**[1], **Nicholas G. Martin**[1], **John Attia**[6,7], **Elizabeth G. Holliday**[6,7], **Mark McEvoy**[6,8,9], **Rodney J. Scott**[7,10,11], **Stephen H. Kennedy**[12], **Susan A. Treloar**[13], **Stacey A. Missmer**[14], **Sosuke Adachi**[15], **Kenichi Tanaka**[15], **Yusuke Nakamura**[2], **Krina T. Zondervan**[5,12,17], **Hitoshi Zembutsu**[2,17], and **Grant W. Montgomery**[1,17]

[1]Queensland Institute of Medical Research, Brisbane, QLD 4029, Australia.

[2]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan.

[3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 2HH, UK.

[4]First Department of Surgery, Sapporo Medical University, School of Medicine, Sapporo, Japan.

[5]Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.

Correspondence should be addressed to D.R.N (dale.nyholt@qimr.edu.au), K.T.Z. (krina.zondervan@well.ox.ac.uk), H.Z (zembutsh@ims.u-tokyo.ac.jp) or G.W.M (grant.montgomery@qimr.edu.au).; **Corresponding (submitting) author:** A/Prof Dale R Nyholt Queensland Institute of Medical Research Locked Bag 2000, Royal Brisbane Hospital, Herston, Qld 4029, Australia Tel: +61 7 3362 0258 Fax: +61 7 3362 0101 dale.nyholt@qimr.edu.au.

[16]These authors contributed equally to this work.

[17]These authors jointly directed this work.

Author contributions:

Manuscript preparation and final approval: D.R.N., S.-K.L., C.A.A., J.N.P., S.U., A.P.M., S.M., S.D.G., A.K.H., N.G.M., J.A., E.G.H., M.M., R.J.S., S.H.K., S.A.T., S.A.M., S.A., K.T., Y.N., K.T.Z., H.Z. & G.W.M.

Study conception and design: D.R.N., S.M., Y.N., K.T.Z., H.Z. & G.W.M.

GWAS data collection, sample preparation and clinical phenotyping: J.N.P., S.U., A.K.H., N.G.M., J.A., E.G.H., M.M., R.J.S., S.H.K., S.A.T., K.T.Z., H.Z. & G.W.M.

Replication data collection, sample preparation and clinical phenotyping: S.A., K.T. & H.Z.

Replication genotyping: H.Z.

Data analysis: GWA analysis: D.R.N., C.A.A. & S.-K.L.; imputation and replication analysis: D.R.N. & S.-K.L.; polygenic prediction, gene-based and meta-analysis: D.R.N.

Obtaining study funding: D.R.N., S.M., N.G.M., S.H.K., S.A.T., S.A.M., Y.N., K.T.Z. & G.W.M.

Competing financial interests The authors declare no competing financial interests.

URLs.

A Catalog of Published Genome-Wide Association Studies, www.genome.gov/gwastudies;
Gene Expression Omnibus (GEO) database, http://www.ncbi.nlm.nih.gov/gds/;
GENe Expression VARiation (Genevar) database, http://www.sanger.ac.uk/resources/software/genevar/;
GWAMA, http://www.well.ox.ac.uk/gwama/;
MaCH, http://www.sph.umich.edu/csg/abecasis/MaCH/;
Mammalian Gene Expression Uterus database (MGEx-Udb), http://resource.ibab.ac.in/cgi-bin/MGEXdb/microarray/scoring/interface/Homepage.pl;
METAL, http://genome.sph.umich.edu/wiki/METAL_Program;
Metasoft, http://genetics.cs.ucla.edu/meta/index.html; minimac, http://genome.sph.umich.edu/wiki/Minimac;
minimac, http://genome.sph.umich.edu/wiki/Minimac;
Minimac: 1000 Genomes Imputation Cookbook, http://genome.sph.umich.edu/wiki/Minimac:_1000_Genomes_Imputation_Cookbook; 1000 Genomes Project, http://www.1000genomes.org/;
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/;
SNPSpD, http://genepi.qimr.edu.au/general/daleN/SNPSpD/;
Wellcome Trust Case-Control Consortium, http://www.wtccc.org.uk;

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

[6]Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, Newcastle, NSW 2308, Australia.

[7]Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine, Hunter Medical Research Institute, Newcastle, NSW 2305, Australia.

[8]School of Medicine and Public Health, University of Newcastle, Newcastle, NSW 2308, Australia.

[9]Public Health Research Program, Hunter Medical Research Institute, Newcastle, NSW 2305, Australia.

[10]School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, NSW 2308, Australia.

[11]Division of Genetics, Hunter Area Pathology Service, Newcastle, NSW 2305, Australia.

[12]Nuffield Department of Obstetrics and Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK.

[13]Centre for Military and Veterans' Health, University of Queensland, Mayne Medical School, 288 Herston Road, QLD 4006, Australia.

[14]Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA.

[15]Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, 1-757 Asahimachi-dori, Niigata, Japan.

Abstract

We conducted a genome-wide association (GWA) meta-analysis of 4,604 endometriosis cases and 9,393 controls of Japanese[1] and European[2] ancestry. We show that rs12700667 on chromosome 7p15.2, previously found in Europeans, replicates in Japanese ($P = 3.6 \times 10^{-3}$), and confirm association of rs7521902 on 1p36.12 near *WNT4*. In addition, we establish association of rs13394619 in *GREB1* on 2p25.1 and identify a novel locus on 12q22 near *VEZT* (rs10859871). Excluding European cases with minimal or unknown severity, we identified additional novel loci on 2p14 (rs4141819), 6p22.3 (rs7739264) and 9p21.3 (rs1537377). All seven SNP effects were replicated in an independent cohort and produced $P < 5 \times 10^{-8}$ in a combined analysis. Finally, we found a significant overlap in polygenic risk for endometriosis between the European and Japanese GWA cohorts ($P = 8.8 \times 10^{-11}$), indicating that many weakly associated SNPs represent true endometriosis risk loci and risk prediction and future targeted disease therapy may be transferred across these populations.

Endometriosis (MIM131200) is a common gynecological disease associated with severe pelvic pain, affecting 6-10% of women in their reproductive years[3,4] and 20-50% of women with infertility[5]. Endometriosis risk is influenced by genetic factors and has an estimated heritability of around 51%[3].

Two large endometriosis GWA studies[1,2] have reported genome-wide significant associations. The first, in a Japanese sample of 1,423 cases and 1,318 controls obtained from the BioBank Japan (BBJ), with 484 cases and 3,974 controls for replication, implicated a SNP (rs10965235) in the *CDKN2BAS* gene on chromosome 9p21.3 (overall odds ratio (OR) = 1.44, 95% CI 1.30–1.59; $P = 5.57 \times 10^{-12}$)[1]. The second, by the International Endogene Consortium (IEC) in a sample of European ancestry from Australia (2,270 cases and 1,870 controls) and the UK (924 cases and 5,190 controls), with 2,392 cases and 2,271 controls from the US for replication, identified an intergenic SNP (rs12700667) on 7p15.2 (overall OR = 1.20, 95% CI 1.13–1.27; $P = 1.4 \times 10^{-9}$)[2]. These two studies did not report replication

Europe PMC Funders Author Manuscripts

of each other's top locus, partly because rs10965235 is monomorphic in Caucasian populations. The European study did find association with rs7521902 (OR = 1.16, 95% CI 1.08–1.25, $P = 9.0 \times 10^{-5}$) near the *WNT4* gene on 1p36.12, that was reported to be suggestively associated in the Japanese (OR = 1.20, 95% CI 1.11–1.29, $P = 2.2 \times 10^{-6}$).

Encouraged by the *WNT4* association and with accumulating evidence for many complex traits that the number of discovered variants is strongly correlated with experimental sample size[6], we sought to increase the ratio of controls to cases in the Australian GWA cohort and to perform a formal meta-analysis of the Australian (QIMR), UK (OX) and Japanese (BBJ) GWA data.

To increase the power of the Australian GWA dataset we matched the existing QIMR cases and controls[2] on ancestry to individuals from the Hunter Community Study (HCS)[7]. After stringent quality control (QC), the combined QIMRHCS GWA cohort consisted of 2,262 endometriosis cases and 2,924 controls, increasing the number of controls by 1,054 and the Australian effective sample size by 24%. We also performed more stringent QC incorporating the OX dataset, resulting in a revised OX GWA cohort of 919 endometriosis cases and 5,151 controls. All cases in the QIMRHCS and OX studies have surgically confirmed endometriosis and disease stage from surgical records using the rAFS classification system[8], subjects are grouped into stage A (stage I or II disease or some ovarian disease with a few adhesions; $n = 1{,}680$, 52.8%) or stage B (stage III or IV disease; $n = 1{,}357$, 42.7%), or unknown ($n = 144$, 4.5%). Details of the final GWA and independent replication case-control cohorts are summarized in Table 1 and a schematic of our study design is provided in Fig. 1.

Meta-analysis of all endometriosis 4,604 cases and 9,393 controls for the 407,632 SNPs overlapping in the QIMRHCS, OX and BBJ GWA data, showed that the A allele of rs12700667 at the European 7p15.2 locus (OR = 1.22, 95% CI 1.13–1.31, $P = 7.2 \times 10^{-8}$) also replicates in the Japanese GWA data (OR = 1.22, 95% CI 1.07–1.39, $P = 3.6 \times 10^{-3}$), producing an overall OR of 1.22 (95% CI 1.14–1.30) and $P = 9.3 \times 10^{-10}$ in the GWA meta-analysis; we also confirmed association with allele A of rs7521902 at the 1p36.12 *WNT4* locus (OR = 1.18, 95% CI 1.11–1.25, $P = 4.6 \times 10^{-8}$) (Table 2).

The GWA meta-analysis identified a novel locus on 12q22 near the *VEZT* gene (allele C of rs10859871 OR = 1.18, 95% CI 1.12–1.25, $P = 5.5 \times 10^{-9}$). We also established association with allele G of rs13394619 in the *GREB1* gene on 2p25.1 (OR = 1.12, 95% CI 1.06–1.18, $P = 2.1 \times 10^{-5}$), previously reported (OR = 1.35, 95% CI 1.17–1.56, $P = 3.8 \times 10^{-5}$) in a small independent Japanese GWA study of 696 cases and 825 controls by Adachi et al (2010)[9]. The G allele of rs13394619 approached conventional genome-wide significance ($P \leq 5 \times 10^{-8}$) in combined analysis of the QIMRHCS, OX, BBJ, Adachi500K and Adachi6.0 GWA data (OR = 1.15, 95% CI 1.09–1.20, $P = 6.1 \times 10^{-8}$) (Table 2). In addition to the three genome-wide significant SNPs on chromosomes 1, 7 and 12 (rs7521902, rs12700667, rs10859871), the Manhattan plot of the all endometriosis GWA meta-analysis results (Supplementary Fig. 1) showed 34 SNPs reached genome-wide *suggestive* association ($P \leq 10^{-5}$).

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 the SNPs reaching genome-wide suggestive association, where the association results from QIMRHCS and OX Stage B cases versus controls, were meta-analyzed with the BBJ association results (stage information not available).

After excluding endometriosis cases with minimal (rAFS stage I-II) or unknown severity in the QIMRHCS and OX cohorts, GWA meta-analysis implicated novel loci on 2p14 (allele C of rs4141819 OR = 1.22, 95% CI 1.14–1.32, $P = 6.5 \times 10^{-8}$), 6p22.3 (allele T of rs7739264 OR = 1.21, 95% CI 1.13–1.30, $P = 5.8 \times 10^{-8}$) and 9p21.3 (allele C of rs1537377 OR = 1.22, 95% CI 1.14–1.30, $P = 1.0 \times 10^{-8}$) (Table 2, Supplementary Fig. 2, Supplementary Table 1-2 and Supplementary Note).

Annotated plots showing evidence for association in the combined QIMRHCS, OX and BBJ GWA data of genotyped SNPs across the seven implicated loci from the analysis of all cases and of stage B cases only are provided in Supplementary Figs. 3-9. Imputation up to the 1000 Genomes reference panel produced more significant *P* values and helped resolve the associated region at the 1p36.12 (rs56318008, $P_{all} = 1.3 \times 10^{-10}$), 2p25.1 (rs77294520, $P_{stageB} = 8.6 \times 10^{-8}$), 2p14 (rs2861694, $P_{stageB} = 7.9 \times 10^{-9}$), 6p22.3 (rs6901079, $P_{all} = 1.9 \times 10^{-8}$), 9p21.3 (rs7041895, $P_{stageB} = 5.1 \times 10^{-10}$) and 12q22 (rs11107968, $P_{all} = 3.9 \times 10^{-9}$) loci (Fig. 2 and Supplementary Figs. 10-16). Of particular note, the most significant imputed SNPs on 1p36.12, rs56318008 and rs3820282 ($P_{all} = 1.6 \times 10^{-10}$), are located 22 bp 5′ and *within* the *WNT4* gene, respectively.

Interestingly, the most associated genotyped SNP at 9p21.3 (rs1537377) is 55 kb centromeric to the genome-wide significant SNP reported in the original BBJ GWA[1] (rs10965235) located in the *CDKN2BAS* gene, and 49 kb 3′ to the transcription end site of *CDKN2BAS*. SNP rs10965235 is monomorphic in Caucasian populations and we investigated the independence of rs10965235 and rs1537377 in the BBJ GWA data. Firstly, in the BBJ GWA data, alleles of rs10965235 and rs1537377 are very weakly correlated, with linkage disequilibrium (LD) metrics of $r^2 = 0.028$ and $D' = 0.461$. Secondly, the allelic association *P* values for rs10965235 and rs1537377 are $P = 1.6 \times 10^{-4}$ and $P = 1.8 \times 10^{-2}$, respectively. After conditioning on rs10965235, weak residual association remains at rs1537377 ($P = 9.0 \times 10^{-2}$). Consequently, the data suggest there may be two independent genetic risk factors near the *CDKN2BAS* locus on 9p21.3. *CDKN2BAS* is a long non-coding RNA adjacent to and transcribed from the opposite strand to *CDKN2B* (p15), *CDKN2A* (p16) and *ARF* (p14). Loss of heterozygosity of *CDKN2A* and hypermethylation of the *CDKN2A* promoter have been reported in endometriosis[10,11].

To further validate the seven SNPs implicated by the meta-analysis, we carried out a replication study using a cohort of 1,044 cases and 4,017 controls obtained from the BioBank Japan independent of the BBJ GWA cohort. As shown in the forest plots of risk allele effects estimated using all cases versus controls (Fig. 3), the effects (ORs) were in the same direction for all seven implicated SNPs across the GWA and replication cohorts. With the exception of rs12700667, which was previously replicated ($P = 1.2 \times 10^{-3}$) in 2,392 cases and 2,271 controls from the US[2], and rs4141819 (with a marginal $P = 5.1 \times 10^{-2}$), all SNPs were replicated at the nominal $P < 0.05$ threshold (Table 2). All seven SNPs surpass the conventional genome-wide significant threshold of $P \leq 5 \times 10^{-8}$ after combined analysis of the GWA and replication cases and controls (Table 2). A conservative adjustment of the rs4141819 total *P* values ($P_{all} = 8.5 \times 10^{-8}$; $P_{stageB} = 4.1 \times 10^{-8}$) for performing two independent GWA studies (all and stage B endometriosis cases versus controls) would produce $P > 5 \times 10^{-8}$ ($P_{all\text{-}adjusted} = 1.7 \times 10^{-7}$; $P_{stageB\text{-}adjusted} = 8.2 \times 10^{-8}$). However, the accurately imputed (Rsq > 0.95) SNP rs2861694 ($P_{stageB} = 7.9 \times 10^{-9}$), in strong LD with rs4141819 ($r^2 = 0.981$, $D' = 1.0$; and $r^2 = 0.867$, $D' = 1.0$, in the 379 European and 286 Asian 1000 Genomes reference samples, respectively), would remain genome-wide significant ($P_{stageB\text{-}adjusted} = 1.6 \times 10^{-8}$).

The Q-Q plots for the QIMRHCS, OX and BBJ GWA data (Supplementary Fig. 17a-c) reflect our stringent quality control, while the GWA meta-analysis Q-Q plot (Supplementary