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

Translational Relevance

Lymphoblastoid cell lines (LCL) have been used in chemotherapeutic pharmacogenomic marker discovery due to their ease of experimental manipulation, extensive genotype catalogs, and lack of the *in vivo* confounders present in clinical samples. One important question is how well these cell-based models generate clinically relevant single-nucleotide polymorphisms (SNP) associated with patient toxicity. We compared genome-wide association study (GWAS) results of paclitaxel-induced cytotoxicity in LCLs and paclitaxel-induced peripheral neuropathy in patients with breast cancer. We observed significant overlap between the clinical and LCL studies, thus confirming a role for the LCL model in the analysis of at least a subset of genes involved in patient paclitaxel response. One overlap gene, *RFX2*, was functionally validated in a nerve cell model of paclitaxel response. Peripheral neuropathy is an often dose-limiting toxicity induced by paclitaxel treatment. If physicians could predict which patients are more likely to experience this severe toxicity, lower doses or alternative treatments could be prescribed.

candidate markers and genes discovered in both preclinical and clinical studies (14, 15). However, a critical question is how well this cell-based model generates clinically relevant markers and genes associated with patient response to drug. Recently, a few chemotherapeutic response single-nucleotide polymorphisms (SNP) discovered in LCLs have been replicated in patient populations by associating with phenotypes such as tumor response and overall survival in patients receiving the same drug (16–19); however, these studies tested the individual variants most associated with the LCL phenotypes. We sought to understand to what extent the overall genetic architecture of patient response to chemotherapy can be captured by LCLs by investigating beyond just the top few signals. In contrast to previous studies that tested single SNPs, we use an enrichment method (20) to determine in a systematic manner whether top genome-wide association study (GWAS) SNPs for paclitaxel-induced sensory peripheral neuropathy in patients with breast cancer (3) are more likely to also be paclitaxel-induced cytotoxicity SNPs identified in LCLs than expected by chance.

In this study, we found that SNPs associated with patient paclitaxel-induced neuropathy are enriched for SNPs associated with paclitaxel-induced cytotoxicity in HapMap LCLs. This significant enrichment confirms that LCLs are a useful model in the study of a subset of shared genes involved in patient toxicity. The overlap SNPs are predominantly expression quantitative trait loci (eQTL) as defined previously (21), therefore supporting an enriched functional role for these significant SNPs. We show a functional role for one eQTL host gene (*RFX2*) in paclitaxel toxicity, using a cellular model of peripheral neuropathy. These results are consistent with the hypothesis that the cell-based models capture components of the underlying genetic architecture for paclitaxel-induced sensory peripheral neuropathy.

Materials and Methods

Cytotoxicity assays

HapMap LCLs from a population with Northern and Western European ancestry from Utah (HAPMAPPT01, CEU, $n = 77$), a Yoruba population in Ibadan, Nigeria (HAPMAPPT03, YRI, $n = 87$), and an African-American population from the Southwest of the United States (HAPMAPPT07, ASW, $n = 83$) were treated with 12.5 nmol/L paclitaxel and cytotoxicity was determined using an AlamarBlue (Invitrogen) cellular growth inhibition assay as described (22). The cytotoxicity phenotype used in the LCL GWAS was mean percentage survival at 12.5 nmol/L paclitaxel determined from 6 replicates from 2 independent experiments. Percentage survival values for each cell line were $\log_2$-transformed before statistical analysis to form an approximately normal distribution in each population.

LCL genome-wide meta-analysis

A GWAS of paclitaxel-induced cytotoxicity was conducted on each of the 3 populations separately. Greater than 2 million SNPs from HapMap r27 [minor allele frequency (MAF) > 0.05 within the panel, no Mendelian errors and in Hardy–Weinberg equilibrium ($P > 0.001$)] were tested for association with paclitaxel cytotoxicity in each population, using the quantitative trait disequilibrium test total association model (23). To control for population structure in the admixed ASW population, local ancestry at each genotyped SNP locus was estimated using HAPMIX (24) and to increase genome coverage of the ASW, ungenotyped markers were imputed using BEAGLE (25) as previously described (26). Genomic control lambda (λ_{GC}) values (27) were calculated for the GWAS of each population. Studies with λ_{GC} values greater than 1 were corrected for residual inflation of the test statistic by dividing the observed test statistic at each SNP by the λ_{GC} (27), and then the corresponding P values were carried through the meta-analysis.

Using the software METAL, we combined SNP P values across the 3 population studies, taking into account a study-specific weight (sample size) and direction of effect (positive or negative β; ref. 28). This approach converted the direction of effect and P value observed in each study into a signed Z-score, such that very negative Z-scores indicate a small P value and an allele associated with higher drug sensitivity, whereas large positive Z-scores indicate a small P value and an allele associated with higher drug resistance. Z-scores for each SNP were combined across studies in a weighted sum, with weights proportional to the square-root of the sample size for each study (28).

Patient samples and GWAS

Cancer and Leukemia Group B (CALGB) 40101 is a phase III trial comparing the efficacy of standard therapy cyclophosphamide and doxorubicin with single-agent paclitaxel

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

as adjuvant therapy for breast cancer in women with 0 to 3 positive axillary lymph nodes. All study participants were enrolled in CALGB 40101 and gave their additional consent to participate in the pharmacogenetic companion study (CALGB 60202), which has been published (3). All patient research met state, federal, and Institutional Review Board guidelines. Germline DNA was isolated from 1,040 patients on the paclitaxel arm of CALGB 40101 and genotyped using the Illumina 610-Quad platform as described previously (3). Following quality control analysis, genotypes were available for 520,679 SNPs. Principal component (PC) analysis identified 855 genetic Europeans that were used in a GWAS of sensory peripheral neuropathy (3). A dose-to-event analysis was conducted, with an event defined as grade 2 or greater sensory peripheral neuropathy. The Cox score test, powered for additive genetic effects, was used to test these marginal associations. Only SNPs with MAFs more than 0.05 in the patient population and in Hardy–Weinberg equilibrium in the CEU ($P > 0.001$) were used in the LCL GWAS comparisons.

Enrichment analysis

We conducted a permutation resampling analysis (29) to test for an enrichment of cytotoxicity-associated SNPs (LCLs) among the paclitaxel-induced sensory peripheral neuropathy-associated SNPs (patients). To this end, the patient outcomes (cumulative dose and event indicator vectors) were randomly shuffled while keeping the genotype data fixed to preserve linkage disequilibrium. On the basis of this permutation replicate, the standardized Cox score statistics were recalculated for all the SNPs. This process was conducted 1,000 times. For each of the 1,000 permutation replicates, the number of SNPs that had $P < 0.05$ in the patient data, $P < 0.001$ in the LCL data, and the same direction of effect (the same allele associated with increased neuropathy and increased cytotoxicity) was calculated. The overlap distribution from the permutations was compared with the observed SNP overlap to generate an empirical P value, calculated as the proportion of permutations in which the number of LCL/patient overlap SNPs exceeds the observed number. To test the robustness of our findings, we calculated an empirical P value across a range of inclusion thresholds from $P < 0.001$ to $P < 0.1$. We also tested for enrichment of patient SNPs among the LCL SNPs by generating 1,000 randomized SNP sets the same size and MAF distribution as the observed LCL data at a range of P value thresholds to calculate empirical P values. In addition to the paclitaxel LCL cytotoxicity data, we compared the patient sensory peripheral neuropathy data with LCL cytotoxicity GWAS data from capecitabine (30) and carboplatin (13) as negative controls.

To test for eQTL enrichment in the LCL, patient, and LCL/patient overlap SNPs, we generated 10,000 randomized SNP sets each of the same size as the observed set of LCL cytotoxicity ($P < 0.001$), patient neuropathy ($P < 0.05$), or LCL/patient overlap SNPs. The randomized SNP sets were matched on MAF distribution of the observed list and sampled (without replacement) from the set of SNPs on the Illumina 610-Quad platform, similar to the method of Gamazon and colleagues (31). We grouped the platform SNPs into discrete MAF bins of a width of 5%, from which the SNPs used in the simulations were selected. For each of the 10,000 sets, we determined the number of eQTLs ($P < 10^{-4}$) and calculated an empirical P value for enrichment. The eQTLs were defined previously and are available in the SCAN database (21, 31).

Filtering procedure for functional analysis

First, we determined which of the LCL/patient overlap SNPs from the enrichment analysis were located in or near (within 2 kb) gene transcripts (dbSNP build 129, human genome assembly build 36). Eleven of 24 overlap SNPs were in or near genes and genotyping intensity plots for these SNPs in the patient data are available in Supplementary Fig. S1. Second, we determined which SNPs within genes were also eQTLs (31) and prioritized by which had the most target genes ($P < 10^{-4}$). We also tested whether the expression of the eQTL target genes associated with paclitaxel-induced cytotoxicity ($P < 0.05$) using previously published exon array data (32). A general linear model was constructed between gene expression and paclitaxel-induced cytotoxicity with growth rate (33) and population as covariates. A Toeplitz covariance structure with 2 diagonal bands was used to allow for familial dependencies in the data as previously described (9).

siRNA

Neuroscreen-1 (NS-1) rat pheochromacytoma cells (Cellomics Inc.) were maintained in NS-1 media (RPMI supplemented with 10% horse serum, 5% fetal calf serum and 1% L-glutamine). Cells were seeded at a density of 1×10^5 cells/mL on collagen I–coated plates and induced to differentiate by adding 20 ng/mL nerve growth factor (NGF, BD Biosciences) to the media 24 hours before transfection. Cells for cytotoxicity assays were plated in 96-well collagen I–coated plates, whereas cells for expression quantification and neurite-outgrowth assays were plated in 6-well collagen I–coated plates. Pooled *Rfx2* siRNA (25 nmol/L; Qiagen; S101639659, S101639666, S101639673, and S101639680) or nontargeting control siRNA (Qiagen; 1027292) was transiently transfected into the NS-1 cells using DharmaFECT Reagent #1 (Dharmacon). Quantitative reverse transcription PCR (qRT-PCR) was conducted for *Rfx2* (Rn00501380_m1) and control gene *Gapdh* (4352338E) using TaqMan Gene Expression Assays (Applied Biosystems) 24 hours posttransfection in the neurite-outgrowth assays and 24, 48, 72, and 96 hours posttransfection in the cytotoxicity assays to assess *Rfx2* knockdown in NS-1 cells. Expression of the potential *Rfx2* target genes *Cyp51* (Rn01526553_m1), *Bach1* (Rn01477344_m1), and *Cbara1* (Rn01644475_m1) was also measured by qRT-PCR at 24 hours post-*siRfx2* transfection. Each qRT-PCR was run in duplicate and individual samples were run in triplicate on each plate. Percentage knockdown was calculated by dividing the relative *Rfx2* expression levels in the *siRfx2* sample by those in the nontargeting control sample.

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

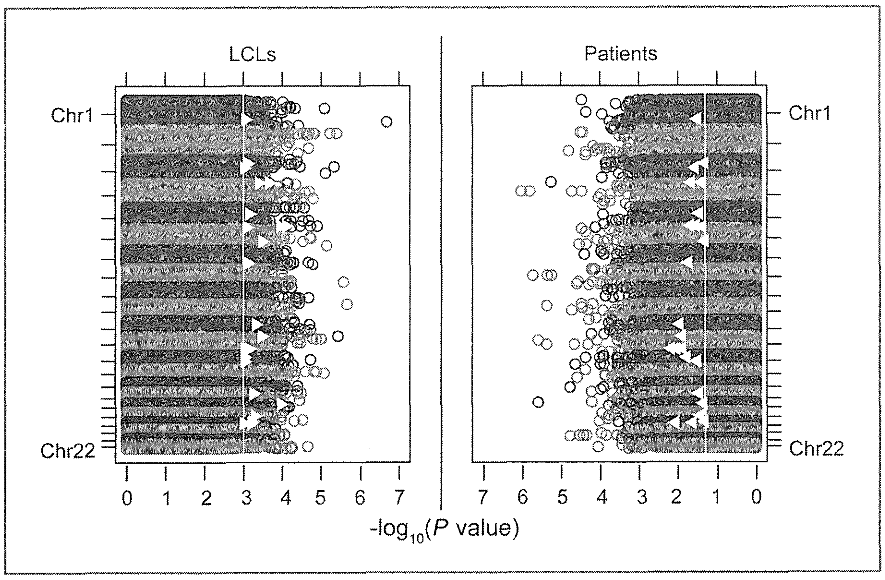

Figure 1. Comparison of individual GWAS results. Left, paclitaxel-induced cytotoxicity in LCLs. Right, paclitaxel-induced sensory peripheral neuropathy in patients. White lines represent the overlap thresholds used in the primary enrichment analysis ($P < 0.001$ for LCLs and $P < 0.05$ for patients)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

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

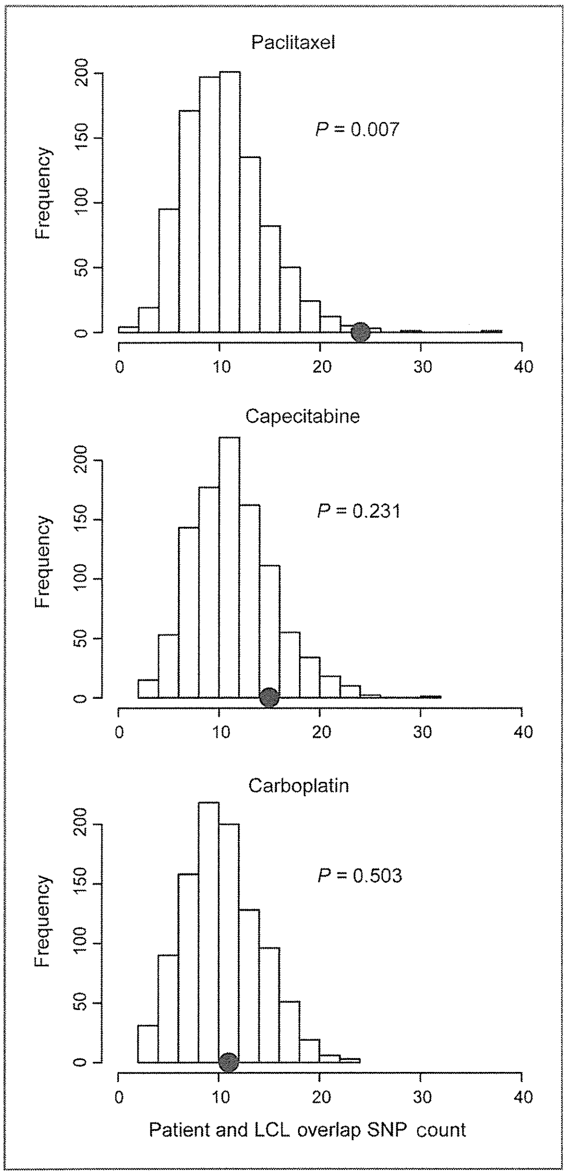

Figure 2. Patient paclitaxel-induced sensory peripheral neuropathy SNPs are enriched for SNPs associated with paclitaxel-induced cytotoxicity in LCLs. Distribution of chemotherapeutic-induced cytotoxicity SNP ($P < 0.001$) count in 1,000 permutations of neuropathy phenotype–genotype connections ($P < 0.05$). 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

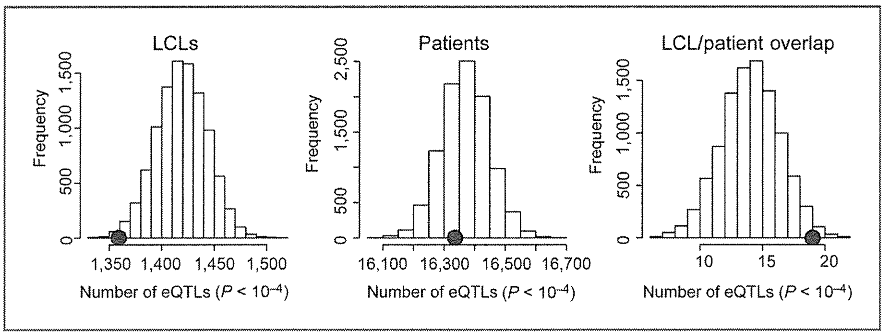

Figure 3. SNPs associated with both patient paclitaxel-induced sensory peripheral neuropathy and LCL paclitaxel-induced cytotoxicity are enriched for eQTLs. Distribution of eQTL ($P < 10^{-4}$) count in 10,000 simulations, each matching the MAF distribution of either LCL paclitaxel SNPs ($P < 0.001$), 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 $P = 0.0447$).

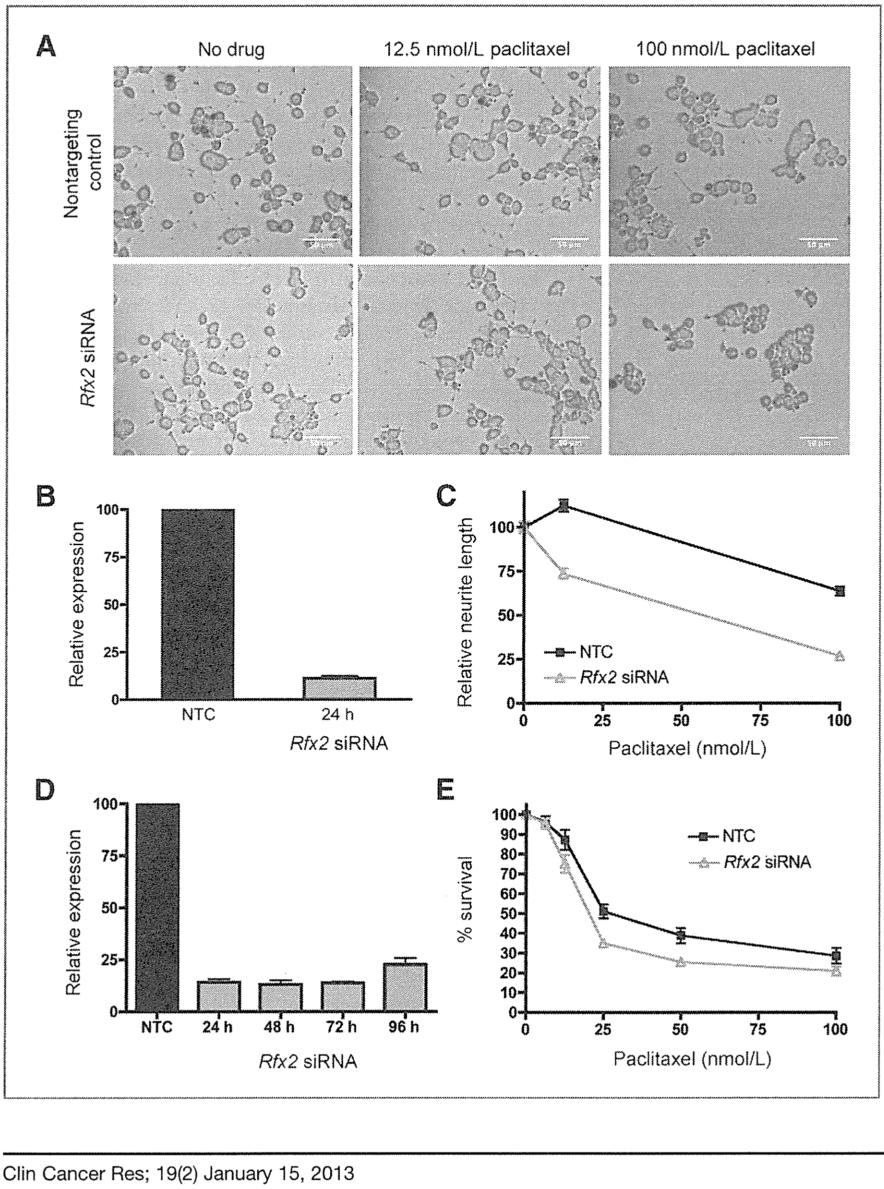

Figure 4. Functional validation of *RFX2* in paclitaxel response using a peripheral neuropathy cell model. A, representative micrographs comparing neurite lengths of NS-1 cells upon siRNA knockdown of *Rfx2* and treatment with paclitaxel (×10 phase-contrast). B, relative gene expression 24 hours posttransfection 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.

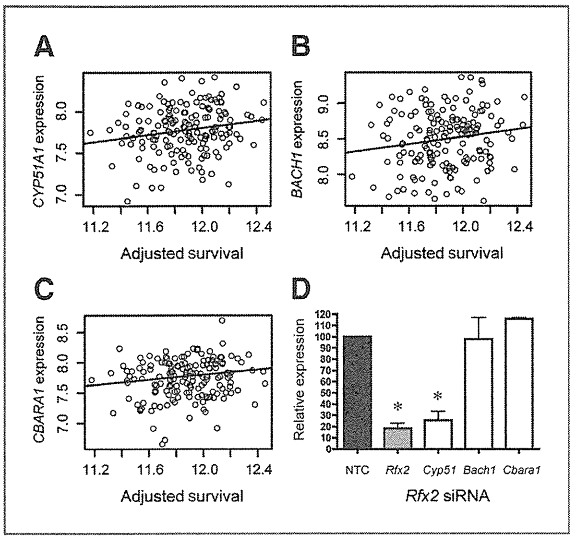

Figure 5. Target genes of *RFX2* eQTL rs7254081 in paclitaxel response. Increased baseline expression of the rs7254081 target genes (A) *CYP51A1*, (B) *BACH1*, and (C) *CBARA1* associate with increased cellular survival (adjusted for growth rate) of LCLs treated with 12.5 nmol/L paclitaxel ($P < 0.05$). 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;

[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 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, $P_{\text{all}} = 1.3 \times 10^{-10}$), 2p25.1 (rs77294520, $P_{\text{stageB}} = 8.6 \times 10^{-8}$), 2p14 (rs2861694, $P_{\text{stageB}} = 7.9 \times 10^{-9}$), 6p22.3 (rs6901079, $P_{\text{all}} = 1.9 \times 10^{-8}$), 9p21.3 (rs7041895, $P_{\text{stageB}} = 5.1 \times 10^{-10}$) and 12q22 (rs11107968, $P_{\text{all}} = 3.9 \times 10^{-9}$) loci (Fig. 2 and Supplementary Figs. 10-16). Of particular note, the most significant imputed SNPs on 1p36.12, rs56318008 and rs3820282 ($P_{\text{all}} = 1.6 \times 10^{-10}$), 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 ($P_{\text{all}} = 8.5 \times 10^{-8}$; $P_{\text{stageB}} = 4.1 \times 10^{-8}$) for performing two independent GWA studies (all and stage B endometriosis cases versus controls) would produce $P > 5 \times 10^{-8}$ ($P_{\text{all-adjusted}} = 1.7 \times 10^{-7}$; $P_{\text{stageB-adjusted}} = 8.2 \times 10^{-8}$). However, the accurately imputed (Rsq > 0.95) SNP rs2861694 ($P_{\text{stageB}} = 7.9 \times 10^{-9}$), 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 ($P_{\text{stageB-adjusted}} = 1.6 \times 10^{-8}$).

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

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

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 $N_{\mathrm{eff}} = 4 / (1 / N_{\mathrm{cases}} + 1 / N_{\mathrm{controls}})$[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]

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]

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.

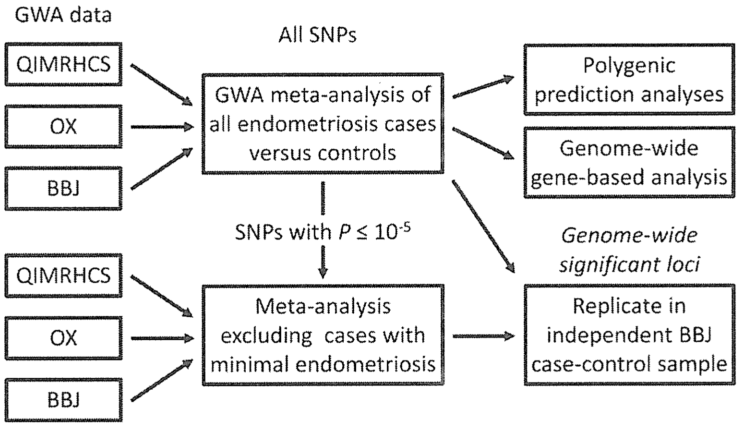
GWA data
All SNPs
QIMRHCS
OX
BBJ
GWA meta-analysis of all endometriosis cases versus controls
Polygenic prediction analyses
Genome-wide gene-based analysis
SNPs with $P \leq 10^{-5}$
Genome-wide significant loci
QIMRHCS
OX
BBJ
Meta-analysis excluding cases with minimal endometriosis
Replicate in independent BBJ case-control sample

Figure 1.
Study design.