Pharmacogenomic Discovery to Function and Mechanism: Breast Cancer as a Case Study

Liewei Wang[1], James Ingle[2] and Richard Weinshilboum[1]

Biomedical research is undergoing rapid change, with the development of a series of analytical omics techniques that are capable of generating Biomedical Big Data. These developments provide an unprecedented opportunity to gain novel insight into disease pathophysiology and mechanisms of drug action and response—but they also present significant challenges. Pharmacogenomics is a discipline within Clinical Pharmacology that has been at the forefront in defining, taking advantage of, and dealing with the opportunities and challenges of this aspect of the Post-Genome Project world. This overview will describe the evolution of germline pharmacogenomic research strategies as we have moved from an era of candidate genes to agnostic genome-wide association studies (GWAS) coupled with the functional and mechanistic pursuit of GWAS signals. Germline pharmacogenomic studies of breast cancer endocrine therapy will be used to illustrate research strategies that are being applied broadly to omics studies of drug response phenotypes.

Biomedical research is in the midst of a period of rapid change, with the incorporation of a variety of omics technologies, the generation of extremely large datasets that result from the application of these analytical methods, and the need for novel computational approaches to deal with these large datasets. Pharmacogenomics represents one discipline within Clinical Pharmacology that has benefited significantly from these advances, advances that now make it possible to scan across the entire genome to identify genes associated with variation in drug response phenotypes and which will ultimately make it possible to sequence the entire genome of every patient being studied. As a result of these rapid technical developments, pharmacogenetics, a discipline that originated over half a century ago,[1] and which initially focused on candidate genes that encode drug-metabolizing enzymes, drug transporters, or drug targets,[2,3] has evolved during the past decade into pharmacogenomics, with genome-wide association studies (GWAS) that have identified genes that influence drug response with unfamiliar names such as *ZNF423*, *MIR2052HG*, and *TCL1A*.[4–6] Those genes had not previously been associated with drug effect. However, the identification of sequence or structural variants in novel genes associated with variation in drug response is only the first step in a process that has reversed the standard pharmacogenomic approach that was applied only a few years ago. During the pre-Genome Project era, we knew that phase I and phase II enzymes catalyzed the biotransformation of drugs, so we cloned and sequenced genes encoding drug metabolizing enzymes to determine whether variation in DNA sequence within or near those genes might be associated with variation in drug response.[3,7–11] Today, we are able to use GWAS or next-generation DNA sequencing (NGS) to discover unanticipated genes or DNA sequence variants that contribute to variation in drug response, but it is then necessary to pursue the underlying function of those variants and genes as well as mechanisms responsible for their association with drug response phenotypes. This reverse strategy, as illustrated subsequently by the results of germline pharmacogenomic studies of the endocrine therapy of breast cancer, can lead to novel insight into function and mechanisms that will facilitate the achievement of true precision medicine, either by enabling better selection of patients for a given therapy or by identifying new therapeutic targets. These principles and approaches will be illustrated by the results of a series of GWAS studies of the pharmacogenomics of the endocrine therapy of estrogen receptor positive (ER+) breast cancer. The focus in subsequent paragraphs will be on GWAS using the germline genome, although, obviously, the tumor somatic genome is also an area of intense study in breast cancer and many candidate gene studies have also been performed.[12] Clearly, crosstalk between these two related genomes also contributes to variation in drug response. Therefore, the examples described here represents only one facet of genome-wide pharmacogenomic discovery; in the case of the examples discussed subsequently, always followed by functional validation, mechanistic pursuit, and, eventually, clinical translation. Finally, we always need to bear in mind the fact that pharmacogenomic discovery, translation, and implementation are intimately interrelated processes, with each dependent on the other two (**Figure 1**).

[1]Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota, USA; [2]Division of Medical Oncology, Department of Oncology, Mayo Clinic, Rochester, Minnesota, USA. Correspondence: R. Weinshilboum (Weinshilboum.richard@mayo.edu)

Received 25 August 2017; accepted 7 October 2017; advance online publication 1 December 2017. doi:10.1002/cpt.915

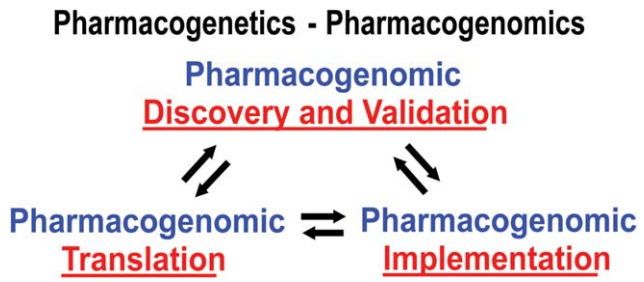

Figure 1 Pharmacogenomics discovery and validation, translation and implementation. The figure illustrates the fact that these components of pharmacogenomic research are interdependent and that each serves to inform the other two.

Breast cancer is the number one invasive cancer of women worldwide.[13,14] In spite of major advances that have been made in the treatment of breast cancer, over 40,000 women die each year in the United States alone as a result of this disease (http://ww5.komen.org/BreastCancer/Statistics.html). A major advance in both the treatment and the chemoprevention of breast cancer was the realization that the growth and origin of many of these tumors is driven, at least in part, by estrogens.[15,16] That insight was followed by the development and application to the treatment and chemoprevention of breast cancer of selective estrogen receptor modulators (SERMs) such as tamoxifen and, subsequently, of aromatase inhibitors (AIs), drugs that inhibit the biosynthesis of estrogens.[15,16] Both SERMs and AIs dramatically reduce the recurrence of ER+ tumors after surgery,[15,16] as well as the occurrence of breast cancer in women with increased risk for this disease.[17–20] However, there are large individual differences in clinical response to SERM and AI therapy, both with regard to therapeutic efficacy, i.e., prevention of either the occurrence or recurrence of breast cancer, and the frequency of adverse drug reactions. A portion of that variation is determined by inheritance, often as a result of common germline genetic polymorphisms. During the past decade, the development of genome-wide genotyping has made it possible to perform hypothesis-generating genome-wide discovery studies that have broadened our understanding of the role of genomics in breast cancer risk, breast cancer chemoprevention, the efficacy of drugs used to treat ER+ breast cancer, and the occurrence of adverse responses to therapy with those drugs. Subsequent paragraphs briefly summarize the application of GWAS to identify genes associated with response to drugs used to prevent or treat breast cancer, always followed by functional validation and the mechanistic pursuit of GWAS signals. Finally, this type of research requires significant funding, and the examples described below were funded by the US NIH NIGMS through the Pharmacogenomics Research Network (PGRN) and involved both national and international collaboration, in the case of the studies described below, collaboration with the RIKEN Center for Integrative Medical Sciences in Japan was particularly important.

BREAST CANCER SERM CHEMOPREVENTION GWAS

Two different classes of drugs have demonstrated efficacy in the chemoprevention of breast cancer: SERMs and AIs.[17–20] The US Food and Drug Administration (FDA) has approved the use of tamoxifen and raloxifene for breast cancer chemoprevention. However, even though in both the US and the UK public health authorities have recommended SERM chemoprevention, very few subjects are being treated with these drugs for that purpose. Included among the reasons are the fact that ~50 patients need to be treated to prevent one case of breast cancer and the occurrence of extremely rare, life-threatening side effects.[17–19] Therefore, reliable biomarkers to identify the patients most likely to benefit from SERM chemoprevention are needed. Beyond SERMs, AIs have also been shown to prevent breast cancer. The Canadian Cancer Trials Group MAP.3 study included 4,560 women at moderate risk for breast cancer and reported a 65% reduction in the occurrence of invasive breast cancer when compared to placebo after 3 years of treatment with the steroidal AI exemestane.[20] However, the drugs most commonly used for breast cancer chemoprevention are the SERMs, and the most compelling evidence in support of their efficacy came from the National Surgical Adjuvant Breast and Bowel Project (NSABP) double-blind, placebo-controlled P-1 trial of tamoxifen and the double-blind P-2 comparison of tamoxifen with raloxifene.[17–19] These two studies included over 33,000 women, and their results were the major basis for FDA approval of these two drugs for breast cancer prevention. Therefore, we used samples from women who participated in P-1 and P-2 to perform the GWAS described subsequently.

A nested matched case–control GWAS for P-1 and P-2 that included 526 cases, i.e., women who developed breast cancer while on SERM chemoprevention therapy, and 1,176 matched controls who did not was published in 2013.[4] The top hit single nucleotide polymorphism (SNP) signals in these women were in the *ZNF423* gene on chromosome 16 (rs8060157, $P = 1.11E\text{-}06$) and near the *CTSO* gene on chromosome 4 (rs15701923, $P = 8.49E\text{-}07$) (**Figure 2**). These SNP signals indicated that these genes might be associated with risk for the occurrence of breast cancer in spite of up to 5 years of SERM preventive therapy. Neither of the top SNPs in these two signals was genome-wide significant but, since P-1 and P-2 were the largest SERM chemoprevention trials ever

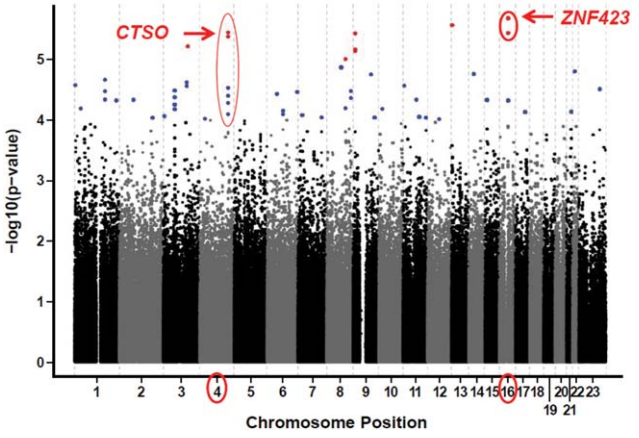

Figure 2 P-1 and P-2 SERM chemoprevention GWAS Manhattan plot of *P* values for conditional logistic regression adjusted for nine eigenvectors. Black: $P \geq 1E\text{-}04$, blue: $P < 1E\text{-}04$ to $1E\text{-}05$, red: $P < 1E\text{-}05$. Figure republished with permission from ref. 4.

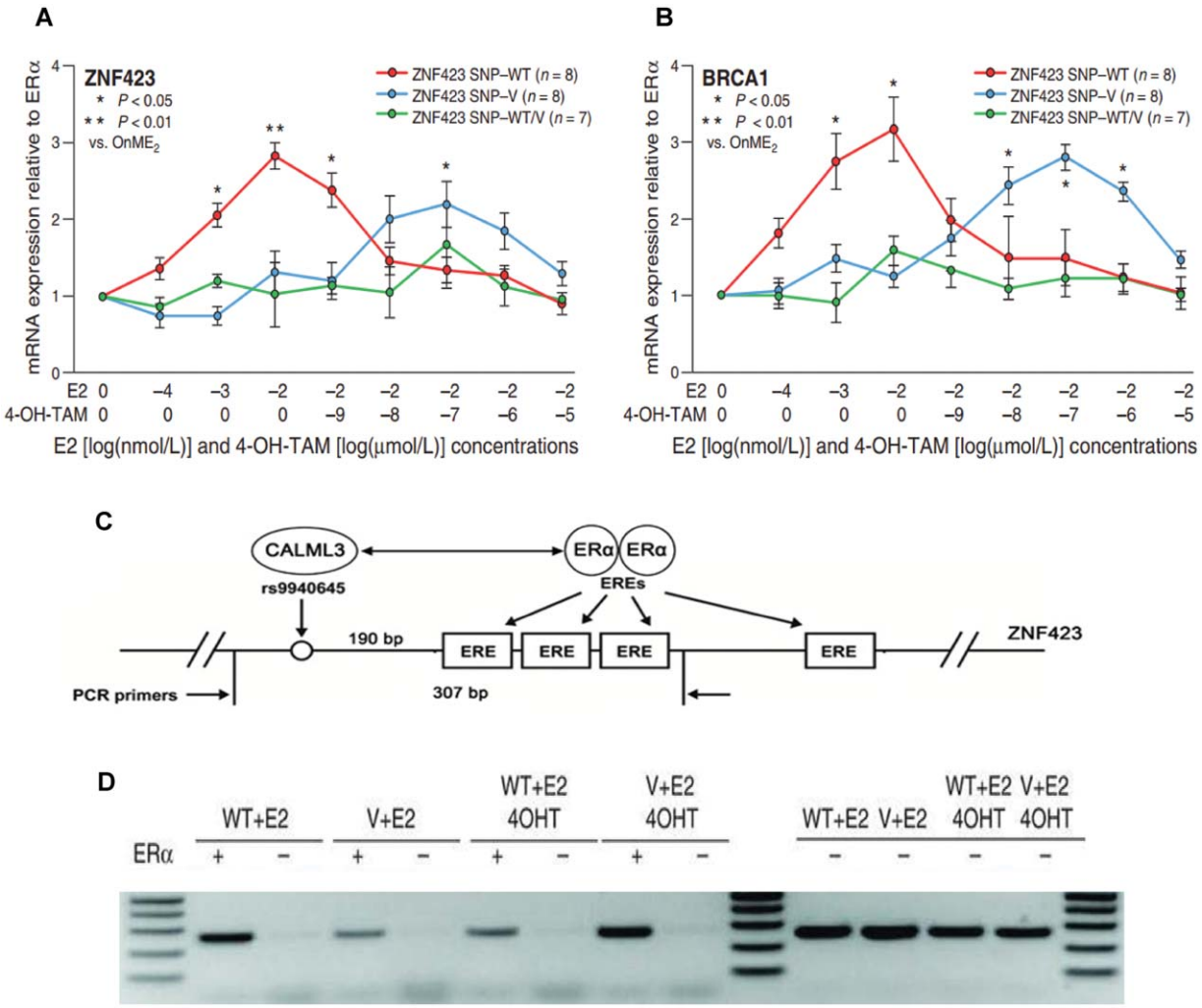

Figure 3 mRNA expression for (**a**) ZNF423 and (**b**) BRCA1 for lymphoblastoid cell lines with WT/WT (8 cell lines), WT/V (7 cell lines), and V/V (8 cell lines) genotypes for the chromosome 16 ZNF423 SNPs after exposure to E2 alone or E2 with increasing concentrations of 4-hydroxytamoxifen (4-OH-TAM). Error bars represent SEM. (**c**), Schematic depiction of the ZNF423 intron 2-area near the rs9940645 SNP. The locations of EREs are shown as boxes, and arrows show the locations of primers used to conduct the ChIP amplifications. (**d**) ChIP assays for the area of ZNF423 that contains the rs9940645 SNP. WT, wildtype; V, variant. Modified from Ref. 4.

conducted, and since they included ~60% of all such samples worldwide, we chose to pursue the functional implications of the SNP signals, a decision that led to strong evidence in support of their biological plausibility based on functional genomic studies. Perhaps of equal importance, the mechanism of action for the *ZNF423* SNPs served to highlight a novel SNP effect that was subsequently shown to have implications that extended well beyond breast cancer.

Specifically, functional genomic studies demonstrated that both *ZNF423* and *CTSO* were estrogen inducible in an SNP-dependent fashion and, in both cases, *BRCA1* expression was induced downstream in parallel with the induction of these genes, also in an SNP-dependent fashion.[4] These results indicated that the two top SNP signals observed in the GWAS might be related to individual variation in the expression of *BRCA1*, a gene known to play a role in breast cancer risk. It was also demonstrated experimentally that ZNF423 was a transcription factor for *BRCA1*. Finally, by using a "Human Variation Panel" of lymphoblastoid cell lines (LCLs) from 300 different individuals of three ethnic groups (100 each) for which dense genome-wide genomic data were available, it was possible to test the effect of treatment with 4-hydroxytamoxifen (4-OH-TAM), an active metabolite of tamoxifen, on the induction of both *ZNF423* and *BRCA1* in LCLs with differing genotypes for the *ZNF423* SNPs. As shown graphically in **Figure 3a,b**, when LCLs with wildtype (WT) genotypes for the *ZNF423* SNPs were exposed to increasing concentrations of estradiol (E2), as anticipated, there was an SNP genotype-dependent induction of ZNF423 mRNA expression and, in parallel, of that of BRCA1. It should be emphasized that the SNPs were in *ZNF423*, not in the *BRCA1* gene. However, when increasing concentrations of 4-OH-TAM were added to E2 to mimic the clinical situation, the SNP genotype dependence of the induction pattern reversed. In that case, it was the variant rather than the WT *ZNF423* genotype that was

associated with the induction of both ZNF423 and BRCA1, while cell lines homozygous for the WT genotype returned to baseline levels of expression (**Figure 3a,b**). It should be pointed out that the variant SNP genotype was protective during 5 years of SERM prevention therapy, compatible with the results shown in **Figure 3a,b**. Furthermore, the *ZNF423* SNPs mapped to intron 2 of that gene, an intron that contained a series of estrogen response elements (EREs), DNA sequences known to bind the ERα dimer (see **Figure 3c**). Therefore, chromatin immunoprecipitation (ChIP) assays for ERα binding to those ERE motifs were performed that demonstrated that the binding of ERα to EREs near the rs9940645 SNP (one of the top SNPs in the *ZNF423* "signal," $P = 1.30E-06$) displayed a pattern of binding that matched the reversal of the expression pattern shown in **Figure 3a**, i.e., the binding pattern of ERα to the EREs reversed in the presence of 4-OH-TAM even though the SNP was located 190 bp away from the EREs (see **Figure 3d**). In a subsequent study designed to determine whether this behavior was unique or might be more general, next-generation DNA sequencing was performed across both the *CTSO* and *ZNF423* genes and that study identified an additional pair of SNPs in intron 5 of the *ZNF423* gene, rs12918288 and rs746157, that were located 196 and 401 bp, respectively, from an ERE that behaved exactly as did the rs9940645 SNP in intron 2, i.e., there was a genotype and E2-dependent induction of ZNF423 expression associated with these SNPs that was mirrored by ER binding as determined by ChIP assay and which could be reversed with 4-OH-TAM.[21] The *ZNF423* intron 5 SNPs mapped over 250 kb distant from the intron 2 SNPs, and all of the studies of the intron 5 SNPs used samples that were homozygous WT for the SNPs in intron 2. These results indicated that SNPs at a distance from transcription factor binding motifs, i.e., response elements, can have profound effects on the ultimate molecular phenotype in the presence of different drugs or other compounds, a conclusion that has potential implications for future pharmacogenomic studies.

The results of the experiments described in the preceding paragraph raised the question of how SNPs at a distance from a transcription factor binding site like an ERE might influence transcription factor binding and subsequent mRNA expression. Recently, in follow-up experiments designed to pursue this question with regard to the common *ZNF423* intron 2 SNP, it was reported that a "sensor protein," CALML3, is a component of a protein complex that includes ERα and that CALML3 is required to sense the presence of the SNP located 190 bp from the EREs, a situation that is depicted graphically in **Figure 3c**.[22] Furthermore, if CALML3 was knocked down, the SNP-dependent reversal phenomenon shown in **Figure 3a,b** was abolished.

The observations described in preceding paragraphs have both clinical and mechanistic implications. Although very few subjects are currently being treated with SERMs for breast cancer chemoprevention, this series of studies points to a strategy based on replicated biomarkers for response by which we might increase the number of women treated with SERMs for chemoprevention and also decrease the number of women exposed to the rare but serious side effects of these drugs, i.e., deep vein thrombosis with pulmonary emboli and endometrial carcinoma.[17–19] From a mechanistic perspective, these experiments demonstrated that SNP and drug-dependent transcription regulation through differential binding to specific genomic regions, regions that might be hundreds of bp from known binding motifs like EREs, can have dramatic effects on downstream transcription and drug response phenotypes. As a result, these observations will have to be considered in future pharmacogenomic studies, particularly when determining SNP effects on gene expression, not only at baseline but also in the presence of drug therapy. These points will be emphasized subsequently when we describe the results of a GWAS designed to study musculoskeletal adverse responses to AI therapy.

BREAST CANCER AROMATASE INHIBITOR THERAPY GWAS

Over the past three decades, endocrine therapy has been the standard of care for the adjuvant therapy of ER+ breast cancer. Approximately 70% of breast cancers in the United States are hormone receptor-positive, and the majority of those tumors occur in postmenopausal women.[23] Initially, these women were treated with tamoxifen, which reduced the rate of recurrence by ~50%. However, during the past decade increasing evidence has accumulated that shows an advantage for AI adjuvant therapy, and today the majority of postmenopausal women with ER+ breast cancer are treated with one of the third-generation AIs, anastrozole, letrozole, or exemestane,[24] drugs which also reduce recurrence rates by ~50%. It would be a major advance if we had biomarkers that would allow us to identify breast cancer patients who will or will not respond to adjuvant AI therapy. That was one of the goals as we set out to perform a GWAS using samples obtained from postmenopausal women with ER+ breast cancer who had been treated with AI therapy in the adjuvant setting, i.e., after the surgical resection of their tumors.[5]

The Canadian Cancer Trials Group MA.27 trial is the largest adjuvant endocrine therapy trial focused exclusively on AI therapy.[25] MA.27 enrolled 7,576 women in the United States, Canada, and Europe in an effort to compare outcomes between anastrozole and the steroidal AI exemestane. The results showed no differences between these two AIs after 5 years of adjuvant therapy.[25] A GWAS was performed for breast cancer recurrence among MA.27 patients using only samples from North American subjects. Of the 6,827 North American subjects included in the trial, 5,221 (76.5%) donated blood samples and provided consent for genomic testing.[6] The Manhattan plot for that GWAS showed an SNP signal on chromosome 8 with *P* values for the top SNPs that ranged from 2.15E-07 to 6.24E-07 (see **Figure 4a** for the Manhattan plot and **Figure 4b** for a regional plot of the chromosome 8 SNP signal). Variant genotypes for those SNPs were associated with decreased risk for breast cancer recurrence. Specifically, subjects heterozygous or homozygous for the variant alleles exhibited an ~40% or 63% reduction in risk for disease recurrence, respectively. Once again, even though the *P* values for these SNPs were not genome-wide significant, we pursued their functional and mechanistic implications because of the importance of the phenotype for ER+ breast cancer patients and because of the very large size of the MA.27 study.[25]

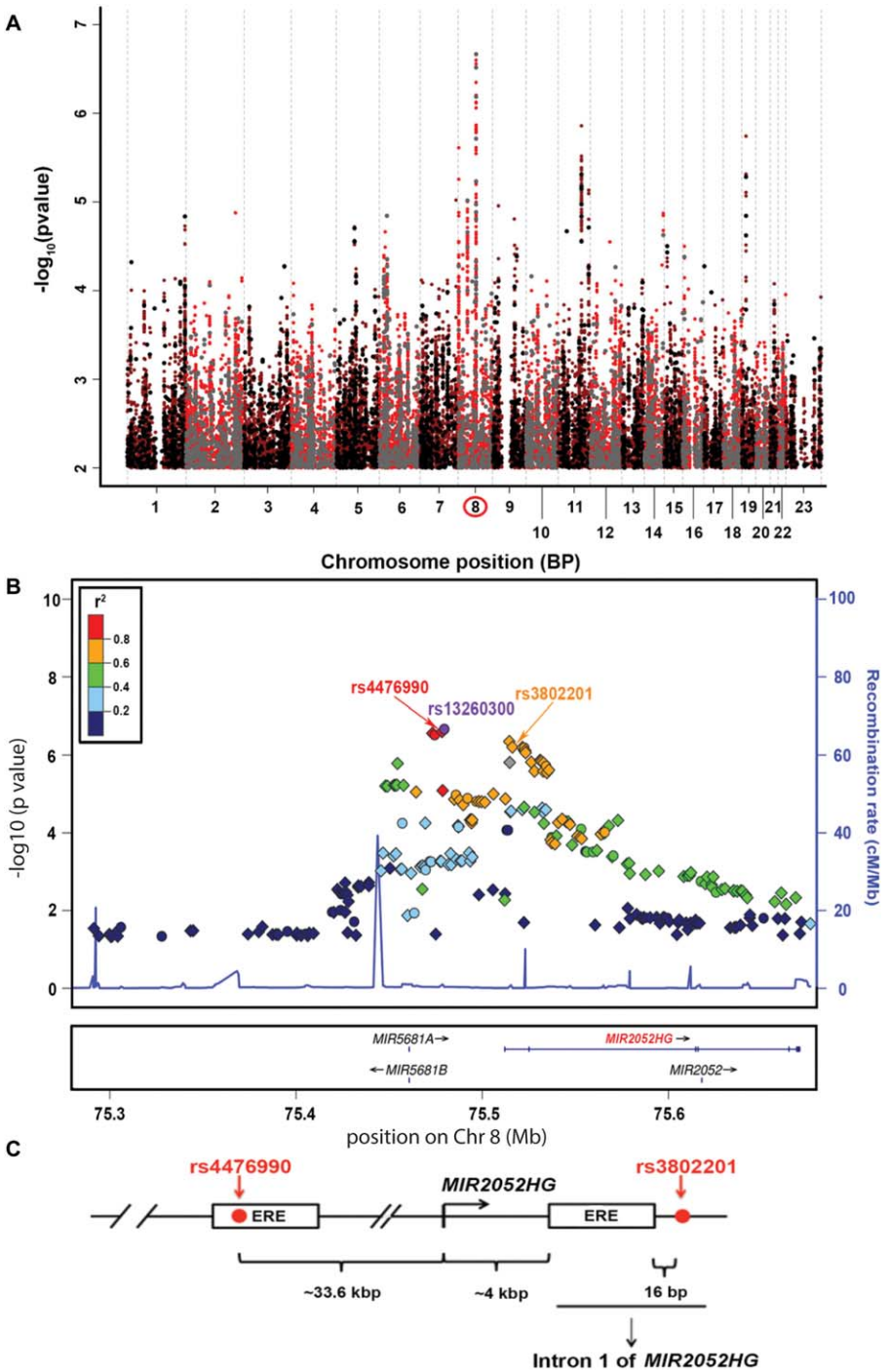

Figure 4 (**a**) Manhattan plot for the MA.27 GWAS for breast cancer recurrence. (**b**) Regional plot of the chromosome 8 region surrounding the *MIR2052HG* gene. (**c**) Schematic depiction of EREs near the rs4476990 and rs3802201 SNPs. The EREs are indicted as boxes and the SNPs as red circles. Modified from Ref. 5.

The chromosome 8 GWAS SNPs were near or just 5′ of a gene encoding a long noncoding RNA (lncRNA), *MIR2052HG*, (also referred to as *FLJ39080* and *LOC44d1355*) (see **Figure 4b**). One of the top SNPs, rs4476990 (P = 2.51E-07) mapped to an ERE and another, rs3802201 (P = 6.24E-07), was located 16 bp from an ERE (**Figure 4c**). Furthermore, *MIR2052HG*, like *ZNF423* and *CTSO*, was estrogen inducible in an SNP genotype dependent fashion, with induction of expression only for variant genotypes. ChIP assays for EREs near the SNPs showed ERα binding only in samples with variant genotypes, compatible with the mRNA expression pattern. In addition, induction of expression for the lncRNA in "Human Variation Panel" LCLs was paralleled by increased expression of the *ESR1* gene that encodes ERα. In ER+ breast cancer cell

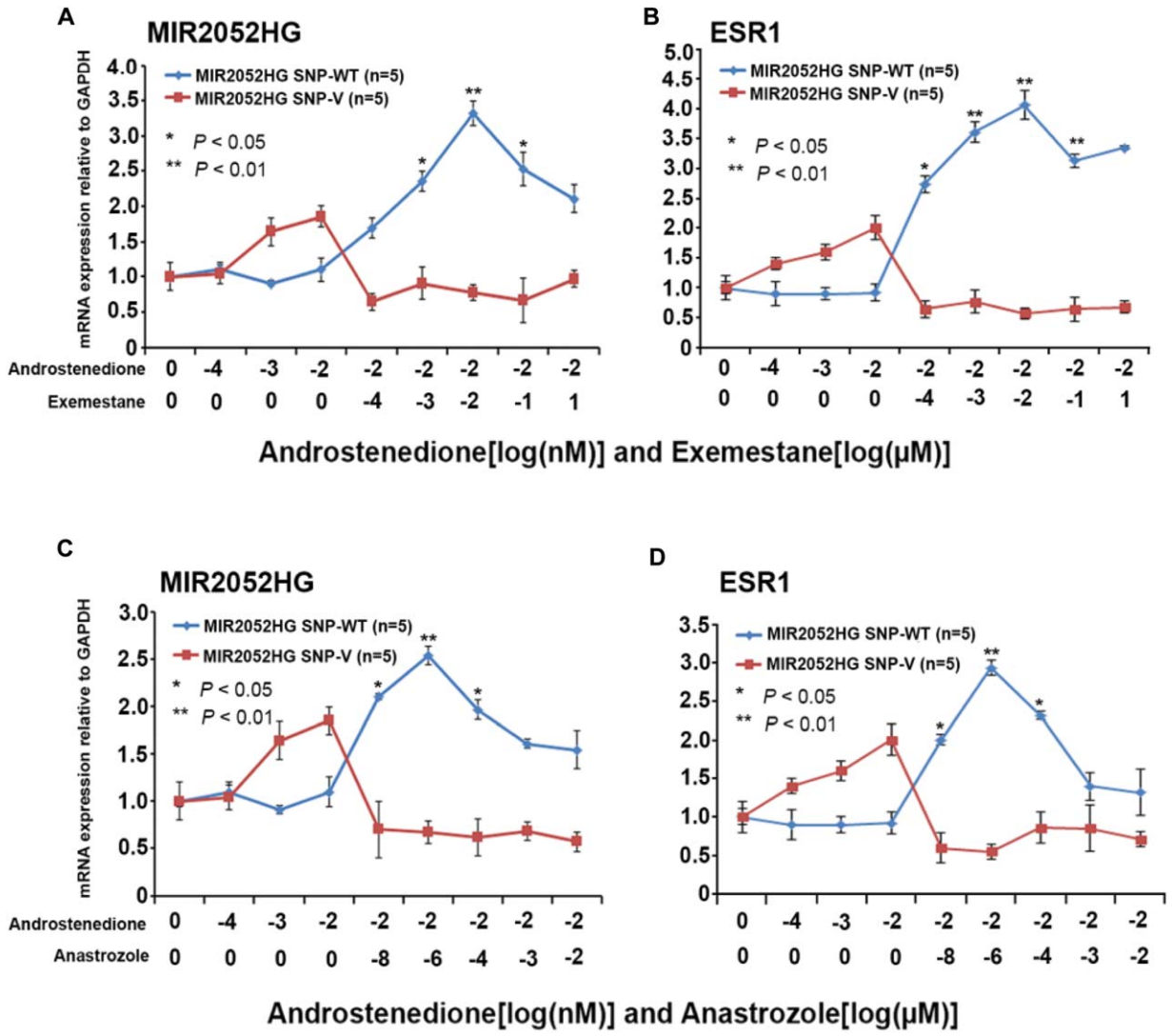

Figure 5 (**a,c**) MIR2052HG RNA expression in LCLs with WT and variant (V) genotypes for both rs4476990 and rs3802201 after exposure to androstenedione alone and with increasing concentrations of exemestane or anastrozole. (**b,d**) *ESR1* mRNA expression in LCLs under the same conditions as in (**a,c**). $*P < 0.05$; $**P < 0.01$. Modified from Ref. 5.

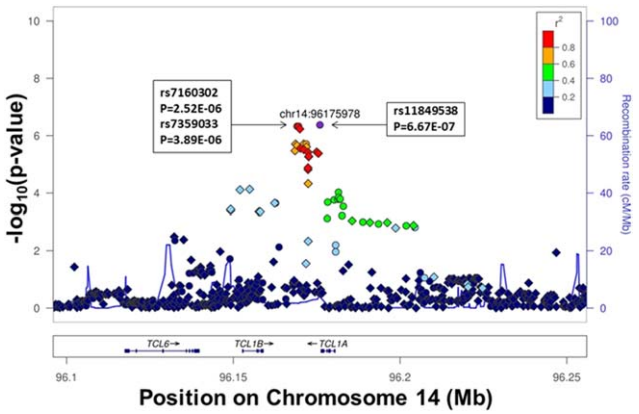

Figure 6 Regional plot showing the association of SNPs identified during the MA.27 AI Trial GWAS for musculoskeletal adverse events within the *TCL* gene cluster on chromosome 14. Modified from Ref. 27.

lines, knockdown of MIR2052HG resulted in a striking decrease of ERα expression, an observation that resulted from an effect on ERα ubiquitination and proteasome-mediated degradation as well as on *ESR1* transcription. Particularly striking was the fact that androstenedione, a precursor for the biosynthesis of estrone catalyzed by aromatase, induced the expression of both the lncRNA and of ESR1 in cells with variant but not in those with WT SNP genotypes (**Figure 5**). However, the addition of either of the two aromatase inhibitors used to treat MA.27 patients, anastrozole or exemestane, reversed this genotype-dependent effect, so that cells homozygous for the WT genotype displayed enhanced expression of both MIR2052HG and ESR1, while expression in cells homozygous for variant genotypes returned to baseline expression levels for both genes (see **Figure 5**). Finally, MIR2052HG overexpression increased ERα expression and enhanced proliferation of ER+ breast cancer cell lines, while

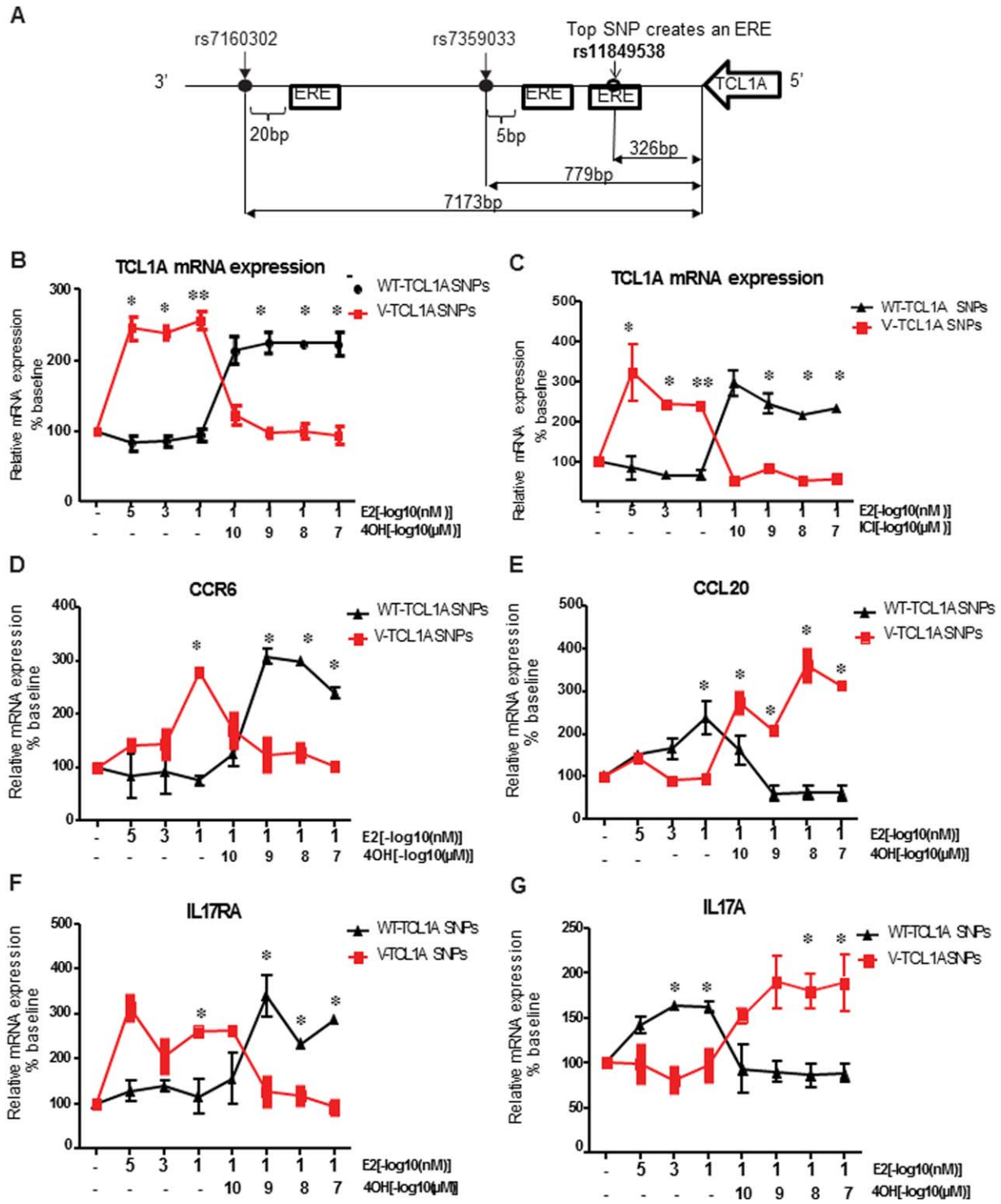

Figure 7 SNP- and estrogen-dependent effects on the mRNA expression of *TCL1A*, genes encoding cytokines, chemokines, and their receptors in LCLs. (**a**) Schematic diagram of the two *TCL1A* SNPs, rs7359033 and rs7160302, in tight LD with rs11849538, the top hit signal from the MA.27 GWAS. Locations of EREs are shown as boxes. These three SNPs map between the 3′-termini of *TCL1A* and *TCL1B*. ER blockade by 4-hydroxytamoxifen (4OH) or fulvestrant (ICI) treatment resulted in the "reversal" of *TCL1A* SNP and estrogen-dependent expression patterns (**b,c**) and downstream effects on the expression of CCR6, CCL20, IL17RA, and IL17A. (**d–g**) Values are mean ± SEM of three assays. $*P < 0.0001$. Modified from Ref. 28.

downregulation of the lncRNA was associated with reduced proliferation. All of these observations were compatible with the GWAS finding that variant genotypes for the chromosome 8 SNPs were associated with reduced risk for breast cancer recurrence.[5] At the mechanistic level, the lncRNA appeared to regulate ER levels both at the transcriptional and protein levels.[5] In

summary, germline SNPs that were associated with AI efficacy led us to the novel finding of a long noncoding RNA that could transregulate *ESR1* gene expression as well as ER protein stability and proteasome-mediated degradation, all of which added additional layers of regulation for ERα. As a result, these studies moved beyond merely establishing a GWAS association, and led to deeper mechanistic understanding that could have implications well beyond AI drug response. That point will be highlighted to an even greater degree by the next series of studies.

BREAST CANCER AROMATASE INHIBITOR ADVERSE RESPONSE GWAS

As mentioned previously, AIs have replaced tamoxifen as the mainstay for the adjuvant therapy of postmenopausal patients with ER+ breast cancer.[15,16,18] However, even though these drugs are highly effective, their clinical use presents challenges. For example, in the MA.27 trial 31.6% of the patients discontinued AI therapy before completing the full 5-year course of therapy. One of the major reasons was severe musculoskeletal adverse events (MS-AEs), i.e., muscle and joint pain. Therefore, we performed a GWAS in 2010 for patients with grade 3 or 4 NCI Common Terminology Criteria for MS-AEs.[6] That GWAS included 293 cases and 585 matched controls from among the MA.27 patients.[6] The top SNP signal mapped to chromosome 14, with the top hit SNP just 3′ of the *TCL1A* gene, one of three genes in a TCL gene cluster (**Figure 6**). The variant sequence for the top hit SNP, rs11849538 (P = 6.67E-07), created an ERE and *TCL1A*, like *ZNF423*, *CTSO*, and *MIR2052HG*, was estrogen inducible, but only in cells with the variant rs11849538 SNP genotype that created an ERE. Of importance, there were at least two other SNPs, rs7160302 and rs7359033, in tight linkage disequilibrium with the top SNP, but those SNPs were not in EREs as was the rs11849538 SNP that created an ERE. However, those two additional SNPs were both near ERE motifs (see **Figure 7a**), an observation that we will return to when we describe the pattern of E2-driven induction of *TCL1A* and the ability of SERMs to reverse that expression pattern. In the 2010 publication, we also showed that TCL1A was capable of regulating the expression of a cytokine receptor, IL-17RA. IL-17RA was selected for study because it had been identified as a therapeutic target for the treatment of rheumatoid arthritis. The date of this initial TCL1A publication is important because it was 3 years prior to the P-1 and P-2 GWAS,[5] so the possibility that SERMs might reverse expression patterns after E2 exposure as a result of SNPs outside of ERE motifs had not yet been reported.

In a subsequent publication,[26] we demonstrated that TCL1A overexpression or knockdown could alter the downstream expression of a series of cytokines and cytokine receptors including IL-17, IL-17RA, IL-12, IL-12RB2, and IL-1R2.[26] In 2016, the list of inflammatory mediators regulated by TCL1A in an SNP-dependent fashion was broadened to include chemokine receptors such as CCR6 and its ligand CCL20.[27] Of equal importance, those experiments showed that exposure to 4-OH-TAM or the ER blocker fulvestrant together with estrogens could reverse the pattern of TCL1A SNP-dependent induction and the downstream induction of cytokine and chemokine receptors (see **Figure 7**). **Figure 7** shows the organization of the three SNPs 3′ of the *TCL1A* gene (**7a**), as well as the E2 induction of TCL1A mRNA, and the reversal of that genotype-dependent pattern by either 4-OH-TAM or ICI (**7b**) and (**7c**) the downstream effects on cytokine and chemokine receptor and ligand expression (**7d–g**). Please note that the SNPs for the data shown in **Figure 7** were in the *TCL1A* gene, not in genes encoding the cytokines and chemokines that were studied. Recently, the range of inflammatory mediators that can be regulated in this fashion by TCL1A was broadened further to include the Toll-like receptors TLR2, TLR7, TLR9, and TLR10—as well as nuclear factor kappa B (NF-κB) through MYD88 signaling.[28] Finally, recent genome-wide RNA-seq studies performed using LCLs identified 357 genes that were regulated in a *TCL1A* SNP, E2, and 4-OH-TAM-dependent fashion. Genome-wide ChIP-seq was then used to validate 74 of those genes.[29] As a result, a GWAS performed originally to study an adverse response to AIs in the clinic resulted in the identification of an unanticipated transcription factor, TCL1A, and raised the possibility of the pharmacological modification by SERMs of the expression of that transcription factor and, downstream, the expression of a series of inflammatory mediators, all informed by knowledge of the *TCL1A* genotype. This series of experiments, like the previous examples of GWAS for breast cancer endocrine therapy, emphasize the importance of the functional pursuit of GWAS signals even if they do not display genome-wide statistical significance. Of course, many signals will be false positives but, by testing borderline results, novel biology can be identified that might generate hypotheses leading to new areas of research. At the same time, these particular experiments were made possible by crosstalk between the P-1 and P-2 SERM chemoprevention GWAS and the MA.27 MS-AEs GWAS—crosstalk that would not have occurred without the technical advances that made genome-wide pharmacogenomic studies possible.

CONCLUSION

This brief overview of the evolution of pharmacogenomic GWAS strategies and approaches in the post-Genome Project world began with the statement that biomedical research is in the midst of a period of rapid change. The validity of that statement was demonstrated by the subsequent description of the observations made in the course of a series of illustrative GWAS focused on the endocrine therapy of breast cancer and their follow-up functional studies. It may be worthwhile to remember that GWAS is only barely more than a decade old, and that some of the techniques described in this overview, e.g., RNA-seq and ChIP-seq, did not exist in a practical sense when the earliest of these GWAS, that for AI-related MS-AEs, was published in 2010.[6] It is also clear that the application of NGS will continue the process of expanding the "reach" of pharmacogenomics and that this brief overview has only addressed studies of germline pharmacogenomic GWAS without addressing the equally important issue of variation in the tumor somatic genome. It might also be helpful to point out the limitations faced by those of us

who are engaged in studies that apply omics science to drug response phenotypes. As mentioned previously, those techniques require a large number of subjects in order to be adequately powered; they produce very large datasets; and none of the studies described in this overview was inexpensive. Furthermore, as opposed to GWAS for disease risk, pharmacogenomic GWAS will rarely include tens or hundreds of thousands of subjects in a single study because of the overwhelming logistic and cost barriers involved in assembling study populations of that size who are on a given program of drug therapy in a reasonably controlled setting. However, there is one great advantage to the study of drug response phenotypes. We begin with insight into the mechanism(s) of action of the drugs that are being studied. It is no coincidence that all of the SNP signals and genes identified during the GWAS described in preceding paragraphs were regulated, at least in part, by estrogens. The drugs used in the endocrine therapy of breast cancer, SERMs and AIs, were designed to either block the estrogen receptor or to block estrogen biosynthesis. As a result, we had a "starting point" for the functional and mechanistic validation studies for those GWAS signals. In essence, we were able to use the drugs that we were studying as "molecular probes," with resultant novel and sometimes surprising insights into both mechanisms of drug action and the pathophysiology of the underlying disease. Included among the lessons learned in the course of the decade-long journey that this illustrative set of studies of breast cancer endocrine therapy taught us was the importance of SNPs outside of exons and splice junctions. Virtually all of the SNP signals that we observed involved regulatory variants, emphasizing the value of public databases like GTEx (https://www.gtexportal.org/home/), TCGA (https://cancergenome.nih.gov/), and ENCODE (https://www.encodeproject.org/). We also learned that the functional validation of signals with borderline statistical validity can be immensely helpful in the pursuit of biological relevance, the ultimate goal of all of our efforts. Decisions with regard to which of these SNP signals to pursue will depend on the potential functional implications of the SNPs, i.e., are they eQTLs or nonsynonymous coding SNPs, the pathways to which the genes identified map, and the results of initial gene knockdown or overexpression experiments performed with appropriate cell lines. Ultimately, each phenotype might require a slightly different set of experiments in order to make a decision with regard to the vigor with which those signals might be pursued.

Finally, the studies described in this brief overview clearly demonstrate that effective translational pharmacogenomic research involves close collaboration between clinical investigators and bench scientists and a process that takes discoveries from clinical studies into the laboratory to gain insight into new biology that can then be returned to patients in the form of more highly individualized diagnosis and therapy. What is certain is that all of us have just begun the journey of discovery that the technical advances that have moved us from candidate genes to genome-wide scans have launched. As we add whole genome sequencing, transcriptomics, proteomics, metabolomics, and the microbiome to these studies, we are certain to see the evolution of pharmacogenetics to pharmacogenomics extend onto pharmacoomics and beyond.

ACKNOWLEDGMENTS

These studies were supported, in part, by the following United States National Institutes of Health grants: U19 GM61388 (The Pharmacogenomics Research Network), P50 CA116201 (Mayo Clinic Breast Cancer Specialized Program of Research Excellence), U54 GM114838, R01 GM28157, R01 CA196648, U10 CA180868, and UO1 CA18967. They were also supported by the Breast Cancer Research Foundation, the Nan Sawyer Breast Cancer Fund and by the Eisenberg Foundation. We also want to express our gratitude to the women who participated in the NSABP P-1 and P-2 SERM chemoprevention trials and in the NCIC MA.27 AI adjuvant therapy trial.

CONFLICT OF INTEREST

The authors declared no conflict of interest.

1. Motulsky, A.G. Drug reactions enzymes, and biochemical genetics. *JAMA* **165**, 835–837 (1957).
2. Weinshilboum, R.M. & Wang, L. Pharmacogenetics and pharmacogenomics: development, science, and translation. *Annu. Rev. Genomics Hum. Genet.* **7**, 223–245 (2006).
3. Wang, L., McLeod, H.L. & Weinshilboum, R.M. Genomics and drug response. *N. Engl. J. Med.* **364**, 1144–1153 (2011).
4. Ingle, J.N. *et al*. Selective estrogen receptor modulators and pharmacogenomic variation in ZNF423 regulation of BRCA1 expression: individualized breast cancer prevention. *Cancer Discov.* **3**, 812–825 (2013).
5. Ingle, J.N. *et al*. Genetic polymorphisms in the long noncoding RNA MIR2052HG offer a pharmacogenomic basis for the response of breast cancer patients to aromatase inhibitor therapy. *Cancer Res.* **76**, 7012–7023 (2016).
6. Ingle J.N, *et al*. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. *J. Clin. Oncol.* **28**, 4674–4682 (2010).
7. Gonzalez, F.J. *et al*. Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. *Nature* 442–446 (1988).
8. Broly, F. *et al*. Debrisoquine/sparteine hydroxylation genotype and phenotype: analysis of common mutations and alleles of CYP2D6 in a European population. *DNA Cell Biol.* **10**, 545–558 (1991).
9. Weinshilboum, R.M. *et al*. Sulfation and sulfotransferases 1. *FASEB J.* Jan, 3–14 (1997).
10. Weinshilboum, R.M., Otterness, D.M. & Szumlanski, C.L. Methylation pharmacogenetics: catechol O-methyltransferase, thiopurine methyltransferase, and histamine N-methyltransferase. *Annu. Rev. Pharmacol. Toxicol.* **39**, 19–52 (1999).
11. Weinshilboum, R.M. & Sladek, S.L. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. *Am. J. Hum. Genet.* **32**, 651–662 (1980).
12. Hertz, D., Henry, N. & Rae, J. Germline genetic predictors of aromatase inhibitor concentrations, estrogen suppression and drug efficacy and toxicity in breast cancer patients. *Pharmgenomics Pers. Med.* **18**, 481–499 (2017).
13. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics. *CA Cancer J. Clin.* **65**, 5–29 (2015).
14. Ferlay, J. *et al*. Cancer incidence and mortality worlwide: sources. *Int. J. Cancer.* **136**, 359–386 (2012).
15. Dalmau, E., Armengol-Alonso, A., Munoz, M. & Segui-Palmer, M.A. Current status of hormone therapy in patients with hormone

receptor positive (HR+) advanced breast cancer. *Breast* **23**, 710–720 (2014).
16. Folkerd, E. & Dowsett, M. Sex hormones and breast cancer risk and prognosis. *Breast* **22**(Suppl 2), S38–43 (2013).
17. Fisher, B. *et al*. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J. Natl. Cancer Inst.* **90**, 1371–1388 (1998).
18. Vogel, V.G. *et al*. Effects of tamoxifen vs raloxifen on the risk of developing invasive breast cancer and other disease outcomes. *JAMA* **295**, 2727–2741 (2006).
19. Vogel, V.G. *et al*. Update of the National Surgical Adjuvant Breast and Bowel Project Study of Tamoxifen and Raloxifene (STAR) P-2 Trial: preventing breast cancer. *Cancer Prev. Res. (Phila.)* **3**, 696–706 (2010).
20. Goss, P.E. *et al*. Exemestane for breast-cancer prevention in postmenopausal women. *N. Engl. J. Med.* **364**, 2381–2391 (2011).
21. Liu, D. *et al*. Breast cancer chemoprevention pharmacogenomics: deep sequencing and functional genomics of the ZNF423 and CTSO genes. *NPJ Breast Cancer.* **3**, (2017).
22. Qin, S. *et al*. Calmodulin-like protein 3 is an estrogen receptor alpha coregulator for gene expression and drug response in an SNP, estrogen, and SERM-dependent fashion. *Breast Cancer Res.* **19**, 95 (2017).
23. Early Breast Cancer Trialists' Collaborative G, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* **378**, 771–784 (2011).
24. Dowsett, F. *et al*. Aromatase inhibitors versus tamoxifen in early breast cancer. *Lancet Infect. Dis.* **386**, 1341–1352 (2015).
25. Goss, P.E. *et al*. Exemestane versus anastrozole in postmenopausal women with early breast cancer. *J. Clin. Oncol.* **31**, 1398–1404 (2013).
26. Liu, M. *et al*. Aromatase inhibitors, estrogens and musculoskeletal pain: estrogen-dependent T-cell leukemia 1A (TCL1A) gene-mediated regulation of cytokine expression. *Breast Cancer Res.* **14**, R41 (2012).
27. Ho, M.F. *et al*. Estrogen, SNP-dependent chemokine expression and selective estrogen receptor modulator regulation. *Mol. Endocrinol.* **30**, 382–398 (2016).
28. Ho, M.F. *et al*. TCL1A single-nucleotide polymorphisms and estrogen-mediated Toll-like receptor-MYD88-dependent nuclear factor-kappab activation: single-nucleotide polymorphism- and selective estrogen receptor modulator-dependent modification of inflammation and immune response. *Mol. Pharmacol.* **92**, 175–184 (2017).
29. Ho, M.-F. *et al*. TCL1A identification as novel transcription factor and a pharmacogenomic link to aromatase inhibitor-induced musculoskeletal adverse event. *Clin. Pharmacol. Ther.* **101**, S5–S99 (2017).