BJCP British Journal of Clinical Pharmacology

DOI:10.1111/bcp.12166

Implications of genome-wide association studies in cancer therapeutics

Jai N. Patel, Howard L. McLeod & Federico Innocenti

UNC Institute for Pharmacogenomics and Individualized Therapy, University of North Carolina, Chapel Hill, NC 27599-7361, USA

Correspondence
Dr Federico Innocenti MD PhD, UNC Institute for Pharmacogenomics and Individualized Therapy, University of North Carolina, 120 Mason Farm Road, CB #7361, Rm 1014, Chapel Hill, NC 27599-7361, USA.
Tel.: +1 919 966 9422
Fax: +1 919 966 5863
E-mail: innocent@email.unc.edu

Keywords
association, cancer, genome-wide, polymorphism, replication, therapeutics

Received
13 February 2013

Accepted
21 April 2013

Accepted Article Published Online
23 May 2013

Genome wide association studies (GWAS) provide an agnostic approach to identifying potential genetic variants associated with disease susceptibility, prognosis of survival and/or predictive of drug response. Although these techniques are costly and interpretation of study results is challenging, they do allow for a more unbiased interrogation of the entire genome, resulting in the discovery of novel genes and understanding of novel biological associations. This review will focus on the implications of GWAS in cancer therapy, in particular germ-line mutations, including findings from major GWAS which have identified predictive genetic loci for clinical outcome and/or toxicity. Lessons and challenges in cancer GWAS are also discussed, including the need for functional analysis and replication, as well as future perspectives for biological and clinical utility. Given the large heterogeneity in response to cancer therapeutics, novel methods of identifying mechanisms and biology of variable drug response and ultimately treatment individualization will be indispensable.

Introduction

Since its completion in 2003, an important feature of the Human Genome Project was the dedication and commitment to expanding technology across private sectors, catalyzing technological advancements and promoting the utility and feasibility of innovative research [1]. Methods for sequencing and scanning the genome have become less costly, resulting in greater discoveries of novel genetic loci that associate with the risk of disease, clinical outcome and/or risk of toxicity [2]. The advent of genome wide association (GWA) technology has enabled researchers to move beyond the limitations of small scale candidate gene studies to the identification of hundreds of loci influencing a wide range of phenotypes [3].

Genome wide association studies (GWAS) provide an agnostic approach to identifying genetic variants or single nucleotide polymorphisms (SNPs), associated with disease susceptibility, prognosis of survival and/or predictive of drug response. Large sample sizes are generally required for a GWAS to have sufficient statistical power to detect associations while using array platforms that can interrogate >500 000 SNPs. The design of common GWAS platforms is based on linkage disequilibrium (LD). Genotypes can be imputed for markers not directly genotyped using reference panels such as 1000 Genomes or HapMap [4]. It is important to consider what proportion of SNPs not directly genotyped are captured, or 'tagged,' by the chip via LD. As a result, genotyping chips have traditionally covered common variants, but more recently coverage has expanded to cover more rare variants [5]. Although these techniques are costly, they do allow for a more unbiased interrogation of the entire genome, resulting in the discovery of novel genes and biological associations [6]. A common disadvantage to GWAS includes high rates of false positives and low statistical power, while interpretation of study results can be extremely challenging, requiring fine mapping and mechanistic studies to understand the biological plausibility of certain findings [6].

This review will focus on the implications of GWAS in cancer therapy, in particular germ-line mutations, including findings from major GWAS which have identified predictive genetic loci for clinical outcome and/or toxicity, as well as challenges and future perspectives for biological

© 2013 The British Pharmacological Society

and clinical utility. Table 1 summarizes these GWAS findings and elucidates the biological plausibility of top hits.

GWAS in cancer therapeutics

To date, the majority of cancer GWAS have characterized mutated cancer predisposition genes accounting for disease susceptibility [2]. Although inherited syndromes account for only a small fraction of the familial risk of cancer, greater than 50 genetic mutations have been associated with high-penetrance cancer susceptibility syndromes [2]. Additionally, debilitating toxicities such as neutropenia and neurotoxicity often necessitate dose reductions, delays or cessation of therapy, resulting in ineffective therapy. SNPs identified through GWAS that are found to be predictive of toxicity or response may be associated with increased effect sizes, ultimately allowing prospective identification of toxicity risk and clinical response.

Pancreatic cancer

Cancer and Leukemia Group B (CALGB) 80303 was a double-blind, placebo-controlled randomized phase III study of gemcitabine with or without bevacizumab, in which 365 patients with advanced pancreatic cancer were genotyped [7]. Results demonstrated a significant association (defined as P value $<1 \times 10^{-7}$ after adjusting for multiple comparisons) between rs763780 in IL17F ($P = 2.61 \times 10^{-8}$) and median overall survival (OS) [3.1 (heterozygotes) *vs.* 6.8 months (wild type), respectively]. This SNP was also in strong LD ($r^2 = 0.955$) with another IL17F SNP (rs7771466), having the second lowest P value ($P = 1.66 \times 10^{-7}$) and a similar effect on OS. However, after adjusting for the stratification factors, such as treatment arm, the SNPs did not meet the criterion for genome-wide statistical significance. A similar trend was observed in a small replication cohort of 26 patients of African ancestry [7]. This study demonstrates the feasibility in conducting a GWAS using prospectively collected specimens from a randomized phase III clinical trial, in which phenotypes are often more accurately recorded. Although predictors of clinical response in pancreatic cancer are direly needed and the IL17F locus may be a promising genetic determinant of response, replication in larger cohorts is necessary. Unfortunately, the chances of obtaining adequate sample sizes for replication in patients treated in the identical manner is minimal and poses a large barrier to further validation.

Kiyotani *et al*. identified four genetic markers for haematological toxicities in a GWAS of 79 Japanese patients with cancer (the majority of which were pancreatic cancer cases) receiving gemcitabine, consisting of 21 patients experiencing grade 3 or 4 leukopenia/neutropenia and 58 controls (not experiencing grade 3 or 4 leukopenia) [8]. In a replication phase, 33 patients experiencing grade 3 or 4 leukopenia/neutropenia and 62 controls were genotyped. Of the top 100 SNPs (P values 2.12×10^{-4} to 6.69×10^{-6}), 70 were genotyped in the replication study, four of which were identified with associations of $P <$ 0.05 before multiple testing (one SNP, rs11141915, was found in the gene DAPK1 and another SNP, rs12046844, was found in the gene PDE4B). The proportion of patients with gemcitabine-induced leukopenia/neutropenia was significantly increased in groups with higher prediction scores (calculated from the combined effect of the four loci) (trend test $P = 1.31 \times 10^{-14}$). The prevalence of grade 3/4 leukopenia/neutropenia was 11.5% (13/113) in the combined group of scores 0 and 1, 60.9% (28/46) in the score 2 group and 86.7% (13/15) in the score 3 group. Correspondingly, the odds ratio (OR) in the score 3 group was as high as 50.00 ($P = 4.13 \times 10^{-9}$) and that of the score 2 group was 11.97 ($P = 6.25 \times 10^{-10}$), compared with that in the group of scores 0 and 1 [8]. Investigators took the approach of utilizing a large biobank as opposed to a clinical trial to select and enrich patients experiencing a phenotype of interest. The small sample size is ultimately insufficient to determine a true causal relationship and a variety of tumour types were included, thus allowing heterogeneity in dosing and treatment regimens. Prediction scores of 0, 1, 2 and 3 (frequency, 29.0, 45.3, 20.8 and 4.9%, respectively) demonstrated a possible cumulative effect on the risk of gemcitabine-induced severe haematologic toxicity, which remains to be validated in a larger and more homogeneous cohort [9].

Breast cancer

CALGB 40101, a phase III randomized study comparing cyclophosphamide and doxorubicin *vs.* single agent paclitaxel in breast cancer patients, was the basis of another GWAS in which 855 genetically-defined European patients treated on the paclitaxel arm were genotyped for >500 000 SNPs. An additional 154 self-declared European patients and 117 African American patients from the same study were genotyped for internal replication [9]. No SNPs analyzed for association with initial onset of sensory peripheral neuropathy reached genome-wide significance (defined as P value $<1 \times 10^{-7}$ after adjusting for multiple comparisons). Of these top SNPs, biologic relevance was apparent for polymorphisms in EPHA5 (rs7349683, $P = 9.6 \times 10^{-7}$) and FGD4 (rs10771973, $P = 2.6 \times 10^{-6}$). Ordinal logistic regression identified a SNP within FZD3 demonstrating a significant association with grade of sensory peripheral neuropathy (rs7001034, $P = 3.1 \times 10^{-9}$, OR, 0.57). Association for FGD4 was confirmed in both the European and African American samples in the internal replication set (HR 1.72, $P = 0.013$ and HR 1.93, $P = 6.7 \times 10^{-3}$, respectively). To assess the potential translational implications of this finding to clinical practice, investigators estimated the cumulative dose level triggering an event for each FGD4 genotype. The tolerated cumulative paclitaxel dose level for patients with zero, one and two copies of the risk allele was 1047 mg m^{-2}, 877 mg m^{-2} and 710 mg m^{-2}, respectively [9]. Similar to CALGB 80303 [7], the GWAS for 40101 collected

Table 1

Summary and biological plausibility of cancer GWAS findings

Tumour	Chemotherapy	Sample size (ethnicity)	Top SNP(s), gene	*P* value	Replication? (sample size)	Biological plausible association
Pancreatic [7]	Gemcitabine	365 (Caucasian)	rs763780, IL17F	2.61×10^{-8}	Yes (26 African American)	IL17F is a cytokine with the ability to induce inflammation. In contrast to WT 161H, the 161R variant form lacks the ability to activate the MAPK pathway, restricting cytokine and chemokine production. The variant 161R is a natural antagonist of the anti-angiogenic and proinflammatory effects of WT 161H IL-17F.
Pancreatic, lung and bile duct [8]	Gemcitabine	79 (Japanese)	rs11141915, DAPK1	1.27×10^{-6}	Yes (95 Japanese)	DAPK1 is a member of a serine/threonine kinase family that mediates the gamma-interferon-induced cell death and also mediates apoptosis induced by tumour necrosis factor-alpha. DAPK1 is reported to be expressed in bone marrow and peripheral blood.
Breast [9]	Paclitaxel	855 (Caucasian)	rs7349683, EPHA5 rs10771973, FGD4 rs7001034, FZD3	9.6×10^{-7} (EPHA5) 2.6×10^{-6} (FGD4)	Yes (154 Caucasian; 117 African American)	EPHA5 encodes an ephrin receptor gene implicated in the process of neuronal regeneration following nerve injury. FGD4 encodes a Rho-GTPase guanine nucleotide exchange factor previously implicated in congenital peripheral neuropathies and development of Schwann cell function. Previous studies implicate FGD4 mutations in the congenital disease, Charcot Marie Tooth syndrome, resulting in chronic peripheral neuropathy. FZD3 encodes a Wnt receptor with reported roles in neurite outgrowth.
Breast [10]	Tamoxifen	240 (Japanese)	rs10509373, C10orf11 gene on 10q22	1.06×10^{-7}	Yes (two sets of 105 and 117, both Japanese)	C10orf11 protein, comprising 198 amino acids, is predicted to contain a leucine-rich repeat domain and to have the capacity of protein binding, although no reports have elucidated its functional role.
Breast and liver [11]	Epirubicin	270 (Japanese)	rs2916733, MCPH1	1.59×10^{-7}	Yes (48 Japanese)	Little is known regarding the mechanism of MCHP1. MCPH1 may play an important role for repair of DNA damage induced by epirubicin.
Breast [12]	Anastrozole, exemestane	878 (Caucasian)	rs7158782, rs7159713, rs2369049. TCL1A	$<1 \times 10^{-6}$	None	Relationship of TCL1A to IL17RA expression indicates that the association of TCL1A expression with cytokine function may be a notable mechanism of inducing arthritis-related adverse events.
NSCLC [13]	Irinotecan	101 (Korean)	rs2745761, PLCB1 (for diarrhoea) rs7779029, SEMA3C (for neutropenia)	8.7×10^{-6} (PLCB1) 7.1×10^{-5} (SEMA3C)	Yes (146 Korean)	PLCB1 catalyzes the formation of inositol 1,4,5-trisphosphate and diacylglycerol, which play a role in the intracellular transduction of extracellular signals. SEMA3C may be involved in cell survival. A GWAS meta-analysis revealed that variants in SEMA3C are associated with bilirubin levels. Because the plasma concentration of SN-38 is strongly correlated with total bilirubin level, SEMA3C variants may lead to neutropenia after irinotecan treatment.
NSCLC [14]	Cisplatin	327 (Caucasian)	rs1878022, CMKLR1 rs10937823	1.42×10^{-6} (CMKLR1) 1.80×10^{-6} (rs10937823)	Yes (315 Caucasia and 420 Spanish)	CMKLR1 encodes for a seven transmembrane G-protein coupled receptor that is involved in several cellular pathways, including inflammation, adipogenesis, and osteoblastogenesis. The rs10937823 variant is located within an intron, a gene encoding for the poorly characterized SORCS2 protein.
NSCLC [17]	Cisplatin, carboplatin	535 (Japanese)	rs7629386, CTNNB1 rs3850370, SNW1	3.63×10^{-5} (CTNNB1) 2.92×10^{-5} (SNW1)	Yes (340 Japanese and 409 Caucasian)	SNW1 is a nuclear matrix-associated co-activator that may couple vitamin D receptor-mediated transcription and RNA splicing. It was also shown to be required for TGF-beta-1 induced epithelial–mesenchymal transition and invasiveness in transformed cells. CTNNB1 a key component of the Wnt signaling pathway, increasing cell proliferation and tumor growth.
NSCLC [18]	Carboplatin	105 (Japanese)	rs1656402, EIF4E2 rs1209950, ETS2 rs9981861, DSCAM	8.4×10^{-8} (EIF4E2) 2.8×10^{-7} (ETS2) 3.5×10^{-6} (DSCAM)	None	EIF4E2 functions to recruit mRNAs to the ribosome. Although EIF4E2 is ubiquitously expressed, the expression in metastatic tumours in increased, suggesting an active role in the prognosis of NSCLC. ETS has shown to be an important downstream target in cellular transformation. Evidence suggests ETS activity may strongly influence the phenotype of prostate tumours. DSCAM encodes for the Down syndrome cell adhesion molecule and is expressed widely in the developing nervous system. It has been shown to mediate neurite processes and spacing of neuronal cell bodies.
Colorectal [19]	5-fluorouracil, oxaliplatin, leucovorin	221 (Caucasian)	rs10876844, upstream of METTL7B and ITGA7	4.78×10^{-5}	Yes (791, Caucasian)	Unknown mechanism
Acute lymphoblastic leukemia [21]	Methotrexate	434 (Caucasian)	rs11045879, rs4149081, Both found in SLCO1B1	1.7×10^{-10} 1.7×10^{-9}, respectively	Yes (206 Caucasian)	SLCO1B1 mediates uptake of substrates from sinusoidal blood, resulting in their net excretion from blood (likely via biliary excretion). SLCO1B1 has been shown to transport methotrexate *in vitro*.

samples from a prospective phase III randomized clinical trial. Although evidence suggests a dose to event association with variations in FGD4, which was also replicated in two additional cohorts of patients from the same study, the trend towards genome-wide significance in the discovery population requires that further validation in independent cohorts is required.

Another study by Kiyotani *et al.* enrolled a total of 462 Japanese patients with primary breast cancer treated with adjuvant tamoxifen monotherapy [10]. Two hundred and forty patients were genotyped for the discovery phase and 105 and 117 cases were used as two replication sets. Significant associations were found between recurrence-free survival (RFS) and 15 SNPs in nine genetic regions that satisfied a genome-wide significant threshold of P value $<5 \times 10^{-7}$ (after adjusting for multiple testing of 470 796 SNPs). Results demonstrated that rs10509373 in the C10orf11 gene on 10q22 was the only SNP significantly associated with RFS in the replication stage (log-rank $P = 4.18 \times 10^{-4}$). Investigators genotyped 130 tag SNPs for fine mapping on chromosome 10.q22, which identified a 172-kb LD block associated with RFS in patients receiving tamoxifen therapy. Combined analysis of C10orf11, CYP2D6 and ABCC2 (the latter two being previously identified to impact RFS) revealed that variants of the three genes have cumulative effects on RFS ($P = 2.28 \times 10^{-12}$) [risk of recurrence increased from 6.51-fold (three risk alleles) to 119.51-fold (five risk alleles) compared with those carrying one risk allele] [10]. The large LD block associated with RFS, along with the evidence of a cumulative effect with previously known genetic variations, may provide a novel locus to identify patients most likely to respond to tamoxifen therapy. However, further replication and validation in larger cohorts is necessary.

Srinivasan *et al.* examined epirubicin-induced leukopenia in 270 Japanese patients comprising 67 severe leukopenia cases (the majority of which were breast cancer cases) and 203 controls (non-leukopenia) (the majority of which were breast or liver cancer cases) [11]. Samples were taken from the same Japanese biobank used in the study by Kiyotani *et al.* [8]. The top finding identified a locus having three SNPs (rs2916733, rs2440399, and rs2920656) associated with leukopenia, in the MCPH1 gene on chromosome 8 ($P = 1.59 \times 10^{-7}$, OR = 2.92; $P = 4.51 \times 10^{-7}$, OR = 2.81 and $P = 6.59 \times 10^{-7}$, OR = 4.86, respectively). No SNPs remained significant after Bonferroni correction in a replication phase of 48 ADR patients. A haplotype constructed from the landmark SNP (rs2916733) and rs1031309, which was in LD with rs2916733 ($r^2 = 0.64$), showed a stronger association ($P = 2.20 \times 10^{-10}$, OR = 2.88) than the single landmark SNP [11]. These results may provide insight into genetic determinants of epirubicin-induced leukopenia. However, the suggested haplotype should be validated in a larger controlled, homogenous cohort, as the dosing and treatment regimens may vary with a heterogenous biobank of samples.

Ingle *et al.* [12] utilized a nested case-control design to select breast cancer patients enrolled onto the MA.27 phase III trial comparing anastrozole with exemestane [13]. Cases ($n = 293$) were matched to controls ($n = 585$) in a 1:2 design and were defined as patients with grade 3 or 4 musculoskeletal adverse events (MS-AEs). Although non-significant after testing for multiple comparison, the top three SNPs (P values $<1 \times 10^{-6}$) were found on chromosome 14, and were found to be in high LD ($r^2 > 0.8$), while an additional imputed SNP (rs11849538) also showed an association with MS-AEs (OR 2.21, $P = 6.67 \times 10^{-7}$). The three genotyped SNPs (rs7158782, rs7159713, and rs2369049) and the imputed SNP were all in high LD in the T-cell leukemia 1A (TCL1A) gene ($r^2 = 0.85$). Functional studies demonstrated an eight-fold induction of TCL1A expression in ER-transfected cells and significantly higher expression after exposure to varying concentrations of o–estrogen in lymphoblastoid cell lines (LCLs) containing the variant SNPs when compared with cells having the wild type sequence [12]. Although a functional analysis was completed providing insight into novel molecular mechanisms of MS-AEs, no replication studies were completed thus necessitating validation in a larger independent cohort.

Non-small cell lung cancer (NSCLC)

Han *et al.* conducted a GWAS for irinotecan-related severe toxicities in patients with stage III/IV NSCLC [13]. Cases (grade 3 diarrhoea or grade 4 neutropenia) and controls were obtained from two phase II studies for the discovery phase ($n = 101$). Results demonstrated 137 SNPs for diarrhoea and 42 SNPs for neutropenia with $P < 1.0 \times 10^{-4}$ [only one reaching the established genome-wide significance level at 10^{-7}, after adjusting for multiple testing of 440 094 SNPs (best $P = 6.9 \times 10^{-8}$)]. In a replication cohort of 146 irinotecan-treated patients, rs1517114 in C8orf34, rs1661167 in FLJ41856, and rs2745761 in PLCB1 showed strong associations with diarrhoea. Rs11128347 in PDZRN3 and two SNPs (rs11979430 and rs7779029) located on chromosome 7 showed strong associations with neutropenia. High LD was observed between rs11979430 and rs7779029 (r2 = 0.911). Although UGT1A1*6 failed to reach genome-wide significance, the combined analysis showed that homozygous UGT1A1*6 was significantly associated with neutropenia. Additionally, the ABCC2 3972C>T and SLCO1B1 521T>C variants were significantly associated with diarrhoea and neutropenia, respectively, in the replication stage [13]. As 24% and 14% of patients from the discovery set and replication set experienced grade 4 neutropenia, respectively, and only 11% and 9% experienced grade 3 diarrhoea, the sample sizes are likely insufficient to identify true causal variants and validation in larger independent cohorts is necessary [14].

Wu *et al.* completed a GWAS of survival in 327 stage III/IV NSCLC patients receiving platinum-based chemotherapy [14], selected from a previous GWAS [15] of lung cancer risk. To determine which associations identified in

the discovery phase were robust, investigators identified 60 SNPs to perform a validation study using an independent NSCLC patient cohort following the same eligibility criteria as the discovery population ($n = 315$). Results demonstrated that rs1878022, found in CMKLR1, was statistically significantly associated with poorer OS in the discovery population (HR = 1.59, $P = 1.42 \times 10^{-6}$ after adjusting for multiple testing of 31 751 SNPs), and the difference in median survival time (MST) was dependent on the number of variant alleles. However, this association did not reach significance in the validation population. Another SNP, rs10937823, was associated with statistically significantly poorer OS in the discovery dataset (HR = 2.40, $P = 1.80 \times 10^{-6}$), the validation dataset (HR = 1.45, $P = 0.04$) and the combined dataset (1.82, $P = 1.73 \times 10^{-6}$). In the combined analysis, the MST for wild type patients (16.05 months) was statistically significantly longer than the MST for those carrying the variant genotypes (10.72 months, $P = 6.76 \times 10^{-5}$). To validate these findings further, both rs1878022 and rs10937823 were analyzed in similar patients enrolled in the PLATAX clinical trial (single arm study of 420 advanced NSCLC patients receiving platinum therapy). Only rs1878022 was validated in the PLATAX population (HR = 1.23, $P = 0.05$) [14]. The three stage study design in multiple independent populations allows for validation across cohorts. However, further studies to replicate these data in other patient populations with similar treatments regimens are necessary to confirm the conclusions of this study.

Similar to Wu *et al.* [14], Hu *et al.* conducted a GWAS in a total of 535 stage III/IV NSCLC patients, taken from two independent cohorts from a previous GWAS study on lung cancer susceptibility [16], treated with first-line platinum-based chemotherapy without surgery [17]. Investigators selected 33 top SNPs that had $P < 1 \times 10^{-4}$ and a consistent association with survival at $P < 0.05$ in both cohorts, when tested individually. Twelve SNPs were selected for validation using an independent NSCLC cohort from Southeastern China with baseline characteristics similar to the discovery population ($n = 340$), five of which were found to be associated with OS in all cohorts. These five SNPs were then validated in a separate Caucasian cohort from Massachusetts General Hospital with similar baseline characteristics ($n = 409$). Four SNPs (rs7629386, rs969088, rs12000445 and rs3850370) had the same direction of association, but only rs762938 and rs3850370, found in CTNNB1 and SNW1-ALKBH1-NRXN3, respectively, had significant association with NSCLC survival (HR 1.50, $P = 0.04$ and HR 1.22, $P = 0.03$). Investigators attempted to replicate the two SNPs from the previous NSCLC GWAS [14], rs1878022 and rs10937823. However both were non-significantly associated with NSCLC death risk [17]. In another three stage GWAS, investigators were able to replicate findings across patient cohorts including Han Chinese and Caucasian samples, although the heterogeneity in genetic background and treatment to NSCLC between Han Chinese and Caucasian populations may be confounding. Additionally the effect of rs7629386 may be unstable secondary to low frequency of homozygous Caucasian patients resulting in increased effect sizes in Han Chinese *vs.* Caucasians.

Sato *et al.* [18] completed a GWAS in 105 stage III/IV NSCLC patients treated on a clinical trial of single arm carboplatin and paclitaxel [19]. The following three SNPs were significantly associated with lower OS after multiple comparison adjustment (109 365 SNPs tested): rs1656402 in the EIF4E2 gene [MST for AG ($n = 50$) + AA ($n = 40$) and GG ($n = 15$) were 18.0 and 7.7 months, respectively, $P = 8.4 \times 10^{-8}$, HR = 4.22], rs1209950 in the ETS2 gene [MST for CC ($n = 94$) and CT ($n = 11$) + TT ($n = 0$) were 17.7 and 7.4 months, respectively, $P = 2.8 \times 10^{-7}$, HR = 4.96] and rs9981861 in the DSCAM gene [MST for GG ($n = 75$) + AG ($n = 26$) and AA ($n = 4$) were 17.1 and 3.8 months, respectively, $P = 3.5 \times 10^{-6}$, HR = 16.1]. Although major differences in MST were noted, initial sample sizes were insufficient and no replication studies were completed post-discovery; therefore, further validation in larger, controlled cohorts is necessary [18].

Colorectal cancer

Fernandez-Rozadilla *et al.* completed a GWAS to predict toxicity after treatment with 5-fluorouracil (5-FU) or FOLFOX (5-FU, oxaliplatin, leucovorin) in stage III/IV colorectal cancer patients [19] collected through the EPICOLON II project (a multicentre study of the prevalence and clinical pathological features of colorectal cancer) [20]. The discovery and replication phase consisted of 221 and 791 patients from the same cohort, respectively (all Caucasian Europeans, 58% treated with FOLFOX and 42% treated with 5-FU). Results demonstrated that association *P* values were very modest (best $P = 1.076 \times 10^{-5}$), with none reaching the established genome-wide significance level at 10^{-7} after adjusting for multiple comparisons. Investigators then genotyped the top five association hits in the replication phase. Copy number variation (CNV) regions showed similar *P* values but were not well tagged by SNP markers, and thus could not be replicated. Only one of these association signals was consistent in both phases: rs10876844 at 12q13.2 with 5-FU-induced diarrhoea (pooled $P = 0.010$, OR 6.5, 95%CI 1.5, 27.2). Investigators also evaluated the association signals for seven SNP variants on four genes (DPYD, MTHFR, TYMS and GSTP1) that had been linked to either 5-FU or FOLFOX related toxicity in the literature. Four of these variants had good proxy SNPs in the GWAS, although none of them showed a statistically significant association [19]. The large confidence interval in the association between rs10876844 and toxicity may reflect a possible type I error. Verification by replication in large independent data sets and fine-mapping studies, including functional assays, is necessary to clarify the biological mechanisms underlying these associations.

Acute lymphoblastic leukemia (ALL)

Trevino *et al*. utilized a GWAS approach to identify germline genetic variations associated with methotrexate pharmacokinetics and clinical effects [21]. The discovery cohort included 434 children with acute lymphoblastic leukemia (ALL) enrolled and treated on St Jude Children's Research Hospital Total XIIIB and Total XV protocols, with methotrexate and 6-mercaptopurine, followed by a validation cohort of 206 patients with similar characteristics. The strongest association was on chromosome 12 with the top two SNPs associated with the transporter gene, SLCO1B1/OATP1B1, and ultimately methotrexate clearance. These two SLCO1B1 SNPs remained significant (rs11045879, $P = 1.7 \times 10^{-10}$ and rs4149081, $P = 1.7 \times 10^{-9}$) after correction for multiple testing, and were in complete LD ($r^2 = 1$) with each other. Eight additional SLCO1B1 SNPs were associated with clearance ($P < 0.05$). These SNPs were not in LD with the top two SNPs and were encompassed by multiple haplotype blocks. SLCO1B1, rs11045879 T allele (OR 16.4, $P = 0.004$) and the G allele at rs4149081 (OR 15.3, $P = 0.03$) were each significantly associated with gastrointestinal toxicity. Seven of the 10 interrogated SNPs remained associated in the validation cohort. SLCO1B1 T521C was associated with methotrexate clearance in the discovery and the combined (discovery plus validation) patient cohorts ($P = 1.9 \times 10^{-7}$ and $P = 1.2 \times 10^{-7}$, respectively). SLCO1B1 genetic variation accounted for 9.3% of the inter-patient variability in the discovery cohort and 11.3% in the validation cohort [21]. This locus remained a significant predictor after adjusting for gender, race, renal function and age. Identifying patients at risk of low methotrexate clearance, particularly as it relates to toxicity, may be useful for monitoring and supportive care during high dose methotrexate.

Lessons and challenges in the post-GWAS era

Although several of the GWAS noted above have identified potential variants associated with treatment outcome, the biological basis of such an association is often difficult to identify in the majority of scenarios [22]. Additionally, functional studies have not always been used to aid in the interpretation of the findings. Since SNP arrays were designed to capture LD structure rather than functional variants, many GWAS identify an association between the phenotype of interest and a surrogate marker (tag SNP) rather than the causal variant [22]. Alternatively, tag SNPs may be used to identify LD blocks, which may be correlated with the phenotype of interest, as noted in the study by Kiyotani *et al*. [10]. Additionally, at some loci, multiple associated risk alleles rather than a single risk allele may be responsible for the occurrence of a particular toxicity or influence clinical outcome. Performing functional assays provides the mechanistic basis for the observed associations and could point towards the causal variants among the variants within the LD block [23].

A potential solution to the identification of predictive SNPs may involve the application of physiologically-based modelling and neural network analyses, as seen with the discovery of SNPs associated with human diseases [24]. The traditional approach to modelling the association between genetic predictors and phenotype is logistic regression. However, the number of possible interaction terms grows exponentially as each additional variable is included in the regression model. Alternatively, neural networks may offer an advantage by taking what is learned on a given dataset about the relationship between independent variables and an outcome variable and make predictions on data where the outcome variable is unknown (i.e. optimizes the input from a larger pool of variables, weights and connectivity of the network to generate 'optimal neural network architectures' for a given data set) [25].

While, the 1000 Genomes (http://www.1000genomes.org) and HapMap (hapmap.org) projects aim to capture common and rare variant information in diverse ethnic cohorts [26], they do not provide complete SNP coverage across the entire genome, including intergenic and non-coding regions where several of the GWAS associations have been mapped. This may suggest that for rare variants, or those found in intergenic regions, fine mapping by targeted sequencing may be necessary to capture the causal SNP that is in LD with the associated SNP [22]. Regulatory sequences, including enhancers, promoters, insulators and silencers, are becoming more important as we begin to characterize the landscape of susceptibility regions. For example, although many GWAS findings tend to be located in non-coding regions and intergenic, several of these SNPs may be in LD with another gene which may be controlled by a regulatory element or motif. Alternatively, they can be involved in distant regulation of a gene, acting as trans-eQTLs (expression quantitative trait loci, see later) [22, 27]. Furthermore, CNV regions may not be well tagged by SNP markers, and thus may not be replicated, as noted in the study by Fernandez-Rosadilla *et al*. [19].

Epigenetic influence on gene expression can also contribute to gene regulation [28]. Promoter methylation, histone tail modifications, and altered expression of non-coding RNAs may regulate the activity of SNPs and functional genetic associations. In fact, epigenetic silencing has been shown to be the predominant mechanism of gene silencing during tumour development for a number of genes [29]. eQTLs, genomic loci that regulate expression levels of mRNAs or proteins, have been shown to explain a greater proportion of trait variance than is typically seen for risk alleles and clinical traits. Mapping of eQTLs are another method to test the linkage between variation in expression and genetic polymorphism [30].

Encyclopedia of DNA Elements (ENCODE) (genome.ucsc.edu/ENCODE), a project funded by the National

Human Genome Research Institute, resulted in the identification of nearly all regions of transcription, transcription factor association, chromatin structure and histone modification in the human genome sequence [31]. Approximately 80% of the components of the human genome, in particular those regions outside of the well-studied protein-coding regions, now have at least one biochemical function associated with them and nearly 90% of associated SNPs from GWAS have been reported to be either intergenic or intronic [23]. With the goal of delineating all functional elements encoded in the human genome, ENCODE provides a valuable resource in determining biological associations of GWAS hits found in non-coding regions. In fact, 95% of the genome lies within 8 kilobases (kb) of a DNA-protein interaction, and 99% is within 1.7 kb of at least one of the biochemical events identified by ENCODE [31]. Other programs, such as Haploview (http://www.broadinstitute.org/haploview), are designed to facilitate LD and haplotype block analysis, permutation testing and tag SNP analysis. HaploReg (http://www.broadinstitute.org/mammals/haploreg) and RegulomeDB (regulome.stanford.edu) are additional tools used for exploring annotations of the non-coding genome at variants on haplotype blocks, including regulatory SNPs [32]. These databases utilize information available from ENCODE and LD information from the 1000 Genomes Project to help researchers develop mechanistic hypotheses of the impact of non-coding variants on clinical phenotypes.

Replication in independent cohorts is critical, preferably in a separate comparable cohort of patients [33]. The value of replication is to guard against the vulnerability of false positives observed with common alleles with low effect sizes [34]. Only a limited number of variants are genuine risk alleles, and replication helps to sift these out [35]. The initial phase of replication should test the identical index SNPs at the same direction in as similar a population as possible, with the purpose of ruling out false positivity. Replication in different populations can become complicated as there are often differences in LD pattern across populations (i.e. African Americans tend to have smaller LD blocks), in particular if population stratification is not taken into account. If a SNP fails to replicate in different populations, as is commonly seen, this may indicate the original SNP is not a causal variant nor in LD with causal variants in the replication population [35]. In cancer GWAS, replication studies are particularly challenging as the study often requires individuals with the same tumour type, treated in the same manner, and with knowledge of confounders, all within a setting where the standard of care tends to fluctuate as new therapies are tested. Additionally, the majority of GWAS are conducted from pre-marketing studies (i.e. phase II and III clinical trials) where the stringent inclusion/exclusion criteria may hinder the replication of results and, more relevant, their translation into the real world clinical arena. For example, the effect of co-administered medications, age, performance status, disease stage, comorbidities, etc. may influence the generalizability of results and make replication of findings very difficult. Replication of findings made in a GWAS from a large randomized phase III clinical trial would not only be expensive, but ethical considerations may also indicate that a second identical trial is not feasible, as noted in CALGB 80303 [7, 36]. However, clean and objective phenotype data taken from prospective phase III studies provide a more reliable and accurate association with a genetic variant, when found. Studies by Wu *et al*. [14] and Hu *et al*. [17] demonstrated the feasibility in utilizing a novel three stage design for replication. Once replicated, SNPs with a biologically plausible association, and preferably surpassing genome-wide significance, must then be validated in a candidate gene setting, ideally from a cohort of patients with DNA obtained prospectively [37].

Future directions

Although new biologic insights do not guarantee rapid translation into clinical practice, each discovery is a potential first step in a translational expedition. A complete collection of relevant genes and pathways in the hope of shortening the gap between biological knowledge and patient care is needed.

The advent of next generation sequencing techniques is now allowing a systematic analysis of rare variants and copy number changes across the entire genome. The goal of targeted sequencing is to capture the causal SNP that is in LD with the associated SNP. The probability of identifying the causal SNP may be affected by the location and boundaries of the region to be sequenced, as well as the depth of sequence coverage across the region [22]. The region to be sequenced can be guided by LD structure, given that the strength of correlation between the associated SNP and causal SNP is high. Sequencing coverage of 25 times or greater may be required, particularly in the case of sequencing-based genotyping rather than variant discovery [22].

A reservoir of observational studies and randomized clinical trials provides an extensive resource for pharmacogenomic GWAS of drug safety and efficacy [38]. Particularly, large prospective randomized clinical trials allow for improved measurement of drug exposure and concise ascertainment of phenotype. Randomization produces unbiased treatment assignments and ensures balanced baseline factors, eliminating unmeasured confounding factors [39]. Obtaining DNA prospectively and performing GWAS within clinical trials improves efficiency by identifying, measuring, and controlling for potentially interacting variables. Experts stress that effects measured in GWAS will be accurate only when gene–environment interactions are taken into account [40]. While retrospective studies are possible and may allow for larger sample sizes, they

include inconsistent treatments and noise in the data collection may confound results [36].

Secondary to the difficulty in conducting replication studies in oncology, several groups have turned to utilizing cell lines for interpreting GWAS findings [37]. Cell-based models have demonstrated that chemotherapy-induced cytotoxicity is heritable and susceptible to genetic changes [41,42]. In particular, HapMap lymphoblastoid cell lines (LCLs) are rich in genetic information and offer the opportunity to study functional *in silico* discovery, as well as replication [4]. Their ease of manipulation, extensive genotype catalogues and lack of *in vivo* confounders found in patient samples make LCLs attractive alternatives for discovery and/or replication [43]. Wheeler *et al*. identified significant overlap between 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 [43]. Brown *et al*. performed a GWAS utilizing 516 LCLs to assess cytotoxicity to temozolomide. Results identified a genome-wide significant association ($P < 10^{-8}$) with 20 SNPs (in high LD, average $r^2 > 0.9$) in the O6-methylguanine-DNA methyltransferase (MGMT) gene. Furthermore, these SNPs were found to be highly significantly associated with both IC_{50} values and MGMT transcript levels ($P < 10^{-25}$), along with a direct correlation with temozolomide response [44]. Other studies conducted in LCLs have identified potential SNPs correlating with cytotoxicity, which are subsequently confirmed or denied in a replication cohort of patients, as well as further validation in a knockdown *in vitro* model [45,46]. Utilizing LCLs as a method for functional *in silico* discovery and/or replication will prove useful, efficient, and cost-effective in capitalizing on the benefits from GWAS. Alternatively, limitations of this method exist, such as the inability to capture environmental and non-genetic influences inherent within the human body and tumour. While variables such as histology, stage, performance status, gender, smoking status, etc. may be adjusted in a Cox regression analysis to account for confounding variables, the statistical power may be low [45].

Conclusion

The success of any pharmacogenomic GWAS will be contingent upon the effect size and allele frequency of genetic variants that influence the phenotype, the sample size, the particular population, as well as the study design. Although the majority of cancer GWAS to date focus on disease susceptibility, the need for better understanding the biological evidence behind drug response, both clinical outcome and risk of toxicity, is essential. Given the large heterogeneity, debilitating toxicity and excessive treatment costs, novel methods of identifying mechanisms of variable drug response and ultimately treatment individualization will be indispensable.

Competing Interests

All authors have completed the Unified Competing Interest form at http://www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no support from any organization for the submitted work, no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years and no other relationships or activities that could appear to have influenced the submitted work.

REFERENCES

1 Metzker ML. Sequencing technologies – the next generation. Nat Rev Genet 2010; 11: 31–46.

2 Stadler ZK, Thom P, Robson ME, Weitzel JN, Kauff ND, Hurley KE, Devlin V, Gold B, Klein RJ, Offit K. Genome-wide association studies of cancer. J Clin Oncol 2010; 28: 4255–67.

3 McCarthy MI, Hirschhorn JN. Genome-wide association studies: past, present and future. Hum Mol Genet 2008; 17: R100–1.

4 International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC,

Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007; 449: 851–61.

5 Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet 2009; 5: e1000477.

6 Ku CS, Loy EY, Pawitan Y, Chia KS. The pursuit of genome-wide association studies: where are we now? J Hum Genet 2010; 55: 195–206.

7 Innocenti F, Owzar K, Cox NL, Evans P, Kubo M, Zembutsu H, Jiang C, Hollis D, Mushiroda T, Li L, Friedman P, Wang L, Glubb D, Hurwitz H, Giacomini KM, McLeod HL, Goldberg RM, Schilsky RL, Kindler HL, Nakamura Y, Ratain MJ. A genome-wide association study of overall survival in pancreatic cancer patients treated with gemcitabine in CALGB 80303. Clin Cancer Res 2012; 18: 577–84.

8 Kiyotani K, Uno S, Mushiroda T, Takahashi A, Kubo M, Mitsuhata N, Ina S, Kihara C, Kimura Y, Yamaue H, Hirata K, Nakamura Y, Zembutsu H. A genome-wide association study identifies four genetic markers for hematological toxicities in cancer patients receiving gemcitabine therapy. Pharmacogenet Genomics 2012; 22: 229–35.

9 Baldwin RM, Owzar K, Zembutsu H, Chhibber A, Kubo M, Jiang C, Watson D, Eclov RJ, Mefford J, McLeod HL, Friedman PN, Hudis CA, Winer EP, Jorgenson EM, Witte JS, Shulman LN, Nakamura Y, Ratain MJ, Kroetz DL. A genome-wide association study identifies novel loci for paclitaxel-induced sensory peripheral neuropathy in CALGB 40101. Clin Cancer Res 2012; 18: 5099–109.

10 Kiyotani K, Mushiroda T, Tsunoda T, Morizono T, Hosono N, Kubo M, Tanigawara Y, Imamura CK, Flockhart DA, Aki F, Hirata K, Takatsuka Y, Okazaki M, Ohsumi S, Yamakawa T, Sasa M, Nakamura Y, Zembutsu H. A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. Hum Mol Genet 2012; 21: 1665–72.

11 Srinivasan Y, Sasa M, Honda J, Takahashi A, Uno S, Kamatani N, Kubo M, Nakamura Y, Zembutsu H. Genome-wide association study of epirubicin-induced leukopenia in Japanese patients. Pharmacogenet Genomics 2011; 21: 552–8.

12 Ingle JN, Schaid DJ, Goss PE, Liu M, Mushiroda T, Chapman JA, Kubo M, Jenkins GD, Batzler A, Shepherd L, Pater J, Wang L, Ellis MJ, Stearns V, Rohrer DC, Goetz MP, Pritchard KI, Flockhart DA, Nakamura Y, Weinshilboum RM. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. J Clin Oncol 2010; 28: 4674–82.

13 Han JY, Shin ES, Lee YS, Ghang HY, Kim SY, Hwang JA, Kim JY, Lee JS. A genome-wide association study for irinotecan-related severe toxicities in patients with advanced non-small-cell lung cancer. Pharmacogenomics J 2012; 1–6. Epub ahead of print.

14 Wu X, Ye Y, Rosell R, Amos CI, Stewart DJ, Hildebrandt MA, Roth JA, Minna JD, Gu J, Lin J, Buch SC, Nukui T, Ramirez Serrano JL, Taron M, Cassidy A, Lu C, Chang JY, Lippman SM, Hong WK, Spitz MR, Romkes M, Yang P. Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy. J Natl Cancer Inst 2011; 103: 817–25.

15 Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 2008; 40: 616–22.

16 Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, Yang L, Dai J, Hu L, Tan W, Li Z, Deng Q, Wang J, Wu W, Jin G, Jiang Y, Yu D, Zhou G, Chen H, Guan P, Chen Y, Shu Y, Xu L, Liu X, Liu L, Xu P, Han B, Bai C, Zhao Y, Zhang H, Yan Y, Ma H, Chen J, Chu M, Lu F, Zhang Z, Chen F, Wang X, Jin L, Lu J, Zhou B, Lu D, Wu T, Lin D, Shen H. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat Genet 2011; 43: 792–6.

17 Hu L, Wu C, Zhao X, Heist R, Su L, Zhao Y, Han B, Cao S, Chu M, Dai J, Dong J, Shu Y, Xu L, Chen Y, Wang Y, Lu F, Jiang Y, Yu D, Chen H, Tan W, Ma H, Chen J, Jin G, Wu T, Lu D, Christiani DC, Lin D, Hu Z, Shen H. Genome-wide association study of prognosis in advanced non-small cell lung cancer patients receiving platinum-based chemotherapy. Clin Cancer Res 2012; 18: 5507–14.

18 Sato Y, Yamamoto N, Kunitoh H, Ohe Y, Minami H, Laird NM, Katori N, Saito Y, Ohnami S, Sakamoto H, Sawada J, Saijo N, Yoshida T, Tamura T. Genome-wide association study on overall survival of advanced non-small cell lung cancer patients treated with carboplatin and paclitaxel. J Thorac Oncol 2011; 6: 132–8.

19 Fernandez-Rozadilla C, Cazier JB, Moreno V, Crous-Bou M, Guino E, Duran G, Lamas MJ, Lopez R, Candamio S, Gallardo E, Pare L, Baiget M, Paez D, Lopez-Fernandez LA, Cortejoso L, Garcia MI, Bujanda L, Gonzalez D, Gonzalo V, Rodrigo L, Rene JM, Jover R, Brea-Fernandez A, Andreu M, Bessa X, Llor X, Xicola R, Palles C, Tomlinson I, Castellvi-Bel S, Castells A, Ruiz-Ponte C, Carracedo A. Pharmacogenomics in colorectal cancer: a genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. Pharmacogenomics J 2013; 13: 209–17.

20 Pinol V, Castells A, Andreu M, Castellvi-Bel S, Alenda C, Llor X, Xicola RM, Rodriguez-Moranta F, Paya A, Jover R, Bessa X, Gastrointestinal Oncology Group of the Spanish Gastroenterological A. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. JAMA 2005; 293: 1986–94.

21 Trevino LR, Shimasaki N, Yang W, Panetta JC, Cheng C, Pei D, Chan D, Sparreboom A, Giacomini KM, Pui CH, Evans WE, Relling MV. Germline genetic variation in an organic anion

transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. J Clin Oncol 2009; 27: 5972–8.

22 Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, James M, Liu P, Tichelaar JW, Vikis HG, You M, Mills IG. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet 2011; 43: 513–8.

23 Glubb DM, Innocenti F. Architecture of pharmacogenomic associations: structures with functional foundations or castles made of sand? Pharmacogenomics 2013; 14: 1–4.

24 Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. Am J Hum Genet 2010; 86: 6–22.

25 Motsinger AA, Lee SL, Mellick G, Ritchie MD. GPNN: power studies and applications of a neural network method for detecting gene-gene interactions in studies of human disease. BMC Bioinformatics 2006; 7: 39–48.

26 Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: new opportunities for research and social challenges. Genome Med 2010; 2: 1–3.

27 Glubb DM, Dholakia N, Innocenti F. Liver expression quantitative trait loci: a foundation for pharmacogenomic research. Front Genet 2012; 3: 153.

28 Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007; 128: 683–92.

29 Raval A, Tanner SM, Byrd JC, Angerman EB, Perko JD, Chen SS, Hackanson B, Grever MR, Lucas DM, Matkovic JJ, Lin TS, Kipps TJ, Murray F, Weisenburger D, Sanger W, Lynch J, Watson P, Jansen M, Yoshinaga Y, Rosenquist R, de Jong PJ, Coggill P, Beck S, Lynch H, de la Chapelle A, Plass C. Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. Cell 2007; 129: 879–90.

30 Cheung VG, Spielman RS. Genetics of human gene expression: mapping DNA variants that influence gene expression. Nat Rev Genet 2009; 10: 595–604.

31 Consortium EP, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kelllis M, Khatun J, Kheradpour P, Kundaje A, Lassman T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SC, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LA, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elinitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigo R, Hardison RC, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Snyder M, Starnatoyannopoulos JA, Tennebaum SA, Weng Z, White KP, Wold B, Khatun J, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Giddings MC, Bernstein BE, Epstein CB, Shoresh N, Ernst J, Kheradpour P, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altshuler RC, Eaton ML, Kellis M, Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena HP, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Risk BA, Robyr D, Ruan X, Sammeth M, Sandu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Hayashizaki Y, Harrow J, Gerstein M, Hubbard TJ, Reymond A, Antonarakis SE, Hannon GJ, Giddings MC, Ruan Y, Wold B, Carninci P, Guigo R, Gingeras TR, Rosenbloom KR, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kent WJ, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Furey TS, Song L, Grasfeder LL, Giresi PG, Lee BK, Battenhouse A, Sheffield NC, Simon JM, Showers KA, Safi A, London D, Bhinge AA, Shestak C, Schaner MR, Kim SK, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Birney E, Iyer VR, Lieb JD, Crawford GE, Li G, Sandhu KS, Zheng M, Wang P, Luo OJ, Shahab A, Fullwood MJ, Ruan X, Ruan Y, Myers RM, Pauli F, Williams BA, Gertz J, Marinov GK, Reddy TE, Vielmetter J, Partridge EC, Trout D, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, King B, Muratet MA, Antoshechkin I, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Gunter C, Newberry JS, Levy SE, Absher DM, Mortazavi A, Wong WH, Wold B, Blow MJ, Visel A, Pennachio LA, Elnitski L, Margulies EH, Parker SC, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Chrast J, Davidson C, Derrien T, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Howald C, Hunt T, Jungreis I, Kay M, Khurana E, Kokocinski F, Leng J, Lin MF, Loveland J, Lu Z, Manthravadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tanzer A, Tapanan E, Tress ML, van Baren MJ, Walters N, Washieti S, Wilming L, Zadissa A, Zhengdong Z, Brent M, Haussler D, Kellis M, Valencia A, Gerstein M, Raymond A, Guigo R, Harrow J, Hubbard TJ, Landt SG, Frietze S, Abyzov A, Addleman N, Alexander RP, Auerbach RK, Balasubramanian S, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Cheng C, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyenger S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Larnarre-Vincent N, Leng J, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Pei B, Raha D, Ramirez L, Reed B, Rozowsky J, Sboner A, Shi M, Sisu C, Slifer T, Witt H, Wu L, Xu X, Yan KK, Yang X, Yip KY, Zhang Z, Struhl K, Weissman SM, Gerstein M, Farnham PJ, Snyder M, Tenebaum SA, Penalva

LO, Doyle F, Karmakar S, Landt SG, Bhanvadia RR, Choudhury A, Domanus M, Ma L, Moran J, Patacsil D, Slifer T, Victorsen A, Yang X, Snyder M, White KP, Auer T, Centarin L, Eichenlaub M, Gruhl F, Heerman S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Weng Z, Whitfield TW, Wang J, Collins PJ, Aldred SF, Trinklein ND, Partridge EC, Myers RM, Dekker J, Jain G, Lajoie BR, Sanyal A, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Ebersol AK, Frum T, Garg K, Gist E, Hansen RS, Boatman L, Haugen E, Humbert R, Jain G, Johnson AK, Johnson EM, Kutyavin TM, Lajoie BR, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sabo P, Sanchez ME, Sandstrom RS, Sanyal A, Shafer AO, Stergachis AB, Thomas S, Thurman RE, Vernot B, Vierstra J, Vong S, Wang H, Weaver MA, Yan Y, Zhang M, Akey JA, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Kaul R, Dekker J, Stamatoyannopoulos JA, Dunham I, Beal K, Brazma A, Flicek P, Herrero J, Johnson N, Keefe D, Lukk M, Luscombe NM, Sobral D, Vaquerizas JM, Wilder SP, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Kundaje A, Hardison RC, Miller W, Giardine B, Harris RS, Wu W, Bickel PJ, Banfai B, Boley NP, Brown JB, Huang H, Li Q, Li JJ, Noble WS, Bilmes JA, Buske OJ, Hoffman MM, Sahu AO, Kharchenko PV, Park PJ, Baker D, Taylor J, Weng Z, Iyer S, Dong X, Greven M, Lin X, Wang J, Xi HS, Zhuang J, Gerstein M, Alexander RP, Balasubramanian S, Cheng C, Harmanci A, Lochovsky L, Min R, Mu XJ, Rozowsky J, Yan KK, Yip KY, Birney E. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489: 57–74.

32 Glubb D, Paugh WS, van Schaik RHN, Innocenti F. A Guide to the Current Web-Based Resources in Pharmacogenomics. Totowa, NJ: Humana Press, Inc, 2013.

33 Studies N-NWGoRiA, Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS. Replicating genotype-phenotype associations. Nature 2007; 447: 655–60.

34 Chung CC, Magalhaes WC, Gonzalez-Bosquet J, Chanock SJ. Genome-wide association studies in cancer – current and future directions. Carcinogenesis 2010; 31: 111–20.

35 Huo D, Olopade OI. Interpretation of genome-wide association study results. Oncology (Huntingt) 2010; 24. Available at http://www.cancernetwork.com/cancer-genetics/content/article/10165/1595344# (last accessed 19 June 2013).

36 Wheeler HE, Maitland ML, Dolan ME, Cox NJ, Ratain MJ. Cancer pharmacogenomics: strategies and challenges. Nat Rev Genet 2012; 14: 23–34.

37 Huang RS, Dolan ME. Approaches to the discovery of pharmacogenomic markers in oncology: 2000-2010-2020. Pharmacogenomics 2010; 11: 471–4.

38 Guessous I, Gwinn M, Khoury MJ. Genome-wide association studies in pharmacogenomics: untapped potential for translation. Genome Med 2009; 1: 46.1–46.4.

39 Maitland ML, Ratain MJ, Cox NJ. Interpreting P values in pharmacogenetic studies: a call for process and perspective. J Clin Oncol 2007; 25: 4513–5.

40 Khoury MJ, Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies – challenges and opportunities. Am J Epidemiol 2009; 169: 227–30; discussion 234–5.

41 Dolan ME, Newbold KG, Nagasubramanian R, Wu X, Ratain MJ, Cook EH Jr, Badner JA. Heritability and linkage analysis of sensitivity to cisplatin-induced cytotoxicity. Cancer Res 2004; 64: 4353–6.

42 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.

43 Wheeler HE, Gamazon ER, Wing C, Njiaju UO, Njoku C, Baldwin RM, Owzar K, Jiang C, Watson D, Shterev I, Kubo M, Zembutsu H, Winer EP, Hudis CA, Shulman L, Nakamura Y, Ratain MJ, Kroetz D, Cox NJ, Dolan ME. Integration of cell line and clinical trial genome-wide analyses supports a polygenic architecture of paclitaxel-induced sensory peripheral neuropathy. Clin Cancer Res 2013; 19: 491–9.

44 Brown CC, Havener TM, Medina MW, Auman JT, Mangravite LM, Krauss RM, McLeod HL, Motsinger-Reif AA. A genome-wide association analysis of temozolomide response using lymphoblastoid cell lines shows a clinically relevant association with MGMT. Pharmacogenet Genomics 2012; 22: 796–802.

45 Niu N, Schaid DJ, Abo RP, Kalari K, Fridley BL, Feng Q, Jenkins GD, Batzler A. Genetic association with overall survival of taxane-treated lung cancer patients – a genome-wide association study in human lymphoblastoid cell lines followed by a clinical association study. BMC Cancer 2012; 12: 422–34.

46 Tan XL, Moyer AM, Fridley BL, Schaid DJ, Niu N, Batzler AJ, Jenkins GD, Abo RP, Li L, Cunningham JM, Sun Z, Yang P, Wang L. Genetic variation predicting cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based chemotherapy. Clin Cancer Res 2011; 17: 5801–11.