GENOME-WIDE ASSOCIATION STUDIES

Genome-wide association studies in pharmacogenomics

Ann K. Daly

Abstract | Genome-wide association (GWA) studies for pharmacogenomics-related traits are increasingly being performed to identify loci that affect either drug response or susceptibility to adverse drug reactions. Until now, only the largest effects have been detected, partly because of the challenges of obtaining large numbers of cases for pharmacogenomic studies. Since 2007, a range of pharmacogenomics GWA studies have been published that have identified several interesting and novel associations between drug responses or reactions and clinically relevant loci, showing the value of this approach.

The term pharmacogenomics, which first appeared in the literature in 1997 (REF. 1), is often described as the whole-genome application of pharmacogenetics, and there is a large overlap between these two disciplines. Until the 1990s, pharmacogenetic studies were concerned with the effects of single genes on drug disposition and drug response but, in the era of pharmacogenomics, the effects of several genes on a particular phenotype are typically investigated, especially when inter-individual variation exists in the response to a particular drug or when adverse drug reactions are known to occur.

Until recently, such phenotypes have been related to particular genotypes through candidate-gene studies. The genes investigated include those encoding enzymes involved in drug metabolism, such as the cytochrome P450 oxidases, and genes coding for drug targets, typically receptors or enzymes. However, in the case of adverse drug reactions, it is not just genes encoding metabolic enzymes that are relevant; a wide range of genes, including those related to the immune response and mitochondrial functions, can be involved. These candidate studies have led to the identification of genes with important contributions to drug response — for example, those that influence the effect of anticoagulant drugs, such as warfarin[2].

However, since 2007, genome-wide association (GWA) studies have increasingly been applied to pharmacogenomics. The large-effect phenotypes shown by pharmacogenomics traits allow powerful studies to be carried out on smaller sample sizes (BOX 1).

A survey of the published literature reveals that approximately 70% of published GWA studies in pharmacogenomics are concerned with drug response, and the remaining studies are concerned with adverse drug reactions. In studies on drug response, genome-wide significant associations have been detected for interferon-α[3–5] and clopidogrel[6] response, and for anticoagulant dose requirement[7–9]. For adverse drug reactions, significant associations have been reported for statin-induced myopathy[10] and flucloxacillin-induced liver injury[11]. In most cases these studies have reported novel findings and have made important contributions to the field, with some potentially influencing clinical practice.

This article reviews the range of published GWA studies in pharmacogenomics, with particular reference to those reporting genome-wide significant associations. The potential problems of conducting such studies are also considered together with some specific differences between pharmacogenomics GWA studies and those on susceptibility to complex diseases.

Study designs and challenges

Candidate-gene studies. Candidate-gene studies have provided valuable data in the areas of pharmacogenetics and pharmacogenomics. This is especially the case for adverse drug reactions attributable to alleles of a single gene, often one that encodes an enzyme contributing to metabolism of the drug. For example, thiopurine drugs (such as azathioprine), which are used as immunosuppressants and in cancer treatment, can cause bone-marrow suppression associated with inactivating polymorphisms in thiopurine *S*-methyltransferase (*TPMT*)[12]. There is also an association between a particular uridine diphosphate glucuronosyltransferase 1A1 (*UGT1A1*) allele (*UGT1A1*28*) and neutropenia induced by irinotecan, another anticancer drug[13]. The US Food and Drug Administration (FDA) now recommends testing for *TPMT* variants in patients being treated with azathioprine and suggests a lower dose of irinotecan for patients known to be homozygous for the *UGT1A1*28* allele[14]. Susceptibility to hypersensitivity reactions induced by the anti-HIV drug abacavir was also shown to relate to a single gene in several candidate studies[15,16]. The gene involved is not a metabolic gene but the class I human leukocyte antigen gene *HLA-B*; most hypersensitivity cases carry the *HLA-B*5701* allele, so genotyping for *HLA-B*5701* before commencing abacavir treatment is recommended by the FDA[14] and other drug regulators worldwide. There is also limited evidence from candidate-gene studies showing that individuals with polymorphisms that result in the absence of activity of cytochrome P450, family 2, subfamily D, polypeptide 6 (CYP2D6) may show more adverse events during antidepressant drug treatment[17]. Dose adjustments for individual antidepressant drugs on the basis of the *CYP2D6* genotype have been proposed[18], although these have not been tested prospectively and are not generally recommended by regulatory agencies.

There are several examples of the successful use of candidate-gene approaches for investigating drug toxicity, but there are fewer examples of the successful use of such approaches for understanding overall drug response, possibly because several genes — not necessarily obvious candidates, such as the metabolic genes — may contribute to response. An exception to this is the response to coumarin anticoagulants; here, there is extensive literature from candidate-gene

Box 1 | **Differences between pharmacogenomics and complex-disease GWA studies**

In pharmacogenomics, some issues relating to performing genome-wide association (GWA) studies may be different to those encountered in studies on complex diseases. These differences include:

Sample size. A number of published GWA studies on pharmacogenomics have failed to show a large enough effect for genome-wide significance; the main reason for this is probably that they mainly involve small sample sizes with insufficient power to detect small or moderately sized effects. Obtaining adequate numbers of cases for pharmacogenomics GWA studies is more challenging than for most studies on complex-disease genetics. Drug non-responders are often rarer than responders, and serious adverse drug reactions, depending on the precise drug involved, often only affect one in every 10,000 to 100,000 patients treated.

Phenotypic characterization. Heterogeneity in phenotype is likely to be a problem in some studies, especially because drug response often cannot be measured in a fully quantitative way. Exceptions include the case of drug response with coumarin anticoagulants: here, treatment is already individualized on the basis of phenotypic response, and so allows the continuous variable of drug dose to be related to genotype. Other exceptions include drugs used in hypertension and diabetes treatment, in which reliable clinical measures of response are used routinely.

Replication of findings. Replication is an important feature of GWA studies[47]. However, the problem with small sample sizes in many pharmacogenomics studies makes replication of findings more challenging than in studies on common complex diseases. TABLE 1 and TABLE 2 list some associations for which replication has been possible, even if the replication set is smaller than the original sample set. Pharmacogenomic GWA studies often involve randomized clinical trials, which are large and very expensive. Therefore, high costs and, in some cases, ethical considerations may mean a second randomized trial is not feasible and no replication set is available. As an alternative to using a replication cohort, functional validation of a SNP showing a significant association may be possible.

Effect size. Several recent pharmacogenomic studies involving around 100 cases have detected genome-wide significant associations, suggesting that at least some pharmacogenomic effects tend to be larger and involve fewer genes than those detected in GWA studies for complex diseases. This may not be universally true for pharmacogenomics and, if larger sample sizes are available, lower effect sizes similar to those seen in most complex-disease GWA studies may also be detected.

studies on genetic factors that affect coumarin dose requirement, with only minor variations in findings among studies (see below)[2]. Other examples of drugs for which candidate-gene studies have provided insights into response include clopidogrel — the *CYP2C19* genotype affects the response directly[19] and also influences the treatment outcome[20,21] — and tamoxifen. In the case of tamoxifen, there is a replicated effect of the *CYP2D6* genotype on treatment outcome[22,23].

GWA studies. Candidate-gene studies have laid a solid foundation for personalized medicine, and several findings from the studies described above are now in clinical use. However, the availability of GWA approaches enables contributions from novel and less obvious genes to be detected, especially in the area of drug-target genetics, which is more complex and less well understood than the pharmacogenetics of drug metabolism[24]. Examples of these success stories are described in the following sections.

GWA studies relating to pharmacogenomics present some different challenges to those concerned with complex diseases (BOX 1).

GWA studies on drug response

Patients can vary in their response to prescribed drugs. For example, some patients do not respond at all to certain drug treatments, and for other drugs the precise dose required to achieve a response may vary. Until now, GWA studies on drug response have been mainly concerned with drugs for which the dose needs to individualized (coumarin anticoagulants[7–9]) or for which failure to respond presents an important clinical problem (for example, interferon-α treatment in hepatitis C infection[3,4,8] or anti-tumour necrosis factor (TNF) treatment in rheumatoid arthritis[25]). In some cases, GWA has also been applied to phase III clinical trials of new drugs for which DNA samples and detailed data on response are already being collected as part of the study (for example, iloperidone[26]; see TABLE 1).

At least 16 different GWA studies on drug response have appeared since the first was published in late 2007. For the reasons explained in BOX 1, less than half of these have shown genome-wide significance, although for several of the studies in this category some potentially interesting associations that come close to significance have been detected that should be investigated in larger cohorts. Despite these limitations, TABLE 1 lists some important success stories showing the value of GWA studies in pharmacogenomics. The salient findings are described in detail below. The studies on coumarin anticoagulants, together with a study on another drug used in cardiovascular disease, clopidogrel, showed genome-wide significance for at least one marker[6–9]. Three recent studies on response to interferon-α treatment in hepatitis C infection have also reported genome-wide significance for SNPs in a single novel gene[3–5].

Response to coumarin anticoagulants. Before any GWA study was initiated, there was good evidence that up to 50% of variability in dose requirement for the coumarin anticoagulants warfarin and acenocoumarol could be explained by patient-specific factors, both genetic and non-genetic (for example, age and body mass index). One-third (~35%) of this variation related to two genes: *CYP2C9*, which encodes the main metabolizing enzyme for both drugs, and vitamin K epoxide reductase complex 1 (*VKORC1*), which encodes the drug target[2]. GWA studies were performed with the main objective of identifying additional genetic factors that might explain a higher percentage of inter-patient variability.

The first of two published GWA studies on warfarin involved 181 patients and found significance at the *VKORC1* gene only[7]. *CYP2C9* did not show genome-wide significance, but one SNP in this gene showed a p value of ~10^{-5}. Further genotyping in a replication cohort confirmed the strong effect for *VKORC1* and the weaker *CYP2C9* effect; it was concluded that markers in these two genes were the main genetic predictors of dose, with only minor roles for other, unknown genes. In a subsequent study involving 1,053 patients, SNPs in both *VKORC1* and *CYP2C9* showed genome-wide significance with lower p values, reflecting the increased study size[8]. This study also found genome-wide significance for a SNP in a third gene, *CYP4F2*, through an additional multivariate regression analysis that involved adjustment for all known genetic and non-genetic factors. Although the effect from *CYP4F2* had also been identified in studies using candidate-gene approaches[27–29], the demonstration of its relevance in the GWA study was novel. Furthermore, the new data showed that the *CYP4F2* SNP explains ~1.5% of warfarin dose variability. No additional gene was detected that affected at least 1.5% of the dose variability.

Table 1 | **Published pharmacogenomics GWA studies on drug response**

Drug	Response	Number of cases	Genome-wide significance/ replication	Lowest *p* value	SNP genotyping platform	Significant gene(s)	Refs
Warfarin	Maintenance dose	181; 1,053	Yes/Yes	6×10^{-13}; 1×10^{-78}	Illumina 550K; Illumina CNV370	*VKORC1*, *CYP2C9*, *CYP4F2*	7,8
Acenocoumarol	Maintenance dose	1,451	Yes/Yes	2×10^{-123}	Illumina 550K	*VKORC1*, *CYP2C9*, *CYP4F2*, *CYP2C18*	9
Interferon-α	Response in hepatitis C infection assessed by absence of viral RNA in serum	1,137; 293; 154	Yes/Yes	1×10^{-25}; 9×10^{-9}; 3×10^{-15}	Illumina 610 quad	*IL28B*	3–5
Clopidogrel	Antiplatelet effect assessed by platelet aggregometry	429	Yes/Yes	4×10^{-11}	Affymetrix 500K or 1M	*CYP2C19*	6
Methotrexate	Drug clearance and incidence of gastrointestinal toxicity in childhood leukaemia patients	434	Yes/Yes	1.7×10^{-10}	Affymetrix 500K	*SLCO1B1*	37
Thiazide	Diuretic response assessed by diastolic blood pressure change	389*	Yes/No	2×10^{-7} ‡	Affymetrix 100K	Region of chromosome 12q15	49
Interferon-β	Response in multiple sclerosis assessed by clinical scoring	206	No/No	0.004	Affymetrix 100K	None found	48
Anti-tumour necrosis factor	Response in rheumatoid arthritis assessed by clinical scoring	89	No/No	0.00009	Illumina 300K	None found	25
Methylphenidate	Response in attention-deficit hyperactivity syndrome assessed by clinical scoring	309	No/No	3×10^{-6}	Affymetrix 6.0	None found	50
Various	Response in childhood leukaemia based on presence of minimal residual disease	487	No/No	1.5×10^{-6}	Affymetrix 100K and 500K	None found	51
Iloperidone	Response in schizophrenia assessed by clinical scoring	210	No/No	1×10^{-7} §	Affymetrix 500K	None found	26
Antidepressants	Response in depression assessed by clinical scoring	339	No/No	8×10^{-7}	Illumina Human-1 and Hap300	None found	52
Citalopram	Response in depression assessed by clinical scoring	1,948	No/No	5×10^{-7} (response); 4×10^{-7} (remission)	Affymetrix 500K and 5.0	None found	53

*50% good responders, 50% poor responders. ‡Genome-wide significant difference between good responders and poor responders of African-American origin only. §Not formally significant after correction for multiple testing. CYP, cytochrome P450; GWA, genome-wide association; *IL28B*, interleukin-28B (also known as interferon-λ3); *SLCO1B1*, solute carrier organic anion transporter family, member 1B1; *VKORC1*, vitamin K epoxide reductase complex 1.

Similarly, a study for another coumarin anticoagulant, acenocoumarol, showed genome-wide significance across 1,451 patients for the *VKORC1* and *CYP2C9* SNPs, and *CYP4F2* reached significance after adjustment for the two strongest genetic factors[9], in agreement with recent findings for warfarin[8]. A novel feature of the acenocoumarol study was an additional genome-wide significant association of dose with SNPs in the *CYP2C18* gene, after adjustment for *VKORC1* and *CYP2C9*. As no substrates for CYP2C18 have so far been identified, the association may relate to the adjacent *CYP2C19* gene, which codes for an enzyme with a minor role in acenocoumarol metabolism[30].

The three published GWA studies on coumarin dose requirement have therefore been valuable in confirming previous observations from candidate-gene associations. However, they have not detected novel genetic factors that affect dosage, apart from the suggestion of a small additional contribution from *CYP2C19* to acenocoumarol dose. Additional GWA studies in this area are in progress and there is hope of detecting further small-effect SNPs or genetic factors that affect aspects of coumarin anticoagulant response other than dosage.

Response to antiplatelet drugs. Response to another drug used widely in the treatment of cardiovascular disease, the antiplatelet drug clopidogrel, has also been investigated by a GWA study[6]. The main study concerned response to the drug in a healthy volunteer group. The most significant SNP ($p = 1.5 \times 10^{-13}$ for response in an additive model) is in strong linkage disequilibrium (LD) with a polymorphism in the *CYP2C19*2* allele, which is known to be associated with absence of CYP2C19 enzyme activity[31]. No SNPs outside the *CYP2C* locus showed genome-wide significance, providing no evidence for a strong effect by other genetic factors on clopidogrel response. The association of clopidogrel response with *CYP2C19* had already been reported (see section on candidate-gene studies) so, like the studies on warfarin, the GWA study provided confirmation of earlier results rather than novel data.

Hepatitis C: response to interferon-α. A recent example of the successful application of GWA to the study of drug response is in the area of infectious diseases research. The main treatment for hepatitis C virus infection involves administration of a modified form of interferon-α together with the

Table 2 | **Published pharmacogenomics GWA studies on adverse drug reactions**

Drug	Response	Number of cases/ controls	Genome-wide significance/ replication	Lowest *p* value	SNP genotyping platform	Significant gene	Refs
Simvastatin	Myopathy assessed by symptoms and elevated creatine kinase	85/90	Yes/Yes	4×10^{-9}	Illumina Hap 300K	*SLCO1B1*	10
Flucloxacillin	Liver injury assessed by elevated liver enzymes and/or bilirubin	51/282	Yes/Yes	8.7×10^{-33}	Illumina 1M	*HLA-B*5701*	11
Etoposide	Secondary leukaemia diagnosed cytogenetically	13/169	No	<0.001	Affymetrix 100K	None found	54
Antipsychotics	Tardive dyskinesia assessed by clinical scoring	50/50	No	0.00005	Illumina Human 1 (approx. 40,000 tag SNPs)	None found	55
Ximelagatran	Transaminitis* diagnosed by elevated alanine transaminase activity	74/169	No	6×10^{-6}	Perlegen 250K	None found	39
Bisphosphonates	Osteonecrosis of the jaw diagnosed by clinical examination	25/65	No	1×10^{-6}	Affymetrix 500K	None found	41
Iloperidone	QT‡ prolongation diagnosed by electrocardiography	183/0	No	1.6×10^{-6}	Affymetrix 500K	None found	56

*The elevation of plasma alanine aminotransferase or aspartate aminotransferase activity above the normal maximum level. This indicates the possibility of liver damage. ‡The QT interval is a measure of the time between the start of the Q wave and the end of the T wave in the heart's electrical cycle. GWA, genome-wide association; HLA, human leukocyte antigen; *SLCO1B1*, solute carrier organic anion transporter family, member 1B1.

antiviral drug ribavirin. Depending on ethnic origin, up to 50% of patients show a positive response to this treatment and become essentially virus-free. Three independent GWA studies show that polymorphisms adjacent to interleukin-28B (*IL28B*, also known as interferon-λ3) predict the likelihood of a positive response to interferon-α[3–5]. The findings were genome-wide significant and were validated further by the fact that the three studies involved different ethnic groups, all of whom showed the significant *IL28B* effect. This finding provides interesting new insights into the disease process and may be valuable in determining treatment. However, patient genotype is not a perfect predictor of response to interferon-α, and the current absence of alternative treatments for hepatitis C infection is a limitation in using the *IL28B* genotype to personalize treatment[32].

GWA studies on adverse drug reactions

There are currently fewer published GWA studies on adverse drug reactions than there are on adverse drug response; this is probably because many serious adverse drug reactions are rare, and are therefore a more difficult area to investigate than drug response. Although there have been many more published candidate-gene studies on adverse drug reactions than on overall drug response, the reverse is true when it comes to GWA studies. At least seven separate GWA studies on drug toxicity have now appeared[33,34], but only two of these report genome-wide significance and, with one exception, the number of cases in each study is below 100 (TABLE 2). The two studies that showed significant associations also included small replication cohorts that confirmed the associations. As with studies on drug response, there were additional SNPs that came close to achieving genome-wide significance.

Simvastatin-induced myopathy. Of the two studies that reported genome-wide significance, the first was concerned with simvastatin-induced myopathy. The study analysed 310,000 SNP markers in 85 cases of myopathy and 90 controls who had been exposed to the drug without any evidence of myopathy[10]. A highly significant association was seen for a single SNP in solute carrier organic anion transporter family, member 1B1 (*SLCO1B1*) with an odds ratio of 4.5 per copy of the variant allele. This effect was confirmed in 21 cases of myopathy from a separate replication cohort. *SLCO1B1* encodes a transporter located on the sinusoidal face of the hepatocyte, and is known to be the main inward transporter for a number of different statins. The significant SNP was in strong LD with a non-synonymous SNP that had previously been shown to impair statin transport and was associated with higher plasma levels of statins[35]. This association is therefore highly plausible; however, it could also have been detected by candidate-gene approaches, as the role of the SLCO1B1 drug transporter in statin transport is well established. In addition, the significant polymorphism is common (variant allele frequency is 0.13 among controls) and possession of the variant only explains around 18% of attributable risk, as substantial numbers of myopathy cases (~40%) are homozygous for wild-type *SLCO1B1*. The study had limited power to detect variants with a smaller effect than *SLCO1B1*, but no suggestion of significant effects for a list of candidate genes was obtained. There is a need for further, larger studies with the power to detect smaller effects so that a higher proportion of risk for this toxicity can be explained. The association of muscle injury with *SLCO1B1* has recently been confirmed for milder toxicity in a candidate-gene study[36]. The clearance of an unrelated drug, methotrexate (which is used in the treatment of cancer and as an immunosuppressant), has recently been shown to be predicted by the same *SLCO1B1* SNP as simvastatin-induced myopathy, and a form of toxic response to methotrexate has also been linked to this SNP[37] (TABLE 1).

Flucloxacillin-induced liver injury. The second example of a serious adverse reaction showing a genome-wide significant association in a GWA study is liver injury induced by flucloxacillin. Flucloxacillin is an antimicrobial used widely in Europe and Australia in the treatment of staphylococcal infections. Its use has been associated with a rare but characteristic cholestatic hepatitis. In a GWA study involving 51 UK cases and 282 controls, a strong association ($p = 8.7 \times 10^{-33}$) was obtained with a SNP in HLA complex P5 (*HCP5*), which is located in the major histocompatibility complex region[11]. This SNP had been previously shown to be in complete LD with the *HLA-B*5701* allele. Possession of the *HLA-B*5701* allele was associated with an 80-fold increased risk of adverse reaction development, a finding that was

replicated in a small additional cohort of 23 cases. The *HLA-B*5701* association has also been reported for another adverse drug reaction, abacavir-related hypersensitivity, using candidate-gene approaches[15] (see above), but has not been associated previously with drug-induced liver injury.

The GWA study also suggested that a non-HLA gene, ST6 β-galactosamide α2,6-sialyltranferase 1 (*ST6GAL1*), which has a possible role in B cell immune responses, also contributed to flucloxacillin toxicity but not as strongly as *HLA-B*5701*. To observe genome-wide significance for the *ST6GAL1* SNP, it was necessary to exclude any samples that were negative for *HLA-B*5701*. The *HLA-B*5701* association was previously detected in the same populations by candidate-gene approaches[38], but it is unlikely that an association with *ST6GAL1* would have been considered in a candidate-gene study. The relatively small but interesting contribution from this gene to toxicity shows the value of the GWA approach.

Other adverse reactions. The other published GWA studies on adverse drug reactions focused on important toxicities but failed to find SNPs with genome-wide significance. Despite this limitation, two of these studies reported findings of interest in relation to existing pharmacogenetic data.

A study of adverse reaction to the drug ximelagatran reported an association between raised alanine aminotransferase levels (a marker for liver injury) and the class II HLA allele *DRB1*0701* (REF. 39). This association is similar to that detected for flucloxacillin-induced liver injury, in which the frequency of this class II HLA allele is also increased[11]. However, the ximelagatran study found no associations in the class I HLA region, and it seems likely that this association involves a different HLA haplotype to that involved in flucloxacillin toxicity. Nevertheless, further investigation involving more detailed study of the HLA genes in relation to both drug toxicities would be of interest.

The second study concerned osteonecrosis of the jaw induced by bisphosphonates. Bisphosphonates are widely used in the treatment of cancer and osteoporosis to limit the loss of bone mass, but in some patients their use is associated with necrotic damage to bone tissue in the jaw, which is often triggered by dental problems[40]. In a small GWA study, four SNPs in the *CYP2C8* gene showed altered frequencies in the cases that narrowly escaped genome-wide significance[41]. This finding is of interest because *CYP2C8* encodes an enzyme that has an important role in drug metabolism[42] (athough probably not in bisphosphonate metabolism) and that might also have a role in inflammation[43], but follow-up studies with a larger cohort will be needed to prove any association. It is possible that the association relates to the more general biological role for this enzyme — which is expressed in a range of tissues in addition to the liver — in the metabolism of inflammatory mediators[43].

Future directions

As with studies on complex diseases, the number of GWA studies in the area of pharmacogenomics is increasing rapidly, with at least 12 studies appearing during 2009. Interesting findings have emerged from small cohorts, mainly due to large effects for some genes. Potentially, the findings from some of these studies on drug response could be used to develop simple genetic tests to determine the most suitable drug for treatment or the most appropriate dose to use. Subsequent, larger studies, ideally involving international collaborations, will be needed to widen our understanding of genetic factors that affect drug response and susceptibility to serious adverse drug reactions.

Glossary

Abacavir
Reverse transcriptase inhibitor used in the treatment of HIV.

Additive model
A genetic model with an increased disease risk of r for heterozygous genotypes and 2r for homozygous variant genotypes.

Cholestatic hepatitis
Obstruction of bile secretion and dysfunction of bile canaliculi together with inflammation of the liver. There are a number of possible causes, including exposure to certain drugs.

Drug disposition
(Also known as pharmacokinetics.) The process by which a drug is handled by the body following administration. It includes the processes of absorption, distribution, metabolism and excretion of a drug.

Drug metabolism
The biochemical modification or degradation of drugs through specialized enzymatic systems; examples include the cytochrome P450 oxidases and the uridine diphosphate glucuronosyltransferases. The rate of drug metabolism is an important determinant of the duration and intensity of the pharmacological action of drugs.

Drug response
The way an individual patient or cell responds to drug administration. It can be assessed in different ways, either qualitatively or quantitatively depending on the particular drug. The term pharmacodynamics can also be used to describe patient response.

Class I human leukocyte antigen genes
Genes within the human major histocompatibility complex that encode polymorphic membrane glycoproteins that are expressed on all nucleated cells. They form complexes with antigenic peptides and present them to cytotoxic T cells.

Linkage disequilibrium
The non-random association of alleles. For example, alleles of SNPs that reside near one another on a chromosome often occur in non-random combinations owing to infrequent recombination. Linkage disequilibrium is useful in genome-wide association studies as it reduces the number of SNPs that must be interrogated to determine genotypes across the genome. Conversely, strong linkage disequilibrium can complicate the identification of functional variants.

Multivariate regression analysis
Regression analysis includes any technique for modelling and analysing several variables when the focus is on the relationship between a dependent variable (y) and one or more independent variables (x). In multivariate regression analysis, two or more dependent variables are included in one analysis.

Neutropenia
An abnormally low number of neutrophils, the most important type of white blood cell, in the blood.

Odds ratio
A measurement of association that is commonly used in case–control studies. It is defined as the odds of exposure to the susceptible genetic variant in cases compared with the odds of exposure in controls. If the odds ratio is significantly greater than one, the genetic variant is associated with the disease.

Phase III clinical trials
Randomized, controlled, multicentre trials on large patient groups that aim to assess the effectiveness of a new drug under development compared with the current best treatment.

Replication cohort
In the context of a genome-wide association study, a cohort of cases and controls that is drawn from a population comparable to the original population and that is analysed to test the validity of the initial association.

Simvastatin
A particular statin drug. Statins are prescribed to decrease blood cholesterol level. They act by inhibiting the enzyme 3-hydroxy-3-methyl-glutaryl-CoA (HMGCoA) reductase.

Sinusoidal
Hepatocytes are polarized cells: one plasma membrane face points towards the sinusoids, or local blood vessels, in the liver and the other face points towards the bile canaliculi. Drugs and other compounds are normally transported into the hepatocyte from the general circulation across the sinusoidal membrane.

In the longer term, there is considerable interest in applying whole-genome sequencing to pharmacogenomics with a view to identifying rare sequence variants. This may involve sequencing of all coding sequences (exome sequencing)[44] or entire genomes using the new sequencing technologies, which are developing rapidly[45]. Whole-genome sequencing is likely to be of particular relevance to the study of adverse drug reactions, in which, due to a contribution from rare variants, GWA studies may not necessarily identify all risk factors, even if the number of cases can be increased. For example, in drug-induced cardiotoxicity, some evidence of an important role for rare variants already exists from limited sequencing studies on candidate genes[46]. Additional rare variants may also be important in other forms of idiosyncratic toxicity. Although genetic tests for rare variants may not be useful for making treatment decisions, knowledge of the underlying mechanisms of serious adverse drug reactions will inform the drug-development process and may enable better prediction of the potential toxicity of new drugs.

Ann K. Daly is at the Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, UK.
e-mail: a.k.daly@ncl.ac.uk

doi:10.1038/nrg2751

1. Meyer, U. A. Pharmacogenetics — five decades of therapeutic lessons from genetic diversity. *Nature Rev. Genet.* **5**, 669–676 (2004).
2. Daly, A. K. Pharmacogenomics of anticoagulants: steps toward personal dosage. *Genome Med.* **1**, 10 (2009).
3. Ge, D. *et al.* Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
4. Suppiah, V. *et al. IL28B* is associated with response to chronic hepatitis C interferon-α and ribavirin therapy. *Nature Genet.* **41**, 1100–1104 (2009).
5. Tanaka, Y. *et al.* Genome-wide association of *IL28B* with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. *Nature Genet.* **41**, 1105–1109 (2009).
6. Shuldiner, A. R. *et al.* Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* **302**, 849–857 (2009).
7. Cooper, G. M. *et al.* A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* **112**, 1022–1027 (2008).
8. Takeuchi, F. *et al.* A genome-wide association study confirms *VKORC1*, *CYP2C9*, and *CYP4F2* as principal genetic determinants of warfarin dose. *PLoS Genet.* **5**, e1000433 (2009).
9. Teichert, M. *et al.* A genome-wide association study of acenocoumarol maintenance dosage. *Hum. Mol. Genet.* **18**, 3758–3768 (2009).
10. Link, E. *et al. SLCO1B1* variants and statin-induced myopathy — a genomewide study. *N. Engl. J. Med.* **359**, 789–799 (2008).
11. Daly, A. K. *et al. HLA-B*5701* genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nature Genet.* **41**, 816–819 (2009).
12. Eichelbaum, M., Ingelman-Sundberg, M. & Evans, W. E. Pharmacogenomics and individualized drug therapy. *Annu. Rev. Med.* **57**, 119–137 (2006).
13. Kweekel, D., Guchelaar, H. J. & Gelderblom, H. Clinical and pharmacogenetic factors associated with irinotecan toxicity. *Cancer Treat. Rev.* **34**, 656–669 (2008).
14. Naisbitt, D. J. *et al.* Characterization of drug-specific T cells in lamotrigine hypersensitivity. *J. Allergy Clin. Immunol.* **111**, 1393–1403 (2003).
15. Mallal, S. *et al.* Association between presence of *HLA-B*5701*, *HLA-DR7*, and *HLA-DQ3* and hypersensitivity to HIV-1 reverse transcriptase inhibitor abacavir. *Lancet* **359**, 727–732 (2002).
16. Mallal, S. *et al. HLA-B*5701* screening for hypersensitivity to abacavir. *N. Engl. J. Med.* **358**, 568–579 (2008).
17. Seeringer, A. & Kirchheiner, J. Pharmacogenetics-guided dose modifications of antidepressants. *Clin. Lab. Med.* **28**, 619–626 (2008).
18. Kirchheiner, J. *et al. CYP2D6* and *CYP2C19* genotype-based dose recommendations for antidepressants: a first step towards subpopulation-specific dosages. *Acta Psychiatr. Scand.* **104**, 173–192 (2001).
19. Hulot, J. S. *et al.* Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. *Blood* **108**, 2244–2247 (2006).
20. Simon, T. *et al.* Genetic determinants of response to clopidogrel and cardiovascular events. *N. Engl. J. Med.* **360**, 363–375 (2009).
21. Collet, J. P. *et al.* Cytochrome P450 2C19 polymorphism in young patients treated with clopidogrel after myocardial infarction: a cohort study. *Lancet* **373**, 309–317 (2009).
22. Schroth, W. *et al.* Association between *CYP2D6* polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA* **302**, 1429–1436 (2009).
23. Hoskins, J. M., Carey, L. A. & McLeod, H. L. *CYP2D6* and tamoxifen: DNA matters in breast cancer. *Nature Rev. Cancer* **9**, 576–586 (2009).
24. Roden, D. M. *et al.* Pharmacogenomics: challenges and opportunities. *Ann. Intern. Med.* **145**, 749–757 (2006).
25. Liu, C. *et al.* Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Mol. Med.* **14**, 575–581 (2008).
26. Lavedan, C. *et al.* Association of the *NPAS3* gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. *Mol. Psychiatry* **14**, 804–819 (2009).
27. Caldwell, M. D. *et al. CYP4F2* genetic variant alters required warfarin dose. *Blood* **111**, 4106–4112 (2008).
28. Zhang, J. E. *et al.* Effects of *CYP4F2* genetic polymorphisms and haplotypes on clinical outcomes in patients initiated on warfarin therapy. *Pharmacogenet. Genomics* **19**, 781–789 (2009).
29. Borgiani, P. *et al. CYP4F2* genetic variant (rs2108622) significantly contributes to warfarin dosing variability in the Italian population. *Pharmacogenomics* **10**, 261–266 (2009).
30. Thijssen, H. H., Flinois, J.-P. & Beaune, P. H. Cytochrome P450 2C9 is the principal catalyst of racemic acenocoumarol hydroxylation reactions in human liver microsomes. *Drug Metab. Dispos.* **28**, 1284–1290 (2000).
31. de Morais, S. M. F. *et al.* The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. *J. Biol. Chem.* **269**, 15419–15422 (1994).
32. Iadonato, S. P. & Katze, M. G. Genomics: hepatitis C virus gets personal. *Nature* **461**, 357–358 (2009).
33. Becquemont, L. Pharmacogenomics of adverse drug reactions: practical applications and perspectives. *Pharmacogenomics* **10**, 961–969 (2009).
34. Wilke, R. A. *et al.* Identifying genetic risk factors for serious adverse drug reactions: current progress and challenges. *Nature Rev. Drug Discov.* **6**, 904–916 (2007).
35. Niemi, M. Transporter pharmacogenetics and statin toxicity. *Clin. Pharmacol. Ther.* **87**, 130–133 (2010).
36. Voora, D. *et al.* The *SLCO1B1*5* genetic variant is associated with statin-induced side effects. *J. Am. Coll. Cardiol.* **54**, 1609–1616 (2009).
37. Trevino, L. R. *et al.* Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J. Clin. Oncol.* **27**, 5972–5978 (2009).
38. Donaldson, P. T. *et al.* Flucloxacillin-induced liver injury: the extended MHC 57.1 haplotype as a major risk factor. *Hepatology* **48** (Suppl. 1), 396–397 (2008).
39. Kindmark, A. *et al.* Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J.* **8**, 186–195 (2008).
40. Khan, A. A. *et al.* Bisphosphonate associated osteonecrosis of the jaw. *J. Rheumatol.* **36**, 478–490 (2009).
41. Sarasquete, M. E. *et al.* Bisphosphonate-related osteonecrosis of the jaw is associated with polymorphisms of the cytochrome P450 *CYP2C8* in multiple myeloma: a genome-wide single nucleotide polymorphism analysis. *Blood* **112**, 2709–2712 (2008).
42. Totah, R. A. & Rettie, A. E. Cytochrome P450 2C8: substrates, inhibitors, pharmacogenetics, and clinical relevance. *Clin. Pharmacol. Ther.* **77**, 341–52 (2005).
43. Fleming, I. Cytochrome P450 epoxygenases as EDHF synthase(s). *Pharmacol. Res.* **49**, 525–533 (2004).
44. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl Acad. Sci. USA* **106**, 19096–19101 (2009).
45. Strausberg, R. L., Levy, S. & Rogers, Y. H. Emerging DNA sequencing technologies for human genomic medicine. *Drug Discov. Today* **13**, 569–577 (2008).
46. Paulussen, A. D. *et al.* Genetic variations of *KCNQ1*, *KCNH2*, *SCN5A*, *KCNE1*, and *KCNE2* in drug-induced long QT syndrome patients. *J. Mol. Med.* **82**, 182–188 (2004).
47. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Rev. Genet.* **9**, 356–369 (2008).
48. Byun, E. *et al.* Genome-wide pharmacogenomic analysis of the response to interferon β therapy in multiple sclerosis. *Arch. Neurol.* **65**, 337–344 (2008).
49. Turner, S. T. *et al.* Genomic association analysis suggests chromosome 12 locus influencing antihypertensive response to thiazide diuretic. *Hypertension* **52**, 359–365 (2008).
50. Mick, E., Neale, B., Middleton, F. A., McGough, J. J. & Faraone, S. V. Genome-wide association study of response to methylphenidate in 187 children with attention-deficit/hyperactivity disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **147B**, 1412–1418 (2008).
51. Yang, J. J. *et al.* Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA* **301**, 393–403 (2009).
52. Ising, M. *et al.* A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch. Gen. Psychiatry* **66**, 966–975 (2009).
53. Garriock, H. A. *et al.* A genomewide association study of citalopram response in major depressive disorder. *Biol. Psychiatry* **67**, 133–138 (2010).
54. Hartford, C. *et al.* Genome scan implicates adhesion biological pathways in secondary leukemia. *Leukemia* **21**, 2128–2136 (2007).
55. Inada, T. *et al.* Pathway-based association analysis of genome-wide screening data suggest that genes associated with the γ-aminobutyric acid receptor signaling pathway are involved in neuroleptic-induced, treatment-resistant tardive dyskinesia. *Pharmacogenet. Genomics* **18**, 317–323 (2008).
56. Volpi, S. *et al.* Whole genome association study identifies polymorphisms associated with QT prolongation during iloperidone treatment of schizophrenia. *Mol. Psychiatry* **14**, 1024–1031 (2009).

Competing interests statement
The author declares no competing financial interests.

DATABASES

Entrez Gene: http://www.ncbi.nlm.nih.gov/gene
CYP2C8 | *CYP2C9* | *CYP2C18* | *CYP2C19* | *CYP4F2* | *HCP5* | *HLA-B* | *IL28B* | *SLCO1B1* | *ST6GAL1* | *TPMT* | *UGT1A1* | *VKORC1*
UniProtKB: http://www.uniprot.org
CYP2D6

FURTHER INFORMATION

Ann K. Daly's homepage:
http://www.ncl.ac.uk/icm/people/profile/a.k.daly
***Nature Reviews Genetics* series on Genome-wide association studies:** http://www.nature.com/nrg/series/gwas/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF