TABLE 2

Listing of family and case-control studies ongoing at GSK as of September 2004

Disease	No. of sites	Family or case-control	No. of subjects collected	Target no. of subjects
Asthma	14	Family	5909	5909
Alzheimer's disease	9	Case-control	1504	2000
COPD	11	Case-control and family	5121	5460
Schizophrenia	4	Case-control	1317	1953
Metabolic syndrome	6	Case-control and family	4842	4842
Osteoarthritis	8	Case-control and family	4137	5685
Rheumatoid arthritis	1	Case-control	1736	2600
Parkinson's disease	1	Case-control and family	3039	3039
Unipolar depression	9	Case-control and family	3781	3861
Obesity	2	Case-control	2019	2000

be to use knockout and conditional knockin mice with genetic variants as additional target validation for genes that are statistically associated with the expression of human diseases.

Industry-initiated partnerships, such as The SNP Consortium, and efforts to sequence the human genome rapidly catalyzed SNP discovery [44]; this information was used to identify validated SNP assays for the ~1800 targets that constitute, for example, the GPCRs and kinases. For large-scale, disease-association analyses, the testing of these variants requires two sets of reagents: (i) validated SNP assays of each gene; and (ii) DNA from clinically well-phenotyped patients, appropriate controls and integration with extensive bioinformatic support systems. The standard candidate gene association experiment is to test variations of a specific gene for association to a single defined human disease. For HiTDIP, it is possible to extend this experiment to hundreds of genes (~1800), using thousands of SNP variations (~7000), and to several diseases (17+) in independent test and secondary screens.

The clinical data

The rate-limiting step in the HiTDIP program is the examination, acquisition of consent and collection of DNA from large collections of well-phenotyped patients and controls. Over the past decade, the key practical problems encountered are that the majority of patient collections of DNA reside in academic laboratories or small countries, and chemical-screening libraries are located in industry. Institutional review boards require that patient populations have specific informed consent for the commercial use of DNA, thus rendering many collections of patients enrolled in academic institutions largely unusable for commercial HTS. Patient and control clinical evaluations are not high throughput, because they are performed on an individual basis. Therefore, even if the technical screening capacity were generally available, as it is now, the essential clinical populations with available, consented DNA samples are not.

Disease phenotype does not simply depend on a diagnostic label

Over the past seven years, GSK has sponsored an extensive series of disease-specific clinical collaborations. These networks eventually involved ~200 specialist physician collaborators and currently comprise at least 17 diseases in several ethnic populations. Large groups of highly phenotyped patients and controls have been collected for these associations: ten of the case-control or case-control family studies currently completed (asthma and type 2 diabetes were the first in 2003 and 2004, respectively) or scheduled to be finalized are summarized in Table 2. Productivity gains have resulted in an increase in genotyping capacity, thus six of the listed diseases are scheduled to be analyzed in 2005, with other diseases to follow on completion of clinical enrollments. The lag phase for HiTDIP analyses is the process of registering subjects (patients and controls). After that, the process is 'industrialized' with planned overlapping primary, secondary and, in some cases, tertiary screening blocks.

Unlike more common retrospective patient collections, where information is gleaned from records, the network physicians were required to agree on diagnostic criteria in advance of subject collection. Data are often missing in retrospective collections, which can give rise to speculation on the implications of a 'blank' answer in the record of an individual. GSK created a clinical database that encapsulates the complete clinical and demographic information about each patient and each control prospectively: a prospective study ensures that data on all individuals are collected, and the database is as comprehensive as possible.

Each network of clinical experts established a core database of phenotypes with defined clinical descriptions for each disease, with the added caveat that additional supplementary clinical information that any participating clinical investigator wanted to collect would also be included in the phenotypic database. This resulted in disease-specific databases encompassing clinical parameters agreed by all network physicians, in addition to the sub-sets collected with respect to the particular research interests of an investigator.

This database format provides the opportunity to test genetic associations with more granularity than simply diagnosis. The data can be analyzed using the physicians' agreed disease diagnosis, single symptoms or signs or groups of clinical findings. There can be several contributing pathogenic processes in complex diseases, any of which might not be expressed concurrently to produce active disease, but each of which might be associated at some level with similarly diagnosed patient populations.

For example, in studying type 2 diabetes mellitus, inclusion of ophthalmologic examinations, renal function indicators and a defined neurological physical examination enables specific sub-type association studies. If the association of SNP variants in those patients with peripheral neuropathy was required, then performing and specifically recording the presence or absence of ankle jerks on all patients and controls is more accurate than trying to select patients with retrospective lack of data. Similarly, if diabetic retinopathy were an interest, few retrospective-controlled studies would have this information.

Illustrative study example: asthma

Asthma was the first disease to be analyzed in HiTDIP against multiple phenotypic variations. The association of thousands of SNP variants was measured against five separate but correlated asthma-related traits in large family sets including: (i) physicians' diagnosis of asthma; (ii) atopy (positive skin prick tests); (iii) atopic asthma (physicians' diagnosis plus atopy); (iv) strict asthma (two or more classical symptoms and a positive methacholine challenge test or bronchodilator test); and (v) bronchial hyper-responsiveness (positive methacholine response at or below 10 mg/ml of methacholine) [37]. The subjects were ascertained prospectively, but retrospectively grouped for each set of clinical criteria. Association studies using high-throughput genotyping of SNPs from tractable targets to define associations with various clinical definitions have been productive.

When thinking of asthma as a complex disease and/or syndrome, it is understandable that there might be multiple genetic and environmental susceptibilities. Certainly, there should be no expectation that all or most of the phenotypic variables would be present in all cases, as might be expected for a specific genetic mutation in a highly penetrant, rare inherited disease. A comprehensive approach that improves this situation would be to test each of the clinical forms that occur within families for genetic association with particular genetic variants. Confirmation of association could be performed in a subsequent set of families as well as in series of sporadic cases.

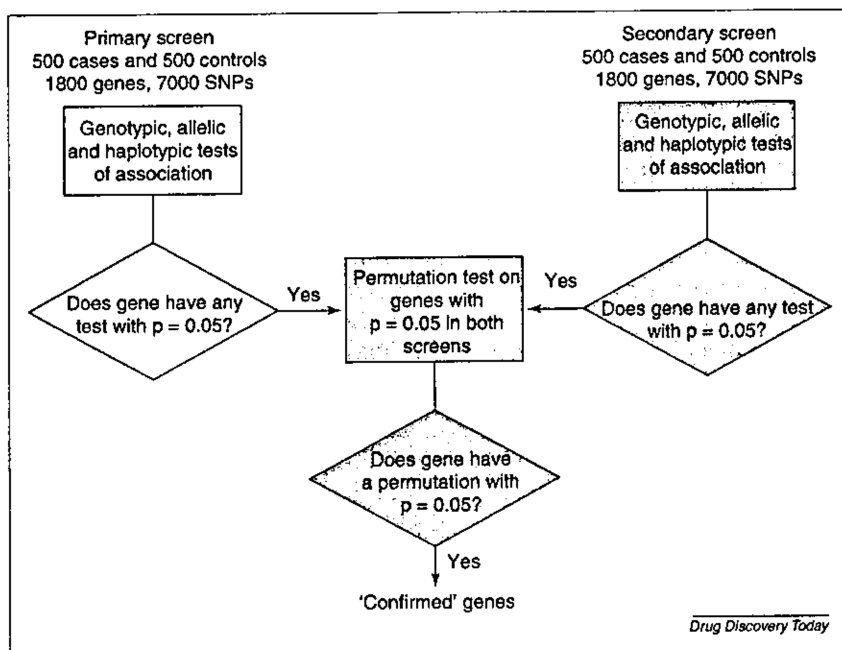
When the initial pilot asthma-screening was performed in 2001 with ~2700 SNPs from 1244 genes, interesting patterns of association were observed. Many of the associated gene variants were related to three or more of the selected clinical phenotype definitions. The association of some genes with physicians' diagnoses tended to separate to bronchial sensitivity-related signs and symptoms and atopy-related phenotypes. These early data regarding phenotypic criteria seemed to support the convergence of multiple susceptibility loci for the production of symptomatic complex disease.

In genetic linkage or association literature, there are frequently tacit assumptions, for example, that the expert physicians' diagnoses are sacrosanct, or that inclusion and/or exclusion criteria of any sort could often narrow the disease populations to be non-representative of all patients with the disease. Many of the conflicting reports of associations in the literature, where independent scientists have not 'confirmed' published linkage or association results with the same 'disease', could be the result of variations in the selected phenotypes or even the critical controls. With the same corps of physicians collecting the large prospective patient and control test series, as well as the subsequent confirmation series, variation between individual physician diagnostic skills can be minimized and the phenotype definitions can be stabilized.

Illustrative study example: metabolic syndrome

Metabolic syndrome is usually described as a combination of obesity, diabetes mellitus, hypertension and dyslipidemia – thus providing considerable opportunities for selection of who might be included in genetic studies. In 2002, The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults, Adult Treatment Panel III (ATP-III) published criteria f or a clinical diagnosis of the metabolic syndrome*. However, the GSK Genetics of Metabolic Syndrome (GEMS) patient collections for target association and susceptibility gene studies were initiated before this ATP-III report was created. By comparing the detailed phenotypic definitions for the collected patient information with those of the samples, and subsequently contrasting this with the definitions provided by a high level, independent expert panel, it was possible to analyze the relevance of our contemporary operational research definitions. In GEMS, two simple lipid-based criteria were used to define the disease-affected individuals: low high-density lipoprotein (HDL)-cholesterol and a concomitant elevation of plasma triglyceride concentrations. These criteria were selected because they are primary features of atherogenic dyslipidemia, are closely related to insulin resistance, detectable at an early stage in the development of metabolic syndrome, highly heritable and easy to measure. Wyszynski *et al.* [45] (clinical investigators supporting the GEMS Network) reported that 86% of individuals greater than 35 years of age met both the ATP-III and GEMS criteria. Conclusions based on genetic linkage, genetic association of susceptibility loci and PGx studies can thus be more accurately interpreted across studies. Accurate, interpretable, reproducible phenotypic definitions, even of complex syndromes mixing several complex diseases, can increase the value of clinical collections and facilitate analysis. Metabolic syn-

*NCEP criteria for clinical diagnosis of metabolic syndrome requires any three of the following: fasting plasma glucose of at least 110 mg/dl (6.10 mmol/l); serum triglycerides of at least 150 mg/dl (1.70 mmol/l); serum HDL cholesterol of less than 40 mg/dl (1.04 mmol/l) and 50 mg/dl (1.30 mmol/l) for males and females, respectively; and blood pressure of at least 130 mm Hg systolic and 85 mm Hg diastolic, or waist circumference (a measure of central adiposity) of more than 102 cm and 88 cm for males and females, respectively.



Experimental design algorithm for HiTDIP analyses. The basic design of the HiTDIP program was to screen approximately 500 patients and 500 controls, all prospectively examined, with full clinical information and commercial informed consent obtained. The current gene panel consists of approximately 1800 genes and 7000 validated SNP assays. Although highly significant p-values will be observed for some genes in the primary screen, the probability for occurrence of false positives is high. A secondary study of approximately 500 patients and 500 controls (obtained prospectively) is also tested. A gene with a p-value of ≤0.05 in primary and secondary screens is assessed by permutation (Box 1). Any gene with a permutation of p ≤0.05 is considered 'confirmed'. Calculations of random gene association assessed by permutation were confirmed (Figure 4).

drome is an excellent example of variable clinical definitions and hypotheses defining the disease.

Experimental data, analyses and statistical significance

During the past two years, up to ~1800 tractable genes represented by up to ~7000 SNPs were genotyped against several large patient and unrelated control groups for disease associations. The SNPs were identified predominantly from public databases, with focused SNP discovery performed as necessary. The SNPs selected were common, with 28% average minor allele frequency, and mapped typically to intergenic and promoter regions. Genotyping was performed using a multiplexed, bead-based approach published previously (Figure 2) [46,47]. A bioinformatic system, called SubjectLand, was designed and implemented to contain patient data, SNP (and other genetic variation) data, analyses programs and analytical results, with the ability to add-on data-mining capabilities. The standard experiment for each disease was designed to test ~500 well-phenotyped patients and ~500 matched controls in the initial, primary screen and follow-up with a secondary screen using an independent set of patients and controls ascertained by the same group of physicians.

The data were statistically analyzed using a gene-based approach (Figure 3). Allelic, genotypic and haplotypic tests of association were conducted in the primary and secondary screens. A 'fast fishers exact' test in SAS/BASICS® software was used for the allelic and genotypic tests [48], whereas the 'composite haplotype method' developed by Zaykin (unpublished results) was used for haplotypic tests. The use of a gene-based approach for analysis and replication, as used in HiTDIP, has recently been championed by Neale and Sham [49]. Cardon and Bell [50] stress that incorrectly adjusting for multiple testing can either unnecessarily reduce statistical power if too stringent a correction is applied or increase the false-positive rate if too weak a correction is used. As a result of the large number of tests conducted, adjustments for multiple testing were made using a gene-based permutation approach (Box 1).

The type 2 diabetes study made use of legacy (GlaxoWellcome, Burroughs-Wellcome, Glaxo, Smith Kline Beecham or Beecham) in-house collections – ~400 cases and controls were examined in the primary screen and >1100 cases and controls in the secondary screen. Among the 1405 genes examined in the primary screen, 256 genes had a p-value of ≤0.05. Of these 256 genes, 53 also had a p-value of ≤0.05 in the independent secondary screen but only 21 of the 53 genes were confirmed by passing the permutation process (Figure 4, Box 1). As well as conducting further investigations regarding the 21 confirmed genes that passed the permutation test and the 32 genes that did not, analyses of the probability that random genes would demonstrate similar statistical significance were also performed. Of the 21 permutation-confirmed genes, ten could be identified in pathways directly related to precedented mechanisms or metabolic pathways associated with disease-specific hypotheses. In the case of type 2 diabetes mellitus, four of the 21 genes had already been chemically screened in legacy companies and provided several leads. Statistical analyses of each disease will be submitted for formal peer-review in specialty journals by the appropriate network physicians.

In some diseases, such as type 2 diabetes mellitus, there are reasonable prior hypothesis concerning pathogenesis. For example, it is generally agreed that glucose metabolism and insulin sensitivity play a role. Among the HiTDIP type 2 diabetes confirmed genes, several are supported by prior published hypotheses and appear in expanded metabolic pathways. Such coincidences resulting from

The permutation testing process

The objective of this study was to identify tractable genes associated with disease. Some small genes might have only one or two SNPs analyzed, whereas 30–40 SNPs could be assessed in the case of some of the larger genes, such as those encoding ion channels. The greater the number of SNPs and tests performed on a gene, the greater the probability that the gene will appear significant than by chance alone. To account and correct for the variable number of tests conducted across genes, a gene-based permutation test was applied. Permutation testing is a standard method used to assess significance in the statistical analysis of genetic data [57]. Any gene significant at a p-value of ≤0.05 in the primary and secondary screens was further assessed by performing this permutation process on the data from the secondary screen. For each permutation, affection status was shuffled among the cases and controls. The genetic data for all SNPs across each subject was not altered. This maintains the underlying correlation between SNPs within a gene. All the SNPs within the gene were analyzed using allelic, genotypic and haplotypic association tests via the same methods used for the observed data. The smallest p-value across all tests was recorded for each permutation. The permutations were repeated up to 5000 times per gene. The empirical p-value is the proportion of minimal p-values from the permutations that are less than the observed minimal p-value from the actual data. The empirical p-value is estimated using Equation i.

$$\frac{r+1}{n+1} \qquad \text{[Eqn i]}$$

where r is the number of permutation p-values as small as or smaller than in the actual data and n is the number of permutations [58,59].

A gene with an empirical p-value of ≤0.05 was considered to be 'confirmed' with respect to statistical association with disease. For example, if the smallest p-value of gene A in the observed data was 0.004 and among 5000 permutation (n = 5000) a p-value of ≤0.004 was observed 50 times (r = 50), then the permutation process would generate an empirical p-value for gene A of 0.01, and it would be classed as a 'confirmed' gene.

For the type 2 diabetes study assessed by permutation, the minimal p-value from the actual secondary dataset for each of the 53 genes analyzed and their empirical p-values from the permutation process are shown in Figure i (observed and gene-based permutation results). Of the 53 genes assessed, 21 genes had empirical p-values of ≤0.05 and thus were considered 'confirmed'. The observed and empirical p-values for gene 7, a confirmed gene, were highly similar, differing by only ~0.0001. For gene 27, which was not confirmed, the corresponding values showed striking differences: the permutated p-value was approximately 0.33, whereas the observed p-value was <0.02.

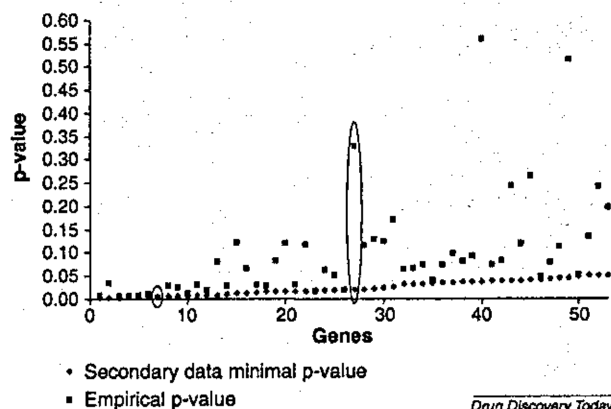


FIGURE I

Gene-based results for 53 type 2 diabetes mellitus genes assessed by permutation. The observed and gene-based permutation results for the 53 genes assessed by permutation in the type 2 diabetes data are rank ordered according to their observed p-values before permutation. Two of the 53 genes (genes 7 and 27) are circled to illustrate how the observed and empirical p-value can differ. It is useful to note that the first 12 genes were confirmed by permutation; it was expected – and quite reassuring – that the genes with consistently the lowest p-values would be confirmed. Permutation assessments confirmed four of the subsequent ten genes. Of the remaining 31 genes, five were confirmed. Hirschhorn *et al.* [18] reported that, by meta-analyses, only six of 166 genes replicated consistently across studies – again not unexpected in single-arm experiments, which are many studies with much smaller numbers of cases and controls and no opportunity to assess permutation (Box 2).

The gene-based permutation process controls for multiple correlated SNPs and multiple tests performed for each SNP in a gene in the secondary screen. A related question is – how many of the 1405 genes examined in the type 2 diabetes study would be expected to confirm under the null hypothesis of no association? That is, how many genes would have a minimal p-value of <0.05 in the primary screen and a permutation p-value of <0.05 in the secondary screen? To address the question of the permutation p-value, an additional round of 1000 permutations was conducted using the type 2 diabetes primary screen data. At each permutation, the affection status of the cases and controls was shuffled and the number of genes observed to have at least one p-value of ≤0.05 was recorded and multiplied by 0.05, the α level already enforced by the gene-based permutation applied to the secondary dataset. Of 1000 permutations performed on 1405 genes, the average number of genes that was confirmed by chance was 9.8 (95% confidence interval 8.6–11.0).

screening so many hypothesis-independent genes are, at the very least, interesting. This provides increased priority for screening specific targets in defined pathways and, subsequently, for designing relevant clinical trials. In diseases where there is a paucity of information to form the basis of supposition, for example, schizophrenia, specific genes and pathways identified in HiTDIP can provide insights to new hypotheses. Results for four other secondary screens (schizophrenia, obesity, migraine and unipolar depression) will be available in early 2005.

Additional experimental data for the association of each gene can also be generated by testing whether the target gene is within a region of extended linkage disequilibrium (LD) [51–53]. This is also important because the particular SNP used in the HiTDIP analysis might not be the variant responsible for the disease association, but has simply evolved concomitantly with the disease-variant. SNPs that are in LD with the causal variant can provide positive association data, as is the case for susceptibility genes for Alzheimer's disease, migraine, Crohn's disease and psoriasis [52–55]. Analyzing for regions of extended LD is also important to define whether the association signal is driven by the tested HiTDIP gene – and not by one of its neighboring genes [51]: documenting extended

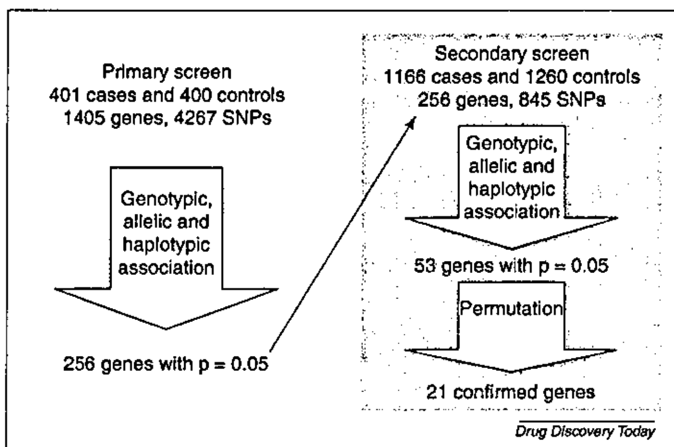


FIGURE 4

Type 2 diabetes mellitus HiTDIP results. The type 2 diabetes mellitus HiTDIP experiment was the first to be performed with two case-control screens. The prior asthma screens used a family-based series, thus the statistical analyses were significantly different. The primary screen consisted of 401 cases and 400 controls and used 4267 validated SNP assay for 1405 genes. A set of 256 genes yielded a p-value of ≤0.05 (a large number illustrating the high false-positive rate that is probable with a single screen). The secondary screen enrolled 1166 patients and 1260 controls. At the time that the secondary screen was initiated, the genotyping capacity was limited and thus only the 256 genes and their 845 SNPs were included. A set of 53 genes yielded a p-value of ≤0.05, and were assessed by permutation, with 21 genes being confirmed (Figure i). Estimation of the number of genes being confirmed by chance in this study yielded a rate of approximately ten genes per 1400 tested (Box 1).

LD around a HiTDIP gene can increase (or decrease) confidence in its validity.

It is reasonable to expect that a large screen with many tested variables could be subject to false positives. Although the same limitations also characterize susceptibility genes resulting from linkage and association studies, they are not usually tractable targets for chemical screening. Susceptibility gene studies provide new insights that are relevant to disease pathogenesis: we anticipate that the HiTDIP genes will also generate new theories and provide support for existing hypotheses. One immediate strategy could be to provide validation support by concentrating thorough phenotypic studies on knockout or otherwise manipulated mice [16]. For the type 2 diabetes example, this would mean knockouts and conditional knockins for ~21 genes and their specific variants associated with disease. Each mouse line could be examined with highly focused and complete biochemical and physiological phenotyping in parallel with high-throughput chemical screening.

There is no documented method for predicting which HiTDIP genes will point to lead candidates for medicines. The selection of genes with which to initiate high-throughput assay development is empirically based on several criteria, one of which is the relative ease of creating the chemical screening assay. A few confirmed genes for type 2 diabetes mellitus were screened previously in GSK legacy companies because of pre-existing literature rationales for potential involvement in disease pathogenesis. Enlarging or repeating the chemical screens with the current GSK compound libraries is an immediate option, as is re-evaluating earlier lead programs. In addition, there are several genes that the literature indicates are involved in interesting, potentially novel mechanisms. There are other genes, particularly those for which the design of high-throughput assays is relatively uncomplicated, that can be prioritized. In the GSK R&D structure, scientists in the Centers for Excellence in Drug Discovery (CEDDs), who are most familiar with the disease, prioritize the targets.

It is particularly pertinent to mention here that GSK is now dealing with a relatively large, staggered flow of new target genes for which there is excellent support for genetic association. If the leads from screened HiTDIP genes reduce attrition, this will be evidenced in a larger proportion of positive Phase IIA efficacy studies over the next 3–5 years. Remember, hits and leads must be optimized and undergo preclinical regulatory testing before human testing can begin.

HiTDIP provides genetic validation for target selection to drive the pipeline. The confirmed targets are genetically associated in some manner with specific human diseases or therapeutic indications that can be encompassed by the disease diagnostic label. Indeed, genes that are associated with specific clinical variables could help to define complex disease heterogeneity. A significant difference in success compared with historical target selection can be tested against benchmarks. Attrition that is the result of a lack of relevance to a particular human disease might only be appreciated when there is an absence of clinical efficacy many years down the pipeline. Reducing attrition at the proof of efficacy stage (Phase IIA) will increase the efficiency of the pipeline and could provide a sustainable stream of effective medicines related to human diseases. Furthermore, additional hypotheses might be generated at Phase IIA by the application of efficacy pharmacogenetics. These hypotheses can be tested reiteratively in Phases IIB, III and IV. Of course, the consented DNA samples remain available for testing new target classes as new chemical synthesis capabilities are developed.

HiTDIP to discovery shunt – some matches are made in heaven

Trying to decide what to wear for a particular occasion frequently involves a quick scan of your collection of clothes. Large pharmaceutical companies, particularly those with legacy compound libraries resulting from long histories of drug discovery, have a lot of unworn items in their collections. Many discovery programs are initiated and numerous targets are screened only for leads to be 'left on the shelf' (some with the price tags still on!) when times and circumstances changed. The initial assumption for HiTDIP was that by screening all the known tractable targets against several well-defined diseases, new targets would be identified. Each would then require the design of a high-throughput screen and a new screen of the chemical compound library.

However, once targets with highly statistically significant associations with a particular disease were defined, another surprisingly rich supply of existing leads, which required no novel chemical screening, was identified immediately. Indeed, several targets were already screened against the legacy company chemical libraries, with resulting leads left on the shelf because the program was dropped, or abandoned after a molecule failed for a particular therapeutic indication. Thus, immediately after completing confirmation analyses for the initial HiTDIP programs, several targets were identified for which there were lead – or better – quality molecules on the shelf. At present, several molecules have either entered preclinical testing for Phase I or have already been tested in clinical trials designed for different therapeutic indications.

The match between the molecule and disease is the key data gained from confirmed genetic association. Using molecules for which considerable data and work already exists will no doubt accelerate pipeline dynamics. More shots on goal quickly – or better shots at more defined goals – results from being able to mine data generated from years of previous drug discovery programs. Occasionally, after considerable prior hard work, serendipity plays a part – which is indeed food for thought for those who wonder what the advantages of corporate size could be! Although legacy company names might disappear, contributions to capabilities, such as prior screens of targets, are maintained.

BOX 2

Learning of the power and statistical significance of association studies

Hirschhorn *et al.* [18] conducted an interesting meta-analysis of 166 initial associations to determine the probability of their being reliably replicated. A particularly notable finding was that of 166 initial associations from multiple studies, only six replicated consistently across investigations (i.e. had p-values <0.05 in 75% or more of the studies identified), whereas 97 were observed to have at least one significant replication. Replication rates of 16–30%, depending on the definition of replication, have been identified in the association literature [20,21]. Replication of an association study has not usually been included in the experimental design, nor have the numbers of patients and controls been sufficient to expect replication.

The initial gene-based HiTDIP scans of type 2 diabetes mellitus, using a large cohort of patients and controls, identified five genes with a statistical significance of $p \leq 0.0005$, 40 genes with $p \leq 0.0050$ and 211 with $p \leq 0.0500$ (Table i). Although these data might comprise false negatives, or undetected genes, the vital question is how to eliminate false positives. Because of the large number of tests performed, the majority of the signals identified from the initial, primary test-screen, across all levels of significance, were expected to be false positives – it is not reasonable to anticipate that the testing of less than 2000 genes will result in 45 genes with a p-value of <0.005 being real. It should be emphasized that most published association studies are performed with much smaller patient and control groups. Although meta-analyses of small and heterogeneic studies might have been the best available secondary analysis method in the past, it was never considered ideal.

The factor that distinguishes the HiTDIP study design is the availability of an equally powerful set of patients and controls, all of whom have been examined and clinically characterized by the same physicians, thereby reducing phenotypic heterogeneity. Thus, the question can be posed – how many of the genes identified in the initial association studies with highly significant p-values are actually confirmed by a second confirmation study using statistical analyses that are appropriate for complex genome-based association studies? The answer provides a fascinating insight into the poor track record of association study confirmations. A significant proportion of the genes from the large initial association study with a p-value of <0.005 were not confirmed after the secondary screen of similar size (Table i). However, it is extremely important to realize that the published association literature undergoing meta-analyses traditionally use data that are typically much less significant – and the only confirmation comes from other small studies or takes the form of the meta-analyses.

TABLE i

Comparison of numbers of confirmed genes from primary and secondary screens based on p-values (before the permutation testing process)

Primary screen		Secondary screen[a]	Permutation[a]
p-value	No. of genes	No. of genes	No. of genes
≤0.0005	5	1	0
≤0.0050	40	8	3
≤0.0500	211	44	18
Total	256	53	21

[a]Performed at a p-value of 0.05.

Whole-genome testing and pathway analyses

Although nothing in science is unanimous, there is a growing belief that high-density, whole-genome SNP-screening, using newer statistical methods and powerful computing capacities, can identify susceptibility genes for a disease. From a practical viewpoint, and the experience of studying almost 2000 genes selected solely because they could be drug targets, it would seem reasonable to perform additional genome-wide (all genes) SNP-screening association studies. Family linkage studies of the past two decades have defined large regions of chromosomes (1–10 Mb) using 'log of the odds' scores that successfully identified disease and susceptibility genes. By narrowing these large regions using candidate genes, many positive results were reported. Much smaller regions of extended LD (50–250 kb) containing disease susceptibility genes were identified using large patient and control collections for case-control association studies. The DNA is there, the methods are increasingly more feasible economically and there is a growing literature of statistical methods to support these large studies [56]. From an academic point of view, whole-genome screening studies might be too expensive to be readily available, but that does not mean they will not provide scientifically valid data and disease insight. Therefore, after the timetable of HiTDIP confirmation screening is completed, whole-genome SNP-screening will follow. Making the right choice of targets at the beginning of the pipeline will be the first step down the long road of creating innovative medicines.

Within three years

The first practical metric that will provide an estimate of the success of HiTDIP will be comparing Phase II attrition

with historical benchmarks. If the target genes selected result in more frequent early-phase efficacy successes, then the 'quantal step-up in discovery' described by Weisberg will be apparent [32]. There will be other parallel strategies for selecting targets. However, comparisons of the success and attrition rates of drug candidates resulting from screens of genetically associated targets can be compared with benchmarks for other historical or current strategies. A retrospective view will be of interest but, for the present, the first indications will come from decreased attrition at Phase II. Along the way, there will be a stream of lead molecules with which to prime a high-throughput pipeline. Combining the choice of the right target and the application of efficacy pharmacogenetics at proof of concept, when necessary, should result in better defined medicines for patients with complex diseases [3]. The objective of the process is to reduce attrition, cycle times and expense, as well as provide safe and effective medicines.

Acknowledgements

This research was supported by GSK R&D. We wish to acknowledge the work of hundreds of GSK scientists, almost two hundred collaborating clinicians, many academic consultants and others who have contributed to the HiTDIP programs. As each set of disease-specific results is published, it is anticipated that many of the contributors will be identified and will be co-authors. Alison Gatherum and Mindy McPeak were extremely helpful in preparing the manuscript. Several important visionaries must also be acknowledged for a program that has taken seven years to begin to deliver sustained value to the pipeline: James Niedel, who was the Chairman of R&D in GlaxoWellcome and supported the initiation of the clinical networks (1997–2000); Tachi Yamada, Chairman of R&D in GSK, who backed and expanded this strategy as a core objective of the R&D organization of GSK (2000–present); Jean-Paul Garnier, CEO of GSK, whose support for the early, middle and late parts of the R&D pipeline was crucial in sustaining these efforts; and the thousands of people in the GSK organization who believe in 'Focus on the Patient' to create safe and effective medicines.

References

1 Chapman, T. (2004) Drug discovery – the leading edge. *Nature* 430, 109
2 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–715
3 Roses, A. (2002) Genome-based pharmacogenetics and the pharmaceutical industry. *Nat. Rev. Drug Discov.* 1, 541–549
4 Aono, S. *et al.* (1995) Analysis of genes for bilirubin UDP-glucuronosyltransferase in Gilbert's syndrome. *Lancet* 345, 958–959
5 Fijal, B.A. *et al.* (2000) Clinical trials in the genomic era: effects of protective genotypes on sample size and duration of trial. *Control. Clin. Trials* 21, 7–20
6 Monaghan, G. *et al.* (1996) Genetic variation in bilirubin UPD-glucuronosyltransferase gene promoter and Gilbert's syndrome. *Lancet* 347, 578–581
7 Chanda, S. and Caldwell, J. (2003) Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov. Today* 8, 168–174
8 Glassman, R. and Sun, A. (2004) Biotechnology: identifying advances from the hype. *Nat. Rev. Drug Discov.* 3, 177–183
9 Johnston, P.A. and Johnston, P.A. (2002) Cellular platforms for HTS: three case studies. *Drug Discov. Today* 7, 353–363
10 Austin, C.P. *et al.* (2004) The knockout mouse project. *Nat. Genet.* 36, 921–924
11 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
12 Marecki, S. and Kirkpatrick, P. (2004) Efalizumab. *Nat. Rev. Drug Discov.* 3, 473–474
13 Food and Drug Administration (2004) Innovation stagnation, challenge and opportunity on the critical path to new medical products (http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html)
14 Berns, A. (2001) Cancer – Improved mouse models. *Nature* 410, 1043–1044
15 Liggett, S.B. (2004) Opinion: genetically modified mouse models for pharmacogenomic research. *Nat. Rev. Genet.* 5, 657–663
16 Zambrowicz, B.P. *et al.* (2003) Predicting drug efficacy: knockouts model pipeline drugs of the pharmaceutical industry. *Curr. Opin. Pharmacol.* 3, 563–570
17 Zambrowicz, B.P. and Sands, A. (2003) Knockouts model the 100 best-selling drugs – will they model the next 100? *Nat. Rev. Drug Discov.* 2, 38–51
18 Hirschhorn, J. *et al.* (2002) A comprehensive review of genetic association studies. *Genet. Med.* 4, 45–61
19 Neale, B. and Sham, P. (2004) The future of association studies: gene-based analysis and replication. *Am. J. Hum. Genet.* 75, 353–362
20 Lohmueller, K. *et al.* (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* 33, 177–182
21 Ioannidis, J. *et al.* (2003) Genetic associations in large versus small studies: an empirical assessment. *Lancet* 361, 567–571
22 Tabor, H.K. *et al.* (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet.* 3, 391–397
23 Thomas, D. and Clayton, D. (2004) Betting odds and genetic associations. *J. Natl. Cancer Inst.* 96, 421–423
24 Weiss, K. and Terwilliger, J. (2000) How many diseases does it take to map a gene with SNPs. *Nat. Genet.* 26, 151–157
25 Risch, N. (2000) Searching for genetic determinants in the new millennium. *Nature* 405, 847–856
26 Colhoun, H.M. *et al.* (2003) Problems of reporting genetic associations with complex outcomes. *Lancet* 361, 865–872
27 Tufts Center for the Study of Drug Development (2003) Post-approval R&D raises total drug development costs to $897 million. *Impact Report: Analysis and Insight into Critical Drug Development Issues* 5, May–June
28 Venter, J.C. (2000) Genomic impact on pharmaceutical development. *Novartis Foundation Symposium* 229, 14–18
29 Collins, F. and McKusick, V. (2001) Implications of the Human Genome Project for medical science. *J. Am. Med. Assoc.* 285, 540–544
30 Lander, E. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
31 Venter, J. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304
32 Fishman, M.C. (2004) An audience with. *Nat. Rev. Drug Discov.* 3, 292
33 Lindsay, M. (2003) Innovation – target discovery. *Nat. Rev. Drug Discov.* 2, 831–838
34 Goldstein, D. *et al.* (2003) Pharmacogenetics goes genomic. *Nat. Rev. Genet.* 4, 937–947
35 Altshuler, D. *et al.* (2000) The common PPAR gamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* 26, 76–80
36 Martin, E. *et al.* (2001) Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. *J. Am. Med. Assoc.* 286, 2245–2250
37 Van Eerdewegh, P. *et al.* (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 418, 426–430
38 Funke, B. *et al.* (2004) Association of the DTNBP1 Locus with Schizophrenia in a U.S. Population. *Am. J. Hum. Genet.* 75, 891–898
39 Duan, J. *et al.* (2004) Polymorphisms in the trace amine receptor 4 (TRAR4) gene on chromosome 6q23.2 are associated with susceptibility to schizophrenia. *Am. J. Hum. Genet.* 75, 624–638
40 Giallourakis, C. *et al.* (2003) IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease

and identification of a novel association with ulcerative colitis. *Am. J. Hum. Genet.* 73, 205–211

41 Schmith, V. *et al.* (2003) Pharmacogenetics and disease genetics of complex diseases. *Cell. Mol. Life Sci.* 60, 1636–1646

42 Drews, J. (1996) Genomic sciences and the medicine of tomorrow. *Nat. Biotechnol.* 14, 1516–1518

43 Hopkins, A.L. and Groom, C. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730

44 Sachidanandam, R. *et al.* (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933

45 Wyszynski, D.F. *et al.* The relationship between atherogenic dyslipidemia and the adult treatment program-III definition of metabolic syndrome: The Genetic Epidemiology of Metabolic Syndrome Project. *Am. J. Cardiol.* (in press)

46 Taylor, J.D. *et al.* (2001) Flow cytometric platform for high-throughput single nucleotide polymorphism analysis. *Biotechniques* 30, 661–669

47 Chen, J. *et al.* (2000) A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res.* 10, 549–557

48 Mehta, C. and Patel, N. (1983) A network algorithm for performing Fisher's Exact Test in r×c contingency tables. *J. Am. Stat. Assoc.* 78, 427–434

49 Neale, B. and Sham, P. (2004) The future of association studies: gene-based analysis and replication. *Am. J. Hum. Genet.* 75, 353–362

50 Cardon, L. and Bell, J. (2001) Association study designs for complex diseases. *Nat. Rev. Genet.* 2, 91–99

51 Roses, A. (2002) SNPs – where's the beef? *Pharmacogenomics J.* 2, 277–283

52 McCarthy, L. *et al.* (2001) Single-nucleotide polymorphism alleles in the insulin receptor gene are associated with typical migraine. *Genomics* 78, 135–149

53 Hewett, D. *et al.* (2002) Identification of a psoriasis susceptibility candidate gene by linkage disequilibrium mapping with a localized single nucleotide polymorphism map. *Genomics* 79, 305–314

54 Martin, E. *et al.* (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am. J. Hum. Genet.* 67, 146–154

55 Rioux, J. *et al.* (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* 29, 223–228

56 Carlson, C. *et al.* (2004) Mapping complex disease loci in whole-genome association studies. *Nature* 429, 446–452

57 North, B.V. *et al.* (2002) A note on the calculation of empirical P values from Monte Carlo procedures. *Am. J. Hum. Genet.* 71, 439–441

58 Davison, A.C. and Hinkley, D.V. (1997) The Basic Bootstraps. In *Bootstrap Methods and their Application* (Gill, R. *et al.*, eds), pp. 11–66, Cambridge University Press

59 Davison, A.C. and Hinkley, D.V. (1997) Tests. In *Bootstrap Methods and their Application* (Gill, R. *et al.*, eds), pp. 136–187, Cambridge University Press

60 Roses, A. (2004) Pharmacogenetics and drug development: the path to safer and more effective drugs. *Nat. Rev. Genet.* 5, 645–655

Executive Summary

The remarkable development and application of new genetic technologies over the past 25 years has been accompanied by profound changes in the way in which research is commercialised in the life sciences. Many thousands of patents which assert rights over DNA sequences have been granted to researchers across the public and private sector. In general, we acknowledge the benefits that have accrued to society from the patent system, but we ask whether the application of the patent system to DNA sequences is achieving its goals, namely the stimulation of innovation for the public good, and the rewarding of people for useful new inventions.

We note that many patents that assert rights over DNA sequences have already been granted but are of doubtful validity. The effects of many of these patents are extensive, because inventors who assert rights over DNA sequences obtain protection on all uses of the sequences.

We conclude that in the future, the granting of patents that assert rights over DNA sequences should become the exception rather than the norm. The patent system currently regards DNA sequences as eligible for patenting. However, as computational techniques replace cloning as the main route to identifying genes, we consider that the issue of the eligibility for patenting of DNA sequences needs to be reopened. Even if DNA sequences are considered eligible for patenting, they must then satisfy the criteria of being novel, inventive and useful. We consider that the application of these criteria to DNA sequences has not been sufficiently stringent. We note, further, that the fact that DNA sequences are essentially just genetic information distinguishes them from other chemical compounds, with regard to the patent system.

We distinguish four different uses to which DNA sequences can be put: in diagnostic tests based on genes, as research tools, in gene therapy and for the production of therapeutic proteins. We conclude that patents that assert rights over DNA sequences and their uses are, in some cases, supportable, but in others, should be treated with great caution.

Diagnostic tests

- **We recommend that the criteria already in place within existing patent systems for the granting of patents, particularly the criterion of inventiveness, be stringently applied to applications for product patents which assert, inter alia, rights over DNA sequences for use in diagnosis. We recommend that the European Patent Office (EPO), the United States Patent and Trademark Office (USPTO) and the Japan Patent Office (JPO) together examine ways in which this may be achieved.** If this recommendation is implemented, we expect that the granting of product patents which assert rights over DNA sequences for use in diagnosis will become the rare exception, rather than the norm.
- We consider the grant of use patents for diagnostic tests and **conclude that the protection by use patents of specific diagnostic tests which are based on DNA sequences could provide an effective means of rewarding the inventor while providing an incentive for others to develop alternative tests.**
- We consider that, in the case of patents that have been granted for diagnostic tests based on genes, compulsory licensing may be required to ensure reasonable licensing terms are available to enable alternative tests to be developed.

Research tools

- **We consider, that in general, the granting of patents which assert rights over DNA sequences as research tools should be discouraged.** We have taken the view that that the best way to discourage the award of such patents is by stringent application of the criteria for patenting, particularly utility.

Gene therapy

- **We consider that once a gene which is associated with a disease is identified, the use of the relevant DNA sequences in gene replacement therapy, to alleviate the effects of mutations in that gene, is obvious (particularly when such use is claimed on a purely speculative basis). Therefore, we recommend that protection by product patents should seldom be permissible.** We believe that patent protection should be concentrated on developing safe and effective methods of appropriate gene delivery. This is where the real inventiveness and investment will be required, rather than in simply defining the sequence of the genes to be used in treatment.

Therapeutic proteins

- **We take the view that while rights asserted over DNA sequences which are used to make new medicines that are therapeutic proteins are generally acceptable, they should be narrowly defined. By this we mean that the rights to the DNA sequence should extend only to the protein described.**

We consider that the adoption of the recommendations that we put forward here will serve to guide patent offices and the courts to a more rational use of the system which reserves patent protection for those patents that assert rights over DNA sequences that reflect a significant contribution by the researcher.

PHARMACOGENETICS AND DRUG DEVELOPMENT: THE PATH TO SAFER AND MORE EFFECTIVE DRUGS

Allen D. Roses

Abstract | Pharmacogenetics provides opportunities for informed decision-making along the pharmaceutical pipeline. There is a growing literature of retrospective studies of marketed medicines that describe efficacy or safety on the basis of patient genotypes. These studies emphasize the potential prospective use of genome information to enhance success in finding new medicines. An example of a prospective efficacy pharmacogenetic Phase-IIA proof-of-concept study is described. Inserting a rapidly performed efficacy pharmacogenetic step after initial clinical data are obtained can provide confidence for a commitment to full drug development. The rapid identification of adverse events during and after drug development using genomic mapping tools is also reviewed.

As a consequence of the sequencing and mapping of the human genome, pharmacogenetics (PGx) — the use of genetic analysis to predict drug response, efficacy and toxicity — is becoming the first drug-discovery pipeline technology to affect the structure and economics of the pharmaceutical industry. In particular, high-throughput SNP mapping is allowing human disease-associated drug targets to be identified and is reducing attrition in early-phase clinical trials[1–3]. The recent history and background of PGx has been reviewed elsewhere[4], so here I focus on current experimental applications of PGx technologies. The application of genome-wide mapping and screening methods to clinical medicine is accelerating[5–8]. However, there have been few published examples of genome-wide applications of SNP mapping to drug development or to the drug safety of licensed medicines. Nonetheless, many proof-of-concept projects are progressing within the pharmaceutical industry, which has direct access to clinical trial patients and the capabilities to apply genome-wide technologies. What was predicted a few years ago is applicable now[8–10].

Understanding the structure of the pharmaceutical pipeline (BOX 1), including the decisions for selecting drug candidates and progressing drug development, is not a core capability of many academic scientists. Although segments of the process for producing a marketed drug might be pursued in academic laboratories and biotechnology companies, the entire pharmaceutical pipeline — from target selection to full development and launch — is fully represented only in the pharmaceutical industry. Companies that take drugs through the pipeline range from the very large (for example, Pfizer and GlaxoSmithKline) to much smaller, younger companies (for example, Amgen and Elan). Much of this industry-based research never reaches the academic literature because, for the most part, it is about the repeated serial failures of scientific projects. It might take many years after launching new chemical entities with proven therapeutic advantages to recognize the rare successes.

It was hoped that the Human Genome Project would provide a flood of new targets for drug discovery[11–14]. However, at that time, most scientists and business commentators did not realize the difficulties that are inherent in moving from a genome sequence to the identification of disease-relevant targets for drug discovery. In general, most genes are not drug targets that can be chemically screened with current technologies. Disease-related targets for drug discovery are usually chosen on the basis of

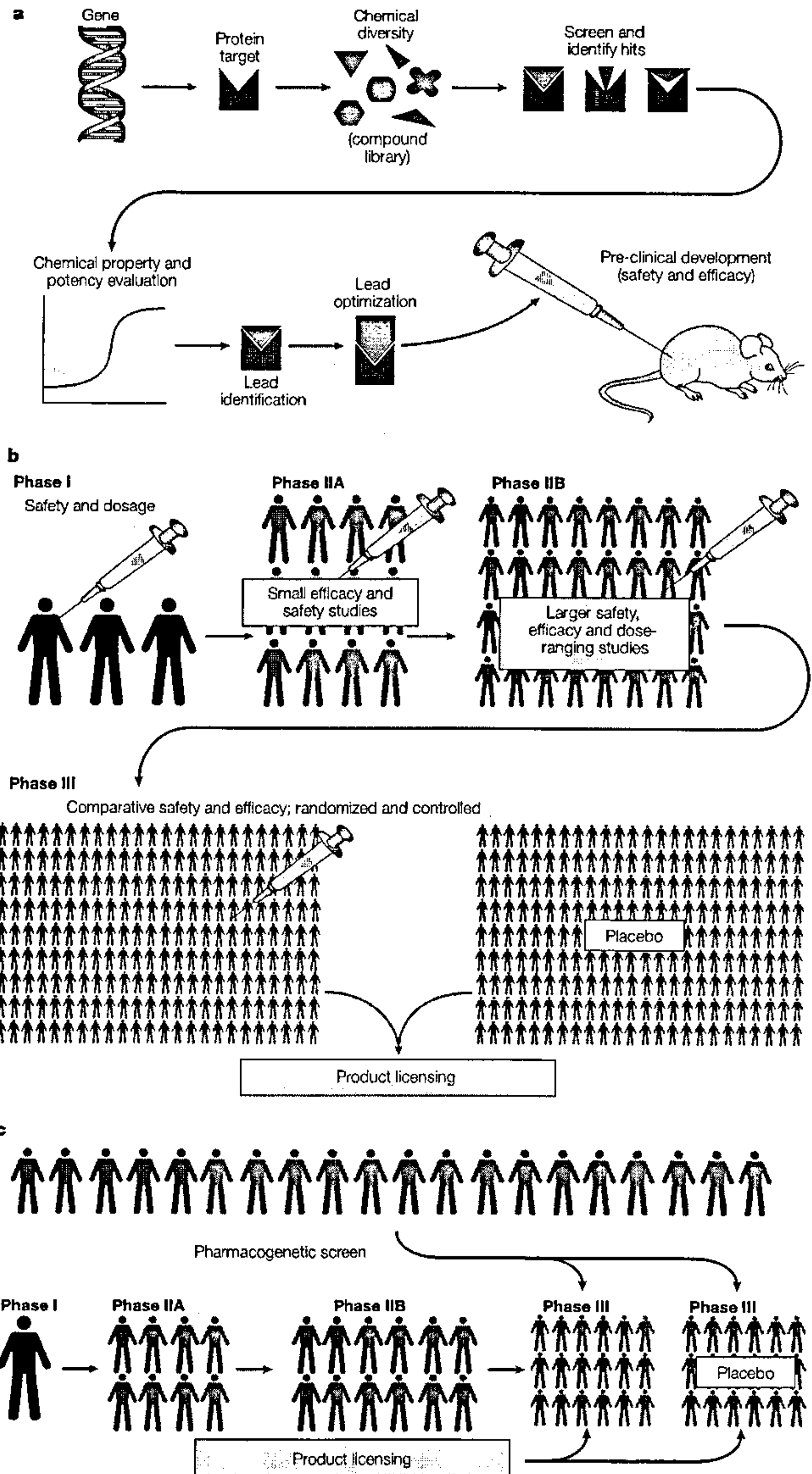
Genetics Research, GlaxoSmithKline, Five Moore Drive, Research Triangle Park, North Carolina 27709, USA. e-mail: allen.d.roses@gsk.com
doi:10.1038/nrg1432

©2004 Nature Publishing Group

Box 1 | **Pharmacogenetics and the pharmaceutical pipeline**

From having the sequence of a gene that encodes a potential target for drugs to actually having a medicine that interacts with that target involves a lengthy, expensive and complicated pharmaceutical pipeline that usually requires several years of basic science for target validation before chemical screening (see figure part a). Once there is a positive decision to progress a target, an effective screening assay that can allow the high throughput of many thousands of known chemical entities must be designed and implemented. Molecules that affect the target (hits) must then be evaluated for chemical properties and potency before those that are worth pursuing (leads) can be identified, synthesized, evaluated and modified for drug qualities (lead optimization). When a lead is identified, considerable pre-clinical development must also occur, particularly in the fields of toxicology, drug kinetics and drug metabolism. All of these processes must occur before the first dose of any new molecule can be tested in humans (part b).

Following initial clinical testing for safety in humans (PHASE I), the molecule enters the most crucial phase, PHASE IIA, during which the desired clinical effect (that is, efficacy) is addressed in a relatively small, but still expensive, clinical trial. For example, if 100 patients participate in a Phase-IIA study at a minimum cost of US $10,000 per patient, the study would cost US $1 million. Even after a molecule demonstrates efficacy, more extensive dose-ranging PHASE-IIB studies, as well as PHASE-III studies, cost several hundreds of millions of dollars. With costs as high as this, only molecules for which there is good evidence of efficacy and a reasonable biological rationale for its mechanism of action are selected for full development. Prospective efficacy pharmacogenetics early in clinical development can be used to select patients who would be predicted to be potential responders (green figures in parts b and c) and hyper-responders (blue figures). By selecting out patients in Phase-II and -III clinical trials who would be predicted to be non-responders (red figures in parts b and c), subsequent clinical trials can be made smaller, faster and less expensive (part c).



©2004 Nature Publishing Group

the pathogenesis of the disease and/or the relevant biology of the target, and not because they are simply a newly sequenced gene[11]. To study the genetics of human disease or even to confirm targets in patients from those identified in animal models, large resources of well-phenotyped patients and controls are necessary[15–18]. Despite this seemingly obvious need for well-phenotyped patients, the support of clinical research has not kept pace in academic institutions[19]. Conversely, the pharmaceutical industry, as a whole, does not have sufficient clinical expertise or the patience to invest in long-term strategic research of this type[19,20].

My aim here is to examine the current uses of PGx in drug discovery and development, particularly through high-throughput, genome-wide screening applications. With respect to early drug discovery, the disclosure of early targets and leads in the pipeline is limited by regulatory and commercial concerns. Nonetheless, here I attempt to illustrate, with examples, how high-throughput genotyping, PGx and pharmacogenomics can be applied across the pharmaceutical pipeline (BOX 1).

PHASE I
Initial clinical studies usually involve small numbers of normal human volunteers and small doses to assess safety, metabolism and excretion of a drug or drug combination.

PHASE IIA
After the initial Phase-I studies, these randomized and controlled clinical studies are used to assess efficacy (proof of concept) and to continue to assess a drug or drug combination in a relatively small number of patients (tens of patients). These early studies can be non-comparative or double-blind comparisons.

PHASE IIB
After Phase-IIA studies, these randomized and controlled clinical studies are used to assess the efficacy and dose-ranging of a drug or drug combination in larger groups of patients (hundreds of patients). In some designs, there is a positive interim end point so that Phase IIB can be extended into Phase-III registration trials. Safety continues to be assessed.

PHASE III
Randomized and controlled clinical studies in patients (thousands of patients) designed to evaluate the comparative safety and efficacy of a drug or drug combination. Also includes the principal data used by regulatory agencies to approve or reject a product-licensing application.

EFFICACY PHARMACOGENETICS
The study of genetic variation that underlies variability in the efficacy of drugs for treating disease.

Target discovery

Target discovery is important in supplying the drug-development pipeline with disease-relevant molecules. Although there might be approximately 30,000 human genes, few currently have the capacity to be screened against large chemical libraries to identify specific interactions. Those that can be screened form groups of so-called 'tractable' or 'drug-able' targets. Widely appreciated target classes, such as nuclear receptors, kinases and 7-transmembrane repeats (7-TMs), constitute approximately 1,800 genes, but there are additional custom screens for many enzymes. The availability of advanced drafts of the human and mouse genome sequences means that genes can now be rapidly analysed *in silico* for characteristic coding regions that are similar in tractable gene classes[21–26]. Before this capability, the pharmaceutical industry knew of perhaps 500 targets[26]. There might be up to 5,000 additional tractable targets, of which perhaps a few hundred might be relevant to the treatment of human disease[24,25].

The standard genetic association experiment is to test variations of specific candidate target genes for association to a defined human disease. Using high-throughput genotyping, bioinformatic support and robotics, it is now possible to do such experiments on hundreds of genes, using thousands of SNP variations[27,28]. Two sets of reagents are required for large-scale, disease-association analyses: validated SNP assays and DNA from clinically well-phenotyped patients and appropriate controls[27–31]. The practical problems over the past decade have centred on patient collections of DNA, which mainly reside in academic laboratories, whereas chemical screening, optimization and regulatory compliance capabilities mostly reside in industry.

For a major pharmaceutical company that is attempting to screen thousands of SNPs from potential target-class genes against thousands of patients, obtaining the DNA reagents from well-characterized patients was recognized as a functional bottleneck[31,32]. Patient and control individuals must be evaluated one at a time, and cannot be evaluated clinically in a high-throughput model. Although technical genotyping capacity is now generally available, the essential patient and control populations with DNA, detailed prospective clinical evaluations and adequate informed consent are rare (see BOX 2). Nonetheless, strategies for high-throughput experiments to identify disease-associated tractable drug targets are now available[33].

Box 2 | The importance of phenotype

The accuracy and the amount of detailed phenotypic data that can be compared in disease-association experiments is crucial to the interpretation of the data. In general, the phenotypic data that clinicians collect determine what questions can be addressed with such experiments. Detailed specific phenotypic data provide many more opportunities for phenotype-specific questions compared with simple diagnostic designations. Complex-disease diagnoses are expected to be heterogeneous. For example, depending on the clinical data obtained, up to 30% of patients designated as having possible Alzheimer disease might actually have other forms of dementia[66]. It would therefore be helpful to collect more detailed data, including pathological information and follow-up data on the patient's clinical course.

Similar to disease diagnoses, the clinical measurement of drug efficacy is variable. Graded clinical responses as well as placebo effects can complicate such measurements. Standardization of clinical protocols for the comparison of drug responses would therefore be both useful and more accurate. Generally, an adverse event from a drug can be assessed more objectively. Such events usually occur a short time after the patient experiences the drug, so the precise phenotypic response to the drug can be specifically, sensitively and accurately documented. Pharmacogenetic analyses can be used to differentiate phenotypic heterogeneity, to segment populations that seem to be responsive or unresponsive to a medicine, or to accurately define individuals who might be at higher personal risk of an adverse event.

Prospective application of efficacy PGx

Usually, PGx has been applied retrospectively to develop molecular diagnostics to optimize drug selection and dosing[34–38]. However, here I focus on the prospective application of EFFICACY PGX to reduce attrition of effective drugs in early drug development. In this context, efficacy PGx is not being used to define diagnostic markers to select or predict individual patients who will respond to therapies. Rather, these prospective strategies can be used to increase confidence in a molecule at a crucial stage in its development. Phase IIA is the first time that a molecule is tested for its desired clinical effect in humans. Phase-IIA trials are not exact, nor are they usually large enough for convincing statistical significance. Rather, they provide confidence that a molecule would be effective in subsequent clinical trials. In practice, most candidate drugs that reach human testing fail in Phase IIA. The clinical measurement of efficacy is also

©2004 Nature Publishing Group

not exact, as placebo effects can cloud the assessment of phenotypes[39,40]. Typically, there are apparent responders and non-responders: the goal of prospective efficacy PGx in this context is to differentiate these groups and select responders for subsequent clinical trials. For 'INTENT-TO-TREAT' decisions in subsequent clinical trials, differentiated patient-responding groups can provide confidence in the next (expensive) clinical trial investments.

The output of pharmaceutical pipelines over the past decade has been disappointing[41,42]. Prospective efficacy PGx offers the possibility of reducing the size and expense of clinical trials by differentiating the patient groups who are more likely to respond (BOX 1). There are essentially three types of result for candidate therapies in Phase-IIA clinical trials: first, the comparison of drug versus placebo demonstrates clear clinical efficacy; second, there is no appreciable efficacy; or, third, there is some efficacy but it seems to be in a treated subgroup. The first two results are easy to interpret: the drug-development programme either continues or stops. The third is problematic: more supportive data are usually required before a major financial commitment to Phases IIB and III (full development) is made. However, owing to the huge costs of development, those programmes with 'borderline' efficacy present an increased risk for development and are generally stopped[41,42]. However, the cost of additional clinical trials might be reduced if the population of responders and non-responders could be segmented on the basis of their genetic profiles early in Phases IIA and IIB. Removing apparent non-responders could enrich the remaining patient population for subsequent clinical trials. It can then become feasible to consider further testing of a molecule with subgroup efficacy rather than cancelling the programme.

There are no published clinical examples of prospective applications of efficacy PGx to the early Phase-II pipeline: all of the published, peer-reviewed literature of efficacy PGx has been retrospective[34–38]. For the purpose of guiding efficient drug development, a PGx strategy that can be applied prospectively at the early stages of clinical development is necessary. However, one of the most relevant examples of a retrospective study that subsequently allowed the prospective application of a diagnostic test to progress a drug at the later clinical stages is the case of trastuzumab (Herceptin; Genetech). This drug was developed in late Phase III on the basis of the differentiation of patients whose tumours would be more responsive to therapy[43–45] (BOX 3).

The essential feature of any drug-efficacy analysis is the definition of phenotypic response. Although all phenotypes are recorded during the double-blind trials, efficacy can be assessed only after the double-blind coding is broken. In early Phase-IIA studies, with a limited number of participating patients and controls, any successful evidence-based methods to increase confidence in a decision to continue development are always helpful. Whether or not the SNPs identified in Phase IIA as being associated with efficacy are really useful in subsequent clinical studies can be confirmed in reiterative studies. In fact, follow-up trials also provide the opportunity for reiterative screening for additional regions of phenotypic associations mapped across the genome in regions of extended LINKAGE DISEQUILIBRIUM (LD).

Proof-of concept study. Molecules entering Phase IIA are generally not the subject of publications because the information at that stage is proprietary and the studies are not meant to be diagnostic. However, it is important to illustrate for this review that prospective efficacy PGx in early development is currently a pipeline strategy. FIGURE 1 illustrates the results of a recent Phase-IIA trial of a molecule to treat obesity, with 40 drug-treated and 41 placebo-treated obese patients. These results are unpublished but have previously been presented publicly in several forums (see, for example, the Food and Drug Administration (FDA) Science Board presentation in the Online links box). The two curves plot weight change in the treated and placebo patients during the double-blind clinical trial. This clinical trial was performed over a 2-month period with a defined therapeutic target of weight loss on the basis of current regulatory guidance for the approval of obesity medicines.

Box 3 | **ERBB2 and Herceptin for breast cancer: an example of pharmacogenomics**

ERBB2 is a 185-kDa (mw) tyrosine kinase receptor that might be overexpressed in 25–30% of human breast cancers. Overexpression of ERBB2 is associated with enhanced tumour aggressiveness and a high risk of relapse and death[43,44]. It is also a clinically relevant and validated drug target for breast cancer. Trastuzumab (Herceptin; Genetech) is a humanized monoclonal antibody against the ERBB2 receptor that is approved for the treatment of breast cancer. It was developed in the 1990s as a means of blocking the ERBB2 receptor, and therefore the aggressive spread of the cancer, and of selectively tagging cancer cells for the body's immune system to attack. Retrospective re-examination of the original clinical trials of Herceptin showed that a therapeutic response was more likely in patients with tumours that overexpress ERBB2 protein[44]. So, the measurement of ERBB2 overexpression is useful for assessing whether treatment with Herceptin can help specific patients. The availability of a diagnostic test for a subgroup with a better chance of responding to treatment with Herceptin allowed this drug to progress through the pipeline to approval, and now it is helping many patients with cancer. However, even in tumours that overexpress ERBB2, the therapeutic response to Herceptin therapy is only approximately 45%, with a median time to observable tumour progression of 7–8 months[44]. So, ERBB2 is now a target for other potential therapeutic agents. The Herceptin example highlights how useful a pharmacogenetic test can be for progressing drugs through the pipeline. Moreover, as discussed in the text, the earlier the stratification of patients can be used, the more efficient the design of subsequent clinical trials and the greater the number of partially effective molecules that can be progressed.

INTENT TO TREAT
Analysis of clinical-trial results that includes all data from patients in the groups to which they were randomized (that is, assigned through random distribution) even if they never received the treatment.

LINKAGE DISEQUILIBRIUM
The condition in which the frequency of a particular haplotype for two loci is significantly different from that expected under random mating. The expected frequency is the product of observed allelic frequencies at each locus.

©2004 Nature Publishing Group

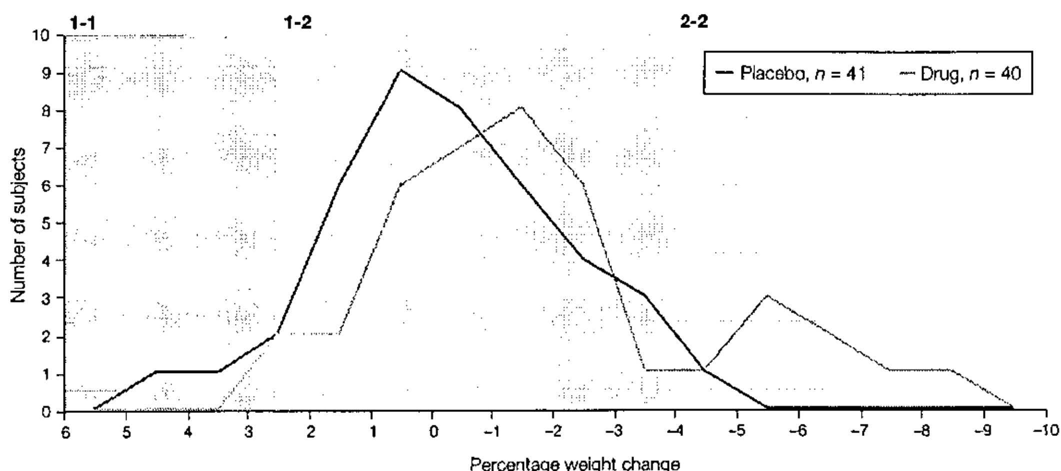


Figure 1 | **Efficacy pharmacogenetics for an obesity drug.** The graph illustrates the results of a small double-blind Phase-IIA efficacy clinical trial of a molecule intended for the treatment of obesity. Weight change in the 40 drug-treated patients (purple line) is compared with 41 placebo-treated patients (blue line) during the 2-month double-blind trial. There is an obvious hyper-responder subgroup in the treated patients, with patients losing as much as 9% weight during the trial. Two SNPs from two candidate genes that are related to the proposed mechanism of action of this molecule, and one SNP from another candidate gene that is thought to be implicated in theories of obesity, segregated with the hyper-responders. Each hyper-responder was homozygous for a single allele (labelled 2-2; light shaded section of the graph) with patients on the left side of the curves being more likely to be 1-1 homozygous (dark green section). Heterozygous patients (1-2; light green section) clustered in the middle. In this experiment, treated patients with the 2-2 genotype for any of the three SNPs on average lost ~3.3 Kg, whereas treated patients with the 1-1 genotype gained an average of ~1.3 Kg. This pattern reassures us that the molecule has efficacy and that subsequent Phase-IIB trials might be enriched by using only patients who are 1-2 heterozygous and 2-2 homozygous, with the exclusion of 1-1 homozygous patients. In addition, although the 1-1 subgroup might be less responsive to this specific treatment, this subgroup could be used in clinical trials of other obesity-drug candidates, which might subsequently allow a drug to be developed that is complementary to the first. *n*, total number of patients in study group.

It is obvious that there is a large group of hyper-responders who represent approximately 20% of the treated patients. There is no similar delineation of a hyper-responder group with placebo treatment. Perhaps of even greater medical significance is the shift of the treated curve towards weight loss. This broader distribution of weight loss provides additional confidence in progressing to full drug development using a population of potential responders as well as hyper-responders.

Proof of concept: candidate-gene approach. The fundamental question concerns whether it is possible to differentiate the hyper-responding population, the patients with potentially smaller responses and those with the greatest risk of gaining weight. Can this be accomplished using simple, rapid and inexpensive analysis that is applicable to a typical Phase-IIA trial in which the decision of whether or not to move development ahead needs to be made rapidly? One approach to these questions is to try to identify SNPs that segregate with drug efficacy in candidate genes that relate to the mechanism of the molecule or that might be involved in the disease being studied. A retrospective analysis of 42 variants that have been associated with a drug response at least twice supports the use of receptors as candidate genes for prospective efficacy PGx: 21 of these variants were the target or were in a known pathway of the target[4,46].

For the example illustrated in FIG. 1, 21 candidate genes and 112 SNPs that represent potential polymorphic sites in the mechanism of action were tested. A relatively simple analysis of allele segregation of the 21 candidate genes across the drug-treated and placebo populations identified 3 SNPs that segregated with the variation in drug responses. One allele of each SNP associated with the hyper-responsive group (2-2) and the other allele associated with the least responsive group (1-1). By contrast, heterozygous patients segregated in the middle where the treated subject curve shifted towards higher weight loss compared with the placebo. So, in this case, our unpublished observations indicate that it is possible to segment the group using SNP variants in candidate genes related to the molecular receptor of metabolism of the molecule.

Proof of concept: genome-wide screening. Another consideration is whether whole-genome analyses can be used to identify markers of efficacy. For example, had the initial candidate-gene screens failed in this case, a whole-genome scan for homozygous polymorphic SNPs that differentiate the hyper-responder group could have been another option.

To assess this option, local extended SNP maps — similar to those made for susceptibility genes for apolipoprotein E4 (APOE4) and Alzheimer disease, migraine and psoriasis[10,47,48] — were created to determine whether the candidate-gene SNPs resided in regions of extended LD that were detectable by whole-genome screening. These unpublished observations indicate that each genetic susceptibility effect resided in

©2004 Nature Publishing Group

a detectable region of extended LD. So, in this case, it would have been possible for high-density whole-genome SNP mapping (100,000–200,000 SNP-mapping density) to be used to define regions of extended LD that select the hyper-responders (2-2; FIG. 1) and the non-responders (1-1). The false-positive LD regions that might be expected when testing 100,000–200,000 SNPs could be validated in subsequent clinical studies.

So, a few SNPs identified from candidate genes or whole-genome scans of Phase-IIA-tested patients could allow non-responders to be excluded from subsequent clinical trial studies, therefore allowing enriched, smaller, faster, less expensive clinical studies on patients with a better chance of responding favourably. A larger Phase-IIB study (~25% larger) is currently in progress in which patients who would have been selected on the basis of the SNP variants (that is, excluding the 1-1 genotypes) will be compared with the larger, all-inclusive population. This experiment is designed to confirm the strategy of applying efficacy PGx prospectively at an early stage of the pipeline and to provide an estimate of cost savings for subsequent trials. However, already this proof-of-concept study shows that with prospective efficacy PGx it might be possible for medicines with variable efficacy to continue through the pipeline, when in the past their development would have been too expensive. Identification of non-responders early in the development process might also mean that other potential drugs can be targeted for development for this segment of the patient population. Of course, the general applicability of a prospective PGx strategy to reduce the size and to enrich responders in subsequent clinical trials is still the subject of debate. It is certainly not likely that in every case the genetic basis of lack of efficacy will be as clear-cut as that discussed here. As in all PGx studies, the results with a specific drug and a selected patient population will be context-dependent. However, the unpublished observations discussed here do at least show that this strategy is possible.

To apply this strategy more broadly, GlaxoSmithKline now store patient DNA in every new Phase-IIA study so that when trials indicate that a molecule is efficacious for part of the patient population, we can proceed with efficacy PGx analyses. There are indications that other companies are also applying efficacy PGx[49,50]. For these companies, the application of prospective efficacy PGx with SNP genotyping does not demand a new model for clinical trial designs: currently accepted clinical chemistry or imaging studies can also be used to select patients for trial inclusion or exclusion. However, this strategy might provide a more precise and more generally applicable decision tool.

Clinical and commercial implications. The purpose of the prospective efficacy strategies discussed above is not to design 'personalized medicines', but to enrich subsequent clinical trials with patients who are more likely to respond[51–53]. This application of prospective efficacy PGx can be viewed as an early version of the same-patient selection process that is exemplified by the selection for treatment with Herceptin of patients who overexpress ERBB2 and who are more susceptible to breast cancer (BOX 3). The proof-of-concept study discussed here illustrates that this use of prospective PGx does not necessarily limit the available market for a drug. In this case, heterozygotes still have the capacity to respond to the medicine, so, by excluding predicted homozygous non-responders from subsequent clinical trials, it is more likely that a drug with broad applicability is eventually registered.

Medicines — particularly those that are relatively safe — are currently prescribed for segments of the population even when no efficacy is expected[54]. For example, montelukast (Singulair; Merck Sharpe and Dohme), a leukotriene-receptor antagonist that is used as a treatment for asthma, is not therapeutic for 14% of the population that lack the specific leukotriene-receptor activity[54,55]. Regulators have not required leukotriene-receptor testing to identify this group of non-responsive patients as the medicine is relatively safe. So, efficacy PGx analyses for drug development do not, *de facto*, present new challenges for prescribing medicines. Moreover, although the trial results will be disease- and treatment-specific, in general there is a greater chance of predicting effective responses and reducing the time and expense of drug development[11].

Few currently available medicines actually treat all patients effectively: some medicines are licensed on the basis of only 30% efficacy in clinical trials[56]. The lower the percentage of responders that can be detected, the larger and more expensive the registration clinical trials. At a cost of approximately US $1 billion to bring a drug to the market, clinical trials must be of a manageable size[42]. As medicines are generally not effective in all people, patients and physicians accept that in practice many medicines are prescribed by trial and error, testing first one medicine, then another, then another. Unfortunately, the adverse risks that are associated with each medicine can be additive. So, there is an important medical advantage in determining whether a patient is likely to respond, and in determining an individual's personal risk of adverse events (AEs)[57,58]. Prospective efficacy PGx studies on specific molecules can also provide data that can feedback into subsequent research efforts such that additional lead-candidate molecules can be targeted for patients with continued unmet need — that is, non-responders to the first molecule tested.

Regulators consider the risk/benefit ratio in medicine approvals, but their first and most important concern is individual safety[10,11]. In clinical development, efficacy PGx should be viewed as distinct from safety PGx: the former is useful for patient segmentation whereas the latter is highly specific for each individual. However, in some specific cases, such as cancer therapies, efficacy PGx might be used more specifically to select patients to avoid adverse treatment effects in those in whom there is little chance of efficacy. Early efficacy PGx (in Phases I and II) can also lay the foundations for identifying patients who require a different dosage regime — either higher to achieve efficacy or lower due to early safety signals.

©2004 Nature Publishing Group

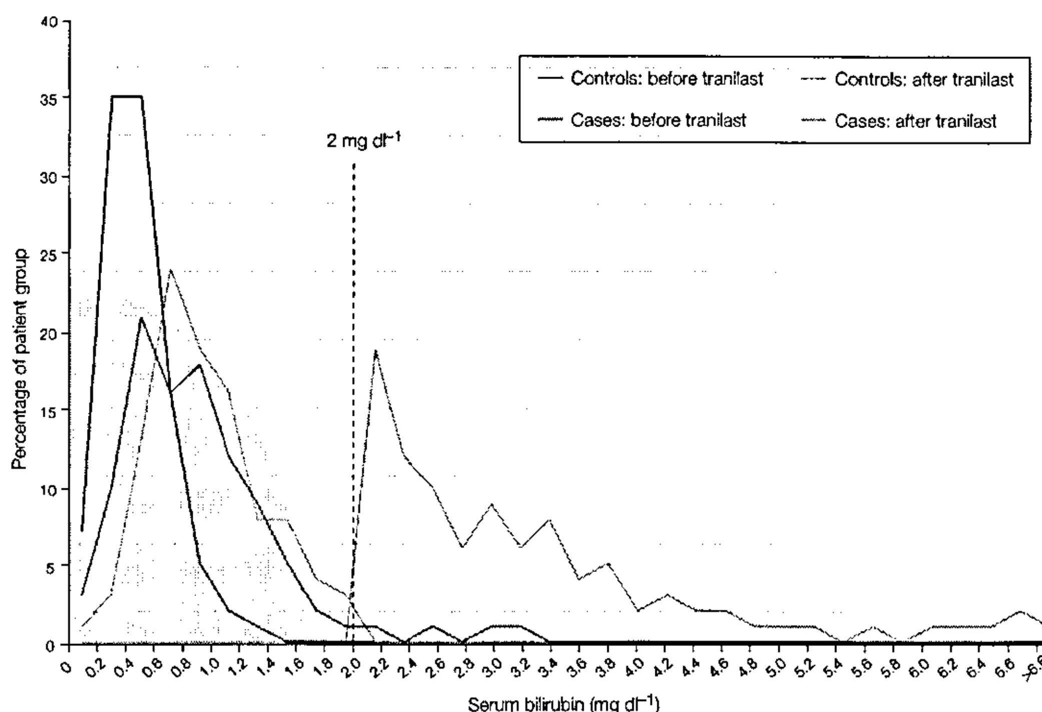


Figure 2 | **Bilirubin levels in patients treated with tranilast during a Phase-III clinical trial.** This figure illustrates an example of a mild adverse event with a clear genetic component that can be used as a model for the use of safety pharmacogenetics. Patients in a trial for TRANILAST who developed bilirubin levels of >2 mgdl^{-1} are considered to be HYPERBILIRUBINAEMIC. The scale of the figure is adjusted to show 147 patients who developed hyperbilirubinaemia (to the right of the dotted line), and 1,054 subjects who did not develop bilirubinaemia after tranilast treatment (to the left of the dotted line). The dark green peak on the left represents the 'controls' (that is, those who did not develop hyperbilirubinaemia after tranilast treatment) before they were administered the drug, and the light green peak represents the same patient group after treatment with tranilast. The red peak represents pre-treatment patients who became hyperbilirubinaemic and the orange peak represents their post-treatment data. High levels of bilirubin were most common in the 7-repeat homozygotes and low levels were most common in the 6-repeat homozygotes, whereas patients with 6-6 and 6-7 genotypes more frequently had intermediate levels of bilirubin (see text). Approximately 96% of the tranilast-treated patients who were hyperbilirubinaemic carried the 7-7 genotype at the *UGT1A1* gene, whereas almost all the treated subjects who did not develop hyperbilirubinaemia carried the 6-6 or 6-7 genotypes.

TRANILAST
The name that was used for a specific anti-restenosis drug while it was being investigated by GlaxoSmithKline.

HYPERBILIRUBINAEMIA
A high level of bilirubin in the blood. This can cause yellowing of the skin (jaundice).

PERIPHERAL NEUROPATHY
A problem in peripheral nerve function (any part of the nervous system except the brain and spinal cord) that might cause pain, numbness, tingling, swelling and muscle weakness in various parts of the body. Neuropathies might be caused by physical injury, infection, toxic substances, disease (for example, cancer, diabetes, kidney failure or malnutrition) or drugs such as anti-cancer drugs.

CHARCOT–MARIE–TOOTH NEUROPATHY
A genetic disease that is characterized by progressive peripheral neuropathy and debilitating muscular weakness, particularly of the limbs. There are multiple inherited mutations of several genes that result in similar phenotypes.

Safety PGx

Once a molecule is in clinical development, patient safety is the main concern. There are two important places in the development pipeline where PGx studies can contribute to safety. The first occurs during early clinical trials in which indications of a potential future AE can occur. Observations of AEs can present considerable risks to a development programme. However, these risks can be effectively managed during clinical trials to allow 'Go/No Go' development decisions to be made in a timely manner, cutting the dead time between the steps in the progression of a drug through the pipeline. The second application of safety PGx occurs post-launch when AEs begin to be observed only after tens of thousands of patients experience the drug. This is the most dangerous time for patients for new safety concerns to arise, and the most expensive time for drug developers after full development and launch costs. Serious AEs might lead to product withdrawal. It is at this stage that the accurate identification of individuals at risk for uncommon AEs would be invaluable. This implies an improved system for post-marketing surveillance in which DNA (or PGx panel data in the future) from AE patients is available, enabling the rapid clarification of a diagnostic profile (or test) of extremely high sensitivity and high specificity that can be identified rapidly. It is in this context that the use of whole-genome PGx is particularly relevant. As an aside, safety PGx would also be applicable at this stage for generic, over-the-counter drugs and other remedies (such as dietary supplements).

How many AE patients? One of the key questions for safety PGx is how few patients will be required to create a useful predictive profile of AE susceptibility on the basis of whole-genome SNP mapping. Subsequent considerations would include how to relate these data to clinical practice and drug surveillance.

Specific AEs can be associated with carriers of rare disease genes. An old example is the acceleration of PERIPHERAL NEUROPATHY in patients with hereditary diseases (for example, CHARCOT–MARIE–TOOTH NEUROPATHY) that use anti-cancer medicines[59]. In this case, testing for the disease gene provides a diagnostic surrogate marker for those cancer patients who might be at increased risk of accelerated neuropathy. There are many other cases in which inherited disease-gene variants, or those from multiple susceptibility genes, can increase susceptibility

©2004 Nature Publishing Group

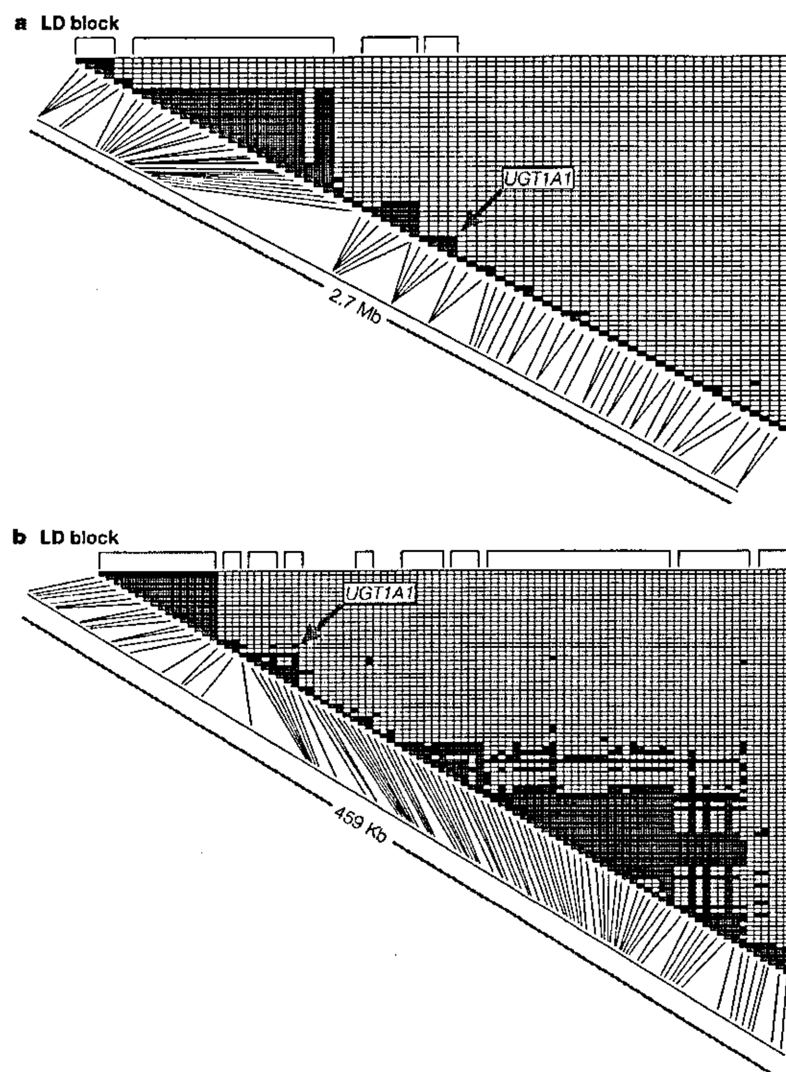


Figure 3 | **Association mapping of adverse-event susceptibility: tranilast and hyperbilirubinaemia. a** | A diagrammatic representation of pairwise linkage disequilibrium (LD) between 76 SNPs across a 2.7-Mb region that contains the *UGT1A1* gene, representing a mean SNP density of approximately 1 SNP per 35 Kb. SNPs were typed in a group of 1,054 patients who received tranilast but who did not experience clinical hyperbilirubinaemia and in a group of 147 patients with hyperbilirubinaemia who received tranilast. SNPs are ordered using the NCBI map build 30. Red squares indicate strong LD (ABSOLUTE D′ >0.8); clear squares indicate D′ <0.8. The bracketted regions along the top of the figure indicate LD blocks. The arrow points to the location of *UGT1A1*. The long diagonal black line along the figure indicates the physical length of the region, with short black lines showing the position of SNPs. Note that the large region of LD at this density is apparently adjacent to *UGT1A1* (REF. 10). **b** | Higher-density (~5-Kb) SNP map around the *UGT1A1* gene. As a follow-up, a higher-density scan with more SNPs across a smaller region was performed around the region identified in a typical whole-genome screen. A diagrammatic representation of pairwise LD (D′) between the 99 SNPs across a 459-Kb region, with a mean SNP density of approximately 1 SNP per 5 Kb. This density is approximately that which is the target of the International HapMap Project[49]. This region contains several LD blocks over which a region of extended LD was indicated in part **a**. This type of selective, sequential higher-density screening confirms the regions indicated by less dense scans[10].

to an AE. However, even healthy individuals can have a genetic predisposition that places them at risk of AEs associated with a specific medicine. Given this possibility, it is natural to consider how few AE patients would need to provide DNA samples to identify a predictive SNP profile and whether such an approach to AE risk-management is a practical way of preventing useful drugs from having to be withdrawn from the market.

SNP profiling of AEs: the tranilast example. The anti-RESTENOSIS drug known as tranilast, which for efficacy reasons is no longer in development, provides a proof of concept for successfully addressing safety concerns during clinical trials[9,10] (FIG. 2). During a Phase-III Tranilast clinical trial of 11,500 patients, approximately 4% developed hyperbilirubinaemia. Obviously, this caused considerable concern and created potential labelling restrictions for the drug. In an attempt to identify the genetic factors that underlie this mild AE, candidate genes were screened for variants that associated with it. One of these variants, a DNA repeat polymorphism in the UDP-glucuronosyl transferase 1A1 (*UGT1A1*) gene — a published aetiological factor for GILBERT'S DISEASE — was highly associated with the patients who were hyperbilirubinaemic[60]. *UGT1A1* alleles with 6 or 7 repeats are the most common forms in human populations: the 7-repeat form was highly associated with the AE. This association was identified during the course of the clinical trial. Subsequently, breaking the placebo- versus drug-treated codes at the conclusion of the trial showed that all 7-7 patients who developed hyperbilirubinaemia received the drug, whereas none of the 7-7 patients treated with the placebo developed the AE. Several drug-treated patients with the 6-7 genotype also developed mildly elevated levels of bilirubin, but no treated or placebo patients with the 6-6 genotype became hyperbilirubinaemic (FIG. 2). All treated patients had a clinically insignificant increase in bilirubin while taking tranilast, but hyperbilirubinaemia occurred primarily in those patients with 2 7-repeat alleles.

Although hyperbilirubinaemia is only a mild AE, the large amount of data generated from this trial is useful for addressing the question of how few patients must be observed to develop a highly sensitive, highly selective SNP profile for AEs in general. To model a typical genome scan of 100,000–200,000 SNPs covering the genome, we constructed an extended SNP map around the *UGT1A1* gene, with validated SNPs located at approximately 36-Kb intervals[10]. The SNPs were selected from public SNP maps and were located in a 2.7-Mb region that encompasses the *UGT1A1* gene (FIG. 3a). Initially, 76 validated SNPs were typed in a group of 1,054 patients who received tranilast but who did not experience clinical hyperbilirubinaemia and in a group of 147 patients with hyperbilirubinaemia who received tranilast. The data are analogous to those obtained for the susceptibility genes for Alzheimer Disease, migraine and psoriasis[16,46,47]. A region of extended LD that was limited to approximately 150 Kb was identified[10]. An important point to consider, with respect to whole-genome SNP mapping, is that this small region of LD that contains the *UGT1A1* gene would have been found by an initial 100-Kb-density whole-genome SNP scan.

A second follow-up SNP screen of higher density that involved 99 validated SNPs spanning a 459-Kb region (average SNP interval of 5 Kb — the current International HapMap Project[61]) surrounding the LD

©2004 Nature Publishing Group

Table 1 | **Probabilities of associations between four SNPs near the *UGT1A1* gene and an adverse response to tranilast**

Number of cases	Number of controls	SNP poly ID 4082379	SNP poly ID 3729885	SNP poly ID 3730948	SNP poly ID 3737550
10	3,000	0.10392	0.01542	0.04623	0.00644
20	3,000	0.00143	4.37×10^{-6}	0.00014	9.96×10^{-8}
30	3,000	3.93×10^{-6}	2.91×10^{-7}	4.14×10^{-5}	5.59×10^{-9}
50	3,000	8.69×10^{-8}	7.39×10^{-8}	2.47×10^{-5}	1.32×10^{-10}

The statistical significance data obtained from 4 SNPs that surround the *UGT1A1* gene (not the 6-7 repeat polymorphism itself) are illustrated. Each row indicates the association data with sets of adverse-event patients of increasing size. The first row compares 10 Caucasian patients with tranilast-induced hyperbilirubinaemia with 3,000 Caucasian controls. The second, third and fourth rows compare different sets of 20, 30 and 50 patients with tranilast-induced hyperbilirubinaemia, respectively, with the same 3,000 controls. SNP poly ID, SNP polymorphism identification number.

region that was identified in the primary screen demonstrated that the SNPs in LD around *UGT1A1* were located across several LD blocks (FIG. 3b). The associated SNPs were mapped to several adjacent LD blocks (extended LD), but still in a region that allowed LD with the *UGT1A1* gene to be demonstrated.

This study shows the practicality of a two-phase SNP-mapping strategy for identifying AE-susceptibility loci. An initial whole-genome scan with a SNP density of 36-Kb might be practical for routine initial screens, followed by a higher-density map that is limited to LD regions indicated by the initial scan. Of course, the size of the LD blocks associated with other AE loci will also influence whether they could be detected using this approach.

The practicality of PGx surveillance: the tranilast example. The SNPs identified as being in LD with the 6-7 repeat polymorphism were tested to determine how few patients would have been required to identify the region[10]. The statistical significance of associations between 4 SNPs that surround the *UGT1A1* gene (but not the 6-7 repeat polymorphism itself) and hyperbilirubinaemia was calculated for different numbers of cases compared with many controls (TABLE 1). These data show, in this case at least, that with a large control population, relatively few patients with AEs are required, and can involve as few as 10 patients. With 20 patients, the association is highly statistically significant to the extent that it would still be significant after correction for the 100,000 SNP-association tests required in a whole-genome screen. With 30 patients, the association data are overwhelming.

In cases in which a second or third locus is necessary to manifest the clinical AE, the number of patients required for a highly sensitive and specific identification of the SNP LD association profile might be even less. The relevance of these data for other drugs will depend on relative gene–environment reactions. For example, variants with stronger, more penetrant phenotypic effects might be easier to detect, whereas other AEs might be more difficult to recognize clinically and might need more patients[10]. Similarly, it should be noted that in cases in which the effect of a given variant is different in different genetic backgrounds, more patients might be required. Moreover, for cases in which the inheritance of a variable drug response is more complex (that is, influenced by many genes of much smaller effect than *UGTA1*), more patients would be needed.

The theory and practice of PGx surveillance. Recently, a simple statistical framework, based on likelihood ratio and empirical BAYES FACTOR analyses, has been developed for PGx surveillance[62]. This method estimates the average objective positive evidence that a reported individual gives in classifying AE cases as genetically different from controls. This theoretical analysis indicates that it would not be until the fourth or fifth AE that these events would be recognized as being linked to a drug. Unfortunately, in the real world, all patients who take a new medicine are not under direct observation, nor are sufficiently rapid and reliable reporting surveillance procedures available. In fact, the reporting process for AEs is based on selective self-reported, physician-reported or attorney-reported cases. If the rate of such an AE is actually very low — in the range of 1 out of every 10,000 patients — then a very large cohort of treated patients would need to be observed.

A potential solution to improve surveillance might be sampling and privacy-protected DNA banking from the first 250,000 patients who are prescribed a newly marketed medicine, or a similar risk-management system coordinated with regulatory agencies[9]. Once patients with an AE are diagnosed, their banked DNA (anonymous to the pharmaceutical company) would be immediately available for SNP screening for associated regions of LD. Of course, this strategy would be only an interim measure in anticipation of a day when individuals would carry their standardized SNP profiles on the magnetic information strip of an identification card for use with each new prescription[57]. In this situation, an updated databank with PGx information could be made immediately accessible from a physician's office as any new medication is being considered. The patient's SNP profile could then be compared with known drug-response SNP LD profiles, or could be used to identify new AE profiles[63,64].

An important consideration in any association study is the definition of controls. In practice, the association of a disease gene or a drug-specific clinical event requires that the control population represents the at-risk population. The more representative the control population, the fewer patients with AEs are necessary to define deviation or differences. The number of AE cases required also decreases as the genetic influence of the variant increases (FIG. 4). An analogous example is the medical diagnosis of a rare hereditary disease. In this situation, control groups are generally

ABSOLUTE D′
For specified alleles at two distinct loci, D′ is the absolute difference between the observed and expected haplotype frequencies, divided by the maximum value that the difference could possibly attain.

RESTENOSIS
A re-narrowing or blockage of an artery at the same site at which treatment, such as an angioplasty or stent procedure, has already taken place.

GILBERT'S DISEASE
A benign syndrome of increased sensitivity to external drugs that causes hyperbilirubinaemia but does not progress to severe liver impairment.

BAYES FACTOR
The ratio of the posterior odds to the prior odds.

©2004 Nature Publishing Group

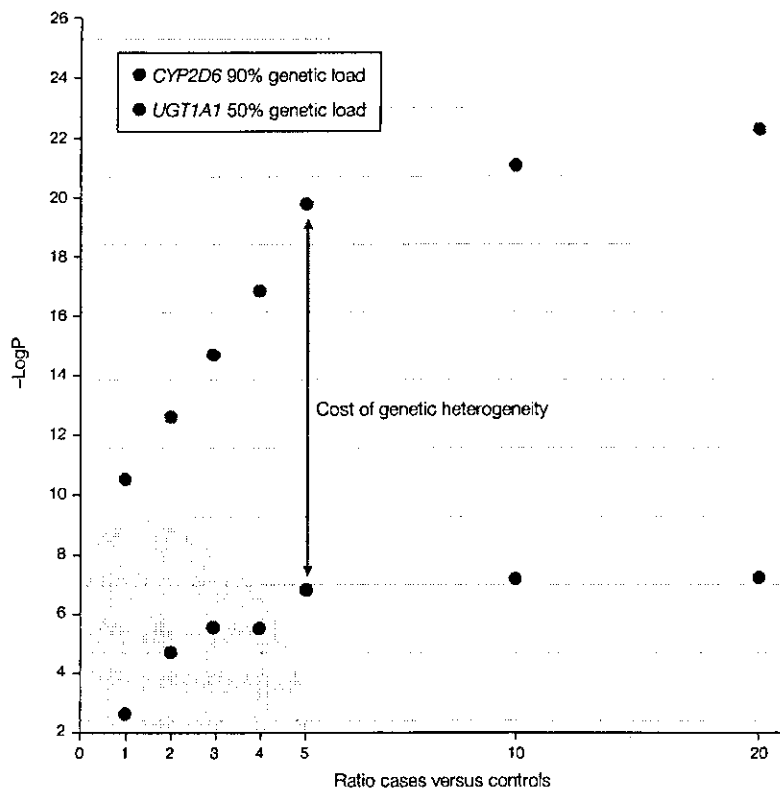


Figure 4 | **Association as a function of the genetic load of the polymorphism.** The tranilast case data provide a comparative estimate of the number of controls that would be required if 41 hyperbilirubinaemia adverse-event (AE) patients were tested (see also TABLE 1) compared with a more penetrant trait. The minus log of P-value is plotted against the ratio of controls to cases for two genes: *CYP2D6* (a metabolizing enzyme polymorphism that is expressed as a dominant variant) and *UGT1A* (with a variant that is expressed as a recessive allele; see FIG. 2)[10,67]. At a ratio of five controls to each case, these data demonstrate that the association of these variants to the respective AEs was highly significant, although the difference was −log 7 for tranilast with 50% GENETIC LOAD, and −log 22 for *CYP2D6* with >90% genetic load. Note that for the diagnosis of dominant expressed traits, such as *CYP2D6*, controls are not generally necessary.

no longer necessary for diagnosis because the sensitivity and specificity of the measurement of the disease allele is so high. Consider the 147 tranilast hyperbilirubinaemic patients illustrated in FIG. 2 who were compared with 1,054 people who did not receive tranilast. If population controls were used, rather than treated controls from a large trial, a few control individuals would be predicted to carry the 7-7 genotype. The fact that virtually all of the 147 patients with AEs carried at least one 7-repeat allele, and that a much larger proportion of the controls did not, would make the association highly statistically significant[10]. This is analogous to the situation that currently defines clinical diagnostics. It is applicable to rare diseases or to multi-locus diseases, for which the practical diagnostic utility is based on prior diagnostic sensitivity and specificity data, and controls are not repeated for each new patient. If a standardized PGx panel is implemented, the data from several anonymous ethnic or regional control groups could be made available for standardized testing and regulatory validation.

GENETIC LOAD
The degree to which a given trait can be attributed to genetic variation by the proportion of cases that carry the allele.

Conclusion, or the beginning

Prospective efficacy PGx and safety PGx are frequently viewed in the scientific and lay literature as future technologies. However, efficacy PGx has already entered into drug development. A rapid transition from proof-of-concept investigations to regularly applied pipeline technology is in progress. The ability to segment patients by therapeutic efficacy quickly and in real time during early-phase development provides the opportunity to progress multiple compounds that can treat overlapping groups of patients with the same disorder. So, there might be three or four molecules developed to cover most patients specifically, rather than the present hit-or-miss approach of recurrent treatment with a series of medicines. 'The right medicine for the right patient at the right dose' does not translate to a personally designed medicine for each patient, but to an individualized guide to proper therapeutic options among predictable segments of the patient population. The initial applications are already in use for segmenting cancers for therapy options (BOX 3). The application of candidate-gene screening, or whole-genome screening when necessary, for segmenting patients with a greater or lesser chance of therapeutic effect is currently a pipeline strategy. Nevertheless, there are still sceptics who will not acknowledge that this new strategy has been successful until there are marketed products. It is useful to remember that the period between positive proof of concept for efficacy to marketed drug includes approximately 3–7 years of clinical trials, depending on the length of time to measure clinical end points and regulatory review. The effect of efficacy PGx will first be widely appreciated by the size and breadth of the full-development pipelines, and the decrease in attrition at the proof-of-concept stage in early drug development[42].

Safety PGx carries with it some of the more popular myths. One of these myths is that molecules that were withdrawn from the market place for safety reasons will be resurrected with accompanying tests for the AE. This prospect is unlikely, at least for commercially-driven pharmaceutical companies, as most of these molecules will no longer be patent-protected. Even if other sources of funding become available for resurrecting such drugs, to develop the AE PGx tests required, DNA would need to have been collected from patients with AEs before the drug's withdrawal. Unless a specific drug mechanism constitutes a unique treatment for some serious ailment and can qualify for a risk-management programme, obtaining samples and identifying associated SNP LD panels for AEs would not be possible unless a prospective surveillance programme was initiated[9]. Rarely discussed, but of a much broader concern, is the possibility that off-patent, generic drugs can be made safer — which would require greater accountability from companies that market generic and over-the-counter medications. Home remedies and herbal supplements could also be more closely evaluated for safety.

One ethical concern with respect to prospective surveillance that uses genome-wide SNP-testing panels to recognize LD-associated regions is that these screens might also be used to diagnose susceptibility to complex

©2004 Nature Publishing Group