

# Genome-based biomarkers for adverse drug effects, patient enrichment and prediction of drug response, and their incorporation into clinical trial design

Thomas J White<sup>1†</sup>,  
Andrew G Clark<sup>2</sup> &  
Samuel Broder<sup>3</sup>

<sup>†</sup>Author for correspondence

<sup>1</sup>Celera Genomics,  
1401 Harbor Bay Parkway,  
Alameda, CA 94502, USA  
Tel.: +1 510 749 4200;  
Fax: +1 510 749 6248;  
E-mail: Thomas.white  
@celeradiagnostics.com

<sup>2</sup>Cornell University,  
Molecular Biology and  
Genetics, Ithaca,  
NY 14853, USA

Tel.: +1 607 255 0527;  
Fax: +1 607 255 6249;  
E-mail: ac347@cornell.edu

<sup>3</sup>Celera Genomics, Inc. &  
Celera Diagnostics,  
45 W. Gude Drive, Rockville,  
MD 20850, USA  
Tel.: +1 240 453 3300;  
E-mail: Samuel.broder  
@celera.com

Classic examples of pharmacogenomic biomarkers for drug efficacy include genetic variation in the drug target (including its expression level) and drug metabolizing enzymes (DMEs). Recent US FDA approvals of tests for cytochrome P450 2D6/2C9 and uridine diphosphate glucuronosyltransferase (UGT)1A1 have given regulatory endorsement to biomarkers that can improve drug safety by identifying individuals at risk for drug toxicity. Markers that predict risk for disease can identify patients who will have a greater than average benefit from therapy. This creates a new opportunity to enrich clinical trials with patients who are likely to have more events and to achieve earlier drug approval. Markers that predict for risk of cardiovascular, thrombotic and liver diseases may also identify a subset of individuals at substantially elevated risk for adverse drug effects. The adaptive clinical trial design provides a mechanism for incorporating genomic information during clinical trials, while providing sufficient time for diagnostic product development and co-registration with a new drug application.

Genetic biomarkers for drug efficacy and adverse effects have traditionally derived from two sources, genetic variation in the target on which the drug acts, and variation in enzymes that metabolize the drug [1,2]. Well-known examples of the former are the association of a mutation in the 5-hydroxytryptamine receptor and response to clozapine [3] and HER2/neu overexpression and response to trastuzumab (Herceptin<sup>®</sup>) [4]. Variation in drug-metabolizing enzymes (DMEs), such as the cytochrome P450 (CYP) family, include the CYP2D6 variants associated with slow or rapid metabolism of codeine, tricyclic antidepressants and  $\beta$ -blockers [5], and dihydropyrimidine dehydrogenase (DPD) deficiency and increased risk of toxicity in cancer patients receiving 5-fluorouracil [6]. The US FDA has recently cleared the first molecular diagnostic products for CYP2D6 and CYP2C9 variants [101] and has also required Pfizer (NY, USA), the manufacturer of Camptosar<sup>®</sup> (irinotecan), to change its label to recommend testing of patients for the promoter mutation in uridine glucuronosyltransferase (UGT)1A1 prior to prescribing the drug [102]. Despite the general acceptance of the clinical utility of these biomarkers by the US FDA and medical community, some pharmaceutical companies consider them to be isolated cases that, except for oncology, have little relevance to most drugs in development. This has resulted in a very cautious approach to pharmaceutical–diagnostic (Rx–Dx) co-development, where most

biomarker identification is performed by in-house groups. The most common pharma strategy is to consider licensing diagnostic labs or partnering with product manufacturers in late Phase III clinical trials and only if a pharmacogenomic biomarker might be required for drug approval (Box 1). This approach underestimates the complexity of diagnostic product development and registration and may miss opportunities for earlier drug approval. However, the landscape may be changing with a new category of biomarkers for drug efficacy arising from a subset of genetic variants associated with disease predisposition. Furthermore, the initiatives taken by the US FDA via the guideline on voluntary genomic data submissions (VGDS) and the concept of an adaptive clinical trial design (ACTD) will create new requirements and incentives for drug and diagnostic co-development [103–106].

## Improve drug efficacy or reduce adverse effects?

Most drugs are effective in only 30–60% of the people who take them, which would seem to present a burgeoning opportunity for targeted medicine. As a new molecular entity moves through the stages of clinical trials it might have an efficacy/toxicity profile or therapeutic index that falls in any of the three clusters shown in Figure 1. Highly effective and nontoxic compounds would not benefit from pharmacogenomic markers, nor would those that are highly

**Keywords:** adaptive clinical trial design, adverse drug events, biomarkers, drug metabolizing enzymes

future  
medicine

**Box 1. Current pharma strategies for co-development of pharmaceuticals (Rx) and diagnostics (Dx).**

- Watchful waiting: except for cancer, biomarkers have not been shown to be useful in drug development for any major disease.
- Develop biomarker technology for in-house use and hope a diagnostic test won't be necessary (or will be ready if needed)
- Out-license biomarkers to clinical reference laboratories to offer as a diagnostic service.
- Partner with a diagnostic product manufacturer
  - Dx must have commercial value independent of drug; no Rx royalty.
  - Dx company must pay its own costs of test development.
  - Access to pharma samples provides enough value to Dx partner.

toxic and poorly effective. However, the drugs clustered in the center of the figure might be significantly improved if a group of individuals at elevated risk for serious side effects could be identified with biomarkers. In addition, the efficacy of less effective, but relatively safe compounds might be improved. In concert with its legislative mandate to ensure drug safety, the first molecular diagnostic tests to receive US FDA clearance from the Office of *In Vitro* Devices (OIVD) have all focused on variants of DMEs that could be used in conjunction with other medical information to aid physicians in selecting a safe and effective dose of drugs that are metabolized by these enzymes. In the review process, the agency did not require a prospective clinical trial of the device, but allowed the manufacturers to reference peer-reviewed articles in the scientific and medical literature to demonstrate clinical utility for the claimed indications for use. This initiative was welcome news to diagnostic companies, since the duration and cost of such prospective trials for these biomarkers would have been prohibitive. Nonetheless, it remains to be seen whether US FDA-cleared diagnostic kits for these relatively simple biomarkers will be profitable for the manufacturers, since many clinical laboratories already offer similar tests developed and validated in-house, i.e., so-called 'homebrew' assays. Another barrier to developing tests for similar biomarkers of drug toxicity, such as thiopurine methyltransferase (TPMT), which is useful in predicting potential toxicity of certain anticancer drugs, is difficulty in obtaining access to the intellectual property rights at a reasonable cost.

#### Latent genomic stratification: an example from the Women's Health Initiative

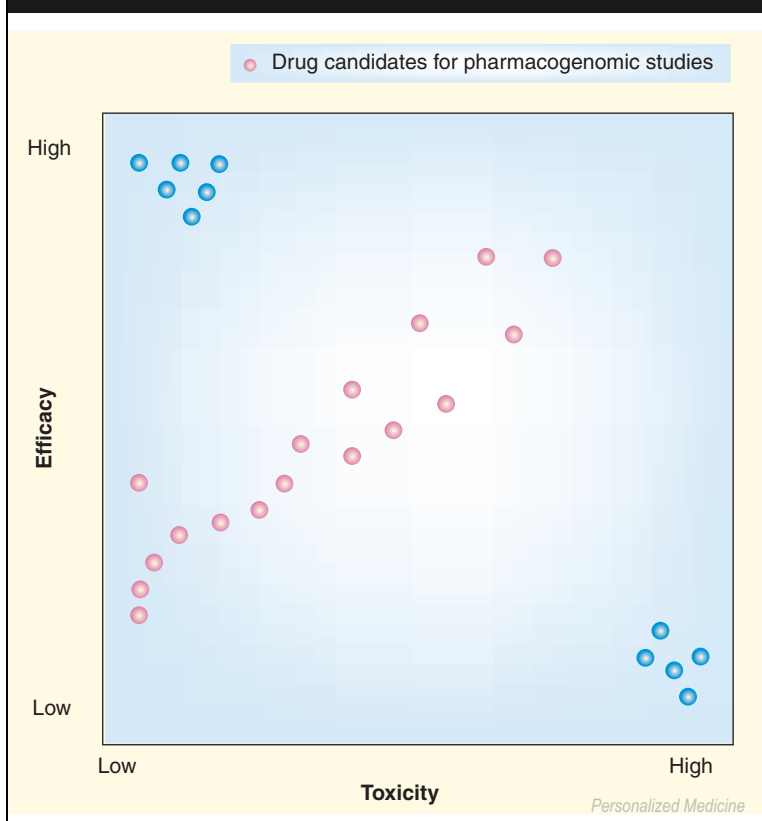
Epidemiologists are comfortable with the idea that a population sample may be stratified in multiple ways, and that the different strata may face different risks. For example, it is well estab-

lished that smokers have inflated risk of cardiovascular disease. Interactions of this sort result in a different set of causal factors for smokers and nonsmokers, and it seems almost inevitable that there will be additional unknown and unmeasured genetic and environmental variables that may interact with the measured variable to influence risk. These issues can be addressed by appropriate stratification of populations based on known variables (co-variables). The challenge arises when we realize that there must inevitably be unobserved causal variables in the form of single nucleotide polymorphisms (SNPs), which are so numerous that each patient is unique.

Following a highly successful tradition of large-scale clinical trials, the National Institutes of Health initiated the Women's Health Initiative (WHI) to assess the role of hormone replacement therapy (HRT) on cardiovascular and other risks. The hormone replacement arm of the trial involved 16,608 women in a randomized trial comparing the effects of conjugated equine estrogens plus medroxyprogesterone acetate (Prempro™) to a placebo, on the incidence of cardiovascular disease. Another arm of the study included 10,739 postmenopausal women divided into a group receiving conjugated equine estrogen (unopposed estrogen) and a placebo group.

In July 2002, the Prempro study was terminated early, as women in the estrogen/progestin treatment had an increased risk of breast cancer (mean hazard ratio = 1.26) and an elevated overall measure of harm, called the global index [7]. Other risks were also identified (Table 1): the estrogen/progestin arm showed increased coronary heart disease (mean hazard ratio = 1.29), stroke (mean hazard ratio = 1.41), and venous thromboembolism (mean hazard ratio = 2.11). However, the estrogen/progestin arm revealed some positive outcomes as well, including decreases in colorectal cancer (mean hazard ratio = 0.63) and total fractures (mean hazard ratio = 0.76). It was concluded that the potential for adverse outcomes outweighed the possible benefits for reduction in colorectal cancer and fractures. In March 2004, the unopposed estrogen study was also terminated before the planned date, and a report was published indicating a significantly inflated risk of stroke (mean hazard ratio = 1.39), but other clinical events and the global index showed no inflated risk from unopposed estrogen [8].

Despite the massive scale of the WHI study, the number of observed cases of deep vein thrombosis (DVT) was only 167, raising the question

**Figure 1. Balance of efficacy and toxicity.**

of whether there might be a relatively small number of individuals who carry a specific confounding risk factor. One plausible factor that could be driving the elevated DVT risk is a major polymorphism in the gene for Factor V Leiden (Arg506Gln) or risk variants in other genes yet to be discovered. This variant allele has a frequency of 0.05 in Europeans, and has been shown to result in greater resistance to Factor V degradation. The resulting hypercoagulation results in a strong association with risk of DVT [9]. Of special interest to this discussion, there is a significant interaction between Factor V Leiden and oral contraceptive (OC) use on the risk of venous thromboembolism in women [10]. Thus, the rate of DVT was 0.8 per 10,000 person-years in a population not endowed with Factor V Leiden and without OC use. OC use, but no Factor V Leiden, yielded a rate of 3.0; no OCs, but the presence of Factor V Leiden, yielded a rate of 5.7; and the combination of OC use and presence of Factor V Leiden produced a risk of approximately 29 cases per 10,000 person-years. Similarly, in another study, there was more than a 36-fold increase in risk of DVT among Factor V Leiden, OC takers, compared with non-Factor V, non-OC takers [11].

This presents us with the question – could the apparent inflated risk estimated from the whole sample in the WHI study be entirely caused by the interaction with Factor V Leiden and a few other genetic risk factors? If the WHI sample is considered as a mixture, with 90% having the same risk as the control and the 10% with at least one copy of Factor V Leiden having all the elevated risk, the relative risk of DVT in that 10% subsample would have to be 12.47 in order to produce the observed number of cases (Table 1). This is in fact a relative risk that is comfortably within the observed 95% confidence interval of the risk for combined Factor V Leiden plus OC use. This means that it is conceivable that HRT has no inflated DVT risk unless one also bears the Factor V Leiden allele. Recently, Cushman and colleagues tested several known risk variants for DVT in samples from the WHI [12]. They found that the presence of the Factor V Leiden mutation in the HRT arm conferred a 6.7-fold increased risk compared with the placebo arm of the trial. Variants in prothrombin, 5,10-methylenetetrahydrofolate reductase (*MTHFR*) and three other genes, had no modifying effect on the risk of DVT in women using HRT. Although Factor V Leiden variations account for approximately half of the effect predicted to be due to latent stratification, known risk factors for DVT are thought to account for only half of the genetic risk for this disease, so it is likely that other unknown risk variants may contribute to the full effect.

All population samples have hidden strata, and the challenge we face as we enter the era of massive genotyping and identification of markers associated with disease risk is when to be satisfied that the associations we observed are not spuriously caused by associations with some unobserved variable. The problem is made more critical when considering treatments that may be genuinely beneficial to the large majority of people, but whose net effect appears deleterious due to a high-risk minority. In the case of the WHI study, one would certainly like to know whether variants in other genes associated with thrombosis explain the inflated DVT risk among the Prempro group. While the benefits of HRT might still be available to women who have no elevated risk for these adverse events, certainly the women who may be at a much higher (> tenfold) risk than average should be aware of this when considering whether to continue or begin HRT. These questions could be readily investigated by genotyping the WHI samples with additional novel genetic variants associated with risk for DVT and the other adverse outcomes.

**Table 1. Hazard ratios of hypothetical substrata.**

	No stratification	10% high risk	5% high risk	3% high risk
Coronary heart disease	1.29	3.80	6.61	9.42
Stroke	1.41	5.23	9.46	13.71
Venous thromboembolism	2.11	12.47	23.93	35.44
Invasive breast cancer	1.26	3.75	6.50	9.27

*Figures in the table report the hazard ratios of substrata of various sizes needed to drive the hazard ratio of the whole population to the observed level. Assume that the hormone replacement group is stratified into two subgroups, one whose risk is the same as that of the placebo, and a second stratum that shows elevated risk only when given Prempro. The column headed '10% high risk' gives the hazard ratio of the high-risk group assuming they constitute 10% of the treatment group. Similarly the columns headed 5 and 3% assume the respective proportions of the high-risk group. In all cases we assume that the placebo group is essentially homogeneous, implying that the at-risk genotype(s) only face the high risk when given hormone replacement therapy.*

### Genetic risk factors for common complex diseases

Although the draft sequence of the human genome was published 5 years ago [13,14], only a small number of associations of genetic variation with risk for common complex diseases have been published and repeatedly confirmed. The draft sequences involved only a few individuals, and therefore the genome-wide studies that could be conducted were mostly limited to testing common variants. To address the paucity of potential disease-causing polymorphisms, the National Human Genome Research Institute embarked on a comprehensive survey of single nucleotide polymorphisms (SNPs) in regions of high linkage disequilibrium in the genome. This International HapMap project focused on completing a haplotype map to reduce the cost of future disease association studies [15]. Since only half of the genome appears to contain these conserved regions, the remainder will likely require an alternate approach. Scientists at Perlegen (CA, USA) also characterized whole-genome patterns of common variation by genotyping 1,586,383 SNPs in 71 Americans of European, African and Han Chinese ancestry [16]. To enrich the database with less frequent, but putatively functional SNPs, Celera Genomics (MD, USA) resequenced the coding regions of approximately 24,000 known and predicted genes from 20 Caucasians, 19 African-Americans and a chimpanzee [17]. Missense, nonsense, and regulatory SNPs were then used in a gene-centric approach to genotype multiple large case-control sample collections in Alzheimer's, cardiovascular, autoimmune and other diseases [18–20].

With this catalog of human variation and rapid cost-effective methods of analysis, it is finally possible to conduct very large-scale gene-disease association studies by genotyping

multiple, appropriately consented sample collections where anonymized, detailed clinical phenotype information is available. The absolute necessity for confirming any initial associations with multiple replication studies has been extensively reviewed [21,22]. Some recent examples of confirmed associations with genetic variation are: complement factor H with age-related macular degeneration [23], protein tyrosine phosphatase N22 with juvenile diabetes, rheumatoid arthritis and other autoimmune diseases [18,24], transcription factor 7-like 2 gene with type 2 diabetes [25] and paraoxonase 1 with stroke [26].

### Identification of new disease indications for existing drugs

The identification of additional disease indications beyond that in an initial drug registration is a major contributor to the overall medical and commercial success of new therapies. The decision on the next disease indications for clinical trials often rests on information regarding disease phenotype rather than on knowledge of an underlying disease mechanism. In addition to identifying potential new targets for drug discovery, gene-disease associations may discover possible new disease indications for existing drugs. For example, variation in the 5-lipoxygenase-activating protein (FLAP) was found to be associated with risk for myocardial infarction in a genome-wide linkage study conducted in Iceland [27]. Subsequently, deCODE Genetics (Reykjavik, Iceland) licensed a *FLAP* inhibitor (DG-031) that had previously been shown to be safe and well tolerated in clinical trials for asthma. In a clinical trial of DG-031 in Icelandic myocardial infarction patients who carried a risk allele in the *FLAP* gene, production of the biomarkers myeloperoxidase and leukotriene B<sub>4</sub>, a potent chemokine mediator of arterial inflammation, were

significantly reduced [28]. Similarly, Perlegen Sciences recently licensed a peroxisome proliferator-activated receptor (PPAR) agonist (MCC-555), which they hope to personalize by using genomic information to guide the treatment of diabetes and other metabolic disorders toward those most likely to benefit [107].

### A source of new biomarkers for enrichment and efficacy

Overexpression of the *HER2/neu* gene is both prognostic for a poor outcome in women with breast cancer and predictive of response to treatment with trastuzumab. Patients who are carriers of the apolipoprotein ε4 allele are at elevated risk for heart disease and also show an increased survival benefit from statin therapy compared with noncarriers [29]. These examples, along with the FLAP study described above, may portend an increase in the discovery of new biomarkers for drug efficacy. That is, as new genetic variants that predict risk for disease are identified and confirmed from multiple case-control studies, they may constitute a new category of variants worth testing in prospective, randomized, placebo-controlled clinical trials of new drugs. Genotyping of 1000 functional SNPs in the Cholesterol and Recurrent Events Study (CARE) [30] identified variants in several genes that predicted risk for recurrent myocardial infarction (RMI) in the placebo arm of this secondary prevention trial [Jakoubova O. PERS. COMMUN.]. Carriers of a risk allele in the myeloid immunoglobulin A Fc receptor gene (*FCAR*), an important mediator of inflammation, had approximately a 1.5-fold increased risk compared with noncarriers. This excess risk was ameliorated by pravastatin in the treatment arm of the trial; in fact, those at elevated risk had approximately a twofold greater improvement in survival over 5 years compared with the average or unstratified survival benefit (Figure 2). Furthermore, this efficacy would have reached statistical significance 9 months earlier than was observed in the CARE study. These results were subsequently confirmed in a primary myocardial infarction prevention study of pravastatin. Thus, new biomarkers for drug efficacy may be discovered in a subset of the genetic variants that predict risk for disease. Risk variants for cardiovascular, thrombotic and liver diseases may also predict those individuals at elevated risk for certain adverse drug events (ADEs), for example, those at increased risk for myocardial infarction and stroke as a result of treatment with cyclooxygenase-2 inhibitors for rheumatoid arthritis [31].

### Impact on clinical trials

How might biomarkers for enrichment and efficacy be incorporated into clinical trials of new drugs? If known or probable valid biomarkers are used as an inclusion criterion to stratify a Phase II or III trial, the US FDA has indicated that a diagnostic test for the biomarkers must be available for approval of the drug [103]. In addition, a trial must continue to assess ADEs in the group predicted to be nonresponders in order to provide information pertaining to inappropriate prescriptions of the drug.

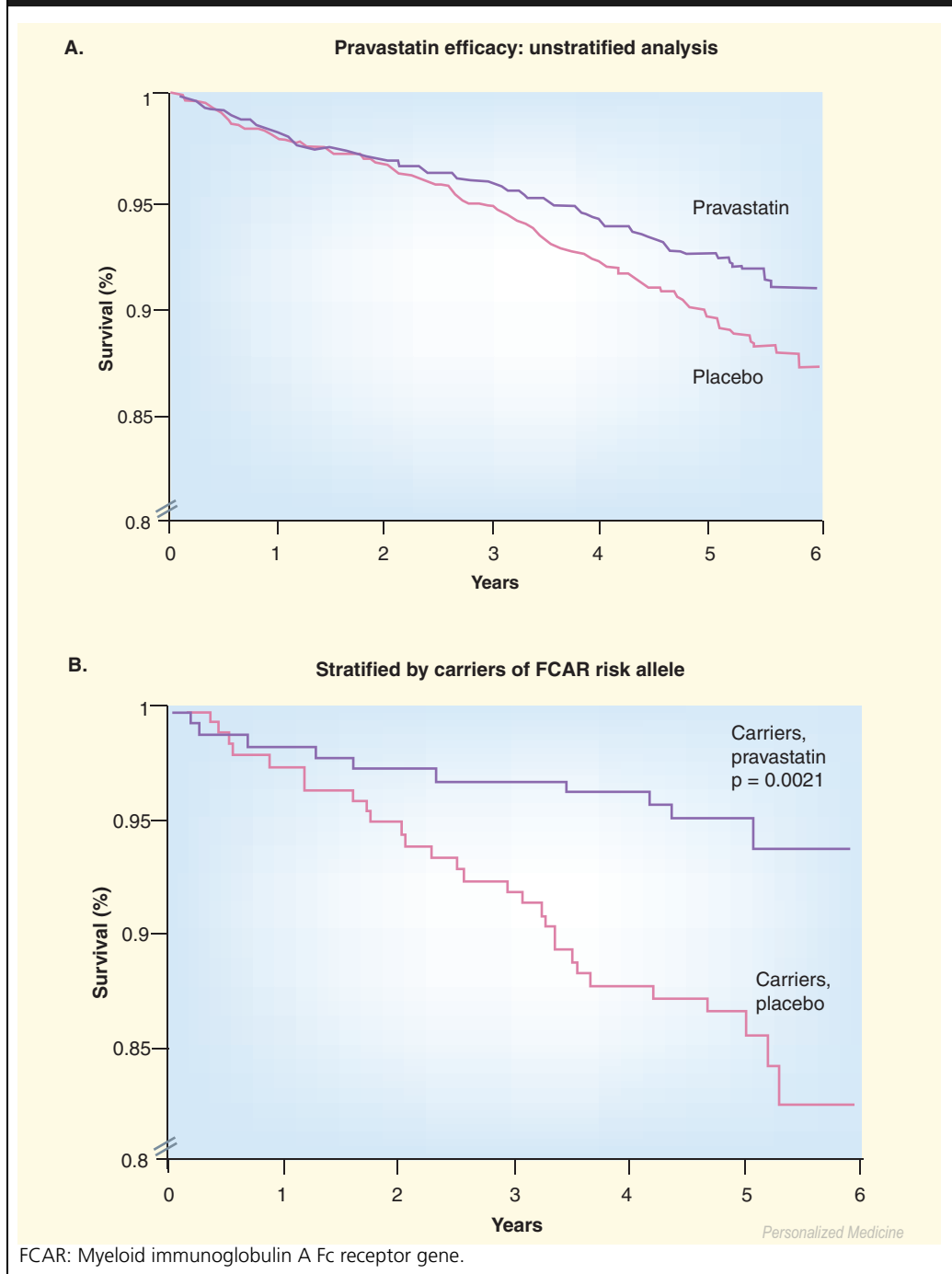
If an unstratified efficacy trial is conducted and biomarkers are measured in parallel, Temple and Simon have recently proposed an innovative Adaptive Clinical Trial design [105,106] that could aid in early approval if the drug is only efficacious in a subgroup (Figure 3). In this model, the results are analyzed after half of the patients have reached the endpoint. If efficacy in all patients is observed at a p-value of less than 0.04, the drug is approvable and no biomarkers are necessary. However, if the drug is only efficacious in the subgroup defined by the biomarkers, the second half of the trial is continued, with an approvable drug if efficacy is observed at a p-value of less than 0.01 in the subgroup. A critical requirement for this design is *a priori* planning of the data analysis. Note that the accelerated approval in the subgroup does not preclude the ultimate approval of the drug for the entire population depending on its overall efficacy.

Finally, if pharmacogenomic testing identifies a subgroup in which the drug is efficacious after completion of the trial, an additional clinical trial would be required for validation and drug approval. Comparable guidance documents have not yet been issued by European and other regulatory authorities.

### Impact of voluntary genomic data submissions guidelines

Table 2 summarizes the FDA Guidelines on VGDS, where known valid biomarkers are from an accepted test system with widespread acceptance of results [103]. Although the agency has not yet published a list of known biomarkers, it would likely include those mentioned above: *DPD*, *TPMT*, *UGT1A1*, *CYP2D6*, *CYP2C9*, *CYP2C19*, the *bcr-abl* translocation, Factor V Leiden, and *APOE4*. Similarly, no list of probable valid biomarkers, defined as ‘from an accepted test system, but not conclusive or requires replication’, has been published, but it might include C-reactive protein, FLAP, paraoxonase 1 (PON1)

**Figure 2. Recurrent myocardial infarction-free survival in the Cholesterol and Recurrent Events Study (CARE).**

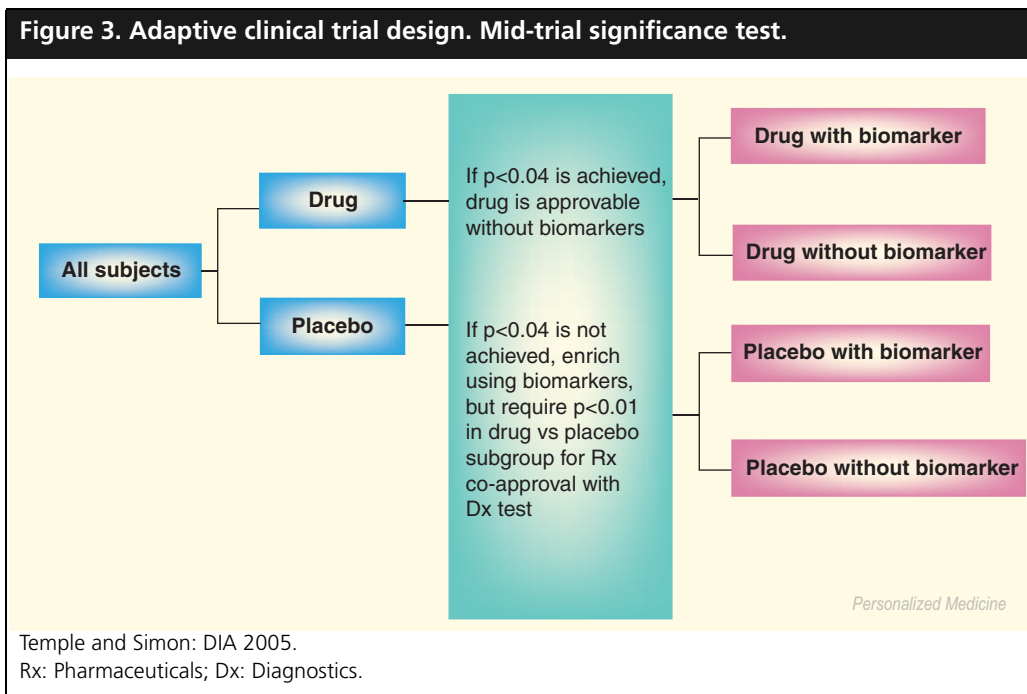


and cholesterol esterase transfer protein, among others ([2] see Table 1). The point is that the number of known and probable valid biomarkers will increase dramatically over the next few years as more novel gene–disease associations are discovered, published and confirmed. With the US FDA recommendation that data on such biomarkers must be submitted, and the requirement for diagnostic tests to be broadly available if they are

to be used for treatment selection, a substantial economic opportunity for new diagnostic products will be created.

**Outlook**

The US FDA’s initiatives on genomic data submissions will change the current pharma strategy on personalized medicine. Companies that are close to submitting new drug applications



(NDAs) on drugs that involve known valid biomarkers must submit such data for drug registration. For example, companies hoping to register second generation or mixed kinase inhibitors competitive with imatinib mesylate (Gleevec®) for the treatment of chronic myeloid leukemia will probably have to submit quantitative *BCR-ABL* data as well as DNA sequence information on the *ABL* gene if they hope to claim a use in Gleevec-resistant patients. Although no diagnostic test for the *BCR-ABL* translocation has been cleared by the US FDA, there will now be an incentive for diagnostic manufacturers to submit a 510(k) application for such a device, as well as for other known valid biomarkers.

Genome-wide genotyping and expression studies are now yielding novel confirmed associations with common complex diseases. These studies will be a potential new source of biomarkers for the enrichment of drug clinical trials

and for identifying patients who may have a greater than average benefit from therapy or who have a substantially elevated risk for certain adverse effects. As these biomarkers are published in the peer-reviewed literature, they will evolve from research biomarkers to probable valid biomarkers to known valid biomarkers with a significant impact on new drug development and drug–diagnostic co-registration. The adaptive clinical trial design provides a strategy for incorporating such genetic information in parallel with unstratified prospective drug trials and a mechanism for obtaining earlier drug approval in a subset of patients. Ultimately, there will be a vast opportunity for new diagnostic tests resulting from the convergence of biomarkers for disease risk and drug efficacy.

**Acknowledgments**

*The authors wish to thank John Sninsky for helpful comments on the manuscript and Eric Lai for the concept of Figure 1.*

**Table 2. FDA guidance on PGx – new driver for diagnostic biomarkers.**

	IND	New NDA	Previous NDA
<b>From an accepted test system with widespread acceptance of results</b>			
Known valid biomarker	Must be submitted	Must be submitted	Must be submitted
<b>From an accepted test system but not conclusive or requires replication</b>			
Probable valid biomarker	Does not need to be submitted, but welcomes submission	Recommends submission	Must be submitted
<b>From a broad screen of RNA or DNA markers</b>			
Exploratory PGx data	Welcomes voluntary submission	Recommends submission	Welcomes voluntary submission

*IND: Investigational new drug; NDA: New drug application; PGx: Pharmacogenetic.*

## Highlights

- Testing for known valid biomarkers is gaining broad acceptance. For example, the US FDA's guidance document on voluntary genomic data submission requires that data be submitted for investigational new drugs, novel new drug applications (NDAs) and previous NDAs. It is hoped this will increase the opportunity for new diagnostic and pharmacogenomic tests.
- Novel disease associations are being discovered from whole-genome scans and confirmed with multiple replications. Genetic variants that predict risk for disease may identify patients who will benefit most from therapy. They may also serve as biomarkers for predicting some adverse events.
- Adaptive clinical trial design provides a strategy to incorporate biomarkers for efficacy in clinical trials and the possibility of earlier drug approval.
- Retrospective genetic analysis of prospective randomized drug trials may be sufficient to prove pharmacogenomic utility of new biomarkers.

## Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Evans WE, McLeod HL: Pharmacogenomics – drug disposition, drug targets, and side effects. *N. Engl. J. Med.* 348(6), 538–549 (2003).
2. Goldstein DB, Tate SK, Sisodiya SM: Pharmacogenomics goes genomic. *Nature Rev. Genet.* 4, 937–947 (2003).
- An excellent overall review, especially on variants associated with drug response.
3. Arranz MJ, Collier D, Kerwin RW: Pharmacogenetics for the individualization of psychiatric treatment. *Am. J. Pharmacogenomics* 1(1), 3–10 (2001).
4. Fornier M, Riso M, Van Poznak C, Seidman A: HER2 testing and correlation with efficacy of trastuzumab therapy. *Oncology* 16(10), 1340–1352 (2002).
5. Roberts R, Sullivan P, Joyce P, Kennedy MA: Rapid and comprehensive determination of cytochrome P450 *CYP2D6* poor metabolizer genotypes by multiplex polymerase chain reaction. *Hum. Mutat.* 16(1), 77–85 (2000).
6. Wei X, McLeod HL, McMurrrough J, Gonzalez FJ, Fernandez-Salguero P: Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. *J. Clin. Investigation* 98(3), 610–615 (1996).
7. Rossouw JE, Anderson GL, Prentice RL *et al.*: Risks and benefits of estrogen plus Progestin in healthy postmenopausal women: Principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 288, 321–333 (2002).
8. Anderson GL, Limacher M, Assaf AR *et al.*: Effects of conjugated equine estrogen in postmenopausal women with hysterectomy. *JAMA* 291, 1701–1712 (2004).
9. Price DT, Ridker RM: Factor V Leiden mutation and the risk for thromboembolic disease: A clinical perspective. *Annals Int. Med.* 127, 895–903 (1997).
10. Vandenbroucke JP, Rosing J, Bloemenkamp KWM *et al.*: Oral contraceptives and the risk of venous thrombosis. *NEJM* 344(20), 1527–1535 (2001).
- A good example of latent genetic stratification for risk of deep vein thrombosis (DVT) in women using oral contraceptives.
11. Bloemenkamp KW, Rosendaal FR, Helmerhorst FM *et al.*: Enhancement by Factor V Leiden mutation of risk of deep-vein thrombosis associated with oral contraceptives containing a third-generation progestagen. *Lancet* 346, 1593–1596 (1995).
12. Cushman M, Kuller LH, Prentice R *et al.*: Estrogen plus progestin and risk of venous thrombosis. *JAMA* 292(13), 1573–1580 (2004).
13. Venter JC, Adams MD, Myers EW *et al.*: The sequence of the human genome. *Science* 291, 1304–1351 (2001).
14. Lander ES, Linton L, Birren B *et al.*: Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).
15. Goldstein DB, Cavalleri GL: Understanding human diversity. *Nature* 437, 1241–1242 (2005).
16. Hinds DA, Stuve LL, Nilsen GB *et al.*: Whole-genome patterns of common DNA variation in three human populations. *Science* 307, 1072–1079 (2005).
17. Nielsen R, Bustamante C, Clark AG *et al.*: A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3(6), 1–10 (2005).
18. Botstein D, Risch N: Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genet.* 33, 228–237 (2003).
- A key paper on using a gene-centric, functional single nucleotide polymorphism (SNP) approach to genome-wide disease association studies.
19. Begovich AB, Carlton VEH, Honigbert LA *et al.*: A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (*PTPN22*) is associated with rheumatoid arthritis. *Am. J. Hum. Genet.* 75, 330–337 (2004).
20. Shiffman D, Ellis SG, Rowland CM *et al.*: Identification of four gene variants associated with myocardial infarction. *Am. J. Hum. Genet.* 77, 596–605 (2005).
21. Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG: Replication validity of genetic association studies. *Nature Genet.* 29, 306–309 (2001).
22. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K: A comprehensive review of genetic association studies. *Genet. Med.* 4(2), 45–61 (2002).
- Essential reading on the importance of replicating initial association results in multiple independent studies.
23. Klein RJ, Zeiss C, Chew EY *et al.*: Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385–389 (2005).
24. Bottini N, Musumeci L, Alonso A *et al.*: A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nature Genet.* 36(4), 337–338 (2004).
25. Grant SFA, Thorleifsson G, Reynisdottir I *et al.*: Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nature Genet.* 38(3), 320–323 (2006).
26. Ranade K, Kirchgessner TG, Iakoubova OA *et al.*: Evaluation of the paraoxonases as candidate genes for stroke: Gln192Arg polymorphism in the paraoxonase 1 gene is associated with increased risk of stroke. *Stroke* 36, 2346–2350 (2005).
27. Helgadottir A, Manolescu A, Thorleifsson G *et al.*: The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nature Genet.* 36(3), 233–239 (2004).



28. Hakonarson H, Thorvaldsson S, Helgadóttir A *et al.*: Effects of a 5-lipoxygenase-activating protein inhibitor on biomarkers associated with risk of myocardial infarction: a randomized trial. *JAMA* 293(18), 2245–2256 (2005).
29. Gerdes LU, Gerdes C, Kervinen K *et al.*: The apolipoprotein  $\epsilon 4$  allele determines prognosis and the effect on prognosis of simvastatin in survivors of myocardial infarction: A substudy of the Scandinavian simvastatin survival study. *Circulation* 101, 1366–1371 (2000).
- **An example of a genetic variant for risk of disease that also predicts drug efficacy.**
30. Sacks FM, Pfeffer MA, Moyer LA *et al.*: The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. *N. Engl. J. Med.* 335(14), 1001–1009 (1996).
31. Topol EJ: Arthritis medicines and cardiovascular events – “House of coxibs.” *JAMA* 293(3), 366–368 (2005).
- Websites**
101. FDA draft guidance document for industry and FDA staff: Class II special controls guidance document: drug metabolizing enzyme genotyping system. [www.fda.gov/cdrh/oivd/guidance/1551.html](http://www.fda.gov/cdrh/oivd/guidance/1551.html)
102. FDA approval of a pharmacogenetic test related to Irinotecan (Camptosar ) toxicity [www.fda.gov/medwatch/SAFETY/2005/Jun\\_PI/Camptosar\\_PI.pdf](http://www.fda.gov/medwatch/SAFETY/2005/Jun_PI/Camptosar_PI.pdf).
103. FDA draft guidance document for industry and FDA staff: pharmacogenomic data submissions. [www.fda.gov/cder/guidance/6400fnl.pdf](http://www.fda.gov/cder/guidance/6400fnl.pdf)
- **Essential reading on the US FDA’s thinking on how pharmacogenomic data will be reviewed in new drug submissions.**
104. FDA draft guidance document for industry and FDA staff: Pharmacogenomic and genetic tests for heritable markers. [www.fda.gov/cdrh/oivd/guidance/1549.html](http://www.fda.gov/cdrh/oivd/guidance/1549.html)
105. R Temple: A roadmap for developing and validating genomic classifiers for treatment selection. DIA Genomic Biomarkers Conference, Oct 7, 2005. [www.fda.gov/cder/genomics/presentations\\_20051006/051007\\_10\\_Temple.pdf](http://www.fda.gov/cder/genomics/presentations_20051006/051007_10_Temple.pdf)
- **Very insightful article on biomarkers for classification of disease, enrichment of drug trials and efficacy including the concept of adaptive clinical trial design.**
106. Freidlin B, Simon R: Use of genomic biomarkers in a regulatory environment. [www.fda.gov/cder/genomics/presentations\\_20051006/051007\\_15\\_Simon.pdf](http://www.fda.gov/cder/genomics/presentations_20051006/051007_15_Simon.pdf)
107. Perlegen Sciences licenses late-stage compound from mitsubishi pharma. [www.perlegen.com](http://www.perlegen.com)