ELSEVIER

European Journal of Pharmacology 410 (2000) 107–120

ejp

www.elsevier.nl/locate/ejphar

Extreme discordant phenotype methodology: an intuitive approach to clinical pharmacogenetics

Daniel W. Nebert *

Center for Environmental Genetics, Department of Environmental Health, and Department of Pediatrics & Developmental Biology / Division of Human Genetics, University of Cincinnati Medical Center, P.O. Box 670056, Cincinnati, OH 45267-0056, USA

Accepted 20 October 2000

Abstract

Pharmacogenetics represents "the study of variability in drug response due to heredity". Of the more than six dozen pharmacogenetic differences described in the medical literature, the majority of these variations occurs in drug-metabolizing enzyme genes and others in drug receptor and drug transporter genes, whereas many others have not yet been explained on a molecular basis. It is clear that "drug efficacy" or "drug toxicity" represents a multiplex phenotype, i.e. interaction between the drug (or its metabolites) and the gene products (enzymes, receptors, other targets) of two, and usually many more than two, genes. Because there is a gradient in these phenotypes (efficacy or toxicity), it is extremely important to select patients having the most unequivocal phenotype possible—if one wishes to find the gene(s) responsible for the trait. The method of "extreme discordant phenotype" (EDP) is therefore highly recommended. Using EDP methodology, DNA sequence variants (genotype) can be unconditionally correlated with drug efficacy or toxicity (phenotype). EDP methodology is mathematically intuitive and, in essence, has been used in a number of previous clinical pharmacogenetic studies. This EDP approach should be applicable to virtually any pharmaceutical agent in patient populations.

Keywords: Pharmacogenetic; Pharmacogenomic; Molecular epidemiology; Genotype; Phenotype; DNA sequence variants; Population genetic; Patient-pair studies

1. Introduction

It is well known that when several patients receive the same "recommended" prescribed dose of a particular drug, often the effect is efficacious in most, of little or no effect in others, and seriously toxic to a third subset. The reason for these clinical differences in drug response reflects each individual's underlying genetic predisposition to the way each drug is handled. This research field of studying gene–drug interactions was first named "pharmacogenetics" by Vogel (1959). In the past several years, the term "pharmacogenomics" has been introduced, in which there is an increased emphasis on the development of novel drugs based on new knowledge derived from the Human Genome Project. The broader field of studying all gene–environment interactions was first described as ecogenetics in the 1970s by Brewer.

In this mini-review, there is first a definition of terms. Next, I discuss several of the most prominent factors that one needs to consider in pharmacogenetics/pharmacogenomics research. The importance in identifying a quantitative phenotype (trait) is then emphasized, and the statistical power of using extreme disordant phenotype (EDP) methodology is examined in detail. Finally, how to identify a DNA sequence variant presumably associated with the defined trait (drug efficacy or toxicity), and then how to prove a functional correlation between the genotype and phenotype, are described.

2. Definition of terms

Healthy humans have 23 pairs of chromosomes: 22 autosomal pairs plus the sex chromosomal pair (XX or XY). A gene denotes the location (stretch of DNA) on each of a chromosome pair that encodes a gene product (enzyme, or the protein); the "gene" also includes any regulatory regions upstream or downstream of the coding sequence (Nebert, 2000). A locus indicates the location of

* Tel.: +1-513-558-4347; fax: +1-513-558-3562.
E-mail address: dan.nebert@uc.edu (D.W. Nebert).

0014-2999/00/$ - see front matter
PII: S0014-2999(00)00809-8

a segment of DNA on each of a pair of chromosomes that need not necessarily code for a gene product. Diploid refers to having paired chromosomes; haploid refers to one active chromosome of each pair (e.g. as found in the sperm and egg). In diploid organisms, each gene (or locus) is made up of two alleles, one from the father and one from the mother; the combination of these two alleles is called the genotype. An allele can transmit a dominantly inherited trait (e.g. pigmented skin) or a recessive trait (e.g. blue eyes). Another term for "trait" is phenotype. Homozygous means having two identical alleles at the locus under study; heterozygous means having two different alleles. Proband refers to the first patient in a family tree to have been studied, or identified.

As discussed later in this article, it is not clear whether there is really ever a "simple Mendelian" (single-gene) trait—such as the alleles colored (C) and noncolored (c) described in Mendel's original studies of the garden pea in the 1860s. Allele C (red phenotype) is dominant to allele c (white phenotype). The Hardy–Weinberg distribution [$(p+q)^2 = p^2 + 2pq + q^2$] states that if the allelic frequencies of $C(=p)$ and $c(=q)$ in the population are, say, 0.7 and 0.3, respectively, this would mean that $(0.3)^2 = 9\%$ of the population would have the recessive white trait, i.e. homozygous for the *cc* genotype. Crossing two *Cc* heterozygous garden peas (which have the red phenotype) would give a 1:2:1 ratio of the *CC*:*Cc*:*cc* genotypes, but a 3:1 ratio of the red to white phenotypes. In the snapdragon, inheritance of these two colors is additive (codominant, gene–dose); crossing two *Cc* heterozygous snapdragons would give a 1:2:1 ratio of the *CC*:*Cc*:*cc* genotypes, and a 1:2:1 ratio of the red to pink to white phenotypes.

A phenotype that is dependent on two or more genes is called polygenic, multifactorial, or a multiplex phenotype. Examples of a polygenic trait would include blood pressure, height, weight, or formation of the face during embryonic development, but might also include "efficacy of a drug" or "toxicity of a drug" (at a given dose and/or time). The "1:2:1" ratio for two alleles at one locus becomes "1:4:6:4:1" for two alleles at two loci, and "1:6:15:20:15:6:1" for two alleles at three loci. If one realizes that usually for any given gene, there are many more than two alleles in a human population, and that all human diseases (as well as drug efficacy or toxicity) are traits that are influenced by dozens if not hundreds of genes, one can readily appreciate how quickly and complicated the genotypes and corresponding phenotypes can be. As will be detailed later, the outliers, or individuals at the extreme ends of the spectrum of any phenotype, are the most informative patients to those who wish to dissect the genes responsible for any phenotype.

A polymorphism denotes two or more phenotypes in a population (e.g. poor, extensive, and intermediate metabolizers). Often, the molecular epidemiologist or pharmacogeneticist considers 1% as the "cut-off" in describing a polymorphism. However, this is a completely arbitrary choice, and "a polymorphism" exists *ipso facto*—even if only one case is seen in an entire population. Thus, the most common variant (called "consensus" or "wild-type") occurs with $\leq 99\%$ frequency, meaning that if the locus is biallelic, the rarer allele must occur with a frequency of $\geq 1\%$ in the population. This would mean (by the Hardy–Weinberg distribution) the allelic frequency q for an autosomal recessive trait would be ≥ 0.1.

Polymorphisms arise from DNA sequence variation, of which there are several types: single-nucleotide polymorphisms (SNPs); insertions or deletions of sometimes a single DNA base, but other times insertions/deletions of stretches of hundreds, or thousands, of bases (arising from unequal crossing-over and problems during DNA recombination); and insertions or deletions of repetitive DNA ['variable number of tandem repeats' (VNTRs), microsatellites or 'simple tandem repeats' (STRs), and *Alu* I segments].

3. Factors to consider in a multiplex phenotype

Correlation of a pharmacogenetic trait (phenotype) with a particular DNA sequence variation (genotype) is extremely important, because (*a*) adverse drug reactions (in the genetically "sensitive" patient) might be predicted and therefore avoided and (*b*) new drugs might be designed from such new knowledge. Not only are there multiple alleles at each locus among individuals in any population, and sometimes striking ethnic differences (Kalow and Bertilsson, 1994; Nebert, 1999) in allelic frequencies, but it must be appreciated that drug toxicity is probably always a multiplex phenotype, i.e. a multifactorial trait (factors include dose, length of dosage time, therapeutic window of the drug, competing metabolic pathways, drug–drug interactions, ethnic differences, age, gender, hormonal balance, nutrition, etc.) that involves two, and often many more than two, genes. Several decades ago, the physician was taught that the "recommended prescribed dose" of a drug meant that all patients responded similarly (Fig. 1, top). However, we now realize that the efficacy might be 0.5 mg/kg in one patient and 50 mg/kg in another patient (Fig. 1, bottom)—due to the multiple factors listed above. It is proposed in this mini-review first to tease apart this complexity by identifying groups with extreme responses to a drug (sensitivity versus resistance to drug efficacy or toxicity); this method is called "extreme discordant phenotype" (EDP). Second, once a DNA sequence variant is discovered between patients having the two extreme phenotypes and proof of function is established, it is then important to study the impact of these allelic differences in larger populations and in different ethnic groups.

3.1. Fate of a drug

There are four pharmacokinetic actions subsequent to administration of a drug to a patient: absorption, binding

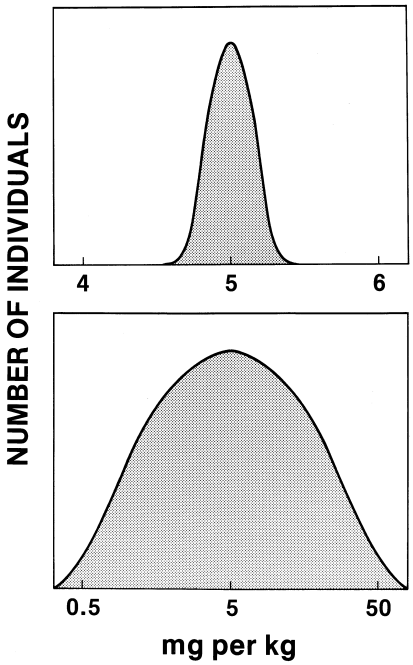

Fig. 1. Hypothetical number of patients achieving "efficacy" at doses of drug between 0.5 and 5.0 mg per kg. This diagram is used simply to illustrate that most drugs do not have nearly the same therapeutic dosage in all patients (top), but rather a 40-fold or larger difference might exist between the "most sensitive" and "most resistant" individuals in any population (bottom).

and distribution, biotransformation, and excretion. At the level of the whole animal (physiology), transporter genes, ion channel genes and receptor genes would be expected to affect absorption, binding and distribution, and/or excretion, whereas metabolism genes would principally affect biotransformation, but also binding and distribution, and/or excretion. For example, a defective ion-channel gene product (membrane protein) might be manifested as delayed excretion, while an amino-acid change in a metabolism-gene product (enzyme) might lead to greatly diminished metabolism in the liver, and so on. At the level of the cell and subcellular organelles (cell biology), almost any category of genes might be considered as candidates to explain a drug's therapeutic or toxic response (Nebert, 1999). Virtually, all pharmaceutical agents are metabolized by the "phase I" (functionalization), followed by the "phase II" (conjugation) drug-metabolizing enzymes—encoded by metabolism genes (Fig. 2). Incoming drugs can be regarded as exogenous signals that are "detected" by the cell —either by means of well characterized endogenous receptors, or by "reception mechanisms" not yet understood; these drugs/signals can displace the naturally occurring endogenous ligands and act either as agonists to up- or down-regulate phase I and phase II drug-metabolizing enzyme genes (Nebert, 1991, 1997a). For example, phenobarbital induces subsets of numerous drug-metabolizing enzyme superfamilies (Nebert and Gonzales, 1987); phenytoin up-regulates its own metabolism (Talas et al., 1999). The movement of drugs and metabolites in and out of the cell by transporters is also dynamic.

The oxygenated reactive intermediates following phase I metabolism, as well as many incoming nonmetabolized drugs (and even conjugated products whose conjugation group has been cleaved), are capable of causing toxicity. In the broadest of terms, toxicity occurs via two mechanisms: (a) perturbation of signal transduction pathways, including oxidative stress that has striking effects on the cell cycle; and (b) covalent binding of reactive endogenous or foreign compound intermediates to cellular proteins and nucleic acids (Nebert, 1999). What might be included as a candidate "environmental susceptibility gene"? A gene in any category (e.g. metabolism, receptor, transporter, ion channel, cell cycle, nucleic acid and protein repair, histocompatibility locus antigen (HLA) protein, other signal transduction pathway genes, etc.) might directly, or indirectly, be involved in drug toxicity (phenotype). In other words, practically any gene in the human genome might be the major gene responsible for, or a modifier gene that contributes to, the multiplex phenotype of clinical drug toxicity.

3.2. Therapeutic window

Drugs that cause adverse reactions usually have a relatively narrow therapeutic window, i.e. the dose causing toxicity is not much greater than the dose needed for efficacy (Leeder, 1998). Most pharmacogenetic studies are successful because they focus on drugs with narrow therapeutic windows. If a drug has a wide therapeutic window, it is unlikely to cause toxicity in a significant portion of

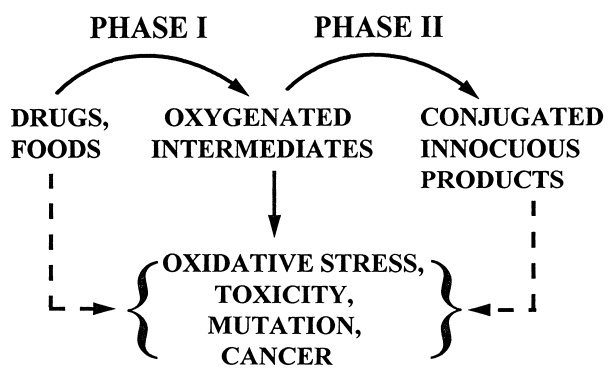

Fig. 2. Diagram of the relationship between the phase I and phase II drug-metabolizing enzymes (DMEs) and cancer, mutation, toxicity and oxidative stress caused by reactive intermediates. Drugs enter the cell by passive diffusion or active transporters. Either the parent drug (without requiring metabolism) or metabolite might reach its clinical target (efficacy), or might cause toxicity via perturbation of the cell cycle (or disturbance of other signal transduction pathways) or covalent binding. Reception mechanisms appear to "detect" the incoming drug, often leading to the up- or down-regulation of certain phase I and phase II drug-metabolizing enzymes. Transporters also assist in moving the parent drug and metabolites in and out of the cell (reproduced with permission from Nebert et al., 1996).

any human population and therefore would be of little concern to public health and the desire to prevent adverse drug reactions. For example, if the dose causing toxicity is 100 times greater than the dose needed to be efficacious, and genetic differences in handling this drug are no greater than 30-fold across the entire human population, this drug would be of little concern to the pharmacogeneticist.

3.3. Drug–drug interactions

Interactions between drugs are another important consideration, because the nonhospitalized adult takes on average about five drugs each day and the average hospitalized patient about 14 drugs each day (Nebert, 1999). Drug–drug interactions are likely to be greatly magnified by allelic differences not only in metabolism genes, but also in other types of genes. If a patient is taking drug X and drug Y and these two drugs compete for the same site of action (Fig. 3), and if the patient is, for example, a poor metabolizer for drug Y, then the "recommended" prescribed dose of drug Y will result in abnormally elevated levels of drug Y—thereby leading to an exaggerated drug–drug interaction, i.e. greater inhibition of the efficacy (or toxicity) of drug X. The same might also apply to the interactions of drugs with receptors, drug transporter proteins, and ion-channel membrane proteins. In pharmacogenetic studies, one must be constantly on guard against any possibility of drug–drug interactions.

3.4. Multiple enzymes means involvement of multiple genes

It should also be appreciated that just as a metabolic pathway goes from compound "A" to "B", there usually are enzymes that will take the compound "B" back to "A" (Fig. 4). Therefore, if a patient is genetically "high" for methyltraferase activity and a "poor metabolizer" for *N*-demethylase activity (Fig. 4, bottom), his methyltransferase activity would appear to be even greater. Multiple genes, of course, can participate together in the activation or degradation of a drug. For example (Fig. 5), caffeine is detoxified by the combined action of at least five enzymes:

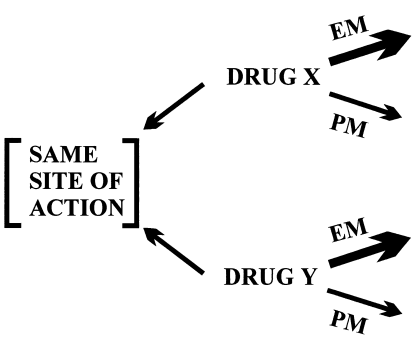

Fig. 3. Diagram of possible genetic effects on drug–drug interactions. EM, extensive-metabolizer phenotype. PM, poor-metabolizer phenotype (modified from Nebert, 2000).

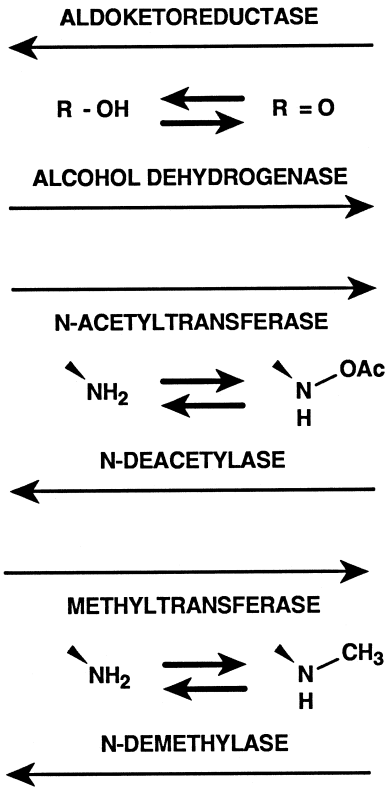

Fig. 4. Three examples of metabolic pathways catalyzed by drug-metabolizing enzymes, illustrating that different drug-metabolizing enzymes can work in one or the other direction in the same pathway (modified from Nebert, 2000).

cytochrome P450 1A2 (CYP1A2), CYP2E1, CYP2A6, *N*-acetyltransferase-2 (NAT2), and xanthine oxidase. Given the fact that each enzyme is encoded by a gene, and there are multiple alleles of each gene in any human population, it becomes easy to understand that most phenotypes of drug efficacy or toxicity will probably be very complicated, i.e. a multiplex phenotype representing the contribution of two, and probably many more than two, genes.

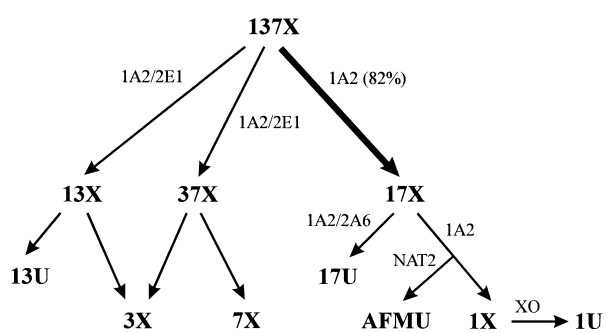

Fig. 5. Major metabolic pathways of caffeine in humans (reviewed in Kalow and Bertilsson, 1994). 137X, 13X, 37X, 17X, 3X, 7X, 1X = 1,3,7-trimethyl, 1,3-dimethyl, 3,7-dimethyl, 1,7-dimethyl, 3-methyl, 7-methyl, and 1-methylxanthine, respectively. 13U, 17U, 1U = 1,3-dimethyl, 1,7-dimethyl, and 1-methyluric acid, respectively. AFMU = 5-acetylamino-6-formylamino-3-methyluracil. Cytochromes P450 1A2, 2E1, 2A6 = CYP1A2, CYP2E1, CYP2A6. NAT2 = *N*-acetyltransferase-2. XO = xanthine oxidase.

3.5. Ethnic differences in drug response

It has also become increasingly apparent that ethnic differences in metabolism genes are of great importance in the therapeutic efficacy, as well as toxicity, of drugs (Kalow and Bertilsson, 1994; Ingelman-Sundberg et al., 1999). For example, approximately 92% of Egyptians, but less than 10% of Asians, are "isoniazid-slow-acetylators" (Weber, 1997). Although not nearly so well studied, it is anticipated that similar striking ethnic differences in allelic frequencies will be found in receptor genes, transporter genes, ion channel genes, and probably any other category of genes in the human genome. By knowing the allelic frequencies for DNA sequence variants associated with a particular trait (drug efficacy, toxicity) in distinct ethnic populations, therefore, the physician would be able to increase (or avoid) usage of a particular drug, thereby improving the rate of efficacy, or diminishing the risk or adverse drug reactions in one or another ethnic group.

4. Need for a quantitative phenotype

In a human population receiving a particular drug, the critical issue is to identify and classify the phenotypic variation (therapeutic or toxic response) into quantitative, unequivocal groups. If an entire population is examined for, say, toxicity as a function of drug dosage, there will be a general increase–with means, and standard-deviations-of-the-means, at each level of drug dose (Fig. 6). If statistically significant outliers can be found, these extreme genetic variants are far more useful—to the investigator who seeks to identify a mechanism for the trait—than minor inter-individual differences in the general population. The "resistant outlier" represents a patient with little therapeutic (or toxic) response to a large dose of drug, whereas the "sensitive outlier" exhibits an exaggerated response to a low dose of drug.

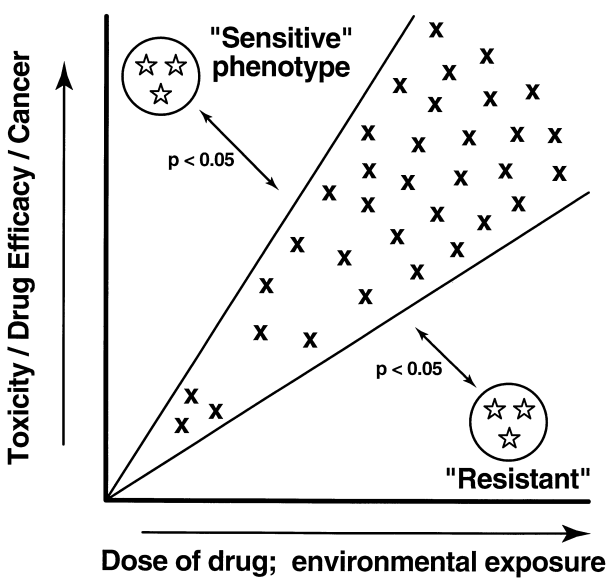

Fig. 6. Generic "dose–response" relationship in which the dosage of any drug, or exposure to any environmental agent, is plotted on the abscissa. Toxicity, drug efficacy, or cancer, as examples, can be plotted on the ordinate. Whereas one might see a gradual increase in toxicity/efficacy/cancer as a function of increasing drug or environmental dose, "resistant" or "sensitive" outliers—significantly different from the general population—can be particularly valuable to the pharmacogeneticist (reproduced with permission from Nebert, 1999).

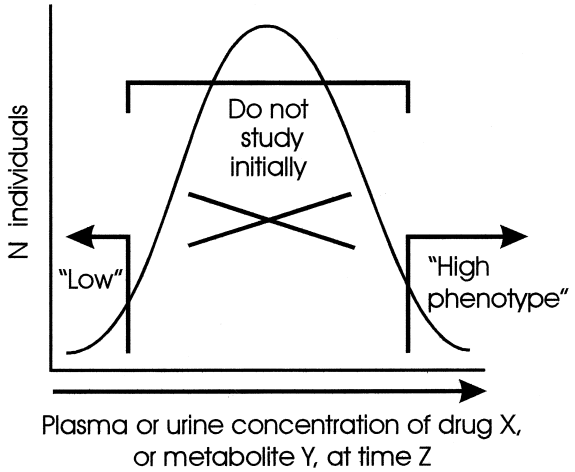

Fig. 7. Hypothetical distribution of numbers of patients (N) in a study population in which their plasma or urine concentration of "drug X" or "metabolite Y" has been determined at "time Z", after a particular dose of drug had been given. Because the majority of the population is neither at the extreme high end nor low end of the curve, to study all these patients invites problems in complexity, due to the polygenic nature of the defined phenotype. Although this hypothetical curve is Gaussian, it is realized that curves in a clinical population need not be Gaussian (reproduced with permission from Nebert, 1999).

Experiments like those depicted in Fig. 6 are commonly carried out in laboratory animals, however, and prospective studies involving a large dose range are usually not ethically possible in human trials. Because drug efficacy or toxicity always represents a gradient, one approach to quantitate the clinical phenotype would be to examine the extreme ends of a distribution, in much the same way as genetic studies have been done in the dissection and identification of genes responsible for blood pressure homeostasis (Jacob et al., 1991; Brown et al., 1996; Halushka et al., 1999). Fig. 7 shows the number of individuals as a function of the "concentration of drug X, or metabolite Y, at time Z". Patients who are arguably "intermediate" should not (at least, initially) be included, for the sake of reducing the complexity of the two subsets being studied. Hence, it is possible to examine relatively small numbers of highly informative patients having an unequivocal quantitative trait—following which one can attempt to correlate the phenotype with a genotype. For example, selecting only patients from the top and bottom 2.5th percentile of a normalized blood-pressure distribution, Halushka et al. (1999) examined 75 candidate genes in 74 patients (148 alleles); the remaining 95% in the middle

were not included in this study. The same approach can be highly successful in pharmacogenetic studies.

4.1. Method of extreme discordant phenotype (EDP)

Fig. 8 illustrates the concept of EDP methodology. As the dose of any drug under study is increased, we would expect more patients to exhibit toxicity. Hence, as the dose of the drug is increased and the incidence of toxicity rises, there would be fewer (still taking this drug) in the population who have no toxicity. Those patients receiving the lowest doses who then exhibit an exaggerated toxic response (i.e. a dose-independent response) are classified as the "most sensitive" trait, and those patients receiving the highest doses who exhibit no toxicity are classified as the "most resistant" trait. In order to identify the gene(s) responsible for this difference, it is therefore very important to focus on these most-sensitive and most-resistant phenotype groups, and initially, one should disregard the vast majority of patients who are being given intermediate doses (of the drug under study) or who exhibit intermediate responses (showing no toxicity at low doses, or small degrees of toxicity at high doses). When selecting for, say, the "top 5%" and "bottom 5%" extremes, a possible caveat includes the possibility that we would be selecting for an "extreme phenotype gene" responsible for the most-sensitive or the most-resistant trait; this unlikely possibility can be ruled in or out later, however, when the pharmacogeneticist can assess the impact of any established DNA sequence variant(s) on "intermediary" patients (i.e. those in the gradient between the bottom 5% most-sensitive and the top 5% most-resistant groups as illustrated in Fig. 8).

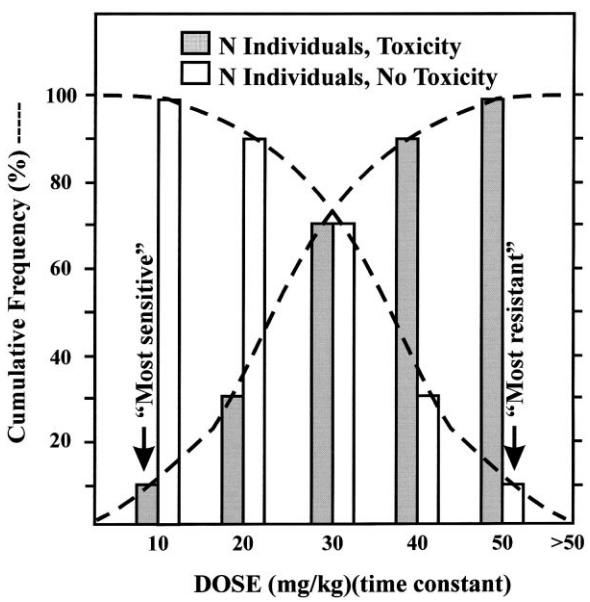

Fig. 8. Generic dose–response relationship from an idealized clinical population. The bars depict theoretical numbers of patients (exhibiting toxicity or no toxicity). The percent cumulative frequency is also plotted, as a function of drug dosage, in patients exhibiting a particular type of toxicity (increasing from left or right) versus patients showing no toxicity (increasing from right to left). The curves shown here are symmetrical but, of course, would not necessarily be expected to be symmetrical in the typical clinical population. Also, the dose (10 to >50 mg/kg) is completely arbitrary, simply to make the point; this fivefold dose range might easily be only twofold for certain drugs. As discussed in the text, rigorous parameters for unequivocal "toxicity" versus "no toxicity" must be set ahead of time. The same approach could also be used for drug efficacy—if rigorous quantitative endpoints were defined and established ahead of time.

Fig. 8 shows data in which the dose is increasing, whereas the time of administered drug (months, years) is constant. Conceivably, a similar estimation could be illustrated in which the time of drug being administered is increasing and the dose of drug is constant. Obviously, however, dose and time are not dependent on one another; in other words, due to the pharmacokinetics of virtually all drugs, 10 mg/kg/day for 100 days is almost never equivalent to 100 mg/kg/day for 10 days. Also, 10 mg/kg/day might be ineffective as far as efficacy (as well as toxicity), whereas the "threshold" of this hypothetical drug might be, say, 60 mg/kg/day—where one might see both efficacy in some and toxicity in other patients. It is also possible that one might see toxicity in some individuals at, say, 40 mg/kg/day, while the drug is not efficacious in most individuals below 60 mg/kg/day.

This EDP approach has previously been carried out, in principle, in numerous pharmacogenetic studies—but not in any rigorous mathematically systematic way that has been recognized and described in detail in this mini-review. For an example of previous studies, isoniazid-poor-acetylator patients during the late 1940s and 1950s were found to exhibit isoniazid-induced peripheral neuropathy (reviewed in Weber, 1997), and this trait was shown to be inherited as autosomal recessive (Evans et al., 1960), which subsequently was demonstrated to reflect DNA sequence variants in the *N*-acetyltransferase-2 (NAT2) gene and amino-acid alterations in the NAT2 enzyme protein (Blum et al., 1990). DNA sequence variants were thus correlated with the extreme phenotypes of slow versus rapid acetylation. As another example, "poor metabolizers" of the antihypertensive agent debrisoquine were found to become hypotensive when given the recommended prescribed dose (Idle and Smith, 1979). This affliction is now known to reflect various aberrations in the cytochrome P450 2D6 (*CYP2D6*) gene, which result in alterations in the CYP2D6 enzyme protein (Gonzalez et al., 1988). Again, the genotypes were ultimately correlated with the extreme clinical phenotypes of poor versus extensive metabolism. These two above-mentioned examples represent metabolism differences. Multi-drug resistance P-glycoproteins (MRPs) are part of a large and ancient ATP-binding cassette (ABC) superfamily of drug transporter genes; mutations in one of these genes are known to be responsible for Dubin–Johnson syndrome (Tsujii et al.,

1999), an inherited form of hyperbilirubinemia and an example of a pharmacogenetic defect in a transporter gene. Normal patients—the other extreme phenotype, opposite of Dubin–Johnson syndrome—do not show these mutations.

If one goes back to the original studies that describe each of these polymorphisms, one can appreciate that none of these phenotypes represents a "clear-cut" example of a single-gene trait. In these and all other examples of pharmacogenetic differences, there is always a gradient—suggesting the involvement of two or more genes responsible for the phenotype.

Dozens of other pharmacogenetic differences have been described clinically (summarized in Nebert, 1999), but the physician's ability to predict which patient might be affected, in advance of receiving the drug, remains almost impossible. For example, halothane maintenance anesthesia causes a dose-dependent depression of white-cell precursors as a function of time the anesthetic was used, but a dose-independent toxicity of the bone marrow (aplastic anemia) in the rare patient; this latter trait is known to run in families (and hence, most likely reflects a genetic predisposition). Likewise, the antibiotic chloramphenicol causes a dose-dependent depression of the bone marrow as a function of the length of time of treatment, but a dose-independent aplastic anemia in the occasional individual. Of patients in the US who receive glucocorticoid-based eye drops, about one in 20 developed glaucoma, this is an autosomal recessive trait. More than six dozen examples of pharmacogenetic differences have been described (Vesell, 1969; Kalow and Bertilsson, 1994; Weber, 1997; Nebert, 1997b, 1999; Caraco, 1998), many with the molecular mechanisms still unexplained. This proposed approach of EDP methodology is a realistic means for dissecting most, if not all, of these complex gene–drug interactions.

In the past 40 years, the usual clinical study generally includes all patients on a given medication, and often "control patients" are "matched for age, gender, social status, etc.," by random telephone-dialing or some other random statistical means. Often there is little regard for the dosage or time the drug has been administered, the threshold at which toxicity is observed in the most sensitive patient group, varying degrees (i.e. gradient) of toxicity, or the relative dose needed for efficacy versus dose that causes toxicity in some patients. It should be obvious that, in the field of pharmacogenetics, if a patient has never been treated with the drug under study, there is no way of knowing in advance that person's phenotype with regard to his degree of sensitivity or resistance. Although a patient might have a particularly dramatic underlying genetic predisposition to sensitivity or resistance to any specific drug or other environmental agent, unless the patient has been sufficiently challenged with this chemical, one might never know his covert genotype with regard to specific metabolism (or receptor, transporter, ion channel, cell cycle, nucleic acid and protein repair, HLA protein, other signal transduction pathway) genes. The EDP approach is very straightforward and much more likely to identify quantitative phenotypes successfully, following which one can correlate the extreme phenotype (drug efficacy, or sensitivity or resistance to toxicity) with the genotype (DNA sequence variant).

4.2. Statistical power of EDP methodology

In principle, the EDP approach has been described in affected sib-pairs (Risch and Zhang, 1995, 1996; McCarthy et al., 1998; Epstein et al., 2000), and selection of the most appropriate ethnic groups, and numbers of patients needed for an informative study, has also been amply discussed (Kruglyak, 1999; Wright et al., 1999; Pratt et al., 2000). For the EDP approach without using sib-pairs, a random-mating diploid population is assumed, with non-overlapping generations. Consider two sites, B and Q in a gene underlying the pharmacogenetic trait of interest. Let Y be a trait value with mean μ and variance σ_p^2. Let $y = (Y - \mu)/\sigma_p$. The truncated phenotypic selection intensity, γ, is defined as the frequency of the phenotype value at the selection threshold, divided by the percentile representing the threshold (Falconer, 1989, p. 193). The null hypothesis is that the mutation is neutral, i.e. does not have an effect on the trait. This hypothesis can be tested by the proposed statistics. The power to detect the functional mutation depends mainly on the difference between the frequencies of the alleles at the two sites B and Q ($\Delta p = p_1 - p_2$), the genetic effect of the functional mutation (d), and the selection intensity (γ). Let us assume the significance level of the test to be $\alpha = 0.05$. Tables 1 and 2 list the sample sizes required to reach the power 0.8. Clearly, in general, the sample size decreases—as the difference of the allelic frequencies at the two sites increases. When the frequencies of the allele at the two sites are the same, they are in fact in complete linkage disequilibrium, and one would not be able to discriminate whether the change of the frequencies is due to genetic effects or linkage disequilibrium, because in that case, $d = p(1 - p)$. When either genetic effect or selection intensity increases, one can expect that the sample sizes required will decrease. Also, from Tables 1 and 2, it can be seen that the sample sizes necessary for reaching a power of 0.8 with a significance level of $P = 0.05$ are in a very reasonable range (Falconer, 1989).

Another way to define the two sites B and Q within or near a gene is to make one site the trait locus (*TL*), which has been functionally proven to be responsible for the phenotype, and the other site the marker locus (*ML*), which, as an internal control, is a site known to be polymorphic but also known not to segregate with *TL* or be variable as a function of the pharmacogenetic trait (drug efficacy or toxicity); choice of the *ML* can be either an established DNA sequence variant or insertion or deletion of repetitive DNA, as described in Section 2. Thus, it is

Table 1
Sample sizes required for achieving power of 0.8 with $\alpha = 0.05$ for a fixed genetic effect $d = 0.2$ and the selection intensity $\gamma = 1.7$

p_2	p_1								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1		388	145	93	75	69	73	90	153
0.2	309		574	195	118	93	88	100	161
0.3	88	446		708	232	139	112	115	171
0.4	42	110	529		791	258	159	140	185
0.5	23	46	119	560		824	276	184	208
0.6	14	23	46	116	538		809	292	248
0.7	8	12	21	40	100	464		754	339
0.8	5	7	9	16	30	74	340		689
0.9	2	3	4	5	8	16	38	174	

anticipated that, in the most-sensitive patients, one might see a ratio of a specific (unique) DNA sequence variant at the trait locus (*TL*) to that at the marker locus (*ML*) to be, say 0.90, as compared with a ratio of ~ 0.50 in the most-resistant patients, and the number (N) of individuals tested could be used to calculate statistical significance by binomial analysis. By selecting the top and bottom 5.0th percentile of our normalized dose–response EDP population (as illustrated in generic terms in Fig. 8), one can calculate the number of patients needed—all of whom have received the drug under study—to prove a statistically significant genotype–phenotype association.

One should then be able to state with confidence that "a specific DNA sequence variant, which would need to include x number of patients, would have a $0.xx$ power to detect important differences in drug (efficacy or) toxicity response at a $P < 0.05$ level of significance". The EDP patient-pair method would definitely help to increase the statistical power to identify the alleles having relatively small genetic contribution to the trait. The approach described here is basically to associate a DNA sequence variant (genotype) with a drug efficacy or toxicity trait (phenotype), when the allelic frequencies of the two groups of individuals with extreme phenotypes are contrasted to one another.

In the usual power calculation analyses, it is assumed implicitly that the allele under study is the allele that is responsible for the variation in phenotype; this may at first glance appear to be an oversimplification, since (as discussed below), the coauthors of many studies do not really know whether a "significant association" is due to the true contribution of the allele or simply due to linkage disequilibrium between their marker allele and the true mutation responsible for the variation in phenotype (at the trait locus *TL*). By performing experiments that prove function, however, this caveat can be ruled out with confidence. As detailed above, EDP methodology can make the distinction between these two cases. The numbers in the Tables 1 and 2 may look small, but they have been selected from a much larger sample of unrelated individuals. Choosing the bottom 5% most-sensitive and top 5% most-resistant groups means that the 90% in the middle have been considered, making the statistical power approximately nine times greater. Choosing the bottom 1% most-sensitive and top 1% most-resistant groups means that the 98% in the middle have been considered, making the statistical power almost 50 times greater.

Table 2
Sample sizes required for achieving power of 0.8 with $\alpha = 0.05$ for a fixed frequency $p_2 = 0.5$ and the selection intensity $\gamma = 1.7$

p_1	d								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.6	3296	824	366	206	132	92	67	52	41
0.7	1104	276	123	69	44	31	23	17	14
0.8	738	184	82	46	30	21	15	12	9
0.9	831	208	92	52	33	23	17	13	10

The merit of EDP methodology is to screen for genotypic differences in unrelated patient-pair groups exhibiting extreme phenotypes. This approach becomes particularly useful and powerful when the knowledge of the pathway (e.g. metabolism, transporters, etc.) is well understood, from which one can easily identify the "correct" enzymes/genes in which to discover allelic variants. This would markedly increase the statistical power—if we know with great certainty that the particular gene under study is very likely to be involved in the phenotypic variation. In essence, the association study is only used as an initial confirmation, rather than the sole step for gene discovery. As emphasized earlier, the EDP approach should be applicable to virtually any complex disease or multiplex phenotype—including drug efficacy or toxicity.

Let us suppose that individuals with the DNA sequence variant of interest above the 80th percentile and individuals below the 20th percentile are selected. Fig. 9 shows the

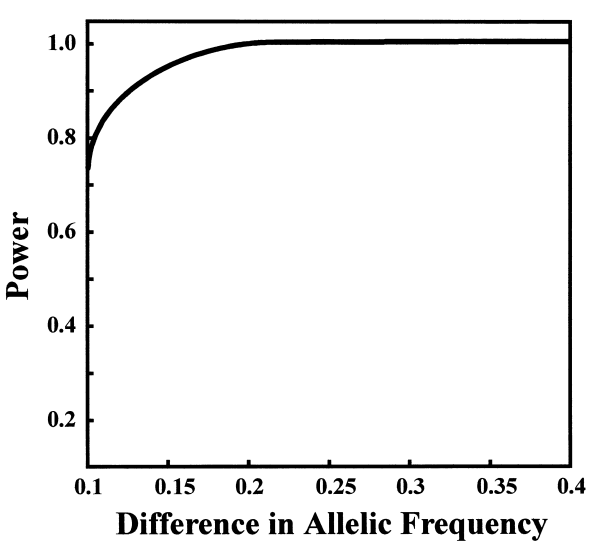

Fig. 9. Statistical power of detecting a DNA sequence variant (trait locus, *TL*) as a function of the difference between the frequencies of the alleles at the two sites *B* and *Q*, as detailed in the text. These data are calculated for the fixed value of $d = 0.2$.

power of detecting a proven functional mutation, as a function of the difference between the frequencies of the alleles at the two sites—assuming that the proposed sampling scheme is used. As illustrated, the proposed phenotypic selection scheme provides considerable power for detecting statistically significant functional variation influencing the phenotype of interest, as long as the frequency differences of the marker alleles in the case and control groups are larger than 15%.

EDP methodology is a powerful and robust approach for studying unrelated individuals (patient-pairs) in drug-treated populations. The advantage of the EDP approach is that it does not depend on the actual values of the quantitative trait, but only on the fact that patient-pairs can exhibit extreme values. In general, association studies with candidate loci can be far more powerful for detecting weak gene effects than linkage studies (Risch, 1987; Greenberg, 1993; Eaves and Meyer, 1994). The limitation of this approach, however, is that it requires either prior identification of the causative genetic variant itself (i.e. the *TL*) or another variant in linkage disequilibrium with it, as well as a neutral nonsegregating polymorphic site (i.e. the *ML*). Linkage disequilibrium spans only short genomic regions (5–10 Mb), thereby limiting its utility in a global genome screen. For candidate loci, however, this approach can be extremely powerful if variation at the locus contributes to phenotypic variation.

5. Proof of a functional genotype

Many molecular epidemiology studies have attempted to demonstrate associations between a phenotype and a DNA sequence variant when the DNA change had not been rigorously proven to reflect an alteration in function of the gene product. For example, associations between the *Alu* I insertion/deletion in intron 16 of the *DCP1* gene and phenotypes such as "risk of myocardial infarction", "risk of arterial hypertension", "risk of left ventricular hypertrophy", "elite athletic performance", and "response to physical training" have been described. However, when more than 24 kb across the entire *DCP1* gene (5′ flanking, exons, introns, 3′ nontranslated) was sequenced in 11 individuals, the conclusion was that the nucleotide change responsible for these cardiovascular phenotypes might well lie outside the sequenced region (Rieder et al., 1999)—although some evidence now suggests several DNA sequence variants within 1 kb of one another in the 3′ region might be associated with elevated angiotensen I-converting enzyme levels (Zhu et al., 2000). Another example is the reported association between a deletion in exon 18 of the α-2 macroglobulin (*A2M*) gene and Alzheimer Disease (Blacker et al., 1998); these results have been disputed (Dow et al., 1999; Rudrasingham et al., 1999; Rogaeva et al., 1999) as possibly due to regional population differences, another functional variant closely linked to the *A2M*2* allele, or another gene in linkage disequilibrium weakly associated with Alzheimer Disease. Another example is the professed evidence of a role for the gene poly(ADP-ribose) polymerase (*PARP*) in susceptibility to systemic lupus erythematosus (Tsao et al., 1999). These data have been refuted by Criswell et al. (2000) and by the original authors themselves (Tsao et al., 2000), illustrating many of the difficulties faced by researchers who are attempting to find associations between genotype and phenotype in a complex disease such as systemic lupus erythematosus; this is elegantly discussed by Risch (2000). Another example involves the reported association of two DNA sequence variants—an amino acid change (I462V) and a nucleotide change 450 bp 3′-ward of the human *CYP1A1* gene—with lung cancer in Japanese but not Caucasian or Black populations (Hayashi et al., 1992; reviewed in Nebert, 1997b). No differences have been found in catalytic activity, however, between the wild-type and variant cDNA-expressed CYP1A1 enzyme (Zhang et al., 1996; Persson et al., 1997)—suggesting that another gene (perhaps *CYP1A2*?) in linkage disequilibrium might be responsible for this association. As a final example, there are reports of a purported association of a particular allele of the vitamin D_3 receptor (*VD3R*) gene with lowered bone mineral density and prostate cancer (Morrison et al., 1994; Cooper and Umbach, 1996). After the cDNA-expressed VD3R protein was demonstrated to have no alterations in function (Gross et al., 1998), it was concluded that another gene in linkage disequilibrium is most likely responsible for the lowered bone mineral density, as well as the increased risk in prostrate cancer. In conclusion, one must recognize that a DNA sequence variant found between the most-sensitive and most-resistant patient groups by EDP methodology might not necessarily be associated with the gene in which the mutation exists. Proof of function is absolutely required.

5.1. Does a single-gene Mendelian trait really ever exist?

It is becoming increasingly appreciated that even phenotypes of patients believed to have "simple" Mendelian disorders virtually always represent multiplex traits—affected by thresholds, modifier genes, and environmental factors (Dipple and McCabe, 2000). Clearly, virtually all pharmacogenetic differences, including drug-efficacy or drug-toxicity phenotypes, depict multiplex phenotypes. One therefore needs to identify patients with an unequivocal phenotype, and then define the associated genotype (DNA sequence variant) in one or more genes that contribute to the drug-associated trait.

Many of the genes that cause so-called "simple" Mendelian disorders have been identified (Collins et al., 1997), and although these diseases can be extremely debilitating, most are rare and collectively affect only a few percent of the population. In contrast, the lifetime risk of having a genetically influenced multifactoral disorder in

western populations has been estimated to be 60% (Baird et al., 1988). If one includes adverse drug reactions in patients who are taking a prescribed pharmaceutical agent, this lifetime risk estimate would be even higher.

5.2. Two approaches in selecting pharmacogenetic loci of interest

Common clinical disorders have significant genetic components, as defined by attributable risk (Lander and Schork, 1994), but environmental susceptibility genes that contribute to drug efficacy or adverse drug reactions have, for the most part, not yet been clearly delineated. There are two fundamental approaches to identifying the loci responsible for adverse drug reactions. One, linkage-disequilibrium mapping in large population samples would have increased power to detect such genes (Jorde, 1995; Risch and Merikangas, 1996), but such a whole-genome-scan approach is constrained by the degree of allelic heterogeneity (Kruglyak, 1997; Terwilliger and Weiss, 1998; Wright et al., 1999). Two, the candidate-gene approach seems more straightforward and promising—especially if "scientifically reasonable" candidate genes can be postulated for the phenotype (drug toxicity) being studied.

Wright et al. (1999) have considered two contrasting models of the allelic diversity that underlies complex diseases: the "restricted polymorphism model" in which there is a relatively small pool of common polymorphic disease alleles (frequency > 0.01), which represent a highly restricted allelic subgroup from a limited number of loci; and the "multi-equivalent risk model" which assumes that there is a large pool of risk alleles, at a large number of loci, with varying effects on risk, but which includes many alleles with equivalent effect. For example, at least 100 loci appear to affect coronary artery disease susceptibility (Sing et al., 1996), and a substantial portion of pigmentary retinopathy is caused by more than 600 mostly-rare-disease alleles at more than 55 loci (Daiger et al., 2000). In the more common complex diseases—such as hypertension, dementia and ischemic heart disease—the diversity is likely to be even more extreme (Wright et al., 1999). Whether the average drug efficacy or adverse drug reaction trait represents a multiplex phenotype having this degree of diversity, or a somewhat less degree of diversity, remains to be determined; one would hope, however, that the average drug efficacy or adverse drug reaction trait might reflect the restricted polymorphism model, i.e. allelic differences at probably a dozen or fewer loci. Such diversity would then make a very strong case to study pharmacogenetic disorders by the candidate-gene approach and to look at subsets of clinical populations that represent the extremely discordant phenotype rather than to perform a total-genome scan, or attempt an analysis of an entire population taking a particular drug. Virtually all drugs cause a dose-dependent toxicity, but by comparing the subsets of most-sensitive and most-resistant patients, one is selecting for further biochemical and genetic analyses individuals who display dose-independent toxicity.

Looking at a list of the more than six dozen pharmacogenetic differences (Nebert, 1999), one can conclude that these are complex traits that are generally quantitative (i.e. measurable on a continuous scale) rather than absolute (measurable on a discrete scale). Each pharmacogenetic disorder can be defined by a threshold applied to a continuous variable (usually drug dosage, but it could also be drug administered as a function of time). Population studies employed to map loci that underlie a quantitative trait (quantitative trait loci, QTL mapping) have been carried out using affected sib-pairs (Risch and Zhang, 1995; McCarthy et al., 1998) and in pedigrees (Pratt et al., 2000). In conventional terms, the squared difference in trait values for a sib pair is regressed on identity-by-descent (IBD) at one or more marker loci. For sib pairs selected at random, however, the power of such an approach is quite low unless the proportion of variance (heritability) due to a single contributing locus is large, e.g. 50% (Blackwelder and Elston, 1982). When sib pairs are selected because of their extremely discordant phenotype, and the unselected sib's trait is regressed on IBD with the proband, the power is increased but still remains low at small values of heritability (Carey and Williamson, 1991). The power of heritability can be more than made up for, however, by (a) studying common rather than rare alleles in populations, and (b) selecting patient pairs that exhibit unequivocal, unique, and extremely discordant phenotypes.

5.3. Search for candidate gene polymorphisms

Because the majority of pharmacogenetic differences described clinically to date represent alterations in metabolism genes (reviewed in Nebert, 1999), the first category of candidate gene polymorphisms one might choose to rule in or out this possibility would involve metabolism. Giving a "test dose" of the drug under study to those most-sensitive and most-resistant patients who have been selected for further study, one could quantitate—at a specified time following this challenge dose—the amount of nonmetabolized parent drug and specific metabolites in the urine and/or blood (and saliva, if appropriate). Obviously, extensive-metabolizer (EM) individuals will degrade the drug more rapidly leading to lower concentrations, and poor-metabolizer (PM) individuals will degrade the drug more slowly leading to higher concentrations, of the parent drug in the urine (or plasma or saliva) at the same specified time. Urinary metabolite formation would be higher in the EM than in the PM patient. If metabolism differences are found between the most-sensitive and most-resistant patient groups, appropriate candidate metabolism genes could then be selected, on the basis of previously published reports in the literature about the established metabolism of the drug under study, following which DNA sequence variants that differ in

these metabolism gene(s) between most-sensitive and most-resistant patients could be sought.

If urinary and plasma drug and metabolite data do not differ between the most-sensitive and most-resistant patient groups, one must then focus on other categories of candidate genes. Receptor gene, transporter gene, or ion channel gene polymorphisms could be fairly easy to detect—if these functions exist in blood cells, buccal mucosa, or urinary epithelial cells (i.e. cells easily available by non-invasive techniques). In other words, the rate of movement of a drug into the cell, or affinity determinations by Scatchard plots, might be measured as differences in these gene products between most-sensitive and most-resistant patients. Receptor, transporter or ion-channel function might exist, however, only in specialized cell types (e.g. hepatic or gastrointestinal cells, or renal or central-nervous-system cells). The same is likely to be true for other classes of candidate genes (encoding cell cycle transcription factors, nucleic acid and protein repair enzymes, HLA proteins, factors participating in other signal transduction pathways). If these functions do not exist in cells that are easily available by noninvasive techniques, there are two alternative approaches. One would be to attempt to culture the appropriate cell types from most-sensitive and most-resistant patients, but this would represent invasive and possibly unethical techniques. The other would be to select appropriate candidate genes—based on their likelihood of success by studies published in the literature—and proceed directly with attempts at detection of DNA sequence variants in these genes by looking at genomic DNA. One might also argue, at this particular juncture, for the possibility of success by a total-genome scan instead of the candidate-gene approach.

5.4. Finding a DNA sequence variant between the most sensitive and most resistant phenotypes

At the present time, high-throughput DNA variant detection/confirmation is carried out primarily by DNA resequencing. Differences in DNA sequence variants in candidate genes should be searched for, between the most-sensitive and most-resistant patient groups. If the candidate genes are already well characterized for DNA sequence variants [e.g. the numerous variant alleles of several *CYP* genes (Oscarson et al., 2000) or the *NAT2* and *NAT1* genes (Hein et al., 2000)], one could set up oligo-PCR primer reactions to screen for these common polymorphisms. In addition to all exons, all introns and at least ~5–10 kb of 5′ upstream promoter of each candidate gene is recommended to be sequenced in several patients from these extreme-phenotype groups—for the potential discovery of additional important polymorphic variation (Nebert, 2000). This is to ensure that the possible contextual dependency to account for the extreme phenotype is not overlooked, i.e. the presence of two or more intragenic non-synonymous mutations/variations may be present and necessary for the expression of (or association with) the phenotype.

Peripheral blood genomic DNA is the most widely used accessible source to conduct DNA resequencing studies. Reference sequences for the initial design of PCR primers might originate from those published in the literature or web sites, from GenBank, from the "working draft" of the human genome, and/or by comparison with genomic sequences in the Celera single-nucleotide polymorphism (SNP) database. Thus, by public and "private" databases, all available genomic data can be searched for known sequence variations, and compared with obtained sequences from the most-sensitive and most-resistant patient groups.

In addition, the human gene-based single-nucleotide polymorphism database, called HGBASE (human genic bi-allelic sequences), is now freely available for academic use and hosted at the Karolinska Institute, Stockholm, Sweden (Brookes et al., 2000). HGBASE provides a highly curated and annotated non-redundant summary of all publicly reported human gene/genome sequence variation, which currently lists more than 6000 single-nucleotide polymorphisms. This program is run by a European consortium involving the Karolinska Institute (Sweden), the European Bioinformatics Institute (UK), and the European Molecular Biology Laboratory (Germany), with support from Interactiva (Germany) and Pharmacia. Data entries may be searched for by text or sequence queries, and full database downloads are available.

One significant question is whether intronic sequence variation can be ignored (i.e. other than splice acceptor and donor sites, are intronic variations nonfunctional?). Another debate concerns DNA resequencing versus DNA-chip high-throughput analysis for the discovey of DNA sequence variants. As of September 2000, several laboratories besides those at this university have come to the same conclusion and/or have shown that DNA resequencing presently is still more reliable than DNA chips (Nickerson et al., 1998; Clark et al., 1998; Wang et al., 1998; Rieder et al., 1999; Cargill et al., 1999; Halushka et al., 1999; Ohnishi et al., 2000; Yamada et al., 2000). There are also problems with the methods for determining homozygous versus heterozygous DNA sequence variants in a chromosome pair from each patient. These are all difficult yet important issues and have recently been discussed in detail (Nebert, 2000).

5.5. Proof that a DNA sequence variant is responsible for the phenotype

Finally, the evaluation of the functional consequence of a particular DNA sequence variant is essential. In particular, if one observes a strong association of a DNA sequence variant with a trait, one must determine whether that DNA sequence variant is responsible for altering any possible transcriptional binding sites if it is in the pro-

moter, whether there are differences in mRNA or protein stability if the mutation is in an intron or the noncoding region, or whether the DNA sequence variant would have significant functional/structural consequences if it is an amino-acid replacement mutation in the coding region. The possibility of a change in transcription termination or alteration of a splicing site is, of course, also important.

How might one obtain proof of function? This can be carried out by measurements of the gene product (protein) in terms of higher or lower metabolism, rate of trans-membrane transport, receptor affinity, and so on. These procedures would include, but not be limited to: (a) design, construction, and expression of recombinant proteins in bacterial, yeast, insect, and mammalian cell cultures or in knockout or other transgenic mouse lines; (b) purification and partial characterization of expressed enzymes and proteins; (c) heterologous cDNA expression—by means of developing retroviral vectors with and without conditional activators/repressors; (d) transient and stable transformations/transfections of selected cDNA constructs in bacterial, yeast, insect and stable/transient mammalian cell systems; (e) Epstein–Barr Virus (EBV) transformation of lymphoblast cultures to ensure a continuous supply of important DNA clinical samples; (f) the yeast two-hybrid system for characterizing transcription factors; (g) partial or full purification of recombinant enzymes/proteins by Ni^{2+}-resin chromatography, and affinity chromatography; and (h) characterization of recombinant enzymes/proteins using kinetic analyses, high-pressure liquid chromatography, intrinsic fluorescence spectrocopy, gas–liquid chromatography/mass spectrometry, and circular dichroism.

6. Conclusions

The field of pharmacogenetics represents studies of the heritable basis of variability in response to drugs. The majority of these pharmacogenetic differences reported to date exist in metabolism genes. Traits such as "drug efficacy" or "drug addiction" represent multiplex phenotypes, i.e. interaction between the drug (or its metabolites) and two, and usually many more than two, genes. Because there is probably always a gradient in these phenotypes (whether it be efficacy or toxicity), it is difficult yet important to select patients that have a relatively unequivocal phenotype. The method of "extreme discordant phenotype" (EDP) is described in detail herein. EDP methodology is a powerful and robust approach for studying unrelated individuals (patient-pairs) in drug-treated populations. With the use of EDP methodology, DNA sequence variants (genotype) can then be unconditionally correlated with drug efficacy or toxicity (phenotype). EDP methodology is intuitive and has been used unknowingly in a number of previous clinical studies. This EDP approach should be applicable to virtually any pharmaceutical agent in patient populations.

7. Note added in proof

As an example of applying EDP methodology to a clinical population, Wanner et al. (1999) studied the possible association between allelic frequencies in the human aromatic hydrocarbon receptor (*AHR*) gene (genotype) and severity of dioxin-induced chloracne (phenotype) in chemical workers. Seven "severe" and two "moderate" chloracne patients were compared with five patients showing no skin rash. The "severe" phenotype individuals exhibited dioxin concentrations in an adipose biopsy between 717 and 12,200 parts per trillon (ppt), whereas the patients showing no chloracne had dioxin concentrations between 30 and 831 ppt. The "most-sensitive" patient would therefore be the person with "severe" chloracne at the lowest dose (717 ppt), and the "most-resistant" patient would be the person with no skin rash and the highest level of adipose dioxin levels (831 ppt). In this study, no association between the genotype and phenotype was found.

Acknowledgements

I thank my colleagues for valuable discussions and a careful reading of this manuscript. I am especially grateful to Jin (Felix) Li and Momaio Xiang (University of Texas, Houston) for helping me understand some of the mathematical calculations that support the method of EDP. This work was funded in part by NIH Grants R01 R01 ES06321, R01 ES08147, R01 ES10416, and P30 ES06096.

References

Baird, P.A., Anderson, T.W., Newcombe, H.B., Lowry, R.B., 1988. Genetic disorders in children and young adults: a population study. Am. J. Hum. Genet. 42, 677–693.

Blacker, D., Wilcox, M.A., Laird, N.M., Rodes, L., Horvath, S.M., Go, R.C., Perry, R., Watson, B. Jr., Bassett, S.S., McInnis, M.G., Albert, M.S., Hyman, B.T., Tanzi, R.E., 1998. α-2 macroglobulin is genetically associated with Alzheimer disease. Nat. Genet. 19, 357–360.

Blackwelder, W.C., Elston, R.C., 1982. Power and robustness of sib-pair linkage tests and extension to larger sibships. Commun. Stat. Theory Methods 11, 449–484.

Blum, M., Grant, D.M., McBride, O.W., Heim, M., Meyer, U.A., 1990. Human arylamine *N*-acetyltransferase genes: isolation, chromosomal localization, and functional expression. DNA Cell Biol. 9, 193–203.

Brookes, A.J., Lehväslaiho, H., Siegfried, M., Boehm, J.G., Yuan, Y.P., Sarkar, C.M., Bork, P., Ortigao, F., 2000. HGBASE: a database of SNPs and other variations in and around human genes. Nucleic Acids Res. 28, 356–360. http://hgbase.interactiva.de/about.html, also [website].

Brown, D.M., Provoost, A.P., Daly, M.J., Lander, E.S., Jacob, H.J., 1996. Renal disease susceptibility and hypertension are under independent genetic control in the fawn-hooded rat. Nat. Genet. 12, 44–51.

Caraco, Y., 1998. Genetic determinants of drug responsiveness and drug interactions. Ther. Drug Monit. 20, 517–524.

Carey, G., Williamson, J., 1991. Linkage analysis of quantitative trait loci: increased power by using selected samples. Am. J. Hum. Genet. 49, 786–796.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N.,

Lane, C.R., Lim, E.P., Kalayanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., Lander, E.S., 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat. Genet. 22, 231–238.

Clark, A.G., Weiss, K.M., Nickerson, D.A., Taylor, S.L., Buchanan, A., Stengard, J., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E., Sing, C.F., 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am. J. Hum. Genet. 63, 595–612.

Collins, F.S., Guyer, M.S., Charkravarti, A., 1997. Variations on a theme: cataloging human DNA sequence variation. Science 278, 1580–1581.

Cooper, G.S., Umbach, D.M., 1996. Are vitamin D receptor polymorphisms associated with bone mineral density? A meta-analysis. J. Bone Miner. Res. 11, 1841–1849.

Criswell, L.A. et al., 2000. *PARP* alleles and SLE: failure to confirm association with disease susceptibility. J. Clin. Invest. 105, 1501–1502.

Daiger, S.P., Sullivan, L.S., Rossiter, B.J.F., 2000. Retinal information network. http://www.sph.uth.tmc.edu/retnet/, [web site].

Dipple, K.M., McCabe, E.R.B., 2000. Phenotypes of patients with "simple" Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. Am. J. Hum. Genet. 66, 1729–1735.

Dow, D.J., Lindsey, N., Cairns, N.J., Brayne, C., Robinson, D., Huppert, F.A., Paykel, E.S., Xuereb, J., Wilcock, G., Whittaker, J.L., Rubinsztein, D.C., 1999. α-2 macroglobulin polymorphism and Alzheimer disease risk in the UK. Nature Genet. 22, 16–17, 21–22.

Eaves, L., Meyer, J., 1994. Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. Behav. Genet. 24, 443–455.

Epstein, M.P., Duren, W.L., Boehnke, M., 2000. Improved inference of relationship for pairs of individuals. Am. J. Hum. Genet. 67, 1219–1231.

Evans, D.A.P., Manley, K.A., McKusick, V.A., 1960. Genetic control of isoniazid metabolism in man. Br. Med. J. 2, 485–491.

Falconer, D.S., 1989. Introduction to Quantitative Genetics. 3rd edn. Wiley, New York.

Gonzalez, F.J., Skoda, R.C., Kimura, S., Umeno, M., Zanger, U.M., Nebert, D.W., Gelboin, H.V., Hardwick, J.P., Meyer, U.A., 1988. Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. Nature 331, 442–446.

Greenberg, D.A., 1993. Linkage analysis of necessary disease loci versus susceptibility loci. Am. J. Hum. Genet. 52, 135–143.

Gross, C., Musiol, I.M., Eccleshall, T.R., Malloy, P.J., Feldman, D., 1998. Vitamin D receptor gene polymorphisms: analysis of ligand binding and hormone responsiveness in cultured skin fibroblasts. Biochem. Biophys. Res. Commun. 242, 467–473.

Halushka, M.K., Fan, J.B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R., Chakravarti, A., 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat. Genet. 22, 239–247.

Hayashi, S., Watanabe, J., Kawajiri, K., 1992. High susceptibility to lung cancer analyzed in terms of combined genotypes of *P450IA1* and mu-class glutathione S-transferase genes. Jpn. J. Cancer Res. 83, 866–870.

Hein, D.W., Grant, D.M., Sim, E., 2000. Arylamine *N*-acetyltransferase (EC 2.3.1.5). http://www.louisville.edu/medschool/pharmacology/NAT.html/, [web site].

Idle, J.R., Smith, R.L., 1979. Polymorphisms of oxidation at carbon centers of drugs and their clinical significance. Drug Metab. Rev. 9, 301–317.

Ingelman-Sundberg, M., Oscarson, M., McLellan, R.A., 1999. Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment. Trends Pharmacol. Sci. 20, 342–349.

Jacob, H.J., Lindpaintner, K., Lincoln, S.E., Kusumi, K., Bunker, R.K., Mao, Y.P., Ganten, D., Dzau, V.J., Lander, E.S., 1991. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell 67, 213–224.

Jorde, L.B., 1995. Linkage disequilibrium as a gene-mapping tool. Am. J. Hum. Genet. 56, 11–14.

Kalow, W., Bertilsson, L., 1994. Interethnic factors affecting drug response. Adv. Drug Res. 23, 1–53.

Kruglyak, L., 1997. What is significant in whole-genome linkage disequilibrium studies? Am. J. Hum. Genet. 61, 810–812.

Kruglyak, L., 1999. Genetic isolates: separate but equal? Proc. Natl. Acad. Sci. U.S.A. 96, 1170–1172.

Lander, E.S., Schork, N.J., 1994. Genetic dissection of complex traits. Science 265, 2037–2048.

Leeder, J.S., 1998. Mechanisms of idiosyncratic hypersensitivity reactions to antiepileptic drugs. Epilepsia 39 (Suppl 7), S8–S16.

McCarthy, M.I., Kruglyak, L., Lander, E.S., 1998. Sib-pair collection strategies for complex diseases. Genet. Epidemiol. 15, 317–340.

Morrison, N.A., Qi, J.C., Tokita, A., Kelly, P.J., Crofts, L., Nguyen, T.V., Sambrook, P.N., Eisman, J.A., 1994. Prediction of bone density from vitamin D receptor alleles. Nature 367, 284–287.

Nebert, D.W., 1991. Proposed role of drug-metabolizing enzymes: regulation of steady-state levels of the ligands that effect growth, homeostasis, differentiation, and neuroendocrine functions. Mol. Endocrinol. 5, 1203–1214.

Nebert, D.W., 1997a. Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? Am. J. Hum. Genet. 60, 265–271.

Nebert, D.W., 1997b. Pharmacogenetics: 65 candles on the cake. Pharmacogenetics 7, 435–440.

Nebert, D.W., 1999. Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? Clin. Genet. 56, 247–258.

Nebert, D.W., 2000. Suggestions for the nomenclature of human alleles: relevance to ecogenetics, pharmacogenetics and molecular epidemiology. Pharmacogenetics 10, 279–290.

Nebert, D.W., Gonzalez, F.J., 1987. P450 genes: structure, evolution and regulation. Annu. Rev. Biochem. 56, 945–993.

Nebert, D.W., McKinnon, R.A., Puga, A., 1996. Human drug-metabolizing enzyme polymorphisms: effects on risk of toxicity and cancer. DNA Cell Biol. 15, 273–280.

Nickerson, D.A., Taylor, S.L., Weiss, K.M., Clark, A.G., Hutchinson, R.G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E., Sing, C.F., 1998. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat. Genet. 19, 233–240.

Ohnishi, Y., Tanaka, T., Yamada, R., Suematsu, K., Minami, M., Fujii, K., Hoki, N., Kodama, K., Nagata, S., Hayashi, T., Kinoshita, N., Sato, H., Sato, H., Kuzuya, T., Takeda, H., Hori, M., Nakamura, Y., 2000. Identification of 187 single nucleotide polymorphisms (SNPs) among 41 candidate genes for ischemic heart disease in the Japanese population. Hum. Genet. 106, 288–292.

Oscarson, M., Ingelman-Sundberg, M., Daly, A.K., Nebert, D.W., 2000. Human cytochrome P450 (CYP) alleles. http://www.imm.ki.se/CYPalleles/, [website].

Persson, I., Johansson, I., Ingelman-Sundberg, M., 1997. In vitro kinetics of two human *CYP1A1* variant enzymes suggested to be associated with interindividual differences in cancer susceptibility. Biochem. Biophys. Res. Commun. 231, 227–230.

Pratt, S.C., Daly, M.J., Kruglyak, L., 2000. Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. Am. J. Hum. Genet. 66, 1153–1157.

Rieder, M.J., Taylor, S.L., Clark, A.G., Nickerson, D.A., 1999. Sequence variation in the human angiotensin-converting enzyme. Nat. Genet. 22, 59–62.

Risch, N.J., 1987. Assessing the role of HLA-linked and unlinked determinants of disease. Am. J. Hum. Genet. 40, 1–14.

Risch, N.J., 2000. Searching for genes in complex diseases: lessons from systemic lupus erythematosus. J. Clin. Invest. 105, 1503–1506.

Risch, N., Merikangas, K., 1996. The future of genetic studies of complex human diseases. Science 273, 1516–1517.

Risch, N.J., Zhang, H., 1995. Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science 268, 1584–1589.

Risch, N.J., Zhang, H., 1996. Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. Am. J. Hum. Genet. 58, 836–843.

Rogaeva, E.A., Premkumar, S., Grubber, J., Serneels, L., Scott, W.K., Kawarai, T., Song, Y., Hill, D.L., Abou-Donia, S.M., Martin, E.R., Vance, J.J., Yu, G., Orlacchio, A., Pei, Y., Nishimura, M., Supala, A., Roberge, B., Saunders, A.M., Roses, A.D., Schmechel, D., Grane-Gatherum, A., Sorbi, S., Bruni, A., Small, G.W., 1999. An α-2 macroglobulin insertion-deletion polymorphism in Alzheimer disease. Nat. Genet. 22, 19–22.

Rudrasingham, V., Wavrant-De Vrieze, F., Lambert, J.C., Chakraverty, S., Kehoe, P., Crook, R., Amouyel, P., Wu, W., Rice, F., Perez-Tur, J., Frigard, B., Morris, J.C., Carty, S., Petersen, R., Cottel, D., Tunstall, N., Holmans, P., Lovestone, S., Chartier-Harlin, M.C., Goate, A., Hardy, J., Owen, M.J., Williams, J., 1999. α-2 macroglobulin gene and Alzheimer disease. Nat. Genet. 22, 17–19.

Sing, C.F., Haviland, M.B., Reilly, S.L., 1996. Genetic architecture of common multifactorial diseases. Variation in the Human Genome (Ciba Foundation Symposium 197). Wiley, Chichester, pp. 211–229.

Talas, G., Brown, R.A., McGrouter, D.A., 1999. Role of phenytoin in wound healing—a wound pharmacology perspective. Biochem. Pharmacol. 57, 1085–1094.

Terwilliger, J.D., Weiss, K.M., 1998. Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr. Opin. Biotechnol. 9, 578–594.

Tsao, B.P. et al., 1999. *PARP* alleles within the linked chromosomal region are associated with systemic lupus erythematosus. J. Clin. Invest. 103, 1135–1140.

Tsao, B.P. et al., 2000. Letter to the editor. J. Clin. Invest. 105, 1501–1502.

Tsujii, H., Konig, J., Rost, D., Stockel, B., Leuschner, U., 1999. Exon-intron organization of the human multidrug-resistance protein-2 (*MRP2*) gene mutated in Dubin–Johnson syndrome. Gastroenterology 117, 653–660.

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Lipshutz, R., Chee, M., Lander, E.S., 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280, 1077–1082.

Weber, W.W., 1997. Pharmacogenetics. Oxford Univ. Press, New York/Oxford.

Wright, A.F., Carothers, A.D., Pirastu, M., 1999. Population choice in mapping genes for complex diseases. Nat. Genet. 23, 397–404.

Vesell, E.S., 1969. Recent progress in pharmacogenetics. Adv. Pharmacol. 7, 1–52.

Vogel, F., 1959. Moderne probleme der Humangenetik. Ergeb. Inn. Med. Kinderheilkd. 12, 52–125.

Yamada, R., Tanaka, T., Ohnishi, Y., Suematsu, K., Minami, M., Seki, T., Yukioka, M., Maeda, A., Murata, N., Saiki, O., Teshima, R., Kudo, O., Ishikawa, K., Ueyosi, A., Tateishi, H., Inaba, M., Goto, H., Nishizawa, Y., Tohma, S., Ochi, T., Yamamoto, K., Nakamura, Y., 2000. Identification of 142 single-nucleotide polymorphisms in 41 candidate genes for rheumatoid arthritis in the Japanese population. Hum. Genet. 106, 293–297.

Zhang, Z.Y., Fasco, M.J., Huang, L., Guengerich, F.P., Kaminsky, L.S., 1996. Characterization of purified human recombinant cytochrome P4501A1-Ile^{462} and -Val^{462}: assessment of a role for the rare allele in carcinogenesis. Cancer Res. 56, 3926–3933.

Zhu, X., McKenzie, C.A., Forrester, T., Nickerson, D.A., Broeckel, U., Schunkert, H., Doering, A., Jacob, H.J., Cooper, R.S., Rieder, M.J., 2000. Localization of a small genomic region associated with elevated ACE. Am. J. Hum. Genet. 67, 1144–1153.