

**DEPARTMENT OF BIOTECHNOLOGY  
MINISTRY OF SCIENCE & TECHNOLOGY**

**GUIDELINES TO DEVELOP PROPOSALS IN THE AREA OF PHARMACOGENOMICS**

**1. Preamble**

The publication of the human genome sequence of 3.2 billion bases in February, 2001 has ushered in challenges in relation to understanding of functions of ~31,000 genes and using these to develop medically beneficial products. The era of genomics has opened new opportunities to discover new drug targets. This would ultimately lead to development of novel drugs and therapies that will be cost effective and safe for diseases/disorders. The current therapy which is evidence based would be replaced by “customized” therapy and thus will be very specific to treat major diseases like infectious diseases, cardio vascular disorders, diabetes, neurogenetic disorders, eye diseases, haemoglobinopathies etc. The area also opened up potential commercial development of genomics research with pharmaceutical industries with the wealth of opportunities popularly known as “Pharmacogenomics”, the use of genetic analysis in the drug development process to understand the interaction between given drug a therapy and an individual genetic makeup; by using this information it is possible to design individual based drugs to reduce side effects and avoid adverse drug reaction. Globally, several pharmaceutical industries are working in this area.

Keeping in view the high priority of developing new drugs based on genomics information and also based on the recommendations during the brain storming session in August, 2000 on Post Genome Era, the Department of Biotechnology has initiated a major programme in the area of “Pharmacogenomics” under Human Genetics & Genome Analysis programme.

Accordingly, a discussion meeting on Pharmacogenomics was organized in July, 2002 under the chairmanship of Prof.G.Padmanaban. After detailed discussion, the committee has suggested to invite project proposals to assess genetic basis of the non

responder/poor responder to known drugs for treatment of tuberculosis, diabetes type-II and depression to assess genotype: phenotype association for “known drugs” in volunteers and also to assess the SNP based studies for failed drugs.

The experts also suggested to develop detailed standard guidelines to undertake such studies in India. including number of samples required for study to generate statistically significant data on responders and non-responders.

## **2. Definition of Pharmacogenomics**

It is well-known that a drug used for treatment of a disease often has differential effects on patients. These effects include the response to the drug in the amelioration of the disease condition and/or adverse reactions resulting from administration of the drug. It is now recognised that these differential effects may wholly or partially be due to differences in the genetic make-up of patients. Arg16 and Gly16 polymorphisms in  $\beta$ 2-adrenergic receptor leading to responsiveness and non-responsiveness, respectively, to albuterol, a commonly used drug for Asthma, is a well known example of differential response to a drug. Ile359Leu polymorphism in CYP2C9 gene resulting in reduced drug clearance of warfarin (an anticoagulant used in patients with heart disease) and consequent death due to brain hemorrhage arising from overdose; CYP2D6 \*4 polymorphism resulting in impaired metabolism of debrisoquin, a commonly used drug for hypertension, and a lethal lowering of BP; low activity allele(s) of thiopurine methyltransferase (TPMT) gene in lymphoblastic leukemia patients and transplant recipients under treatment with azathiopurine leading to life threatening myelosuppression and hepatic toxicity are other common examples of varied drug response as well as adverse drug reactions. Pharmacogenomics is thus the study of identification and analysis of genomic variations that affect the efficacy of a drug. Pharmacogenomic studies can potentially be predictive of an individual's drug-response or adverse reactions or susceptibility to iatrogenic disorders, and may also reveal new targets that can help in the design of new drugs.

### **3. Purpose of the Document**

Scientists in India are increasingly undertaking pharmacogenomic studies. The Department of Biotechnology (DBT) has been many receiving proposals for funding such studies. The scientific aspects of these proposals do not often address all the major issues that need to be taken into account before initiating pharmacogenomic studies. This document, therefore, provides some guidelines regarding the major scientific issues for consideration by scientists planning to undertake such studies. These issues need to be addressed in a pharmacogenomics project-proposal submitted to DBT in addition to the standard issues of reviewing background knowledge on the subject, specifying the goals and objectives of the project, possible impact of the project, etc.

### **4. Justifications for Undertaking a Pharmacogenomic Study**

It is imperative that a study undertaken in India should have national relevance. For a pharmacogenomic study to be of national relevance, it is crucial that (a) the disease under consideration should have a high prevalence in India, (b) the drug under consideration should be one of the more widely-used drugs for treatment of the disease, and (c) the proportion of patients who either do not respond to the drug or elicit adverse reactions should be high. Response to pravastatin, a commonly used cholesterol lowering drug among CVD patients is a good example for such a study. Response to this drug is linked to polymorphism in cholesterol ester transferring protein (CETP) gene, patients with B1B1 and B1B2 genotypes responding better than those with B2B2 genotype. Therefore, the following aspects should be considered before undertaking a pharmacogenomic study and relevant data should be provided in a proposal submitted for funding to DBT.

- a. Prevalence of disease in India
- b. Documented variation in drug response
- c. % of patients not responding to the drug (NR) or developing an adverse drug reaction (ADR) of clinical significance

- d. Seriousness/ Clinical significance of NR or ADR
- e. Inter-population or inter-regional variation in prevalence of the disease or NR/ADR, and
- f. Available pharmacokinetic data (including absorption, distribution, metabolism and excretion) on the drug

We also note that from basic science considerations, sometimes it may be useful to carry out pharmacogenomic studies even on diseases that are not highly prevalent in India or on discontinued (orphan) drugs for which no or scanty data may be available on NR/ADR. In such cases, it will be useful to highlight the reasons for undertaking the study.

## **5. Definition of Disease or Phenotype**

It is well-known that for many common diseases, the clinical or phenotypic manifestations are highly variable. In such cases, the disease may be potentially heterogeneous, that is, there may be different clinical subtypes of what may be generally considered as a homogeneous disease entity. The statistical power of a pharmacogenomic study design critically depends on the definition of the disease, because it is possible that different genes may impact on the different subtypes of the same disease. Therefore, it is extremely important to carefully consider whether there may be any underlying clinical heterogeneity in the manifestation of the disease. For this purpose, it may be useful to consider multiple clinical parameters and to assess these parameters quantitatively (even if on an ordinal scale).

In a proposal submitted for funding to DBT, it will be essential to address this issue and to clearly delineate in the proposal how the possibility of clinical heterogeneity will be handled in the study. We finally note that, in the case of certain drugs, a significant proportion of patients receiving medication have clinically significant side effects which necessitates withdrawal of medication. Identification of such cases is also useful.

## **6. Measurement of Response and Adverse Reactions**

Careful measurement of the response to the drug under consideration and its adverse reactions is of paramount importance to the success of a pharmacogenomics study. There are several issues: (a) even if different individuals take drugs of different brands in which the active ingredient is the same and of the same amount, variations in response may occur simply because different brands often have other components that may render bio-availability of the active ingredient to be different, (b) dosage (including variation in prescribed dose over the course of illness) and compliance can lead to differences in response, (c) dosage, response and adverse reactions may be age-dependent, (d) time to respond or elicit adverse reactions is a time-dependent phenomenon and may vary across patients, (e) when a multi-drug therapy is used, further considerations on the drug-interactions and dosages of each drug need to be given, and (f) even after discontinuing a drug, there may be relapse in some – but not all – patients; the time period to relapse may be variable across patients (that is, short-term response and long-term response may be different across patients). Further, there may be quantitative differences in response/ADR, necessitating their careful measurement on a quantitative (even if ordinal) scale.

Therefore, in proposal submitted for funding to DBT, it will be necessary to address these issues and to specify:

- a. The name of the drug and whether drug of same brand will be used (to avoid confounding effect of variable bioequivalence)
- b. Dosage and dosage-variation over the duration of treatment
- c. If a multi-drug therapy is used, the combinations of drugs and their dosages
- d. Methodology for monitoring compliance, including whether body fluid levels will be measured to assure compliance
- e. Methodology for measuring response/ADR and the scale of measurement

- f. Validation of parameters/scales used to measure response/ADR
- g. Time period during which monitoring of response/ADR will be carried out.

## **7. Choice of Genes/Genomic Regions**

Clear justifications are required for the choice of genes or genomic regions in which variations are to be examined. For most diseases and drugs, the number of genes or genomic regions may be many; hence, it is important to prioritize these genes so that variations in them can be sequentially examined. Justifications for according such priorities also need to be clearly provided. These justifications may include, among other reasons, the results of previous studies conducted in India or elsewhere. Further, it is important to provide some preliminary data on the nature and extent of variation in the chosen genes/genomic regions. Such information may be available in public-domain databases, such as dbSNP, ALFRED, etc., although these data may not pertain to Indian populations. It is desirable that some preliminary data from Indian populations be provided. Therefore, it is expected that a proposal submitted to DBT will provide information on:

- a. Genes/genomic regions to be studied, with priorities and justifications for their choice
- b. Nature (SNP, STR, etc.) and extent (genotype frequency/allele frequency/heterozygosity) of variation in the chosen genes
- c. Number of variant sites to be assayed in each of the chosen genes
- d. Methodologies to be used for assaying and scoring genomic variation

Further, it may be noted that in pharmacogenomic studies it may be crucial to study the effects of haplotypes, in addition to those of individual polymorphisms.

## **8. Unit of Study and Selection of Participants**

Most pharmacogenomic studies rely on a responder/non-responder design. In this study design, a set of patients are selected and their response to a drug is evaluated. However, alternative study designs may also be adopted. The design to be adopted in a study needs to be specified, with identification of the unit of study, such as individual patients (responders and non-responders), affected sibling pairs, etc. The criteria for the selection of a unit of study also need to be specified. Therefore, a proposal submitted to DBT shall provide information on:

- a. Study design, with justification for its choice
- b. Criteria for selection of a unit of study: (i) inclusionary (based on which recruitment of study-units will be made), and (ii) exclusionary (based on which potential study-units will be excluded from recruitment).

Another crucial issue in pharmacogenomic studies is that of population stratification. For example, if responders and non-responders are drawn from pools of individuals (such as, ethnic groups) that are genetically different, then spurious results (false positive or false negative associations) may be obtained. Therefore, it is of utmost importance to guard against this possibility in a pharmacogenomic study. The best safeguard is to choose patients from a genetically-homogeneous population. To provide some degree of validation of the results of pharmacogenomic study, it may be pertinent to assay a set of unlinked markers that are possibly unrelated to the disease or the response/ADR to the drug under consideration and show that with respect to these markers the responders and non-responders do not exhibit significant differences in genotype/allele frequencies (that are expected if there is population stratification). Similarly, response/ADR to a drug is also known to be influenced by the intrinsic physical and physiological differences between the two genders. Therefore, a proposal should carefully consider and address these issues by providing information on steps to be taken to ensure that inferences drawn are not affected by possible population stratification or gender differences.

## **9. Sample Size and Statistical Power**

Unless the sample size is adequate, there may not be sufficient statistical power to detect effects of genotypes/alleles/haplotypes on response/ADR. It is to be noted that an association study pertaining to ADR may need a much larger sample size than a study pertaining to response for attainment of the same statistical power. It is, therefore, important to assess what should be the sample size for detecting an association of a given strength for preassigned values of the level of significance and statistical power. This is not an easy task, since it requires considerable prior information, such as allele frequencies at the locus under consideration for responders/non-responders. The task becomes more difficult if the disease is clinically heterogeneous, if responses/ADRs are quantitatively variable, etc. However, it may be important to assess the sample-size requirement, even if under the simplest scenario, at the time of planning a study. The simplest scenario is: one autosomal locus with two alleles at which genotypes have been determined for a number of responders and an equal number of non-responders. Thus, under this scenario, the sample size (n) is calculated using the following formula:

$$n \text{ (each group)} = \frac{(p_0q_0 + p_1q_1)(z_{1-\alpha/2} + z_{1-b})^2}{(p_1 - p_0)^2}$$

where,

$p_1$  = proportion of genotype among responders =  $p_0 \times RR$

$p_0$  = proportion of genotype among non-responders

RR = Relative Risk =  $p_1 / p_0$

$q_1 = 1 - p_1$

$q_0 = 1 - p_0$

$z_{1-\alpha/2}$  = value of the standard normal distribution corresponding to alpha: e.g., 1.96 for 2-sided test at 0.05

$z_{1-b}$  = value of the standard normal distribution corresponding to desired power level: e.g., 0.84 for a power of 80%

Consider the following example:

Suppose,  $p_0 = 10\%$ ,  $RR = 1.8$ ,  $p_1 = (p_0)(RR) = (0.10)(1.8) = 0.18$ . Hence,

$$q_1 = 1 - p_1 = 1 - 0.18 = 0.82$$

$$q_0 = 1 - p_0 = 1 - 0.10 = 0.90$$

$z_{1-\alpha/2}$  = value of the standard normal distribution corresponding to alpha: e.g., 1.96 for a 2-sided test at level  $\alpha = 0.05$

$z_{1-\beta}$  = value of the standard normal distribution corresponding to desired power level: e.g., 0.84 for a power of 80%.

Then, the sample size n (for each group) is calculated as:

$$n = \frac{[(0.1)(0.9) + (0.18)(0.82)] [1.96 + 0.84]^2}{(0.18 - 0.10)^2}$$

$$= \frac{(0.2376) (7.84)}{0.0064} = 291.06 = 291 \text{ (approx.)}$$

For the above values of  $p_0$  (=0.10),  $\alpha$  (=0.05) and power (=0.80), one can compute sample size requirements for different values of RR, as in the table below.

Postulated RR	Required sample size (n) per group
1.2	3834
1.3	1769
1.5	682
1.8	291
2.0	196
2.5	97
3.0	59

From the above equation, it is also easy to see that the statistical power can be calculated for a given sample size (n) for plausible values of  $p_0$ , alpha and RR. The above formula will require slight modifications when there are multiple alleles.

In a proposal submitted to DBT, it is expected that the statistical power of the proposed sample size be calculated for plausible values of the genotype frequency ( $p_0$ )

among non-responders and for clinically-relevant values of the relative risk (RR). Alternatively, the sample size requirement to attain a reasonable statistical power (say, 80%) may be presented. Plausible values of genotype frequencies may be obtained from pilot studies or from the literature, if available. Plausible values of genotype frequencies may be obtained from pilot studies or from the literature, if available.

## **10. Statistical Analysis**

Statistical analysis of the data for determining the effect of genotypes/alleles/haplotypes on response/ADR will depend on the study design, the nature (binary, ordinal, quantitative) of the variables used to measure response/ADR and on covariates (age, dosage, etc.). Such analyses may include simple association analysis in a contingency table with estimation of relative risk or odds ratio, analysis of variance with adjustment for covariates, variance-components analyses, proportional-hazards analysis, etc. It is imperative that a proposal should clearly outline the nature of statistical analyses to be performed on the data.

## **11. Ethical Considerations**

All proposals submitted to the DBT must adhere to the "*Ethical Policies on the Human Genome, Genetic Research & Services*", Department of Biotechnology, New Delhi.