

Draft proposal January 2, 2004

PGRN Pharmacogenetic Discovery Project

Howard L. McLeod, M. Eileen Dolan, Judy Badner, Michael Province

CREATE Pharmacogenetics Research Network, St. Louis, MO and PAAR Pharmacogenetics Research Network, Chicago, IL

A large number of genes are likely to influence the toxicity and response of an individual medication. Our current candidate gene strategies have made major advances towards understanding this variability, but we have not identified all genes important for drug action. In particular, the pharmacodynamic components of response prediction are particularly unclear. Therefore, an approach is needed in which no *a priori* assumptions about candidate genes are made. This would complement current candidate genes strategies by providing a novel approach to hypothesis generation and to a broader understanding of the genes involved in drug response.

Genome-wide mapping approaches have been applied in the search for genes responsible for clinically important diseases. Strategies using linkage and association studies have been responsible for the identification of the majority of disease genes to date. However, pharmacogenetic discovery for chemotherapeutic agents and other commonly used medications are not amenable to standard pharmacogenetics strategies. Limitations to using these strategies to identify genes important in sensitivity to chemotherapy are that the occurrence of cancer in multiple generations of the same family is rare and one cannot administer chemotherapy to healthy volunteers. In addition, many of the phenotypes of interest are not amenable to these classical approaches due to technical issues (e.g. phenotype is only observed at the cellular level).

Therefore, experimental models systems are needed to address these issues.

The EBV-transformed B-Lymphoblastoid cell lines derived from the Centre d' Etude du Polymorphisme Human (CEPH) panel include a large number of individuals (> 700) from 55 multi-generational families. These cell lines can be grown using standard cell culture facilities, allowing for the measurement of *in vitro* phenotypes. As these cell lines have been used for a myriad of genome mapping and single nucleotide polymorphism evaluations, a large amount of marker data is freely available. Over 13,000 microsatellite markers have been performed using DNA from these families and can be downloaded from the CEPH database. As this is an uncurated dataset, there is a need for great care with annotation of gene location, as multiple names have been used for the same microsatellite markers. There is also a great deal of redundancy of marker information in the database, which will greatly influence the results of linkage analysis. Both the SNP consortium and other large single nucleotide polymorphism efforts have generated over 3,000 SNP markers from these family members into the public databases. Therefore, there is a great deal of genetic information already available on which linkage analysis and association testing can be performed.

PGRN Proposal

In order to further Pharmacogenetic discovery we propose the following research plan:

1. Each PGRN node will genotype 20 variants and/or tag SNPs from candidate genes in at least 20 CEPH families (Family 1334, 1340, 1341, 1344, 1345, 1346, 1347, 1349, 1350, 1358, 1362, 1375, 1408, 1416, 1420, 1444, 1447, 1454, 1459, 1463). These cell lines were chosen using two distinct criteria. Firstly the statistical geneticists involved in this proposal have identified a number of these cell lines as the most informative for the context of heritability analysis, due to the family structure and number of sibships. Secondly, these families represent the families chosen by the International Haplotype Mapping Project and therefore will have a large number of additional genotype and haplotype information available in the near future. These twenty families consist of 277 individuals. We will request that Coriell make these core 20 families available as a PGRN core sample set.
2. Interested PGRN nodes can evaluate the functional utility of the CEPH cell lines for phenotype(s) of interest. The PGRN listserv can then be used to enhance dialogue between laboratory and statistical genetics investigators.
3. All data generated from these studies will be deposited into PharmGKB.

Output

This initiative will be a tangible contribution by the network to the greater scientific community. Our statistical geneticists and computational scientists will coalesce the existing genetic information, enabling greater ease for the conduct of genome wide scanning and association studies. Each node of the network will make significant contributions to the existing genetic information, with important focus on genes of high relevance to pharmacogenetics (where there is currently limited data).

This resource will also be a direct benefit to many of the nodes of the PGRN. This will also allow network investigators to develop physical tools to enhance the discovery of new pharmacogenetics markers. It will be a focal point for new collaborations within the network and will allow the PGRN bioinformatics, statistical genetics, and computational biology investigators the ability to test new hypotheses related to pharmacogenetics.