Commentary

PUBLICATION OF STAND-ALONE SINGLE NUCLEOTIDE POLYMORPHISM (SNP) DISCOVERY DATA

RONALD N. HINES

Medical College of Wisconsin, Department of Pediatrics, Birth Defects Research Center, Milwaukee, Wisconsin

(Received April 24, 2003; accepted June 5, 2003)

In the recent past, several publications have appeared in Drug Metabolism and Disposition that described the results of single nucleotide polymorphism (SNP) discovery projects. The publications were limited to only this aspect of the science and failed to include any studies that extended beyond the discovery phase. No attempt was made to elucidate the possible functional significance of the identified genetic variants, to determine whether or not a given SNP might be in linkage disequilibrium with previously identified functional variants on the same gene, or to validate the frequency of the SNPs in an adequate number of individuals from one or more ethnic groups. It is recognized that such SNP discovery projects initially represented a novel approach and were considered meritorious based on their hypothesis-generating potential. Furthermore, it also is recognized that such studies often represented intensive efforts. However, as has occurred in the past, the field has quickly moved beyond this infant phase, and it is posited that SNP discovery alone no longer meets the criteria for publication in the peer-reviewed literature. Most importantly, this is not to opine that SNP discovery is unimportant in understanding genetic variability; on the contrary, it is a significant and necessary first step.

It is suggested that future manuscript submissions to Drug Metabolism and Disposition meet certain criteria. Any SNP discovery project should be extended such that the resulting report contains information regarding functional impact, frequency, and/or linkage to previously demonstrated functional variants. The high frequency of SNPs in the human genome is consistent with most sequence variants not having a significant impact on gene function. A given SNP might be linked with a functional variant and thus be useful as a marker, but this should be demonstrated and quantified. An exception might be made for SNPs that would have an obvious function, e.g., variants that result in a premature stop codon within the coding sequence of a gene or that alter key residues in consensus splice donor or acceptor sites. Even in these cases, the most rigorous approach would be to definitively demonstrate loss of function, because we are sometimes surprised by the ability of nature to compensate for such changes.

As has been demanded for publications using other approaches, the data presented should represent a substantial advancement for the field. As indicated above, the existence of SNPs and their overall frequency in the human genome has been well established. Furthermore, given the variability in human phenotypes, such genetic variability as an underlying contributing factor is hardly surprising. The challenge is to identify those SNPs that are wholly or partially causative for a given phenotype or are linked to another causative locus.

The data and/or the experimental approach should be novel and informative. Certainly, it is hoped the genetic variability being reported is indeed novel. However, the data on genetic variability should be accompanied with data allowing one to judge impact. Furthermore, although more refined SNP discovery and validation approaches are being developed and/or refined, publication of such methodology alone may not be appropriate for Drug Metabolism and Disposition. A useful analogy can be made to the situation that occurred when DNA sequencing and cloning technologies converged 15 to 20 years ago, allowing rapid advances. If one peruses the literature from that time, one will find many examples of publications containing entire cDNA or gene sequences. With the development of publicly accessible databases and the growing ability of multiple laboratories to take advantage of emerging technologies, the novelty and overall significance of such information quickly became minimal. Thus, journals altered their editorial policies, refusing to publish studies focused solely on determining primary sequence information. They asked that such information be deposited into one of the publicly accessible databases, the accession number be referenced in the publication, and that the publication itself focus on the functional and mechanistic importance of the cDNA and/or gene. Certainly, limited sequence data can be and are included in current publications, but are used to illustrate functional aspects of the information. With the establishment of the publicly accessible SNP databases (e.g., GeneSNPs at http://www.genome.utah.edu/gensnps/ and dbSNP at http://www.ncbi.nlm.nih.gov/) and the inclusion of information on genetic variability with many gene sequence records, it would seem that the science of exploring genetic variability has reached a similar stage.

In summary, Drug Metabolism and Disposition should enthusiastically encourage SNP discovery on genes encoding proteins involved in drug disposition. The journal should ask that the authors include SNP accession numbers to ensure the public availability of the published information on genetic variability. However, before considering a manuscript focused on genetic variability for publication, Drug Metabolism and Disposition also should ask that SNP discovery projects be extended to include some aspect of functional genomics, a validation of the frequency of the sequence variant in one or more subpopulations, and encourage haplotype analysis where appropriate.

Address correspondence to: Dr. Ronald N. Hines, Medical College of Wisconsin, Department of Pediatrics, Birth Defects Research Center, 8701 Watertown Plank Road, Milwaukee, WI 53226. E-mail: rhines@mail.mcw.edu

1 Abbreviation used is: SNP, single nucleotide polymorphism.