Published in final edited form as:

Clin Pharmacol Ther. 2012 October ; 92(4): 425–427. doi:10.1038/clpt.2012.115.

Clinical translation of cell-based pharmacogenomic discovery

Nancy J. Cox, Eric R. Gamazon, Heather E. Wheeler, and M. Eileen Dolan

Department of Medicine, Committee on Clinical Pharmacology and Pharmacogenomics,
University of Chicago, Chicago, IL

The use of cell-based models has emerged as a promising means to discover and validate pharmacologic phenotype-genotype relationships. The availability of large-scale genome studies in both human and model systems is now allowing us unprecedented opportunity to understand how well cell-based models identify clinically relevant genetic variants associated with drug response and toxicity. Here, we review these studies and the emerging translational information.

The use of human Epstein-Barr virus transformed lymphoblastoid cell lines (LCLs) has emerged as a promising model system in the study of the genetics of drug response because the cells provide a cost-effective testing system in which environmental factors such as concomitant medications can be controlled. The advantages of *HapMap* LCLs include a controlled, unlimited resource to evaluate toxic drugs without the in vivo confounders such as concomitant medications. Most importantly, publicly available genotype and sequencing data is available, to allow for genome-wide association studies between HapMap/1000 Genomes variants and pharmacologic phenotypes measured in the LCLs after drug or radiation treatment such as cell growth inhibition, changes in gene expression, intracellular concentration of drug/metabolite and/or apoptosis. Limitations of the model are that they represent one cell type (B-lymphoblasts), they are subject to in vitro confounders and that most CYP450 genes are not expressed and therefore the model is most useful for identifying genes and genetic variants within drug transporter and pharmacodynamic genes as opposed to pharmacokinetic genes. A recent review has described studies from a number of groups using preclinical cell-based models have taken top genome-wide association study (GWAS) hits from these models and validated them in prospective clinical trials (1). For example, top SNPs from a GWAS of cisplatin-induced cytotoxicity in LCLs were found to be associated with overall survival in patients treated with platinum for non-small and small cell lung cancer and a top SNP from a GWAS of carboplatin-induced cytotoxicity in LCLs was shown to be associated with outcome in ovarian cancer patients treated with carboplatin-based induction chemotherapy. In addition, variation in the expression of *FKBP5* was found to associate with cytarabine-induced cytotoxicity in a genome-wide screen in LCLs and follow-up clinical studies showed SNPs within *FKBP5* associated with both event-free and overall survival in pediatric AML patients treated with cytarabine.

Moreover, the cell-based model has been utilized to generate information on associations between germline variation and a wide variety of functional phenotypes, including gene expression and miRNA levels. Comprehensive eQTL maps have been generated in LCLs to enable the identification of *cis*- and *trans*- effects of (pharmacologic) trait-associated SNPs on gene expression (2). Such maps may facilitate the functional interpretation of findings from GWAS studies of drug response phenotypes in both cell-based and clinical studies. These studies led to the finding that SNPs associated with chemotherapeutic-induced cytotoxicity in LCLs are enriched in expression quantitative trait loci (eQTLs) (3). Since most pharmacogenetic studies previous to GWAS were focused on variation in coding regions of known candidate genes, this was an important finding since it opened up the possibility that SNPs in introns or intergenic regions associated with gene expression

contributed significantly to variation in pharmacologic phenotypes. Furthermore, connections between pharmacologically important variants and eQTLs may lay the basis for understanding the mechanisms underlying the observed associations with potentially important implications for individualized therapy. One consideration in the use of eQTLs identified in LCLs for pharmacogenetic studies is that a proportion of eQTLs may be tissue-specific and therefore LCLs may provide a limited subset of functional SNPs of relevance to toxicity (eg. bladder or renal toxicity, peripheral neuropathy). However, a recent study (4) estimated that 65–70% of the *cis* eQTLs identified in LCLs are also present in fibroblasts and T cells after adjusting for statistical power. This report is consistent with the status of current eQTL mapping studies as typically underpowered, which results in an underestimation of shared eQTLs across tissues.

Although it is indeed gratifying to observe that the very top signals from cell-based models can be validated in prospective clinical trials, it is critical to understand how well the overall genetic architecture of the response to chemotherapy can be captured by the cell-based models. Thus, it is important to compare the results of genetic studies in the cell-based models to studies conducted in human clinical trials beyond just the top few signals. A first approach to such a comparison might be to ask whether SNPs with nominally significant associations to cytotoxicity in studies in cell-based models are enriched among the top SNPs (e.g. those at least nominally significant) showing association to outcomes or adverse events in human clinical trials. The genome-wide criteria for significance in GWAS is widely accepted to be on the order of 5×10^{-8} and there are generally a large number of individual SNPs that are nominally associated ($p < 0.05$), and thus, this type of enrichment study fulfills the criteria of looking for overlapping signals over a much larger part of the genome. For example, when comparing a GWAS from patients and a GWAS from LCLs given the same drug, SNPs may not overlap at a stringent threshold such as $p < 10^{-8}$, but it is possible that there might be significant overlap of SNPs at a more relaxed threshold such as $p < 10^{-2}$ (Figure 1). To determine if this overlap is more than expected by chance, enrichment analysis through random sampling can be performed (Figure 2). Such enrichment studies must be carefully designed, however, to insure that reliable inferences can be drawn; for example, it can be important to condition on allele frequencies and distance to nearest gene when making such comparisons. It is also important to assess consistency of the direction of effects. For enrichment studies, negative controls could include identical analysis of the potential overlap of SNPs associated with drugs that are *not* used in the clinical study. Significant enrichment among a relatively large number of top SNPs is consistent with the hypothesis that the cell-based model and the human phenotypes share at least some part of their genetic architecture.

It should be noted that evidence for shared genetic architecture will invariably involve large numbers of SNPs with small effects; only a small minority of this architecture is likely to have direct clinical relevance. But cumulatively, large numbers of common variants with small effects can account for substantial heritability. For example, we identified significant overlap between a GWAS of paclitaxel-induced cytotoxicity performed in 247 LCLs derived from the HapMap Project and a GWAS of peripheral sensory neuropathy in breast cancer patients ($n=859$) treated with paclitaxel in the Cancer and Leukemia Group B (CALGB) 40101 trial (5). We observed an enrichment of LCL cytotoxicity-associated SNPs in the peripheral neuropathy-associated SNPs from the clinical trial with concordant allelic directions of effect (empirical $P = 0.007$). We did not observe any enrichment when evaluating either capecitabine- or carboplatin-induced cytotoxicity SNPs, which were tested as negative controls. To gain mechanistic insight into the overlapping SNPs, one can evaluate functional SNPs, such as eQTLs. Of the 24 SNPs that overlap between the clinical trial ($P < 0.05$) and the preclinical cytotoxicity study ($P < 0.001$), 19 are expression quantitative trait loci (eQTLs), which is a significant enrichment of this functional class

(empirical $P = 0.05$). For functional validation, differential sensitivity to the chemotherapeutic drug and/or differential downstream target gene expression can be measured after inhibition of the eQTL host gene by siRNA. Genome-scale loss of function screens promise to enable high-throughput validation of pharmacologically relevant targets to chemotherapeutic agents.

This significant LCL enrichment result is a critical observation, since it implies that generating a good understanding of the genetic architecture of the cell-based model will inform our understanding of the genetic architecture of the human response to chemotherapy. Better understanding of this shared biology can, of course, have direct clinical utility (as for example, it is that shared biology responsible for the fact that some of the most highly significant SNPs from the cell-based models can be validated in prospective clinical trials), but should have other value as well. Having insight into the overall genetic architecture might well help in identifying novel drug combinations that will be effective in chemotherapy. Similarly, both the analysis of data from past clinical trials and the design of future studies could benefit from a solid understanding of the genetic architecture of response to chemotherapy.

Acknowledgments

The authors are funded by the Pharmacogenomics of Anticancer Agent Research Group U01GM61393 and NIH/NCI Cancer Biology Training grant T32CA09594 (HEW).

References

1. Wheeler HE, Dolan ME. Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation. *Pharmacogenomics*. Jan; 2012 13(1):55–70. [PubMed: 22176622]
2. Duan S, Huang RS, Zhang W, Bleibel WK, Roe CA, Clark TA, et al. Genetic architecture of transcript-level variation in humans. *Am J Hum Genet*. May; 2008 82(5):1101–13. [PubMed: 18439551]
3. Gamazon ER, Huang RS, Cox NJ, Dolan ME. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proc Natl Acad Sci U S A*. May 18; 2010 107(20):9287–92. [PubMed: 20442332]
4. Ding J, Gudjonsson JE, Liang L, Stuart PE, Li Y, Chen W, et al. Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. *Am J Hum Genet*. Dec 10; 2010 87(6):779–89. [PubMed: 21129726]
5. Wheeler, HE.; Gamazon, ER.; Wing, C.; Njiaju, UO.; Njoku, C.; Baldwin, RM., et al. Unpublished Data

Figure 1. Comparison of hypothetical patient and LCL GWAS results

Green dots correspond to SNPs with the same directions of effect and p-values less than 0.01 (blue lines) in both GWAS. Red lines correspond to the typical genome-wide significance threshold of $p < 5 \times 10^{-8}$.

Figure 2. Overlap enrichment analysis

To determine if the overlap between top SNPs in a patient and LCL GWAS is greater than expected by chance, an expected distribution can be generated taking into account the minor allele frequencies and genomic locations of the observed overlap SNPs. This hypothetical example shows the distribution of patient/LCL overlap SNPs (same direction of effect and $p < 0.01$) in 1000 permutations of the patient GWAS results. The dot is the observed SNP overlap and represents a significant enrichment (empirical $p < 0.001$) of phenotype-associated SNPs from the LCL GWAS in the patient GWAS.