Published in final edited form as:

Curr Pharm Des. 2009 ; 15(32): 3782–3795.

Use of Cell Lines in the Investigation of Pharmacogenetic Loci

Wei Zhang¹ and M. Eileen Dolan^{1,2,3,*}

¹ Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, IL 60637, USA

² Committee on Clinical Pharmacology and Pharmacogenomics, The University of Chicago, Chicago, IL 60637, USA

³ Cancer Research Center, The University of Chicago, Chicago, IL 60637, USA

Abstract

Drug response and toxicity, complex traits that are often highly varied among individuals, likely involve multiple genetic and non-genetic factors. Pharmacogenomic research aims to individualize therapy in an effort to maximize efficacy and minimize toxicity for each patient. Cell lines can be used as a model system for cellular pharmacologic effects, which include, but are not limited to, drug-induced cytotoxicity or apoptosis, biochemical effects and enzymatic reactions. Because severe toxicities may be associated with drugs such as chemotherapeutics, cell lines derived from healthy individuals or patients provide a convenient model to study how human genetic variation alters response to these drugs that would be unsafe or unethical to administer to human volunteers. In addition to the traditional candidate gene approaches that focus on well-understood candidate genes and pathways, the availability of extensive genotypic and phenotypic data on some cell line models has begun to allow genome-wide association (GWA) studies to simultaneously test the entire human genome for associations with drug response and toxicity. Though with some important limitations, the use of these cell lines in pharmacogenomic discovery demonstrates the promise of constructing a more comprehensive model that may ultimately integrate both genetic and non-genetic factors to predict individual response and toxicity to anticancer drugs.

Keywords

drug response; toxicity; pharmacogenomics; lymphoblastoid cell lines; HapMap; single nucleotide polymorphism; gene expression

INTRODUCTION

Personalized medicine has the ambitious goal of both maximizing effective therapy and avoiding adverse effects by tailoring medical care based on one's genetic profile. Drug response is often not consistent across patients and ranges from beneficial effects to lack of efficacy to the extreme of fatal adverse drug reactions (ADRs). ADRs are responsible for 5–7% of annual hospital admissions in the Western countries [1] and estimated to be between the 4th and 6th leading cause of death in the United States [2]. Of particular concern are chemotherapeutic drugs because of their narrow therapeutic index. However, the genetic variants that influence clinical responses to these anticancer drugs remain largely unknown. Hence, identifying patients at risk for severe ADRs to chemotherapeutic agents prior to treatment is an important

*Corresponding author: M. Eileen Dolan, Section of Hematology/Oncology, Department of Medicine, 5841 S Maryland Ave. MC 2115, The University of Chicago, Chicago, IL 60637, USA. Tel: edolan@medicine.bsd.uchicago.edu; Tel: 773-702-4441; Fax: 773-702-0963.

area of research and will improve current oncology practice. We will focus this review on the discovery of pharmacogenetic loci that may predict toxicities to drugs using EBV (Epstein-Barr Virus) immortalized lymphoblastoid cell lines (LCLs), with particular emphasis on the use of the International HapMap Project [3,4] samples.

1. Pharmacogenetic and pharmacogenomics

Traditionally, pharmacogenetic studies have focused on a handful of well-recognized pharmacology-related genes (e.g. drug metabolizing enzymes, transporters) to search for genetic variants such as single nucleotide polymorphisms (SNPs) that determine the variability in drug response and toxicity. For example, *UGT1A1**28 (UDP glucuronosyltransferase 1 family, polypeptide A cluster) polymorphism was identified as the major predictive pharmacogenetic marker for severe hematological toxicity (neutropenia) to irinotecan [5], a commonly prescribed anticancer agent both as a single agent or in combination therapy. More recently, with the availability of the whole-genome DNA sequences [6,7] and the data from several large scale human genetic variation projects (e.g. the International HapMap Project [3,4]), the field of pharmacogenomics now aims to examine variation in the entire human genome to identify variants that control susceptibility to therapy-induced toxicity [8,9] as well as to provide drugs to patients who may respond well.

2. Human genetic variation

Human genetic variation is reflected in the observed differences in DNA sequences between the genomes of individuals or different human populations. The most common form of genetic variation is SNPs, which have been implicated in complex traits since the very beginning of the genomic era, marked by the launch of the Human Genome Project [6,7]. To date, common SNPs have been found to be associated with various complex phenotypes such as the risks of common diseases (e.g. breast cancer [10,11], diabetes [12,13], obesity [12,14], coronary artery disease [15] and Alzheimer's disease [16]), adult height [17] and hair color [18]. It is estimated that the average proportion of nucleotide differences between a randomly chosen pair of humans lie between 1 in 1,000 and 1 in 1,500 base pairs [19,20]. Variation can be measured at both the individual level and at the population level. Of the ~0.1% of DNA that varies among individuals, approximately 85–90% of genetic variation is found within three continental groups (Asians, Europeans and Africans), and only an additional 10–15% of variation is found between any two populations [20–22].

In addition to SNPs, there are other forms of genetic variations in the human genome. For example, copy number variants (CNVs), which can be deletions or amplifications of a genomic region, also account for a substantial fraction of natural genetic variation [23,24]. Since many CNVs result in differential levels of gene expression, these structural variants may contribute to a significant proportion of phenotypic variation [25]. Like SNPs, CNVs in the human genome have also been associated with the susceptibility to various human diseases [26]. Because of the complexity of the human genome, to identify the genomic contributors to the common disorders and complex traits, it will be necessary to systematically explore both SNPs and other genetic variations (e.g. CNVs) in association studies.

3. Drug response and toxicity is a heritable, complex trait

Individual response to therapeutic treatment has been shown to be a heritable trait that may be partially attributed to genetic diversity. For example, it was estimated that the heritability for susceptibility to cisplatin-induced cytotoxicity was 0.47, indicating sensitivity to the cytotoxic effects of cisplatin is under appreciable genetic influence [27]. Furthermore, there is evidence that susceptibility to cisplatin-induced cytotoxicity is likely due to multiple loci, with low locus-specific heritability contributing to the trait. At the same time, the variability of individual drug response has been observed by clinical practitioners among different human populations, age

groups and genders [28–30]. Therefore, drug response and toxicity are likely complex traits, which may be determined by multiple genes and possibly other non-genetic (e.g. environment, age, diet) factors.

ADVANTAGES OF CELL LINE-BASED MODELS

1. Challenges of studying toxic drugs

The majority of drugs currently on the market to treat cancer cause adverse side effects, precluding the use of normal, healthy volunteers for genetic studies. For example, cisplatin, a platinating agent commonly used to treat many human cancers, has been associated with platelet activation *in vitro* [31], and there have been reports of serious vascular toxicity [32] as well as nephrotoxicity, neurotoxicity and ototoxicity [33,34] associated with its use in the clinic. Furthermore, many anticancer drugs present a narrow therapeutic index, indicating that small changes in dosage could cause unacceptable toxic response which in extreme cases could be fatal. Moreover, toxicity is observed earlier than the therapeutic effect, therefore, toxic effects represent a major endpoint for pharmacodynamic studies of anticancer agents [35]. Clinically, patients taking these drugs require constant monitoring so that the dosage can be adjusted to assure uniform and safe results.

2. Advantages of cell line-based models

Use of cell line-based models controls for certain non-genetic factors such as administration time and diet, because cells can be grown and treated with drugs under the same condition. Different dosages of a drug (even at a concentration that may cause potential cell death) can be administered to cell lines to observe their response. Compared to clinical trials on human subjects, it is also much easier to repeat experiments using cell lines. In addition, extensive genotypic (e.g. SNP genotypes) data on some commercially available cell lines such as the International HapMap LCL samples have been made public. Particularly, the commercial availability of the HapMap cell lines through the Coriell Institute for Medical Research (Camden, NJ) (<http://www.coriell.org>) has allowed researchers in both human genetics and pharmacogenomics to accumulate various phenotypic data including gene expression and drug response [36] as well as other useful annotations (e.g. positions under recent positive selection [37]) on the same cell lines.

LYMPHOBLASTOID CELL LINE RESOURCES

1. Available genetic variation datasets for LCLs

SNP genotypes for ~1.8 million SNP markers (May, 2003) for LCLs derived from 48 CEPH (Centre d'Etude du Polymorphisme Humain) families [38] can be downloaded from the SNP Consortium database of the SNP Consortium Linkage Map Project (<http://snpdata.cshl.edu>) [39]. Most SNPs in this database are clustered in very closely linked sets (two or three SNPs within 100 kb), with an average intercluster distance of approximately 3 Mb [40]. Samples from some of these CEPH families (e.g. CEPH 1362, 1408) were included in the International HapMap Project [3,4]. Therefore, more extensive genetic variation data are available for those LCLs.

The International HapMap Project [3,4] aims to develop a human haplotype map that catalogues the common patterns of human DNA sequence variation. The HapMap Phase 1/2 Project released a human haplotype map of over 3.1 million SNPs [41] on a panel of 270 EBV-transformed LCLs derived from apparently-healthy individuals of three major continental populations (CEU: 30 CEPH parents-offspring trios from Utah, USA; YRI: 30 Yoruba parents-offspring trios from Ibadan, Nigeria; 45 unrelated Han Chinese from Beijing, China and 45 unrelated Japanese from Tokyo, Japan). In addition to the frequencies of SNP alleles, raw

genotypes in each population and other information (e.g. phased haplotypes, linkage disequilibrium information), which can be viewed and downloaded from the HapMap Project website (<http://www.hapmap.org>) [42]. During the past a few years, researchers working on these samples also generated comprehensive CNV data. The CNV data that involve gains or losses of several kilobases to hundreds of kilobases of genomic DNA among phenotypically normal individuals are publicly available through the Database of Genomic Variants (DGV) (<http://projects.tcag.ca/variation/>) [43,44]. The current DGV can be used to browse 1,447 CNV regions (~12% of the human genome) of the 270 HapMap samples screened using the Affymetrix Human Mapping 500K EA Array as well as clone-based comparative genome hybridization with a Whole Genome TilePath (WGTP) array [24]. The sample level data can be accessed from the Copy Number Variation Project of the Sanger Institute (<http://www.sanger.ac.uk/humgen/cnv/data/>) [24]. Other genetic variation data such as short-tandem-repeat polymorphism (STRP) data [45] have also been generated from these samples and are expected to be publicly available for analyses in the near future. Table 1 lists a summary of some whole-genome resources of genetic variations in LCLs.

Besides the above whole-genome genetic variation data, some ongoing deep resequencing projects (e.g. the SeattleSNPs Project at <http://pga.gs.washington.edu>) have released genotypic data of some gene targets on some HapMap samples. These will be introduced more in the section of “Current Developments and Outlook”.

2. Available expression and other phenotype datasets for LCLs

Using the Affymetrix Human Focus Array (focus array), 233 LCLs from 14 CEPH families [38] (CEPH 1333, 1340, 1341, 1345, 1346, 1347, 1362, 1408, 1416, 1418, 1421, 1423, 1424 and 1454) were phenotyped for transcriptional expression [40]. During the past a couple of years, gene expression data on the HapMap samples have been generated using different microarray platforms. Spielman *et al.* profiled 60 unrelated CEU, 41 CHB and 41 JPT samples using the focus array [46]. The same group recently generated expression data on 60 unrelated YRI, 60 unrelated CEU and 82 non-HapMap Coriell Human Variation Panel AA (African American) samples using the focus array [47]. At a smaller scale (8 CEU and 8 YRI samples), Storey *et al.*, generated expression data using the focus array [48]. In contrast, Stranger *et al.* used the Illumina BeadChips to obtain gene expression data on the entire panel of 270 HapMap samples [49]. Furthermore, our lab used the Affymetrix Human Exon Array (exon array) to obtain both exon-level (probeset) [50] and gene-level (transcript) [51] expression profiles of 87 CEU and 89 YRI samples. Unlike the focus array (a conventional 3' oligonucleotide array) and the Illumina platform, which can only measure gene-level expression, the exon array has the flexibility to measure the expression of the entire human exome. Though lacking more comprehensive comparisons, there is evidence that the exon array may provide more accurate measurements of gene expression than the conventional 3' arrays (e.g. the focus array) [52]. The above whole-genome expression data are available through the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) website (Table 1). In addition, published pharmacology-related phenotypes such as cell growth inhibition, apoptosis after drug treatment and drug metabolizing enzyme activity on both HapMap and some non-HapMap LCLs are available through the PharmGKB (Pharmacogenetics and Pharmacogenomics Knowledge Base) website (<http://www.pharmgkb.org>) [53]. Together with the genetic variation data, these resources are of tremendous value for human genetic and pharmacogenomic studies [36].

3. Human genetic studies using LCLs

Because gene expression is an intermediate phenotype between DNA sequence variation and more complicated cellular or whole-body phenotypes such as risks of common diseases and drug response, studying the genetics of gene expression regulation can provide insights into the human genome and complex traits/diseases. During the past few years, progress has been

made in dissecting the genetic architecture of gene expression using LCLs. Gene expression variation has been observed within a population (e.g. the CEPH cell lines [40,54–57]) or between continental populations (e.g. the HapMap LCLs [46,48,49,51]). In addition, gene expression differences between males and females were reported using both the CEPH [58] and HapMap [59] cell lines. Furthermore, both *cis*-acting (local) and *trans*-acting (distant) eQTLs (expression quantitative trait loci) have been mapped in the human genome [60,61]. Recently, the genetics of transcript isoform variation (alternative splicing) within the CEU samples [62,63] and between the CEU and YRI samples [50] have also been interrogated using the exon array. Because different populations and genders may have different risks of common diseases and response to therapeutic treatments (e.g. some anticancer drugs [64]), these LCLs, particularly the HapMap samples, provide a convenient model for studying the health disparities (e.g. drug response variation) between major human populations as well as between males and females. Interestingly, population differences in gene expression were observed in some of the important pharmacogenes using the HapMap LCLs [65], as maintained by the PharmGKB [53] VIP (Very Important Pharmacogenes) database. In addition, we set up a searchable database, SCAN (SNP and CNV Annotation Database, <http://www.scandb.org>), which can be useful for both gene-centric and SNP-centric queries for eQTLs using the exon array data (gene-level) on the CEU and YRI samples [51,60]. We expect to update the SCAN database to accommodate the exon-level expression data [50] and CNV data [24] in the future.

USE OF CELL LINES FOR PHARMACOGENOMIC DISCOVERY

LCLs derived from targeted populations have been utilized to evaluate genetic factors responsible for radiation sensitivity [66], sensitivity to mitomycin C in patients with Wilms' tumor [67], camptothecin-induced apoptosis [68], bleomycin-induced chromatid breaks and oxidant-induced apoptosis [69] as well as response to ionizing radiation [70]. For example, studies using LCLs indicated that cellular radiosensitivity was a general property of human cells and the patterns of cellular radiosensitivity suggested autosomal dominant inheritance [66]. In another study, LCLs were used as a control to illustrate the difference in cell death between Werner Syndrome (WRN) and WRN-deficient individuals after treatment with the DNA-topoisomerase-I-trapping drug camptothecin [68]. Generally, there are two main categories of approaches (Fig 1) to identifying genetic determinants responsible for drug response and toxicity. Each has its own advantages and limitations. Though the full strength of the LCL model could be reflected more in the whole-genome approaches (including GWA and linkage analysis), combining both the candidate and whole-genome approaches at different stages of pharmacogenomic and pharmacogenetic studies may be necessary for discovering and validating truly functional genetic variants. We will discuss these approaches and present some examples using them in this section. Fig. (1) shows an overview of these two general approaches.

1. Traditional Candidate gene approach

When the function of a gene is well-established, researchers may focus on that gene to search for functional or associated genetic variants for drug response and toxicity. Similarly, if a particular genetic variant (e.g. SNP) has been associated with drug response and toxicity, researchers may focus on that variant to study its relevance in new cohorts (e.g. by resequencing other samples). For example, some LCLs were used in an earlier study to determine the sequence diversity of CYP3A (Cytochrome P450, family 3, subfamily A) enzymes [71], which have been known to act in drug metabolism, influence circulating steroid levels and response to half of all oxidatively metabolized drugs. Another good example for candidate genes in pharmacology is the VIP list assembled by the PGRN (Pharmacogenetics Research Network) [72] members. Of the current list of 39 PGRN-VIP genes (Jan, 2009), most were annotated with evidence of either *in vitro* or *in vivo* functional effect on drug response

(<http://www.pharmgkb.org>) [53]. Some successful examples of this approach in chemotherapy include: 1) The identification of genetic polymorphisms in *TPMT* (thiopurine S-methyltransferase), which lead to decreased TPMT enzyme activity and subsequently increase 6-mercaptopurine toxicity [73]; 2) Decreased activity of *UGT1A1**28 polymorphism is associated with increased risk of neutropenia associated irinotecan treatment [5] and; 3) There is an effect of the 2677G>T/A polymorphism of *ABCB1* (ATP-binding cassette sub-family B member 1) on progression-free survival in ovarian cancer patients who are treated with a taxane/carboplatin, which is dependent on the extent of residual disease, with a better prognosis for patients with the 2677T/A allele and minimal residual disease [74]. Furthermore, using *TPMT* 719>G as an example, Jones et al. showed that HapMap LCL resources were useful tools for pharmacogenetic discovery when the candidate gene is known, by testing whether the *TPMT* 719A>G SNP could be identified as predicting *TPMT* phenotype [75]. The HapMap CEU and YRI samples were also successfully used to determine the functional consequences of genetic variants in *DCK* (deoxycytidine kinase), which is a rate-limiting enzyme in the activation of nucleoside analogs such as Ara-C (cytarabine arabinoside), gemcitabine, clofarabine, and others [76].

Because the candidate gene approach requires the information from well-known physiological, biochemical or functional knowledge of a gene, which is generally finite or sometimes not available at all, it is not suitable for the discovery of new candidate genes or variants. This limitation is especially important for pharmacogenomic studies. Because drug response and toxicity is most likely a complex trait, focusing on a single gene (or a few genes) may not allow one to see the complete picture of the genes involved in the drug response process. Therefore, a broader and unbiased approach may be needed for identifying novel pharmacogenetic loci.

2. Whole-genome approaches

Whole-genome approaches, including the GWA and linkage analysis, do not require *a priori* knowledge of a gene, which is necessary for the candidate gene approach. As mentioned in the previous section, the currently available LCL resources include genome-wide genetic variation data as well as whole-genome gene expression data. During the past several years, genetic variants and genomic regions associated with toxicity to some anticancer drugs have begun to be identified using these whole-genome approaches.

Because the CEPH LCLs include family members from three generations, early studies were able to estimate the heritability of drug response and apply linkage analysis to identify genomic regions linked to drug response. For example, the LCLs derived from 10 CEPH pedigrees were characterized for the degree of cisplatin sensitivity [27]. The authors estimated that the heritability of susceptibility to cisplatin-induced cytotoxicity to be 0.47, indicating an appreciable genetic influence. Watters et al. also used the CEPH LCLs to discover genetic determinants of 5-fluorouracil and docetaxel-induced cytotoxicity [77]. Cytotoxicity as a result of exposure to these chemotherapeutic agents was shown to be heritable. Genome-wide linkage analysis was also used to map a QTL (quantitative trait locus) influencing the cellular effects of 5-fluorouracil to chromosome 9q13–q22, and two QTLs influencing the cellular effects of docetaxel to chromosomes 5q11–21 and 9q13–q22. More recently, more than 300 CEPH LCLs (24 pedigrees) were evaluated for toxicity to daunorubicin [78] and etoposide [79]. Similarly, heritability analysis showed a significant genetic component contributing to the cytotoxic phenotypes and some linked genomic regions (e.g. 4q28.2 to 4q32.3 for daunorubicin [78]) were identified using a whole-genome linkage scan.

Recently, with the availability of a dense map of human genetic variation, the HapMap LCLs have been begun to be used in pharmacogenomic studies to identify genetic loci responsible for drug response and toxicity (Table 2). The exon array gene expression data (~9,000 expressed genes) [51] generated from the 176 CEU and YRI HapMap LCLs were used to

identify genetic variants, acting through their effect on regulating expression, associated with the cytotoxicities to some anticancer drugs, including etoposide [80], daunorubicin [81], carboplatin [82], cisplatin [83] and Ara-C [84]. A genetics-driven “triangle” approach [80] was used to integrate both gene expression and genotypic data in predicting chemotherapeutic-induced cytotoxicity. The first step of the “triangle” approach aims to identify significant SNPs (based on the HapMap SNP data) whose genotypes are associated with drug response (e.g. IC₅₀, the concentration at which 50% cell growth inhibition occurs). The second step then aims to find significant gene expression phenotypes associated with the identified SNPs from the first step (i.e. eQTLs). Finally, the third step determines whether the gene expression phenotypes from the second step are correlated with drug response. In addition to the HapMap LCLs, non-HapMap LCLs were used as a replication step to increase confidence in the genetic variants identified in the discovery phase [81,84]. These results demonstrated the usefulness of the HapMap LCLs as a model for pharmacogenomic discovery. The identified genetic variants contribute to a substantial proportion of the overall variation in cell sensitivity (e.g. cisplatin [83]). Furthermore, this general whole-genome approach using the HapMap samples can be used to elucidate the eQTLs contributing to a wide range of cellular phenotypes (e.g. apoptosis).

Furthermore, because the HapMap LCLs include samples derived from individuals of different ancestries, these cell lines can be used to study differences in drug response between human populations. These samples may also be used to compare the identified pharmacogenetic loci between populations to evaluate the contribution of genetic variants to drug response in a particular population. For example, using the HapMap LCLs, significant differences in carboplatin and daunorubicin IC₅₀s were shown when comparing the YRI and CEU samples [64]. However, this population difference in treatment-induced cytotoxicity was not seen for either cisplatin or etoposide. By comparing pharmacogenetic loci between populations, some interesting patterns were observed. For example, in a study of the pharmacogenomics of Ara-C using the HapMap YRI and CEU LCLs, population-specific signatures were secondary to either: 1) Polymorphic SNPs in one population but monomorphic in the other; 2) Significant associations of SNPs with cytotoxicity or gene expression in one population but not the other [84].

3. The use of cell lines to further study clinical findings

LCLs are also a convenient model to further study clinical findings. For example, using the HapMap LCLs, French et al. demonstrated that acquired genetic variation in leukemia cells has a stronger influence on MTXPG (methotrexate polyglutamates) accumulation than inherited genetic variation by comparing the expression and inherited SNPs between leukemia cells and the HapMap LCLs [85]. In another study, using 90 CEPH cell lines, Pottier et al. identified a SNP in the promoter of *SMARCB1* (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1), a candidate gene from a previous study in ALL (acute lymphoblastic leukemia), which alters *SMARCB1* expression [86]. Hartford et al. validated that adhesion biological pathway is implicated in secondary leukemia using a genome scan for etoposide-induced leukemogenic MLL [myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)] chimeric fusions in 15 HapMap cell lines for a previous study using patient samples [87]. More recently, in order to further address the hypothesis that cytotoxic effects of etoposide are not necessarily linked to the drug-induced leukemogenic *rMLL* (*MLL* rearrangement), Yang et al. screened 144 CEPH cell lines to identify lines with differing intrinsic sensitivity to etoposide [88]. The HapMap LCLs were also used as a validation set when exploring the genetic contribution to gene expression concordant across primary leukemia cells and normal leukocytes [89]. In an attempt to investigate the relationship between *CYP3A4* (cytochrome P450, family 3, subfamily A, polypeptide 4) copy number and expression, Lamba et al. used LCLs representing eight ethnically diverse

populations to compare with other samples (e.g. normal livers, primary and secondary liver tumors) [90].

LIMITATIONS AND POTENTIAL CONFOUNDING FACTORS

LCLs, particularly the HapMap samples, provide a useful model for pharmacogenomic discovery of cellular phenotypes induced by drugs; however, there are some important limitations and challenges that need to be considered when analyzing and interpreting the results using these samples. While some are “intrinsic,” some of these limitations could be addressed by the advancement of technologies and development of ongoing research efforts. A recent study using the HapMap samples for the genetic analysis of drug response suggested that biological noise and other confounding factors may reduce power and have the potential to create spurious association [91]. Because these cell lines have just recently been used in pharmacogenomic studies, debates exist for which drugs are suitable, how to measure drug response as well as how to account for confounding factors and interpret results.

1. Tissue type limitation

One assumption of the current cell line-based model is that LCLs can be used to represent tissues of toxicity. Genetically, LCLs have the same genetic make-up of autosomes as other tissue types in a human body. Therefore, LCLs are suitable for associations using genetic variation data. However, because of the tissue specificity of human gene expression [92], a large proportion of known genes are not reliably expressed in LCLs, including some well-known drug metabolizing enzymes and drug targets such as *EGFR* (epidermal growth factor receptor) [51]. Previous studies have estimated that LCLs reliably express ~50% of all known genes [40,51]. In addition, the current gene expression data are at mRNA-level; proteomics data will be necessary to evaluate the relationship between mRNAs and protein products in LCLs. Hence, care should be taken when making conclusions on associations involving gene expression data.

2. Small sample size and rarer genetic variants

The current pharmacogenomic discoveries made using LCLs are generally hypothesis-generating, partly because of the limited samples available for these studies. For example, there are only 60 unrelated HapMap CEU samples, which are used to represent Caucasians of northern and western European ancestry, while there are only 45 unrelated Han Chinese and 45 unrelated Japanese to represent eastern Asians. Because there are more than 10 million SNPs in the human genome [93], the relative small sample size is especially challenging when the differences in complex traits are being compared between human populations. With small sample sizes, significant difference are more likely to be spurious than with large samples. Though the Phase 2 HapMap SNP data is believed to cover approximately 25–35% of common SNP variation in the populations surveyed [41], rare genetic variants with potential biological functions may not be captured and analyzed in these samples.

3. Technical limitations

The high throughput profiling of gene expression and genotyping relies on the accuracy and efficiency of microarray platforms. Different platforms may have specific technical designs and limitations. For example, the oligonucleotide probes of the Affymetrix exon array differ from those standard 3' expression arrays (e.g. the Affymetrix focus array, U133 and U95 arrays as well as the Illumina platform) in their design, density, and coverage [46,51,94,95]. In contrast to the 3' arrays, the exon array interrogates the entire mRNA transcript and uses DNA targets. Another potential concern with the use of oligonucleotide arrays is the possibility that SNPs located within probes could affect hybridization efficiency [96] and lead to false eQTLs [97]. Earlier studies using these LCLs did not consider this potential problem, while a couple

of recent studies proposed slightly different approaches based on the dbSNP database [98] to account for this factor [99,100]. Though some results showed that the overall reproducibility of the Affymetrix oligonucleotide arrays and the Illumina BeadChip is good [101,102], more comparisons, both statistically and experimentally, will be necessary for a comprehensive evaluation of the reliability of these platforms.

4. Other potential confounding factors

The CEPH LCLs [38] (including the HapMap CEU samples) were collected in the 1980's, decades earlier than other HapMap collections [3,4]. Cell line transformation, culture techniques and protocols have evolved during the past twenty years. In addition, the issue of being in culture and/or frozen may affect some biological parameters. Therefore, some caution is warranted when analyzing and interpreting the data from these Caucasian samples. For example, Akey et al. suggested that batch effects could be a confounding factor when interpreting results from a study of population differences in gene expression between the CEU and ASN (Asian HapMap cell lines including the CHB and JPT) samples [46,103].

The LCLs were derived from B-lymphocytes which were transformed with EBV, a member of the herpesvirus persisting in most adults [104]. When EBV infects B-lymphocytes *in vitro*, LCLs eventually emerge that can proliferate indefinitely. The growth transformation of these cell lines is the consequence of viral protein expression [105]. EBV viral proteins have been found to induce differential gene expression in host cells. For example, LMP1 (latent membrane protein 1) was found to induce differential expression of genes that are characteristically expressed in inflammatory and hyperproliferative epidermis, including chemokines, cytokines and their receptors in cultured keratinocytes [106]. EBNA2 (Epstein-Barr virus nuclear antigen 2), a transactivator, was found to control expression of several viral and cellular genes [107]. Therefore, it is likely that EBV transformation may also affect the gene expression patterns of LCLs. However, the effects and scope of EBV transformation and its affect on expression have not been comprehensively evaluated in these cell lines. A systematic investigation on the effects of baseline EBV copy number and EBV lytic replication following treatment with drugs under study would help ascertain whether EBV influences cellular sensitivity to drug.

Another potential confounding factor is variation in growth rate of LCLs over time as well as differences between individuals. Most chemotherapeutic agents work by impairing mitosis, and therefore their efficacy depends on the rate of cell division. For example, increased proliferation was found to be associated with increased *in vitro* sensitivity to several anticancer agents in ALL [108]. Therefore, cell lines growing faster are likely to be more sensitive to chemotherapy. Though results from one recent study showed limited evidence for the heritability of growth rate [91], more comprehensive evaluation (e.g.—a GWA study between growth rate and SNPs or an evaluation of growth rate in large pedigrees) may be necessary to illustrate the genetic contribution to this factor. If growth rate is genetic and associated with specific polymorphisms, then these polymorphisms may also be associated with drug sensitivity through this intermediate phenotype. Whether growth rate is genetic or non-genetic, considering this factor would be especially useful when interpreting the pharmacogenomic discoveries of anticancer drugs that target cell division.

CURRENT DEVELOPMENTS AND OUTLOOK

Current cell-based studies for pharmacogenomic discovery have largely been utilizing the HapMap Phase 1/2 LCLs and focusing on a particular type of genetic variation, i.e. SNPs as well as gene expression (mRNA-level whole-transcript expression). As discussed in the previous section, these HapMap LCLs are a model, and therefore have some important limitations. In this section, we will highlight the current developments of some ongoing

research efforts and high-throughput technologies that will have an impact on the next wave of pharmacogenomic studies using cell lines and facilitate the realization of personalized medicine.

1. Next-generation sequencing platforms

Recently, several new sequencing instruments, referred to as “next-generation” or “massively parallel” sequencing platforms, are becoming available to researchers. In contrast to the conventional capillary-based sequencing, the biggest advantage of these next-generation sequencers is their ability to process millions of sequencing reads in parallel rather than 96 at a time. Currently, there are three commercially available and relatively mature technologies, including the Roche (454) GS-FLX sequencer (454 Life Sciences, Branford, CT), the Illumina Genome Analyzer (Illumina, San Diego, CA) and the Applied Biosystems SOLiD (Sequencing by Oligo Ligation and Detection) (Applied Biosystems, Foster City, CA). These three platforms differ significantly in terms of cost, sequencing chemistry, amplification approach and performance (e.g. length of reads). Though the accuracy of their sequencing reads and associated quality are not comprehensively evaluated, these technologies are already transforming the field of genomics research [109,110], as evidenced in the achievements of several ongoing deep resequencing projects.

2. Deep resequencing projects

Though extensive, the International HapMap Project data aimed to cover only common genetic variants ($MAF > 0.05$) [3,4]. The recently released Phase 2 data of the HapMap Project is comprised of more than 3.1 million SNPs including approximately 25–35% of common SNP variation in the populations surveyed [41]. On the other hand, potential pharmacogenetic loci associated with drug response and toxicity with smaller allele frequencies could be missed from the current dataset. The several large-scale deep resequencing projects covering the HapMap samples such as the ENCODE (Encyclopedia of DNA Elements) Project [111], the SeattleSNPs Project (<http://pga.gs.washington.edu/>) and the NIEHS (National Institute of Environmental Health Sciences) Environmental Genome Project (<http://egp.gs.washington.edu/>) [112] could be used to identify those relatively rare genetic variants in certain target genes. For example, the ENCODE Project launched by the NHGRI (National Human Genome Research Institute) aims to provide a more biologically informative representation of the human genome by using high-throughput methods to identify and catalogue the functional elements it encodes [111]. Currently, the HapMap-ENCODE Coordination (<http://www.hapmap.org/downloads/encode1.html.en>) aims to produce a dense set of genotypes across large genomic regions in 48 unrelated HapMap LCLs (16 YRI, 8 JPT, 8 CHB, and 16 CEU samples). Approximately 20,000 SNPs were identified in the 10 500kb HapMap-ENCODE regions. Some of these were already represented in the dbSNP [98,113] (a public SNP database maintained by the NCBI) database, while others were discovered during the resequencing. Another interesting project is the 1000 Genome Project (<http://www.1000genomes.org>), which is an international research effort aimed at establishing a detailed catalogue of human genetic variation using the next-generation sequencing technologies [114]. The specified goal of this project is to identify > 95% of the variants with a $MAF > 1\%$ in parts of the human genome that can be sequenced, with > 95% certainty as well as to identify > 95% of the variants with a $MAF > 0.1\text{--}0.5\%$ in exons, with > 95% certainty [115]. The sequence data generated will provide a comprehensive reference for human genetic variation. Under the first pilot project, researchers are sequencing 60 HapMap samples from three different populations at low coverage. The second pilot involves high-coverage sequencing of two trios of the CEU and YRI samples. The third pilot project aims to sequence 1,000 genes in 1,000 individuals at high coverage [115]. Prospectively, once integrated with other HapMap resources (e.g. gene expression), these efforts will greatly benefit the next wave of association studies and data mining using these cell lines [116].

3. Integrating more “-omics” data

In addition to SNPs, other genetic or non-genetic elements have roles in regulating gene expression, which in turn may influence more complex cellular or whole-body phenotypes (e.g. drug response and toxicity). For example, CNVs contribute to a substantial fraction of gene expression variation [117] and have been implicated in drug response [118,119]. Epigenetic changes such as DNA methylation at the CpG sites of the promoter regions can also affect gene expression [120]. Recently, the role of DNA methylation in regulating gene expression in drug response have been investigated [121–123]. For instance, the efficiency of the DNA repair enzyme *O*⁶-methylguanine DNA methyltransferase (*MGMT*) has been associated with an increased susceptibility to alkylating agent toxicity [124–126]. *MGMT* is transcriptionally silenced by promoter hypermethylation in several human cancers such as diffuse large B-cell lymphoma [127] as well as head and neck squamous cell carcinoma [128].

Besides gene-level or transcript-level expression, the differences in alternative splicing, i.e. transcript isoform variation, reflect another layer of complexity of gene regulation. The disruption of specific alternative splicing events has been implicated in several human genetic diseases including cancer [129–131]. Along with new technologies providing more and more “-omics” data, improved bioinformatics tools, statistical methods and systems biology approaches need to be developed so that researchers can comprehensively incorporate the complex networks of various relationships [132] (e.g. gene-environment, gene-drug, gene-gender, gene-age, *cis*- or *trans*- regulators) that affect complex traits such as drug response and toxicity. By integrating all of the relevant “-omics” data such as SNPs, CNVs, gene-level expression, exon-level gene expression (Table 1), DNA methylation changes [133] and other putative elements for gene regulation like STRPs [45], a comprehensive model for predicting drug response and toxicity could be built using these cell lines in the future.

4. More than the HapMap Phase 1/2 LCLs

More cell lines beyond the 270 HapMap Phase 1/2 LCLs will be necessary to evaluate the current findings of the pharmacogenomic discovery in replicate samples and other populations. Therefore, the recently released HapMap Phase 3 sample collection may present more opportunities for applying the same cell-based approach to identify pharmacogenetic loci. The HapMap Phase 3 cell line collection is comprised of 1,301 samples from 11 diverse populations which include the original 270 Phase 1/2 samples and 7 additional populations (ASW: African ancestry in Southwest USA; GIH: Gujarati Indians in Houston, Texas; LWK: Luhya in Webuye, Kenya; MEX: Mexican ancestry in Los Angeles, California; CHD: Chinese in Metropolitan Denver, Colorado; MKK: Maasai in Kinyawa, Kenya; TSI: Tuscans in Italy) (<http://www.hapmap.org>). Extensive SNP genotypic and frequency data (~1 million SNPs) on these samples (draft 2 released on January 7th, 2009) can be publicly accessed at the International HapMap Project website (<http://www.hapmap.org>). Except for the CEU samples, the HapMap Phase 3 cell lines are available through the NHGRI Sample Repository for Human Genetic Research at the Coriell Institute.

Many drug metabolism-related genes have distinct tissue-specific expression patterns. For example, tissue-specific expression patterns have been observed for *UGT1A* isoforms (UDP glucuronosyltransferase 1 family, polypeptide A cluster), which belong to a superfamily of microsomal enzymes responsible for glucuronidation of numerous endogenous and exogenous compounds [134]. Therefore, a more comprehensive understanding of variation of complex traits including drug response and gene expression might require cell lines derived from other tissues. For example, because the human liver is a metabolically active tissue that is important in drug metabolism as well as a number of common diseases (e.g. obesity, and diabetes), the recently published whole genome expression data on >400 normal livers [135] will be of

tremendous value for pharmacogenomic discovery if these same liver tissues could be utilized for drug response profiling.

Furthermore, the NCI-60 data on a panel of 60 different human cancer cell lines (9 tumor types) used for the NCI (National Cancer Institute) Tumor Cell Line Drug Screen (<http://dtp.nci.nih.gov/>) represent another useful resource for cancer pharmacogenomics [136]. The current data release (March, 2007) is comprised of toxicity data on more than 40,000 compounds and whole-genome molecular profile data such as gene expression [137,138], microRNA abundance [139], SNP genotypes [140] and CNVs [140] (<http://dtp.nci.nih.gov/mtargets/download.html>). Though with limitations such as the small sample sizes for the available tumor types (ranging from 2 to 9 cell lines), the collection of NCI-60 pharmacologic and molecular profile data has allowed both GWA and candidate gene approaches to identify genetic variants responsible for toxicities to anticancer drugs [141, 142] and could be used to compare with the findings from the cell lines derived from normal individuals.

5. Beyond discovery: validation and translational research

The current findings of pharmacogenetic loci through both the candidate gene and GWA approaches are largely at the discovery stage, i.e. the identification of associated genetic variants that could be either causal or linked to causal genetic elements determining drug response and toxicity. Because these discoveries are usually made using a limited number of samples (e.g. the 270 HapMap LCLs), validation of the identified pharmacogenetic loci on independent replicates or patient samples will be necessary to establish their real biological function. For example, the findings using the CEU samples could be evaluated for association in other independent Caucasian samples. Replication in other independent samples could also help eliminate any false discoveries because of the usual large number of tests being made during GWA studies. To validate their findings for gene expression relationships with toxicity to Ara-C using the HapMap CEU samples, Hartford et al replicated 2 of 3 significant relationships between gene expression (as identified using the exon array on the initial samples and real-time PCR on the second set of samples) and cytotoxicity on an additional set of 49 unrelated CEPH cell lines [84].

Finally, smooth communication and close collaboration between basic science researchers and clinical scientists need to be established to bring the pharmacogenomic discovery using cell lines into personalized medicine that attempts to benefit patients through individualized treatment. The cell lines do not consider all the confounding variables that are present in humans such as concomitant medications, diet, disease, etc. Therefore, the findings in cell lines need to be validated in a clinical setting for one to appreciate the advantages and limitations of the model.

In summary, cell line-based models have begun to allow researchers to identify genes and genetic variants that can be used to predict drug response and toxicity. The advances in sequencing and other “-omics” profiling technologies will facilitate the next wave of pharmacogenomic discovery, which in turn may help the realization of personalized medicine.

Acknowledgments

Some of the research described in this article was funded through the Pharmacogenetics of Anticancer Agents Research Group by the NIH/NIGMS grant U01GM61393 and NIH/NCI Breast SPORE P50 CA125183. The SCAN database is maintained by the Cox lab supported by the PAAR grant. The authors are grateful to Dr. Marleen Welsh for critically reviewing the manuscript.

ABBREVIATIONS

AA	African American
ABCB1	ATP-binding cassette sub-family B member 1
ADR	Adverse drug reaction
ALL	Acute lymphoblastic leukaemia
Ara-C	Cytarabine arabinoside
ASN	Asian HapMap cell lines including the CHB and JPT samples
ASW	African ancestry in Southwest USA
CEPH	Centre d'Etude du Polymorphisme Humain
CEU	CEPH samples from Utah, USA
CHB	Han Chinese from Beijing, China
CHD	Chinese in Metropolitan Denver, Colorado
CNV	Copy number variant
CYP3A	Cytochrome P450, family 3, subfamily A
CYP3A4	Cytochrome P450, family 3, subfamily A, polypeptide 4
DCK	Deoxycytidine kinase
DGV	Database of Genomic Variants
EBNA2	Epstein-Barr virus nuclear antigen 2
EBV	Epstein-Barr Virus
EGFR	Epidermal growth factor receptor
ENCODE	Encyclopedia of DNA Elements
eQTL	Expression quantitative trait locus
FDA	Food and Drug Administration
GIH	Gujarati Indians in Houston, Texas
GWA	Genome-wide association
IC50	Half maximal inhibitory concentration
JPT	Japanese from Tokyo, Japan
LCL	Lymphoblastoid cell line
LMP1	Latent membrane protein 1
LWK	Luhya in Webuye, Kenya
MAF	Minor allele frequency
MEX	Mexican ancestry in Los Angeles, California

MGMT	<i>O</i> ⁶ -methylguanine DNA methyltransferase
MKK	Maasai in Kinyawa, Kenya
MLL	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
MTXPG	Methotrexate polyglutamates)
NCI	National Cancer Institute
NHGRI	National Human Genome Research Institute
NIEHS	National Institute of Environmental Health Sciences
PGRN	Pharmacogenetics Research Network
QTL	Quantitative trait locus
rMLL	MLL rearrangement
SAHA	Suberoylanilide hydroxamic acid
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1
SNP	Single nucleotide polymorphism
SOLiD	Sequencing by Oligo Ligation and Detection
STRP	Short-tandem-repeat polymorphism
TPMP	Thiopurine S-methyltransferase
TSI	Toscans in Italy
UGT1A	UDP glucuronosyltransferase 1 family, polypeptide A cluster
VIP	Very Important Pharmacogenes
WGTP	Whole Genome TilePath
WRN	Werner Syndrome
YRI	Yoruba people from Ibadan, Nigeria

References

1. Ingelman-Sundberg M. Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. *Naunyn Schmiedebergs Arch Pharmacol* 2004;369:89–104. [PubMed: 14574440]
2. Davies EC, Green CF, Mottram DR, Pirmohamed M. Adverse drug reactions in hospitals: a narrative review. *Curr Drug Saf* 2007;2:79–87. [PubMed: 18690953]
3. The International HapMap Consortium. The International HapMap Project. *Nature* 2003;426:789–96. [PubMed: 14685227]
4. The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–320. [PubMed: 16255080]
5. Biason P, Masier S, Toffoli G. UGT1A1*28 and other UGT1A polymorphisms as determinants of irinotecan toxicity. *J Chemother* 2008;20:158–65. [PubMed: 18467239]
6. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921. [PubMed: 11237011]

7. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304–51. [PubMed: 11181995]
8. Ratain MJ. Personalized medicine: building the GPS to take us there. *Clin Pharmacol Ther* 2007;81:321–2. [PubMed: 17339859]
9. Ratain MJ. Finding the right dose. *Clin Adv Hematol Oncol* 2003;1:517–8. 31. [PubMed: 16258442]
10. Antoniou AC, Spurdle AB, Sinilnikova OM, Healey S, Pooley KA, Schmutzler RK, et al. Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am J Hum Genet* 2008;82:937–48. [PubMed: 18355772]
11. Dunning AM, Healey CS, Baynes C, Scollen S, Vega A, Rodriguez R, et al. Association of ESR1 gene tagging SNPs with breast cancer risk. *Hum Mol Genet*. 2009
12. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadóttir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009;41:18–24. [PubMed: 19079260]
13. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet*. 2008
14. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008;40:768–75. [PubMed: 18454148]
15. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008;40:161–9. [PubMed: 18193043]
16. Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet* 2008;83:623–32. [PubMed: 18976728]
17. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008;40:575–83. [PubMed: 18391952]
18. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* 2007;39:1443–52. [PubMed: 17952075]
19. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001;409:928–33. [PubMed: 11237013]
20. Jorde LB, Wooding SP. Genetic variation, classification and 'race'. *Nat Genet* 2004;36:S28–33. [PubMed: 15508000]
21. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 2000;66:979–88. [PubMed: 10712212]
22. Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. An apportionment of human DNA diversity. *Proc Natl Acad Sci U S A* 1997;94:4516–9. [PubMed: 9114021]
23. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, et al. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* 2007;80:91–104. [PubMed: 17160897]
24. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature* 2006;444:444–54. [PubMed: 17122850]
25. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, et al. Copy number variation: new insights in genome diversity. *Genome Res* 2006;16:949–61. [PubMed: 16809666]
26. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet* 2007;39:S37–42. [PubMed: 17597780]
27. Dolan ME, Newbold KG, Nagasubramanian R, Wu X, Ratain MJ, Cook EH Jr, et al. Heritability and linkage analysis of sensitivity to cisplatin-induced cytotoxicity. *Cancer Res* 2004;64:4353–6. [PubMed: 15205351]
28. Anderson GD. Gender differences in pharmacological response. *Int Rev Neurobiol* 2008;83:1–10. [PubMed: 18929073]

29. Maitland ML, DiRienzo A, Ratain MJ. Interpreting disparate responses to cancer therapy: the role of human population genetics. *J Clin Oncol* 2006;24:2151–7. [PubMed: 16682733]
30. Rodman JH, Relling MV, Stewart CF, Synold TW, McLeod H, Kearns C, et al. Clinical pharmacokinetics and pharmacodynamics of anticancer drugs in children. *Semin Oncol* 1993;20:18–29. [PubMed: 8475406]
31. Olas B, Wachowicz B. Cisplatin-induced changes in biological activity of blood platelets: thiol-related mechanisms. *Anticancer Drugs* 1996;7:476–82. [PubMed: 8826616]
32. Icli F, Karaoguz H, Dincol D, Demirkazik A, Gunel N, Karaoguz R, et al. Severe vascular toxicity associated with cisplatin-based chemotherapy. *Cancer* 1993;72:587–93. [PubMed: 8319192]
33. Verstappen CC, Heimans JJ, Hoekman K, Postma TJ. Neurotoxic complications of chemotherapy in patients with cancer: clinical signs and optimal management. *Drugs* 2003;63:1549–63. [PubMed: 12887262]
34. Daugaard G. Cisplatin nephrotoxicity: experimental and clinical studies. *Dan Med Bull* 1990;37:1–12. [PubMed: 2178884]
35. Chatelut E, Delord JP, Canal P. Toxicity patterns of cytotoxic drugs. *Invest New Drugs* 2003;21:141–8. [PubMed: 12889735]
36. Zhang W, Ratain MJ, Dolan ME. The HapMap resource is providing new insights into ourselves and its application to pharmacogenomics. *Bioinform Biol Insights* 2008;2:15–23. [PubMed: 18392109]
37. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol* 2006;4:e72. [PubMed: 16494531]
38. Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R. Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 1990;6:575–7. [PubMed: 2184120]
39. Thorisson GA, Stein LD. The SNP Consortium website: past, present and future. *Nucleic Acids Res* 2003;31:124–7. [PubMed: 12519964]
40. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, et al. Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004;430:743–7. [PubMed: 15269782]
41. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–61. [PubMed: 17943122]
42. Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. *Genome Res* 2005;15:1592–3. [PubMed: 16251469]
43. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nat Genet* 2004;36:949–51. [PubMed: 15286789]
44. Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet Genome Res* 2006;115:205–14. [PubMed: 17124402]
45. Payseur BA, Place M, Weber JL. Linkage disequilibrium between STRPs and SNPs across the human genome. *Am J Hum Genet* 2008;82:1039–50. [PubMed: 18423524]
46. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 2007;39:226–31. [PubMed: 17206142]
47. Price AL, Patterson N, Hancks DC, Myers S, Reich D, Cheung VG, et al. Effects of cis and trans genetic ancestry on gene expression in African Americans. *PLoS Genet* 2008;4:e1000294. [PubMed: 19057673]
48. Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM. Gene-expression variation within and among human populations. *Am J Hum Genet* 2007;80:502–9. [PubMed: 17273971]
49. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, et al. Population genomics of human gene expression. *Nat Genet* 2007;39:1217–24. [PubMed: 17873874]
50. Zhang W, Duan S, Bleibel WK, Wisel SA, Huang RS, Wu X, et al. Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum Genet* 2008;125:81–93. [PubMed: 19052777]

51. Zhang W, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, et al. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* 2008;82:631–40. [PubMed: 18313023]
52. Kapur K, Xing Y, Ouyang Z, Wong WH. Exon arrays provide accurate assessments of gene expression. *Genome Biol* 2007;8:R82. [PubMed: 17504534]
53. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, Hewett M, et al. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *Pharmacogenomics J* 2001;1:167–70. [PubMed: 11908751]
54. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, et al. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* 2003;33:422–5. [PubMed: 12567189]
55. Cheung VG, Spielman RS. The genetics of variation in gene expression. *Nat Genet* 2002;32 (Suppl): 522–5. [PubMed: 12454648]
56. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet* 2005;1:e78. [PubMed: 16362079]
57. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 2005;437:1365–9. [PubMed: 16251966]
58. Zhang W, Bleibel WK, Roe CA, Cox NJ, Eileen Dolan M. Gender-specific differences in expression in human lymphoblastoid cell lines. *Pharmacogenet Genomics* 2007;17:447–50. [PubMed: 17502836]
59. Zhang W, Huang RS, Duan S, Dolan ME. Gene set enrichment analyses revealed differences in gene expression patterns between males and females. *In Silico Biol* 2009;9:55–63. [PubMed: 19795565]
60. 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* 2008;82:1101–13. [PubMed: 18439551]
61. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet* 2008;24:408–15. [PubMed: 18597885]
62. Kwan T, Benovoy D, Dias C, Gurd S, Serre D, Zuzan H, et al. Heritability of alternative splicing in the human genome. *Genome Res* 2007;17:1210–8. [PubMed: 17671095]
63. Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Beaulieu P, et al. Genome-wide analysis of transcript isoform variation in humans. *Nat Genet* 2008;40:225–31. [PubMed: 18193047]
64. Huang RS, Kistner EO, Bleibel WK, Shukla SJ, Dolan ME. Effect of population and gender on chemotherapeutic agent-induced cytotoxicity. *Mol Cancer Ther* 2007;6:31–6. [PubMed: 17237264]
65. Huang RS, Duan S, Kistner EO, Zhang W, Bleibel WK, Cox NJ, et al. Identification of genetic variants and gene expression relationships associated with pharmacogenes in humans. *Pharmacogenet Genomics* 2008;18:545–9. [PubMed: 18496134]
66. Gipps EM, Kidson C. Cellular radiosensitivity: expression of an MS susceptibility gene? *Neurology* 1984;34:808–11. [PubMed: 6539446]
67. Imray FP, Smith PJ, Relf W, Kidson C. Wilms' tumour: association with cellular sensitivity to mitomycin C in patients and first-degree relatives. *Lancet* 1984;1:1148–51. [PubMed: 6144877]
68. Poot M, Gollahon KA, Rabinovitch PS. Werner syndrome lymphoblastoid cells are sensitive to camptothecin-induced apoptosis in S-phase. *Hum Genet* 1999;104:10–4. [PubMed: 10071186]
69. Cloos J, Nieuwenhuis EJ, Boomsma DI, Kuik DJ, van der Sterre ML, Arwert F, et al. Inherited susceptibility to bleomycin-induced chromatid breaks in cultured peripheral blood lymphocytes. *J Natl Cancer Inst* 1999;91:1125–30. [PubMed: 10393720]
70. Jen KY, Cheung VG. Transcriptional response of lymphoblastoid cells to ionizing radiation. *Genome Res* 2003;13:2092–100. [PubMed: 12915489]
71. Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat Genet* 2001;27:383–91. [PubMed: 11279519]
72. Giacomini KM, Brett CM, Altman RB, Benowitz NL, Dolan ME, Flockhart DA, et al. The pharmacogenetics research network: from SNP discovery to clinical drug response. *Clin Pharmacol Ther* 2007;81:328–45. [PubMed: 17339863]

73. Zhou S. Clinical pharmacogenomics of thiopurine S-methyltransferase. *Curr Clin Pharmacol* 2006;1:119–28. [PubMed: 18666383]
74. Johnatty SE, Beesley J, Paul J, Fereday S, Spurdle AB, Webb PM, et al. ABCB1 (MDR 1) polymorphisms and progression-free survival among women with ovarian cancer following paclitaxel/carboplatin chemotherapy. *Clin Cancer Res* 2008;14:5594–601. [PubMed: 18765553]
75. Jones TS, Yang W, Evans WE, Relling MV. Using HapMap tools in pharmacogenomic discovery: the thiopurine methyltransferase polymorphism. *Clin Pharmacol Ther* 2007;81:729–34. [PubMed: 17329987]
76. Lamba JK, Crews K, Pounds S, Schuetz EG, Gresham J, Gandhi V, et al. Pharmacogenetics of deoxycytidine kinase: identification and characterization of novel genetic variants. *J Pharmacol Exp Ther* 2007;323:935–45. [PubMed: 17855478]
77. Watters JW, Kraja A, Meucci MA, Province MA, McLeod HL. Genome-wide discovery of loci influencing chemotherapy cytotoxicity. *Proc Natl Acad Sci U S A* 2004;101:11809–14. [PubMed: 15282376]
78. Duan S, Bleibel WK, Huang RS, Shukla SJ, Wu X, Badner JA, et al. Mapping genes that contribute to daunorubicin-induced cytotoxicity. *Cancer Res* 2007;67:5425–33. [PubMed: 17545624]
79. Bleibel WK, Duan S, Huang RS, Kistner EO, Shukla SJ, Wu X, et al. Identification of genomic regions contributing to etoposide-induced cytotoxicity. *Hum Genet*. 2008[Epub ahead of print]
80. Huang RS, Duan S, Bleibel WK, Kistner EO, Zhang W, Clark TA, et al. A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc Natl Acad Sci U S A* 2007;104:9758–63. [PubMed: 17537913]
81. Huang RS, Duan S, Kistner EO, Bleibel WK, Delaney SM, Fackenthal DL, et al. Genetic variants contributing to daunorubicin-induced cytotoxicity. *Cancer Res* 2008;68:3161–8. [PubMed: 18451141]
82. Huang RS, Duan S, Kistner EO, Hartford CM, Dolan ME. Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans. *Mol Cancer Ther* 2008;7:3038–46. [PubMed: 18765826]
83. Huang RS, Duan S, Shukla SJ, Kistner EO, Clark TA, Chen TX, et al. Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genomewide approach. *Am J Hum Genet* 2007;81:427–37. [PubMed: 17701890]
84. Hartford CM, Duan S, Delaney SM, Mi S, Kistner EO, Lamba JK, et al. Population specific genetic variants important in susceptibility to cytarabine arabinoside cytotoxicity. *Blood*. 2008[Epub ahead of print]
85. French D, Yang W, Cheng C, Raimondi SC, Mullighan CG, Downing JR, et al. Acquired variation outweighs inherited variation in whole genome analysis of methotrexate polyglutamate accumulation in leukemia. *Blood*. 2008
86. Pottier N, Cheok MH, Yang W, Assem M, Tracey L, Obenauer JC, et al. Expression of SMARCB1 modulates steroid sensitivity in human lymphoblastoid cells: identification of a promoter SNP that alters PARP1 binding and SMARCB1 expression. *Hum Mol Genet* 2007;16:2261–71. [PubMed: 17616514]
87. Hartford C, Yang W, Cheng C, Fan Y, Liu W, Trevino L, et al. Genome scan implicates adhesion biological pathways in secondary leukemia. *Leukemia* 2007;21:2128–36. [PubMed: 17673902]
88. Yang J, Bogni A, Cheng C, Bleibel WK, Cai X, Fan Y, et al. Etoposide sensitivity does not predict MLL rearrangements or risk of therapy-related acute myeloid leukemia. *Clin Pharmacol Ther* 2008;84:691–7. [PubMed: 18509329]
89. French D, Yang W, Hamilton LH, Neale G, Fan Y, Downing JR, et al. Concordant gene expression in leukemia cells and normal leukocytes is associated with germline cis-SNPs. *PLoS ONE* 2008;3:e2144. [PubMed: 18478092]
90. Lamba JK, Chen X, Lan LB, Kim JW, Wei Wang X, Relling MV, et al. Increased CYP3A4 copy number in TONG/HCC cells but not in DNA from other humans. *Pharmacogenet Genomics* 2006;16:415–27. [PubMed: 16708050]
91. Choy E, Yelensky R, Bonakdar S, Plenge RM, Saxena R, De Jager PL, et al. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet* 2008;4:e1000287. [PubMed: 19043577]

92. Dezso Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, Dosymbekov D, et al. A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol* 2008;6:49. [PubMed: 19014478]
93. Crawford DC, Akey DT, Nickerson DA. The patterns of natural variation in human genes. *Annu Rev Genomics Hum Genet* 2005;6:287–312. [PubMed: 16124863]
94. Abdueva D, Wing MR, Schaub B, Triche TJ. Experimental comparison and evaluation of the Affymetrix exon and U133Plus2 GeneChip arrays. *PLoS ONE* 2007;2:e913. [PubMed: 17878948]
95. Robinson MD, Speed TP. A comparison of Affymetrix gene expression arrays. *BMC Bioinformatics* 2007;8:449. [PubMed: 18005448]
96. Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP. Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res* 2005;15:674–80. [PubMed: 15867429]
97. Alberts R, Terpstra P, Li Y, Breitling R, Nap JP, Jansen RC. Sequence polymorphisms cause many false cis eQTLs. *PLoS ONE* 2007;2:e622. [PubMed: 17637838]
98. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11. [PubMed: 11125122]
99. Duan S, Zhang W, Bleibel WK, Cox NJ, Dolan ME. SNPInProbe_1.0: A database for filtering out probes in the Affymetrix GeneChip(R) Human Exon 1.0 ST array potentially affected by SNPs. *Bioinformatics* 2008;24:469–70. [PubMed: 18841244]
100. Benovoy D, Kwan T, Majewski J. Effect of polymorphisms within probe-target sequences on oligonucleotide microarray experiments. *Nucleic Acids Res* 2008;36:4417–23. [PubMed: 18596082]
101. Pradervand S, Paillusson A, Thomas J, Weber J, Wirapati P, Hagenbuchle O, et al. Affymetrix Whole-Transcript Human Gene 1.0 ST array is highly concordant with standard 3' expression arrays. *Biotechniques* 2008;44:759–62. [PubMed: 18476829]
102. Okoniewski MJ, Hey Y, Pepper SD, Miller CJ. High correspondence between Affymetrix exon and standard expression arrays. *Biotechniques* 2007;42:181–5. [PubMed: 17373482]
103. Akey JM, Biswas S, Leek JT, Storey JD. On the design and analysis of gene expression studies in human populations. *Nat Genet* 2007;39:807–8. [PubMed: 17597765]author reply 8–9
104. Speck P, Haan KM, Longnecker R. Epstein-Barr virus entry into cells. *Virology* 2000;277:1–5. [PubMed: 11062029]
105. Yates JL, Warren N, Sugden B. Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. *Nature* 1985;313:812–5. [PubMed: 2983224]
106. Morris MA, Dawson CW, Wei W, O'Neil JD, Stewart SE, Jia J, et al. Epstein-Barr virus-encoded LMP1 induces a hyperproliferative and inflammatory gene expression programme in cultured keratinocytes. *J Gen Virol* 2008;89:2806–20. [PubMed: 18931079]
107. Zimmer-Strobl U, Strobl LJ. EBNA2 and Notch signalling in Epstein-Barr virus mediated immortalization of B lymphocytes. *Semin Cancer Biol* 2001;11:423–34. [PubMed: 11669604]
108. Kaaijk P, Kaspers GJ, Van Wering ER, Broekema GJ, Loonen AH, Hahlen K, et al. Cell proliferation is related to in vitro drug resistance in childhood acute leukaemia. *Br J Cancer* 2003;88:775–81. [PubMed: 12618889]
109. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 2008;9:387–402. [PubMed: 18576944]
110. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet* 2008;24:133–41. [PubMed: 18262675]
111. The ENCODE Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004;306:636–40. [PubMed: 15499007]
112. Livingston RJ, von Niederhausern A, Jegga AG, Crawford DC, Carlson CS, Rieder MJ, et al. Pattern of sequence variation across 213 environmental response genes. *Genome Res* 2004;14:1821–31. [PubMed: 15364900]
113. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res* 2000;28:352–5. [PubMed: 10592272]

114. Kuehn BM. 1000 Genomes Project promises closer look at variation in human genome. *Jama* 2008;300:2715. [PubMed: 19088343]
115. The 1000 Genomes Project. Meeting Report: A workshop to plan a deep catalog of human genetic variation. 2007. [http://www.1000genomes.org]
116. Zhang W, Dolan ME. Beyond the HapMap genotypic data: prospects of deep resequencing projects. *Curr Bioinformatics* 2008;3:178–82.
117. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007;315:848–53. [PubMed: 17289997]
118. Greshock J, Cheng J, Rusnak D, Martin AM, Wooster R, Gilmer T, et al. Genome-wide DNA copy number predictors of lapatinib sensitivity in tumor-derived cell lines. *Mol Cancer Ther* 2008;7:935–43. [PubMed: 18413807]
119. Cappuzzo F, Varella-Garcia M, Shigematsu H, Domenichini I, Bartolini S, Ceresoli GL, et al. Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *J Clin Oncol* 2005;23:5007–18. [PubMed: 16051952]
120. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006;38:1378–85. [PubMed: 17072317]
121. Brown R, Glasspool R. Epigenetic modulation of resistance to chemotherapy? *Ann Oncol* 2007;18:1429–30. [PubMed: 17761701]
122. Stratthdee G. Epigenetic markers and response to chemotherapy in cancer. *Dis Markers* 2007;23:43–9. [PubMed: 17325425]
123. Gius D, Bradbury CM, Sun L, Awwad RT, Huang L, Smart DD, et al. The epigenome as a molecular marker and target. *Cancer* 2005;104:1789–93. [PubMed: 16149092]
124. Kaina B, Fritz G, Ochs K, Haas S, Grombacher T, Dosch J, et al. Transgenic systems in studies on genotoxicity of alkylating agents: critical lesions, thresholds and defense mechanisms. *Mutat Res* 1998;405:179–91. [PubMed: 9748564]
125. Gerson SL. MGMT: its role in cancer aetiology and cancer therapeutics. *Nat Rev Cancer* 2004;4:296–307. [PubMed: 15057289]
126. Rabik CA, Njoku MC, Dolan ME. Inactivation of O6-alkylguanine DNA alkyltransferase as a means to enhance chemotherapy. *Cancer Treat Rev* 2006;32:261–76. [PubMed: 16698182]
127. Esteller M, Gaidano G, Goodman SN, Zagonel V, Capello D, Botto B, et al. Hypermethylation of the DNA repair gene O(6)-methylguanine DNA methyltransferase and survival of patients with diffuse large B-cell lymphoma. *J Natl Cancer Inst* 2002;94:26–32. [PubMed: 11773279]
128. Zuo C, Ai L, Ratliff P, Suen JY, Hanna E, Brent TP, et al. O6-methylguanine-DNA methyltransferase gene: epigenetic silencing and prognostic value in head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 2004;13:967–75. [PubMed: 15184253]
129. Novoyatleva T, Tang Y, Rafalska I, Stamm S. Pre-mRNA missplicing as a cause of human disease. *Prog Mol Subcell Biol* 2006;44:27–46. [PubMed: 17076263]
130. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev* 2003;17:419–37. [PubMed: 12600935]
131. Brinkman BM. Splice variants as cancer biomarkers. *Clin Biochem* 2004;37:584–94. [PubMed: 15234240]
132. Alves R, Vilaprinyo E, Sorribas A. Integrating bioinformatics and computational biology: perspectives and possibilities for In Silico network reconstruction in molecular systems biology. *Curr Bioinformatics* 2008;3:98–129.
133. Zhang W, Huang RS, Dolan ME. Integrating epigenomics into pharmacogenomic studies. *Pharmacogenomics Personalized Med* 2008;1:7–14.
134. Zhang W, Liu W, Innocenti F, Ratain MJ. Searching for tissue-specific expression pattern-linked nucleotides of UGT1A isoforms. *PLoS ONE* 2007;2:e396. [PubMed: 17460763]
135. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008;6:e107. [PubMed: 18462017]

136. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;6:813–23. [PubMed: 16990858]
137. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24:236–44. [PubMed: 10700175]
138. Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24:227–35. [PubMed: 10700174]
139. Blower PE, Verducci JS, Lin S, Zhou J, Chung JH, Dai Z, et al. MicroRNA expression profiles for the NCI-60 cancer cell panel. *Mol Cancer Ther* 2007;6:1483–91. [PubMed: 17483436]
140. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* 2005;436:117–22. [PubMed: 16001072]
141. Zhang W, Liu W, Poradosu E, Ratain M. Genome-wide identification of genetic determinants for the cytotoxicity of perifosine. *Hum Genomics* 2008;3:53–70. [PubMed: 19129090]
142. Liu W, Wu X, Zhang W, Montenegro RC, Fackenthal DL, Spitz JA, et al. Relationship of EGFR mutations, expression, amplification, and polymorphisms to epidermal growth factor receptor inhibitors in the NCI60 cell lines. *Clin Cancer Res* 2007;13:6788–95. [PubMed: 18006781]

Fig. 1. General approaches to pharmacogenetic and pharmacogenomic studies. (A) Candidate gene approach; (B) Whole-genome approach. Though SNPs are shown, other genetic variation data can also be used in these approaches.

Table 1

Some publicly available whole-genome resources of LCLs.

Date Type		Technology Platform	Samples ^a	Website & Reference
Genotype	SNPs	PCR and sequencing by Celera or Motorola	48 CEPH families	http://snpdata.cshl.edu [39]
	SNPs	Affymetrix Human Mapping 100K and 500K Arrays; Perlegen	90 CEU, 90 YRI, 45 CHB and 45 JPT lines	http://www.hapmap.org [41]
	CNVs	Affymetrix Human Mapping 500K EA Array; Whole Genome TilePath Array	90 CEU, 90 YRI, 45 CHB and 45 JPT lines	http://www.sanger.ac.uk/humgen/cnv/ [24,49,117] http://projects.tcag.ca/variation/ [43]
mRNA Expression	Gene-level	Affymetrix Human Focus Array	233 CEPH lines	http://www.ncbi.nlm.nih.gov/geo/ [GSE14851] [40]
	Gene-level	Affymetrix Human Focus Array	60 CEU, 41 CHB and 41 JPT lines	http://www.ncbi.nlm.nih.gov/geo/ [GSE5859] [46]
	Gene-level	Affymetrix Human Focus Array	60 CEU, 60 YRI and 82 AA lines	http://www.ncbi.nlm.nih.gov/geo/ [GSE10824] [47]
	Gene-level	Affymetrix Human Focus Array	8 CEU and 8 YRI lines	http://www.ncbi.nlm.nih.gov/geo/ [GSE7236] [48]
	Gene-level	Affymetrix Human Exon 1.0 ST Array	87 CEU and 89 YRI lines	http://www.ncbi.nlm.nih.gov/geo/ [GSE7851] [51]
	Exon-level	Affymetrix Human Exon 1.0 ST Array	87 CEU and 89 YRI lines	http://www.ncbi.nlm.nih.gov/geo/ [GSE9703] [50]
	Gene-level	Illumina Sentrix Human-6 Expression BeadChip Version 1	90 CEU, 90 YRI, 45 CHB and 45 JPT lines	http://www.ncbi.nlm.nih.gov/geo/ [GSE6536] [49]

^a CEPH, Centre d'Etude du Polymorphisme Humain; CEU, CEPH samples from Utah, USA; YRI, Yoruba people from Ibadan, Nigeria; CHB, Han Chinese from Beijing, China; JPT, Japanese from Tokyo, Japan; AA, Coriell Human Variation Panel (African American).

Table 2

Whole-genome pharmacogenomic studies using the HapMap LCLs.

Drug ^a	Mechanism of Action	Samples ^b	Validation Set	Reference
Etoposide	Topoisomerase II inhibitor	87 CEU and 89 YRI lines		(Huang et al. 2007) [80]
Daunorubicin	Topoisomerase II inhibitor	87 CEU; and 89 YRI lines	50 non-HapMap CEPH	(Huang et al. 2008) [81]
Carboplatin	Alkylating agent	89 YRI lines		(Huang et al. 2008) [82]
Cisplatin	Alkylating agent	87 CEU and 89 YRI lines		(Huang et al. 2007) [83]
Ara-C	RNA/DNA antimetabolite	87 CEU and 89 YRI lines	49 non-HapMap CEPH	(Hartford et al. 2008) [84]
Bortezomib	Proteasome inhibitor	Experiment failed		(Choy et al. 2008) [91]
Rapamycin	mTOR inhibitor	Experiment failed		(Choy et al. 2008) [91]
5-Fluorouracil	RNA/DNA antimetabolite	90 ASN, 84 CEU and 85 YRI lines		(Choy et al. 2008) [91]
Methotrexate	RNA/DNA antimetabolite	90 ASN, 81 CEU and 85 YRI lines		(Choy et al. 2008) [91]
6-Mercaptopurine	DNA antimetabolite	90 ASN, 79 CEU and 85 YRI lines		(Choy et al. 2008) [91]
SAHA	Histone deacetylase inhibitor	90 ASN, 85 CEU and 85 YRI lines		(Choy et al. 2008) [91]
Simvastatin	Lipid-lowering agent	90 ASN, 87 CEU and 85 YRI lines		(Choy et al. 2008) [91]

^a Ara-C, Cytarabine arabinoside; SAHA, Suberoylanilide hydroxamic acid.

^b CEU, CEPH samples from Utah, USA; YRI, Yoruba people from Ibadan, Nigeria; CHB, Han Chinese from Beijing, China; JPT, Japanese from Tokyo, Japan; ASN, Asian HapMap cell lines including the CHB and JPT samples.