Genome **Medicine**

REVIEW

Pharmacogene regulatory elements: from discovery to applications

Robin P Smith[1,2†], Ernest T Lam[2,3†], Svetlana Markova[1], Sook Wah Yee[1]* and Nadav Ahituv[1,2]*

Abstract

Regulatory elements play an important role in the variability of individual responses to drug treatment. This has been established through studies on three classes of elements that regulate RNA and protein abundance: promoters, enhancers and microRNAs. Each of these elements, and genetic variants within them, are being characterized at an exponential pace by next-generation sequencing (NGS) technologies. In this review, we outline examples of how each class of element affects drug response via regulation of drug targets, transporters and enzymes. We also discuss the impact of NGS technologies such as chromatin immunoprecipitation sequencing (ChIP-Seq) and RNA sequencing (RNA-Seq), and the ramifications of new techniques such as high-throughput chromosome capture (Hi-C), chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and massively parallel reporter assays (MPRA). NGS approaches are generating data faster than they can be analyzed, and new methods will be required to prioritize laboratory results before they are ready for the clinic. However, there is no doubt that these approaches will bring about a systems-level understanding of the interplay between genetic variants and drug response. An understanding of the importance of regulatory variants in pharmacogenomics will facilitate the identification of responders versus non-responders, the prevention of adverse effects and the optimization of therapies for individual patients.

Keywords ChIP-Seq, enhancers, miRNA, next-generation sequencing, pharmacogenomics, promoters, RNA-Seq

†These authors contributed equally to this review
*Correspondence: sookwah.yee@ucsf.edu, nadav.ahituv@ucsf.edu
[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA
Full list of author information is available at the end of the article

Pharmacogenomics and gene regulatory elements: an emerging picture

Most pharmacogenomics studies to date have focused on coding variants of pharmacologically important proteins. However, well-supported examples of variants in regulatory elements of genes involved in drug response, such as drug metabolizing enzymes and transporters (see review by Georgitsi *et al.* [1]), show that variants in noncoding regulatory sequences are also likely to be important (Table 1). Three classes of regulatory elements have been studied in this context: promoters, enhancers and microRNAs (miRNAs). Each of these has a direct impact on the abundance of messenger RNA (mRNA) (in the case of promoters and enhancers) and protein (in the case of miRNAs). Genetic variation within each of these elements has been linked to human disease as well as interindividual differences in drug response. For example, a single nucleotide polymorphism (SNP) in the promoter of *VKORC1*, the gene encoding vitamin K epoxide reductase complex subunit 1, radically affects an individual's response to the anticoagulant warfarin [2]. Likewise, a SNP within an enhancer in the vicinity of several solute carrier family (SLC) drug transporters is associated with increased clearance of methotrexate (MTX) [3], and a SNP within a 3'-untranslated region (UTR) miRNA binding site prevents resistance to the chemotherapeutic cisplatin [4].

While a rough estimate of the number of coding genes exists [5], it is unclear how many regulatory elements there are in the genome. The difficulty in defining critical regulatory elements is compounded by the fact that the search space for regulatory elements is vast (98% of our genome is noncoding) and without clear sequence cues such as open reading frames. Next-generation sequencing (NGS) approaches are rapidly changing the status quo, revealing the location and function of regulatory elements on a genomic scale. Robust, high-throughput DNA sequencing platforms emerged as the Human Genome Project came to a close in 2003, as did a desire to establish a reference human epigenome [6]. Key technical advances have brought this goal closer to reality by enabling rapid *de novo* detection of DNA methylation [7,8], enhancers [9,10] and RNA transcripts [11-13] on a genome-wide level (Table 2).

Table 1. Pharmacogene regulatory variants linked to drug response

Gene/miRNA	Drug	Comments	References
Promoters			
MGMT	Temozolomide	Promoter methylation of the gene encoding O-6-methylguanine-DNA methyltransferase (MGMT) is associated with treatment outcome in patients with glioblastoma multiforme and anaplastic astrocytoma treated with temozolomide	[78,79]
UGT1A1	Irinotecan	The UGT1A1*28 polymorphism (insertion of a seventh 'TA' repeat in the UGT1A1 promoter) leads to lower UGT1A1 expression levels, causing lower clearance of the active irinotecan metabolite, SN-38. This results in severe side effects, such as diarrhea and neutropenia	[23,29-31]
VKORC1	Anticoagulant	A common promoter variant in VKORC1, -1639G>A (rs17878363), has significant implications for dosing algorithm of anticoagulant response. Patients with the variant G require higher dose of warfarin and have lower response to acenocoumarol	[22,25-27]
Enhancers			
SLCO1A2	Methotrexate	A screen of ECRs in the vicinity of nine pharmacologically relevant membrane transporters identified several enhancers, notably a region in the first intron of SLCO1A2. Common SNPs within this region were linked to reduced SLCO1A2 mRNA levels (rs4148981C>T) and methotrexate clearance (rs11045981 T>G)	[3]
miRNAs			
miR-24	Methotrexate	A naturally occurring SNP near the miR-24 binding site in the 3' UTR of human DHFR (829C>T) interfered with miR-24 function and resulted in DHFR overexpression and methotrexate resistance	[73]
miR-27a and miR-451	Doxorubicin	P-glycoprotein, encoded by ABCB1, causes resistance in cancer cells to multiple chemotherapeutics, and can be activated by miR-27a and miR-451. Antagonizing these miRNAs results in increased uptake of doxorubicin	[71]
miR-27b	Cyclophosphamide	Overexpression of miR-27b in PANC1cells downregulated CYP3A4 and led to cyclophosphamide resistance	[69,70]
miR-200b/200c/429 family	Cisplatin	A polymorphism (rs1045385) within the 3' UTR of TFAP2 prevents binding of this family of miRNAs, leading to higher levels of its product, transcription factor AP-2α. This variant allows drugs that induce AP-2α, such as cisplatin, to more effectively inhibit cancer cell growth	[4]

ECR, evolutionarily conserved region; miRNA, microRNA; SNP, single nucleotide polymorphism; UTR, untranslated region.

In this review, we discuss the role of each class of regulatory element in drug response variability, and how our understanding of these mechanisms has been impacted by NGS approaches. We also discuss NGS technologies such as deoxyribonuclease I sequencing (DNase-Seq), formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-Seq), high-throughput chromosome capture (Hi-C) and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), which have not yet been applied to pharmacogene regulation but will greatly improve our ability to interpret noncoding genetic variation. Finally, we comment on the need for more efficient functional validation, and discuss other challenges that need to be considered in moving NGS data into the clinic.

Pharmacogene regulatory elements

There are many different classes of gene regulatory elements, including promoters, enhancers, miRNAs, silencers and insulators (Figure 1; for detailed reviews see Maston *et al.* [14] and Noonan *et al.* [15]). In this review, we will focus on the first three classes, each of which has been linked to multiple pharmacogenomic phenotypes

(Table 1). That is not to say that other classes of regulatory elements are not important for pharmacogene regulation; they most likely are, but have not yet been identified. As functional validation assays improve, a more complete picture of regulatory mechanisms will no doubt emerge.

Promoters

Gene promoters are located at the 5' terminus of their target gene, and often have two separate domains, known as the core and proximal promoter regions. The core promoter is where the transcription machinery assembles [16,17] and is usually 35 to 40 base pairs (bp) long. The proximal promoter is located in the immediate vicinity (-250 bp to +250 bp) of the gene's transcription start site (TSS). It contains several transcription factor binding sites (TFBS) and is thought to serve as a tethering element for enhancers, enabling them to interact with the core promoter [18]. Additional elements up to 5 kb upstream of the proximal promoter are often considered to be part of the 'promoter region', and are designated as such in this review. Mammalian promoters often contain CpG islands: sequences at least 200 bp in length with

Table 2. Current next-generation sequencing technologies that are suitable for pharmacogene regulatory element discovery

Technology	Description	References
DNA methylation		
MethylC-Seq	Unmethylated cytosines are converted into uracil by bisulfite treatment, and the converted DNA is sequenced. Methylation is detected with single-base resolution by comparing the sequence of the converted DNA with that of the unconverted DNA, assuming efficient cytosine conversion	[7,8]
RRBS	Restriction digested DNA is size-selected, treated with bisulfite for cytosine conversion, and sequenced. Methylation is detected as described for MethylC-Seq. The reduced representational approach lessens the complexity of analysis and covers a reproducible fraction of the genomic space, but regions with a lack of restriction sites might be missed	[80,81]
MeDIP	Methylated DNA elements are enriched with an antibody that binds to 5-methylcytosine, and then sequenced	[82]
MBD-Seq	Methylated DNA elements are enriched with recombinant methyl CpG binding domain of MBD2, and then sequenced	[83]
CAP-Seq	The chromatography-based method uses the CXXC domain, which has a high affinity for unmethylated CpG sites, to enrich for DNA elements with methylated CpG sites	[84]
MRE-Seq	A methylation-sensitive restriction enzyme is used, leaving unmethylated CpG sites available for sequencing	[85]
Identification of regulatory elements		
ChIP-Seq	DNA is crosslinked and isolated via immunoprecipitation typically using an antibody for a transcription factor, regulatory co-factor, or chromatin mark. Subsequent sequencing allows identification of the binding sites of the DNA-associated protein or histone mark	[9,10,86-90]
DNase-Seq	DNase I hypersensitive sites are known to harbor regulatory elements. In this method, they are identified by sequencing DNase-digested DNA fragments	[121,144, 145]
FAIRE-Seq	Nucleosome-depleted DNA elements associated with regulatory activities are isolated via phenol-chloroform phase separation and sequenced	[122,123]
ChIP-exo	Similar to ChIP-Seq, it is used for identification of DNA elements interacting with a protein of interest. It provides better resolution over ChIP-Seq by using an exonuclease to trim unbound DNA around the transcription factor binding site	[133]
Expression profiling		
RNA-Seq	RNA fragments are converted to cDNA and sequenced for gene expression, isoform, and splicing analysis. Novel transcripts and isoforms can be recovered, but low-expression level transcripts may be missed unless sequencing coverage is sufficient	[11-13, 99-102,109]
Chromosomal interactions		
Hi-C	DNA elements in close spatial proximity are crosslinked, ligated and sequenced. It allows for a genome-wide, high-resolution analysis of interacting DNA elements	[127]
ChIA-PET	DNA elements interacting with a protein of interest are enriched via chromatin immunoprecipitation, crosslinked and ligated so that long-range interactions can be identified by sequencing	[128]

CAP, CXXC affinity purification; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; ChIP, chromatin immunoprecipitation; DNase, deoxyribonuclease I; FAIRE, formaldehyde-assisted isolation of regulatory elements; Hi-C, high-throughput chromosome capture; MBD, methylated DNA binding domain sequencing; MeDIP, methylated DNA immunoprecipitation; MRE, methylation-sensitive restriction enzyme; RRBS, reduced representation bisulfite sequencing.

>50% G/C content [19]. CpG islands are unmethylated in tissues in which their target gene is expressed [20,21], but can be silenced by methylation in disease states.

Numerous studies have shown that genetic variations in promoters have functional effects on drug response. Two well-studied pharmacogene promoter variants are those of genes encoding VKORC1 and UDP glucurono-syltransferase 1 family polypeptide A1 (UGT1A1), which have been linked to the anti-coagulation response of warfarin [22] and diarrhea and neutropenia toxicity caused by irinotecan, respectively [23,24]. *VKORC1* is targeted by warfarin (Figure 2), and a common variant (rs9923231, global minor allele frequency 0.467) in its promoter region (-1639G>A) results in the formation of a novel E-box binding site, leading to lower mRNA expression of *VKORC1* [2], and a lower effective dose of warfarin. This site is thought to recruit transcription factors that suppress gene expression by activating repressive histone modification complexes. This variant can explain much of the variability in average dose requirements among Caucasians, and is incorporated in the warfarin-dosing algorithm to improve warfarin treatment outcome [25-27]. The active form of irinotecan, SN-38, is metabolized through glucuronidation by the UGT1A enzyme [28], which has five to eight copies of 'TA' in its promoter. There is a negative correlation between the number of TA repeats and UGT1A1 expression levels, and the presence of seven repeats
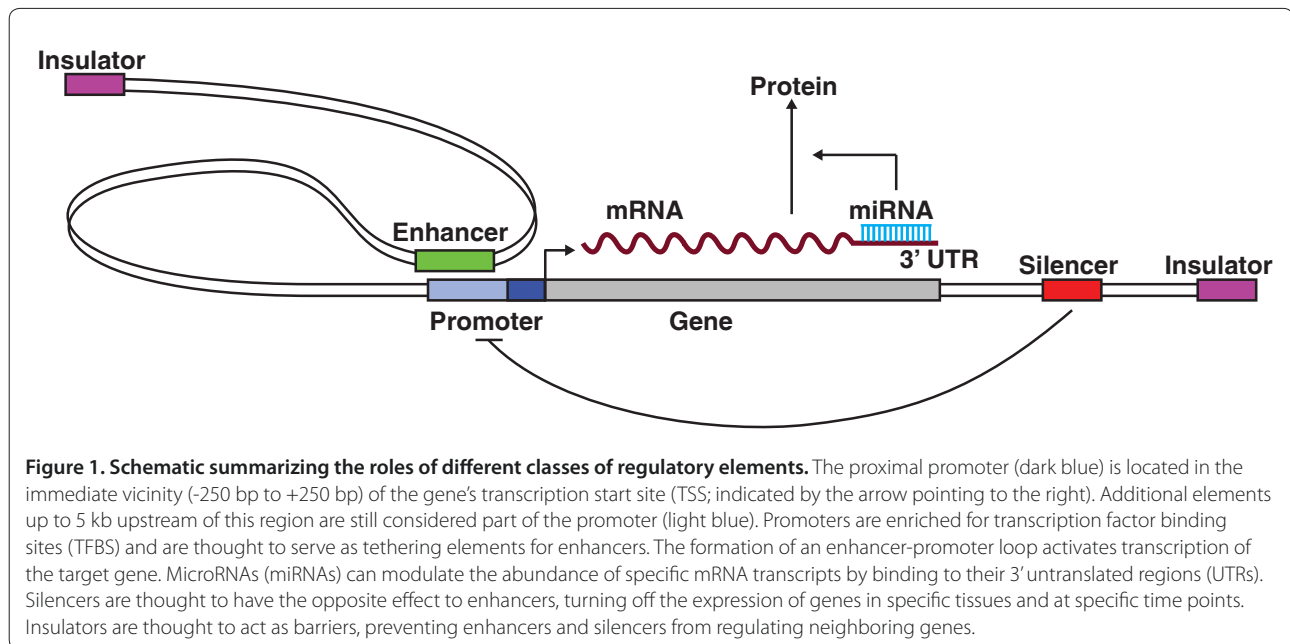
Figure 1. Schematic summarizing the roles of different classes of regulatory elements. The proximal promoter (dark blue) is located in the immediate vicinity (-250 bp to +250 bp) of the gene's transcription start site (TSS; indicated by the arrow pointing to the right). Additional elements up to 5 kb upstream of this region are still considered part of the promoter (light blue). Promoters are enriched for transcription factor binding sites (TFBS) and are thought to serve as tethering elements for enhancers. The formation of an enhancer-promoter loop activates transcription of the target gene. MicroRNAs (miRNAs) can modulate the abundance of specific mRNA transcripts by binding to their 3' untranslated regions (UTRs). Silencers are thought to have the opposite effect to enhancers, turning off the expression of genes in specific tissues and at specific time points. Insulators are thought to act as barriers, preventing enhancers and silencers from regulating neighboring genes.

(denoted as *UGT1A1*28*) was shown to have a significant association with higher-grade neutropenia and diarrhea for patients treated with irinotecan [29-31].

Several large-scale pharmacogene promoter sequencing studies have been conducted, and illustrate the potential advantages of using NGS technologies in the future. In one such example, promoters of 107 different ATP-binding cassette (ABC) transporters and SLC drug-associated transporters were resequenced in an ethnically diverse cohort of 272 individuals, identifying several variants that affect expression levels [32]. Another study systematically identified non-coding expression quantitative trait loci (eQTLs) affecting expression of liver cytochrome P450 superfamily (CYPs) enzymes [33], which play key roles in drug metabolism and toxicity. Studies such as these provide valuable functional annotations that can be mined in future pharmacogenomic association studies and whole-genome datasets.

Enhancers

Enhancers interact with promoters, instructing the promoters when, where and at what level to express their target gene. They can regulate in *cis*, meaning that they regulate a nearby gene on the same chromosome, or in *trans*, regulating genes on a different chromosome [34]. *cis* enhancers can be located 5' or 3' distal to the regulated gene, in introns or even within a coding exon of their target gene [35,36]. The Sonic Hedgehog (*SHH*) limb enhancer is approximately 1,000,000 bp away from its TSS, highlighting the difficulty in linking such elements with a target gene [37]. Enhancers are thought to direct tissue-specific expression in a modular fashion, and

therefore a gene that is active in many tissues is likely to be influenced by multiple enhancers [38,39]. While genetic variation within enhancers can have direct consequences in human disease states [15,40-42], information regarding their role in interindividual drug response is scarce.

A useful example of pharmacogene enhancers from the literature is that of liver CYPs, which metabolize the vast majority of pharmaceutical compounds. Of these, CYP3A4 is the most abundantly expressed in sites of drug disposition in the liver [43] and is also thought to single-handedly catalyze the metabolism of >50% of prescribed pharmaceutical agents. CYP3A4 activity can vary 5- to 20-fold between individuals (depending on the substrate) [44] and its protein expression level can vary up to 40-fold [45]. Of the 28 common SNPs in the CYP3A4 locus, none has been linked to variability in its expression [46], suggesting regulatory variation. Two regions 7.7 kb and 10.5 kb upstream of the gene encoding CYP3A4 were shown to drive its expression in transgenic mouse studies [47,48]. A trinucleotide insertion within this region is present in about 3.1% of the French population, and leads to reduced induction of CYP3A4 expression in cell culture models [48]. Although this insertion is relatively rare in other populations and its effects on adverse drug reactions are unclear, this study provides evidence that enhancer variants can lead to interindividual differences in drug response. Distal enhancers have also been discovered for genes encoding other CYP family members, including *CYP1A1* [49], *CYP1B1* [49], *CYP2E1* [50] and *CYP2B6* [51], as well as genes encoding other liver enzymes, such as alcohol dehydrogenase 4 (ADH4) [52] and UGT1A1 [53]. The phenotypic effects of variants
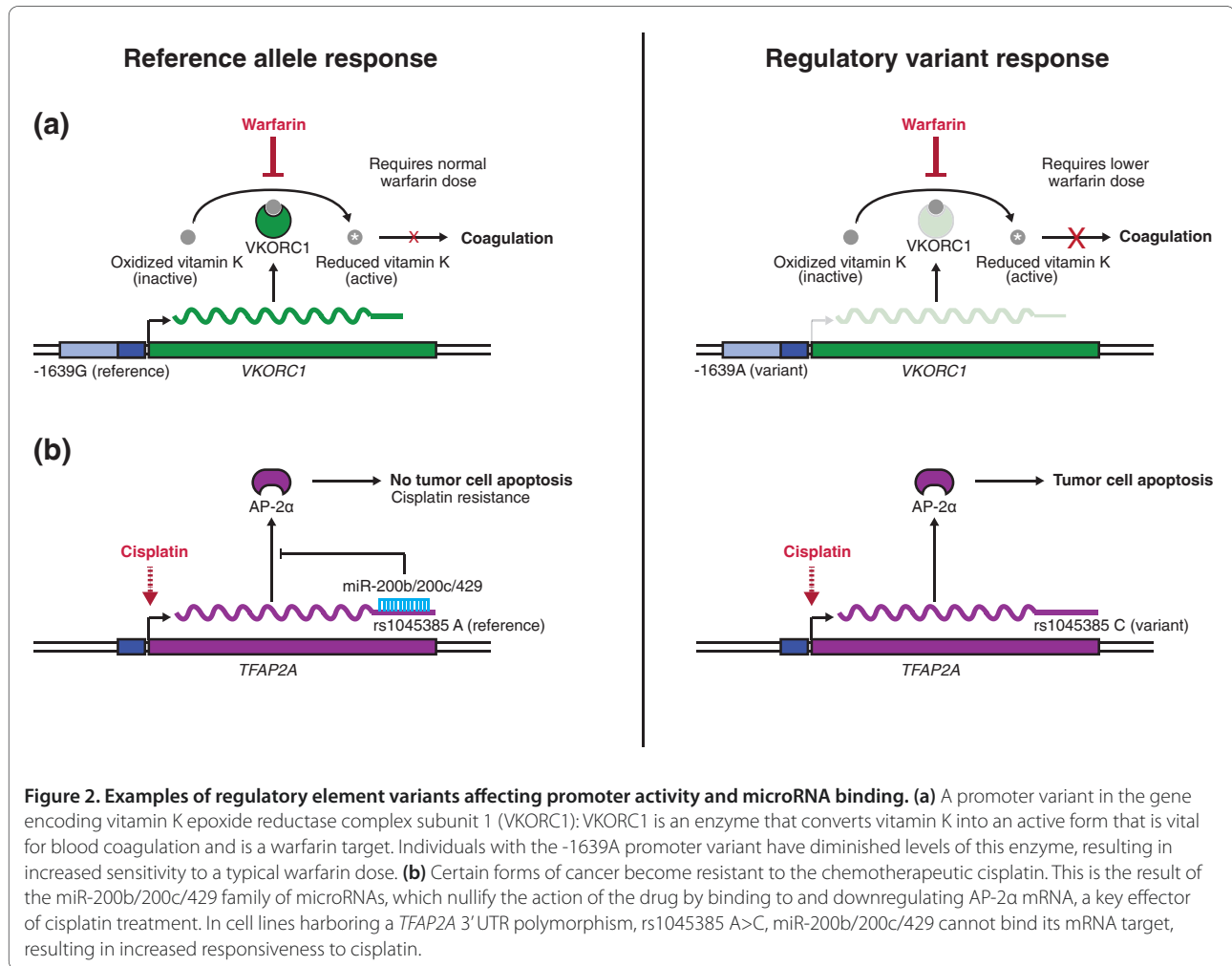
Figure 2. Examples of regulatory element variants affecting promoter activity and microRNA binding. (a) A promoter variant in the gene encoding vitamin K epoxide reductase complex subunit 1 (VKORC1): VKORC1 is an enzyme that converts vitamin K into an active form that is vital for blood coagulation and is a warfarin target. Individuals with the -1639A promoter variant have diminished levels of this enzyme, resulting in increased sensitivity to a typical warfarin dose. **(b)** Certain forms of cancer become resistant to the chemotherapeutic cisplatin. This is the result of the miR-200b/200c/429 family of microRNAs, which nullify the action of the drug by binding to and downregulating AP-2α mRNA, a key effector of cisplatin treatment. In cell lines harboring a *TFAP2A* 3' UTR polymorphism, rs1045385 A>C, miR-200b/200c/429 cannot bind its mRNA target, resulting in increased responsiveness to cisplatin.

within these regions are mostly unknown, mainly due to the difficulty in carrying out physiologically relevant studies.

Our laboratory recently used comparative genomics to identify evolutionarily conserved regions (ECRs) in proximity to nine liver membrane transporters, which were screened *in vivo* using the hydrodynamic tail vein assay [3]. In this technique, a large volume of plasmid-containing delivery solution is rapidly injected into the adult mouse tail vein, causing specific expression of a reporter such as luciferase in the liver. Five ECRs in the vicinity of the genes encoding ABCB11, SLC10A1, SLCO1B1, SLCO1A2 and SLC47A1 were identified as enhancers using this approach. Common human genetic variants within these regions were further functionally characterized, one of which was associated with reduced mRNA expression of *SLCO1A2* in human liver tissue samples. Another variant was associated with increased clearance of MTX, a chemotherapeutic substrate of SLCO1A2 that is used to treat several malignancies, as

well as psoriasis and rheumatoid arthritis. As NGS techniques become more widely adopted, we will be able to rapidly identify distal regulatory elements and key variants within them.

MicroRNAs
miRNAs are small (18 to 25 nucleotide) noncoding RNAs that regulate gene expression by binding to complementary 3' UTRs of target genes. They are endogenously transcribed as precursors and processed [54-56] into mature forms. Mature miRNAs harbor a two to eight nucleotide 'seed' region at the 5' end of the miRNA that is crucial for binding to target mRNA. Upon binding, the miRNA initiates translational repression or cleavage of its target mRNA [57,58]. SNPs within the seed region of the miRNA or the binding site on the target mRNA (miRSNPs) affect targeting of the miRNA and can lead to interindividual expression differences. Rarer variants can occur in genes involved in miRNA biogenesis and maturation, leading to more severe, syndromic phenotypes

[59-62]. Compared with enhancers, miRNAs are relatively easy to identify using computational tools [63,64] and extensive databases of known and predicted miRNAs are publicly available [65].

Despite abundant evidence that miRNAs participate in almost all aspects of cell biology [66-68], there are only a handful of examples of their role in interindividual drug response. Overexpression of miR-27b, which binds the 3' UTRs of *CYP1B1* [69] and *CYP3A4* [70], leads to CYP3A4 downregulation and increased sensitivity to cyclophosphamide [70]. miR-27a and miR-451 activate expression of P-glycoprotein, an *ABCB1* gene product that renders cancer cells resistant to chemotherapeutics. Treatment with miR-27a and miR-451 antagomirs, small synthetic RNAs complementary to their target miRNAs, results in increased accumulation of doxorubicin in drug-resistant cells. [71]. It has been reported recently that miR-200c is downregulated in patients resistant to breast cancer therapy, as well as in human breast cancer cell lines resistant to doxorubicin [72], suggesting additional mechanisms in this pathway.

An example of a mutation affecting pharmacogene miRNA targeting is the C829T variant near the miR-24 binding site in the 3' UTR of the gene encoding human dihydrofolate reductase (*DHFR*). DHFR is a key metabolic enzyme important in DNA synthesis. MTX binds DHFR with high affinity, inhibiting its activity in malignant cells. The C829T variant interferes with miR-24 targeting, resulting in *DHFR* overexpression and MTX resistance [73]. Another example is rs1045385 A>C in the miR-200b/200c/429 binding site in the 3' UTR of the gene encoding transcription factor AP-2α (*TFAP2A*) (Figure 2). AP-2α acts as a tumor suppressor by regulating key genes involved in cell proliferation and apoptosis, and can be induced by the chemotherapeutic agent cisplatin. Cancerous cells from certain endometrial and esophageal tumors can become resistant to cisplatin by upregulating miR-200b, miR-200c and miR-429. These molecules bind to the 3' UTR of AP-2α mRNA, repressing protein translation. The AP-2α rs1045385 C SNP interferes with miR-200b/200c/429 targeting of AP-2α, thus preventing its downregulation and resulting in effective cisplatin treatment [4].

Wang *et al.* [74] recently performed a pair-wise correlation coefficient analysis on expression levels of 366 miRNAs and 14,174 mRNAs in 90 immortalized lymphoblastoid cell lines. They identified 7,207 significantly correlated miRNA-mRNA pairs, with a good representation of metabolic enzymes (for example, CYP family) and drug transporters (for example, ABC and SLC family). Datasets such as these provide excellent troves of candidate regulatory elements for functional validation. The use of NGS methods such as RNA-Seq will greatly aid in the generation of such datasets and will

help shed light on the role of miRNAs in regulating drug responses.

NGS approaches for investigating pharmacogene regulatory elements

Although gene promoters are easily identified by their location, their activity in different cell types and disease states can be altered by a myriad of intrinsic regulatory factors and genetic variants. Epigenetic factors such as DNA methylation can also affect promoter activity, resulting in differential response to drug treatment [75-79]. For example, methylation of the promoter of the gene encoding *O*-6-methylguanine-DNA methyltransferase (MGMT) is a good predictor of the efficacy of temozolomide in the treatment of glioblastoma patients [78,79]. NGS technologies that can analyze DNA methylation on a genomic scale (Table 2), such as MethylC-Seq [7,8], reduced representation bisulfite sequencing (RRBS) [80,81], methylated DNA immunoprecipitation sequencing (MeDIP-Seq) [82], methylated DNA binding domain sequencing (MBD-Seq) [83], CXXC affinity purification plus deep sequencing (CAP-Seq) [84] and methylation-sensitive restriction enzyme sequencing (MRE-Seq) [85] will facilitate future systematic epigenetic studies of pharmacogene regulation.

Regulatory proteins influence gene expression by interacting with specific DNA sequences. Determining which proteins bind to which sites in the genome is the first step to understanding regulatory mechanisms. Chromatin immunoprecipitation (ChIP) approaches have been widely used for this purpose. More recently, coupling ChIP with NGS (ChIP-Seq; Figure 3) [86] has become the *de facto* standard as it provides an unbiased, genome-wide look at enhancer binding with a high signal to noise ratio [87]. In addition to specific regulatory proteins, the availability of specific, high-quality antibodies for histone modification marks has been used to characterize chromatin regulatory states [10,88,89]. The nucleosome core consists of histone proteins, which can be modified post-translationally (for example, by methylation, acetylation, phosphorylation, ubiquitination and sumoylation). These modifications determine the regulatory state of the genomic region they are in (active, silent, and so on) and can be used to detect various gene regulatory elements such as promoters or enhancers. For example, developmentally active enhancers can be identified by the acetylation of the 27th lysine of histone H3 (H3K27ac) [90].

Large-scale, multi-center efforts have mapped the binding of dozens of regulatory proteins in a variety of human cell lines [91]. These include the treatment of cells such as the human hepatocyte cell line HepG2 with factors such as forskolin, insulin and pravastatin. Transcription factors such as signal transducer and activator of transcription 1 (STAT1) or 3 (STAT3) [92,93], co-activators
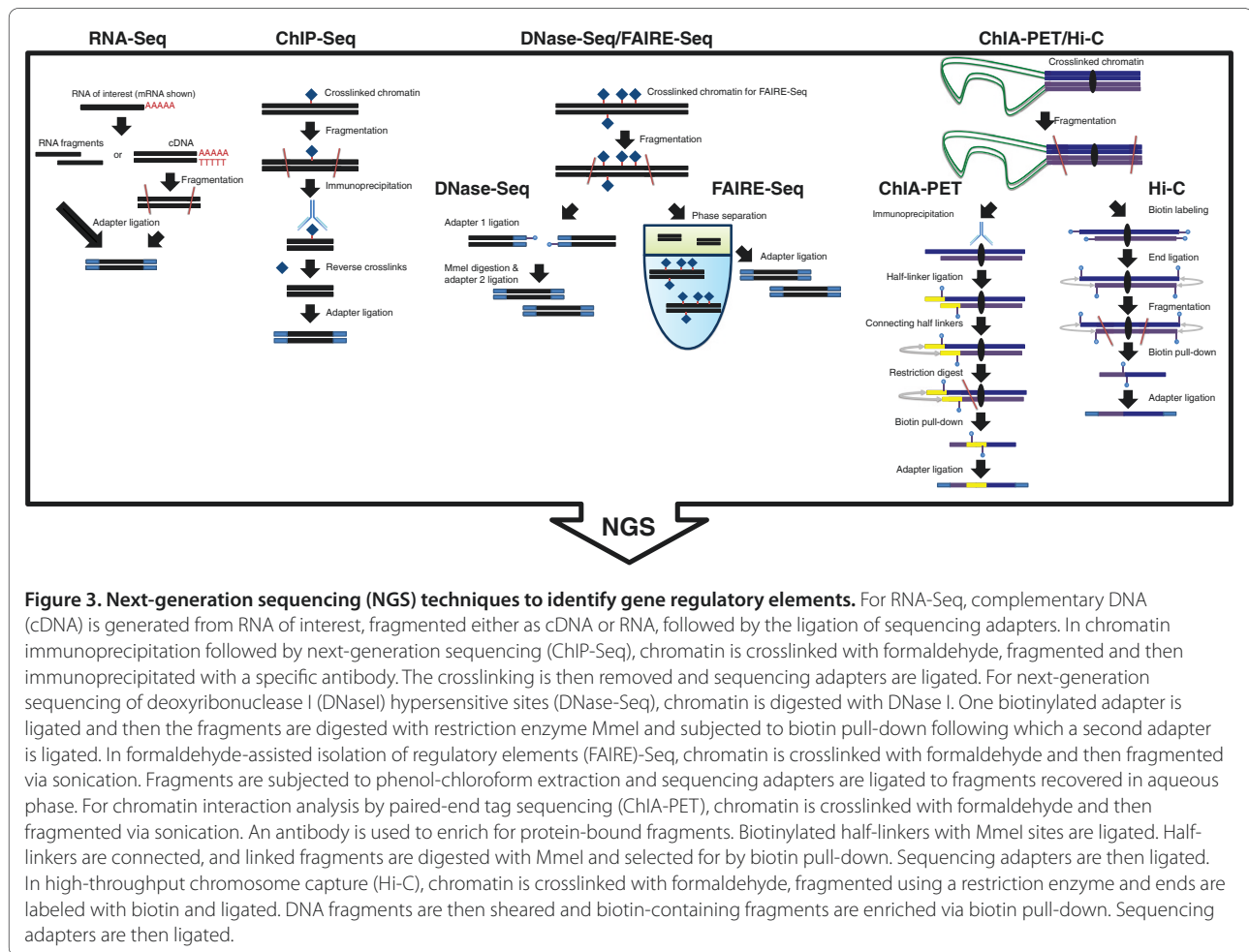
Figure 3. Next-generation sequencing (NGS) techniques to identify gene regulatory elements. For RNA-Seq, complementary DNA (cDNA) is generated from RNA of interest, fragmented either as cDNA or RNA, followed by the ligation of sequencing adapters. In chromatin immunoprecipitation followed by next-generation sequencing (ChIP-Seq), chromatin is crosslinked with formaldehyde, fragmented and then immunoprecipitated with a specific antibody. The crosslinking is then removed and sequencing adapters are ligated. For next-generation sequencing of deoxyribonuclease I (DNaseI) hypersensitive sites (DNase-Seq), chromatin is digested with DNase I. One biotinylated adapter is ligated and then the fragments are digested with restriction enzyme MmeI and subjected to biotin pull-down following which a second adapter is ligated. In formaldehyde-assisted isolation of regulatory elements (FAIRE)-Seq, chromatin is crosslinked with formaldehyde and then fragmented via sonication. Fragments are subjected to phenol-chloroform extraction and sequencing adapters are ligated to fragments recovered in aqueous phase. For chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), chromatin is crosslinked with formaldehyde and then fragmented via sonication. An antibody is used to enrich for protein-bound fragments. Biotinylated half-linkers with MmeI sites are ligated. Half-linkers are connected, and linked fragments are digested with MmeI and selected for by biotin pull-down. Sequencing adapters are then ligated. In high-throughput chromosome capture (Hi-C), chromatin is crosslinked with formaldehyde, fragmented using a restriction enzyme and ends are labeled with biotin and ligated. DNA fragments are then sheared and biotin-containing fragments are enriched via biotin pull-down. Sequencing adapters are then ligated.

such as the CREB binding protein (CREBBP/CBP) and E1A binding protein p300 (EP300/p300) [94,95] that co-localize with enhancers, p160 co-regulators [95] and nuclear receptor proteins such as farnesoid X receptor (FXR/NR1H4) and pregnane X receptor (PXR/NR1I2) [96,97] have been successfully used in ChIP-Seq assays. However, only one study to date has used ChIP-Seq to interrogate a pharmacogenomically relevant drug response *in vivo*. Cui *et al.* [98] mapped PXR binding in mice before and after treatment with pregnenolone-16α-carbonitrile (PCN). PCN is analogous to rifampin, an antibiotic with hepatotoxic side effects, and is thought to activate many of the same targets. In addition to identifying many novel PXR-bound loci, the authors identified a new DNA motif recognized by the factor. Results such as these are invaluable in elucidating complex drug response mechanisms. Furthermore, regulatory regions that are enriched using ChIP-Seq approaches only after the addition of a drug make attractive candidates for interindividual drug response variant discovery.

Over the past two decades, microarray-based methods have substantially improved our ability to quantify gene transcription through gains in throughput. RNA-Seq (Figure 3) has the potential to push the boundaries of our knowledge further by offering an unbiased approach that requires no prior knowledge of transcript variants, and offers single base pair resolution and high dynamic range [11-13]. RNA-Seq is currently the only method that can rapidly detect novel splice isoforms [99-101] and mRNA sequence variants (for example, RNA editing) on a genome-wide scale. Current commercially available RNA-Seq sample preparation kits require as little as 10 pg of total RNA [102], allowing the possibility of strand-specific sequencing of mRNA species from single cells.

A primary use for RNA-Seq in pharmacogenomic studies is the determination of gene expression profiles that can be correlated with drug response phenotypes. Different proteins are responsible for pharmacokinetic interactions with the drug (how the drug enters the cells and reaches its target) and pharmacodynamic interactions (how the drug exerts its cellular effects), and it is

therefore not useful to focus on any one particular gene. For example, breast cancer cell lines demonstrate differential responses to drugs based on their gene expression profiles [103]. In addition, expression levels of several genes were shown to be statistically associated with response to various common chemotherapy agents such as etoposide [104], cisplatin [105] and carboplatin [106]. The Pharmacogenomics Knowledge Base [107,108] project curates pharmacogenomic data from a wide variety of basic and clinical reports, using them to construct drug pathways. The complexity of these pathways, which routinely involve a dozen or more genes and multiple deleterious variants, highlights the need for genome-wide profiling approaches.

For miRNA and miRSNP discovery and profiling, RNA-Seq offers unprecedented scale and depth. For example, Lee *et al.* [109] conducted a comprehensive survey of miRNA sequence variations from human and mouse samples using RNA-Seq. This study demonstrated the complexity of the human miRNA spectrum through deep sequencing of isomiRs (miRNA sequence variants generated from the same precursor by different processing mechanisms). So far, only a few studies have used NGS methodologies to identify miRNAs for diagnostic or prognostic applications. Most of this research was carried out on tumor development and progression [110-114], with sporadic reports of miRNA profiling in non-cancer diseases (for example, endometriosis, pre-eclampsia) [115,116].

The systematic identification and characterization of other classes of regulatory elements will improve our knowledge of how regulatory nucleotide variants affect drug response. Silencers (also termed repressors), which can be thought of as the opposite of enhancers, turn off gene expression at specific time points and locations [14,15,117,118]. Insulators create *cis*-regulatory boundaries that prevent the transcriptional activity of one gene from affecting neighboring genes [14,15,119,120]. Variants in these elements almost certainly influence inter-individual drug responses and remain to be identified.

Future directions for pharmacogene regulatory element discovery

Several emerging NGS techniques have not yet been directly applied to pharmacogenomics, but promise to greatly improve our ability to interpret regulatory variants. Accessible DNA elements residing in active chromatin often harbor regulatory sequences such as promoters, enhancers, silencers and insulators. Deoxyribonuclease I (DNase I) hypersensitive sites (HS) cluster around TSSs, but a significant portion also maps to regions distal to known TSSs [121]. These sites can be mapped genome-wide by DNase-Seq (Figure 3), requiring no prior knowledge of specific transcription factors. A related approach,

known as FAIRE [122], can identify open chromatin by phase separation (Figure 3). There is a high level of correlation between FAIRE and DNase HS sites in general, but unique sites are discovered by each because of slight differences between the techniques [123]. Both DNase-Seq and FAIRE-Seq will be useful in broadly identifying regulatory elements in pharmacologically relevant tissues and will help prioritize SNP discovery efforts.

Enhancers are thought to interact with promoters through chromatin looping (Figure 1) [124]. These looping interactions can be identified through techniques such as chromatin conformation capture (3C) and several of its derivatives (4C [125], 5C [126]). With the advent of NGS, whole-genome adaptations of this technique have been introduced, such as Hi-C [127] and ChIA-PET [128] (Figure 3). A great advantage of these techniques over ChIP-Seq, DNase-Seq and FAIRE is that they can link regulatory elements with their target genes. They could therefore be employed to systematically link variants with individual expression profiles, much like eQTLs but with the power to identify long-distance and *trans* interactions.

NGS technologies are constantly being improved, allowing higher throughput and the ability to ask biological questions on a genomic scale. 'Third-generation', single molecule sequencing platforms are forthcoming and are reviewed in detail elsewhere [129]. Besides higher throughput and longer read lengths, they offer the advantage of eliminating the amplification step, minimizing non-linear biases and thus alleviating some of the informatics and statistical challenges associated with analyses of RNA-Seq and ChIP-Seq data [130-132]. A significant limitation of current ChIP-Seq protocols is the low resolution (about 200 to 300 bp) with which TFBSs within regulatory elements can be identified. ChIP-exo partially eliminates this problem, using lambda exonuclease to facilitate strand-specific 5'-3' degradation, removing DNA not directly involved in the protein-DNA interaction [133]. This modification to the ChIP-Seq protocol significantly increases the signal-to-noise ratio, enabling much more precise peak-calling.

A major obstacle to overcome is our inability to functionally characterize candidate regulatory elements and variants with high throughput. Techniques such as ChIP-Seq, DNase-Seq, FAIRE, Hi-C and ChIA-PET are descriptive in nature. The development of techniques that will allow the functional characterization of thousands of these sequences in a high-throughput manner is critical. One technique that can potentially overcome this hurdle is the use of transcribed barcodes in massively parallel reporter assays, the abundance of which can be measured by RNA-Seq. Using this methodology, thousands of promoter variants were tested in a single experiment [134], and key bases that negatively impacted expression were identified. This methodology has been

recently followed up with enhancers *in vitro* using human cell lines [135] and *in vivo* using the mouse hydrodynamic tail vein assay [136]. Further development of such approaches will permit the high-throughput functional characterization of regulatory elements and nucleotide variants within them.

Translational implications of pharmacogene regulation

As we learn more about how specific variants in regulatory sequences contribute to differential drug responses, it will become more commonplace to personalize drug dosing. Warfarin has become a poster child for pharmacogenomics due to the frequency with which it is prescribed and the importance of genetic testing on proper dosing. Warfarin has a very narrow therapeutic index: there is small difference between clinically beneficial and toxic doses and a large variation in the maintenance dose. Several reports have confirmed VKORC1 as the target of warfarin and CYP2C9 as the principal enzyme responsible for its metabolism [137-139]. Together with non-genetic information, such as age, weight and drug interactions, variants affecting the expression of these genes can explain approximately 50% of the variability of warfarin maintenance dose [26]. A prospective study demonstrated the therapeutic benefits of genotyping known *CYP2C9* and *VKORC1* variant alleles in ensuring an optimal dosage of warfarin [137]. The clinical implications of the *VKORC1* -1639G>A regulatory variant and a coding variant of *CYP2C9* prompted the Food and Drug Administration (FDA) to add this information to warfarin labeling [140].

Another example of a regulatory variant that has been translated to the clinic is the *UGT1A1*28* promoter allele, which alters an individual's response to the anticancer drug irinotecan. The information about the *UGT1A1* variant and summary of the clinically significant adverse reactions, related to severe neutropenia and diarrhea, have been added in the 'Warnings' section of the FDA-approved drug label [140]. If a patient is known to be homozygous for the *UGT1A1*28* allele, clinicians are instructed to reduce the prescribed dose of irinotecan by one level. Patients who are taking irinotecan are often monitored for adverse reactions and to allow early relief of side effects. Genotyping tests for pharmacogene variants are becoming more widely available, along with guidelines to help clinicians with dosing and dosing adjustment [26,141-143].

Despite the fact that we have discovered many functional pharmacogene variants, the uptake of pharmacogenomic testing in the clinic has been slow. There is mounting evidence that pharmacogenomics data can play an important role in identifying responders and non-responders to drugs, avoiding side effects and allowing optimized dosing for patients. However, the link between biologically significant correlations and the therapeutic impact of adopting new clinical practices is unclear. It is vital that we develop a useful framework to sift through and prioritize functional variants for clinical study. At the same time, it will also be necessary to promote training and education among health professionals about the value of pharmacogenomic testing before new policies can be widely adopted.

Concluding remarks

Over the past few years, NGS technologies have greatly accelerated the identification of regulatory elements. However, their use has mainly been limited to genome annotation of physiologically normal cells and tissues. With time, their use in interpreting pharmacogenomic drug-gene interactions will grow rapidly. With each individual genome having millions of nucleotide variants, and reference epigenomic datasets soon to be widely available, there will be a vital need for ways to limit the search space for biologically relevant variants. The use of these technologies in a drug-specific manner, such as the study carried out by Cui *et al.* [98] to map PCN-induced PXR binding sites in mice, will provide the opportunity to highlight drug response-associated regions in these whole-genome datasets.

A major challenge will be to bring these experimental results to a clinical setting. Strong collaboration will be needed between scientists, software engineers, clinicians and pharmacists in order to generate tools to visualize and interpret genomic variations. Several ethical issues will also need to be addressed, such as the privacy and confidentiality of this genomic data, how it will be stored and who can access it, keeping in mind that this information will be extremely important throughout the entire prescription process. In addition, the development of rapid high-throughput assays to functionally characterize variants in pharmacogene-associated regulatory elements is still needed. Techniques that use transcribed barcodes alongside NGS technologies, as mentioned previously [134-136], hold great promise. However, these techniques need to allow the rapid functional assessment of uncharacterized nucleotide variants of each individual. This is necessary to allow the functional consequence of these variants to be analyzed before a drug is prescribed. The existence of such technologies could also be extremely useful in cancer treatment by allowing assessment of how *de novo* mutations alter the efficacy of a drug treatment. The ultimate goal of these studies would be to provide information to a physician or pharmacist regarding an individual's genomic sequence and any drug-associated gene or regulatory element variants so that the most efficient and least harmful drug for each patient can be prescribed.

Abbreviations

ABC, ATP-binding cassette; bp, base pairs; CAP-Seq, CXXC affinity purification plus deep sequencing; cDNA, complementary DNA; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; ChIP-Seq, chromatin immunoprecipitation sequencing; CYP, cytochrome p450; CYP3A4, cytochrome P450 family 3A4; DHFR, dihydrofolate reductase; DNase I, deoxyribonuclease I; DNase-Seq, deoxyribonuclease I sequencing; ECR, evolutionarily conserved region; eQTL, expression quantitative trait loci; FAIRE-seq, formaldehyde-assisted isolation of regulatory elements sequencing; FDA, Food and Drug Administration; Hi-C, high-throughput chromosome capture; HS, hypersensitive sites; kb, kilobase; miRNA, microRNA; miRSNP, single nucleotide polymorphism in the microRNA regulatory pathways; mRNA, messenger RNA; MTX, methotrexate; NGS, next-generation sequencing; PCN, pregnenolone-16α-carbonitrile; PXR, pregnane X receptor; SLC, solute carrier family; SNP, single nucleotide polymorphism; TFBS, transcription factor binding sites; TSS, transcription start site; UGT1A1, UDP glucuronosyltransferase 1 family polypeptide A1; UTR, untranslated region; VKORC1, vitamin K epoxide reductase complex subunit 1.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA. [2]Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA. [3]Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA.

Published: 25 May 2012

References

1. Georgitsi M, Zukic B, Pavlovic S, Patrinos GP: **Transcriptional regulation and pharmacogenomics.** *Pharmacogenomics* 2011, **12**:655-673.
2. Wang D, Chen H, Momary KM, Cavallari LH, Johnson JA, Sadee W: **Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement.** *Blood* 2008, **112**:1013-1021.
3. Kim MJ, Skewes-Cox P, Fukushima H, Hesselson S, Yee SW, Ramsey LB, Nguyen L, Eshragh JL, Castro RA, Wen CC, Stryke D, Johns SJ, Ferrin TE, Kwok PY, Relling MV, Giacomini KM, Kroetz DL, Ahituv N: **Functional characterization of liver enhancers that regulate drug-associated transporters.** *Clin Pharmacol Ther* 2011, **89**:571-578.
4. Wu Y, Xiao Y, Ding X, Zhuo Y, Ren P, Zhou C, Zhou J: **A miR-200b/200c/429-binding site polymorphism in the 3′ untranslated region of the AP-2alpha gene is associated with cisplatin resistance.** *PLoS One* 2011, **6**:e29043.
5. Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, Kellis M, Lindblad-Toh K, Lander ES: **Distinguishing protein-coding and noncoding genes in the human genome.** *Proc Natl Acad Sci U S A* 2007, **104**:19428-19433.
6. Bradbury J: **Human epigenome project - up and running.** *PLoS Biol* 2003, **1**:E82.
7. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR: **Highly integrated single-base resolution maps of the epigenome in Arabidopsis.** *Cell* 2008, **133**:523-536.
8. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462**:315-322.
9. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA: **ChIP-seq accurately predicts tissue-specific activity of enhancers.** *Nature* 2009, **457**:854-858.
10. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473**:43-49.
11. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**:621-628.
12. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM: **Stem cell transcriptome profiling via massive-scale mRNA sequencing.** *Nat Methods* 2008, **5**:613-619.
13. Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, McDonald H, Varhol R, Jones S, Marra M: **Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing.** *BioTechniques* 2008, **45**:81-94.
14. Maston GA, Evans SK, Green MR: **Transcriptional regulatory elements in the human genome.** *Annu Rev Genomics Hum Genet* 2006, **7**:29-59.
15. Noonan JP, McCallion AS: **Genomics of long-range regulatory elements.** *Annu Rev Genomics Hum Genet* 2010, **11**:1-23.
16. Smale ST, Kadonaga JT: **The RNA polymerase II core promoter.** *Annu Rev Biochem* 2003, **72**:449-479.
17. Butler JE, Kadonaga JT: **The RNA polymerase II core promoter: a key component in the regulation of gene expression.** *Genes Dev* 2002, **16**:2583-2592.
18. Calhoun VC, Stathopoulos A, Levine M: **Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex.** *Proc Natl Acad Sci U S A* 2002, **99**:9243-9247.
19. Gardiner-Garden M, Frommer M: **CpG islands in vertebrate genomes.** *J Mol Biol* 1987, **196**:261-282.
20. Deaton AM, Bird A: **CpG islands and the regulation of transcription.** *Genes Dev* 2011, **25**:1010-1022.
21. Illingworth RS, Bird AP: **CpG islands - 'a rough guide'.** *FEBS Lett* 2009, **583**:1713-1720.
22. Rieder MJ, Reiner AP, Gage BF, Nickerson DA, Eby CS, McLeod HL, Blough DK, Thummel KE, Veenstra DL, Rettie AE: **Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose.** *N Engl J Med* 2005, **352**:2285-2293.
23. Ando Y, Saka H, Ando M, Sawa T, Muro K, Ueoka H, Yokoyama A, Saitoh S, Shimokata K, Hasegawa Y: **Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: a pharmacogenetic analysis.** *Cancer Res* 2000, **60**:6921-6926.
24. Iyer L, Das S, Janisch L, Wen M, Ramirez J, Karrison T, Fleming GF, Vokes EE, Schilsky RL, Ratain MJ: **UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity.** *Pharmacogenomics J* 2002, **2**:43-47.
25. Yuan HY, Chen JJ, Lee MT, Wung JC, Chen YF, Charng MJ, Lu MJ, Hung CR, Wei CY, Chen CH, Wu JY, Chen YT: **A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity.** *Hum Mol Genet* 2005, **14**:1745-1751.
26. Johnson JA, Gong L, Whirl-Carrillo M, Gage BF, Scott SA, Stein CM, Anderson JL, Kimmel SE, Lee MT, Pirmohamed M, Wadelius M, Klein TE, Altman RB: **Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing.** *Clin Pharmacol Ther* 2011, **90**:625-629.
27. Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, Lee MT, Limdi NA, Page D, Roden DM, Wagner MJ, Caldwell MD, Johnson JA: **Estimation of the warfarin dose with clinical and pharmacogenetic data.** *New Engl J Med* 2009, **360**:753-764.
28. Iyer L, King CD, Whitington PF, Green MD, Roy SK, Tephly TR, Coffman BL, Ratain MJ: **Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes.** *J Clin Invest* 1998, **101**:847-854.
29. McLeod HL, Sargent DJ, Marsh S, Green EM, King CR, Fuchs CS, Ramanathan RK, Williamson SK, Findlay BP, Thibodeau SN, Grothey A, Morton RF, Goldberg RM: **Pharmacogenetic predictors of adverse events and response to chemotherapy in metastatic colorectal cancer: results from North American Gastrointestinal Intergroup Trial N9741.** *J Clin Oncol* 2010, **28**:3227-3233.

30. Toffoli G, Cecchin E, Gasparini G, D'Andrea M, Azzarello G, Basso U, Mini E, Pessa S, De Mattia E, Lo Re G, Buonadonna A, Nobili S, De Paoli P, Innocenti F: **Genotype-driven phase I study of irinotecan administered in combination with fluorouracil/leucovorin in patients with metastatic colorectal cancer.** *J Clin Oncol* 2009, **28**:866-871.

31. Innocenti F, Undevia SD, Iyer L, Chen PX, Das S, Kocherginsky M, Karrison T, Janisch L, Ramirez J, Rudin CM, Vokes EE, Ratain MJ: **Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan.** *J Clin Oncol* 2004, **22**:1382-1388.

32. Hesselson SE, Matsson P, Shima JE, Fukushima H, Yee SW, Kobayashi Y, Gow JM, Ha C, Ma B, Poon A, Johns SJ, Stryke D, Castro RA, Tahara H, Choi JH, Chen L, Picard N, Sjödin E, Roelofs MJ, Ferrin TE, Myers R, Kroetz DL, Kwok PY, Giacomini KM: **Genetic variation in the proximal promoter of ABC and SLC superfamilies: liver and kidney specific expression and promoter activity predict variation.** *PLoS ONE* 2009, **4**:e6942.

33. Yang X, Zhang B, Molony C, Chudin E, Hao K, Zhu J, Gaedigk A, Suver C, Zhong H, Leeder JS, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich RG, Slatter JG, Schadt EE, Kasarskis A, Lum PY: **Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver.** *Genome Res* 2010, **20**:1020-1036.

34. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R: **Interchromosomal interactions and olfactory receptor choice.** *Cell* 2006, **126**:403-413.

35. Neznanov N, Umezawa A, Oshima RG: **A regulatory element within a coding exon modulates keratin 18 gene expression in transgenic mice.** *J Biol Chem* 1997, **272**:27549-27557.

36. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, Pappalardo Z, Clarke SL, Wenger AM, Nguyen L, Gurrieri F, Everman DB, Schwartz CE, Birk OS, Bejerano G, Lomvardas S, Ahituv N: **Coding exons function as tissue-specific enhancers of nearby genes.** *Genome Res* 2012 [Epub ahead of print].

37. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E: **A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.** *Hum Mol Genet* 2003, **12**:1725-1735.

38. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM: **In vivo enhancer analysis of human conserved non-coding sequences.** *Nature* 2006, **444**:499-502.

39. Visel A, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA: **Functional autonomy of distant-acting human enhancers.** *Genomics* 2009, **93**:509-513.

40. VanderMeer JE, Ahituv N: **cis-regulatory mutations are a genetic cause of human limb malformations.** *Dev Dyn* 2011, **240**:920-930.

41. Kleinjan DA, van Heyningen V: **Long-range control of gene expression: emerging mechanisms and disruption in disease.** *Am J Hum Genet* 2005, **76**:8-32.

42. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Koteliansky V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ: **From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus.** *Nature* 2011, **466**:714-719.

43. Cholerton S, Daly AK, Idle JR: **The role of individual human cytochromes P450 in drug metabolism and clinical response.** *Trends Pharmacol Sci* 1992, **13**:434-439.

44. Flockhart DA, Rae JM: **Cytochrome P450 3A pharmacogenetics: the road that needs traveled.** *Pharmacogenomics J* 2003, **3**:3-5.

45. Dai D, Tang J, Rose R, Hodgson E, Bienstock RJ, Mohrenweiser HW, Goldstein JA: **Identification of variants of CYP3A4 and characterization of their abilities to metabolize testosterone and chlorpyrifos.** *J Pharmacol Exp Ther* 2001, **299**:825-831.

46. Lamba JK, Lin YS, Thummel K, Daly A, Watkins PB, Strom S, Zhang J, Schuetz EG: **Common allelic variants of cytochrome P4503A4 and their prevalence in different populations.** *Pharmacogenetics* 2002, **12**:121-132.

47. Goodwin B, Hodgson E, Liddle C: **The orphan human pregnane X receptor mediates the transcriptional activation of CYP3A4 by rifampicin through a distal enhancer module.** *Mol Pharmacol* 1999, **56**:1329-1339.

48. Matsumura K, Saito T, Takahashi Y, Ozeki T, Kiyotani K, Fujieda M, Yamazaki H, Kunitoh H, Kamataki T: **Identification of a novel polymorphic enhancer of the human CYP3A4 gene.** *Mol Pharmacol* 2004, **65**:326-334.

49. Beedanagari SR, Taylor RT, Bui P, Wang F, Nickerson DW, Hankinson O: **Role of epigenetic mechanisms in differential regulation of the dioxin-inducible human CYP1A1 and CYP1B1 genes.** *Mol Pharmacol* 2010, **78**:608-616.

50. Shadley JD, Divakaran K, Munson K, Hines RN, Douglas K, McCarver DG: **Identification and functional analysis of a novel human CYP2E1 far upstream enhancer.** *Mol Pharmacol* 2007, **71**:1630-1639.

51. Wang H, Faucette S, Sueyoshi T, Moore R, Ferguson S, Negishi M, LeCluyse EL: **A novel distal enhancer module regulated by pregnane X receptor/ constitutive androstane receptor is essential for the maximal induction of CYP2B6 gene expression.** *J Biol Chem* 2003, **278**:14146-14152.

52. Pochareddy S, Edenberg HJ: **Identification of a FOXA-dependent enhancer of human alcohol dehydrogenase 4 (ADH4).** *Gene* 2010, **460**:1-7.

53. Sugatani J, Yamakawa K, Tonda E, Nishitani S, Yoshinari K, Degawa M, Abe I, Noguchi H, Miwa M: **The induction of human UDP-glucuronosyltransferase 1A1 mediated through a distal enhancer module by flavonoids and xenobiotics.** *Biochem Pharmacol* 2004, **67**:989-1000.

54. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN: **The nuclear RNase III Drosha initiates microRNA processing.** *Nature* 2003, **425**:415-419.

55. Yi R, Qin Y, Macara IG, Cullen BR: **Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs.** *Genes Dev* 2003, **17**:3011-3016.

56. Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH: **Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans.** *Genes Dev* 2001, **15**:2654-2659.

57. Diederichs S, Haber DA: **Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression.** *Cell* 2007, **131**:1097-1108.

58. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T: **Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs.** *Mol Cell* 2004, **15**:185-197.

59. Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K: **Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing.** *Genes Dev* 2005, **19**:489-501.

60. Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ: **Dicer is essential for mouse development.** *Nat Genet* 2003, **35**:215-217.

61. Dalzell JJ, Warnock ND, Stevenson MA, Mousley A, Fleming CC, Maule AG: **Short interfering RNA-mediated knockdown of drosha and pasha in undifferentiated Meloidogyne incognita eggs leads to irregular growth and embryonic lethality.** *Int J Parasitol* 2010, **40**:1303-1310.

62. Melo SA, Moutinho C, Ropero S, Calin GA, Rossi S, Spizzo R, Fernandez AF, Davalos V, Villanueva A, Montoya G, Yamamoto H, Schwartz S Jr, Esteller M: **A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells.** *Cancer Cell* 2010, **18**:303-315.

63. Witkos TM, Koscianska E, Krzyzosiak WJ: **Practical aspects of microRNA target prediction.** *Curr Mol Med* 2011, **11**:93-109.

64. Saito T, Saetrom P: **MicroRNAs - targeting and target prediction.** *Nat Biotechnol* 2010, **27**:243-249.

65. miRBase: the microRNA database [http://www.mirbase.org/]

66. *Gambari R, Fabbri E, Borgatti M, Lampronti I, Finotti A, Brognara E, Bianchi N, Manicardi A, Marchelli R, Corradini R:* **Targeting microRNAs involved in human diseases: a novel approach for modification of gene expression and drug development.** *Biochem Pharmacol* 2011, **82**:1416-1429.

67. Janga SC, Vallabhaneni S: **MicroRNAs as post-transcriptional machines and their interplay with cellular networks.** *Adv* 2011, **722**:59-74.

68. Ambros V: **The functions of animal microRNAs.** *Nature* 2004, **431**:350-355.

69. Tsuchiya Y, Nakajima M, Takagi S, Taniya T, Yokoi T: **MicroRNA regulates the expression of human cytochrome P450 1B1.** *Cancer Res* 2006, **66**:9090-9098.

70. Pan YZ, Gao W, Yu AM: **MicroRNAs regulate CYP3A4 expression via direct and indirect targeting.** *Drug Metab Dispos* 2009, **37**:2112-2117.

71. Zhu H, Wu H, Liu X, Evans BR, Medina DJ, Liu CG, Yang JM: **Role of MicroRNA miR-27a and miR-451 in the regulation of MDR1/P-glycoprotein expression in human cancer cells.** *Biochem Pharmacol* 2008, **76**:582-588.

72. Chen J, Tian W, Cai H, He H, Deng Y: **Down-regulation of microRNA-200c is associated with drug resistance in human breast cancer.** *Med Oncol* 2011. doi: 10.1007/s12032-011-0117-4

73. Mishra PJ, Humeniuk R, Longo-Sorbello GS, Banerjee D, Bertino JR: **A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance.** *Proc Natl Acad Sci U S A* 2007, **104**:13513-13518.

74. Wang L, Oberg AL, Asmann YW, Sicotte H, McDonnell SK, Riska SM, Liu W, Steer CJ, Subramanian S, Cunningham JM, Cerhan JR, Thibodeau SN: **Genome-wide transcriptional profiling reveals microRNA-correlated genes and biological processes in human lymphoblastoid cell lines.** *PLoS One* 2009, **4**:e5878.

75. Chiam K, Centenera MM, Butler LM, Tilley WD, Bianco-Miotto T: **GSTP1 DNA methylation and expression status is indicative of 5-aza-2'-deoxycytidine efficacy in human prostate cancer cells.** *PLoS One* 2011, **6**:e25634.

76. Narayan G, Freddy AJ, Xie D, Liyanage H, Clark L, Kisselev S, Un Kang J, Nandula SV, McGuinn C, Subramaniyam S, Alobeid B, Satwani P, Savage D, Bhagat G, Murty VV: **Promoter methylation-mediated inactivation of PCDH10 in acute lymphoblastic leukemia contributes to chemotherapy resistance.** *Genes Chromosomes Cancer* 2011, **50**:1043-1053.

77. Weberpals JI, Koti M, Squire JA: **Targeting genetic and epigenetic alterations in the treatment of serous ovarian cancer.** *Cancer* 2011, **204**:525-535.

78. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO; European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups; National Cancer Institute of Canada Clinical Trials Group: **Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.** *New Engl J Med* 2005, **352**:987-996.

79. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JE, Hau P, Mirimanoff RO, Cairncross JG, Janzer RC, Stupp R: **MGMT gene silencing and benefit from temozolomide in glioblastoma.** *N Engl J Med* 2005, **352**:997-1003.

80. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R: **Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.** *Nucleic Acids Res* 2005, **33**:5868-5877.

81. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES: **Genome-scale DNA methylation maps of pluripotent and differentiated cells.** *Nature* 2008, **454**:766-770.

82. Sorensen AL, Collas P: **Immunoprecipitation of methylated DNA.** *Methods Mol Biol* 2009, **567**:249-262.

83. Serre D, Lee BH, Ting AH: **MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome.** *Nucleic Acids Res* 2010, **38**:391-399.

84. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP: **Orphan CpG islands identify numerous conserved promoters in the mammalian genome.** *PLoS Genet* 2010, **6**:e1001134.

85. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF: **Conserved role of intragenic DNA methylation in regulating alternative promoters.** *Nature* 2010, **466**:253-257.

86. Johnson DS, Mortazavi A, Myers RM, Wold B: **Genome-wide mapping of in vivo protein-DNA interactions.** *Science* 2007, **316**:1497-1502.

87. Ho JW, Bishop E, Karchenko PV, Negre N, White KP, Park PJ: **ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis.** *BMC Genomics* 2011, **12**:134.

88. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B: **Histone modifications at human enhancers reflect global cell-type-specific gene expression.** *Nature* 2009, **459**:108-112.

89. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129**:823-837.

90. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans.** *Nature* 2011, **470**:279-283.

91. Myers RM, Stamatoyannopoulos J, Snyder M, Dunham I, Hardison RC, Bernstein BE, Gingeras TR, Kent WJ, Birney E, Wold B, Crawford GE: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**:e1001046.

92. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab

A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH: **Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.** *Cell* 2008, **133**:1106-1117.

93. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S: **Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing.** *Nat Methods* 2007, **4**:651-657.

94. Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Bristow J, Ren B, Black BL, Rubin EM, Visel A, Pennacchio LA: **ChIP-Seq identification of weakly conserved heart enhancers.** *Nat Genet* 2010, **42**:806-810.

95. Zwart W, Theodorou V, Kok M, Canisius S, Linn S, Carroll JS: **Oestrogen receptor-co-factor-chromatin specificity in the transcriptional regulation of breast cancer.** *EMBO J* 2011, **30**:4764-4776.

96. Chong HK, Infante AM, Seo YK, Jeon TI, Zhang Y, Edwards PA, Xie X, Osborne TF: **Genome-wide interrogation of hepatic FXR reveals an asymmetric IR-1 motif and synergy with LRH-1.** *Nucleic Acids Res* 2010, **38**:6007-6017.

97. Thomas AM, Hart SN, Kong B, Fang J, Zhong XB, Guo GL: **Genome-wide tissue-specific farnesoid X receptor binding in mouse liver and intestine.** *Hepatology* 2010, **51**:1410-1419.

98. Cui JY, Gunewardena SS, Rockwell CE, Klaassen CD: **ChIPing the cistrome of PXR in mouse liver.** *Nucleic Acids Res* 2010, **38**:7943-7963.

99. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.** *Nat Genet* 2008, **40**:1413-1415.

100. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes.** *Nature* 2008, **456**:470-476.

101. Twine NA, Janitz K, Wilkins MR, Janitz M: **Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease.** *PLoS One* 2011, **6**:16266.

102. Gertz J, Varley KE, Davis NS, Baas BJ, Goryshin IY, Vaidyanathan R, Kuersten S, Myers RM: **Transposase mediated construction of RNA-seq libraries.** *Genome Res* 2011, **2011**:29.

103. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, Ng S, Gibb WJ, Wang NJ, Ziyad S, Tong F, Bayani N, Hu Z, Billig JI, Dueregger A, Lewis S, Jakkula L, Korkola JE, Durinck S, Pepin F, Guan Y, Purdom E, Neuvial P, Bengtsson H, Wood KW, Smith PG, Vassilev LT, Hennessy BT, Greshock J, Bachman KE, Hardwicke MA, *et al.*: **Subtype and pathway specific responses to anticancer compounds in breast cancer.** *Proc Natl Acad Sci U S A* 2011, **2011**:14.

104. Huang RS, Duan S, Bleibel WK, Kistner EO, Zhang W, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ, Dolan ME: **A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity.** *Proc Natl Acad Sci U S A* 2007, **104**:9758-9763.

105. Huang RS, Duan S, Shukla SJ, Kistner EO, Clark TA, Chen TX, Schweitzer AC, Blume JE, Dolan ME: **Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genomewide approach.** *Am J Hum Genet* 2007, **81**:427-437.

106. 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**:3038-3046.

107. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, Hewett M, Lin Z, Liu Y, Liu S, Oliver DE, Rubin DL, Shafa F, Stuart JM, Altman RB: **Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base.** *Pharmacogenomics J* 2001, **1**:167-170.

108. PharmGKB: The Pharmacogenomics Knowledgebase [http://www.pharmgkb.org]

109. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, Wang K: **Complexity of the microRNA repertoire revealed by next-generation sequencing.** *RNA* 2010, **16**:2170-2180.

110. Keller A, Backes C, Leidinger P, Kefer N, Boisguerin V, Barbacioru C, Vogel B, Matzas M, Huwer H, Katus HA, Stähler C, Meder B, Meese E: **Next-generation sequencing identifies novel microRNAs in peripheral blood of lung cancer patients.** *Mol Biosyst* 2011, **7**:3187-3199.

111. Beck D, Ayers S, Wen J, Brandl MB, Pham TD, Webb P, Chang CC, Zhou X: **Integrative analysis of next generation sequencing for small non-coding RNAs and transcriptional regulation in myelodysplastic syndromes.** *BMC Med Genomics* 2011, **4**:19.

112. Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, Loman N, Jonsson G, Naya H, Hoglund M, Borg A, Rovira C: **Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene.** *Cancer Res* 2011, **71**:78-86.

113. Xu G, Wu J, Zhou L, Chen B, Sun Z, Zhao F, Tao Z: **Characterization of the small RNA transcriptomes of androgen dependent and independent prostate cancer cell line by deep sequencing.** *PLoS One* 2010, **5**:e15519.

114. Watahiki A, Wang Y, Morris J, Dennis K, O'Dwyer HM, Gleave M, Gout PW: **MicroRNAs associated with metastatic prostate cancer.** *PLoS One* 2011, **6**:e24950.

115. Hawkins SM, Creighton CJ, Han DY, Zariff A, Anderson ML, Gunaratne PH, Matzuk MM: **Functional microRNA involved in endometriosis.** *Mol Endocrinol* 2011, **25**:821-832.

116. Yang Q, Lu J, Wang S, Li H, Ge Q, Lu Z: **Application of next-generation sequencing technology to profile the circulating microRNAs in the serum of preeclampsia versus normal pregnant women.** *Clin Chim Acta* 2011, **412**:2167-2173.

117. Burke LJ, Baniahmad A: **Co-repressors 2000.** *FASEB J* 2000, **14**:1876-1888.

118. Ogbourne S, Antalis TM: **Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes.** *Biochem J* 1998, **331**:1-14.

119. Bushey AM, Dorman ER, Corces VG: **Chromatin insulators: regulatory mechanisms and epigenetic inheritance.** *Mol Cell* 2008, **32**:1-9.

120. Gaszner M, Felsenfeld G: **Insulators: exploiting transcriptional and epigenetic mechanisms.** *Nat Rev Genet* 2006, **7**:703-713.

121. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A, Weaver M, Shafer A, Lee K, Neri F, Humbert R, Singer MA, Richmond TA, Dorschner MO, McArthur M, Hawrylycz M, Green RD, Navas PA, Noble WS, Stamatoyannopoulos JA: **Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays.** *Nat Methods* 2006, **3**:511-518.

122. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD: **FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin.** *Genome Res* 2007, **17**:877-885.

123. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, Sheffield NC, Graf S, Huss M, Keefe D, Liu Z, London D, McDaniell RM, Shibata Y, Showers KA, Simon JM, Vales T, Wang T, Winter D, Zhang Z, Clarke ND, Birney E, Iyer VR, Crawford GE, Lieb JD, Furey TS: **Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity.** *Genome Res* 2011, **21**:1757-1767.

124. Visel A, Rubin EM, Pennacchio LA: **Genomic views of distant-acting enhancers.** *Nature* 2009, **461**:199-205.

125. Vassetzky Y, Gavrilov A, Eivazova E, Priozhkova I, Lipinski M, Razin S: **Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification.** *Methods Mol Biol* 2009, **567**:171-188.

126. van Berkum NL, Dekker J: **Determining spatial chromatin organization of large genomic regions using 5C technology.** *Methods Mol Biol* 2009, **567**:189-213.

127. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science* 2009, **326**:289-293.

128. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li H, Desai KV, Thomsen JS, Lee YK, et al.: **An oestrogen-receptor-alpha-bound human chromatin interactome.** *Nature* 2009, **462**:58-64.

129. Schadt EE, Turner S, Kasarskis A: **A window into third-generation sequencing.** *Hum Mol Genet* 2010, **19**:R227-240.

130. Chen X, Hoffman MM, Bilmes JA, Hesselberth JR, Noble WS: **A dynamic Bayesian network for identifying protein-binding footprints from single molecule-based sequencing data.** *Bioinformatics* 2010, **26**:i334-342.

131. Pepke S, Wold B, Mortazavi A: **Computation for ChIP-seq and RNA-seq studies.** *Nat Methods* 2009, **6**:S22-32.

132. Ghosh D, Qin ZS: **Statistical issues in the analysis of ChIP-Seq and RNA-Seq data.** *Genes* 2010, **1**:317-334.

133. Rhee PJ, Pugh BF: **Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution.** *Cell* 2011, **147**:1408-1419.

134. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J: **High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis.** *Nat Biotechnol* 2009, **27**:1173-1175.

135. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG, Jr, Kinney JB, Kellis M, Lander ES, Mikkelsen TS: **Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay.** *Nat Biotechnol* 2012, **30**:271-277.

136. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, Ahituv N, Pennacchio LA, Shendure J: **Massively parallel functional dissection of mammalian enhancers in vivo.** *Nat Biotechnol* 2012, **30**:265-270.

137. Carlquist JF, Horne BD, Muhlestein JB, Lappe DL, Whiting BM, Kolek MJ, Clarke JL, James BC, Anderson JL: **Genotypes of the cytochrome p450 isoform, CYP2C9, and the vitamin K epoxide reductase complex subunit 1 conjointly determine stable warfarin dose: a prospective study.** *J Thromb Thrombolysis* 2006, **22**:191-197.

138. Cooper GM, Johnson JA, Langaee TY, Feng H, Stanaway IB, Schwarz UI, Ritchie MD, Stein CM, Roden DM, Smith JD, Veenstra DL, Rettie AE, Rieder MJ: **A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose.** *Blood* 2008, **112**:1022-1027.

139. Takeuchi F, McGinnis R, Bourgeois S, Barnes C, Eriksson N, Soranzo N, Whittaker P, Ranganath V, Kumanduri V, McLaren W, Holm L, Lindh J, Rane A, Wadelius M, Deloukas P: **A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose.** *PLoS Genet* 2009, **5**:e1000433.

140. **U.S. Food and Drug Administration** [http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics]

141. Crews KR, Gaedigk A, Dunnenberger HM, Klein TE, Shen DD, Callaghan JT, Kharasch ED, Skaar TC: **Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype.** *Clin Pharmacol Ther* 2012, **91**:321-326.

142. Relling MV, Klein TE: **CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network.** *Clin Pharmacol Ther* 2011, **89**:464-467.

143. Relling MV, Gardner EE, Sandborn WJ, Schmiegelow K, Pui CH, Yee SW, Stein CM, Carrillo M, Evans WE, Klein TE: **Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing.** *Clin Pharmacol Ther* 2011, **89**:387-391.

144. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: **High-resolution mapping and characterization of open chromatin across the genome.** *Cell* 2008, **132**:311-322.

145. Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, Fields S, Stamatoyannopoulos JA: **Global mapping of protein-DNA interactions in vivo by digital genomic footprinting.** *Nat Methods* 2009, **6**:283-289.