Title
Uncovering drug-responsive regulatory elements

Journal
Pharmacogenomics, 16(16)

ISSN
1462-2416

Authors
Luizon, Marcelo R
Ahituv, Nadav

Publication Date
2015-11-01

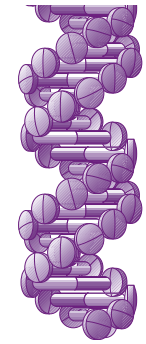
DOI
10.2217/pgs.15.121

Peer reviewed

Review

For reprint orders, please contact: reprints@futuremedicine.com

Pharmacogenomics



Uncovering drug-responsive regulatory elements

Nucleotide changes in gene regulatory elements can have a major effect on interindividual differences in drug response. For example, by reviewing all published pharmacogenomic genome-wide association studies, we show here that 96.4% of the associated single nucleotide polymorphisms reside in noncoding regions. We discuss how sequencing technologies are improving our ability to identify drug response-associated regulatory elements genome-wide and to annotate nucleotide variants within them. We highlight specific examples of how nucleotide changes in these elements can affect drug response and illustrate the techniques used to find them and functionally characterize them. Finally, we also discuss challenges in the field of drug-responsive regulatory elements that need to be considered in order to translate these findings into the clinic.

Keywords: ChIP-seq • enhancers • gene regulatory elements • pharmacogenomics • promoters • RNA-seq • transcriptional regulation

Marcelo R Luizon[1,2] & Nadav Ahituv*[,1,2]

[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158, USA

[2]Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94158, USA

*Author for correspondence: nadav.ahituv@ucsf.edu

Over 98% of the genome do not encode for protein. Within it lie gene regulatory elements that can control gene expression. Genome-wide association studies (GWAS) have led to the identification of over a thousand genomic regions that harbor sequence variants affecting risk for numerous diseases and other phenotypic traits [1,2]. Noteworthy, more than 90% of SNPs in the National Human Genome Research Institute (NHGRI) GWAS catalog [2] are located within noncoding regions [3,4]. This preponderance of noncoding sequence points to a potential role for regulatory variation in the predisposition to many diseases [5,6]. Therefore, the identification and functional characterization of these gene regulatory elements is important for their association with human disease [7].

GWAS has also led to the identification of variants associated with either drug response efficacy or safety [8–12]. However, the typical GWAS sample size for pharmacogenomic assays is in the hundreds, while GWAS of common diseases and other complex traits typically use a few thousand subjects. The allelic odds ratio (OR) values are often well over 3, when drug response is measured as a dichotomous trait, as compared with the effects (OR <1.5) seen in GWAS for common diseases [12]. Moreover, it is not unusual that a single variant accounts for more than 10% of the variation in drug dosing or efficacy, whereas in other complex traits, one variant typically explains no more than a few percent of the variance [12].

The majority of pharmacogenomic-associated SNPs from GWAS are in noncoding regions. Using the NHGRI GWAS catalog [13], we reviewed all available pharmacogenomic GWAS considering both studies for drug response and adverse drug reactions (Table 1 & Supplementary Table 1). We excluded GWAS with no available information regarding the strongest SNP associated with the drug response/trait and the SNP functional class. In total, we reviewed 108 pharmacogenomic GWAS and found 928 associated SNPs, discovering that 96.4% of the associated SNPs are noncoding (Table 1, Supplementary Table 1). These

Future Medicine part of fsg

Table 1. Summary findings of the 108 reviewed pharmacogenomic genome-wide association studies from the NHGRI GWAS catalog.

Location	Context†	928 SNPs	%
Noncoding		895	96.4
	Intergenic	430	46.3
	Intronic	428	46.1
	UTRs, nearGene	33	3.6
	ncRNA	4	0.4
Coding		33	3.6
	Missense	24	2.6
	Synonymous	9	1

See Supplementary Table 1.
†According to SNP functional class (NCBI), and as stated in the 'Context' column of the NHGRI GWAS catalog. The definitions of these terms are stated in Box 1.

results suggest that the majority of common variants leading to interindividual differences in drug response are regulatory. However it is worth noting that most of the SNPs used in the arrays in GWAS are located in noncoding regions, which were selected as informative genetic markers with higher allele frequencies and more likely to have greater power to detect association. Moreover, it is extremely important to note that the associated SNPs identified by GWAS are not necessarily the causal variant, but may be in high linkage disequilibrium (LD) with the causal variant. Finally, only a few of the SNPs found to be associated in pharmacogenomic GWAS have been the focus of follow up studies, and with the techniques and biochemical assays reviewed here these studies can now be performed.

Despite progress through GWAS and other genetic association studies, a large portion of the genetic etiology of pharmacogenomic traits remains unknown [14,15]. Nucleotide changes in gene regulatory elements can play an important role in the variability of individual response to drug treatment [16–18]. The systematic identification of drug-responsive regulatory elements would thus provide a unique resource to discover novel genetic variants that lead to differences in drug response [17]. In this review, we discuss how sequencing technologies have improved our ability to interpret noncoding genetic variation in the human genome, and review techniques used to identify and functionally characterize gene regulatory elements. We outline biochemical assays that take advantage of sequencing technologies such as chromatin immunoprecipitation followed by sequencing (ChIP-seq), chromatin conformation capture (3C) assays and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), which have been applied to the identification of drug-responsive regulatory elements. Finally, we also discuss the challenges of translating the findings of drug-responsive regulatory elements and variants within them to the clinic.

Regulatory elements in the human genome

The most commonly characterized gene regulatory element is the promoter (Figure 1). Promoters are located at the 5' end of the gene, and thus can be readily identified. The core promoter is defined as the minimal stretch of DNA sequence that is sufficient to allow the RNA polymerase II machinery to initiate transcription [19,20]. The proximal promoter is the region that is in the immediate vicinity (-250 to +250 bp) of the transcription start site (TSS) of the gene. It can contain several transcription factor binding sites (TFBS) and is thought to serve as a tethering element for distal regulatory elements, enabling them to interact with the core promoter [21]. Further definitions of these and other terms are shown in Box 1.

Enhancers turn on the promoters at specific locations, times and levels and can be simply defined as the 'promoters of the promoters' [7]. They often have modular expression patterns and a gene that is active in many tissues is likely influenced by multiple enhancers [22,23]. They can regulate in *cis*, meaning that they regulate a gene in a nearby chromosomal region, or in *trans*, regulating a gene that is located on a different chromosome [7]. *Cis* enhancers can be 5′ or 3′ of the regulated gene, in introns or even within the coding exon of the gene they regulate [24–26]. Enhancers can be near the promoter or very far away, and enhancer function is generally considered to be independent of location or orientation relative to the gene they regulate. Enhancers are thought to function through the recruitment of transcription factors (TFs) and subsequent physical interactions with the gene promoter, which are thought to be carried out through DNA looping (Figure 1).

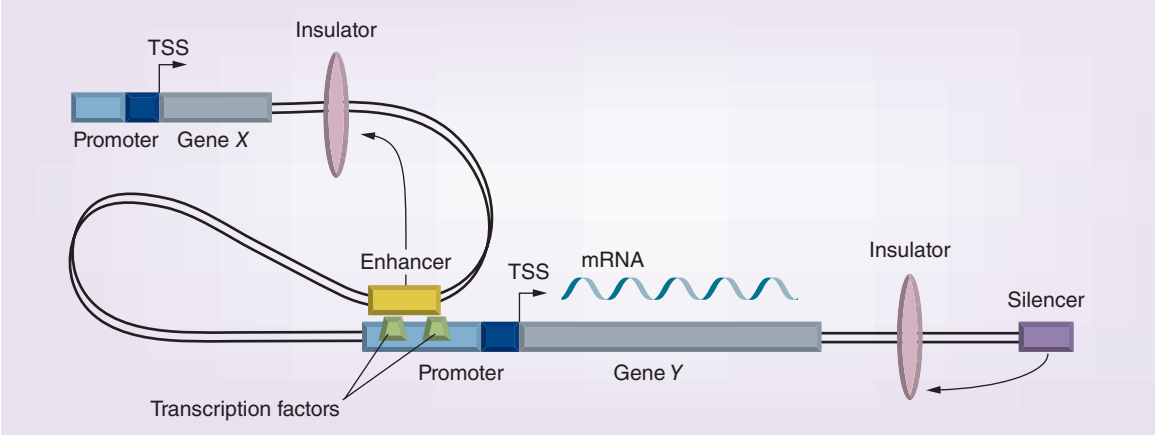


Figure 1. The various types of gene regulatory elements. The proximal promoter (dark blue) is located in the immediate vicinity (-250 to +250 bp) of the TSS of the genes *X* and *Y*. The promoter has additional elements up to 2.5 kb upstream of the proximal promoter (light blue). Promoters are enriched for transcription factor binding sites that are thought to serve as tethering elements for enhancers. The formation of an enhancer–promoter loop activates transcription of the target gene. Silencers are thought to have the opposite effects compared with 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.
TSS: Transcription start site.

Opposite to enhancers, silencers are thought to turn off gene expression at specific time points and locations. Similar to enhancers, silencers can be located almost anywhere with regard to the genes that they regulate and they get activated by the binding of TFs and transcription co-factors. Insulators are DNA sequences that create *cis*-regulatory boundaries that prevent the transcriptional activity of one gene from affecting neighboring genes [7]. Probably, the most widely studied vertebrate insulator-associated protein is the CCCTC-binding factor (CTCF). CTCF is a ubiquitously expressed protein that has 11 zinc fingers and uses different combinations of them to identify and bind different DNA sequences. CTCF-binding sites, which serve as potential insulator regions, have been mapped throughout the human genome in several cell lines using ChIP. Analysis of the location of these sites in various cell lines found that they remain largely unchanged, suggesting that insulator activity stays more or less constant in these different cell lines [27].

Regulatory elements get activated by the binding of TFs. In addition, transcription co-factors bind to the TFs themselves to control transcription. The binding to these specific DNA regulatory sequences changes the nucleosome positioning in that region. The nucleosome consists of 147 bp of DNA wrapped around a histone core. The binding of transcription-related proteins repositions the nucleosome and changes it into a more open state [7]. This change in nucleosome state can be used for the identification of regulatory elements, because the remodeled chromatin state is characterized by markedly heightened accessibility to nucleases [28,29]. Moreover, the nucleosome core consists of histone proteins which can have various post-translational modifications.

The sequencing of DNase I hypersensitive sites (DHSs; Box 2), for example, can take advantage of the changes in nucleosome state and identify all classes of regulatory elements including enhancers, promoters, insulators and silencers [29]. In addition, the post-translational modifications of histones affect the state of the genomic region and can also be used to detect various gene regulatory elements by performing ChIP (Box 2) [7,30]. There are many other techniques that have been developed to identify gene regulatory elements, and since this is not the focus of this review, we only mention the techniques described in this review (Box 2). We also only focus here on the identification of active regulatory elements, mainly promoters and enhancers, and the role of nucleotide variations in these elements on drug response. However, it is important to note that nucleotide variations in silencers and insulators could also have a major role in drug response, and should also be considered.

Promoter variants & drug response

Promoter variants with important clinical pharmacogenomic effect have been associated with drug response. A well-characterized example is the promoter variant of the *UGT1A1* gene and irinotecan (Table 2). Irinotecan is used in the treatment of several solid tumors, including gastrointestinal and lung tumors [39], and it is one of the standard first-line options for patients with metastatic colorectal cancer, in combination with other chemotherapy agents [40]. Irinotecan has a narrow

Box 1. Definitions of terms used in this review

- Allelic expression imbalance (AEI): Analyzes the relative expression between two allelic mRNA transcripts and can be used to detect *cis*-regulatory effects
- Synonymous: A sequence variant within the protein-coding sequence where there is no resulting change to the encoded amino acid (Table 1)
- *Cis*-regulatory element*:* A DNA sequence that regulates the transcription of a nearby gene
- CCTC-binding factor (CTCF): CTCF is a ubiquitously expressed insulator-associated protein that has 11 zinc fingers and uses different combinations of them to identify and bind different DNA sequences
- Co-activator: A protein that assists in transcriptional activation by binding to a transcription factor or factors
- Enhancer: A regulatory element that controls the timing, location and levels of expression of a specific gene via the promoter (Figure 1)
- Expression quantitative trait locus (eQTL): Genomic loci that regulate expression levels of mRNAs
- Histone post-translational modifications: Modifications of histone proteins, which affect the overall chromatin structure
- Intergenic: A DNA sequence that is located in between genes (Table 1)
- Intronic: A DNA sequence located within an intron of a gene (Table 1)
- ncRNA: Noncoding RNA (Table 1)
- Missense: A sequence variant, where at least one base of a codon is changed resulting in a codon that encodes for a different amino acid (Table 1)
- Nucleosome: A segment of DNA wound in an octamer of histone protein cores
- Open chromatin: Regions highly accessible for transcription factors and other proteins
- Promoter: A regulatory region located at the 5′ of the gene at which the transcription machinery binds to initiate transcription. The proximal promoter serves as a tethering element for distal regulatory elements (Figure 1)
- Silencer: A regulatory element that turns off gene expression at specific time points and locations, and can be located almost anywhere with regard to the genes that it regulates (Figure 1)
- SNP: Single nucleotide polymorphism is a DNA sequence variation that occurs in more than 1% of the general population
- Transcription factors (TFs): Proteins that bind to specific DNA sequences and control mRNA transcription (Figure 1)
- *Trans*-regulatory elements: A gene or regulatory element which may regulate the expression of distant genes
- Transcription factor binding site (TFBS): A nucleotide sequence that is recognized and gets bound by a certain TF
- Transcription start site (TSS): The location where RNA polymerase initiates the transcription from the DNA (Figure 1)
- Untranslated regions (UTRs): The sequence at the 3′ and 5′ ends of the mRNA that does not code for protein (Table 1)

therapeutic index and exerts its cytotoxicity by inhibiting topoisomerase I during DNA replication through its active metabolite SN–38, which is metabolized through glucuronidation by the UGT1A1 enzyme. A reduction of UGT1A1 activity causes unconjugated hyperbilirubinemia and Gilbert's syndrome, which is characterized by mild, chronic unconjugated hyperbilirubinemia in the absence of liver disease. The promoter region of the *UGT1A1* gene was first sequenced in patients with Gilbert's syndrome who were found to be homozygous for the *UGT1A1*28* variant [41]. This variant is an elongated TA repeat A[TA]$_7$TAA instead of the more usual A[TA]$_6$TAA (*UGT1A1*1*) located 39 nucleotides upstream of the TSS of the *UGT1A1* gene, which resulted in decreased bilirubin-glucuronidating activity and leads to mildly elevated serum bilirubin levels. The *UGT1A1*28* variant was associated with significantly decreased *UGT1A1* expression levels. This is thought to lead to lower glucuronidation activity, which results in reduced SN–38 clearance and side effects that include diarrhea and neutropenia [41,42].

Several additional promoter variants have been implicated in drug response. For example, promoters variants of well-known genes coding for drug-metabolizing enzymes, such as the *TPMT* and the CYP family of cytochrome P450 oxidases (*CYP1A2, CYP2C9, CYP2C19, CYP2D6* and *CYP3A4*), as well as promoter variants in genes coding for the drug transporters *ABCC2* and *SLCO1B1* are reviewed in detail by Georgitsi *et al.* [49] and have been shown to be associated with drug response.

ChIP (Box 2) is one of the techniques that can be used to uncover and characterize drug response-associated promoters. For example, the human SULT2A1 is a Phase II drug-metabolizing enzyme that catalyzes the sulfonation of endogenous and exogenous

hydroxyl-containing compounds. The expression of *SULT2A1* is transcriptionally regulated by several nuclear receptors, including the liver X receptor (LXR) α (also called NR1H3). LXRα was shown to bind to the *SULT2A1* promoter region [43]. The activation of LXRα by GW3965, a liver X receptor agonist (activator) of human LXRα, induced the expression of *SULT2A1* at mRNA, protein and enzymatic levels. In addition, LXRα transactivated the *SULT2A1* promoter through its specific binding to the -500 to -258 bp region as shown by reporter gene assays and ChIP. The LXRα-responsive reporter activity was abolished when the -500 to -258 bp region was deleted, suggesting that this region was responsible for the LXRα transactivation. Inspection of this promoter region predicted a putative direct repeat spaced by four nucleotides (DR4)-type LXR response element. Mutation of this putative DR4 in the context of the 500 bp promoter abolished the transactivation by LXRα. A positive correlation between the expression of *SULT2A1* and LXRα was shown in primary human hepatocytes, which further supported the regulation of *SULT2A1* by LXRα and establishes human *SULT2A1* as a novel LXRα target gene [43].

Sequencing technologies can improve our ability to uncover promoters associated with drug response. For example, SLFN11 was identified as a critical determinant of response to DNA-damaging agents, such as topoisomerase I inhibitors. Ewing's sarcoma is characterized by the chimeric transcription factor EWS-FLI1, and has notably high *SLFN11* expression. EWS-FLI1-mediated *SLFN11* expression is responsible for high sensitivity of Ewing's sarcoma to camptothecin and combinations of inhibitors of poly(ADP)-ribose polymerase (PARP) with temozolomide. Ewing's sarcoma patients with higher *SLFN11* expression showed better tumor-free survival rate. However, how *SLFN11* is regulated in cancer cells remained largely unknown. To test whether EWS-FLI1 drives *SLFN11* expression and the role of SLFN11 in drug response, A673 Ewing's sarcoma cells were analyzed by ChIP-seq using an FLI1 antibody [50]. EWS-FLI1 was found to bind near the *SLFN11* promoter and act as a positive regulator of *SLFN11* expression in Ewing's sarcoma cells. To determine the relative contribution of the putative FLI1-binding site(s) to the EWS-FLI1-induced *SLFN11* promoter activity, three *SLFN11* promoter mutations were tested by site-directed mutagenesis (mt+91, mt+181, mt+201). Luciferase reporter assays showed that individual mutations at positions +91 and +201 reduced *SLFN11* promoter activity by more than 80%, whereas the +181

Box 2. Techniques used to identify gene regulatory elements discussed in this review

DNase I hypersensitive sites

- Open chromatin regions are associated with nucleosome-free regions and hence can be more attainable to DNase I, an endonuclease that cleaves both single- and double-stranded DNA. These regions are thus termed DNase I hypersensitive sites (DHSs) [7]. DHSs can be discovered on a genomic scale using massively parallel sequencing technologies such as DNase-seq [31]. DHSs do not reveal the identity of the regulatory element, but they can show whether a certain region in a specific cell type or tissue has potential regulatory function [7]

Chromatin immunoprecipitation

- Chromatin immunoprecipitation (ChIP) uses antibodies against DNA-binding proteins to pull down specific DNA sequences to which they bind. DNA-binding proteins are cross-linked to the DNA and an antibody that is specific to a protein of interest is used to pull down the protein along with the bound DNA sequences. Following reversal of the cross links, these DNA sequences can be identified either through quantitative PCR (ChIP-qPCR), by binding to a DNA microarray (ChIP-chip) or using massively parallel sequencing technologies (ChIP-seq) [32–34]. ChIP-seq is becoming the most commonly used tool to map putative regulatory sequences on a genomic scale, and the gold standard in the identification of potential gene regulatory elements [7]. For a more detailed description of the ChIP protocol, see [35,36]

Chromatin conformation capture

- Chromatin loops (Figure 1) have been shown to be one of the major mechanisms by which the various regulatory elements regulate transcription. Chromatin conformation capture (3C) and several derivatives of this technique have been developed to unravel the physical interactions between regulatory elements. They are primarily based on cross-linking DNA-binding proteins with DNA (similar to ChIP), so that both the regulatory elements are bound together (e.g., an enhancer and its target promoter), bridged by the proteins that facilitate this interaction. The DNA is then cut randomly with restriction enzymes or other methods and ligated in dilute conditions where the segments of DNA bridged by the protein cross-linked bundle will preferentially ligate to one another rather than to random-free DNA. These newly ligated DNA segments are then analyzed to identify which regions of DNA have been joined, implying that they physically interact. The specific analysis of these sequences is what determines whether this technique is known as 3C, 4C or 5C [7]. With the advent of massively parallel sequencing technologies, whole-genome adaptations of this technique have been introduced such as high-throughput chromosome capture (Hi-C) [37] and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [38]

Table 2. Examples of drug response-associated regulatory elements.

Gene	Drug	Comments	Ref.
Promoters			
UGT1A1	Irinotecan	The *UGT1A1*28* variant (seven instead of six TA repeats in the *UGT1A1* promoter) is associated with significantly decreased *UGT1A1* expression levels and glucuronidation activity, which results in reduced clearance of the active irinotecan metabolite (SN–38) and, consequently, in severe diarrhea and neutropenia	[41,42]
SULT2A1	GW3965 (agonist)	*SULT2A1* was uncovered as a novel liver X receptor (LXR)α target. The activation of LXRα induced the expression of *SULT2A1*. LXRα transactivated the *SULT2A1* promoter through its specific binding to the -500 to -258 bp region, as shown by promoter reporter and ChIP assays	[43]
SLFN11	Chemotherapy	EWS-FLI1 binds near the TSS of *SLFN11* promoter and acts as a positive regulator of *SLFN11* expression in Ewing's sarcoma (ES) cells, as shown by promoter reporter assays and ChIP-seq. EWS-FLI1–mediated *SLFN11* expression is responsible for high sensitivity of ES to camptothecin and combinations of PARP inhibitors with temozolomide	[44]
Enhancers			
KLF4	Pitavastatin	A novel functional MEF2C-binding site 148 kb upstream of the *KLF4* gene was identified by ChIP-seq and luciferase assays. MEF2C-bound enhancer and TSS of *KLF4* were observed to come into closer spatial proximity after pitavastatin treatment by ChIA-PET and 3C assays	[45]
SLC13A5	Rifampicin	Two enhancer modules located upstream of the TSS of *SLC13A5* gene associated with regulation of PXR-mediated *SLC13A5* induction were identified and functionally characterized by luciferase reporter assays, EMSA and ChIP assays	[46]
CYP2D6	N/A[†]	An enhancer region located ~115 kb downstream of *CYP2D6* was identified. rs5758550 was identified as the regulatory SNP by 4C, reporter gene assays, ChIP assays and CRISPR-mediated enhancer deletion. The minor allele rs5758550 (G) was shown to increase enhancer activity	[47,48]

[†]CYP2D6 metabolizes nearly 25% of clinically used drugs.
ChIP: Chromatin immunoprecipitation; CRISPR: Clustered regularly interspaced short palindromic repeats; EMSA: Electrophoretic mobility shift assay; LXRα: Liver X receptor α; MEF2C: Myocyte enhancer factor-2; PARP: Poly(ADP)-ribose polymerase; TSS: Transcription start site.

mutation had no significant effect. *SLFN11* promoter activity in 293T/EWS-FLI1 cells was also suppressed by approximately 90% for mt+91 and 50% for mt+201, whereas mt+181 did not affect *SLFN11* promoter activity. These results indicate that consensus sequences +91 and +201 are critical for activation of the *SLFN11* promoter by EWS-FLI1. The relationship between *SLFN11* and EWS-FLI1 were further examined in EWS-FLI1-knockdown or -overexpressing cells and in clinical tumor samples. These results suggest that *SLFN11* expression is transcriptionally activated by the ETS transcription factors EWS-FLI1 and ETS1 [44].

The study above highlights how ChIP-seq can be used to identify active promoters. Antibodies against trimethylated histone H3 lysine 4 (H3K4me3) are widely used in ChIP-seq experiments to identify active promoters [51]. RNA polymerase II (PolII) occupancy is also commonly used to identify active promoters. For example, by characterizing PolII-binding profiles using ChIP-seq in mice and humans it was observed that the enrichment of PolII near TSS exhibits a stereotypical bimodal structure, with one peak near active TSSs and a second peak 110 bp downstream from the first peak [52]. Antibodies for specific TFs that are active in the cell line or tissue of interest are also commonly used to identify active promoters.

Enhancers & drug response

Enhancers can also be associated with drug response (Table 2). For example, statins are widely used cholesterol-lowering drugs that exert atheroprotective effects through the induction of specific TFs. In endothelial cells, statin-dependent atheroprotective gene upregulation is mediated by the Kruppel-like factor (*KLF*) family of TFs [45]. The contribution of KLFs in human umbilical vein endothelial cells (HUVECs)

treated with pitavastatin was validated by microarray analyses and *KLF4* was determined to be the most highly induced gene [45]. MEF2 activation is reported to be involved in pitavastatin-dependent *KLF4* induction. A novel functional MEF2C-binding site 148 kb upstream of the *KLF4* gene was identified by ChIP-seq using an MEF2C antibody followed by luciferase assays. To determine which MEF2C-binding sites are functionally active, a series of luciferase reporter assays in HUVECs were carried out with constructs containing the human *KLF4* promoter, and a 3.2-fold increase in reporter gene expression was observed using the construct containing the -148 kb MEF2C-binding region. Moreover, the MEF2C-bound enhancer and TSS of the *KLF4* gene were shown to physically interact following pitavastatin treatment using both ChIA-PET and 3C assays. The conformational change in individual cells was supported by 3D-fluorescence *in situ* hybridization imaging. Thus, these studies showed that dynamic chromatin conformation change mediates pitavastatin-responsive gene induction in endothelial cells [45].

Enhancers can also have an important regulatory role in the expression of drug-associated transporters. For example, the *SLC13A5* is a highly inducible gene associated with rifampicin-stimulated activation of pregnane X receptor (PXR) in human liver [46]. SLC13A5 is a sodium-coupled transporter that mediates cellular uptake of citrate, which plays important roles in the synthesis of fatty acids and cholesterol, and has potential importance in energy homeostasis. However, the transcriptional regulation of the *SLC13A5* gene and whether clinically prescribed drugs could be used to disrupt the expression of this transporter were largely unknown [46]. The selective PXR activator rifampicin markedly induced the mRNA and protein expression of SLC13A5 in human primary hepatocytes. Two enhancers located upstream of *SLC13A5* that are associated with the regulation of PXR-mediated *SLC13A5* induction were identified, and then functionally characterized by using luciferase reporter assays, electrophoretic mobility shift assays (EMSAs) and ChIP assays. Two clusters of potential PXR-binding sites were identified at approximately -1.7 and -22 kb from the TSS with response motifs exhibiting high-sequence homology to PXR response elements (AGGTCA) spaced by 4 nucleotides (DR4). Insertion of a 247 and 239 bp fragment containing the DR4-1 and DR4-2 response motifs within those distal (-22 kb) and proximal (-1.7 kb) regions, respectively, in the *SLC13A5*–1kb promoter resulted in PXR-dependent activation in HepG2 cells. Notably, mutation of DR4-1 completely abolished its response to PXR, whereas the DR4-2 mutant only exhibited a moderate decrease in PXR-based activation. Functional analysis further revealed that *SLC13A5* induction was positively correlated with rifampin-mediated fat accumulation in human primary hepatocytes, whereas knockdown of *SLC13A5* significantly decreased lipid content in HepG2 cells. Combined, these assays uncovered *SLC13A5* as a novel transcriptional target gene of PXR, which could contribute to drug-induced steatosis and metabolic disorders in humans [46].

Enhancers can also control the expression levels of drug-metabolizing enzymes. For example, an enhancer of *CYP2D6* was recently analyzed by using gene reporter assays, chromatin accessibility, ChIP and clustered regularly interspaced short palindromic repeats (CRISPR)-mediated deletion [47]. CYP2D6 metabolizes nearly 25% of clinically prescribed drugs and exhibits great interindividual variability, influencing drug dosing, efficacy and toxicity [47,53]. Two linked SNPs (rs5758550 and rs133333) were identified in an enhancer region located approximately 115 kb downstream of the *CYP2D6* gene that are associated with greater than two-fold increased *CYP2D6* transcription [48]. However, it was not clear which of the two enhancer SNPs are causative, or whether there are additional regulatory regions that could be associated with *CYP2D6* expression levels. Thus, a 4C assay was carried out to identify additional genomic regions that might interact with the *CYP2D6* promoter [47]. The previously identified enhancer region was confirmed as having robust effects on *CYP2D6* expression, and reporter gene assays identified rs5758550 as the regulatory SNP involved in increasing CYP2D6 transcription [47]. CRISPR-mediated deletion in HepG2 cells of the enhancer region surrounding rs5758550 resulted in 70% decreased *CYP2D6* expression. The minor allele of the SNP rs5758550 (G) was shown to increase enhancer activity in HepG2 cells [47]. Combined, these studies show robust effects of both the enhancer element and the SNP rs5758550 on *CYP2D6* expression [47,48].

Identification of drug-responsive regulatory elements

Allelic RNA expression imbalance (AEI) analyzes the relative expression between two allelic mRNA transcripts, and can be used to detect *cis*-regulatory effects [54]. One such example of the use of AEI for pharmacogenomic studies is the *VKORC1* gene and warfarin. A haplotype of *VKORC1* gene carrying the minor allele for five SNPs located in the promoter and intragenic regions, including the promoter SNP -1639G>A (rs9923231, Supplementary Table 1), was associated with lower mRNA expression and lower warfarin maintenance dose [55]. Moreover, the minor -1639A allele of this promoter SNP was shown to have lower activity when compared with the -1639G allele in a luciferase assay [56]. However, how and where the SNP -1639G>A

reduces promoter activity was not known. To search for functional polymorphisms that determine *VKORC1* expression, regulatory polymorphisms of *VKORC1* were analyzed in human liver, heart and B lymphocytes using AEI [57]. The effect of the promoter SNP -1639G>A was shown to be tissue-dependent, observable in the liver but not in B lymphocytes and heart tissue. The -1639A allele was associated with a twofold lower level of *VKORC1* mRNA in human liver [57], which was consistent with the lower warfarin maintenance dose in subjects carrying the haplotype marked by the -1639A allele [55]. ChIP with antibodies against H3K4me3 or acetyl-Histone H3 (Lys 4) (H3K4ac) revealed preferential association of the promoter -1639G allele with active chromatin, consistent with enhanced mRNA expression. The minor -1639A allele is thought to generate a suppressor E-box-binding site, which could lead to lower *VKORC1* expression [57], and a lower effective dose of warfarin. This site is thought to recruit TFs that suppress gene expression by activating repressive histone modification complexes. Thus, the common variant in the *VKORC1* promoter region (-1639G>A, rs9923231) can explain much of the variability in average dose requirements among Caucasians, and it is incorporated in the warfarin-dosing algorithm to improve warfarin treatment outcome [58,59].

Expression quantitative trait loci (eQTLs) have also been used to identify pharmacogenomic-associated variants. Different than AEI, which compares the relative expression of two alleles in the same individual as a phenotype influenced only by *cis*-acting genetic variants [54,60], eQTLs are genomic loci that show a correlation between RNA expression levels and SNPs located in *cis* or in *trans* [60]. SNPs associated with chemotherapeutic agent-induced cytotoxicity for six different anticancer agents were systematically evaluated for their genomic regions and their functional class, such as coding (consisting of missense, nonsense or frameshift polymorphisms), noncoding (such as 3′ UTRs or splice sites) or eQTLs [61]. The chemotherapeutic drug susceptibility-associated SNPs were more likely to be associated with the level of gene expression (as eQTLs). Notably, these SNPs are associated with the transcriptional expression level of multiple genes (≥ten genes), as potential master regulators. This suggests that the pharmacological effects of a drug may depend on differences in the expression level of many genes. For example, the SNP rs1649942 is associated with sensitivity for both carboplatin and cisplatin, and with the transcriptional expression level of 39 genes throughout the genome. Therefore, the function of SNPs associated with chemotherapeutic drug susceptibility may be a result of their role in the regulation of gene expression [61].

Genomic assays can also be used to uncover drug-associated elements in a genome-wide manner. For example, RNA-seq and ChIP-seq for various marks can be used on relevant cell or tissue types treated with the drug of interest or vehicle control. Using this approach, a genome-wide catalog of PXR targets induced by rifampin was generated [17]. RNA-seq and ChIP-seq were performed by using antibodies against PXR and three active regulatory marks (p300, H3K4me1, H3K27ac) on primary human hepatocytes treated with rifampin or vehicle control. Rifampin and PXR were chosen since they are part of the CYP3A4 pathway, which accounts for Phase I metabolism of over 50% of the most commonly prescribed drugs [62]. Annotation of genomic regions bearing a conditional PXR occupancy, as well as all three active regulatory marks, identified 1297 genomic regions, which are enriched near genes that function in the metabolism of xenobiotics, specifically members of the cytochrome P450 family [17].

Functional characterization of gene regulatory elements

It is important to note that the aforementioned techniques (ChIP-seq, DHS-seq) used in the identification of gene regulatory elements are descriptive and only identify potential regulatory elements. In order to assess whether these regulatory elements are functional, one needs to perform functional assays. For promoters, candidate promoter sequences are placed in front of a reporter gene and tested for their ability to drive reporter gene activity in fitting cell lines or tissues. In the PXR example [17] mentioned above, 227 candidate promoters for genes that were found to show rifampin-dependent expression or nearby PXR/p300 occupancy sites were tested for promoter activity by checking their ability to induce luciferase in rifampin-treated HepG2 cells. These experiments found that only 10 (4.4%) of the assayed promoters exhibited drug-dependent activity, suggesting that other factors could be involved in rifampin drug response [17].

Commonly used enhancer assays place the assayed sequence in front of a minimal promoter (a promoter that should only get activated if it has an enhancer in front of it) followed by a reporter gene. If the assayed sequence is an enhancer, it will turn on the minimal promoter, which in turn will drive the reporter gene expression. For example, enhancer assays in rifampin-treated HepG2 cells for 42 potential rifampin-induced sequences as well as 7 sequences that overlap LD blocks defined by associated SNPs from pharmacogenomic GWAS revealed 15/42 and 4/7 to be functional enhancers, respectively [17]. This illustrates that not all ChIP-seq marked regions are necessarily functional enhancers. Combined, these results suggest that enhancers could have a major role in rifampin-induced

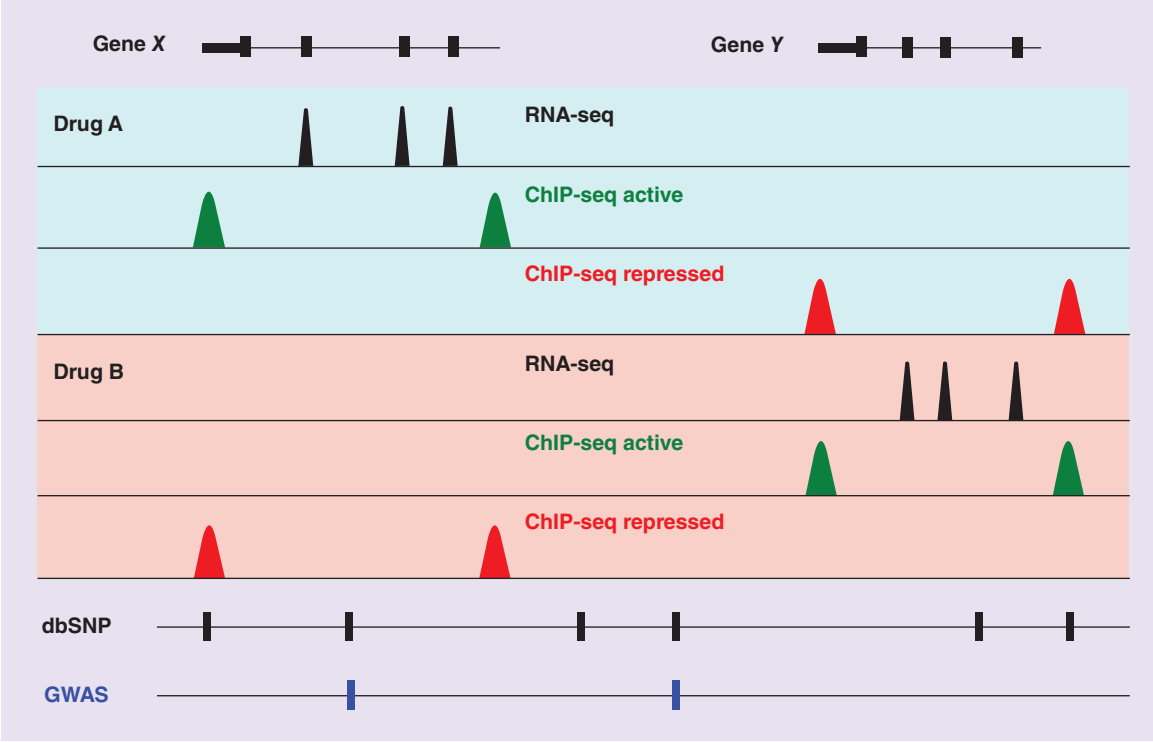


Figure 2. Drug regulatory maps. A genomic region containing the genes *X* and *Y* and tracks for RNA-seq and ChIP-seq data for active (green) and repressed (red) marks after drug A and drug B treatment. Gene *X* is transcriptionally activated by drug A (RNA-seq track), and ChIP-seq tracks show active marks (green) for drug A and repressed marks (red) for drug B in the vicinity of this gene. Conversely, gene *Y* is activated by drug B (RNA-seq track), and the ChIP-seq tracks show active marks for drug B and repressed marks for drug A in the vicinity of gene *Y*. At the bottom, tracks for SNPs from the dbSNP database and GWAS-associated SNPs in this genomic region are shown. These SNPs may overlap with ChIP-seq peaks or may be in linkage disequilibrium with SNPs within those peaks.
ChIP: Chromatin immunoprecipitation; GWAS: Genome-wide association study.

PXR activation, and the genomic techniques used in this study (RNA-seq, ChIP-seq) can serve as a model for the identification of drug-responsive regulatory elements [17].

Conclusion & future perspective

By carrying out a review of published GWAS in the NHGRI catalog, we found that 96.4% of the associated SNPs from 108 pharmacogenomic GWAS reside in noncoding regions. These results suggest that the majority of common variants leading to interindividual differences in drug response could be regulatory. It is also important to note that a GWAS identified associated SNP is not necessarily the causative SNP, and another SNP or a combination of SNPs that are in LD with this SNP could be causative. Given the large number of pharmacogenomic GWAS associations within noncoding regions, follow-up fine genetic mapping and functional studies will be required to identify the causal variants.

Sequencing technologies have improved our ability to identify drug-responsive gene regulatory elements. However, there are still only a few examples in the literature for nucleotide variants in enhancers that are associated with drug response. As the techniques used to identify and functionally characterize nucleotide variants are being continuously improved, we expect an increase in the number of examples within the next few years. It is important to note that the techniques used in the identification of drug-responsive regulatory elements just find predicted regulatory elements, and functional assays need to be done to assess their function. Although promoter and enhancer functional assays are being currently conducted as standard techniques to functionally characterize gene regulatory variants once they are discovered, they can pose as a bottleneck because they are carried out on a 'one-by-one' basis. In order to overcome this hurdle, massively parallel reporter assays, which can assay thousands of sequences and variants for regulatory activity [63,64], could be applied to analyze drug-responsive promoters and enhancers.

Genomic assays could be an efficient means to identify additional drug-responsive regulatory elements in a genome-wide manner. For example, RNA-seq and ChIP-seq for various marks can be used on relevant

cell or tissue types treated with the drug of interest or vehicle control, like the genome-wide catalog of PXR targets induced by rifampin [17]. The future generation of additional 'drug regulatory maps' could provide ideal candidate sequences where nucleotide variation can influence drug response (Figure 2). In addition, these 'drug regulatory maps' will be of extreme importance as individual genomes will become more commonly applied to drug treatments.

We expect that there will be many challenges before functionally confirmed drug response-associated nucleotide variation in regulatory elements could be implemented in the clinic. First, the generation of enormous amounts of sequence data presents a challenge for the extraction of clinically useful and actionable information in order to validate significant genotype–phenotype associations. Second, the variants with potential pharmacogenomics effect uncovered within drug-responsive regulatory elements must be validated for both the molecular mechanism and their clinical effect. The majority of these variants could be rare and unique to a certain individual, thus warranting the need for high-throughput and easy to carry out functional assays. Third, since the effects of these regulatory variants on gene regulation are expected to be complex, the effects of nonlinear interactions between multiple genetic variants must be taken into account, and multigenic predictors of pharmacogenomic response must be developed and established as decision-support tools, so that they can be properly implemented in the clinical practice. However, the future integration of these 'drug regulatory maps', gene–gene and gene–environment interactions, and rare variants, may guide the translation of novel pharmacogenomics findings with potential impact to enhance efficacy and reduce adverse effects in the clinic.

Supplementary data

To view the supplementary data that accompany this paper ,please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/PGS.15.121

Executive summary

- 96.4% of the 928 associated SNPs from the 108 pharmacogenomics genome-wide association studies reviewed are noncoding. These findings suggest that the majority of common variants leading to interindividual differences in drug response are regulatory.

Promoter variants & drug response

- *SULT2A1* was uncovered as a novel liver X receptor α target. The activation of liver X receptor α induced the expression of *SULT2A1.* Liver X receptor α transactivated the *SULT2A1* promoter through its specific binding to the -500 to -258 bp region, as shown by promoter reporter and ChIP assays.
- EWS-FLI1 binds near the transcription start site of the *SLFN11* promoter and acts as a positive regulator of *SLFN11* expression in Ewing's sarcoma cells, as shown by ChIP-seq and promoter reporter assays. EWS-FLI1-mediated *SLFN11* expression is responsible for high sensitivity of Ewing's sarcoma to camptothecin and combinations of PARP inhibitors with temozolomide.

Enhancers & drug response

- A novel functional MEF2C-binding site 148 kb upstream of the *KLF4* gene was identified by ChIP-seq and luciferase assays. MEF2C-bound enhancer and transcription start site of *KLF4* gene were observed to come into closer spatial proximity after pitavastatin treatment by ChIA-PET and 3C assays.
- Two enhancer modules located upstream of the transcription start site of the *SLC13A5* gene associated with regulation of PXR-mediated *SLC13A5* induction were identified and functionally characterized by ChIP and luciferase reporter assays.
- An enhancer region located ~115 kb downstream of *CYP2D6* was initially identified, and a SNP within it (rs5758550) was shown to affect gene expression using 4C, reporter gene assays, ChIP and CRISPR-mediated enhancer deletion.

Identification of drug-responsive regulatory elements

- Allelic expression imbalance and expression quantitative trait loci have been used to identify pharmacogenomic-associated variants.
- Genomic assays can also be used to uncover drug-associated elements. As an example, we describe the genome-wide catalog of PXR targets induced by rifampin that was characterized using RNA-seq and ChIP-seq.

Conclusion & future perspective

- Sequencing technologies have improved our ability to identify drug-responsive gene regulatory elements. However, there are only a few examples for nucleotide variants in enhancers that are associated with drug response.
- Future 'drug regulatory maps' will uncover additional drug-responsive regulatory elements in a genome-wide manner. These findings may guide the translation of novel pharmacogenomics findings with potential impact to enhance efficacy and reduce adverse effects in the clinic.

Financial & competing interests disclosure

This work was supported by a grant from the National Institute of General Medical Sciences GM61390. N Ahituv is also supported in part by grants by the National Institute of Neurological Disorders & Stroke 1R01NS079231, National Institute of Diabetes & Digestive & Kidney Diseases 1R01DK090382 and National Cancer Institute 1R01CA197139. MR Luizon was supported in part by a fellowship of the Coordination of Improvement of Higher Level Personnel (CAPES-Brazil) and also by the Young Talent Attraction Fellowship-BJT from the National Council of Scientific and Technological Development (CNPq-Brazil). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Hindorff LA, Sethupathy P, Junkins HA *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* 106(23), 9362–9367 (2009).

2 Welter D, Macarthur J, Morales J *et al.* The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006 (2014).

• **The National Human Genome Research Institute (NHGRI) catalog of published genome-wide association study (GWAS) catalog. We reviewed this publicly available manually curated collection of published GWASs for 108 pharmacogenomic GWAS and discovered a total of 928 associated SNPs, 96.4% of them are noncoding.**

3 Huang D, Ovcharenko I. Identifying causal regulatory SNPs in ChIP-seq enhancers. *Nucleic Acids Res.* 43(1), 225–236 (2015).

• **A computational model that uses chromatin immunoprecipitation-sequencing intensity variation in response to noncoding allelic change as a proxy for the quantification of the biological role of noncoding SNPs was developed and applied to HepG2 enhancers and detected 4796 enhancer SNPs capable of disrupting enhancer activity upon allelic change.**

4 Kumar V, Westra HJ, Karjalainen J *et al.* Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet.* 9(1), e1003201 (2013).

5 Mathelier A, Shi W, Wasserman WW. Identification of altered *cis*-regulatory elements in human disease. *Trends Genet.* 31(2), 67–76 (2015).

6 Sakabe NJ, Savic D, Nobrega MA. Transcriptional enhancers in development and disease. *Genome Biol.* 13(1), 238 (2012).

7 Ahituv N. Gene regulatory elements. In: *Gene Regulatory Sequences and Human Disease.* Ahituv N (Ed.). Springer-Verlag, NY, USA, 1–17 (2012).

8 Daly AK. Genome-wide association studies in pharmacogenomics. *Nat. Rev. Genet.* 11(4), 241–246 (2010).

9 Daly AK. Using genome-wide association studies to identify genes important in serious adverse drug reactions. *Annu. Rev. Pharmacol. Toxicol.* 52, 21–35 (2012).

10 Low SK, Takahashi A, Mushiroda T, Kubo M. Genome-wide association study: a useful tool to identify common genetic variants associated with drug toxicity and efficacy in cancer pharmacogenomics. *Clin. Cancer Res.* 20(10), 2541–2552 (2014).

11 Motsinger-Reif AA, Jorgenson E, Relling MV *et al.* Genome-wide association studies in pharmacogenomics: successes and lessons. *Pharmacogenet. Genomics* 23(8), 383–394 (2013).

12 Zhou K, Pearson ER. Insights from genome-wide association studies of drug response. *Annu. Rev. Pharmacol. Toxicol.* 53, 299–310 (2013).

•• **In this review, the authors outline the challenges and recent successes of the GWAS approach in disease genetics and the ways in which these can be applied to pharmacogenomics for biological discovery, determination of heritability and personalized treatment.**

13 Hindorff LA, MacArthur J (European Bioinformatics Institute), Morales J (European Bioinformatics Institute. A catalog of published genome-wide association studies. www.genome.gov

14 Chhibber A, Kroetz DL, Tantisira KG *et al.* Genomic architecture of pharmacological efficacy and adverse events. *Pharmacogenomics* 15(16), 2025–2048 (2014).

•• **In this review, the authors highlight several factors that must be taken into account in future pharmacogenomics research, as the combined effects of multiple genetic variants, nonprotein coding functional genetic variants and prioritize laboratory characterization of molecular mechanisms.**

15 Sadee W, Hartmann K, Seweryn M, Pietrzak M, Handelman SK, Rempala GA. Missing heritability of common diseases and treatments outside the protein-coding exome. *Hum. Genet.* 133(10), 1199–1215 (2014).

•• **Genetic factors strongly influence risk of common human diseases and treatment outcomes, but the causative variants remain largely unknown. In this review, the authors propose hypotheses that in combination may be responsible to a large extent for this so-called 'missing heritability'.**

16 Sadee W. The relevance of "missing heritability" in pharmacogenomics. *Clin. Pharmacol. Ther.* 92(4), 428–430 (2012).

17 Smith RP, Eckalbar WL, Morrissey KM *et al.* Genome-wide discovery of drug-dependent human liver regulatory elements. *PLoS Genet.* 10(10), e1004648 (2014).

•• **Genomic assays can be an efficient means to identify drug-responsive regulatory elements in a genome-wide manner. Using RNA-seq and chromatin immunoprecipitation-sequencing for various marks, this article provides a genome-wide catalog of PXR targets induced by rifampin.**

18 Smith RP, Lam ET, Markova S, Yee SW, Ahituv N. Pharmacogene regulatory elements: from discovery to applications. *Genome Med.* 4(5), 45 (2012).

• **A review that outlined examples of how regulatory elements affect drug response via regulation of drug targets, transporters and enzymes, and discussed the impact of sequencing technologies on pharmacogenomics.**

19 Butler JE, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16(20), 2583–2592 (2002).

20 Smale ST, Kadonaga JT. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72, 449–479 (2003).

21 Calhoun VC, Stathopoulos A, Levine M. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the drosophila antennapedia complex. *Proc. Natl Acad. Sci. USA* 99(14), 9243–9247 (2002).

22 Pennacchio LA, Ahituv N, Moses AM *et al. In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* 444(7118), 499–502 (2006).

23 Visel A, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA. Functional autonomy of distant-acting human enhancers. *Genomics* 93(6), 509–513 (2009).

24 Birnbaum RY, Clowney EJ, Agamy O *et al.* Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* 22(6), 1059–1068 (2012).

25 Birnbaum RY, Patwardhan RP, Kim MJ *et al.* Systematic dissection of coding exons at single nucleotide resolution supports an additional role in cell-specific transcriptional regulation. *PLoS Genet.* 10(10), e1004592 (2014).

26 Tumpel S, Cambronero F, Sims C, Krumlauf R, Wiedemann LM. A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2. *Proc. Natl Acad. Sci. USA* 105(51), 20077–20082 (2008).

27 Heintzman ND, Hon GC, Hawkins RD *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459(7243), 108–112 (2009).

28 Felsenfeld G, Boyes J, Chung J, Clark D, Studitsky V. Chromatin structure and gene expression. *Proc. Natl Acad. Sci. USA* 93(18), 9384–9388 (1996).

29 Thurman RE, Rynes E, Humbert R *et al.* The accessible chromatin landscape of the human genome. *Nature* 489(7414), 75–82 (2012).

30 Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414), 57–74 (2012).

31 Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* 2010(2), pdb prot5384 (2010).

32 Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* 316(5830), 1497–1502 (2007).

33 Robertson G, Hirst M, Bainbridge M *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4(8), 651–657 (2007).

34 Visel A, Blow MJ, Li Z *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457(7231), 854–858 (2009).

35 Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* 13(12), 840–852 (2012).

36 Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10(10), 669–680 (2009).

37 Lieberman-Aiden E, Van Berkum NL, Williams L *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950), 289–293 (2009).

38 Fullwood MJ, Liu MH, Pan YF *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462(7269), 58–64 (2009).

39 Innocenti F, Schilsky RL, Ramirez J *et al.* Dose-finding and pharmacokinetic study to optimize the dosing of irinotecan according to the *UGT1A1* genotype of patients with cancer. *J. Clin. Oncol.* 32(22), 2328–2334 (2014).

40 Toffoli G, Cecchin E, Gasparini G *et al.* Genotype-driven Phase I study of irinotecan administered in combination with fluorouracil/leucovorin in patients with metastatic colorectal cancer. *J. Clin. Oncol.* 28(5), 866–871 (2010).

41 Sugatani J. Function, genetic polymorphism, and transcriptional regulation of human UDP-glucuronosyltransferase (UGT) 1A1. *Drug Metab. Pharmacokinet.* 28(2), 83–92 (2013).

42 Hu DG, Meech R, Mckinnon RA, Mackenzie PI. Transcriptional regulation of human UDP-glucuronosyltransferase genes. *Drug Metab. Rev.* 46(4), 421–458 (2014).

43 Ou Z, Jiang M, Hu B *et al.* Transcriptional regulation of human hydroxysteroid sulfotransferase SULT2A1 by LXRα. *Drug Metab. Dispos.* 42(10), 1684–1689 (2014).

44 Tang SW, Bilke S, Cao L *et al.* SLFN11 is a transcriptional target of EWS-FLI1 and a determinant of drug response in ewing's sarcoma. *Clin. Cancer Res.* 1(18), 4184–4193 (2015).

45 Maejima T, Inoue T, Kanki Y *et al.* Direct evidence for pitavastatin induced chromatin structure change in the KLF4 gene in endothelial cells. *PLoS ONE* 9(5), e96005 (2014).

46 Li L, Li H, Garzel B *et al.* SLC13A5 is a novel transcriptional target of the pregnane x receptor and sensitizes drug-induced steatosis in human liver. *Mol. Pharmacol.* 87(4), 674–682 (2015).

47 Wang D, Papp AC, Sun X. Functional characterization of CYP2D6 enhancer polymorphisms. *Hum. Mol. Genet.* 24(6), 1556–1562 (2015).

48 Wang D, Poi MJ, Sun X, Gaedigk A, Leeder JS, Sadee W. Common CYP2D6 polymorphisms affecting alternative splicing and transcription: long-range haplotypes with two regulatory variants modulate CYP2D6 activity. *Hum. Mol. Genet.* 23(1), 268–278 (2014).

49 Georgitsi M, Zukic B, Pavlovic S, Patrinos GP. Transcriptional regulation and pharmacogenomics. *Pharmacogenomics* 12(5), 655–673 (2011).

50 Bilke S, Schwentner R, Yang F *et al.* Oncogenic ETS fusions deregulate E2F3 target genes in ewing sarcoma and prostate cancer. *Genome Res.* 23(11), 1797–1809 (2013).

51 Mikkelsen TS, Ku M, Jaffe DB *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448(7153), 553–560 (2007).

52 Quinodoz M, Gobet C, Naef F, Gustafson KB. Characteristic bimodal profiles of RNA polymerase ii at thousands of active mammalian promoters. *Genome Biol.* 15(6), R85 (2014).

53 Gaedigk A, Simon SD, Pearce RE, Bradford LD, Kennedy MJ, Leeder JS. The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. *Clin. Pharmacol. Ther.* 83(2), 234–242 (2008).

54 Johnson AD, Zhang Y, Papp AC *et al.* Polymorphisms affecting gene transcription and mRNA processing in pharmacogenetic candidate genes: detection through allelic expression imbalance in human target tissues. *Pharmacogenet. Genomics* 18(9), 781–791 (2008).

55 Rieder MJ, Reiner AP, Gage BF *et al.* Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N. Engl. J. Med.* 352(22), 2285–2293 (2005).

56 Yuan HY, Chen JJ, Lee MT *et al.* A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity. *Hum. Mol. Genet.* 14(13), 1745–1751 (2005).

57 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* 112(4), 1013–1021 (2008).

58 International Warfarin Pharmacogenetics C, Klein TE, Altman RB *et al.* Estimation of the warfarin dose with clinical and pharmacogenetic data. *N. Engl. J. Med.* 360(8), 753–764 (2009).

59 Johnson JA, Gong L, Whirl-Carrillo M *et al.* Clinical pharmacogenetics implementation consortium guidelines for *CYP2C9* and *VKORC1* genotypes and warfarin dosing. *Clin. Pharmacol. Ther.* 90(4), 625–629 (2011).

60 Smith RM, Webb A, Papp AC *et al.* Whole transcriptome RNA-seq allelic expression in human brain. *BMC Genomics* 14, 571 (2013).

61 Gamazon ER, Huang RS, Cox NJ, Dolan ME. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proc. Natl Acad. Sci. USA* 107(20), 9287–9292 (2010).

62 Rahmioglu N, Heaton J, Clement G *et al.* Genome-wide association study reveals a complex genetic architecture underpinning-induced CYP3A4 enzyme activity. *Eur. J. Drug Metab. Pharmacokinet.* 38(1), 63–67 (2013).

63 Haberle V, Lenhard B. Dissecting genomic regulatory elements *in vivo*. *Nat. Biotechnol.* 30(6), 504–506 (2012).

64 Patwardhan RP, Hiatt JB, Witten DM *et al.* Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat. Biotechnol.* 30(3), 265–270 (2012).