Title Page:

Cytochrome P450 Transcriptional Regulation by TSPYLs: Identification of Novel Upstream Transcription Factors

Authors:

Suganti Shivaram*, Huanyao Gao*, Sisi Qin, Duan Liu, Richard M. Weinshilboum#, Liewei Wang#

* these two authors contributed equally to this manuscript.

Affiliations:

Division of Clinical Pharmacology, Department of Molecular Pharmacology and Therapeutics, Mayo Clinic, Rochester, Minnesota (S.S., H.G., S.Q., D.L., R.M.W, L.W.)
Running Title: CYP Expression and TSPYLs: Role of Upstream TFs

#Corresponding authors:

Liewei Wang, M.D., Ph.D.
Department of Molecular Pharmacology & Experimental Therapeutics
Mayo Clinic
200 First Street Southwest, Rochester, MN 55905
Tel: +1 (507)284-5264
Fax: +1 (507)284-4455
E-mail: Wang.Liewei@mayo.edu;

Richard M. Weinshilboum, M.D.
Department of Molecular Pharmacology & Experimental Therapeutics
Mayo Clinic
200 First Street SW, Rochester, MN 55905
Tel: +1 (507)284-2246
Fax: +1 (507)284-4455
Email: Weinshilboum.Richard@mayo.edu
The number of text pages: 11

number of tables: 0

number of figures: 5

number of references: 30

number of words in the Abstract: 250

number of words in the Introduction: 679

number of words in the Discussion: 690
Nonstandard abbreviations used in the paper

ADE Adverse Drug Event
ChIP chromatin immunoprecipitation
CYP cytochrome P450
DME drug-metabolizing enzyme
eQTL expression quantitative trait loci
KD knock-down
MAF minor allele frequencies
OE overexpression
PK pharmacokinetic
PD pharmacodynamic
RT-qPCR reverse transcription real time polymerase chain reaction
SD standard deviation
siRNA small interfering RNA
SNP single-nucleotide polymorphism
TF transcription factors
TSPYL Testis-specific Y-encoded-like protein
TSS transcription starting site
Abstract

Cytochrome P450s (CYPs) display significant inter-individual variation in expression, much of which remains unexplained by known CYP single-nucleotide polymorphisms (SNPs). Testis-specific Y-encoded-like proteins (TSPYLs) are transcriptional regulators for several drug-metabolizing CYPs including CYP3A4. However, transcription factors (TFs) that might influence CYP expression through an effect on TSPYL expression are unknown. Therefore, we studied regulators of TSPYL expression in hepatic cell lines and their possible SNP-dependent variation. Specifically, we identified candidate TFs that might influence TSPYL expression using the ENCODE ChIPseq database. Subsequently, the expression of TSPYL1/2/4 as well as that of selected CYP targets for TSPYL regulation were assayed in hepatic cell lines before and after knockdown of TFs that might influence CYP expression through TSPYL-dependent mechanisms. Those results were confirmed by studies of TF binding to TSPYL1/2/4 gene promoter regions. In hepatic cell lines, knockdown of the REST and ZBTB7A TFs resulted in decreased TSPYL1 and TSPYL4 expression and increased CYP3A4 expression, changes reversed by TSPYL1/4 overexpression. Potential binding sites for REST and ZBTB7A on the promoters of TSPYL1 and TSPYL4 were confirmed by chromatin immunoprecipitation. Finally, common SNP variants in upstream binding sites on the TSPYL1/4 promoters were identified and luciferase reporter constructs confirmed SNP-dependent modulation of TSPYL1/4 gene transcription. In summary, we identified REST and ZBTB7A as regulators of the expression of TSPYL genes which themselves can contribute to regulation of CYP expression and—potentially—of drug metabolism. SNP-dependent modulation of TSPYL transcription may contribute to individual variation in both CYP expression and—downstream—drug response phenotypes.
Significance Statement: (55/80 words) Testis-specific Y-encoded-like proteins (TSPYLs) are transcriptional regulators of cytochrome P450 (CYP) gene expression. Here we report that variation in TSPYL expression as a result of the effects of genetically regulated TSPYL transcription factors is an additional factor that could result in downstream variation in CYP expression and potentially, as a result, variation in drug biotransformation.
**Introduction:**

Approximately half of the population of the United States uses prescription drugs every year (CDC, 2015-2018b). Adverse Drug Events (ADEs) and toxicity as a result of prescription drug use could potentially be decreased by enhanced understanding of variation in pharmacokinetic (PK) and/or pharmacodynamic (PD) factors that contribute to inter-individual differences in drug exposure or response (CDC, 2015-2018a). Many studies ranging from candidate gene studies to genome-wide analyses have highlighted the contribution of genomics to individual variation in the occurrence of ADEs and/or inter-individual variability in drug response phenotypes (Nebert et al., 2013; Nelson, 2013; Zanger and Schwab, 2013). The cytochrome P450 (CYP) enzymes play an important role in Phase I drug metabolism and, as a result, have the potential to be major contributors to individual variability in PK. The CYP superfamily includes 18 families of protein encoding human CYP genes including the CYP1, CYP2 and CYP3 families which include many important drug-metabolizing enzymes (DMEs) (Bush et al., 2016; Kozyra et al., 2017).

CYP3A4, CYP2C9 and CYP2C19 are important CYPs with common, functionally significant genetic polymorphisms (Evans and Relling, 1999). These three enzymes have been estimated to contribute to the metabolism of approximately 50%, 20% and 5% of drugs, respectively (Evans and Relling, 1999; Neavin et al., 2019). However, known SNP variants that influence the expression or function of CYP3A4, CYP2C9 and CYP2C19 explain only a portion of inter-individual differences in drug biotransformation catalyzed by these CYPs (Daly, 2010; Liu et al., 2010; Zi et al., 2010; He et al., 2011; Wang et al., 2011; Motsinger-Reif et al., 2013; Wright et al., 2016). Most of the early examples of CYP pharmacogenomic variation involved SNPs that resulted in alterations in the amino acid sequence of the encoded protein, alterations in gene splicing or variation in gene structure (deletions/insertions) but, increasingly, it is becoming
apparent that variants which alter gene transcription represent a major source of pharmacogenomic variation—either directly or indirectly (Wang et al., 2022).

Previous studies from our group reported that testis-specific Y-encoded-like proteins (specifically TSPYL1, 2 and 4) are transcriptional regulators that can influence the expression of CYP3A4, CYP2C9 and CYP2C19 (Qin et al., 2017). Elevated expression of these TSPYLs can suppress the expression of CYP3A4, CYP2C9 and CYP2C19 (Qin et al., 2017). The TSPYL gene family consists of six genes, TSPYL1 to TSPYL6, with TSPYL3 being a pseudogene. The Genotype-Tissue Expression (GTEx) database (https://gtexportal.org/) reports that the TSPYLs are expressed in most human tissues, with isoform-specific variation in their tissue distribution. TSPYLs have multiple cellular functions (de Andrade et al., 2006; Epping et al., 2015), and genetic polymorphisms and/or variation in the methylation status of these genes have been related to disease states (Kim et al., 2006; Guo et al., 2012; Le Gallo et al., 2012). In addition, as stated above, functionally significant polymorphisms in TSPYL genes have been reported to alter their ability to regulate transcription and, as a result, the expression of CYPs, resulting in inter-individual variation in drug biotransformation (Qin et al., 2017; Qin et al., 2019).

Specifically, and of importance for the studies described subsequently, we reported previously that knock-down (KD) of TSPYL1,2 and 4 in HepaRG cells can result in increased expression of CYP2C9, 2C19 and 3A4, while over-expression (OE) of these same TSPYL genes can result in decreased expression of the same CYPs—with the most striking effects for CYP3A4 (Qin et al., 2017; Qin et al., 2019). Given our increasing recognition of the role of the TSPYLs in drug metabolism, it would be important to understand the possible role of upstream regulators of TSPYL gene expression, specifically, transcription factors (TFs) that influence TSPYL gene expression, to help us achieve a more comprehensive understanding of downstream variability in
CYP expression and drug response phenotypes. In the present study, we set out systematically to identify TFs that might be involved in the regulation of TSPYL1, 2 and 4 expression in human hepatic cell lines as a step toward a more comprehensive understanding of the potential contribution of the TSPYLs to individual variation in CYP expression and function.

**Materials and Methods**

**ENCODe ChIP-Seq data:** The ENCODE UCSC genome browser includes ChIP-seq data for HepG2 cells that lists TFs that bind to the promoter regions of TSPYL1, TSPYL2 and TSPYL4, 1kbp upstream or downstream of the transcription starting site. We used that information as a starting point for this series of studies of possible transcriptional regulatory factors that might contribute to variation in the expression of human TSPYLs.

**Hepatic eQTL database association analysis:** We next determined associations between TFs that bind to TSPYL gene promoters and TSPYL1, TSPYL2 and TSPYL4 expression in an hepatic eQTL database (Storey et al., 2011) using Pearson correlation analyses and identified TFs that might bind near TSPYL1, TSPYL2 or TSPYL4 and, as a result, might influence the expression of genes influenced by TSPYL expression with p-values < 0.05.

**Transfection of HepaRG cells and HepG2 cells:** Specific short interfering RNAs (siRNAs) targeting the 30 candidate TFs that we had identified in the ENCODE database were then used to knock down the expression of those TFs in HepaRG and HepG2 cells using specific siRNAs, with non-targeting siRNAs as a control. (see **Supp. Table 1**) The cells were harvested 48 hours after transfection and RNA was extracted for the performance of RT-PCR.

**Gene expression quantification:** Total RNA from HepG2 and HepaRG cells was extracted and was used to perform real time quantitative PCR to assay expression levels of TSPYL1, TSPYL2, TSPYL4, CYP3A4, CYP2C9 and CYP2C19, as well as after the after KD of TFs using the primers.
listed in **Supp. Table 2**. Alterations in the expression of those genes were expressed as fold change from baseline.

**Chromatin Immunoprecipitation (ChIP) for HepaRG cells:** HepaRG cells were used to perform ChIP assays to validate TF binding to promoter regions of the *TSPYL1* and *TSPYL4* genes and the results were analyzed using real-time quantitative PCR. *TSPYL1* and *TSPYL4* were selected for study because they map in relatively close proximity in the genome and because our previous experiments had demonstrated that those two *TSPYL* genes appeared to have significant impact on variation in the expression of *CYP3A4* (Qin et al., 2017).

**Luciferase Reporter Assay:**

The luciferase reporter vector, pGL4.23 (Cat#: E8411) was obtained from Promega with inserts encoding either 2Kbp of the *TSPYL1* or 2Kbp of the *TSPYL4* promoter regions and were used to create *TSPYL1* Wild Type promoter, *TSPYL1* variant promoter, *TSPYL4* wild type promoter or *TSPYL4* variant promoter constructs (**Supp. Table 4**). Those vectors were then used to transfect HepaRG cells. The cells were harvested 48 hrs after transfection to assay relative luciferase and Renilla activities.

Additional methodological details have been provided as **Supplementary Methods**.

**Results:**

The series of studies described subsequently was designed to pursue our previous observation of the potential importance of members of the *TSPYL* gene family in regulation of the expression of drug metabolizing CYPs (Qin et al, 2017, 2019). Specifically:

1. As a first step in the present studies, the ENCODE database was consulted to identify TFs that might bind to the promoters of the *TSPYL1/2* and *4* genes in HepG2 cells. Thirty potential candidate TFs were identified.
2. Those 30 candidate *TSPYL* TFs were then knocked down (KD) in HepaRG cells and the effect of KD on the expression of *CYP3A4*, *CYP2C9* and *CYP2C19* was determined and compared with our previous results after the KD of *TSPYL1, 2* and *4* in this same cell line. The most striking similarities observed related to *CYP3A4* and the putative TSPYL TFs *REST* and *ZBTB7A*. Therefore, the final series of studies focussed on SNPs in the promoters of *TSPYL1* and *TSPYL4*—two genes that map in close proximity to each other in the genome—as well as the effect of *REST* and *ZBTB7A* on their transcription.

3. The final series of experiments addressed the possible binding of *REST* and *ZBTB7A* to the promoters of *TSPYL1* and *TSPYL4* and the influence of SNPs in those genes on that binding and the expression of those two TSPYLs. Neither *REST* nor *ZBTB7A* appeared to bind to the promoter of *TSPYL2*, so *TSPYL2* was not included in this series of experiments.

**Candidate Transcriptional Regulators of *TSPYL1, TSPYL2* and *TSPYL4***: The initial list of candidate TFs that might participate in regulation of the expression of *TSPYL1, TSPYL2* and *TSPYL4* was assembled based on their ability to bind to promoter regions of the genes encoding these three TSPYLs based on ENCODE data for HepG2 cells. Specifically, using HepG2 cell ChIP-seq data, we identified TFs that bound to 2 Kb regions extending 1 Kb on either side of the transcription start sites for *TSPYL1, TSPYL2* or *TSPYL4*. As the next step, significant correlations between expression levels of these candidate TFs and the expression of *TSPYL1, TSPYL2* and *TSPYL4* were determined by Pearson correlation analysis of hepatic eQTL expression data obtained from the GTEx database--with the full understanding that hepatic tissue expression might differ significantly from that for either HepG2 or HepaRG cells, the two cell lines used in our experiments. TFs with correlation coefficients> 0.2 were then advanced to the next step of
the analysis. By applying this step-wise narrowing-down process, we identified the 30 candidate TFs that are listed in Figure 1. Those 30 candidate TFs were then knocked down in HepaRG cells using siRNAs with KD efficiencies as shown graphically in Supp. Figure 1. We used HepaRG rather than HepG2 cells in these experiments because they have been reported to better reflect the biology of hepatocytes than do HepG2 cells (Ramboer et al., 2015). The mRNA expression levels of TSPYL1, TSPYL2, TSPYL4 and of CYP3A4, CYP2C9 and CYP2C19 were then assayed in HepaRG cells by qRT-PCR as shown in Figure 2 for CYP3A4. Panels a), b), and c) in Figure 2 display data for the expression of CYP3A4 versus those of TSPYL1, TSPYL2 and TSPYL4, respectively after KD of the candidate TFs, with each black or red circle representing one of the 30 TFs studied and with error bars showing the impact of KD of the indicated TSPYL as a vertical line and the impact on the expression of CYP3A4 as a horizontal line. We have highlighted points in Figure 2 in red in which the relationship of the expression of those TFs mapped to the lower right quadrant of the four quadrant graphical representations of the data—i.e., these were TFs that were associated with increased expression of CYP3A4 in the setting of decreased TSPYL expression—a relationship that we had reported previously in our KD and OE studies of TSPYL genes in HepaRG cells (Qin et al., 2017). After excluding TFs already known to be involved in the regulation of CYP3A4 expression (Martinez-Jimenez et al., 2007; Jover et al., 2009) as well as TFs known to bind to the promoter region of CYP3A4 based on ENCODE data, 18 TFs were found to significantly influence the mRNA expression of CYP3A4 and at least one of the TSPYLs in HepaRG cells (fold change ≥ 2). However, only the KD of REST, MAFK and ZBTB7A resulted in the down regulation of TSPYL expression coupled with the up regulation of CYP3A4 expression (see Figure 2 panel c), consistent with our previous findings. KD of these three TFs resulted in striking upregulation of CYP3A4 expression.
coupled with the downregulation of TSPYL4 expression. **Supp. Fig. 1 a)** through **f)** shows similar data for CYP2C9 and CYP2C19. However, for CYP2C9 and 2C19, as shown in the Figure, there was a noticeable absence of points in the lower right quadrants of the Figures. As a result, *REST*, *MAFK* and *ZBTB7A* were selected for further study after a rescue experiment designed to verify our initial results and to help determine the mechanism of regulation of *CYP3A4* expression by TSPYLs—in this case focussing on TSPYL4.

**Transcriptional regulation of CYPs by TFs by regulation of TSPYL expression:** As the next step in this series of experiments, a candidate TF KD and *TSPYL4* OE rescue study was performed for *REST*, *MAFK* and *ZBTB7A*, the three TFs that displayed the most striking upregulation of *CYP3A4* expression after the down regulation of TSPYL4 (see **Figure 2c**). The results of that experiment, as shown in **Figure 3**, demonstrated that KD of *ZBTB7A*, *REST* and MAFK consistently resulted in increased expression of *CYP3A4*, while OE of *TSPYL4* decreased CYP3A4 expression. However, TSPYL4 OE was able to reverse the upregulation of *CYP3A4* expression (**Figure 3b** and **e**) only after ZBTB7A or REST KD, but not after MAFK KD (**Figure 3h**).

Based on the results of the rescue experiment, we concluded that ZBTB7A and REST clearly enhanced TSPYL4 expression which, in turn, downregulated the expression of *CYP3A4*. Similar results were observed when we used NCI-H2405 human lung adenocarcinoma cells to perform similar studies (see **Supp. Fig. 2b** and **2c**). Therefore, REST and ZBTB7A appeared to be negative regulators of CYP3A4 expression as a result, at least in part, of an effect on *TSPYL4* expression. As the next step in the analysis, ChIP-qPCR experiments were performed using HepaRG cells to validate specific TF binding sites in the promoter regions of the *TSPYL1* and *TSPYL4* genes (**Figure 4**). Specifically, primers were designed to target the promoter regions of
*TSPYL4*, and *TSPYL1* based on ZBTB7A ChIPseq performed with HepG2 cells (ENCODE experiment ENCSR000BQA) and REST ChIPseq performed with hepatic tissue (ENCODE experiments ENCSR867WPH and ENCSR893QWP) (see Supp. Table 3). We chose to study *TSPYL1* and *TSPYL4* together because those two genes map only 20kbp away from each other. The results of ChIP-qPCR, as shown in Figure 4, demonstrated significant enrichment of the binding of ZBTB7A or REST antibody to the promoter regions of both TSPYL1 and TSPYL4 as compared to IgG, indicating that both ZBTB7A and REST could bind directly to the promoter regions of these two TSPYLs.

SNP dependent modulation of transcriptional regulatory activity for TSPYL expression by REST and ZBTB7A: We next identified 6 common SNPs, rs9400898(G/C), rs3828743(G/A), rs3749895 (C/G), rs910391(T/G), rs17524614 (G/T), and rs2232470 (C/A) that mapped within the ChIPseq peaks for REST (ENCODE experiments ENCSR867WPH and ENCSR893QWP) and ZBTB7A (ENCODE experiment ENCSR000BQA) on the *TSPYL1* and *TSPYL4* gene promoter regions. Specifically, the first three SNPs mapped to the *TSPYL1* promoter region and were in tight linkage disequilibrium, with minor allele frequencies (MAFs) that ranged from 0.26 to 0.30 for the group near the transcription start site (TSS) for *TSPYL1* based on 1000 Genomes Project data (https://www.genome.gov/27528684/1000-genomes-project) (Supp. Table 5), while the latter three SNPs mapped to the *TSPYL4* promoter region with MAF values that ranged from 0.16 to 0.33, as depicted graphically in Figure 5a and Supp. Table 6. Luciferase reporter constructs were then designed that incorporated wild type and variant SNP *TSPYL* promoter region SNPs, as shown schematically in Figures 5b and 5c to study possible SNP-dependent modulation of regulatory activity. Specifically, after transfecting HepaRG cells with a series of luciferase reporter constructs, we found that the *TSPYL1* variant promoter displayed reduced...
luciferase activity (Figure 5b) as compared to the TSPYL1 wild type promoter while the TSPYL4 variant promoter resulted in increased luciferase activity (see Figure 5c) as compared to the TSPYL4 wild type promoter, thus revealing SNP-dependent regulatory differences in their effect on TSPYL gene expression. Very similar results were observed when we transfected luciferase reporter constructs into Caco2 and HepG2 cells, using the same constructs that had been employed in a previous TSPYL study (Qin et al., 2019) (See Supp. Figures 4a to 4d).

Discussion:

Our previous studies had shown that TSPYLs are transcriptional regulators of several CYPs and that down regulation of the expression of TSPYLs can result in the upregulation of CYP3A4, CYP2C9 and CYP2C19 gene expression (Qin et al., 2017; Qin et al., 2019). Furthermore, genetic polymorphisms present in both the TSPYLs and CYPs are known to be associated with variation in drug response phenotypes, but those polymorphisms only explain a portion of the population variability that has been observed in CYP expression. The present study of upstream regulators of TSPYL expression has revealed additional factors that could contribute to individual variation in CYP-dependent drug metabolism pathways. The results of the series of experiments described here may help us better understand molecular factors that contribute to that variation.

We used the HepaRG cell line for most of our studies since those cells have been reported to be more similar biologically to human hepatocytes than are many other widely used hepatic cell model systems such as HepG2, Huh7 or the Hep3B (Zeilinger et al., 2016), especially in terms of basal CYP expression and/or induction (Ramboer et al., 2015). Our mRNA expression results for TSPYLs and CYP3A4 after the silencing of REST and ZBTB7A reflect known eQTL relationships that have been reported between TSPYLs and CYPs in HepaRG cells (Qin et al.,
2017; Qin et al., 2019) and they made it possible for us to identify REST and ZBTB7A as upstream regulators of TSPYL4 expression. Our studies of the expression of TSPYLs and CYPs after silencing these TFs as well as rescue experiments performed with TSPYL overexpression vectors further supported the apparent regulation of TSPYL gene expression by REST and ZBTB7A. Future proteomic analysis (Graves and Haystead, 2002) of HepaRG cells might provide insight into additional transcriptional cofactors involved in the sequential, step-wise regulation of the expression of TSPYLs and, downstream, of CYPs.

In one of our previous reports, 4 missense variants in TSPYL genes were studied, with the identification of one missense SNP, rs3828743 (G/A) (Pro62Ser) in the TSPYL1 open reading frame that abolished the suppression of CYP3A4 expression by TSPYL1 due to loss of the ability of TSPYL variant protein to bind to the CYP promoter region (Qin et al., 2017). In the present study, we observed that the rs3828743 variant genotype results in SNP-dependent modulation of transcriptional regulatory effects on the expression of TSPYL1. That SNP, rs3828743, resides in the binding region for REST and ZBTB7A, as shown by published ChIPseq data and by our ChIP-qPCR data. As a result, variation of TSPYL1 expression based on rs3828743 genotype may contribute, at least in part, to variation in the impact of REST and ZBTB7A on transcriptional activity at this locus.

Based on our previous studies, we know that the TSPYL4 rs910391 SNP, a SNP that is in tight linkage disequilibrium with the TSPYL1 rs10223646 SNP, is associated with baseline depression severity in major depressive disorder patients (Qin et al., 2019). In the present study, we showed that rs910391 maps to the binding region for REST and ZBTB7A, as demonstrated by published ChIPseq data and by our own ChIP-qPCR data. In addition, we observed that the rs910391 genotype variants contribute to SNP-dependent modulation of TSPYL4 expression, which is at
least partially responsible for variation in the transcription activity of *TSPYL4*. Further study of cis eQTL SNPs for additional upstream transcription factors like those identified in the present study might provide mechanistic insight into regulation of the expression of drug metabolizing enzymes and of genetic polymorphisms associated with variability in drug response.

In summary, the novel transcription factors *REST* and *ZBTB7A* appear to be transcriptional regulators of TSPYL gene expression resulting in variation in expression which then plays a role downstream in CYP expression and CYP-mediated variation in drug metabolism. This series of events represents a novel upstream source of variation in downstream CYP expression that may mechanistically help us to better understand variation in CYP expression. Ultimately, this novel SNP-dependent modulation of transcription regulating TSPYL expression and activity may contribute to variability in both CYP expression and, as a result, variation in drug response phenotypes.

**Acknowledgments**

This work was supported, in part, by NIH grants R01GM125633-04 (LW), GM28157 (RW) and by T32GM008685-24 and U01GM61388 (LW and RW).

**Authorship Contributions:**

Participated in research design: Shivaram, Gao, Qin, Liu, Weinshilboum, and Wang.

Conducted experiments: Shivaram, and Qin.

Performed data analysis: Shivaram, Gao, Qin, Liu.

Wrote or contributed to the writing of the manuscript: Shivaram, Gao, Liu, Weinshilboum, and Wang.
References:


mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nature Genetics* **44:**1310-1315.


the Metabolism of Selective Serotonin Reuptake Inhibitors in Patients with Major Depressive Disorder. *Clinical Pharmacology & Therapeutics* **107**:662-670.


Footnotes:

Conflict of Interest

Drs. Wang and Weinshilboum are cofounders of and stock-holders in OneOme LLC. All other authors have no conflict of interest with the contents of this article.

Figures captions

Figure 1: Transcription factors that might transcriptionally regulate TSPYL1/2/4 expression. The Table lists the 30 TFs identified as binding to the promoter regions of TSPYL 1/2/4 as well as those that also displayed significant correlations with TSPYL expression in human liver tissue (Storey et al., 2011). The three TFs that were studied in detail here, ZBTB7A, REST and MAFK, are highlighted in red type in the TF gene list.

Figure 2: RT-qPCR quantification of expression for TSPYL genes and CYP3A4 in HepaRG cells after individual knockdown of 30 candidate TFs. The relative mRNA expression values for TSPYL genes and CYP3A4 after KD of selected TFs, as listed in Figure 1, were plotted on a log2 scale after being normalized to expression of the housekeeping gene GAPDH. Highlighted in red are data for ZBTB7A, REST and MAFK. Each point represents the expression of TSPYLs vs CYP3A4 after KD of one TF, measured in triplicate, and error bars indicate the SD for each point.

Figure 3. Transcriptional regulation of TSPYL4 and CYP3A4 genes by ZBTB7A, REST and MAFK. In HepaRG cells, TSPYL4 and CYP3A4 mRNA levels were determined after co-transfection with non-targeting siRNA (CONTROL) or siRNA targeting a-c) ZBTB7A, d-f) REST, or g-i) MAFK, and empty vector or plasmids overexpressing TSPYL4. The mRNA levels of CYPs in KD-only or OE-only were compared to those for cells transfected with negative siRNA and empty vector, and expression levels of KD+OE-TSPYL4 were compared with KD-only samples in three independent experiments by two-tailed student’s t-test, *p < 0.05, **p < 0.01, ***p < 0.001. Error bars represent the SD of three replicates.
**Figure 4: ChIP assays for ZBTB7A and REST in HepaRG cells.** a) Primer sets targeting TSPYL1 and TSPYL4 promoter regions are indicated as follows: for TSPYL1 P1: -280bp to -30bp; P2: -50bp to +114bp; P3: +92bp to +264bp) and for TSPYL4 (P4: -86bp to +93bp; P5: +262bp to +433bp). Nucleotides have been designated as negative or positive numbers if they are downstream or upstream from the TSS for TSPYL1 or TSPYL4. Bindings of the transcription factors b) ZBTB7A or c) REST to promoter regions of TSPYL1 and TSPYL4 were determined by qRT-PCR and are shown as fold enrichment over Input. Error bars represent the SD of two replicates.

**Figure 5: SNP-dependent modulation of transcriptional activity for TSPYL1 and TSPYL4 promoter regions.** a) Genome line diagrams of the TSPYL1 and TSPYL4 promoter regions showing the locations of common SNPs in Caucasians with MAF values >1% (arrows). Binding regions for ZBTB7A and REST based for published ENCODE ChIPseq data are also indicated. b) Luciferase reporter constructs were created for TSPYL promoter regions either with wild type or variant genotypes for the indicated SNPs to study SNP-dependent transcriptional activity of ZBTB7A and REST binding loci. b) and c) Luciferase Assays using HepaRG cells co-transfected for b) TSPYL4 or c) TSPYL1 promoter firefly luciferase reporters and Renilla luciferase constructs. Transfection efficiencies have been normalized based on Renilla luciferase reporter signals. Differences in normalized luciferase activity between WT and variant TSPYL promoter constructs were then compared in three independent experiments by two-tailed student’s t-test, *p <0.05, ***p <0.001. Error bars represent the SDs for three replicates.
**Figure 1**

<table>
<thead>
<tr>
<th>Candidate TFs</th>
<th>TSPYL1</th>
<th>TSPYL2</th>
<th>TSPYL4</th>
<th>Candidate TFs</th>
<th>TSPYL1</th>
<th>TSPYL2</th>
<th>TSPYL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>FOSL2</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
</tr>
<tr>
<td>TAF1</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>RCOR1</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIN3AK20</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>REST</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MXI1</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>RAD21</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBP</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>SMC3</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEBPB</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>HSF1</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EP300</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>HDAC2</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHD2</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>ARID3A</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFX5</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>TEAD4</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YY1</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>NFIC</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZBTB7A</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td>HNF4A</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRCA1</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td>HNFAG</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZBTB33</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td>Δ=Binds to promoter regions of TSPYL 1/2/4 (Chip-seq data in HepG2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAFF</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td>Δ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAFK</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td>Δ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOXA1</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td>Δ=Binds to promoter regions of TSPYL 1/2/4 (Chip-seq data in HepG2) + shows significant correlation with TSPYL expression in human liver tissue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOXA2</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td>Δ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYBL2</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td>Δ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3

a) **ZBTB7A**

b) **TSPYL4**
c) **CYP3A4**

d) **REST**
e) **TSPYL4**
f) **CYP3A4**

g) **MAFK**
h) **TSPYL4**
i) **CYP3A4**
Figure 4

a) 

```
^  ^
P1  P3
^  ^
P2  P4

TSS

|   |
---|---|
P1  | P2  | P3  | P4  | P5  |
   |   |   |   |   |
   |   |   |   |   |
   |   |   |   |   |
   |   |   |   |   |
20kbp

TSPYL1

```

b) 

```
<table>
<thead>
<tr>
<th>Binding of ZBTB7A</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG</td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
<tr>
<td>P3</td>
</tr>
<tr>
<td>P4</td>
</tr>
<tr>
<td>P5</td>
</tr>
</tbody>
</table>
```

c) 

```
<table>
<thead>
<tr>
<th>Binding of REST</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG</td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
<tr>
<td>P3</td>
</tr>
<tr>
<td>P4</td>
</tr>
<tr>
<td>P5</td>
</tr>
</tbody>
</table>
```
Figure 5

a) TSPYL1 and TSPYL4 gene structures with polymorphisms.

b) Relative luminescence of HepaRG cells transfected with EV, TSPYL1-WT, and TSPYL1-Variant.

c) Relative luminescence of HepaRG cells transfected with EV, TSPYL4-WT, and TSPYL4-Variant.