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 staring 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 (*TSPYL*s) 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 (*TSPYL*s) 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 *TSPYL*s.

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

9

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.

10

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 $\geq$ 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

12

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

14

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 *TSPYL*s are transcriptional regulators of several CYPs and that down regulation of the expression of *TSPYL*s 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 [27,28,29] 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.,

15

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 rs3828743genotype 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

16

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

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:

(CDC) CfDCaP (2015-2018a) <https://www.cdc.gov/medicationsafety/basics.html>.

(CDC) CfDCaP (2015-2018b) <https://www.cdc.gov/nchs/products/databriefs.htm>.

Bush WS, Crosslin DR, Owusu-Obeng A, Wallace J, Almoguera B, Basford MA, Bielinski SJ,
Carrell DS, Connolly JJ, Crawford D, Doheny KF, Gallego CJ, Gordon AS, Keating B,
Kirby J, Kitchner T, Manzi S, Mejia AR, Pan V, Perry CL, Peterson JF, Prows CA,
Ralston J, Scott SA, Scrol A, Smith M, Stallings SC, Veldhuizen T, Wolf W, Volpi S,
Wiley K, Li R, Manolio T, Bottinger E, Brilliant MH, Carey D, Chisholm RL, Chute CG,
Haines JL, Hakonarson H, Harley JB, Holm IA, Kullo IJ, Jarvik GP, Larson EB,
McCarty CA, Williams MS, Denny JC, Rasmussen-Torvik LJ, Roden DM, and Ritchie
MD (2016) Genetic variation among 82 pharmacogenes: The PGRNseq data from the
eMERGE network. *Clinical Pharmacology & Therapeutics* **100:**160-169.

Daly AK (2010) Genome-wide association studies in pharmacogenomics. *Nature Reviews
Genetics* **11:**241-246.

de Andrade TG, Peterson KR, Cunha AF, Moreira LS, Fattori A, Saad STO, and Costa FF
(2006) Identification of novel candidate genes for globin regulation in erythroid cells
containing large deletions of the human β-globin gene cluster. *Blood Cells, Molecules,
and Diseases* **37:**82-90.

Epping MT, Lunardi A, Nachmani D, Castillo-Martin M, Thin TH, Cordon-Cardo C, and
Pandolfi PP (2015) TSPYL2 is an essential component of the REST/NRSF
transcriptional complex for TGFβ signaling activation. *Cell Death & Differentiation*
22:1353-1362.

Evans WE and Relling MV (1999) Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286:**487-491.

Graves PR and Haystead TAJ (2002) Molecular Biologist's Guide to Proteomics. *Microbiology and Molecular Biology Reviews* **66:**39-63.

Guo YI, Gong YI, Shi G, Yang KUN, Pan C, Li MIN, Li Q, Cheng Q, Dai R, Fan L, and Wan H (2012) Single-nucleotide polymorphisms in the TSPYL-4 and NT5DC1 genes are associated with susceptibility to chronic obstructive pulmonary disease. *Molecular Medicine Reports* **6:**631-638.

He Y, Hoskins JM, and McLeod HL (2011) Copy number variants in pharmacogenetic genes. *Trends in Molecular Medicine* **17:**244-251.

Jover R, Moya M, and Gomez-Lechon M (2009) Transcriptional Regulation of Cytochrome P450 Genes by the Nuclear Receptor Hepatocyte Nuclear Factor 4-Alpha. *Current Drug Metabolism* **10:**508-519.

Kim T-Y, Zhong S, Fields CR, Kim JH, and Robertson KD (2006) Epigenomic Profiling Reveals Novel and Frequent Targets of Aberrant DNA Methylation-Mediated Silencing in Malignant Glioma. *Cancer Research* **66:**7490-7501.

Kozyra M, Ingelman-Sundberg M, and Lauschke VM (2017) Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genetics in Medicine* **19:**20-29.

Le Gallo M, O'Hara AJ, Rudd ML, Urick ME, Hansen NF, O'Neil NJ, Price JC, Zhang S, England BM, Godwin AK, Sgroi DC, Hieter P, Mullikin JC, Merino MJ, and Bell DW (2012) Exome sequencing of serous endometrial tumors identifies recurrent somatic

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

Liu D, Gao Y, Wang H, Zi J, Huang H, Ji J, Zhou R, Nan Y, Wang S, Zheng X, Zhu J, Cui Y,
and Chen C (2010) Evaluation of the effects of cytochrome P450 nonsynonymous single-
nucleotide polymorphisms on tanshinol borneol ester metabolism and inhibition potential.
Drug Metab Dispos **38:**2259-2265.

Martinez-Jimenez C, Jover R, Teresa Donato M, Castell J, and Jose Gomez-Lechon M (2007)
Transcriptional Regulation and Expression of CYP3A4 in Hepatocytes. *Current Drug
Metabolism* **8:**185-194.

Motsinger-Reif AA, Jorgenson E, Relling MV, Kroetz DL, Weinshilboum R, Cox NJ, and
Roden DM (2013) Genome-wide association studies in pharmacogenomics.
Pharmacogenetics and Genomics **23:**383-394.

Neavin DR, Lee J-H, Liu D, Ye Z, Li H, Wang L, Ordog T, and Weinshilboum RM (2019)
Single Nucleotide Polymorphisms at a Distance from Aryl Hydrocarbon Receptor (AHR)
Binding Sites Influence AHR Ligand–Dependent Gene Expression. *Drug Metabolism
and Disposition* **47:**983-994.

Nebert DW, Wikvall K, and Miller WL (2013) Human cytochromes P450 in health and disease.
Philosophical Transactions of the Royal Society B: Biological Sciences **368**.

Nelson DR (2013) A world of cytochrome P450s. *Philosophical Transactions of the Royal
Society B: Biological Sciences* **368**.

Qin S, Eugene AR, Liu D, Zhang L, Neavin D, Biernacka JM, Yu J, Weinshilboum RM, and
Wang L (2019) Dual Roles for the TSPYL Family in Mediating Serotonin Transport and

the Metabolism of Selective Serotonin Reuptake Inhibitors in Patients with Major

Depressive Disorder. *Clinical Pharmacology & Therapeutics* **107:**662-670.

Qin S, Liu D, Kohli M, Wang L, Vedell PT, Hillman DW, Niu N, Yu J, Weinshilboum RM, and
Wang L (2017) TSPYL Family Regulates CYP17A1 and CYP3A4 Expression: Potential
Mechanism Contributing to Abiraterone Response in Metastatic Castration-Resistant
Prostate Cancer. *Clinical Pharmacology & Therapeutics* **104:**201-210.

Ramboer E, Vanhaecke T, Rogiers V, and Vinken M (2015) Immortalized Human Hepatic Cell
Lines for In Vitro Testing and Research Purposes, in: *Protocols in In Vitro Hepatocyte
Research*, pp 53-76.

Storey JD, Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, Ramirez
J, Liu W, Lin YS, Moloney C, Aldred SF, Trinklein ND, Schuetz E, Nickerson DA,
Thummel KE, Rieder MJ, Rettie AE, Ratain MJ, Cox NJ, and Brown CD (2011)
Identification, Replication, and Functional Fine-Mapping of Expression Quantitative
Trait Loci in Primary Human Liver Tissue. *PLoS Genetics* **7**.

Wang H, An N, Wang H, Gao Y, Liu D, Bian T, Zhu J, and Chen C (2011) Evaluation of the
effects of 20 nonsynonymous single nucleotide polymorphisms of CYP2C19 on S-
mephenytoin 4'-hydroxylation and omeprazole 5'-hydroxylation. *Drug Metab Dispos*
39:830-837.

Wang L Scherer SE Bielinski SJ Muzny DM Jones LA Black JL, 3rd Moyer AM Giri J Sharp
RR Matey ET Wright JA Oyen LJ Nicholson WT Wiepert M Sullard T Curry TB Rohrer
Vitek CR McAllister TM St Sauver JL Caraballo PJ Lazaridis KN Venner E Qin X Hu J
Kovar CL Korchina V Walker K Doddapaneni H Wu TJ Raj R Denson S Liu W
Chandanavelli G Zhang L Wang Q Kalra D Karow MB Harris KJ Sicotte H Peterson SE

Barthel AE Moore BE Skierka JM Kluge ML Kotzer KE Kloke K Vander Pol JM Marker H Sutton JA Kekic A Ebenhoh A Bierle DM Schuh MJ Grilli C Erickson S Umbreit A Ward L Crosby S Nelson EA Levey S Elliott M Peters SG Pereira N Frye M Shamoun F Goetz MP Kullo IJ Wermers R Anderson JA Formea CM El Melik RM Zeuli JD Herges JR Krieger CA Hoel RW Taraba JL St Thomas SR Absah I Bernard ME Fink SR Gossard A Grubbs PL Jacobson TM Takahashi P Zehe SC Buckles S Bumgardner M Gallagher C Fee-Schroeder K Nicholas NR Powers ML Ragab AK Richardson DM Stai A Wilson J Pacyna JE Olson JE Sutton EJ Beck AT Horrow C Kalari KR Larson NB Liu H Wang L Lopes GS Borah BJ Freimuth RR Zhu Y Jacobson DJ Hathcock MA Armasu SM McGree ME Jiang R Koep TH Ross JL Hilden MG Bosse K Ramey B Searcy I Boerwinkle E Gibbs RA and Weinshilboum RM (2022) Implementation of preemptive DNA sequence-based pharmacogenomics testing across a large academic medical center: The Mayo-Baylor RIGHT 10K Study. *Genet Med*.

Wright GEB, Carleton B, Hayden MR, and Ross CJD (2016) The global spectrum of protein-coding pharmacogenomic diversity. *The Pharmacogenomics Journal* **18:**187-195.

Zanger UM and Schwab M (2013) Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics* **138:**103-141.

Zeilinger K, Freyer N, Damm G, Seehofer D, and Knöspel F (2016) Cell sources for in vitro human liver cell culture models. *Experimental Biology and Medicine* **241:**1684-1698.

Zi J, Liu D, Ma P, Huang H, Zhu J, Wei D, Yang J, and Chen C (2010) Effects of CYP2C9*3 and CYP2C9*13 on Diclofenac Metabolism and Inhibition-based Drug-Drug Interactions. *Drug Metab Pharmacokinet* **25:**343-350.

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 $\log_2$ 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.

24

Figure 1

Candidate TFs	TSPYL1	TSPYL2	TSPYL4	Candidate TFs	TSPYL1	TSPYL2	TSPYL4
SP1	Δ	Δ	Δ	FOSL2	Δ		Δ
TAF1	Δ	Δ	Δ	RCOR1	Δ		
SIN3AK20	Δ	Δ	Δ	REST	Δ		
MXI1	Δ	Δ	Δ	RAD21		Δ	
TBP	Δ	Δ	Δ	SMC3		Δ	
CEBPB	Δ	Δ	Δ	HSF1			Δ
EP300	Δ	Δ	Δ	HDAC2			Δ
CHD2	Δ	Δ	Δ	ARID3A			Δ
RFX5	Δ	Δ	Δ	TEAD4			Δ
YY1	Δ		Δ	NFIC			Δ
ZBTB7A	Δ		Δ	HNF4A			Δ
BRCA1	Δ		Δ	HNFAG			Δ
ZBTB33	Δ		Δ	Δ=Binds to promoter regions of TSPYL 1/2/4 (Chip-seq data in HepG2)			
MAFF	Δ		Δ				
MAFK	Δ		Δ				
FOXA1	Δ		Δ	**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**			
FOXA2	Δ		Δ				
MYBL2	Δ		Δ				

Figure 2

a)

b)

c)
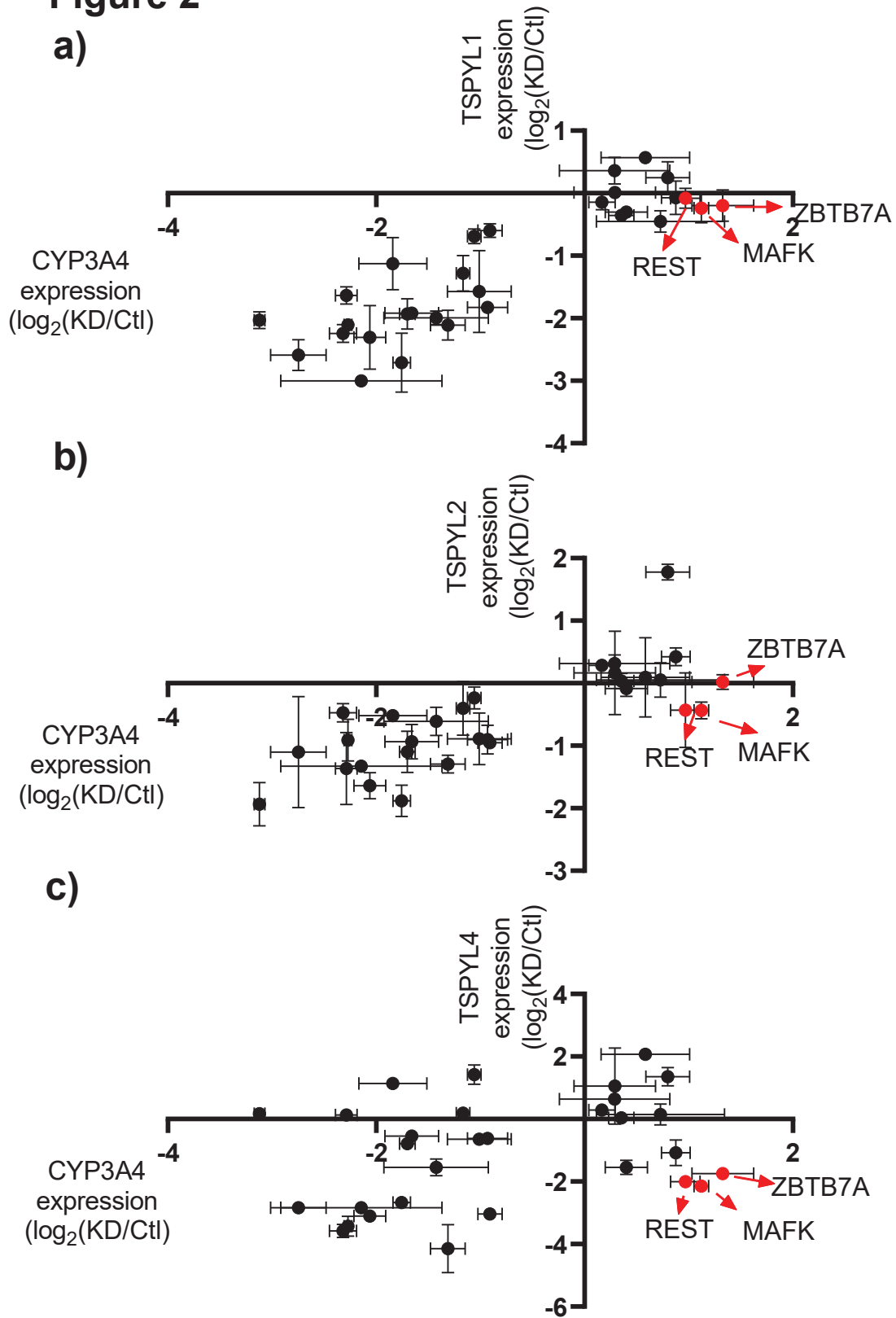
Figure 3
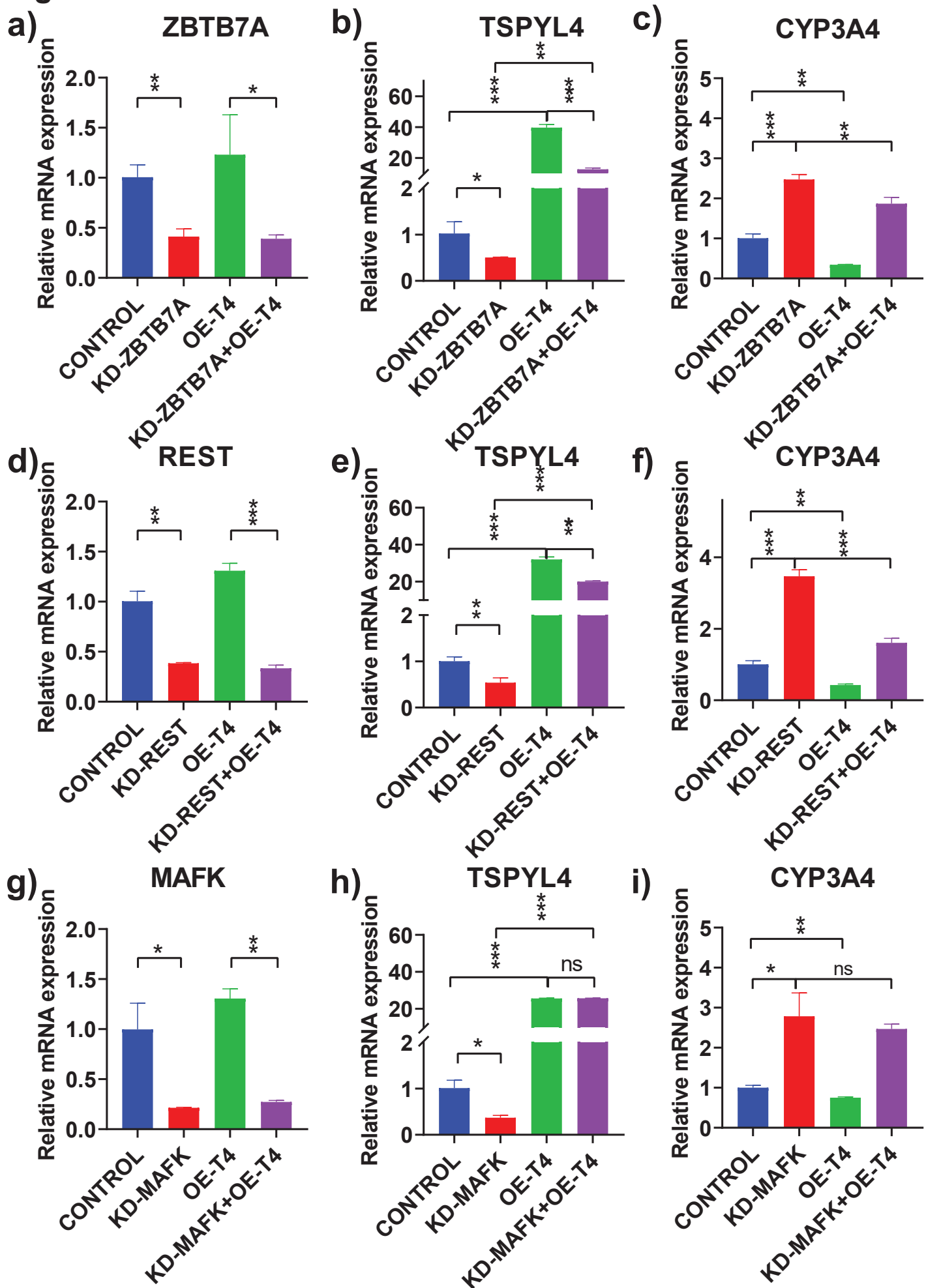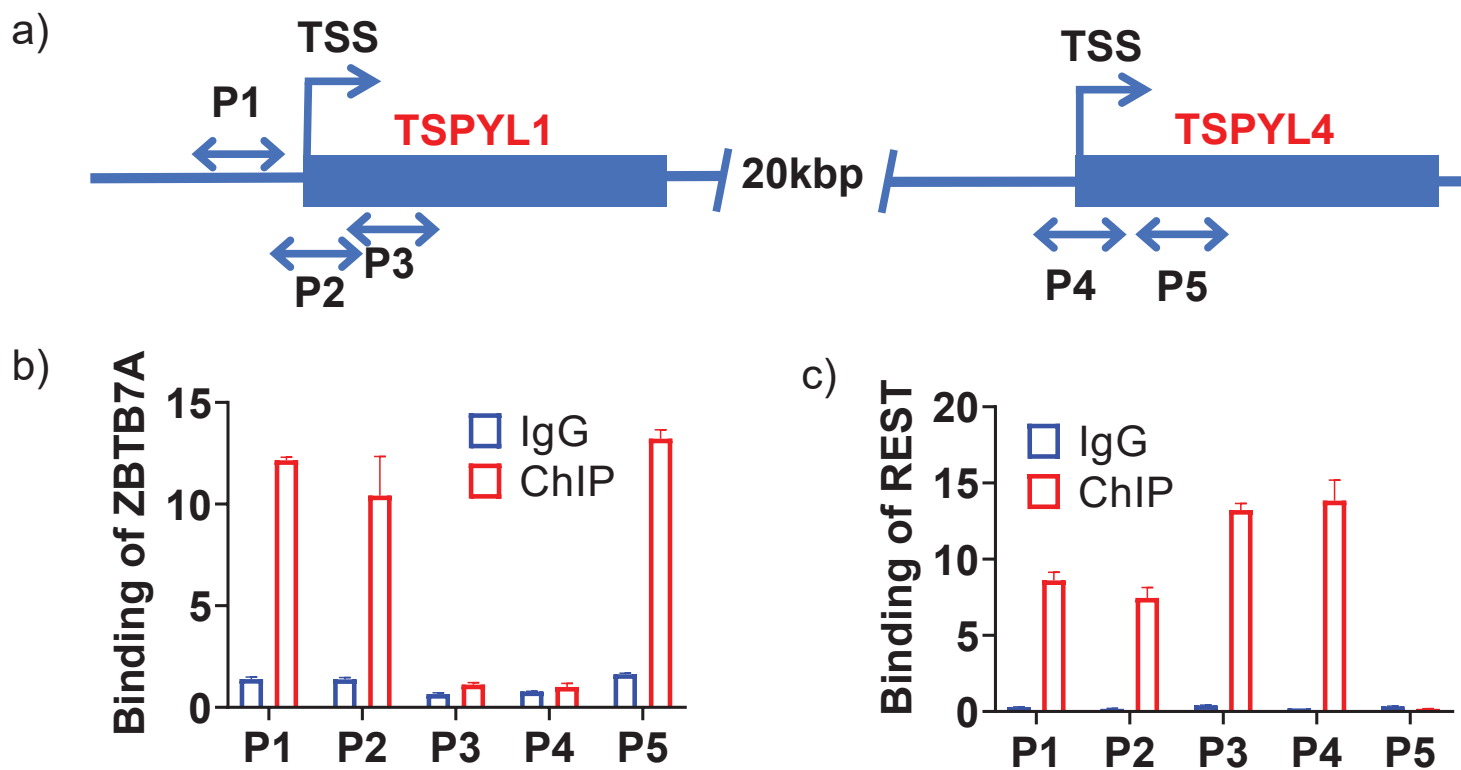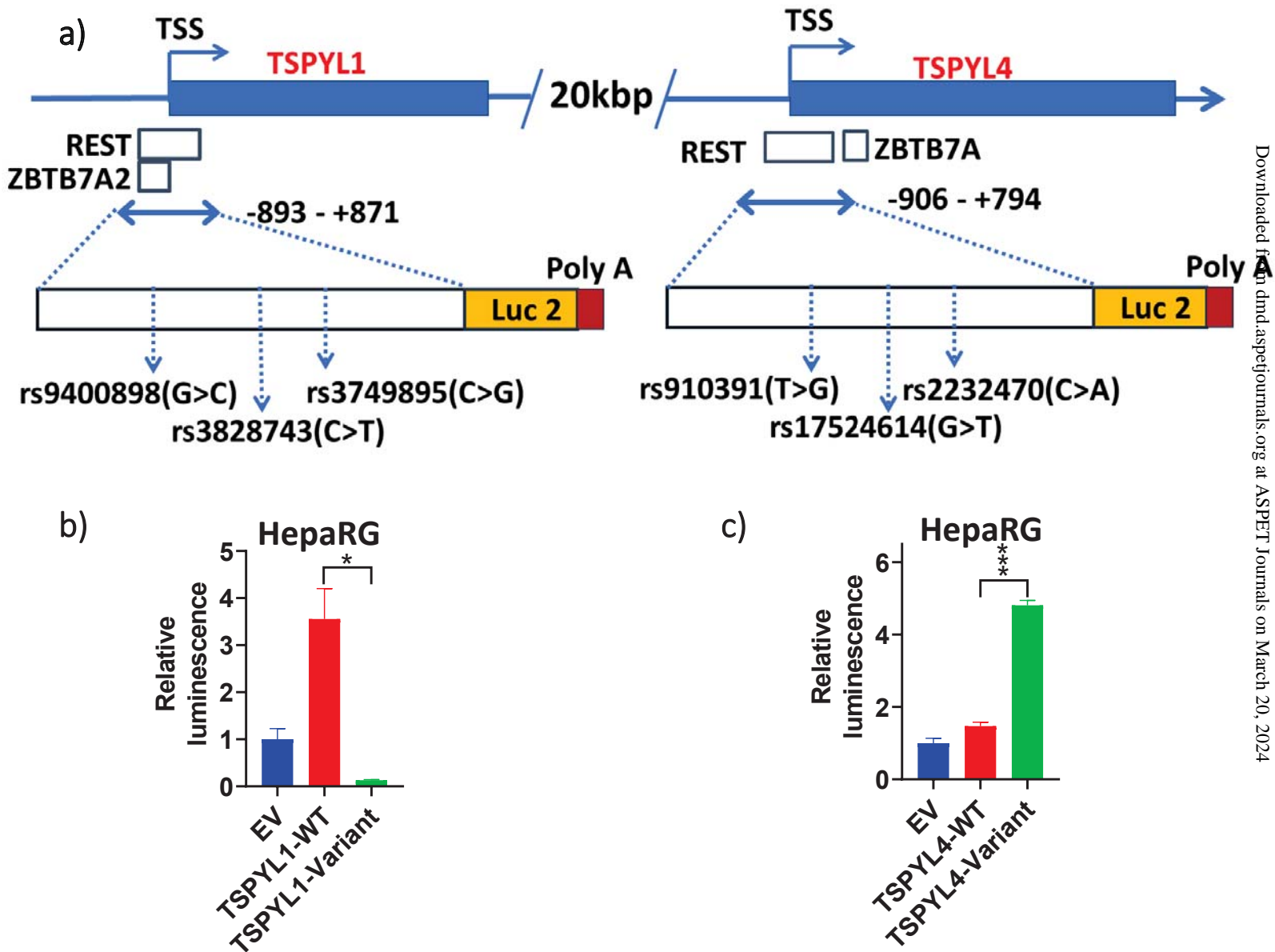
Figure 4

Figure 5

Drug Metabolism and Disposition

Supplementary Materials

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.

Article number: DMD-AR-2022-000945

Supplementary Table 1: SiRNAs from Dharmacon/Horizon Discovery

Gene	Dharmacon Catalog number
SP1	M-026959-00
TAF1	M-005041-01
SIN3A	M-012990-00
MXI1	M-009947-02
TBP	M-011790-01
CEBPB	M-006423-03
EP300	M-003486-04
CHD2	M-008948-01
RFX5	M-011103-01
YY1	M-011796-02
ZBTB7A	M-020818-00
BRCA1	M-003461-02
ZBTB33	M-019982-01
MAFF	M-003903-00
MAFK	M-008580-01
FOXA1	M-010319-01
FOXA2	M-010089-01
MYBL2	M-010444-01
FOSL2	M-004110-00
RCOR1	M-014076-01
REST	M-006466-02
RAD21	M-006832-01

SMC3	M-006834-01
HSF1	M-012109-01
HDAC2	M-003495-02
ARID3A	M-012032-01
TEAD4	M-019570-03
NFIC	M-008362-00
HNF4A	M-003406-02
HNF4G	M-003407-02

Supplementary Table 2: Prime time primers for qRT-PCR

Gene	Forward	Reverse
TSPYL1	CTAGAAATCCAAGTGACAGCAAC	AGCCAAGCAAATGATACTGTAGA
TSPYL2	ATGTCTTCATCACGTCGGTT	GTCTCAAGCGCAAGTTCATC
TSPYL4	CTTCCTCAATGCCCCAACTC	ACTCCTAGCCAACAAACCATC
CYP3A4	ATCATGTCAGGATCTGTGATAGC	GGGAAATATTTTGTCCTACCATAAGG
CYP2C19	GGTAATCACTGCAGCTGACT	TCAATCTCTTCCTGGACTTTAGC
CYP2C9	CTGCAGTTGACTTGTTTGGAG	GTTTCTGCCAATCACACGTTC
ARID3A	GACTTGTTCAGCTTCATGCAG	CTTCTCCGTCACCAGCAC
BRCA1	AATGGAAGGAGAGTGCTTGG	ATACCTGCCTCAGAATTTCCTC
CEBPB	AGAAACGTCTATGTGTACAGATGA	GATTGCATCAACTTCGAAACCG
CHD2	ATCCGAAGGTTCATCAAGGC	TGCCACCGACTTATCTACCA
EP300	GCGGCCTAAACTCTCATCTC	GTAAGTCGTGCTCCAAGTCA
FOSL2	GATCAAGACCATTGGCACCA	CAGCCAGCTTGTTCCTCT
FOXA1	CCAGGATGTTAGGAACTGTGAA	CTGAGTTCATGTTGCTGACC
FOXA2	GGAGCGGTGAAGATGGAAG	TGTTCATGCCGTTCATCCC
HNF4A	GCCATCATCTTCTTTGACCCA	GATGTAGTCCTCCAAGCTCAC
HNF4G	GAGCAACAGGAAAACACTATGG	CAACACATTGCCGACTGAAC
HSF1	CCTCCACCCCTGAAAAGTG	GGAGTCCATAGCATCCAAGTG
MAFF	GCTAGGAGTGAGGGATGTGA	CTCAGCTCTCGCTTGATCT
MAFK	GACGCCAGCTACGAGTTC	GACACCAGCTCATCATCGC
MXI1	AGCACCCAAGTCTAAGTCAAC	CCGCTGCTGTGTTTCTGT
MYBL2	GATTCCTGTAACAGCCTCACG	CTCTCCAGCTCCAATGTGTC
NFIC	GGCGGCGATTACTACACTTC	CTTGTCCATCTCTGTCTTCTTCA
RAD21	GGAAAGAGACAGGAGGAGTAGA	GTGTAAGACAGCGTGTAAAGAG
RCOR1	CCCAGATAATTCCATTAAGATGCC	TAACACAGTAGTCCACACCAAG
REST	ACTAGACATATGCGTACTCATTCAG	CCATTGTGAACCTGTCTTGC
RFX5	ACCTACCACCTTCTTCAGA	GTAGAGATACAGCTTGTCATTGTC
SIN3A	AGATGTTCACCATTCATGCCT	CACACAGATCTCATCACTCACG
SMC3	TGAGTTTAGTCATCTTCGTCCAG	TCCACAAAAGCAGAAATAACACG
SP1	GGTACTTCAGGAATCCAGGTG	GCTGTGTCATCATGTATTCCATC
TAF1	ATGGTTTGGAGGATAGCAACA	CTCCTCATCTTCTTCCTCCTCT
TBP	GATAAGAGAGCCACGAACCAC	CAAGAACTTAGCTGGAAAACCC
TEAD4	GACTCCTTGGAACTGGCTTAG	GATGTGGCTGGAGACCTG
YY1	CAGAATTTGCTAGAATGAAGCCA	CCGAGTTATCCCTGAACATCT
ZBTB33	TGGAGCGACGTTTAAAGAAGG	GCCAAGTGAAGTCAACAGACA
ZBTB7A	GAAGCCCTACGAGTGCAAC	GGTTCTTCAGGTCGTAGTTGTG
GAPDH	ACATCGCTCAGACACCATG	TGTAGTTGAGGTCAATGAAGGG

Supplementary Table 3: ChIP-qPCR primer sets for promoter regions of TSPYL1 and TSPYL4

Gene	Sequences
TSPYL1 - P1	Forward: TTCCAGACTCAGCACAATCG Reverse: TTTCCTCAGAGGCCGAACT
TSPYL1 - P2	Forward: AGTTCGGCCTCTGAGGAAA Reverse: AGCCTCAGGTACTGGTGTGC
TSPYL1 - P3	Forward: GACGCACACCAGTACCTGAG Reverse: CAACTCGGATCTGGGGAGTA
TSPYL4 - P4	Forward: AGTAAAGGAGGGGTGGTGCT Reverse: GGATCTCCTGAGGCATGGT
TSPYL4 - P5	Forward: GGATGCTCCACCTTCTACGA Reverse: TTCTTCCCCGGTATCATCTG

Supplementary Table 4: Primer sets for TSPYL1 and TSPYL4 promoter region amplification

| Name	Restriction Site	Primer sequence
TSPYL1 - forward	Kpn1	aatgGGTACCAGCCTTTCCTCACAAAATGG
TSPYL1 - reverse	HindIII	actgAAGCTTGATGTAGTTCCTCCGCTCCA
TSPYL4 - forward	Kpn1	aatgGGTACCTCCACAGGCATTATGAAGCA
TSPYL4 - reverse	HindIII	actgAAGCTTCTCCTCCACCTCCAAATTGA

Supplementary Table 5: Linkage disequilibrium (LD) and minor allele frequencies (MAFs) for selected

SNPs for the *TSPYL1* promoter regions

| Promoter Region	SNPid	LD (r²)			Ref	Alt	MAF (Eur)
		rs3749895	rs3828743	rs9400898			
TSPYL1	rs3749895		0.81	0.81	C	G	0.3
TSPYL1	rs3828743			1	C	T	0.26
TSPYL1	rs9400898				G	C	0.26

Supplementary Table 6: LD and MAFs for selected SNPs for the *TSPYL4* promoter regions

| Promoter Region	SNPid	LD(r²)			Ref	Alt	MAF (Eur)
		rs2232470	rs910391	rs17524614			
TSPYL4	rs2232470		0.96	<0.2	C	A	0.67
TSPYL4	rs910391			<0.2	T	G	0.67
TSPYL4	rs17524614				G	T	0.16

Supplementary Figures
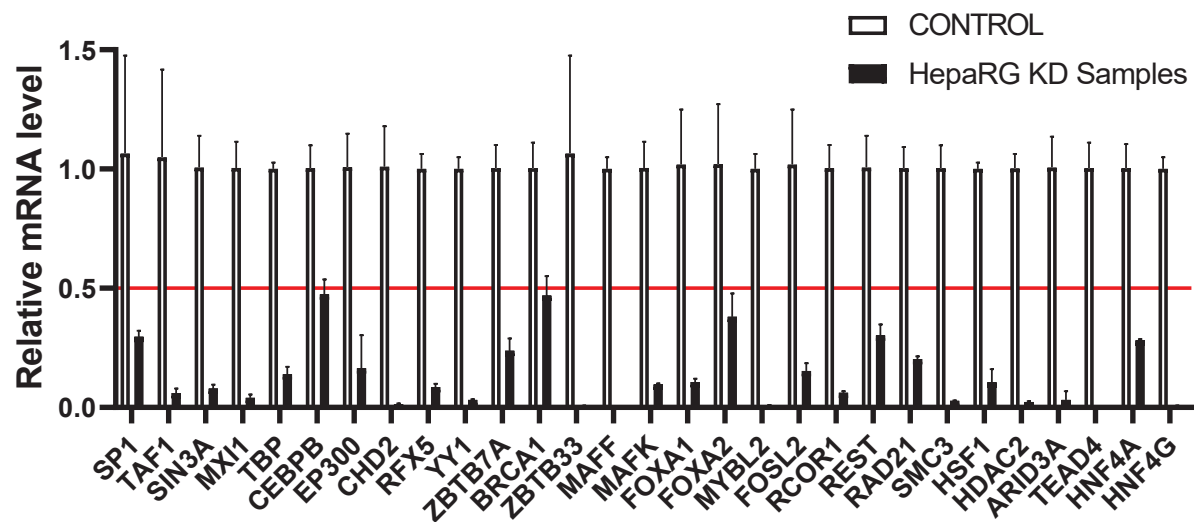
Supplementary Figure 1: Knockdown efficiency achieved in HepaRG cells. The mRNA levels of candidate TFs were compared to those in cells transfected with negative SiRNA in three independent experiments. Error bars represent SDs of three replicates.
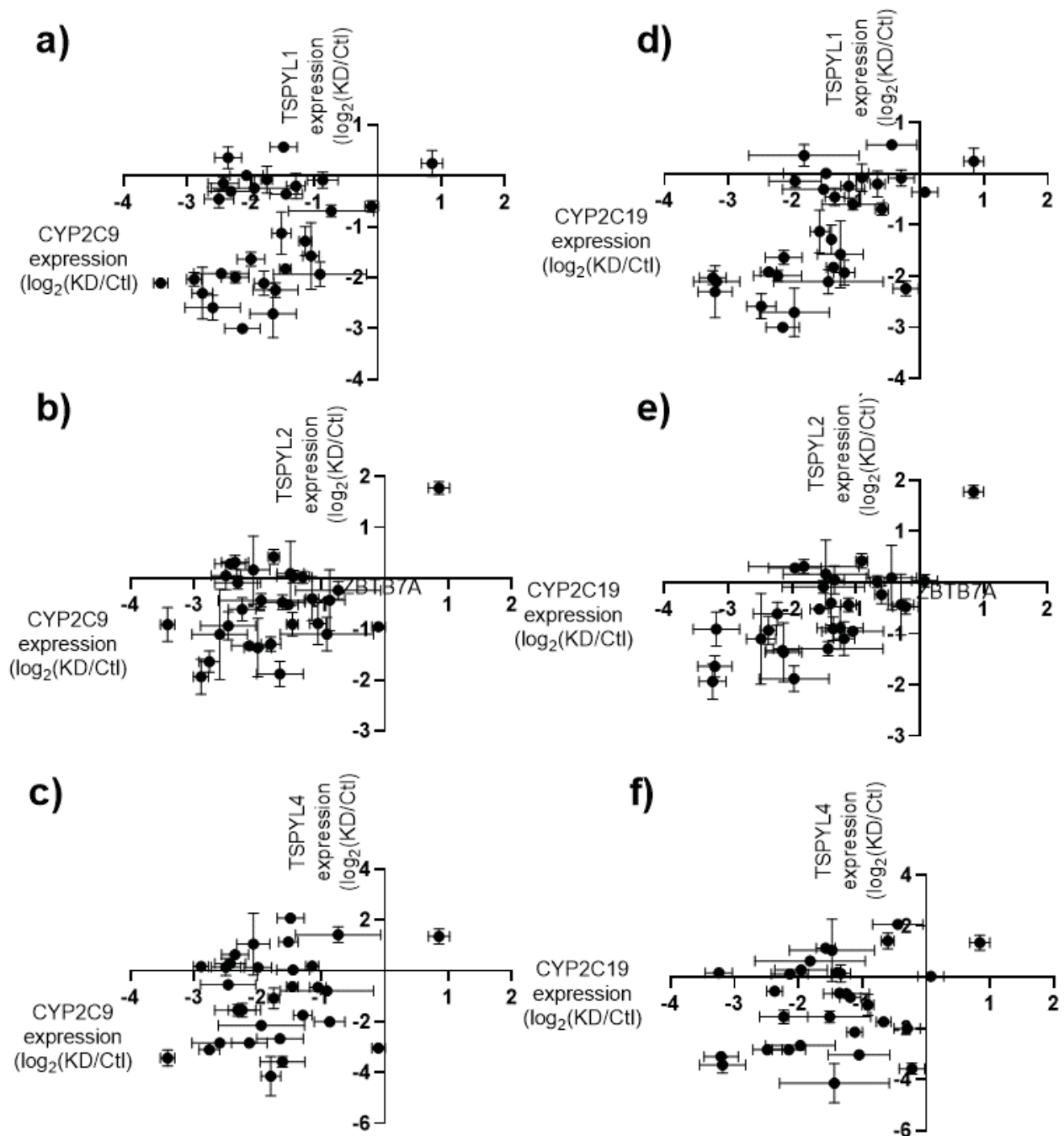
Supplementary Figure 2: RT-qPCR quantification of expression for TSPYL genes and *CYP2C9* and *CYP2C19* in HepaRG cells after selected TF knockdown. The relative mRNA expression values for TSPYL genes and CYP genes after selected TF Knock Down were plotted on log scales after being normalized to the expression of the housekeeping gene *GAPDH*. a) CYP2C9 vs TSPYL1; b) CYP2C9 vs

TSPYL2; c) CYP2C9 vs TSPYL4; d) CYP2C19 vs TSPYL1; e) CYP2C19 vs TSPYL2; f) CYP2C19 vs

TSPYL4. Error bars represent SDs for three independent replicates.
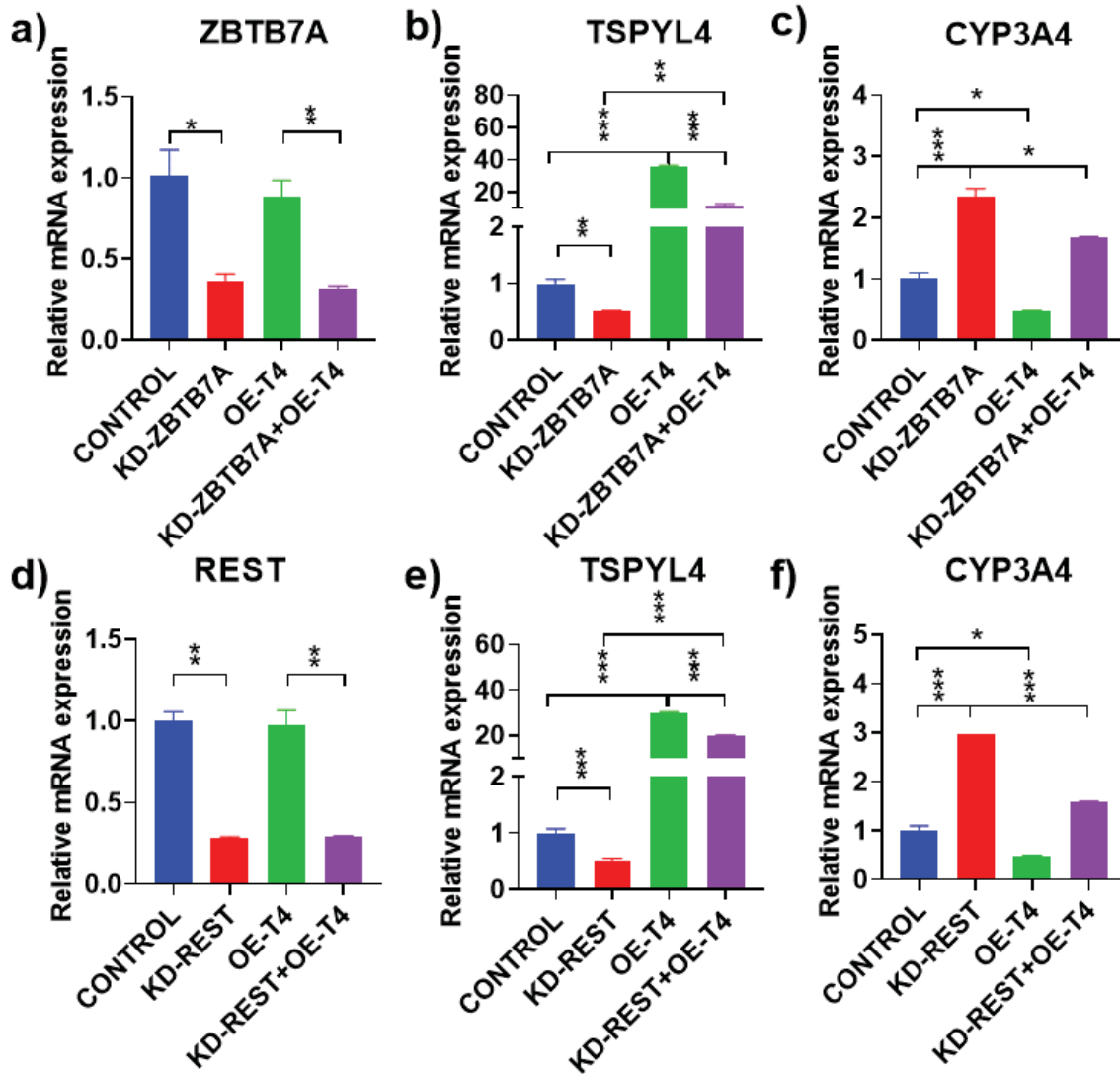
Supplementary Figure 3: Transcriptional regulation of the expression of *TSPYL4* and *CYP3A4* by ZBTB7A and REST. NCI-H2405 cells were co-transfected with non-targeting siRNA (CONTROL) or siRNA targeting **a-c**) ZBTB7A or d-f) REST, and empty vector or a plasmid overexpressing *TSPYL4*. mRNA levels for **a**) *ZBTB7A* **b, e**) *TSPYL4*, **c, f**) *CYP3A4* and **d**) REST were analyzed by RT-qPCR. The mRNA levels after KD-only or OE-only samples were compared to those in cells transfected with

negative siRNA and empty vector, and mRNA levels of KD+OE-TSPYL4 were compared with KD-only samples in three independent experiments with significance tested by two-tailed student's t-test, *p <0.05, **p < 0.01, ***p <0.001. Error bars represent the SDs of three independent replicates.
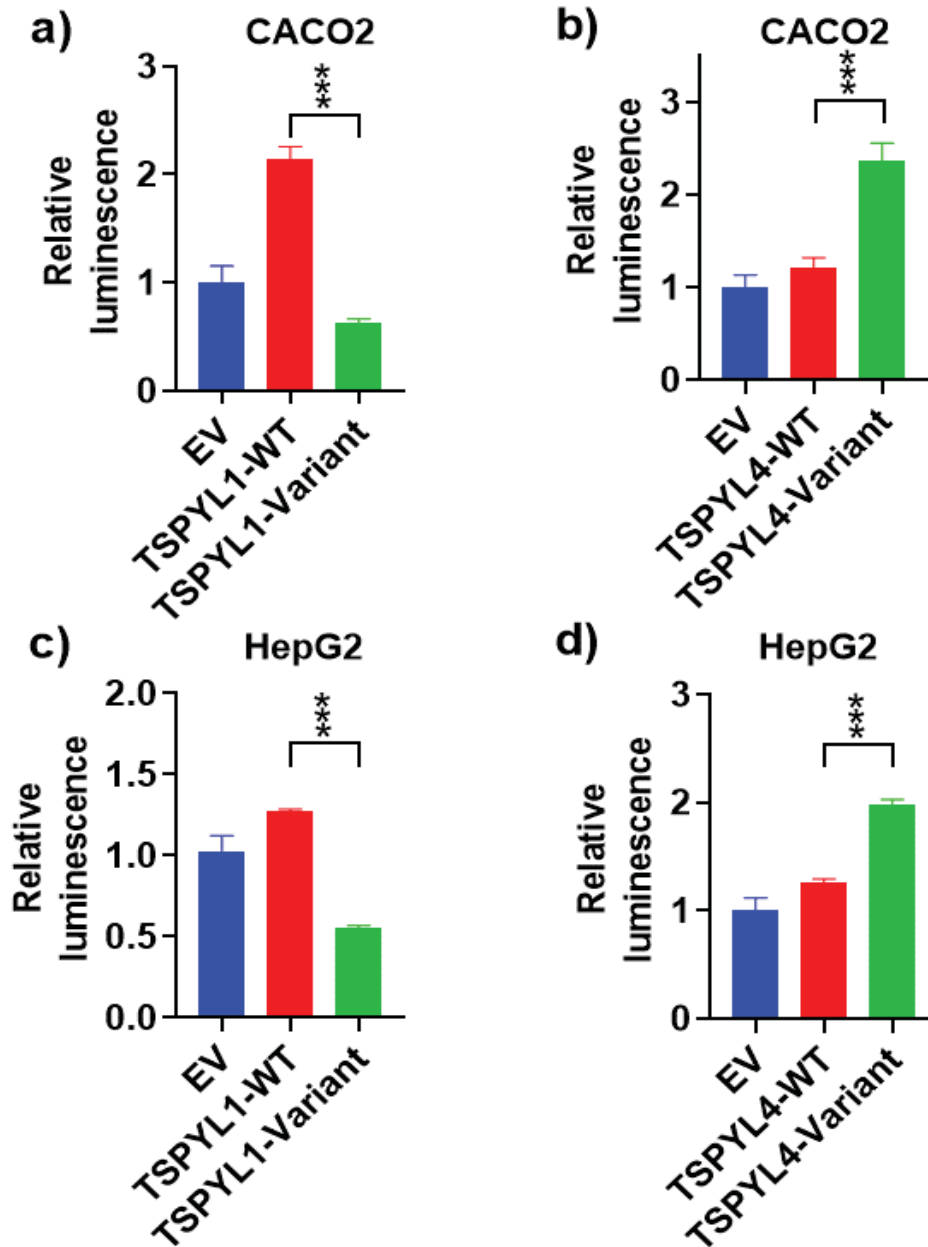
Supplementary Figure 4: SNP-dependent modulation of the transcriptional regulation of TSPYLs at *REST* and *ZBTB7A* binding sites. a) and **b**) Luciferase Assays for **a**) *TSPYL4* and **b**) *TSPYL1* promoter reporter constructs in *CACO2* cells and for luciferase Assays for **c)** *TSPYL4* and **d)** *TSPYL1* promoter reporter constructs in HepG2 cells. Transfection efficiencies were normalized using Renilla luciferase reporter genes. The differences in normalized luciferase activity between WT and variant TSPYL promoter constructs were compared in three independent experiments by two-tailed student's t-test, ***p <0.001. Error bars represent SDs of three independent replicates.

Supplementary Methods and Materials:

ENCODE ChIP Seq data:

The ENCODE UCSC genome browser includes ChIP-seq data for HepG2 cells that shows TF binding to the promoter regions of *TSPYL1,* TSPYL2 and *TSPYL4*.

LIVER eQTL data-based association analysis:

We determined associations between expression of these TFs and *TSPYL1, TSPYL2* or *TSPYL4* in liver eQTL databases using Pearson correlations and identified TFs that were significantly associated with *TSPYL1, TSPYL2* or *TSPYL4* expression with p- values < 0.05.

Cell Culture:

Undifferentiated HepaRG ™ cells (HPR101) were purchased from Biopredic Internationals, Saint-Grégoire, France. Undifferentiated HepaRG cells are human hepatic stem cell lines with the ability to express the full array of cytochrome P450s with function that mimicks that of primary human hepatocytes. The HepaRG cells were cultured in base Williams' Medium E (12551032) with growth medium supplementation (ADD711C) (Biopredic internationals, Saint-Grégoire, France) for 2 weeks. After obtaining a sufficient number of cells, they were cultured in base Williams' Medium E (12551032) with differentiation medium supplementation

(ADD721C) (Biopredic internationals, Saint-Grégoire, France) for 2 weeks to transform them into mature human hepatocytes.

HepG2 cells (ATCC® HB-8065™), NCI H2405 cells (ATCC® CRL-5944™) and Caco2 cells (ATCC® HTB-37™) were purchased from American Type Culture Collection (ATCC, Manassas, VA). HepG2 cells are a human hepatoma cell line which was cultured in base medium Eagle's minimum essential medium EMEM (ATCC 30-2003) with 10% Fetal Bovine Serum FBS (ATCC 30-2020).

NCI H2405 cells are human lung adenocarcinoma cells which were cultured in RPMI-1640 Medium (ATCC 30-2001) with 5% Fetal Bovine Serum (FBS; ATCC 30-2020).

Caco2 cells are a colorectal adenocarcinoma cell line. These cells were cultured in base medium Eagle's minimum essential medium EMEM (ATCC 30-2003) with 10% Fetal Bovine Serum FBS (ATCC 30-2020).

KD screening of HepaRG cells:

Specific targeting short interfering RNAs (siRNAs) for 30 potential TFs and nontargeting siRNA were ordered from Dharmacon (Lafayette, CO). Differentiated HepaRG cells were seeded in 6-well plates at a density of 2×10^5 cells/well with appropriate medium overnight. The next day, cells were transfected with either nontargeting siRNA or siRNAs targeting one of the 30 TFs with lipofectamine RNAiMAX Reagent (Thermo Fisher Scientific, Waltham, MA) in OptiMEM media following the manufacturer's instructions. Cells were harvested 48 hours after transfection and RNA was extracted.

Co-transfection of siRNA and *TSPYL4* overexpression plasmids:

Specific targeting short interfering RNAs (siRNAs) for ZBTB7A, REST and nontargeting siRNAs were ordered from Dharmacon (Lafayette, CO). *TSPYL4* OE plasmid construct and

respective empty vector pCMV6-XL4 were purchased from Origene Technologies (Rockville, MD). Differentiated HepaRG, HepG2, or NCI H2405 cells were seeded in 6-well plates at a density of 4×10^5 cells/well overnight. The next day, the cells were co-transfected using nontargeting siRNA or siRNAs targeting *REST* or *ZBTB7A, as well as* empty vector or TSPYL4 OE plasmids using lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA) in OptiMEM media following the manufacturer's instructions. Cells were harvested after 48 hours and RNA was extracted.

Gene expression quantification:

Total RNA was extracted from HepG2 and HepaRG cells with Zymo research the Quick RNA prep kit (Irvine, CA). The concentration of RNA was measured using a Nano drop 8000. PCR reactions were performed with100 ng of total RNA, 5 µl of 2X SYBR green qPCR master mix (Life technologies, CA USA), 1 µl of gene specific primer and distilled water up to 10 µl final volume per reaction. Primer sets for real time PCR are listed in **Supplementary Table 2**. Real time PCR reactions were performed in triplicate using the Applied Biosystems Vii A 7™ real-Time PCR System (Life Technologies, Carlsbad, CA, USA). RNA concentrations across all KD samples and negative controls were normalized to 100ng/µl. Expression levels for TSPYL1, TSPYL 2, TSPYL4, CYP3A4, CYP2C9, and CYP2C19 were tested in all TF KD samples and in negative controls by using the ΔΔCT method in qRT-PCR. GAPDH was used as an internal control for quantification. Alteration of gene expression levels for TSPYL1, TSPYL 2, TSPYL4, CYP3A4, CYP2C9 and CYP2C19 were noted as fold changes as well as statistically significant p-values using student's t- test with < 0.05 being significant.

Chromatin Immunoprecipitation (ChIP) in HepaRG:

HepaRG cells were used to perform ChIP assays to validate TF binding to the promoter regions of the *TSPYL1* and *TSPYL4* genes. Primer sets (see **Supplemental Table 3**) covered 500bp upstream and downstream of exon 1 of the *TSPYL1* and *TSPYL4* genes. ChIP assay was performed using the SimpleChIP® Enzymatic Chromatin IP Kit (Magnetic Beads) (Cell Signaling Technology®, Boston, MA) followed by quantitative PCR using Takara SYBR Green PCR Master Mix reagent (Takara Bio, Saint-Germain-en-Laye, France). The real time Quantitative PCR assays were performed with 2µl of template, 1 µl of primers, 5 µl of PCR Master Mix reagent and nuclease free water to a 10 µl reaction volume. Real time PCR reactions were performed in triplicate using the Applied Biosystems Vii A 7™ real-Time PCR System (Life Technologies, Carlsbad, CA, USA). IP efficiency was calculated using the Percent Input method in which signals obtained from each immunoprecipitation (IP) were expressed as a percent of the total input chromatin using the following formula: Percent Input $= 2\% \times 2^{(CT\ of\ 2\%\ input\ sample - CT\ of\ IP\ samples)}$. Input enrichment values for IP samples were compared with IgG control samples using student's t-test with a significant p-value threshold < 0.05.

TSPYL Promoter Constructs:

Lymphoblastoid cell lines (LCL) with known *TSPYL1* and *TSPYL4* promoter SNP genotypes were used as PCR templates. The primers were designed to amplify 2Kbp promoter regions of the *TSPYL1* and *TSPYL4* genes with KPNI and HINDIII restriction sites at the ends of the primers using the primer3 tool. The primer sets are reported in **Supplementary Table 4**. The PCR amplification was performed with Kapa HIFI hot start ready-mix reagents (KR0370) (Kapa Biosystems, Wilmington, MA) and with primers designed to obtain inserts for ligation with vector. The PCR reactions were set up with 1 ng of genomic DNA, 25µl of Kapa HIFI hot start ready-mix reagent, 1µl of forward and reverse primers and distilled water up to 50 µl. PCR

products were purified with the QIAquick PCR Purification Kit (28104) (QIAGEN, Germantown, MD). KPNI and HINDIII Restriction sites for both pGL4.23 [luc2/minP] vector (Promega, Madison, WI) and purified PCR products were digested at $37\degree$C with restriction enzymes in smart cut buffer (New England Biolabs, Ipswich, MA). Digested PCR products and digested luciferase vectors were separated in 1% agarose gel. Appropriate bands were cut, and gel purified with the QIAquick Gel Extraction Kit (28704) (QIAGEN, Germantown, MD). The ligation reaction included digested 100ng luciferase vector DNA, 48ng digested insert DNA, 1µl Ligase 10X Buffer, 0.1–1unit of T4 DNA Ligase and Nuclease-Free Water to a final volume of 10µl at $4\degree$C overnight (New England Biolabs, Ipswich, MA). The sequences of luciferase constructs (*TSPYL4* promoter Wild Type (WT) and *TSPYL4* promoter Variant (VT)) were confirmed by Sanger sequencing.

Luciferase Reporter Assay:

The firefly luciferase reporter vector, pGL4.23 (Cat#: E8411) was obtained from Promega with inserts encoding either 2Kbp of *TSPYL1* or 2Kbp of *TSPYL4* promoter regions and were used to create *TSPYL1* Wild Type promoter, *TSPYL1* variant promoter, *TSPYL4* wild type promoter or *TSPYL4* variant promoter constructs. HepaRG cells, HepG2 cells and Caco2 cells were seeded in 6 well plates at a density of 4×10^5 for overnight. The next day, the cells were co-transfected with firefly luciferase constructs and Renilla luciferase vector pRL-TK (Promega, Madison, WI) at a ratio of 20:1 using lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA) in OptiMEM media following manufacturer's instructions. 48 hrs after transfection, luciferase assays were performed on these cells with the Dual-Luciferase Reporter assay system (Promega) following the manufacturer's instructions. Luciferase activities were normalized to Renilla

luciferase activities. The relative luciferase activity of variant samples was compared with WT

samples using student's t-test with a significant p-value threshold < 0.05.