



## INTRODUCTION

Drug disposition encompasses the processes of drug absorption into the bloodstream, drug metabolism into different chemicals (mainly in the liver and intestine), distribution of the various chemicals into different tissues, and removal of the chemicals from the body through excretion. The organism's genetic makeup might play a part in the activity of any of these processes and might underlie chemical toxicity and adverse drug reactions (Meyer, 2000). To address this problem, a key goal of predictive medicine is to identify the DNA loci, termed quantitative trait loci (QTLs), which can be used to predict the response to a given medication and its toxicity in a particular patient (e.g., Rost et al., 2004; Harrill and Rusyn, 2008). This challenge can be easily and systematically addressed by utilizing specific mouse models in preclinical pharmacogenetics research.

Mouse pharmacogenetic studies have typically been applied across F2 progeny and backcross populations (Rusyn et al., 2010; Frick et al., 2013); across recombinant inbred (RI) lines derived by crossing two founder strains and inbreeding during many generations (Cook et al., 2004; Hitzemann et al., 2004); and across a predefined collection of classical inbred lines (Montgomery et al., 2013; Yoo et al., 2015). Although these approaches have proved useful in many studies, they are derived mainly from *Mus musculus domesticus* and thus reflect only a partial repertoire of adverse effects, limiting pharmacogenetic investigation.

A promising new model organism has been provided by the recently developed Collaborative Cross (CC) strains, a large, genetically diverse mouse reference population. The CC panel is a collection of RI mouse lines that combine the genomes of eight genetically and phenotypically diverse founder strains—A/J, C57BL/6J [B6], 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ (Aylor et al., 2011; Consortium, 2012). Whereas the first five founders are classical *Mus musculus domesticus* subspecies, the last three are wild-derived strains representing the three *Mus musculus* subspecies, namely *M. m. musculus*, *M. m. domesticus*, and *Mus m. castaneus*. Thus, like classical RI strains, CC mice are genetically reproducible with balanced allele frequencies, but in addition they incorporate a large amount of genetic variation and dense recombination sites (Valdar et al., 2006; Roberts et al., 2007; Yang et al., 2011; Chesler, 2014). The availability of genotyping data across all CC lines opened the way to analyses of QTLs in a variety of traits, including behavioral and morphological phenotypes (Aylor et al., 2011; Chesler, 2014; Mao et al., 2015; Percival et al., 2016), homeostatic immune processes (Kelada et al., 2012; Phillippi et al., 2014), susceptibility to infectious diseases (Durrant et al., 2011; Ferris et al., 2013; Vered et al., 2014; Graham et al., 2015; Gralinski et al., 2015; Lorè et al., 2015), and liver-related functionalities (Kelada et al., 2012; Svenson et al., 2012; Thaisz et al., 2012).

While the CC population has been proven phenotypically diverse, the extent to which drug disposition varies across these strains is still largely unknown (Rusyn et al., 2010; Gelinas et al., 2011; Frick et al., 2013). One of the many mechanisms through which variation in drug disposition can arise is the biotransformation of drugs in the liver. In such

biotransformation, drug metabolizing enzymes (DMEs) and drug transport proteins (DTPs) catalyze the biochemical modification and transport of exogenous chemicals and other xenobiotics (Katz et al., 2008). With regard to hepatic drug metabolism in the CC lines, two key questions arise: (i) Should a large diversity be expected in hepatic biotransformation of particular drugs? (ii) Can CC mice be used to evaluate the crosstalk between drug metabolism and other functionalities of the liver, especially those related to lipid and fatty acid metabolism?

Here we exploit transcriptional mechanisms to dissect genetic variation in hepatic drug metabolism of the CC lines. We focus on *cis*-regulatory variants underlying inter-individual variation in gene expression. Such genetic variants, referred to as “proximal eQTLs,” are central to the understanding of metabolic diversity owing to their relatively large genetic effect size (Wittkopp and Kalay, 2012; Bryois et al., 2014) and the plasticity of proximal factors over short evolutionary time scales (Wray, 2007). Proximal regulatory variation is known to affect metabolic traits (e.g., Hsieh et al., 2007; Zhong et al., 2010; Kang et al., 2012), but the organization and architecture of such elements in governing hepatic drug metabolism has not yet been subjected to comprehensive characterization.

We leverage RNA-sequencing technology (Mortazavi et al., 2008; Rosenkranz et al., 2008; Parkhomchuk et al., 2009) to generate transcription profiles of the liver tissues from 29 CC lines (55 individuals). By applying eQTL analysis on these data we characterized the genetic control over expression of drug disposition enzymes—either control of the overall expression of genes or of the expression of particular transcript isoforms. Using these predictions we compiled a network of connectivity between genomic loci and the metabolism of specific drugs, highlighting potential joint genetic effects and drug-drug interactions. Further analysis of nuclear receptors participating in the regulation of cellular-biochemical pathways provided DNA variants that affect the crosstalk between lipid metabolism and drug metabolism. Thus, our study supports the premise that the CC population is a valuable model system for the investigation of genetic variation in response to a wide range of chemicals and drugs, and may further offer mechanisms by which DNA variation contributes to the relationship between lipid metabolism and adverse clinical effects. Moreover, our approach provides a general strategy for a system-level mapping of eQTL-drug connectivity across a genetically diverse population.

## MATERIALS AND METHODS

### CC Lines

The Collaborative Cross (CC) recombinant inbred mouse lines were used as described elsewhere (Iraqi et al., 2008; Durrant et al., 2011; Consortium, 2012). Animals were housed on hardwood chip bedding in open top cages at the animal facility of Tel-Aviv University (TAU) under a 12-h light/dark cycle. Mice were given tap water and rodent chow *ad libitum* throughout the experiment. Liver tissues were collected from 8- to 10-week old male CC mice from the TAU cohort at inbreeding generation between 16 and 42. A total of 55 mice from 29 CC lines were

used (Table S1). All experimental protocols were approved by the Institutional Animal Care and Use Committee (IACUC) at TAU (approved protocol M-13-033) according to the national guidelines.

## RNA Extraction, RNA-Seq Library Preparation, and Sequencing

The liver tissues were dissected and subsequently stored in sterilized tubes at  $-196^{\circ}\text{C}$  (in liquid nitrogen). The RNA was extracted using the RNeasy Plus kit procedure (Qiagen). RNA quality was tested on a 2100 BioAnalyzer (Agilent); in all samples, RNA Integrity number (RIN) exceeded 7. The RNA-Seq libraries were prepared using the TruSeq Stranded mRNA library preparation kit (Illumina). Libraries were pooled and sequenced on the Illumina HiSeq 2000 and 2500 sequencers with Illumina v3 sequencing chemistry. Paired-end sequencing was performed by reading 50 bases at each end of a fragment. Overall, each sample consisted of 24–37.5 M RNA-sequencing fragments with an average of 31.5 M fragments. This data is accessible through GEO Series accession number GSE77715. A detailed description of our data analysis appears below (see a summary of the computational pipeline in Table S2).

## RNA-Seq Quantification

RNA-Seq data was mapped and quantified using RSEM version 1.2.18 (Li and Dewey, 2011) with the mouse genome (UCSC, mm9, NCBI37) and annotation file (Ensembl version 37.67). The reference was created by the RSEM `rsem-prepare-reference` command, followed by calculation of the expression level of genes using the `rsem-calculate-expression` command. The analysis was applied with default parameters and using Bowtie 2 version 2.1.0 (Langmead and Salzberg, 2012). Overall, the average percentage of unmapped fragments was 11.9%, (min = 9.4%, max = 24.3%); the average percentage of fragments aligned to a single gene was 59.9% (min = 49.6%, max = 69.5%), and among them, fragments aligned to just one isoform were 32.6% (min = 26.9%, max = 38.7%).

*Total expression levels* were measured by RSEM's FPKM metric, defined as the number of fragments mapped to the genomic region of a gene per kilobase of the gene's exons and per million mapped fragments. An *isoform-ratio level*, in contrast, is defined as the percentage of detectable fragments mapped to a given alternatively spliced isoform of a gene, out of the total number of fragments mapped to that gene (the RSEM's IsoPct metric). The *expression trait* of a gene refers to either the overall expression (a *total-expression trait*) or the percentage of a specific isoform (an *isoform-ratio trait*). In the following we describe the transformation and association test steps, which are applied separately for each trait. For simplicity, we omit the index of the gene (either a total-expression trait or an isoform-ratio trait) whenever possible.

## Data Transformation

For a given expression trait, denote its level in individual  $i$  by  $v_i$ . The maximal level of a trait, denoted  $v_M$ , is  $\max\{v_i\}_{i=1}^n$ , where  $n$  is the number of individuals. The *bin size*  $\delta$  of a trait is defined as  $\max\{0.5, \frac{v_M}{200}\}$ , where the first term (0.5) represents an accuracy

threshold and the second term represents 0.5% of the maximum level.

Let  $v_i^*$  be the rounded level of  $v_i$  to the accuracy of bin size  $\delta$ . Based on these rounded levels, the van der Waerden normal scores transformation (Lehmann, 1975; Aylor et al., 2011) is defined by

$$u_i = \phi^{-1}\left(\frac{r_i^*}{n+1}\right),$$

where  $u_i$  is the adjusted level of the trait in individual  $i$ ,  $\phi^{-1}(p)$  is the quantile function of the normal distribution with probability  $p$ , and  $r_i^*$  is the rank of  $v_i^*$  among the  $n$  values  $\{v_1^*, v_2^*, \dots, v_n^*\}$  with ties resolved by the average rank. Throughout this study, we refer to the adjusted traits  $\{u_i\}_{i=1}^n$  rather than to the original measurements.

## Association Tests

An association score between a given expression trait and a given genome interval was calculated by regressing the trait on the contribution of their eight CC founder strains as previously described (Mott et al., 2000). More formally, for a given genome interval and a trait, the *association score* is the likelihood ratio (LR) between the null model  $u_i = \mu + \beta_{m_i} + \varepsilon_i$  and the genetic model:

$$u_i = \mu + \sum_{k=1}^8 \alpha_k g_{k,m_i} + \beta_{m_i} + \varepsilon_i, \quad (1)$$

where  $m_i$  is the CC line of the  $i$ -th individual;  $g_{k,m_i}$  is the haplotype probability of the  $k$ -th founder ( $k \in \{1, \dots, 8\}$ ) in CC line  $m_i$ ,  $\mu$  is the intercept value,  $\alpha_k$  is the genetic fixed effect of the  $k$ -th founder,  $\beta_m$  is the random effect of CC line  $m$ , and an error term  $\varepsilon_i$  is assumed to be normally distributed  $\varepsilon \sim N(0, \sigma^2)$ . No other covariates were used. The mixed model regression was implemented in R by using the `lme4` package version 1.1-7 (Bates et al., 2015).

The false discovery rate (FDR) was estimated by comparing the LR values in real data to the distribution of the LR values using randomly permuted datasets. Throughout this study, the *permutation FDR* was calculated as  $C_{perm}/C_{real}$  where  $C_{real}$  and  $C_{perm}$  denote the number of traits with LR values exceeding a certain threshold in the original (real) dataset and in the permuted dataset, respectively ( $C_{perm}$  is averaged over 100 permuted datasets). In order that the permutations will specifically randomize the fixed effect (the genetics) rather than the random effect, the shuffling was applied on the CC line labels.

We applied the association tests to genes that are expressed in the liver ( $v_M \geq 2$ ) and to a single isoform-ratio traits from each of these genes: the isoform with the highest number of different rounded levels in the set  $\{v_i^*\}_{i=1}^n$ . Since we are interested in the variation in expression and to reduce the running time of the analysis, we further excluded traits with a low ( $<12$ ) number of different rounded levels across the 55 individuals. Altogether, the association tests were applied on 5950 total-expression traits and 4712 isoform-ratio traits.

The association tests were applied on genotyping dataset that was obtained from the UNC systems genetics repository

(<http://csbio.unc.edu/CCstatus>) and included measurements from MegaMUGA—a 77K marker genotyping array based on the Illumina Infinium platform. The genotyping data was computationally validated using comparison with the polymorphic loci in the RNA-Seq data (Figure S1). For a given genome interval, the haplotype probabilities  $g_{k,m}$  of each of founder  $k$  in each CC line  $m$  were calculated using the HAPPY package version 2.4 (Mott et al., 2000). Altogether, the association tests were applied on 23,217 genome intervals for which the haplotype probabilities were calculated. Unless stated otherwise, the analysis was focused on proximal genome intervals. To that end, association scores were calculated using genomic intervals whose distance to the gene's transcription start site is less than 5 Mbp.

Throughout the manuscript, we use the following terminology: a *proximal eQTL* is defined as a nearby genome interval whose FDR is lower than 0.01 (based on 100 permutations). There are two types of proximal eQTLs: *total-expression eQTLs* that are associated with the expression level of total-expression traits, and *isoform-ratio eQTLs* that are associated with the percentage of alternatively spliced isoforms (see Tables S3, S4 for full lists).

We note that it is possible to use the optimized regression parameters (from equation 1) to determine the contribution of each founder to the overall regulatory variation. As previously described (Aylor et al., 2011; Durrant et al., 2011), we define the *founder effect* as  $\frac{\alpha_k}{\max\{\alpha_k\}_{k=1}^8}$ : the higher the (absolute) founder effect on a certain eQTL target, the stronger the contribution of the DNA changes in the eQTL region of this founder strain.

## Construction of an eQTL-Drug Connectivity Map

The eQTL-drug connectivity map was generated in several steps. In step 1 we assembled a reference collection of manually curated drug-specific sets of enzymes. Each set in this collection includes a group of genes that play a role in the metabolic reactions of a particular drug, based on direct experimental evidence. The reference collection consists of 881 gene sets relating to 165 different drugs, which were assembled from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and from the IPA database (QIAGEN, Redwood City, CA). In step 2, additional enzymes were added to each set based on indirect evidence. Specifically, each of the manually curated sets was further expanded with the alternative genes of the same chemical reaction (that is, with the same EC numbers). Altogether, steps 1 and 2 produced a reference collection of drug-specific gene sets, where the assignment of a gene to a particular set is based on either direct (step 1) or indirect (step 2) experimental evidence. Next, in step 3 we removed genes that were not associated with a proximal eQTL. In particular, given the reference collection from step 2, we retained only those genes that were significantly associated with at least one proximal eQTL (using the same thresholds as detailed above; see final collection in Tables S5, S6).

Building on this collection, the eQTL-drug connectivity map consists of three types of nodes: drugs with at least one non-empty curated gene set, genes in these non-empty sets, and the

proximal eQTLs of these genes. The map contains edges between each drug node and its corresponding nodes of genes, as well as edges between each gene node and the nodes of its proximal eQTLs.

## Demonstration of Genetic Variation in Splicing Events

To demonstrate the alternative splicing events of a particular gene in a given sample, reads were aligned to the genome using Tophat version 2.0.9 (Kim et al., 2013) with  $-\text{max-intron-length } 20,000$  (thus avoiding most cases of aligning the same fragment to two nearby genes). Uniquely mapped reads were extracted by filtering out those reads carrying poor mapping quality ( $<10$ ). Using the TopHat alignment, the IGV software (Thorvaldsdóttir et al., 2013) was used to show the reads' coverage together with the amount of splice junctions (see details below).

## RESULTS

### Characterization of Proximal eQTLs Acting on Hepatic Drug Disposition Enzymes

To investigate transcription diversity in the livers of CC mice, we sequenced total RNA from the livers of 29 distinct CC lines (1–3 individuals per strain, 55 individuals in total; Table S1). On the basis of these data we quantified the total expression of each gene (“total-expression traits”) as well as the relative expression of the annotated isoforms (“isoform-ratio traits”; see Materials and Methods and Table S2). We found that the global expression profiles are similar between individuals of the same strains for both total-expression and isoform-ratio traits (Figures S2, S3); this dataset is therefore amenable to our study. In the following we focus on 5950 total expression traits and 4712 isoform-ratio traits that were highly variable across the CC mice.

We tested the association of each trait separately against all polymorphisms. A genome-wide map indicates, as expected, that this analysis has no spurious *trans*-eQTL bands (Figure S4) and do not show inflation of the association test statistics (Figure S5). The analysis is focused on proximal eQTLs, using a cutoff of 5 Mbp as evidence for *cis* regulatory variation. This permissive genomic range was selected to account for the lack of precision in the associated genome intervals (Figure S6). We used permutation to establish the null distribution of the association test statistics and then exploited the null distribution to calculate a permutation-based false discovery rate (“permutation FDR”) score (see Materials and Methods). At a permutation FDR of 0.01 we identified proximal eQTLs associated with the total expression of 365 genes (“total-expression eQTLs”) and associated with the expression of 243 specific transcript isoforms (“isoform-ratio eQTLs”), a total of 608 significantly associated traits (see Materials and Methods and Tables S3, S4). We note that the similar numbers of total-expression eQTLs and isoform-ratio eQTLs is in accordance with previous studies in human cohorts (González-Porta et al., 2012; Battle et al., 2014). Of the 3400 genes found to have dual annotation (both total-expression and isoform-ratio traits), in 43 we obtained both total-expression

eQTL and isoform-ratio eQTL, not necessarily in the same genomic interval.

To characterize biochemical networks in the context of inherited transcriptional variation, we examined the 608 eQTL-associated traits and found them to be enriched in a curated list of biotransformation reactions from the “Ingenuity Pathway Analysis” (IPA) database ( $P < 0.03$ , hyper-geometric test). Specifically, 40 drug disposition genes were found to be associated with a proximal eQTL, which could be further classified according to their particular type of enzymatic reaction (Table 1):

- (i) Functionalization reactions by oxidation, reduction and hydrolysis, which either activate or detoxify the drug. Among the eQTL-associated functionalization DMEs, oxidation is catalyzed by cytochrome P450 (*Cyp2a/2c/2d/3a*), alcohol and aldehyde dehydrogenases (*Adhfe1* and *Aldh8a1/16a1*), thiol-disulfide oxidoreductase (*Glx2*), and FMO (*Fmo1*); reduction is catalyzed by aldo-keto reductases (*Akr1c13/19*); and hydrolysis is catalyzed by various esterases (*Ces1g/2h/3a*, *Siae*, *Sulf2*), epoxide hydrolase (*Ephx2*), dihydropyrimidinase (*Dpys*), glucuronidase (*Gusb*), and glyoxalase (*Glo1*).
- (ii) Conjugation reactions that transfer a functional group from a cofactor to a substrate chemical, resulting in detoxification followed by excretion. The eQTL-associated conjugation DMEs catalyze the transfer of various functional groups, including UDP-glucuronosyl, amino acid, N-acetyl, methyl and glutathione-S (*Ugt1a/Ugt3a*, *Ccbl1/2*, *Nat8*, *Tpmt*, and *Gsta2/m6/z1/Mgst3*, respectively).
- (iii) Transport reactions, mediated by DTPs that have a role in the facilitated carrying of drugs across cellular membranes (Katz et al., 2008; Penner et al., 2012). The eQTL targets in this class include two types of DTP families: an ATP-binding cassette (*Abcc6*) and a solute-linked carrier (*Slco1a1*).
- (iv) Transcription regulation. The identified eQTL targets include CAR, a nuclear receptor that regulates the transcription of drug disposition enzymes.

## Mapping the Connectivity between eQTLs and Drug Metabolism

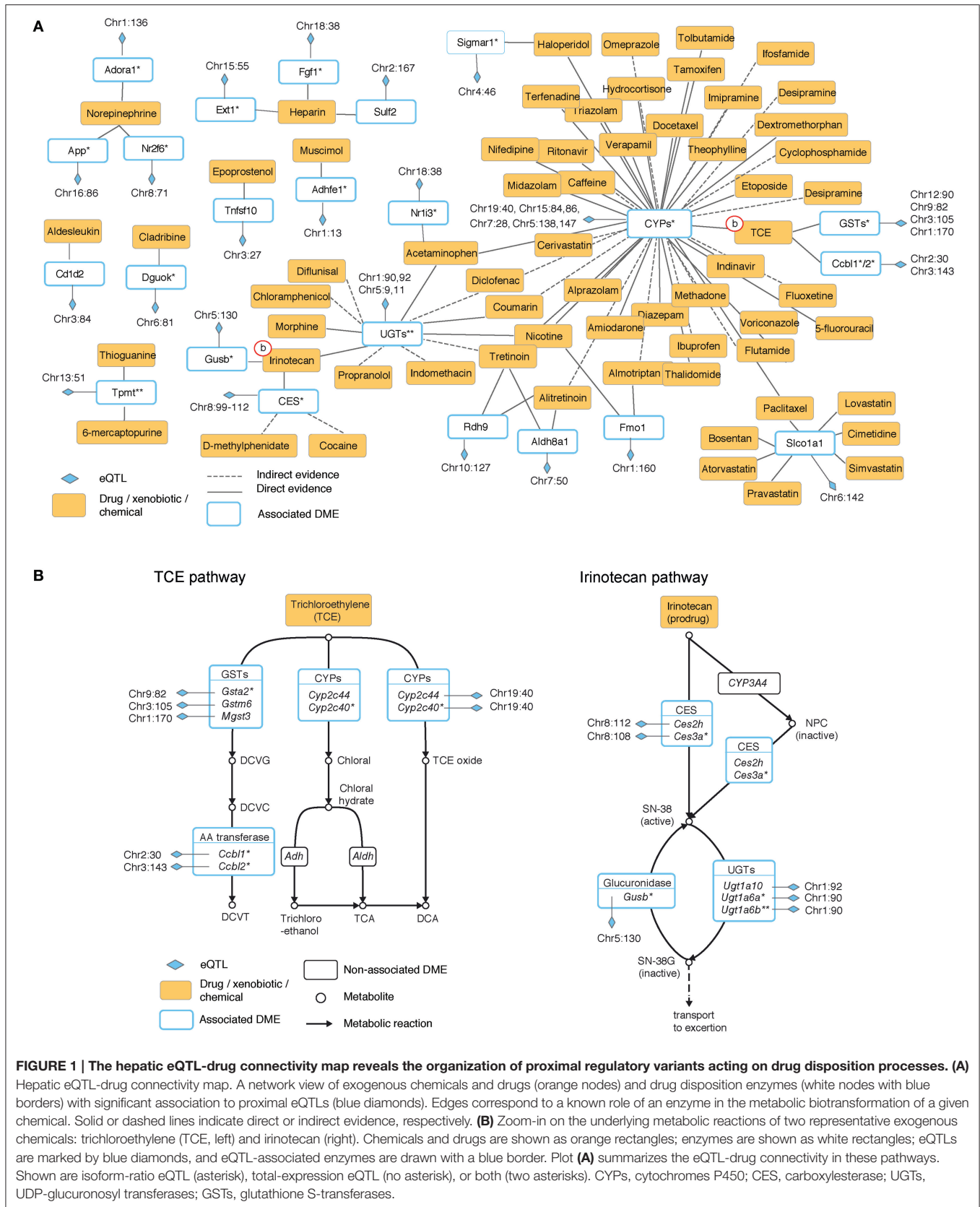
To obtain a global perspective on the participation of eQTLs in drug metabolism we used expert-curated drug-specific sets of enzymes, where each set is a collection of enzymes that play a role in the biotransformation of one particular drug (see Materials and Methods). By analyzing these sets we identified 63 drugs whose biotransformation is perturbed by at least one proximal eQTL (Table S5) and of which 49 are affected by two or more eQTLs that reside in a distinct genomic location (at least 10 Mb apart; Table S6). We organized this information as a network, referred to as the “eQTL-drug connectivity map” (Figure 1A). The map is composed of three types of nodes: drugs, eQTLs and enzymes. Each drug is connected to its metabolizing enzyme nodes and each enzyme is connected to its underlying eQTL nodes.

Characterization of drugs in terms of their perturbing eQTLs provides potential joint effects that may underlie the response to

**TABLE 1 | Summary of proximal eQTLs underlying the biotransformation of drugs in livers of the CC mouse population.**

Reaction	Enzyme class	Enzyme	eQTL	
Functionalization: Cytochromes P450 (CYPs) oxidation		<i>Cyp2a22</i>	chr7:28	
		<i>Cyp2c40*</i>	chr19:40	
		<i>Cyp2c44</i>	chr19:40	
		<i>Cyp2d11</i>	chr15:84	
		<i>Cyp2d12</i>	chr15:86	
		<i>Cyp3a13</i>	chr5:138	
		<i>Cyp3a16</i>	chr5:147	
		<i>Cyp3a25**</i>	chr5:147	
		Alcohol dehydrogenase	<i>Adhfe1*</i>	chr1:12
		Aldehyde dehydrogenase	<i>Aldh8a1</i>	chr10:22
		<i>Aldh16a1</i>	chr7:50	
	Thiol-disulfide oxidoreductase	<i>Glx2**</i>	chr1:146	
	Flavin-containing monooxygenase (FMO)	<i>Fmo1</i>	chr1:160	
Functionalization: Aldo-keto reductase reduction		<i>Akr1c13</i>	chr13:3	
		<i>Akr1c19</i>	chr13:5	
Functionalization: Carboxylesterase (CES) hydrolysis		<i>Ces1g</i>	chr8:99	
		<i>Ces2h</i>	chr8:112	
		<i>Ces3a*</i>	chr8:108	
		Sialic acid acetyltransferase	<i>Siae</i>	chr9:40
		Sulfatase (esterase)	<i>Sulf2</i>	chr2:167
		Epoxide hydrolase	<i>Ephx2</i>	chr14:62
		Dihydropyrimidinase	<i>Dpys*</i>	chr15:39
		Glucuronidase	<i>Gusb*</i>	chr5:130
		Glyoxalase	<i>Glo1**</i>	chr17:32
	Conjugation	UDP-glucuronosyltransferase (UGT)	<i>Ugt1a6a*</i>	chr1:92
<i>Ugt1a6b**</i>			chr1:90*,92	
<i>Ugt1a10</i>			chr1:92	
<i>Ugt3a1</i>			chr15:11	
<i>Ugt3a2</i>			chr15:9	
Methyltransferase		<i>Tpmt**</i>	chr13:51	
N-acetyl transferase		<i>Nat8</i>	chr6:86	
Amino acid (AA) transferase		<i>Ccbl1*</i>	chr2:30	
		<i>Ccbl2*</i>	chr3:143	
Glutathione S-transferase (GST)		<i>Mgst3</i>	chr1:170	
		<i>Gsta2*</i>	chr9:82	
		<i>Gstm6</i>	chr3:105	
		<i>Gstz1</i>	chr12:90	
Transport	ATP-binding cassette	<i>Abcc6</i>	chr7:51	
	Solute-linked carrier	<i>Slco1a1</i>	chr6:142	
Transcription factors	Nuclear receptor	<i>Nr1i3 (CAR)*</i>	chr1:174	

Shown for each drug disposition reaction (column 1) are various enzyme classes (column 2), the identified eQTL-associated genes in each class (column 3), and the top-ranking proximal eQTL of each gene (genomic positions; column 4). The second number indicates the distance in Mb from the start of the chromosome. Expression of each enzyme is associated with an isoform-ratio eQTL (asterisk), total-expression eQTL (no asterisk), or both (two asterisks). Overall, we found 30 total-expression eQTLs and 15 isoform-ratio eQTLs that underlie the expression of 40 different drug disposition enzymes.



administration of a drug. For example, trichloroethylene (TCE) is a small molecule ( $C_2HCl_3$ ) with carcinogenic potential. In the connectivity map, TCE is connected to three eQTL targets (cytochrome P450 [CYPs], glutathione S-transferases [GSTs] and amino acid [AA]-transferases). The connectivity map is based on the known biochemical pathway involving the same combination of eQTL targets (**Figure 1B**, left). The biochemical pathway suggests a potential process by which joint genetic effects may influence the response to TCE.

The irinotecan pathway provides another example of potential genetic interactions within the eQTL-drug connectivity map (**Figure 1B**, right). Irinotecan is a drug used for the treatment of cancer (Nagar and Blanchard, 2006), and there are known individual differences in susceptibility to its effect (Nagar and Blanchard, 2006; Guo et al., 2007; Marsh and Hoskins, 2010). The pathway consists of activation of irinotecan to SN-38 (by members of the carboxylesterases [CES] family) and deactivation of SN-38 into SN-38G (by the UDP-glucuronosyltransferase [UGTs] and the glucuronidase families). Overall, whereas the eQTL-drug connectivity map (**Figure 1A**) summarizes the identity of the relevant metabolizing enzymes, mechanistic visualization (**Figure 1B**) suggests the existence of genetic interactions between different eQTLs along the cascade of metabolic reactions.

In addition, the connectivity map indicates that some of the associated enzymes (8 of 23; 34%) participate in the biotransformation of two or more drugs, highlighting potential regulatory variation that may lead to drug-drug interactions. One example is the solute-linked carrier *Sco1a1*, whose proximal regulatory variation probably has an effect on at least seven drugs, including lovastatin, bosentan and cimetidine. Another example is the family of UDP-glucuronosyl transferases (UGTs), which is connected to 11 different drugs. This suggests specific proximal regulatory variation that has an influence on a large repertoire of drugs.

We note that it is possible to identify groups of CC lines on the basis of their co-variation in the expression of drug-specific metabolizing enzymes. For example, as in the case of the irinotecan pathway (**Figure S7**), there is a clear grouping of the CC lines based on the co-variation of their expression across the relevant enzymes (e.g., a distinct expression of lines IL-670 and IL-785 compared to lines IL-611 and IL-3438 across the expression of *Ces2h/3b*, *Ugt1a10/6a/6b*, and *Gusb*). Based on this grouping it is possible to select a non-redundant subset of lines that covers a large variety of responses (e.g., by choosing a single line from each group). Because distinct drugs initiate different biotransformation pathways perturbed by distinct combinations of eQTLs, this strategy may provide a tailored drug-specific selection of CC lines, opening the way to future expression-based selection of representative CC lines that can be used to test adverse effects in pharmacogenetic studies.

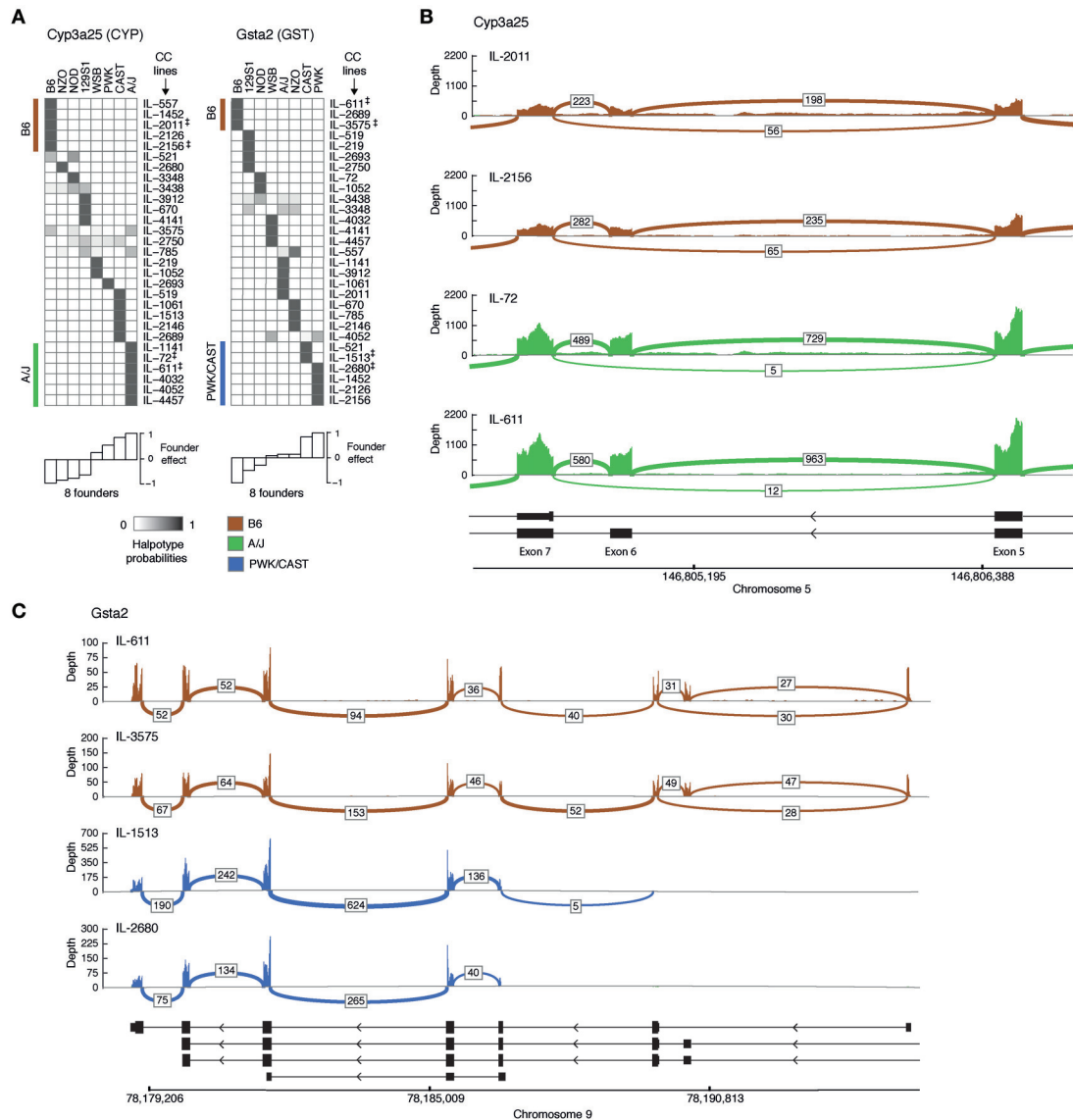
## Genetic Variation in Alternative Splicing of Hepatic Drug Disposition Enzymes

Our data showed that a substantial fraction of the identified associations are a result of splicing variability. Among 608

associated traits, we found 243 isoform-ratio traits (~40%; Tables S3, S4). Furthermore, we observed a similar percentage of isoform-ratio traits within the associated drug disposition genes (33%, 15 from 45 traits; **Table 1**). In the TCE pathway, for instance, out of seven associated genes we identified four (57%) whose association was due to variation in splicing events spanning four different metabolic transformations (ending with TCE oxide, chloral, DCVG and DCVT; **Figure 1B**, left). Similarly, among six associated genes in the irinotecan pathway, four are controlled at the level of alternative splicing (*Ces3a*, *Gusb*, *Ugt1a6a/b*, 66%; **Figure 1B**, right). These results are comparable with the reported isoform-ratio percentages of 38% (496 out of 1290) and 44% (529 out of 1191) in human Caucasian and Yoruba cohorts (60 and 69 individuals, respectively, González-Porta et al., 2012), and are in agreement with studies indicating that isoform-ratio eQTLs are prevalent but less abundant than total-expression eQTLs (e.g., Battle et al., 2014).

We then turned to characterizing representative examples of isoform-level associations. For each eQTL-associated gene we first selected founder strains that differ substantially in their effects, and then chose CC lines that carry the haplotype of the selected founders in the associated locus (eQTL) of the gene. We note that CC lines for a gene are selected because they carry the genetic information of founders, which leads to a major variation in the fraction of expression of at least one isoform; the selected CC lines should therefore show a distinct distribution of the spliced junctions. Here we focus on two genes, *Cyp3a25* and *Gsta2*. For each of these genes we first exemplify the abovementioned CC selection procedure and then show the splice junctions in each of the selected strains.

1. *Cyp3a25* is a functionalization enzyme that oxidizes a variety of compounds including xenobiotics and steroids. We found that the founder B6 and A/J strains differ substantially in their effects on at least one isoform of this gene (**Figure 2A**, bottom left). We therefore distinguish between CC lines that carry the B6 haplotype (IL-557, IL-1452, IL-2011, IL-2126, IL-4156 strains) and those carrying the A/J haplotype (IL-1141, IL-72, IL-611, IL-4032, IL-4052, IL-4457 strains) in the genomic region of *Cyp3a25* (**Figure 2A**, left). Specifically, we focus on two representative B6-carrying lines (IL-2011 and IL-2156) and two A/J-carrying lines (IL-72 and IL-611). **Figure 2B** displays the read coverage of the exons and the numbers of junction reads of the selected strains. We found multiple reads aligned to the junction-spanning exons 5 and 7 in B6-carrying strains (56 and 65 reads in lines IL-2011 and IL-2156, respectively). Notably, the numbers of these exon-skipping reads was substantially lower in the A/J-carrying individuals (5 and 12 reads in lines IL-72 and IL-611, respectively). The widespread exon-skipping events observed in B6-carrying strains raise the possibility that these events reflect a higher level of expression in B6-carrying than in A/J-carrying strains. However, we observed the opposite effect, with a higher coverage level and a higher number of non-skipping splicing reads in the A/J-carrying individuals. For instance, 729 and 963 junction reads between exons 6 and 7



**FIGURE 2 | Genetic variation in alternative splicing of drug disposition enzymes. (A)** Haplotype probabilities of the eight founder lines (columns) in 29 CC lines (rows), calculated in the isoform-ratio eQTLs of two drug disposition genes, *Cyp3a25* (left) and *Gsta2* (right). The gray scale indicates haplotype probabilities, ranging between zero (white) and 1 (dark gray). The calculated effect of each founder is shown in white bars (bottom). Groups of CC lines with the largest (positive and negative) founder effect of their haplotype are marked in brown and green (B6- and A/J-carrying lines; *Cyp3a25*) or brown and blue (B6- and CAST/PWK-carrying lines; *Gsta2*). Double daggers indicate two representative CC lines in each of these groups, which were used for displaying the raw sequencing reads in plots (B,C). **(B,C)** Raw reads of selected strains for the genes *Cyp3a25* and *Gsta2*. The read's coverage over exon is displayed as bar graph, and the number of reads across splice junctions (junction depth) are displayed by arcs. Arcs with junction depth < 5 were omitted. The known isoforms are indicated in black (bottom). **(B)** The *Cyp3a25* locus, focusing on exons 5, 6, and 7 in CC individuals that carry the B6 haplotype (brown) and the A/J haplotype (green) in the associated eQTL. **(C)** Entire *Gsta2* locus (excluding exon 1 of the longest isoform) in CC individuals that carry the B6 haplotype (brown) and the PWK or CAST haplotype (blue) in the associated locus. B6, C57BL/6J; 129S, 129S1/SvIm; NOD, NOD/ShiLtJ; WSB, WSB/EiJ; NZO, NZO/HILtJ; CAST, CAST/EiJ; PWK, PWK/PhJ.

were mapped in the two A/J-carrying strains compared to 198 and 235 junction reads in B6-carrying ones. This confirms that the larger numbers of exon 6-skipping events in B6-carrying strains are not merely a result of the overall level of expression.

2. *Gsta2* is a conjugation enzyme that adds a reduced glutathione to hydrophobic electrophiles. Relying on the observation that the founder effect of the B6 line differs substantially from

the PWK and CAST lines (Figure 2A, right), we focused on two B6-carrying CC lines (IL-611 and IL-3575), one CAST-carrying line (IL-1513) and one PWK-carrying line (IL-2680; Figure 2C). We observed alternative 5' start sites in which only the B6-carrying individuals have alternative upstream start sites. This could not be ascribed merely to the higher overall expression of *Gsta2* in the B6-carrying individuals,

since the coverage of reads (across all exons) is lower in B6-carrying than in CAST/PWK-carrying individuals. The difference between the haplotypes can be attributed to an alternative location of the 5' start site, or alternatively, to a haplotype-specific 5'–3' RNA degradation. Previously annotated isoforms do not fully explain the observed reads in the PWK and CAST CC lines.

## Substantial Transcriptional Diversity in the Crosstalk between Drug Metabolism and Lipids Metabolism

We next analyzed regulatory programs of ligand-activated transcription factors called nuclear receptors (NRs). NRs were selected not only because they play a major role in transcription regulation, but also since many DMEs and DTPs are induced by their own substrates through the activity of NRs. Chemical signals (ligands) of NRs consist of exogenous drugs and xenobiotics, as well as endogenous small molecules such as steroid hormones and cholesterol (e.g., Evans and Mangelsdorf, 2014). Using the extensive mapping of regulatory programs within the curated IPA, we identified seven different NRs (or receptor complexes) whose targets are enriched among the 608 associated traits (Figure 3). Specifically, we found enrichment of targets in xenobiotics- and drug-activated NRs, including the xenobiotics sensor PXR—with or without its heterodimerization partner RXR $\alpha$  ( $P < 0.0024, 0.0158$ , respectively; hyper-geometric test)—and the xenobiotics sensor CAR as an heterodimer with RXR $\alpha$  ( $P < 0.0165$ ). We also found that the hepatic eQTL targets were enriched with known targets of ROR $\alpha$ , ROR $\gamma$ , and PPAR $\alpha$ , three NRs that are activated by cholesterol and fatty acid ligands ( $P < 0.0012, 0.0012, 0.0081$ , respectively).

The map of NR-dependent eQTL targets shows a clear overlap between targets of xenobiotics-sensitive regulators and targets of lipid-sensitive regulators (Figure 3). Specifically, six of 16 genes that are regulated by xenobiotics sensors (PXR and CAR) are also regulated by lipid sensors (ROR $\alpha$ , ROR $\gamma$ , and PPAR $\alpha$ ). A close examination of the map showed that the overlap between lipid metabolism and drug metabolism is even more prominent: the known DME genes *Sulf2*, *Cyp3a25*, *Cyp2c40*, *Gstz1*, and *Mgst3* are targeted by lipid-sensing nuclear receptors; similarly, two genes that are controlled by xenobiotics-activated receptor are well-established lipid metabolizing enzymes (*Ugt1a6* and *Cyp2c44*).

These results are consistent with known biochemical reactions that may act on unrelated compounds. For example, *Cyp3a16* has a role in degradation of steroids, in addition to its role in oxidation of a variety of drugs, including tamoxifen, paclitaxel, haloperidol and tretinoin; similarly, *Ugt1a6b* participates in hepatic degradation of C<sub>18</sub>-steroids and also in conjugation of levothyroxine, nicotine and irinotecan (Figure 1A). Furthermore, our analysis is consistent with previous reports, both in human and mouse, where disordered lipid metabolism has an effect on clinical drug disposition and the other way around. For example, non-alcoholic fatty liver disease (NAFLD) patients exhibit altered metabolism of drugs (Buechler and Weiss, 2011), and DTPs are down-regulated in non-alcoholic steatohepatitis (NASH) subjects (Lake et al., 2011). Furthermore, several genetic modifiers of NAFLD and NASH

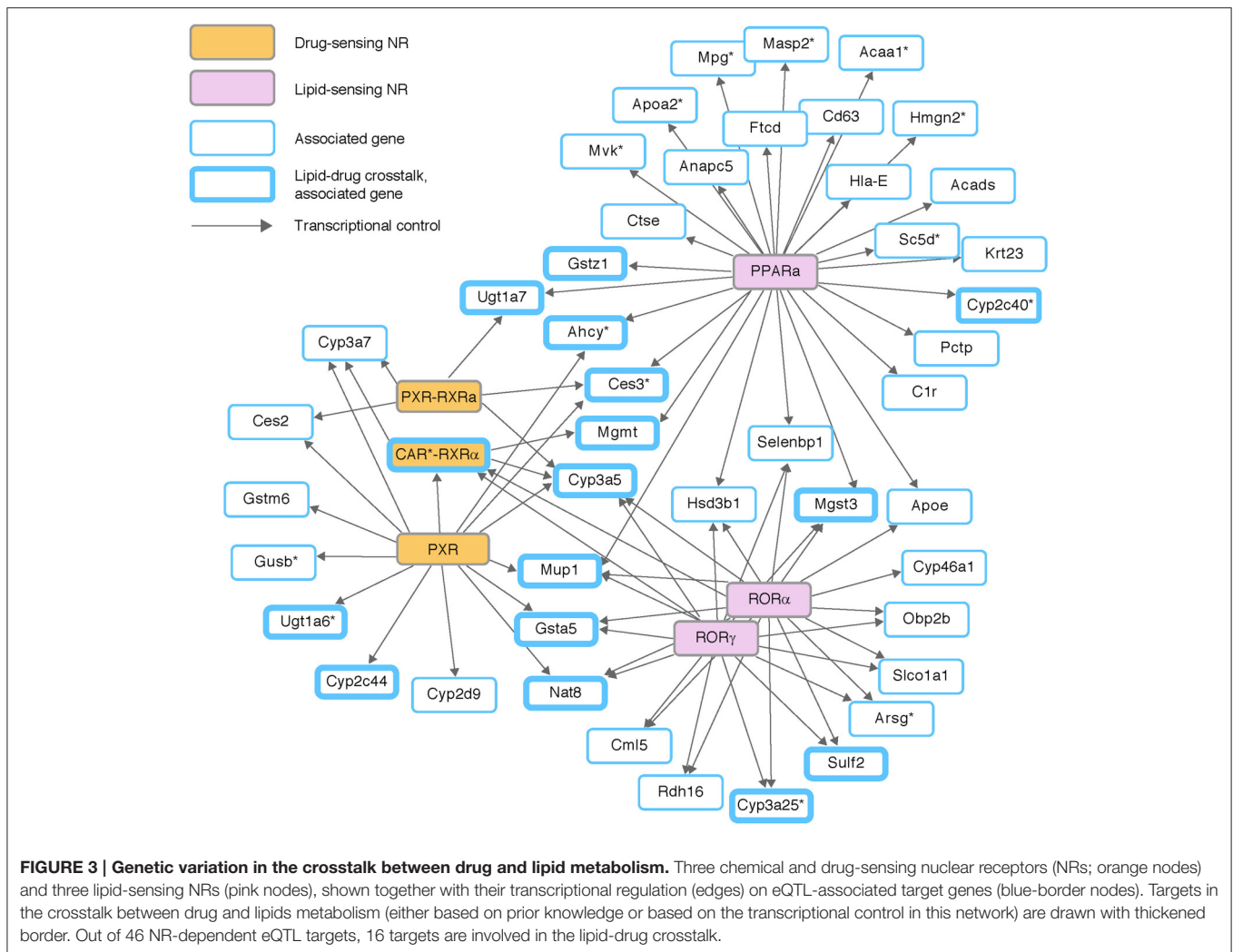
have a documented effect on the efficacy of drugs (reviewed in Naik et al., 2013). Our results in the CC panel demonstrate the prevalence of genetic variation acting on the lipid-drug crosstalk. CC mice can therefore be used to investigate the relationships between these factors and their roles in the interaction between lipid metabolism and metabolism of drugs and xenobiotics.

## DISCUSSION

A central challenge in pharmaceutical research has been to investigate genetic variation in response to drugs, chemicals and xenobiotics. The panel of CC lines is a promising model for pharmacogenomic studies because of the large amount of genetic variation and the ability to investigate the molecular response to drugs within specialized internal tissues. However, the effect of genetic variation on drug disposition enzymes remain to be elucidated. Here we defined the diversity and complexity of genetic variation on the expression of drug disposition genes across the CC lines. Our results indicate a previously unknown overrepresentation of hepatic drug disposition genes that are affected by proximal regulatory variation ( $P < 0.03$ ; Table 1). Further inspection showed the complexity of inter-relationships between regulatory variants, highlighting various potential interactions between drugs due to shared eQTLs (Figures 1–3). This analysis, therefore, provides an informative view to guide future pharmacogenetics and mechanistic studies across the CC strains. For example, we found a significant effect of lipid metabolism on pharmacokinetic parameters (Figure 3). This suggests that the CC mice are a suitable model for studying the lipid-drug crosstalk in human metabolic disorders.

Measuring toxicity and adverse effects across a large population of genetically distinct individuals is costly. Our analysis offers a potential strategy for selecting a subset of CC lines (designed for a specific drug) that can be used in pharmacogenomic studies. In particular, a given drug corresponds to a group of eQTL targets that play a role in deposition of the drug (e.g., six DMEs in the irinotecan pathway; Figure 1B). The transcription profiles of CC lines across the genes in this group may point to a small subset of non-redundant CC lines (e.g., for the irinotecan pathway, a single CC line from each of the eight groups in Figure S7). In subsequent pharmacogenetic studies of this drug, only this subset of CC lines is analyzed. This strategy is in contrast to the standard approach of choosing the lines according to genotyping. The two selection methods complement each other, but the expression-based approach may reveal, in addition, the functional effects of candidate genomic loci.

Although it is not yet known whether the selection of CC mice based on proximal eQTLs reflects drug toxicity and adverse effects *in vivo*, it provides a plausible strategy of a rationalized selection of strains that (i) relies on variation in transcript isoforms in addition to the overall expression of genes; (ii) is tailored to each specific drug; and (iii) allows selection of additional CC mice not included in the reference database (e.g., using the inferred trait levels based on the optimized parameters in Equation 1). Future studies should determine whether an expression-based CC selection strategy is predictive of drug toxicity and adverse outcome. A more complete approach



will require integration with additional genomic data, including proteomics, *trans*-acting polymorphic loci, and epigenetic data.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This work was supported by the European Commission [FP7/2007-2013, under grant agreement no. 262055 (ESGI)], by the Israeli Centers of Research Excellence (I-CORE): Center No. 41/11, by the Broad-ISF grant 1168/14, and by the Wellcome Trust grants 090532/Z/09/Z, 085906/Z/08/Z, 083573/Z/07/Z, and 075491/Z/04. Research in the IG lab was supported by the European Research Council (637885), and by the Israeli Science Foundation (ISF) grant 1643/13. YS was partially supported by the Edmond J. Safra Center for Bioinformatics at Tel Aviv University IG is a Faculty Fellow of the Edmond J. Safra Center

for Bioinformatics at Tel Aviv University and an Alon Fellow. RS was supported in part by the Raymond and Beverly Sackler Chair in Bioinformatics.

## ACKNOWLEDGMENTS

Sequencing and primary data quality control was performed at the Genomics Unit of the Centre for Genomic Regulation (CRG) in Barcelona, Spain. We thank Anna Menoyo, Anna Ferrer, Magda Montfort, Irene González, Manuela Hummel, and Sarah Bonnin for technical assistance, Avital Brodt and Roni Wilentzik for computational analyses, and Shirley Smith for scientific editing. We thank Tel-Aviv University for core funding and technical support.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2016.00172>

## REFERENCES

- Aylor, D. L., Valdar, W., Foulds-Mathes, W., Buus, R. J., Verdugo, R. A., Baric, R. S., et al. (2011). Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res.* 21, 1213–1222. doi: 10.1101/gr.111310.110
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Battle, A., Mostafavi, S., Zhu, X., Potash, J. B., Weissman, M. M., McCormick, C., et al. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* 24, 14–24. doi: 10.1101/gr.155192.113
- Bryois, J., Buil, A., Evans, D. M., Kemp, J. P., Montgomery, S. B., Conrad, D. F., et al. (2014). Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* 10:e1004461. doi: 10.1371/journal.pgen.1004461
- Buechler, C., and Weiss, T. S. (2011). Does hepatic steatosis affect drug metabolizing enzymes in the liver? *Curr. Drug Metab.* 12, 24–34. doi: 10.2174/138920011794520035
- Chesler, E. J. (2014). Out of the bottleneck: the Diversity Outcross and Collaborative Cross mouse populations in behavioral genetics research. *Mamm. Genome* 25, 3–11. doi: 10.1007/s00335-013-9492-9
- Consortium, C. C. (2012). The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics* 190, 389–401. doi: 10.1534/genetics.111.132639
- Cook, D. N., Wang, S., Wang, Y., Howles, G. P., Whitehead, G. S., Berman, K. G., et al. (2004). Genetic regulation of endotoxin-induced airway disease. *Genomics* 83, 961–969. doi: 10.1016/j.ygeno.2003.12.008
- Durrant, C., Tayem, H., Yalcin, B., Cleak, J., Goodstadt, L., de Villena, F. P., et al. (2011). Collaborative Cross mice and their power to map host susceptibility to *Aspergillus fumigatus* infection. *Genome Res.* 21, 1239–1248. doi: 10.1101/gr.118786.110
- Evans, R. M., and Mangelsdorf, D. J. (2014). Nuclear receptors, RXR, and the big bang. *Cell* 157, 255–266. doi: 10.1016/j.cell.2014.03.012
- Ferris, M. T., Aylor, D. L., Bottomly, D., Whitmore, A. C., Aicher, L. D., Bell, T. A., et al. (2013). Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. *PLoS Pathog.* 9:e1003196. doi: 10.1371/journal.ppat.1003196
- Frick, A., Suzuki, O., Butz, N., Chan, E., and Wiltshire, T. (2013). *In vitro* and *in vivo* mouse models for pharmacogenetic studies. *Methods Mol. Biol.* 1015, 263–278. doi: 10.1007/978-1-62703-435-7\_17
- Gelinas, R., Chesler, E. J., Vasconcelos, D., Miller, D. R., Yuan, Y., Wang, K., et al. (2011). A genetic approach to the prediction of drug side effects: bleomycin induces concordant phenotypes in mice of the collaborative cross. *Pharmgenomics. Pers. Med.* 4, 35–45. doi: 10.2147/pgpm.s22475
- González-Porta, M., Calvo, M., Sammeth, M., and Guigó, R. (2012). Estimation of alternative splicing variability in human populations. *Genome Res.* 22, 528–538. doi: 10.1101/gr.121947.111
- Graham, J. B., Thomas, S., Swarts, J., McMillan, A. A., Ferris, M. T., Suthar, M. S., et al. (2015). Genetic diversity in the collaborative cross model recapitulates human West Nile virus disease outcomes. *MBio* 6:e00493-15. doi: 10.1128/mBio.00493-15
- Gralinski, L. E., Ferris, M. T., Aylor, D. L., Whitmore, A. C., Green, R., Frieman, M. B., et al. (2015). Genome wide identification of SARS-CoV susceptibility loci using the Collaborative Cross. *PLoS Genet.* 11:e1005504. doi: 10.1371/journal.pgen.1005504
- Guo, Y., Lu, P., Farrell, E., Zhang, X., Weller, P., Monshouwer, M., et al. (2007). *In silico* and *in vitro* pharmacogenetic analysis in mice. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17735–17740. doi: 10.1073/pnas.0700724104
- Harrill, A. H., and Rusyn, I. (2008). Systems biology and functional genomics approaches for the identification of cellular responses to drug toxicity. *Expert Opin. Drug Metab. Toxicol.* 4, 1379–1389. doi: 10.1517/17425255.4.11.1379
- Hitzemann, R., Reed, C., Malmanger, B., Lawler, M., Hitzemann, B., Cunningham, B., et al. (2004). On the integration of alcohol-related quantitative trait loci and gene expression analyses. *Alcohol. Clin. Exp. Res.* 28, 1437–1448. doi: 10.1097/01.ALC.0000139827.86749.DA
- Hsieh, T. Y., Shiu, T. Y., Huang, S. M., Lin, H. H., Lee, T. C., Chen, P. J., et al. (2007). Molecular pathogenesis of Gilbert's syndrome: decreased TATA-binding protein binding affinity of UGT1A1 gene promoter. *Pharmacogenet. Genomics* 17, 229–236. doi: 10.1097/FPC.0b013e328012d0da
- Iraqi, F. A., Churchill, G., and Mott, R. (2008). The Collaborative Cross, developing a resource for mammalian systems genetics: a status report of the Wellcome Trust cohort. *Mamm. Genome* 19, 379–381. doi: 10.1007/s00335-008-9113-1
- Kang, H. P., Morgan, A. A., Chen, R., Schadt, E. E., and Butte, A. J. (2012). “Coanalysis of GWAS with eQTLs reveals disease-tissue associations,” in *AMIA Joint Summits on Translational Science Proceedings. AMIA Summit on Translational Science* (San Francisco, CA), 35–41.
- Katz, D. A., Murray, B., Bhatthena, A., and Sahelijo, L. (2008). Defining drug disposition determinants: a pharmacogenetic-pharmacokinetic strategy. *Nat. Rev. Drug Discov.* 7, 293–305. doi: 10.1038/nrd2486
- Kelada, S. N., Aylor, D. L., Peck, B. C., Ryan, J. F., Tavarez, U., Buus, R. J., et al. (2012). Genetic analysis of hematological parameters in incipient lines of the collaborative cross. *G3 (Bethesda)* 2, 157–165. doi: 10.1534/g3.111.001776
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:r36. doi: 10.1186/gb-2013-14-4-r36
- Lake, A. D., Novak, P., Fisher, C. D., Jackson, J. P., Hardwick, R. N., Billheimer, D. D., et al. (2011). Analysis of global and absorption, distribution, metabolism, and elimination gene expression in the progressive stages of human nonalcoholic fatty liver disease. *Drug Metab. Dispos.* 39, 1954–1960. doi: 10.1124/dmd.111.040592
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco, CA: Holden-Day.
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323
- Loré, N. I., Iraqi, F. A., and Bragonzi, A. (2015). Host genetic diversity influences the severity of *Pseudomonas aeruginosa* pneumonia in the Collaborative Cross mice. *BMC Genet.* 16:106. doi: 10.1186/s12863-015-0260-6
- Mao, J. H., Langley, S. A., Huang, Y., Hang, M., Bouchard, K. E., Celniker, S. E., et al. (2015). Identification of genetic factors that modify motor performance and body weight using Collaborative Cross mice. *Sci. Rep.* 5:16247. doi: 10.1038/srep16247
- Marsh, S., and Hoskins, J. M. (2010). Irinotecan pharmacogenomics. *Pharmacogenomics* 11, 1003–1010. doi: 10.2217/pgs.10.95
- Meyer, U. A. (2000). Pharmacogenetics and adverse drug reactions. *Lancet* 356, 1667–1671. doi: 10.1016/S0140-6736(00)03167-6
- Montgomery, M. K., Hallahan, N. L., Brown, S. H., Liu, M., Mitchell, T. W., Cooney, G. J., et al. (2013). Mouse strain-dependent variation in obesity and glucose homeostasis in response to high-fat feeding. *Diabetologia* 56, 1129–1139. doi: 10.1007/s00125-013-2846-8
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C., and Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12649–12654. doi: 10.1073/pnas.230304397
- Nagar, S., and Blanchard, R. L. (2006). Pharmacogenetics of uridine diphosphoglucuronosyltransferase (UGT) 1A family members and its role in patient response to irinotecan. *Drug Metab. Rev.* 38, 393–409. doi: 10.1080/03602530600739835
- Naik, A., Belic, A., Zanger, U. M., and Rozman, D. (2013). Molecular interactions between NAFLD and xenobiotic metabolism. *Front. Genet.* 4:2. doi: 10.3389/fgene.2013.00002
- Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., et al. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* 37:e123. doi: 10.1093/nar/gkp596
- Penner, N., Woodward, C., and Prakash, C. (2012). “Drug metabolizing enzymes and biotransformation reactions,” in *ADME-Enabling Technologies in Drug Design and Development*, eds D. Zhang and S. Surapaneni (Hoboken, NJ: John Wiley & Sons, Inc.).
- Percival, C. J., Liberton, D. K., Pardo-Manuel de Villena, F., Spritz, R., Marcucio, R., and Hallgrímsson, B. (2016). Genetics of murine craniofacial morphology: diallel analysis of the eight founders of the Collaborative Cross. *J. Anat.* 228, 96–112. doi: 10.1111/joa.12382

- Phillippi, J., Xie, Y., Miller, D. R., Bell, T. A., Zhang, Z., Lenarcic, A. B., et al. (2014). Using the emerging Collaborative Cross to probe the immune system. *Genes Immun.* 15, 38–46. doi: 10.1038/gene.2013.59
- Roberts, A., Pardo-Manuel de Villena, F., Wang, W., McMillan, L., and Threadgill, D. W. (2007). The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. *Mamm. Genome* 18, 473–481. doi: 10.1007/s00335-007-9045-1
- Rosenkranz, R., Borodina, T., Lehrach, H., and Himmelbauer, H. (2008). Characterizing the mouse ES cell transcriptome with Illumina sequencing. *Genomics* 92, 187–194. doi: 10.1016/j.ygeno.2008.05.011
- Rost, S., Fregin, A., Ivaskevicius, V., Conzelmann, E., Hörtnagel, K., Pelz, H. J., et al. (2004). Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* 427, 537–541. doi: 10.1038/nature02214
- Rusyn, I., Gatti, D. M., Wiltshire, T., Kleeberger, S. R., and Threadgill, D. W. (2010). Toxicogenetics: population-based testing of drug and chemical safety in mouse models. *Pharmacogenomics* 11, 1127–1136. doi: 10.2217/pgs.10.100
- Svenson, K. L., Gatti, D. M., Valdar, W., Welsh, C. E., Cheng, R., Chesler, E. J., et al. (2012). High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics* 190, 437–447. doi: 10.1534/genetics.111.132597
- Thaisz, J., Tsaih, S. W., Feng, M., Philip, V. M., Zhang, Y., Yanas, L., et al. (2012). Genetic analysis of albuminuria in collaborative cross and multiple mouse intercross populations. *Am. J. Physiol. Renal Physiol.* 303, F972–F981. doi: 10.1152/ajprenal.00690.2011
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* 14, 178–192. doi: 10.1093/bib/bbs017
- Valdar, W., Flint, J., and Mott, R. (2006). Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* 172, 1783–1797. doi: 10.1534/genetics.104.039313
- Vered, K., Durrant, C., Mott, R., and Iraqi, F. A. (2014). Susceptibility to *Klebsiella pneumoniae* infection in collaborative cross mice is a complex trait controlled by at least three loci acting at different time points. *BMC Genomics* 15:865. doi: 10.1186/1471-2164-15-865
- Wittkopp, P. J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13, 59–69. doi: 10.1038/nrg3095
- Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8, 206–216. doi: 10.1038/nrg2063
- Yang, H., Wang, J. R., Didion, J. P., Buus, R. J., Bell, T. A., Welsh, C. E., et al. (2011). Subspecific origin and haplotype diversity in the laboratory mouse. *Nat. Genet.* 43, 648–655. doi: 10.1038/ng.847
- Yoo, H. S., Bradford, B. U., Kosyk, O., Uehara, T., Shymonyak, S., Collins, L. B., et al. (2015). Comparative analysis of the relationship between trichloroethylene metabolism and tissue-specific toxicity among inbred mouse strains: kidney effects. *J. Toxicol. Environ. Health A* 78, 32–49. doi: 10.1080/15287394.2015.958418
- Zhong, H., Beaulaurier, J., Lum, P. Y., Molony, C., Yang, X., Macneil, D. J., et al. (2010). Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet.* 6:e1000932. doi: 10.1371/journal.pgen.1000932

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Nachshon, Abu-Toamih Atamni, Steuerman, Sheikh-Hamed, Dorman, Mott, Dohm, Lehrach, Sultan, Shamir, Sauer, Himmelbauer, Iraqi and Gat-Viks. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.