Systematic Genetic and Genomic Analysis of Cytochrome P450 Enzyme Activities in Human Liver

Xia Yang[‡1*], Bin Zhang[‡1], Cliona Molony[2], Eugene Chudin[2], Ke Hao[2], Jun Zhu[1], Andrea Gaedigk[3], Christine Suver[1], Hua Zhong[1], J. Steven Leeder[3], F. Peter Guengerich[4], Stephen C. Strom[5], Erin Schuetz[6], Thomas H. Rushmore[7], Roger G. Ulrich[8], J. Greg Slatter[9], Eric E. Schadt[10], Andrew Kasarskis[1*], Pek Yee Lum[11*]

[1]Rosetta Inpharmatics, LLC, a wholly owned subsidiary of Merck & Co., Inc., Seattle, WA 98109, USA. Current address: Sage Bionetworks, Seattle, WA98109, USA. [2]Rosetta Inpharmatics, LLC, a wholly owned subsidiary of Merck & Co., Inc., Seattle, WA 98109, USA. [3]Clinial Pharmacology and Therapeutics, Children's Mercy Hospital, Kansas City, MO 64108, USA. [4]Department of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA. [5]Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15219, USA. [6]Department of Pharmaceutical Sciences, Saint Jude Children's Research Hospital, Memphis, TN 38105, USA. [7]Drug Metabolism, Merck & Co., Inc., West Point, PA 19486, USA. [8]Calistoga Pharmaceuticals, Inc. Seattle, WA 98121, USA. [9]Pharmacokinetics and Drug Metabolism, Amgen Inc., Seattle, WA 98119, USA. [10]Rosetta Inpharmatics, LLC, a wholly owned subsidiary of Merck & Co., Inc., Seattle, WA 98109, USA. Current address: Pacific Biosciences, Menlo Park, CA 94025, USA. [11]Rosetta Inpharmatics, LLC, a wholly owned subsidiary of Merck & Co., Inc., Seattle, WA 98109, USA. Current address: iPierian, Inc. South San Francisco, CA 94080, USA.

[‡]These authors contributed equally.

*Corresponding authors:

Pek Yee Lum, Ph.D.

Rosetta Inpharmatics, LLC

401 Terry Ave N

Seattle, WA 98109

Email: pek_lum@ipierian.com

Andrew Kasarskis, Ph.D.

Rosetta Inpharmatics, LLC

401 Terry Ave N

Seattle, WA 98109

Email: andrew@sagebase.org

Xia Yang, Ph.D.

Rosetta Inpharmatics, LLC

401 Terry Ave N

Seattle, WA 98109

Email: xia.yang@sagebase.org

Running head: Genetics and Genomics of Human Liver P450s

Supplementary Figure Legends

Figure S1. QQ plots and histograms of P450 activity traits. A) QQ-norm plots. Red line represents normal distribution and black dots represent data distribution. Strong deviation of the data from the red line indicates non-normal distribution. B) Histograms of P450 activity measurements. Peaks shift towards the low activity end for all activity traits.

Figure S2. Correlation plots between age and P450 enzyme activities. Only significant correlations at $p<0.05$ are shown. For each plot, age is on the x-axis, P450 activity measurement is on the y-axis, and the specific gender group is indicated in the plot title (females as "in female", males as "in male", and both genders if not specified). The correlations are all in positive direction, i.e., the more advanced the age, the higher the P450 activity.

Figure S3. Plots of P450 enzyme activities grouped by age and gender. Owning to the non-normal distribution of P450 enzyme activity measurements, log transformation of these traits was used to for better visualization. Each P450 trait was grouped by males (M), Females (F), and various age groups: <10 year old (<10yr), 10-19 year old (10-19yr), 20-39 year old (20-40yr), 40-59 year old (40-50yr), and >60 year old (>60yr).

Figure S4. Representative correlation plots between P450 enzyme activity and gene expression. Expression is on the x-axis and activity is on the y-axis and all correlations are in positive direction.

Figure S5. Gene network connectivity is predictive of gene relevance to P450 activity traits. A) Gene relevance to CYP2C8 enzyme activity versus connectivity measure k.in

for genes in the turquoise module. B) Gene relevance to CYP2C8 enzyme activity versus connectivity measure k.in for genes in the pink module. C) Gene relevance to CYP2C8 enzyme activity versus connectivity measure k.in for genes in the brown module. D) Gene relevance to CYP2C8 enzyme activity versus connectivity measure k.in for genes in the red module. The within module connectivity k.in is on the x-axis and the gene relevance (the absolute value of correlation between a gene and P450 activity trait) is on the y-axis. In each case, the network connectivity of genes within each P450-correlated module is positively correlated with the gene relevance to P450 activity.

Figure S6. Correlation between three CYP2D6 aSNP/eSNPs with CYP2D6 activity measured with log10 urinary ratio of dextromethorphan (DM) to its metabolite dextrorphan (DT; log10_DMDT).

List of Supplementary Tables (full tables supplied as separate excel files)

Table S1. Characteristics of P450 enzyme activity measurements.

Table S2. Pair-wise Spearman correlation between P450 enzyme activity traits.

Table S3. Pair-wise Pearson correlation between P450 gene expression traits.

Table S4. Spearman correlation between known P450 regulators and levels of P450 enzyme activity and gene expression.

Table S5. Spearman correlation between P450 activity measurements and the transcriptionally active genes selected from the 44k microarray.

Table S6. Gene membership and characteristics of the HLC co-expression network.

Table S7. Enriched functional categories of the HLC co-expression modules.

Table S8. Correlation between co-expression modules and P450 activity traits.

Table S9. Correlation between co-expression modules and the expression of P450 genes.

Table S10. *cis*-eSNPs of P450 genes at FDR<10%.

Table S11. aSNPs of P450 activities at FDR<10%.

Table S12. Overlap between aSNPs of P450 activities at FDR<10% and *cis*-eSNPs of P450 genes at $p<0.05$.

Table S13. Overlap between *cis*-eSNPs of P450 genes at FDR<10% and aSNPs of P450 activities at $p<0.05$.

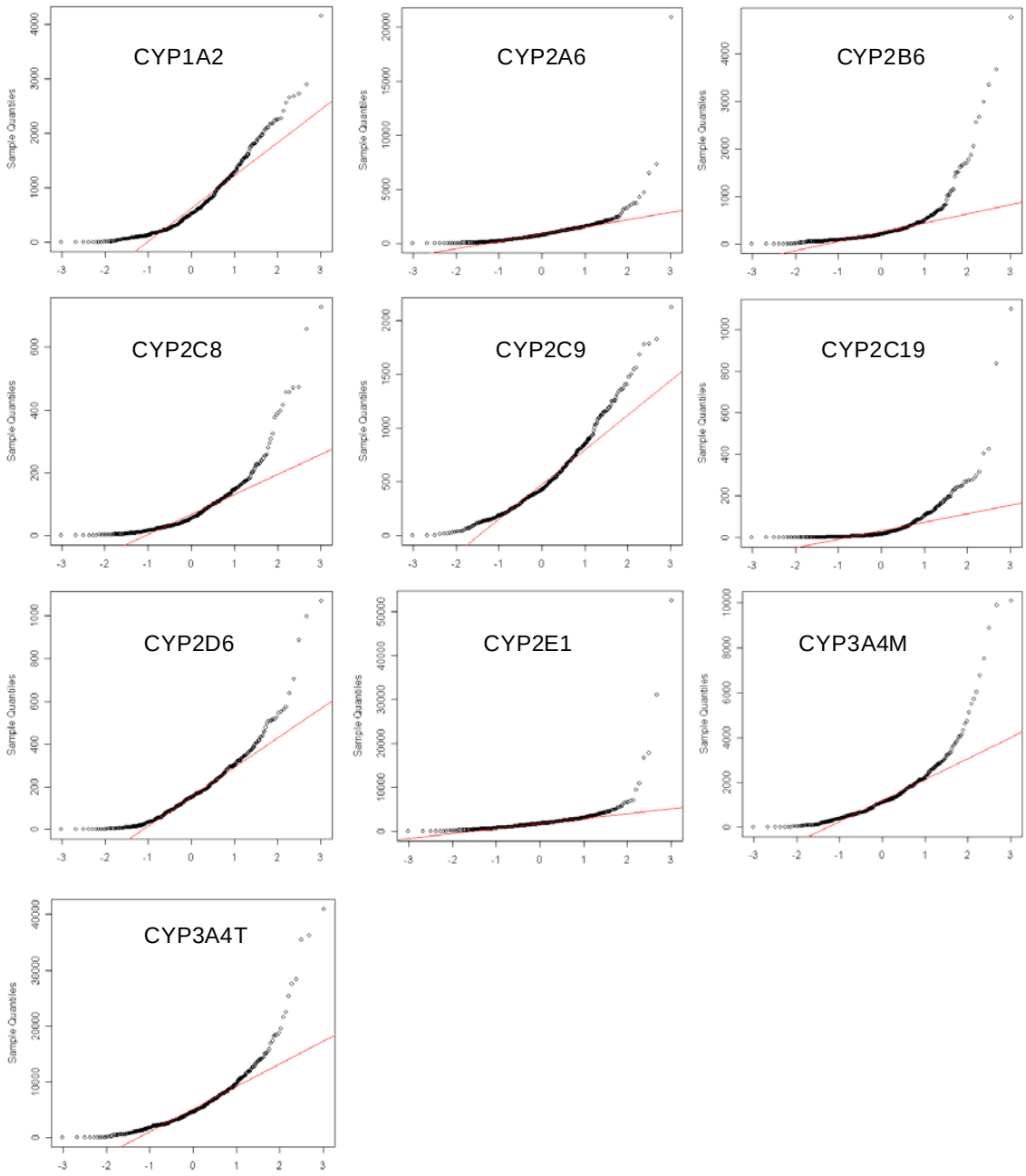
Table S14. Strong LD between three CYP2D6 aSNP/eSNPs and a functional deletion variant CYP2D6 *4.

Table S15. Overlap of CYP450 subenetwork shown in Figure 5A with mouse and rat genes responsive to ligands of AHR, CAR, PXR, and all three receptors (Slatter et al., 2006).

Supplementary Figures

Figure S1. QQ plots and histograms of P450 activity traits. A) QQ plots. B) Histograms.

A



B

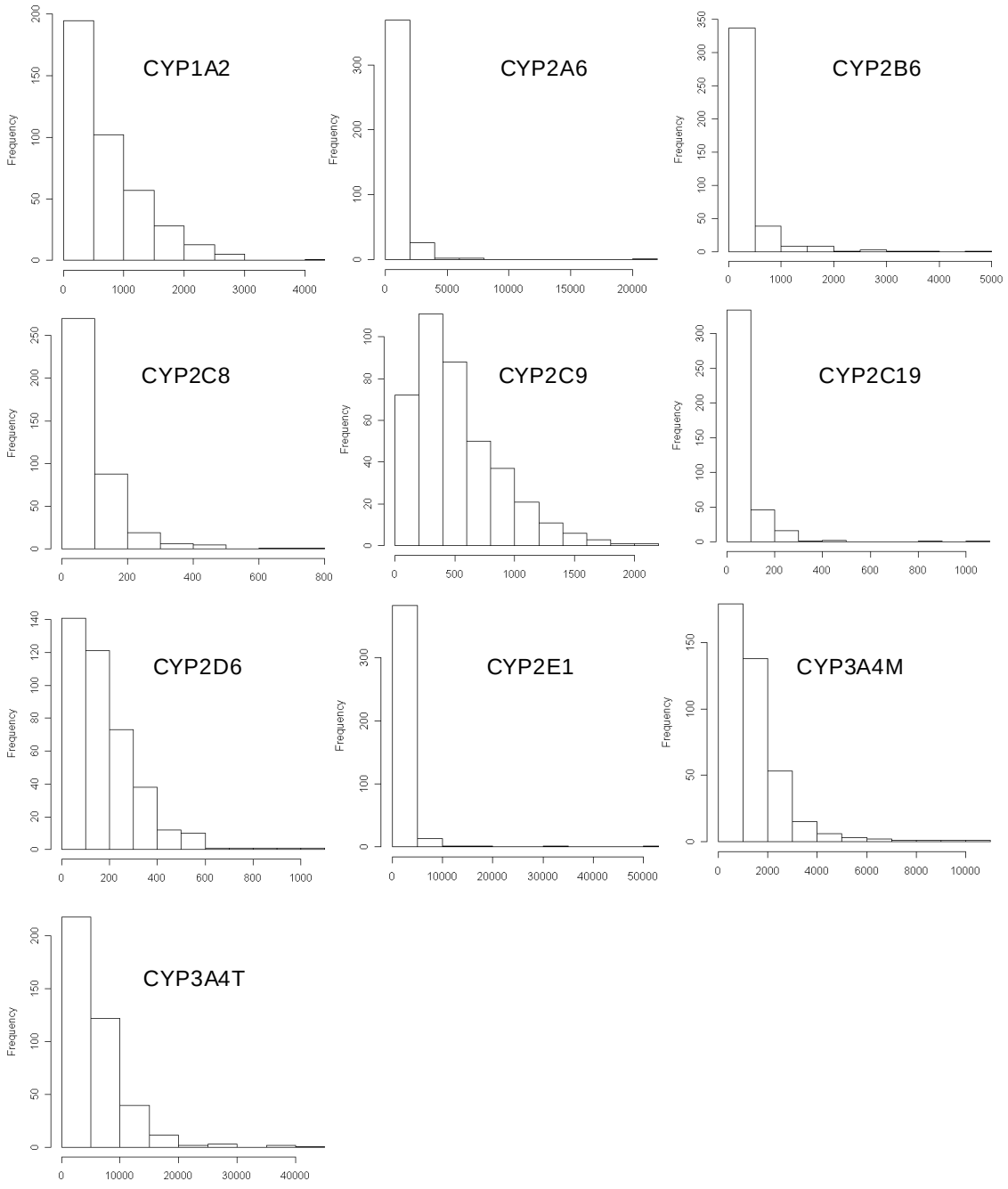

CYP1A2
CYP2A6
CYP2B6
CYP2C8
CYP2C9
CYP2C19
CYP2D6
CYP2E1
CYP3A4M
CYP3A4T
Frequency

Figure S2. Correlation plots between age and P450 enzyme activities.

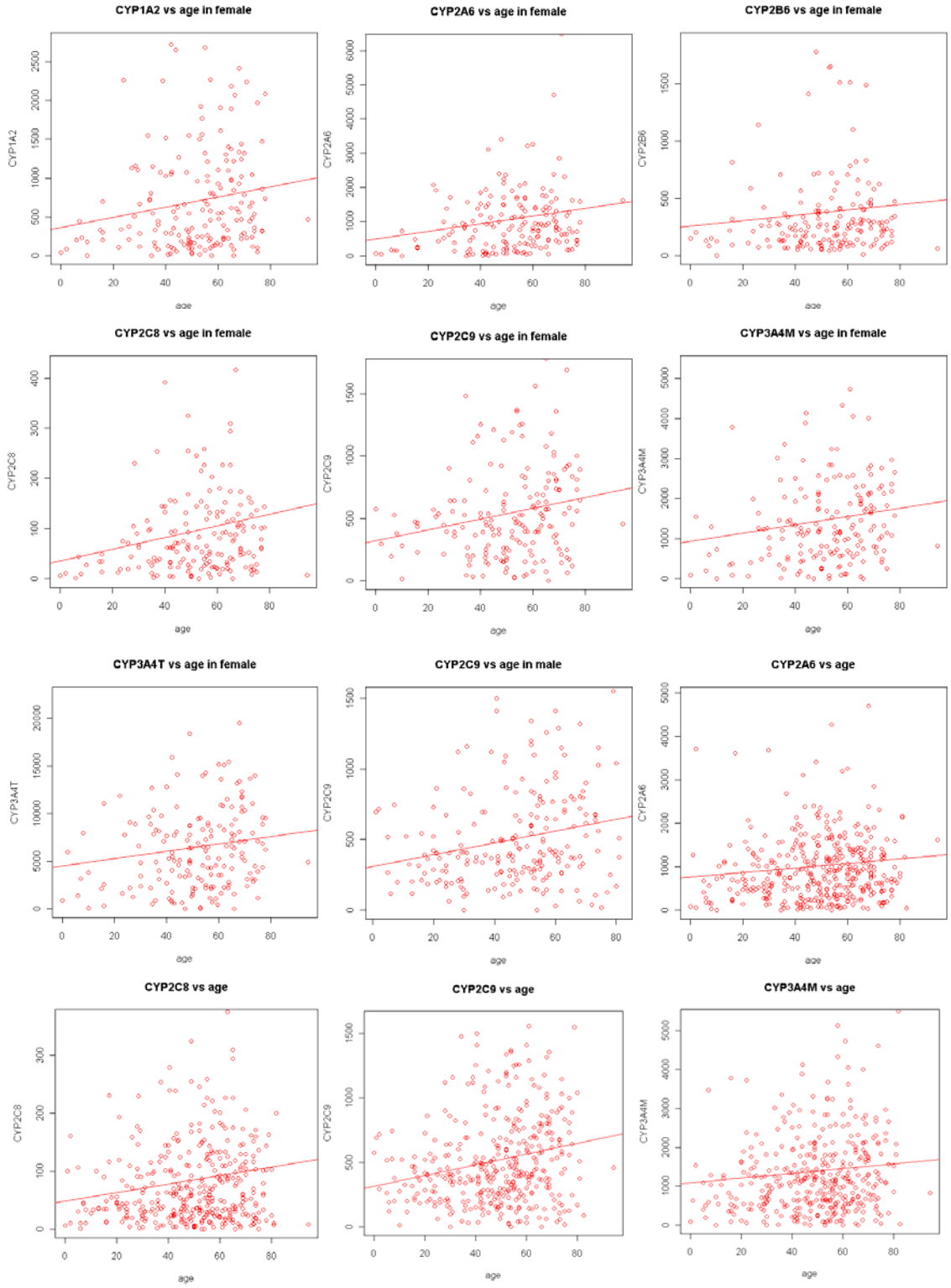
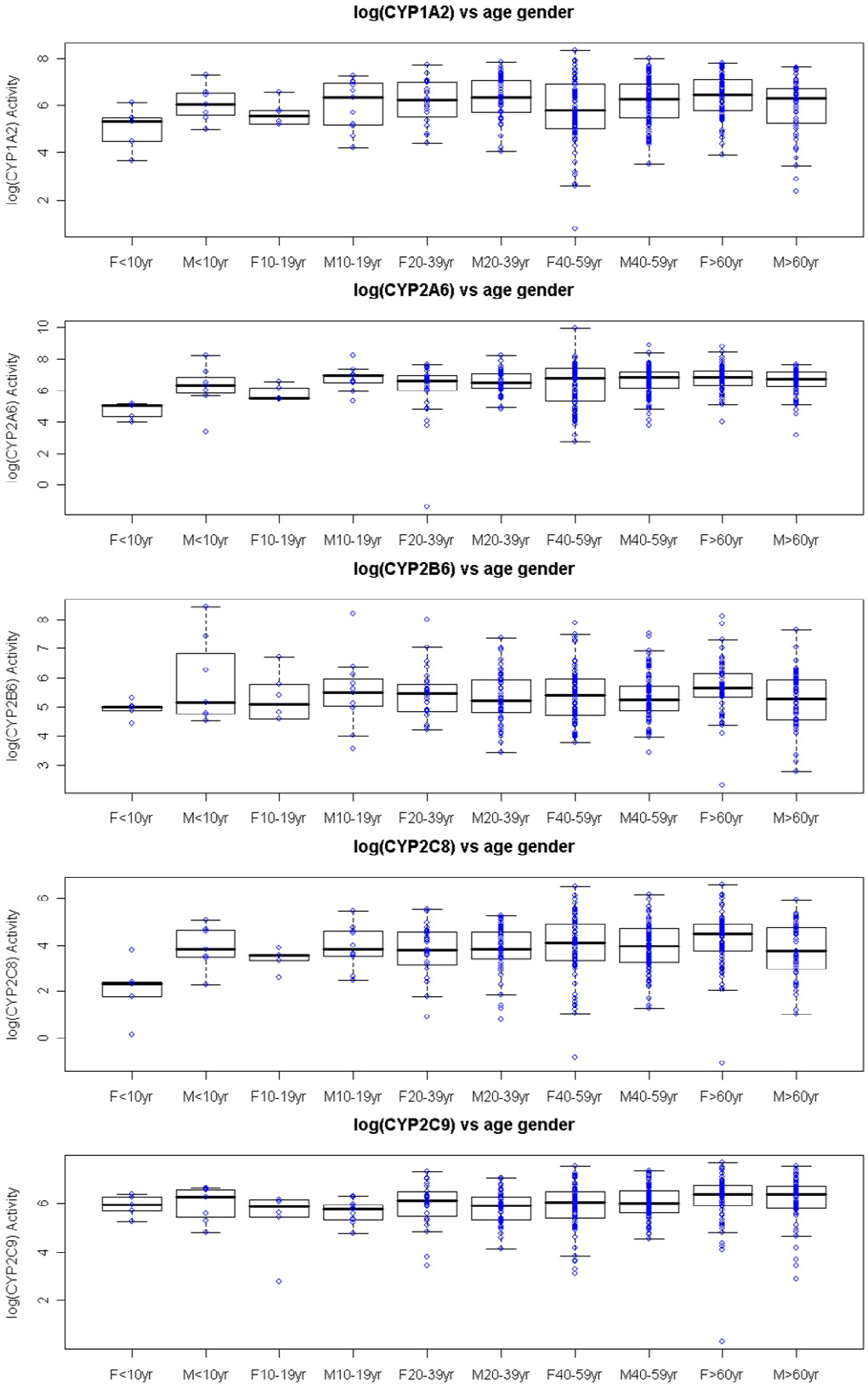
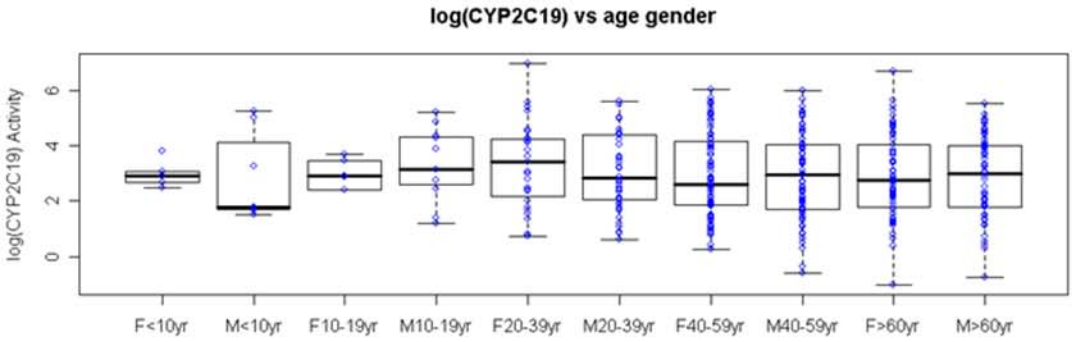
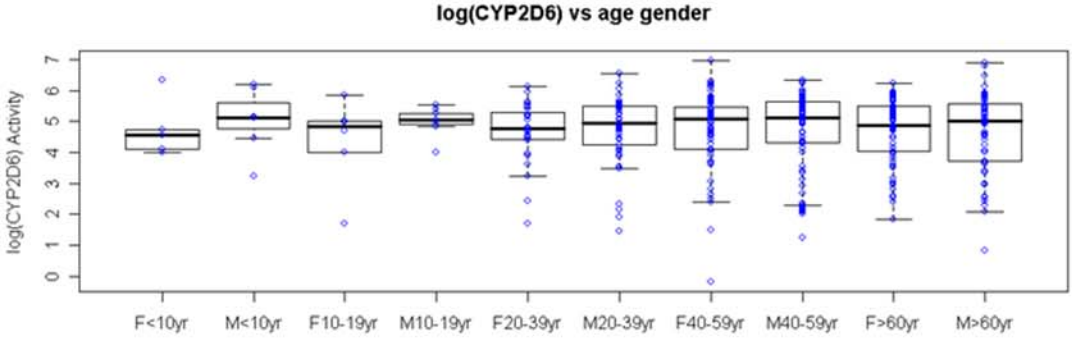
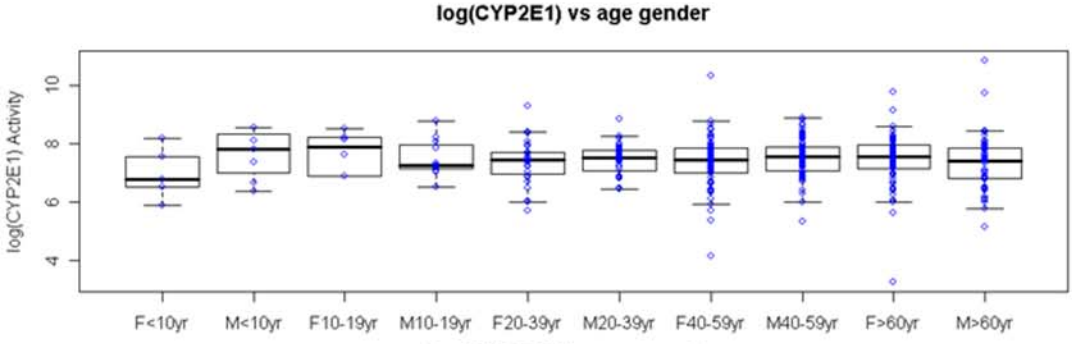


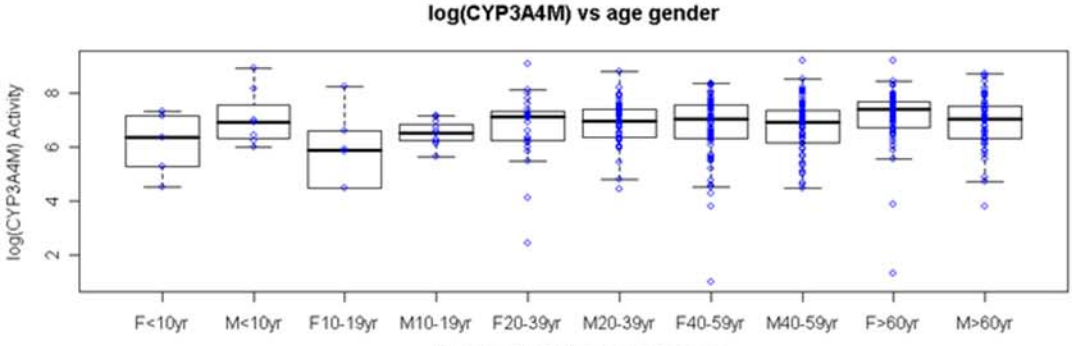
Figure S3. Plots of P450 enzyme activities grouped by age and gender.




log(CYP2C19) vs age gender
log(CYP2C19) Activity
0
2
4
6
F<10yr
M<10yr
F10-19yr
M10-19yr
F20-39yr
M20-39yr
F40-59yr
M40-59yr
F>60yr
M>60yr


log(CYP2D6) vs age gender
log(CYP2D6) Activity
0
1
2
3
4
5
6
7
F<10yr
M<10yr
F10-19yr
M10-19yr
F20-39yr
M20-39yr
F40-59yr
M40-59yr
F>60yr
M>60yr


log(CYP2E1) vs age gender
log(CYP2E1) Activity
4
6
8
10
F<10yr
M<10yr
F10-19yr
M10-19yr
F20-39yr
M20-39yr
F40-59yr
M40-59yr
F>60yr
M>60yr


log(CYP3A4M) vs age gender
log(CYP3A4M) Activity
2
4
6
8
F<10yr
M<10yr
F10-19yr
M10-19yr
F20-39yr
M20-39yr
F40-59yr
M40-59yr
F>60yr
M>60yr

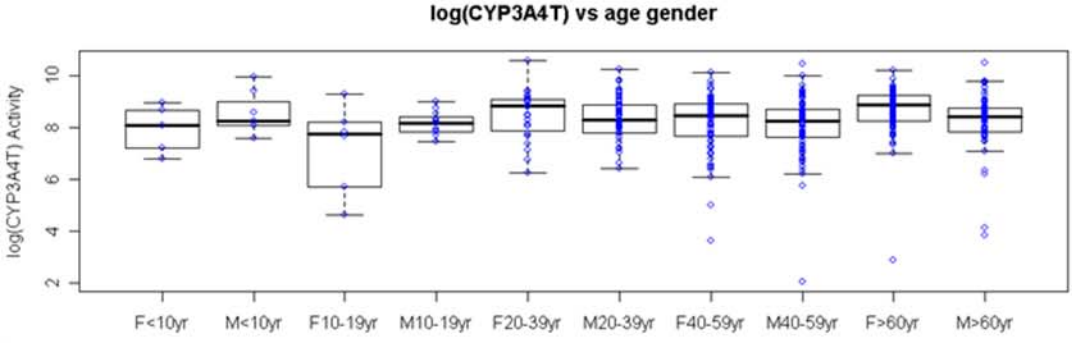

log(CYP3A4T) vs age gender
log(CYP3A4T) Activity
2
4
6
8
10
F<10yr
M<10yr
F10-19yr
M10-19yr
F20-39yr
M20-39yr
F40-59yr
M40-59yr
F>60yr
M>60yr

Figure S4. Representative correlation plots between P450 enzyme activity and gene expression.

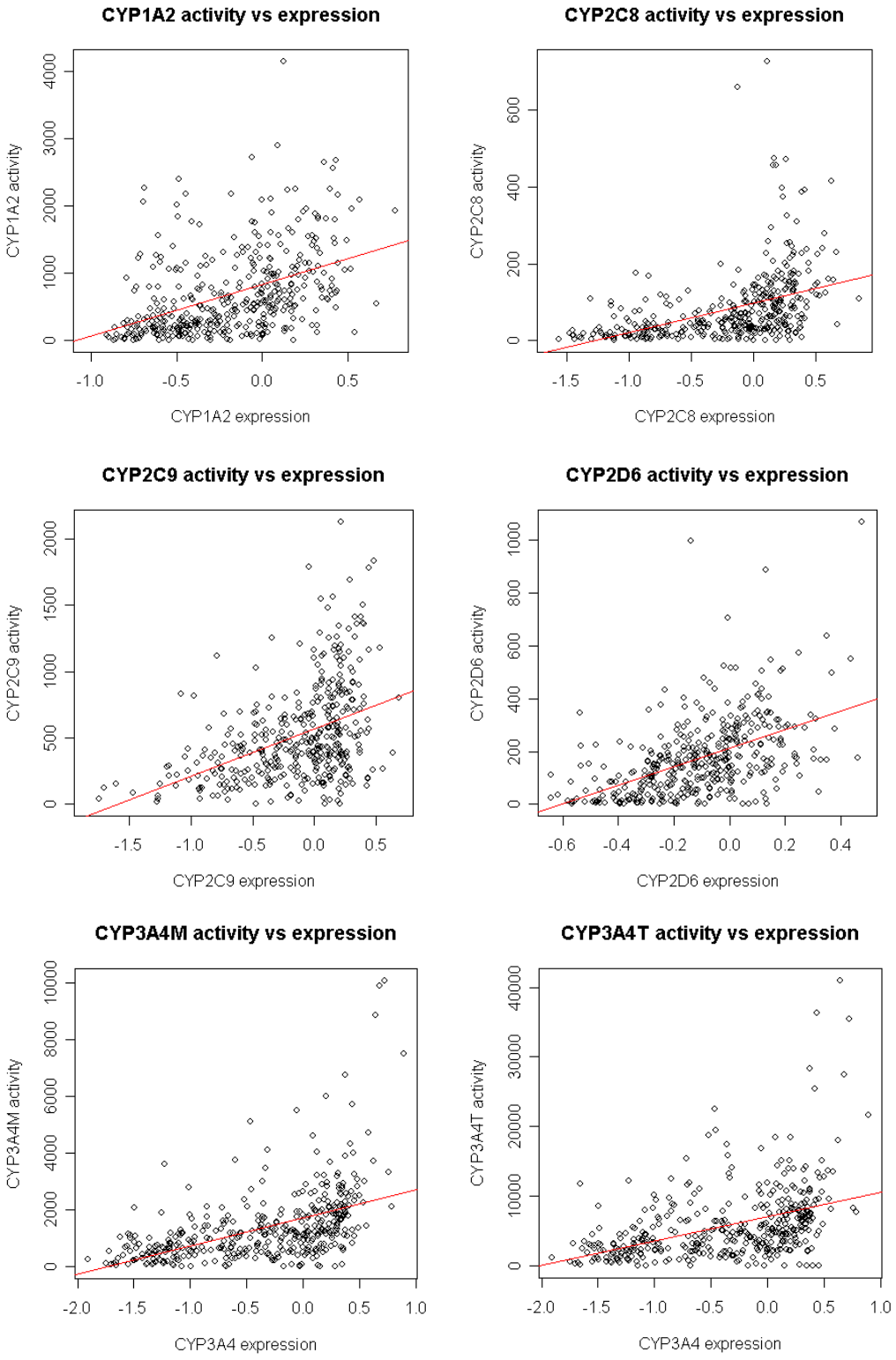


Figure S5. Gene network connectivity is predictive of gene relevance to P450 activity trait.

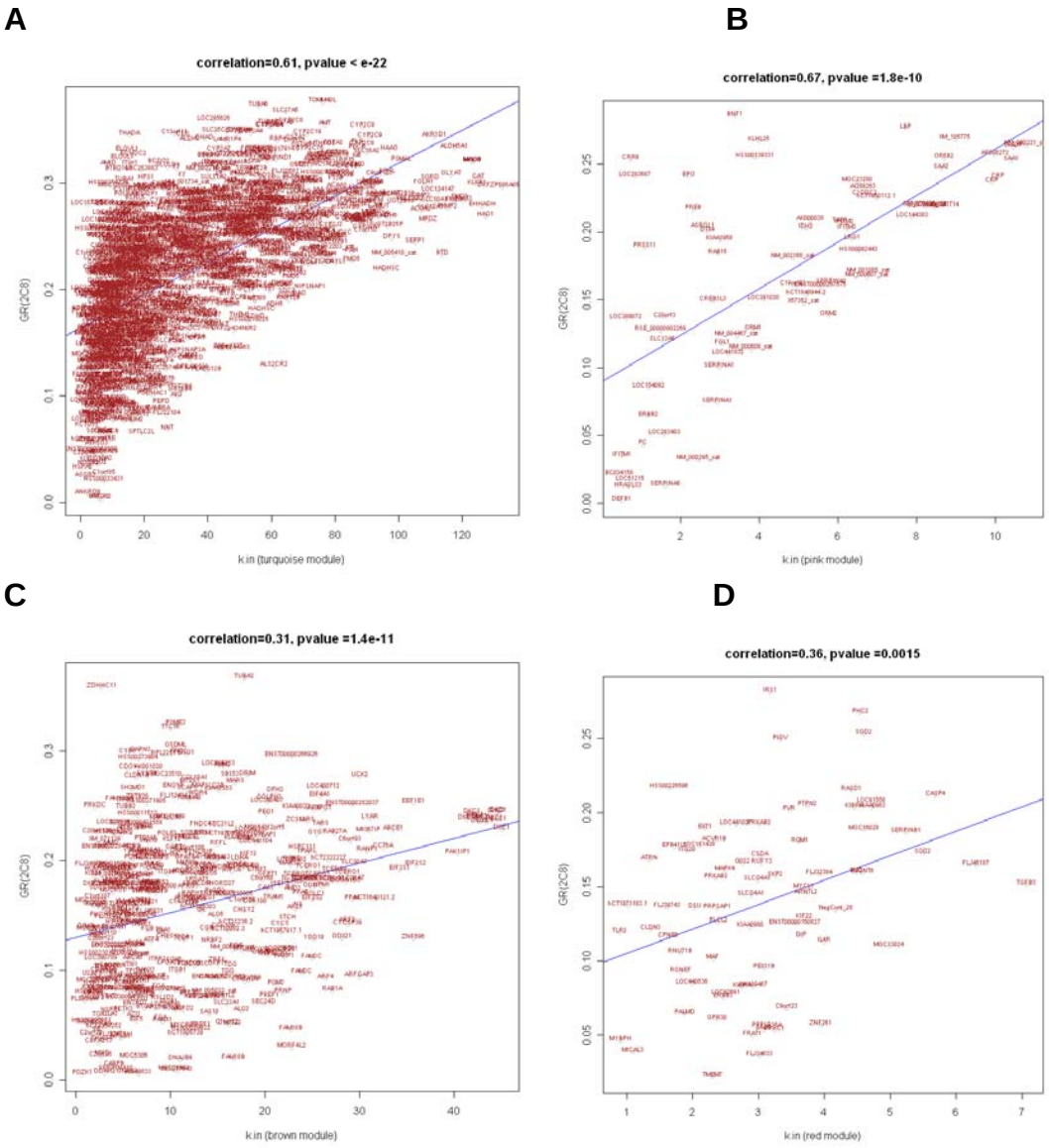


Figure S6. Correlation between three CYP2D6 aSNP/eSNPs with CYP2D6 activity measured with log10 urinary ratio of dextromethorphan (DM) to its metabolite dextrorphan (DT; log10_DMDT).