Published in final edited form as:

Pharmacogenet Genomics. 2009 August ; 19(8): 577–587. doi:10.1097/FPC.0b013e32832c15c6.

Catechol O-Methyltransferase Pharmacogenomics: Human Liver Genotype-Phenotype Correlation and Proximal Promoter Studies

Jianping Zhang^{a,*}, Yuan Ji^a, Irene Moon^a, Linda L. Pellegymounter^a, O. Ezequiel Salavaggione^a, Yanhong Wu^b, Gregory D. Jenkins^c, Anthony J. Batzler^c, Daniel J. Schaid^c, and Richard M. Weinshilboum^a

^a Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905 USA

^b Molecular Genetics Research Laboratory, Mayo Clinic, Rochester, MN 55905 USA

^c Division of Biostatistics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905 USA

* Department of Pharmaceutical Sciences, University at Buffalo. The State University of New York, 527 Cooke Hall, Buffalo, NY

Abstract

Objectives—Catechol O-methyltransferase (COMT) is expressed as both soluble (S) and membrane-bound (MB) isoforms, with S-COMT predominantly expressed in liver. A common nonsynonymous SNP, 472G>A (108/158Val>Met, S/MB), has been associated with variation in levels of COMT enzyme activity and thermal stability. We set out to test the hypothesis that additional *COMT* polymorphisms might also be associated with phenotypic variation.

Methods—We phenotyped 268 liver biopsy samples for S-COMT activity and thermal stability, resequenced a portion of the gene that had not previously been resequenced, and genotyped DNA from these same samples for 16 *COMT* polymorphisms.

Results—There was a significant association between the two COMT phenotypes and genotype at the codon 108 SNP. A haplotype-based approach was then used to assess the possible association of other polymorphisms with phenotype. Specifically, codon 108 SNP explained 20.4% of variance in activity ($P < 10^{-6}$), and 59% of variance in thermal stability ($P < 10^{-6}$). Haplotypes that included SNPs at cDNA nucleotides 408 and 472 explained additional variance in enzyme activity (up to 24.4%), and the addition to the haplotype of a SNP at intron 2 (51) explained a total of 27.5% of the variance. However, no SNPs beyond that at nucleotide 472G>A polymorphism were associated with variation in thermal stability. We also observed a three-fold variation in the ability of reporter gene constructs for “proximal promoter” haplotypes to drive transcription.

Conclusions—The common *COMT* 108Val>Met polymorphism is associated with human liver S-COMT activity and thermal stability, but additional *COMT* SNPs also contribute to variation in activity.

Please address all correspondence to: Dr. Richard Weinshilboum, Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, Telephone: (507) 284-2246; Fax: (507) 538-6670, Weinshilboum.richard@mayo.edu.

The new *COMT* gene resequencing data described in this manuscript have been deposited in the National Institutes of Health database PharmGKB with submission ID: PS206659.

Keywords

catechol O-methyltransferase; COMT; genetic polymorphism; SNP; haplotype; thermal stability

Introduction

Methyl conjugation catalyzed by COMT (E.C.2.1.1.6) plays an important role in the metabolism of catecholamine neurotransmitters, catecholestrogens and catechol drugs [1]. The *COMT* gene encodes both soluble (S) cytoplasmic and membrane-bound (MB) forms of the enzyme [2,3]. MB-COMT differs from S-COMT by the presence of an additional 50 hydrophobic amino acids at its N-terminus as a result of the existence of two translation initiation codons, both located in exon 3, as well as two promoters – a “proximal” promoter for S-COMT which is located in intron 2 and a “distal” promoter for MB-COMT in the 5'-FR of the gene (Figure 1A) [2]. With the exception of the central nervous system and lymphocytes, S-COMT is thought to be the major form of the enzyme expressed in most human tissues, including liver [2–4].

Over 30 years ago, we reported that level of S-COMT enzyme activity was genetically polymorphic in human tissues, and – subsequently – that low level of activity was associated with the trait of thermolabile enzyme [5–8]. Later, a common nonsynonymous SNP in *COMT*, 472G>A, that results in a 108Val>Met alteration in the S-COMT amino acid sequence (corresponding to codon 158 of MB-COMT) was reported to be associated with variation in levels of both COMT thermal stability and activity [9]. Val108 was associated with high levels of enzyme activity and protein as well as elevated thermal stability, while Met108 was associated with decreases in all three phenotypes [9,10]. Because of the important role of COMT in the metabolism of catecholamine neurotransmitters and of pro-carcinogenic catecholestrogens, *COMT* has been the focus for a large number of molecular epidemiology studies – especially for breast cancer (over 20 case-control associations studies since the codon 108/158 polymorphism was reported) [11–14] and schizophrenia (over 200 studies) [15–18].

Recently, the portion of the gene that encodes S-COMT was resequenced using 60 DNA samples each from Caucasian-American (CA) and African-American (AA) subjects. Twenty-four polymorphisms were observed in the 240 alleles analyzed [10]. The availability of that detailed gene sequence information makes it possible to test the hypothesis that sequence variation within the human *COMT* gene beyond the codon 108 polymorphism might also influence phenotype. Therefore, in the present study, we performed a phenotype-genotype correlation analysis for S-COMT in an important drug-metabolizing organ, the human liver. Because of gender-dependent variation in level of hepatic COMT activity [19], all of the samples studied were obtained from women. We also extended the resequencing of *COMT* to include exon 1 and the 5'-FR to make it possible to interrogate a wider range of sequence variation across the gene.

Specifically, 268 hepatic surgical biopsy samples were phenotyped, followed by genotyping for 12 “tagging” SNPs and 4 polymorphisms selected on the basis of previous reports of their possible functional importance [20,21]. We found that at least two additional SNPs beyond the codon 108 polymorphism were associated with variation in level of activity. We also created and studied luciferase reporter gene constructs for common, naturally occurring haplotypes present in the proximal promoter and observed striking variation in the ability of these constructs to drive transcription in HepG2 cells, but those results were not systematically related to phenotype in the tissue biopsy samples. These studies extend our

understanding of the relationship between variation in *COMT* sequence and hepatic phenotype for this important phase II enzyme.

Materials and Methods

Human DNA and tissue samples

Sixty DNA samples each from the Coriell Institute Cell Repository (Camden, New Jersey) CA and AA sample sets (HD100AA and HD100CAU) were used to resequence the 5'-FR and exon 1 of *COMT*. These DNA samples were deposited in the Coriell Institute by the National Institute of General Medical Sciences and have been widely used to perform gene resequencing studies [22–24]. The subjects who donated the samples provided written consent for their DNA to be used for research purposes, and the samples were anonymized.

Human hepatic surgical biopsy samples were obtained from 268 CA women who had clinically-indicated surgery at the Mayo Clinic – predominantly for the diagnosis and/or treatment of metastatic carcinoma. These samples were anonymized, and only information with regard to diagnosis, gender, race and age was provided. Hepatic tissue uninvolved by tumor was used to perform these studies. The Mayo Clinic Institutional Review Board (IRB) reviewed and approved all of these experiments.

COMT gene resequencing

Approximately 1.3 kb of DNA sequence upstream from *COMT* exon 1, exon 1 itself and approximately 750 bp of the 5'-end of intron 1 were amplified as described previously [10]. PCR primer sequences used in the resequencing studies can be obtained upon request. DNA sequencing was performed using dye terminator chemistry, and sequence chromatograms were analyzed using Polyphred 3.0 [25] and Mutation Surveyer 2.41 [26].

Human liver cytosol *COMT* enzyme assay

Hepatic cytosol was prepared and isolated as described previously [8], and *COMT* enzyme activity was measured with a radiochemical enzymatic assay that uses 3,4-dihydroxybenzoic acid and ¹⁴C-methyl-S-adenosyl-L-methionine (¹⁴C-methyl AdoMet) as substrates [27]. *COMT* thermal stability was determined as a heated/control (H/C) ratio by assaying enzyme activity before and after incubation at 49°C for 15 min [6].

COMT genotyping

Sixteen polymorphisms were selected for inclusion in the genotyping studies. Twelve “tagging” polymorphisms were selected by using either the linkage disequilibrium-tag (LD-tag) selection method of Carlson et al. [28] or the haplotype-tagging (ht-tag) method described by Stram et al. [29]. LD tag SNPs were required to have a minimum frequency of 5% and an 80% correlation within “bins”. Ht-tag SNPs were required to have a minimum frequency of 2%, a haplotype frequency of 1%, and an R² value of 0.9. Four additional SNPs were selected for inclusion in the genotyping studies on the basis of their possible functional implications. Specifically, the codon 108 nucleotide 472G>A nonsynonymous polymorphism, a nucleotide 1140G>A SNP in intron 2, a synonymous nucleotide 408C>G SNP in exon 4 that had previously been reported to be associated with variation in pain perception [21], and an insertion/deletion (indel) at nucleotide 820 in the 3'-UTR of the gene were added to the tag SNPs. The locations of polymorphisms selected for use in the genotyping studies are depicted graphically in Fig. 1B. Most of the polymorphisms were genotyped using the SNPstream platform, an assay system that utilizes single-base extension (Beckman Coulter, Fullerton, CA) [30,31]. Primers for those assays were designed using web-based software (<http://www.autoprimer.com>, Beckman Coulter Inc. Fullerton, CA). Pyrosequencing was used to assay the indel at 3'-UTR nucleotide 820. The 5'-flanking

region (5'-FR) SNPs at nucleotides (−628) and (−485) were genotyped by dye-primer DNA sequencing, and the 5'-FR SNP at nucleotide (−1420) and the intron 2 SNP at nucleotide (935) were genotyped by performing restriction fragment length polymorphism assays with *HindIII* and *HbaI*, respectively, (New England Biolabs, Ipswich, MA) because those SNPs could not be assayed with the SNPstream platform.

Luciferase reporter gene studies

Luciferase reporter gene constructs were created for common haplotypes in the *COMT* proximal promoter (intron 2). Haplotypes were inferred both on the basis of data obtained by resequencing the Coriell DNA samples and by genotyping the human liver biopsy samples. The gene resequencing studies had identified 6 SNPs in intron 2 [10]. Those polymorphisms resulted in 9 “common” (frequency 2% or greater) haplotypes in AAs, 7 of which were also observed in CA subjects. Since the patients from whom the biopsies were obtained were CA, the 7 common haplotypes in CA subjects were used to create reporter gene constructs for the *COMT* proximal promoter. Specifically, the PCR was used to amplify a 1,560 bp amplicon that included all of intron 2, as well as 91 bp at the 3'-end of exon 2 and 180 bp from the 5'-end of exon 3, using as template DNA from samples having the haplotypes of interest. Amplicons from those reactions were ligated into pGL-3 basic (Promega, Madison, WI) upstream of the firefly luciferase ORF. A few of the constructs had to be created by site-directed mutagenesis using circular PCR. All constructs were sequenced in both directions to ensure that no PCR-induced errors in sequence had been introduced.

These constructs were then used to transfect HepG2 cells. Specifically, 2 µg of reporter gene construct DNA were transiently transfected into HepG2 cells, together with 200 ng pRL-TK DNA (Promega), using the Transfast™ reagent (Promega). The *Renilla* construct served as a control for transfection efficiency. The cells were harvested 48 h after transfection, and cell lysates were used to assay luciferase activity with a Dual-Luciferase Reporter Assay System (Promega). Luciferase activity was measured with a Turner BioSystems Model TD-20/20 Luminometer (Sunnyvale, CA), and was expressed as the ratio of firefly luciferase light units to *Renilla* luciferase light units.

COMT expression array analyses

Total RNA was extracted from 60 Coriell Institute “Human Variation Panel” Caucasian-American (CA) lymphoblastoid cell lines, the same cell lines from which the DNA that had been used in the *COMT* gene resequencing was extracted. Basal expression array analyses were performed for these cell lines using Affymetrix U133 Plus 2.0 GeneChips (Affymetrix, Santa Clara, CA) to identify the individual variation in *COMT* expression. The procedures used for sample preparation and expression array analyses have been described previously [32,33]. A linear model was then used to determine the association of the intron 2(51G>C) *COMT* SNP with variation in expression for the two *COMT* probes (208817_at and 208818_s_at) present on the U133 Plus 2.0 arrays.

Data analysis

Spearman rank correlation coefficients were used to evaluate the association of phenotypes with age and with other covariates. We used linear regression of phenotype on the number of copies of the least frequent allele to evaluate the association of each SNP genotype with enzyme activity and thermal stability, assuming an additive effect of SNP variants on level of the phenotype. That approach made it possible to determine the most significant marginal associations among all SNPs.

To screen for the strongest association of combinations of variants in haplotypes, we used a sequential procedure to determine the set of variants within haplotypes that were most

strongly associated with a phenotype. That procedure began with the single SNP most highly associated with a phenotype, followed by the addition of one SNP at a time to determine the most significant two-SNP haplotypes; with additional SNPs being added sequentially to create longer haplotypes as long as the incremental SNPs were associated with phenotype at a “liberal” level ($P = 0.10$). This process was then followed by a “backward” procedure to remove SNPs from haplotypes at a more stringent level ($P = 0.01$). The more stringent level was used to control for multiple testing of SNPs during the selection procedure. For these methods, we used the expectation-maximization algorithm to estimate all possible pairs of haplotypes for each subject, given their observed unphased data, together with posterior probabilities of the various haplotype pairs [34], using the software HaploStats. The posterior probabilities were then used to compute expected haplotype scores to test for the effects of haplotypes in linear regression analyses [35]. The forward method added SNPs, one at a time, to the growing haplotypes, while comparing the larger haplotype at each step with the smaller haplotype at each prior step, making it possible to determine whether the addition of a SNP to the growing haplotype was statistically significant. The backward method did just the opposite, removing SNPs that did not contribute significantly.

Results

The experiments described subsequently were performed in an attempt to extend a previous report based on a much small number of samples, that the *COMT* nucleotide 472G>A, codon 108 Val>Met polymorphism was associated with variation in both S-COMT activity and thermal stability in human liver [9], and to test the hypothesis that additional polymorphisms in *COMT* might also be associated with variation in S-COMT phenotypes. As a first step, we extended our earlier *COMT* gene resequencing effort [10] to include the 5'-end of the gene. Subsequently, 268 human surgical liver biopsy samples were phenotyped for level of enzyme activity and thermal stability and DNA from the same samples was genotyped for 16 *COMT* polymorphisms to make it possible to perform genotype-phenotype and haplotype-phenotype correlation analyses. Finally, since the proximal promoter for S-COMT is located in intron 2, reporter gene constructs for common, naturally occurring *COMT* proximal promoter haplotypes were tested in HepG2 cells to determine their possible relationship to results obtained with the hepatic biopsy samples.

COMT gene resequencing

We had previously resequenced *COMT* from exon 2 through the final exon, exon 6 [10]. However, exon 1 and the 5'-FR of the gene were not resequenced because the DNA sequence in that area of the gene was highly GC-rich. As the first step in the present study, we resequenced the 5'-end of the gene, and those data – for the same 60 CA and 60 AA DNA samples used in the previous resequencing study – are listed in Table 1 and are depicted graphically in Fig. 1A (see “boxed” arrows). We observed a total of 31 polymorphisms in the newly resequenced area of the gene. All of the 13 polymorphisms in this portion of the gene in the CA DNA samples, and 20 of the 31 observed in samples from AA subjects had frequencies of greater than 1% and, as a result, would be considered “common” polymorphisms in these populations. When these 31 polymorphisms were compared with those in the public database dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>), 18 of the 31 were not represented in dbSNP. Sixteen of those 18 novel polymorphisms were only observed in DNA from AA subjects (Table 1). All of these polymorphisms, the 31 polymorphisms observed at the 5'-end of the gene and the 24 observed previously, were subjected to pairwise linkage disequilibrium analysis, and those results are depicted graphically in Fig. 2. These data show clearly that *COMT* contains at least two haplotype blocks – including one at the 5'-end of the gene that is clearly distinct from the block

encoding S-COMT. Obviously, we are not able to rule out additional haplotype blocks in the long (~20 kb) intron 1. These *COMT* resequencing data, plus those obtained during our initial gene resequencing effort [10], were then used to perform the human liver genotype-phenotype correlation analysis described subsequently. However, to make that correlation analysis possible, we first had to phenotype the liver cytosol preparations.

Hepatic COMT activity and thermal stability

COMT enzyme activity and thermal stability were measured in 268 human liver cytosol preparations. COMT activity averaged 3.26 ± 1.28 units/mg tissue, with an average H/C ratio of 0.52 ± 0.15 . Neither of these phenotypes were significantly associated with patient age ($P = 0.165$ for level of activity, and $P = 0.133$ for H/C ratio), so subsequent analyses did not correct for age as a potential confounder. The distribution of both level of hepatic COMT activity (Fig. 3A) and H/C ratio (Fig. 3B) fit a mixture distribution that consisted of two underlying normal distributions better than a single distribution ($P = 0.028$ and $P < 10^{-6}$ by likelihood ratio tests, respectively), suggesting that an underlying genotype might explain at least a portion of these distributions. As anticipated, there was a significant correlation between level of activity and thermal stability ($R_s = 0.413$, $P < 0.001$, Fig. 4A). Furthermore, this correlation was largely explained by the 108Val>Met polymorphism, as reported previously for a much smaller population sample [9] – with an association of Met108 with decreases in both thermal stability and enzyme activity (Fig. 4B).

COMT genotype-phenotype correlation analysis

Sixteen *COMT* SNPs were genotyped, and their allele frequencies in the 268 liver biopsy samples and in the Coriell samples are listed in Table 2. The frequencies in the two tissue sources were generally quite similar. Based on single SNP analyses, level of COMT enzyme activity was associated with a number of SNPs, as illustrated in Fig. 5. The top panel illustrates that the nucleotide 472G>A polymorphism (108Val>Met) had the strongest association, explaining 20.2% of the variance in level of activity. The A allele encoding Met108 was associated with a reduced activity level relative to the G allele that encodes Val108, as reported previously [9], with each A allele reducing activity by 0.81 ± 0.10 (mean $\pm$ SE). There was no evidence that the effect of the A allele was not additive ($P = 0.43$).

Because the marginal associations shown in Fig. 5 were not adjusted for the known effects of the nucleotide 472 polymorphism on level of activity, and because the effects of these SNPs were correlated among themselves as a result of linkage disequilibrium (Fig. 2), it was crucial to test the effect of each SNP after adjusting for the effect of the nucleotide 472 polymorphism. That was done by performing a forward stepwise selection of SNPs that contributed significantly to level of enzyme activity, after adjusting for the effect of prior chosen SNPs to the haplotype. Fig. 5 illustrates the sequential steps that were used to evaluate the sequentially expanded two- and three-locus haplotypes, with each analysis adjusted for the prior step. The analysis terminated at the three-locus haplotypes because of lack of statistical significance when comparing the effects of three-locus versus four-locus haplotypes. With a significance level of 0.01, the initial set of SNPs that were significantly associated with activity level were those at exon 4(408) and exon 4 (472). Those two SNPs created four haplotypes. Estimates of the frequencies of those four haplotypes, as well as their effect on enzyme activity, are listed in Table 3A. These haplotypes explained 24.4% of the variation in activity level, an increase of 4.2% above that accounted for by the exon 4 (472) polymorphism alone. As illustrated by the regression coefficients listed in Table 3A, when controlling for the exon 4 (472) allele (i.e., comparing regression coefficients for haplotypes with the same exon 4 (472) allele), the C allele of the exon 4 (408) SNP had lower levels of activity than did the G allele at this SNP. Therefore, although the exon 4

(472) polymorphism explained the majority of variation in level of activity, there were significant residual effects of other SNPs that also influenced activity level. In addition, the intron 2 (51G>C) polymorphism, when adjusted for the effect of the exon 4 (408–472) haplotype, was also marginally significant ($P = 0.02$). Although this effect did not achieve the more stringent level of 0.01, an attempt to correct for multiple testing, it suggested that the intron 2 (51) SNP might warrant further evaluation. When this polymorphism was combined into a set of three-locus haplotypes together with the exon 4 (408) and exon 4(472) SNPs, there were seven haplotypes that explained 27.3% of the variation in level of COMT activity in these samples (Table 3B). After controlling for the effects of alleles at the exon 4 (408) and exon 4(472) SNPs, the C allele for the intron 2 (51) polymorphism was associated with higher activity than was the G allele.

When similar analyses were performed for the thermal stability phenotype, we found that the exon 4 (472G>A) polymorphism had the strongest association ($P < 10^{-6}$), explaining 59% of the variance in thermal stability. After adjusting for this SNP, no other SNPs contributed significantly to any of the haplotype associations. This result was not surprising since Fig. 4 showed that exon 4 (472) genotypes were more tightly clustered for thermal stability than they were for level of activity.

Finally, since we had genotyped 5 SNPs in the *COMT* proximal promoter, and since one of those SNPs, intron 2 (51G>C), appeared to be marginally significant before correction for multiple comparisons, we created and tested reporter gene constructs for naturally occurring proximal promoter haplotypes for *COMT* and then compared results obtained using those constructs with the genotype-phenotype associations described in preceding paragraphs for the hepatic biopsy samples.

***COMT* proximal promoter haplotype studies**

The most common proximal promoter haplotypes that we observed during *COMT* gene resequencing are listed in Table 4. Reporter gene constructs for the seven common intron 2 (proximal promoter) haplotypes listed in Table 4 were created and tested after transfection into HepG2 cells. The results of those studies are shown graphically in Fig. 6. The next step was to compare these data with the results of an analysis of the association of proximal promoter haplotypes with phenotypes in the 268 liver biopsy samples. Specifically, our initial genotype-phenotype correlation analysis had indicated that the C allele for the intron 2 (51) polymorphism was associated with higher activity in the liver samples than was the G allele at that nucleotide. It was possible to directly compare the role of this polymorphism for two pairs of the haplotypes that are listed in Table 4 – haplotype pair #1 and #3, and haplotype pair #4 and #7. Both members of these pairs of haplotype were identical with the exception of the nucleotide at position 51 in intron 2. In both cases, the haplotype with “C” at that position displayed significantly higher reporter gene activity than did the comparable haplotype with “G” at the same position (i.e., #1 > #3 and #7 > #4) (Fig. 6) – compatible with results of the analysis that we performed with the hepatic tissue samples. Search of the AliBaba2.1 [36] transcription factor binding site database predicted that the intron 2 (51) polymorphism occurred within an Sp1 binding motif. Since we had performed expression array analyses with the lymphoblastoid cell lines from which the DNA that we used to resequence the *COMT* gene had been obtained – we were also able to determine whether this polymorphism might be associated with basal COMT expression in those cell lines – as described in subsequent paragraphs.

Expression array and exon array analyses

We obtained *COMT* expression array data using Affymetrix U133 Plus 2.0 GeneChips for lymphoblastoid cell lines from the 60 CA subjects from which the DNA had been extracted

that was used to perform the *COMT* gene resequencing studies. There was a large variation in basal *COMT* expression using either of the two *COMT*-specific probe sets on those chips (208817_at and 208818_s_at). Single SNP vs. *COMT* expression correlation analysis revealed that the Intron 2 (51) “C” allele was significantly associated with increased *COMT* expression for probe set 208817_at (p-value = 0.02, r=0.3), but not for the 208818_s_at probe set. These results for probe set 208817_at were also compatible with our reporter gene results (Fig. 6) and with the results found with the liver biopsy tissue – both of which indicated the possibility of higher levels of transcription for the variant “C” allele. Obviously, expression regulation is tissue dependent, but it was of interest that all of these results were consistent.

Discussion

COMT catalyzes an important reaction in catecholamine, catecholestrogen and catechol drug metabolism [37,38]. It was shown 30 years ago that level of S-*COMT* activity varies widely among individuals and that this variation is due – in part – to the effects of inheritance [5,27]. Furthermore, it was reported a decade ago that this variation was due, at least in part, to the effects of the nonsynonymous exon 4(472G>A), Val108Met SNP [9]. Low level of enzyme activity was associated with the allele encoding Met108 as a result of a decrease in level of S-*COMT* immunoreactive protein [10]. In the present study, we set out to determine whether additional sequence variation in *COMT* beyond the polymorphism resulting in the Val108Met alteration in amino acid sequence might influence hepatic phenotype. As a first step, we resequenced exon 1, the 5'-FR and a portion of intron 1 of *COMT*. Those results, when added to our previous resequencing data, made it possible to select LD and ht tag SNPs – as well as *COMT* polymorphisms previously reported to have possible functional implications [20,21] – for inclusion in our genotyping studies. We then determined S-*COMT* phenotypes in 268 human hepatic surgical biopsy samples. Specifically, S-*COMT* activity and thermal stability were measured, and DNA from the same samples was genotyped for the 16 *COMT* SNPs selected for inclusion in the genotyping studies. The results confirmed the association of the nucleotide 472 108Val>Met polymorphism with both level of hepatic enzyme activity and thermal stability (see Fig. 4). In fact, it appeared – not too surprisingly – that this polymorphism that resulted in an alteration in the amino acid sequence was the only SNP associated with variation in the thermal stability phenotype (Fig. 5). However, two additional SNPs, exon 4 (408C>G), and intron 2 (51G>C) also influenced level of activity, although to a lesser degree than did the codon 108 polymorphism. These observations with regard to nucleotide 408 are compatible with, and support recent reports that this synonymous SNP influences S-*COMT* mRNA structure and stability and, as a result, expression [21]. Finally, our attempts to determine whether SNPs or haplotypes in the proximal promoter located in intron 2 influenced phenotype indicated that those polymorphisms were associated with level of enzyme activity but not with the thermal stability phenotype. When reporter gene constructs were created for intron 2 (the location of the “proximal promoter”), reporter gene activities for haplotype pairs that differed only for nucleotides at the intron 2 (51) polymorphism were compatible with the results found during the analysis of data for the liver biopsy samples (Table 4 and Fig. 6). Specifically, “C” at that position was associated with higher activity than was “G” – results that were also supported by expression array analyses performed with the lymphoblastoid cell lines from which DNA had been extracted for the *COMT* gene resequencing studies.

In summary, our study has confirmed that the 472G>A, 108Val>Met polymorphism appears to be the most important genetic factor influencing S-*COMT* activity and thermal stability in human hepatic tissue [9]. That SNP was associated with 20.2% of the total variation in level of enzyme activity and 59% of the variation in thermal stability. However, after adjusting

for the effect of genotype at nucleotide 472, additional *COMT* polymorphisms, including the synonymous nucleotide 408C>G SNP in exon 4 and the intron 2 (51G>C) SNP also displayed residual association with level of COMT activity. Therefore, those two polymorphisms should be included in future molecular epidemiology studies of the possible relationship of genotype for this important drug, neurotransmitter and hormone-metabolizing enzyme with variation in either drug response phenotypes or disease risk.

Acknowledgments

This study was supported in part by National Institutes of Health (NIH) grants R01 GM28157, U01 GM61388, The Pharmacogenetics Research Network, and by a PhRMA Foundation Center of Excellence in Clinical Pharmacology Award. We thank Mrs. Luanne Wussow for her assistance with the preparation of this manuscript.

References

1. Weinshilboum RM, Otterness DM, Szumlanski CL. Methylation pharmacogenetics: catechol O-methyltransferase, thiopurine methyltransferase, and histamine N-methyltransferase. *Annu Rev Pharmacol Toxicol* 1999;39:19–52. [PubMed: 10331075]
2. Tenhunen J, Salminen M, Lundström K, Kiviluoto T, Savolainen R, Ulmanen I. Genomic organization of the human catechol O-methyltransferase gene and its expression from two distinct promoters. *Eur J Biochem* 1994;223:1049–59. [PubMed: 8055944]
3. Lundström K, Tenhunen J, Tilgmann C, TK, Panula P, Ulmanen I. Cloning, expression and structure of catechol-O-methyltransferase. *Biochim Biophys Acta* 1995;1251:1–10. [PubMed: 7647086]
4. Jeffery DR, Roth JA. Characterization of membrane-bound and soluble catechol-O-methyltransferase from human frontal cortex. *J Neurochem* 1984;42:826–832. [PubMed: 6693904]
5. Weinshilboum RM, Raymond FA. Inheritance of low erythrocyte catechol O-methyltransferase activity in man. *Am J Human Genet* 1977;29:125–135. [PubMed: 848488]
6. Scanlon PD, Raymond FA, Weinshilboum RM. Catechol-O-methyltransferase: thermolabile enzyme in erythrocytes of subjects homozygous for allele for low activity. *Science* 1979;203:63–5. [PubMed: 758679]
7. Weinshilboum R, Dunnette J. Thermal stability and the biochemical genetics of erythrocyte catechol-O-methyl-transferase and plasma dopamine-beta-hydroxylase. *Clin Genet* 1981;19:426–37. [PubMed: 7296933]
8. Boudíková B, Szumlanski C, Maidak B, Weinshilboum R. Human liver catechol O-methyltransferase pharmacogenetics. *Clin Pharmacol Ther* 1990;48:381–389. [PubMed: 2225698]
9. Lachman HM, Papolos DF, Saito T, Yu YM, Szumlanski CL, Weinshilboum RM. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics* 1996;6:243–50. [PubMed: 8807664]
10. Shield AJ, Thomae BA, Eckloff BW, Wieben ED, Weinshilboum RM. Human catechol O-methyltransferase genetic variation: gene resequencing and functional characterization of variant allozymes. *Mol Psychiatr* 2004;9:151–160.
11. Lavigne JA, Helzlsouer KJ, Huang HY, Strickland PT, Bell DA, Selmin O, et al. An association between the allele coding for a low activity variant of catechol-O-methyltransferase and the risk for breast cancer. *Cancer Res* 1997;57:5493–7. [PubMed: 9407957]
12. Millikan RC, Pittman GS, Tse CK, Duell E, Newman B, Savitz D, et al. Catechol-O-methyltransferase and breast cancer risk. *Carcinogenesis* 1998;19:1943–7. [PubMed: 9855007]
13. Thompson PA, Shields PG, Freudenheim JL, Stone A, Vena JE, Marshall JR, et al. Genetic polymorphisms in catechol-O-methyltransferase, menopausal status, and breast cancer risk. *Cancer Res* 1998;58:2107–10. [PubMed: 9605753]
14. Huang CS, Chern HD, Chang KJ, Cheng CW, Hsu SM, Shen CY. Breast cancer risk associated with genotype polymorphism of the estrogen-metabolizing genes CYP17, CYP1A1, and COMT: a multigenic study on cancer susceptibility. *Cancer Res* 1999;59:4870–5. [PubMed: 10519398]

15. Fan JB, Zhang CS, Gu NF, Li XW, Sun WW, Wang HY, et al. Catechol-O-methyltransferase gene Val/Met functional polymorphism and risk of schizophrenia: a large-scale association study plus meta-analysis. *Biol Psychiatry* 2005;57:139–144. [PubMed: 15652872]
16. Fabisch H, Kroisel PM, Fabisch K. Genetic risk factors in schizophrenia. *Fortschr Neurol Psychiatr* 2005;73 (Suppl 1):S44–50. [PubMed: 16270244]
17. Mátyás T. Gene polymorphism and gene expression in schizophrenia. *Psychiatr Hung* 2006;21:404–412. [PubMed: 17438657]
18. Tunbridge EM, Harrison PJ, Weinberger DR. Catechol-o-methyltransferase, cognition, and psychosis: Val158Met and beyond. *Biol Psychiatry* 2006;60:141–151. [PubMed: 16476412]
19. Boudikova B, Szumlanski C, Maidak B, Weinshilboum R. Human liver catechol-O-methyltransferase pharmacogenetics. *Clin Pharmacol Ther* 1990;48:381–9. [PubMed: 2225698]
20. Shifman S, Bronstein M, Sternfeld M, Pisante-Shalom A, Lev-Lehman E, Weizman A, et al. A highly significant association between a COMT haplotype and schizophrenia. *Am J Hum Genet* 2002;71:1296–302. [PubMed: 12402217]
21. Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskiy O, Makarov SS, et al. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 2006;314:1930–1933. [PubMed: 17185601]
22. Ji Y, Salavaggione OE, Wang L, Adjei AA, Eckloff B, Wieben ED, et al. Human phenylethanolamine N-methyltransferase pharmacogenomics: gene resequencing and functional genomics. *J Neurochem* 2005;95:1766–1776. [PubMed: 16277617]
23. Ma CX, Adjei AA, Salavaggione OE, Coronel J, Pellemounter L, Wang L, et al. Human aromatase: gene resequencing and functional genomics. *Cancer Res* 2005;65:11071–11082. [PubMed: 16322257]
24. Martin YN, Salavaggione OE, Eckloff BW, Wieben ED, Schaid DJ, Weinshilboum RM. Human methylenetetrahydrofolate reductase pharmacogenomics: gene resequencing and functional genomics. *Pharmacogenet Genomics* 2006;16:265–277. [PubMed: 16538173]
25. Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 1997;25:2745–51. [PubMed: 9207020]
26. Bardelli A, Parsons DW, Silliman N, Ptak J, Szabo S, Saha S, et al. Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* 2003;300:949. [PubMed: 12738854]
27. Raymond FA, Weinshilboum RM. Microassay of human erythrocyte catechol O methyltransferase: removal of inhibitory calcium ion with chelating resin. *Clin Chim Acta* 1975;58:185–194. [PubMed: 235384]
28. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–120. [PubMed: 14681826]
29. Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, et al. Choosing haplotype-tagging SNPs based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. *Hum Hered* 2003;55:27–36. [PubMed: 12890923]
30. Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, Kopla R, et al. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques* 2002;(Suppl):70–2. 74, 76–77. [PubMed: 12083401]
31. Denomme GA, Van Oene M. High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. *Transfusion* 2005;45:669–666.
32. Moyer AM, Salavaggione OE, Wu TY, Moon I, Eckloff BW, Hildebrandt MA, et al. Glutathione S-transferase P1: gene sequence variation and functional genomic studies. *Cancer Res* 2008;68:4791–4801. [PubMed: 18559526]
33. Ji Y, Olson J, Zhang J, Hildebrandt M, Wang L, Ingle J, et al. Breast cancer risk reduction and membrane-bound catechol O-methyltransferase genetic polymorphisms. *Cancer Res* 2008;68:5997–6005. [PubMed: 18632656]

34. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425–434. [PubMed: 11791212]
35. Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered* 2002;53:79–91. [PubMed: 12037407]
36. Grabe N. AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol* 2002;2:S1–15. [PubMed: 11808873]
37. Axelrod J, Senoh S, Witkop B. O-Methylation of catechol amines in vivo. *J Biol Chem* 1958;233:697–701. [PubMed: 13575439]
38. Axelrod J, Tomchick R. Enzymatic O-methylation of epinephrine and other catechols. *J Biol Chem* 1958;233:702–705. [PubMed: 13575440]

Figure 1.

Human *COMT* polymorphisms. Black rectangles represent coding exons, and open rectangles represent portions of exons that encode UTR sequences. The cross-hatched area encodes protein in MB-COMT but not in S-COMT. “CA” indicates Caucasian-American and “AA” African-American subjects. (A) Polymorphisms observed during our *COMT* resequencing studies are shown. Arrows show polymorphism locations, with frequencies indicated by the color of the arrows. The “box” labeled “New Resequencing Data” contains the novel resequencing data obtained during the present study. (B) Polymorphisms selected for inclusion in the genotyping panel. Arrows indicates locations of those polymorphisms, and are “color-coded” based on their frequencies in CA subjects. The two “boxed” polymorphisms were not within the area resequenced, but were included because they had been reported to be of possible functional significance during previous studies [20].

Figure 2. Linkage disequilibrium for *COMT* polymorphisms in DNA samples from AA (panel A) and CA (panel B) subjects based on both our original [10] and the present resequencing data.

Figure 3. Human liver S-COMT activity and thermal stability. (A) Frequency distribution histogram of hepatic biopsy S-COMT activity in samples from 268 Caucasian women. (B) Frequency distribution histogram of hepatic biopsy S-COMT thermal stability in the same samples.

Figure 4. Human liver S-COMT activity and thermal stability. (A) Scattergram of S-COMT activity on the X-axis and S-COMT thermal stability on the Y-axis (expressed as an H/C ratio). (B) Average values $\pm$ SD for each genotype shown in (A).

Figure 5. Sequential selection of SNPs for their association with human hepatic S-COMT activity, using haplotype analyses, adjusted for sub-haplotype effects.

Figure 6.

Human *COMT* “proximal promoter” (intron 2) reporter gene studies. Activities of luciferase reporter gene constructs that included different *COMT* proximal promoter haplotypes were assayed after the transfection of HepG2 cells are expressed as a percentage of the value for the most common haplotype (construct #1). Each bar represents the average of 6 independent assays (mean $\pm$ SEM); * $p < 0.05$; ** $p < 0.01$ when compared with the value for construct #1.

Human *COMT* gene resequencing data. Polymorphism locations, alterations in nucleotide sequences, minor allele frequencies (MAFs) for the 2 ethnic groups studied and rs numbers – when available – are listed. Numbering of nucleotides in the 5'-FR begins at the "A" in the first "ATG" for the MB-COMT cDNA, with all of these nucleotides assigned negative numbers. Nucleotides located within intron 1 were numbered on the basis of their distance from the nearest splice site. Note that the "intron 1, nucleotide 517" location could have either a rare insertion or a common SNP.

Table 1

Human COMT Gene Resequencing Data					
Location	Nucleotide	Sequence Change	MAF		Rs ID.
			AA	CA	
5'FR of MB-	-1420	A>G	0.425	0.358	rs2075507
5'FR of MB-	-1187	G>A	0.008	0.000	
5'FR of MB-	-1158	G>A	0.008	0.000	
5'FR of MB-	-1037	T>C	0.025	0.000	
5'FR of MB-	-963	G>A	0.017	0.000	
5'FR of MB-	-930	G>A	0.025	0.017	
5'FR of MB-	-901	G>A	0.033	0.000	
5'FR of MB-	-706	G>A	0.075	0.250	rs4646310
5'FR of MB-	-628	C>T	0.150	0.317	rs2020917
5'FR of MB-	-566	G>A	0.033	0.000	
5'FR of MB-	-485	C>T	0.025	0.133	rs13306278
5'FR of MB-	-369 to -368	insertion of G	0.150	0.000	
5'FR of MB-	-307	C>T	0.008	0.000	
5'FR of MB-	-227	G>A	0.008	0.000	
5'FR of MB-	-226	C>A	0.008	0.000	
Intron 1	99	G>A	0.050	0.017	
Intron 1	131	G>A	0.492	0.383	rs3788319
Intron 1	470	G>A	0.008	0.000	
Intron 1	517	insertion of GGAGGC	0.008	0.000	
Intron 1	517	C>A	0.333	0.283	rs3788321
Intron 1	530	C>T	0.008	0.000	
Intron 1	582	C>G	0.325	0.250	rs9306231
Intron 1	594	G>T	0.150	0.000	rs9306232
Intron 1	595	G>T	0.008	0.000	

Human COMT Gene Resequencing Data						
Location	Nucleotide	Sequence Change	MAF		CA	Rs ID.
			AA	CA		
Intron 1	597	G>C	0.142	0.000		rs9306233
Intron 1	689	T>C	0.158	0.333		rs737866
Intron 1	701	A>G	0.158	0.333		rs737865
Intron 1	718	C>G	0.008	0.000		
Intron 1	739	C>T	0.167	0.350		rs737864
Intron 1	757	C>T	0.008	0.000		
Intron 1	770	deletion of C	0.058	0.192		rs5844401

COMT polymorphisms selected for inclusion in the genotyping studies. The table lists the locations of the polymorphism in the gene as well as minor allele frequencies (MAFs) in the Coriell CA DNA samples and in the 268 surgical liver biopsy samples.

Table 2

dbSNP ID	Location In <i>COMT</i>	Nucleotide	Nucleotide Change	MAF in CA Coriell Samples	MAF in Liver Samples
rs52075507	5'-FR	-1420	A>G	0.36	0.44
rs4646310	5'-FR	-706	G>A	0.25	0.18
rs2020917	5'-FR	-628	C>T	0.32	0.27
rs13306278	5'-FR	-485	C>T	0.13	0.14
rs737865	Intron 1	701	A>G	0.33	0.27
rs165656	Intron 2	51	G>C	0.43	0.47
rs165722	Intron 2	201	C>T	0.44	0.33
rs3810595	Intron 2	832	G>C	0.50	0.35
rs11569716	Intron 2	935	T>C	0.02	0.05
rs6269	Intron 2	1140	G>A	0.49	0.60
rs4633	Exon 3	186	C>T	0.41	0.50
rs4818	Exon 4	408	C>G	0.49	0.62
rs4680	Exon 4	472	G>A	0.43	0.49
rs4646315	Intron 5	75	G>C	0.13	0.19
rs362204	3'-UTR	820	I>D	0.23	0.27
rs165599	3'-FR	1338	A>G	0.26	0.31

Table 3

Linear regression results for level of COMT activity based on the best fitting (A) two-locus haplotypes (explaining 24.4% of activity variance) or (B) three-locus haplotypes (explaining 27.3% of activity variance). All P values are based on the F-statistic, comparing the effect of each haplotype to the “baseline” haplotype, i.e. the most common haplotype on the basis of frequency (shaded).

(A)						
No.	Exon 4 (408)	Exon 4 (472)	Frequency	Regression Coefficient	SE	P-value
1	C	A	0.002	-1.27	1.13	0.2620
2	C	G	0.384	0.64	0.11	<0.0001
3	G	A	0.486	Baseline	---	---
4	G	G	0.128	1.21	0.15	<0.0001

(B)							
	Intron 2 (51)	Exon 4 (408)	Exon 4 (472)	Frequency	Regression Coefficient	SE	P value
1	C	C	G	0.014	1.63	0.46	0.0004
2	C	G	A	0.434	Baseline	---	---
3	C	G	G	0.018	1.19	0.35	0.0009
4	G	C	A	0.0002	-1.38	1.11	0.2152
5	G	C	G	0.370	0.52	0.11	<0.0001
6	G	G	A	0.053	-0.45	0.21	0.0373
7	G	G	G	0.109	1.21	0.16	<0.0001

Table 4

Human *COMT* proximal promoter haplotypes. The table lists sequences of the most common *COMT* exon 2 (proximal promoter) haplotypes observed in the course of the gene resequencing studies – as well as their frequencies in DNA samples from CA and AA subjects, and frequencies in the genotyped human liver biopsy samples. Boxed rows indicate haplotype pairs that differ only at nucleotide 51.

Haplotype	Intron 2 Nucleotide Designation					Frequency		
	51	201	832	935	1140	AA	CA	Liver Samples
#1	C	T	G	T	A	0.271	0.363	0.328
#2	G	C	C	T	G	0.201	0.448	0.285
#3	G	T	G	T	A	0.152	0.052	0.003
#4	G	C	G	T	A	0.115	0.042	0.183
#5	G	T	C	T	G	0.057	0.008	0.002
#6	C	C	C	C	G	0.054	0.017	0.033
#7	C	C	G	T	A	0.022	----	0.074