Human Catechol-O-Methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure

ARTICLE *in* SCIENCE · JANUARY 2007

Impact Factor: 33.61 · DOI: 10.1126/science.1131262 · Source: PubMed

8 AUTHORS, INCLUDING:

Inna E Tchivileva
University of North Carolina at Chapel Hill
13 PUBLICATIONS **893** CITATIONS
SEE PROFILE

Sergei S. Makarov
Attagene, Inc.
23 PUBLICATIONS **4,200** CITATIONS
SEE PROFILE

William Maixner
University of North Carolina at Chapel Hill
288 PUBLICATIONS **10,933** CITATIONS
SEE PROFILE

Luda Diatchenko
University of North Carolina at Chapel Hill
101 PUBLICATIONS **9,945** CITATIONS
SEE PROFILE

creased, and body mass was lower when squirrels were feeding on these items as compared with feeding on seed (*26*, *27*). Lastly, food supplementation experiments of American reds have failed to produce increases in litter size or a second litter, providing strong evidence that the increased reproductive investment observed in our study cannot be triggered by increased energy availability alone (*17*, *28*, *29*). We hypothesize that, rather than following a resource-tracking strategy where reproductive investment is determined by current resource amounts, reproductive rates are driven by future fitness payoffs. During years of low seed production, competition among juveniles for available resources is intense, and, although litter augmentation experiments in American reds show that females are capable of supporting larger litters (*30*), they refrain from doing so because offspring recruitment is low (*30*). However, when mast years occur, competition among juveniles is reduced, and females produce more offspring, which successfully recruit into the population (*31*, *32*). Further, the increased production of young by females in mast years does not come with any obvious cost to the female, because overwinter survival is not reduced after years of increased reproductive investment (for American reds, offspring production in the previous year versus proportion of adult females surviving to spring has slope = –0.023 ± 0.034, $t_{15} = -0.7$, and $P = 0.51$; for Eurasian reds, proportion of estrous females in the previous year versus adult female survival to spring has slope = 0.37 ± 0.27, $t_{15} = 1.4$, and $P = 0.19$).

If masting has evolved as a swamp-and-starve adaptation against seed predation, then anticipatory reproduction and population growth represent a potent counteradaptation by the predators. Given that increased reproductive output in these systems coincides with low current but high future resources, our results suggest that reproductive investment in these systems is more responsive to future fitness prospects than present energetic constraints. The evolution of seed masting in trees is also driven by the survival prospects for progeny rather than simple resource tracking, suggesting an intriguing parallel in reproductive strategies of trees and the predators that consume their seed.

References and Notes

1. J. L. Gittleman, S. D. Thompson, *Am. Zool.* **28**, 863 (1988).
2. T. H. Clutton-Brock, *Reproductive Success* (Univ. of Chicago Press, Chicago, 1988).
3. T. H. Clutton-Brock, *The Evolution of Parental Care* (Princeton Univ. Press, Princeton, NJ, 1999).
4. G. Caughley, in *Theoretical Ecology*, R. M. May, Ed. (Blackwell, Oxford, 1981), ch. 6.
5. P. Bayliss, D. Choquenot, *Proc. R. Soc. London Ser. B* **357**, 1233 (2002).
6. C. G. Jones, R. S. Ostfeld, M. P. Richard, E. M. Schauber, J. O. Wolff, *Science* **279**, 1023 (1998).
7. R. S. Ostfeld, F. Keesing, *Trends Ecol. Evol.* **15**, 232 (2000).
8. D. Kelly, V. L. Sork, *Annu. Rev. Ecol. Syst.* **33**, 427 (2002).
9. L. M. Curran, M. Leighton, *Ecol. Monogr.* **70**, 101 (2000).
10. W. D. Koenig, J. M. H. Knops, *Am. Sci.* **93**, 340 (2005).
11. D. H. Janzen, *Annu. Rev. Ecol. Syst.* **2**, 465 (1971).
12. A. F. Hedlin, *For. Sci.* **10**, 124 (1964).
13. M. J. McKone, D. Kelly, A. L. Harrison, J. J. Sullivan, A. J. Cone, *N.Z. J. Zool.* **28**, 89 (2001).
14. P. R. Wilson, B. J. Karl, R. J. Toft, J. R. Beggs, *Biol. Conserv.* **83**, 175 (1998).
15. T. Ruf, J. Fietz, W. Schlund, C. Bieber, *Ecology* **87**, 372 (2006).
16. F. H. Bronson, *Reproductive Biology* (Univ. of Chicago Press, Chicago, 1989).
17. Materials and methods are available on *Science* Online.
18. J. O. Wolff, *J. Mammal.* **77**, 850 (1996).
19. A. S. McNeilly, *Reprod. Fertil. Dev.* **13**, 583 (2001).
20. J. D. Ligon, *Nature* **250**, 80 (1974).
21. P. J. Berger, N. C. Negus, E. H. Sanders, P. D. Gardner, *Science* **214**, 69 (1981).
22. S. Eis, J. Inkster, *Can. J. For. Res.* **2**, 460 (1972).
23. B. M. Fitzgerald, M. J. Efford, B. J. Karl, *N.Z. J. Zool.* **31**, 167 (2004).
24. M. C. Smith, *J. Wildl. Manag.* **32**, 305 (1968).
25. D. A. Rusch, W. G. Reeder, *Ecology* **59**, 400 (1978).
26. L. A. Wauters, C. Swinnen, A. A. Dhondt, *J. Zool.* **227**, 71 (1992).
27. L. A. Wauters, A. A. Dhondt, *J. Zool.* **217**, 93 (1989).
28. C. D. Becker, *Can. J. Zool.* **71**, 1326 (1993).
29. K. W. Larsen, C. D. Becker, S. Boutin, M. Blower, *J. Mammal.* **78**, 192 (1997).
30. M. M. Humphries, S. Boutin, *Ecology* **81**, 2867 (2000).
31. A. G. McAdam, S. Boutin, *Evolution* **57**, 1689 (2003).
32. L. A. Wauters, E. Matthysen, F. Adriaensen, G. Tosi, *J. Anim. Ecol.* **73**, 11 (2004).

33. We thank J. Herbers, C. Aumann, J. LaMontagne, M. Andruskiw, J. Lane, R. Boonstra, K. McCann, B. Thomas, Kluane Red Squirrel Project 2005 annual meeting participants, and three anonymous reviewers for very useful comments. We are indebted to the many field workers who helped collect the data and to A. Sykes and E. Anderson for management of the long-term data. Financed by Natural Sciences and Engineering Research Council of Canada, NSF (American reds), a Concerted Action of the Belgian Ministry of Education, and a Training and Mobility of Researchers grant from the Commission of the European Union (Eurasian reds). This is publication 30 of the Kluane Red Squirrel Project.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5807/1928/DC1
Materials and Methods
SOM Text
Fig. S1
Table S1
References

25 September 2006; accepted 15 November 2006
10.1126/science.1135520

Human Catechol-*O*-Methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure

A. G. Nackley,[1] S. A. Shabalina,[2] I. E. Tchivileva,[1] K. Satterfield,[1] O. Korchynskyi,[3] S. S. Makarov,[4] W. Maixner,[1] L. Diatchenko[1]*

Catechol-*O*-methyltransferase (COMT) is a key regulator of pain perception, cognitive function, and affective mood. Three common haplotypes of the human *COMT* gene, divergent in two synonymous and one nonsynonymous position, code for differences in COMT enzymatic activity and are associated with pain sensitivity. Haplotypes divergent in synonymous changes exhibited the largest difference in COMT enzymatic activity, due to a reduced amount of translated protein. The major *COMT* haplotypes varied with respect to messenger RNA local stem-loop structures, such that the most stable structure was associated with the lowest protein levels and enzymatic activity. Site-directed mutagenesis that eliminated the stable structure restored the amount of translated protein. These data highlight the functional significance of synonymous variations and suggest the importance of haplotypes over single-nucleotide polymorphisms for analysis of genetic variations.

The ability to predict the downstream effects of genetic variation is critically important for understanding both the evolution of the genome and the molecular basis of human disease. The effects of nonsynonymous polymorphisms have been widely characterized; because these variations directly influence protein function, they are relatively easy to study statistically and experimentally (*1*). However, characterizing polymorphisms located in regulatory regions, which are much more common, has proved to be problematic (*2*). Here, we focus on the mechanism whereby polymorphisms of the cathechol-*O*-methyltransferase (*COMT*) gene regulate gene expression.

COMT is an enzyme responsible for degrading catecholamines and thus represents a critical component of homeostasis maintenance (*3*). The human *COMT* gene encodes two distinct proteins: soluble COMT (S-COMT) and membrane-bound COMT (MB-COMT) through the use of alternative translation initia-

[1]Center for Neurosensory Disorders, University of North Carolina, Chapel Hill, NC 27599, USA. [2]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA. [3]Thurston Arthritis Center, University of North Carolina, Chapel Hill, NC 27599, USA. [4]Attagene, Inc., 7030 Kit Creek Road, Research Triangle Park, NC 27560, USA.

*To whom correspondence should be addressed. E-mail: lbdiatch@email.unc.edu

tion sites and promoters (*3*). Recently, COMT has been implicated in the modulation of persistent pain (*4*–*7*). Our group demonstrated that three common haplotypes of the human *COMT* gene are associated with pain sensitivity and the likelihood of developing temporomandibular joint disorder (TMJD), a common chronic musculoskeletal pain condition (*4*). Three major haplotypes are formed by four single-nucleotide polymorphisms (SNPs): one located in the *S-COMT* promoter region (A/G; rs6269) and three in the *S-* and *MB-COMT* coding region at codons $his^{62}his$ (C/T; rs4633), $leu^{136}leu$ (C/G; rs4818), and $val^{158}met$ (A/G; rs4680) (Fig. 1A). On the basis of subjects' pain responsiveness, haplotypes were designated as low (LPS; GCGG), average (APS; ATCA), or high (HPS; ACCG) pain sensitive. Individuals carrying HPS/APS or APS/APS diplotypes were nearly 2.5 times as likely to develop TMJD. Previous data further suggest that *COMT* haplotypes code for differences in COMT enzymatic activity (*4*); however, the molecular mechanisms involved have remained unknown.

A common SNP in codon 158 ($val^{158}met$) has been associated with pain ratings and μ-opioid system responses (*7*) as well as addiction, cognition, and common affective disorders (*3*, *8*–*10*). The substitution of valine (Val) by methionine (Met) results in reduced thermostability and activity of the enzyme (*3*). It is generally accepted that $val^{158}met$ is the main source of individual variation in human COMT activity; numerous studies have identified associations between the low-activity *met* allele and several complex phenotypes (*3*, *8*, *10*). However, observed associations between these conditions and the *met* allele are often modest and occasionally inconsistent (*3*). This suggests that additional SNPs in the *COMT* gene modulate COMT activity. Indeed, we found that haplotype rather than an individual SNP better accounts for variability in pain sensitivity (*4*). The HPS and LPS haplotypes that both code for the stable val^{158} variant were associated with the two extreme-pain phenotypes; thus, the effect of haplotype on pain sensitivity in our study cannot be explained by the sum of the effects of functional SNPs. Instead, the $val^{158}met$ SNP interacts with other SNPs to determine phenotype. Because variation in the S-COMT promoter region does not contribute to pain phenotype (Fig. 1A), we suggest that the rate of mRNA degradation or protein synthesis is affected by the structural properties of the haplotypes, such as haplotype-specific mRNA secondary structure.

Previous reports have shown that polymorphic alleles can markedly affect mRNA secondary structure (*11*, *12*), which can then have functional consequences on the rate of mRNA degradation (*11*, *13*). It is also plausible that polymorphic alleles directly modulate protein translation through alterations in mRNA secondary structure, because protein translation efficiency is affected by mRNA secondary structure (*14*–*16*). To test these possibilities, we evaluated the affect of LPS, APS, and HPS haplotypes on the stability of the corresponding mRNA secondary structures (*17*).

Secondary structures of the full-length LPS, APS, and HPS mRNA transcripts were predicted by means of the RNA Mfold (*18*, *19*) and Afold (*20*) programs. The mRNA folding analyses demonstrated that the major *COMT* haplotypes differ with respect to mRNA secondary structure. The LPS haplotype codes for the shortest, least stable local stem-loop structure, and the HPS haplotype codes for the longest, most stable local stem-loop structure in the val^{158} region for both *S-COMT* and *MB-COMT*. Gibbs free energy (ΔG) for the stem-loop structure associated with the HPS haplotype is ~17 kcal/mol less than that associated with the LPS haplotype for both *S-COMT* and *MB-COMT* (Fig. 1B and fig. S1A). Additional evidence supporting predicted RNA folding structures was obtained by generating consensus RNA secondary structures based on comparative analysis of *COMT* sequences from eight mammalian species (fig. S2). The consensus RNA folding structures were LPS-like and did not contain highly stable local stem-loop structures analogous to the human HPS-like form. Thus, substantial deviation from consensus structure, as observed for the HPS haplotype, should have notable functional consequences. Additional studies were conducted to test this molecular modeling.

We constructed full-length S- and MB-COMT cDNA clones in mammalian expression vectors that differed only in three nucleotides corresponding to the LPS, APS, and HPS haplotypes (*17*, *21*). Rat adrenal (PC-12) cells were transiently transfected with each of these six constructs. COMT enzymatic activity, protein expression, and mRNA abundance were measured. Relative to the LPS haplotype, the HPS haplotype showed a 25- and 18-fold reduction in enzymatic activity for S- and MB-COMT constructs, respectively (Fig. 1C and fig. S1B). The HPS haplotype also exhibited marked reductions in S- and MB-COMT protein expression (Fig. 1D and fig. S1C). The APS haplotype displayed a moderate 2.5- and 3-fold reduction in enzymatic activity for S- and MB-COMT constructs, respectively, while pro-

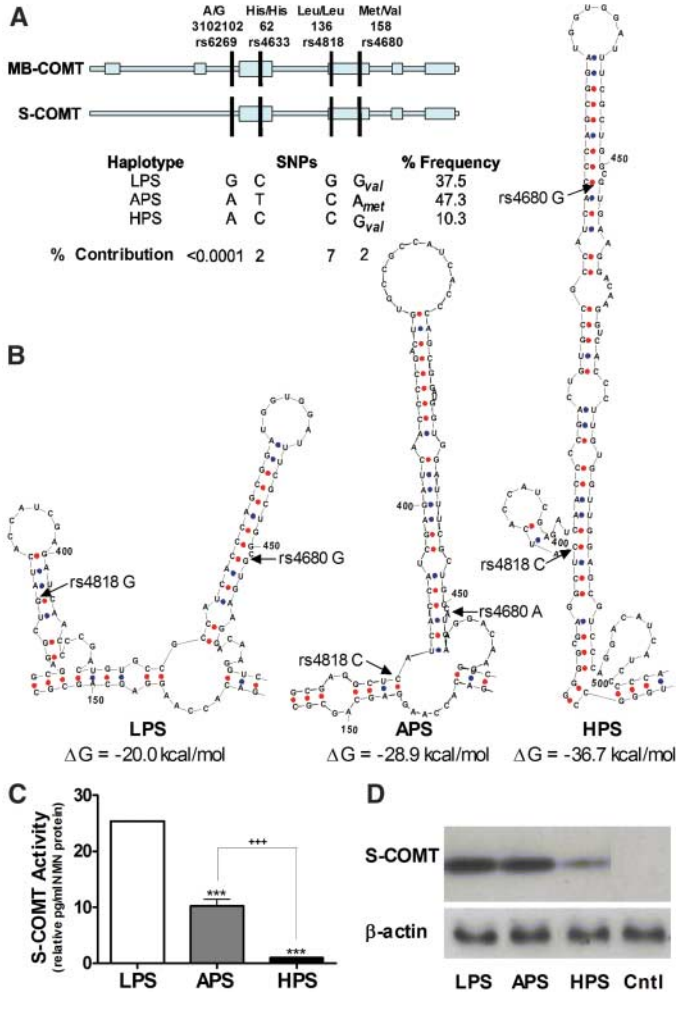

Fig. 1. Common haplotypes of the human *COMT* gene differ with respect to mRNA secondary structure and enzymatic activity. (**A**) A schematic diagram illustrates *COMT* genomic organization and SNP composition for the three haplotypes. Percent frequency of each haplotype in a cohort of healthy Caucasian females, and percent independent SNP contribution to pain sensitivity, are indicated. (**B**) The local stem-loop structures associated with each of the three haplotypes are shown. Relative to the LPS and APS haplotypes, the HPS local stem-loop structure had a higher folding potential. (**C** and **D**) The LPS haplotype exhibited the highest, while the HPS haplotype exhibited the lowest enzymatic activity and protein levels in cells expressing COMT. $^{***}P < 0.001$, ≠ LPS. $^{+++}P < 0.001$, ≠ APS.

tein expression levels did not differ. The moderate reduction in enzymatic activity produced by the APS haplotype is most likely due to the previously reported decrease in protein thermostability coded by the *met*^{158} allele (*3*). These data illustrate that the reduced enzymatic activity corresponding to the HPS haplotype is paralleled by reduced protein levels, an effect that could be mediated by local mRNA secondary structure at the level of protein synthesis and/or mRNA degradation. Because total RNA abundance and RNA degradation rates did not parallel COMT protein levels (fig. S2), differences in protein translation efficiency likely results from differences in the local secondary structure of corresponding mRNAs.

To directly assess this hypothesis, we performed site-directed mutagenesis (*17*). The stable stem-loop structure of *S-* and *MB-COMT* mRNA corresponding to the HPS haplotype is supported by base pairs between several critical nucleotides, including 403C and 479G in *S-COMT* and 625C and 701G in *MB-COMT* (Fig. 2A and fig. S3A). Mutation of 403C to G in *S-COMT* or 625C to G in *MB-COMT* destroys the stable stem-loop structure and converts it into a LPS haplotype–like structure (HPS Lsm). Double mutation of mRNA in position 403C to G and 479G to C in *S-COMT* or 625C to G and 701G to C in *MB-COMT* reconstructs the original long stem-loop structure (HPS dm). The single- and double-nucleotide HPS mutants (HPS Lsm and HPS dm, respectively) were transiently transfected to PC-12 cells. As predicted by the mRNA secondary-structure folding analyses, the HPS Lsm exhibited increased COMT enzymatic activity and protein levels equivalent to those of the LPS haplotype, whereas the HPS dm exhibited reduced enzymatic activity and protein levels equivalent to those of the original HPS haplotype (Fig. 2, B and C; fig. S3, B and C). These data rule out the involvement of RNA sequence recognition motifs or codon usage in the regulation of translation. In contrast to the HPS haplotype, protein levels did not parallel COMT enzymatic activity for the APS haplotype and site-directed mutagenesis confirmed that the *met*^{158} allele, not a more stable mRNA secondary structure, drives the reduced enzymatic activity observed for the APS haplotype (fig. S4). This difference is moderate relative to the mRNA structure–dependent difference coded by LPS and HPS haplotypes. These results were verified by an alternate approach of modifying mRNA secondary structure (fig. S5).

Our data have very broad evolutionary and medical implications for the analysis of variants common in the human population. The fact that alterations in mRNA secondary structure resulting from synonymous changes have such a pronounced effect on the level of protein expression emphasizes the critical role of synonymous nucleotide positions in maintaining mRNA secondary structure and suggests that the mRNA secondary structure, rather than independent nucleotides in the synonymous positions, should undergo substantial selective pressure (*22*). Furthermore, our data stress the importance of synonymous SNPs as potential functional variants in the area of human medical genetics. Although nonsynonymous SNPs are believed to have the strongest impact on variation in gene function, our data clearly demonstrate that haplotypic variants of common synonymous SNPs can have stronger effects on gene function than nonsynonymous variations and play an important role in disease onset and progression.

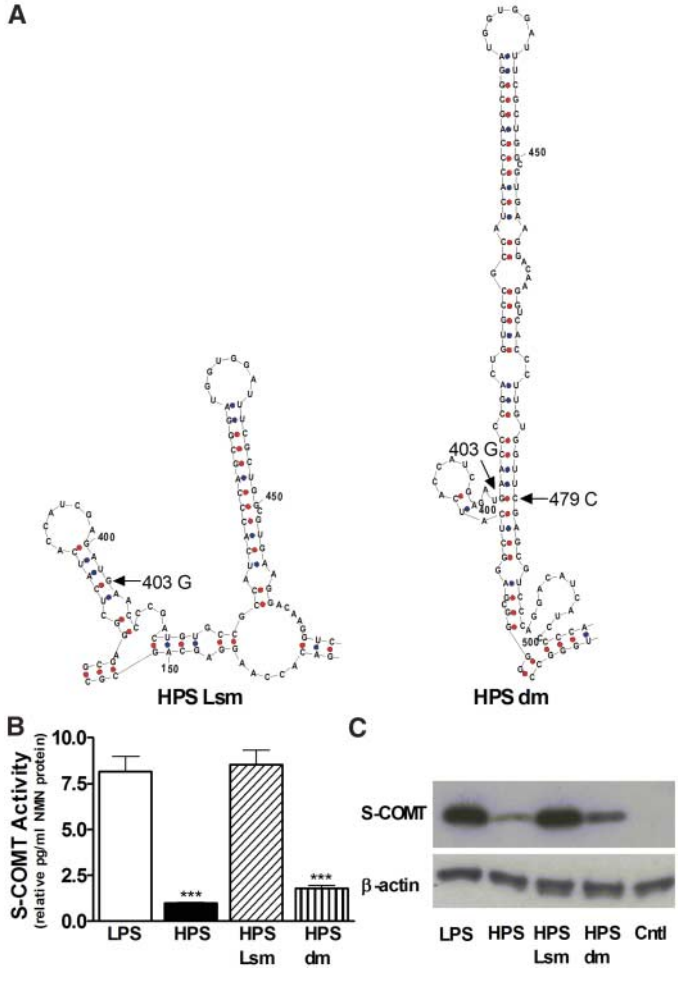

Fig. 2. Site-directed mutagenesis that destroys the stable stem-loop structure corresponding to the HPS haplotype restores COMT enzymatic activity and protein expression. (**A**) The mRNA structure corresponding to the HPS haplotype was converted to an LPS haplotype-like structure (HPS Lsm) by single mutation of 403C to G. The original HPS haplotype structure (HPS dm) was restored by double mutation of interacting nucleotides 403C to G and 479G to C. (**B** and **C**) The HPS Lsm exhibited COMT enzymatic activity and protein levels equivalent to those of the LPS haplotype, whereas the HPS dm exhibited reduced enzymatic activity. ***$P < 0.001$, ≠ LPS.

References and Notes

1. L. Y. Yampolsky, F. A. Kondrashov, A. S. Kondrashov, *Hum. Mol. Genet.* **14**, 3191 (2005).
2. J. C. Knight, *J. Mol. Med.* **83**, 97 (2005).
3. P. T. Mannisto, S. Kaakkola, *Pharmacol. Rev.* **51**, 593 (1999).
4. L. Diatchenko *et al.*, *Hum. Mol. Genet.* **14**, 135 (2005).
5. J. J. Marbach, M. Levitt, *J. Dent. Res.* **55**, 711 (1976).
6. T. T. Rakvag *et al.*, *Pain* **116**, 73 (2005).
7. J. K. Zubieta *et al.*, *Science* **299**, 1240 (2003).
8. G. Winterer, D. Goldman, *Brain Res. Brain Res. Rev.* **43**, 134 (2003).
9. B. Funke *et al.*, *Behav. Brain Funct.* **1**, 19 (2005).
10. G. Oroszi, D. Goldman, *Pharmacogenomics* **5**, 1037 (2004).
11. J. Duan *et al.*, *Hum. Mol. Genet.* **12**, 205 (2003).
12. L. X. Shen, J. P. Basilion, V. P. Stanton Jr., *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7871 (1999).
13. I. Puga *et al.*, *Endocrinology* **146**, 2210 (2005).
14. K. Mita, S. Ichimura, M. Zama, T. C. James, *J. Mol. Biol.* **203**, 917 (1988).
15. T. D. Schmittgen, K. D. Danenberg, T. Horikoshi, H. J. Lenz, P. V. Danenberg, *J. Biol. Chem.* **269**, 16269 (1994).
16. A. Shalev *et al.*, *Endocrinology* **143**, 2541 (2002).
17. Materials and methods are available as supporting material on *Science* Online.
18. D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, *J. Mol. Biol.* **288**, 911 (1999).
19. M. Zuker, *Nucleic Acids Res.* **31**, 3406 (2003).
20. A. Y. Ogurtsov, S. A. Shabalina, A. S. Kondrashov, M. A. Roytberg, *Bioinformatics* **22**, 1317 (2006).
21. Sequence accession numbers: S-COMT clones BG290167, CA489448, and BF037202 represent LPS, APS, and HPS haplotypes, respectively. Clones corresponding to all three haplotypes and including the transcriptional start site were constructed by use of the unique restriction enzyme Bsp MI. The gene coding region containing all three SNPs was cut and inserted into the plasmids through use of the S-COMT (BG290167) and MB-COMT (BI835796) clones containing the entire COMT 5′ and 3′ ends as a backbone vector.
22. S. A. Shabalina, A. Y. Ogurtsov, N. A. Spiridonov, *Nucleic Acids Res.* **34**, 2428 (2006).

23. We are grateful to the National Institute of Child Health and Human Development, National Institute of Neurological Disorders and Stroke, and National Institute of Dental and Craniofacial Research at the NIH for financial support of this work. This research was also supported by the Intramural Research Program of the NIH, National Center for Biotechnology Information.

Supporting Online Material
www.sciencemag.org/cgi/content/full/314/5807/1930/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
References

13 June 2006; accepted 16 November 2006
10.1126/science.1131262

Lineages of Acidophilic Archaea Revealed by Community Genomic Analysis

Brett J. Baker,[1] Gene W. Tyson,[2] Richard I. Webb,[3] Judith Flanagan,[2]* Philip Hugenholtz,[2]† Eric E. Allen,[2]‡ Jillian F. Banfield[1,2]§

Novel, low-abundance microbial species can be easily overlooked in standard polymerase chain reaction (PCR)–based surveys. We used community genomic data obtained without PCR or cultivation to reconstruct DNA fragments bearing unusual 16*S* ribosomal RNA (rRNA) and protein-coding genes from organisms belonging to novel archaeal lineages. The organisms are minor components of all biofilms growing in pH 0.5 to 1.5 solutions within the Richmond Mine, California. Probes specific for 16*S* rRNA showed that the fraction less than 0.45 micrometers in diameter is dominated by these organisms. Transmission electron microscope images revealed that the cells are pleomorphic with unusual folded membrane protrusions and have apparent volumes of <0.006 cubic micrometer.

Our understanding of the variety of microorganisms that populate natural environments was advanced by the development of polymerase chain reaction (PCR)–based, cultivation-independent methods that target one or a small number of genes (*1*–*3*). Genomic analyses of DNA sequence fragments derived from multispecies consortia (*4*–*6*) and whole environments (*7*–*9*) have provided new information about diversity and metabolic potential. However, PCR-based methods have limited ability to detect organisms whose genes are significantly divergent relative to gene sequences in databases, and most cultivation-independent genomic sequencing approaches are relatively insensitive to organisms that occur at low abundance. Consequently, it is likely that low-abundance microorganisms distantly related to known species will be undetected members of natural consortia, even in low complexity systems such as acid mine drainage (AMD) (*10*).

An important way in which microorganisms affect geochemical cycles is by accelerating the dissolution of minerals. For example, microorganisms can derive metabolic energy by oxidizing iron released by the dissolution of pyrite (FeS_2). The ferric iron by-product promotes further pyrite dissolution, leading to AMD generation. AMD solutions forming underground in the Richmond Mine at Iron Mountain, California, are warm (30° to 59°C), acidic (pH ~0.5 to 1.5), metal-rich [submolar Fe^{2+} and micromolar As and Cu (*11*)] and host active microbial communities. Extensive cultivation-independent sequence analysis of functional and rRNA genes (*11*, *12*) revealed that biofilms contain a significant number of Archaea, but the diversity reported to date has been limited to the order Thermoplasmatales (*10*). Current models for AMD generation thus include only these species.

The genomes of the five dominant members of one biofilm community from the "5-way" region of the Richmond Mine (fig. S1) were largely reconstructed through the assembly of 76 Mb of shotgun genomic sequence (*4*). Previously unreported is a genome fragment that encodes part of the 16*S* rRNA gene of a novel archaeal lineage: Archaeal Richmond Mine Acidophilic Nanoorganism (ARMAN-1). Using an expanded data set that now comprises more than 100 Mb of genomic sequence, we reconstructed a contiguous 4.2-kb fragment adjacent to this gene. A second 13.2-kb genome fragment encoding a 16*S* rRNA gene from an organism that is related to ARMAN-1 (ARMAN-2) was reconstructed from 117 Mb of community genomic sequence derived from a biofilm from the A drift (fig. S1). Within the data sets from each site, results to date indicate that each ARMAN population is near-clonal.

Comparison of the ARMAN-1 and -2 DNA fragments revealed some gene rearrangements, insertions, and deletions (Fig. 1). Genes present in both organisms encode putative inorganic pyrophosphatases, a transcription regulator, and a gene shown to be an arsenate reductase (*13*). Comparative analysis of these genes with sequences in the public databases consistently

[1]Department of Earth and Planetary Sciences, University of California, Berkeley, CA 94720, USA. [2]Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA. [3]Centre for Microscopy and Microanalysis and Department of Microbiology and Parasitology, University of Queensland, Brisbane 4072, Australia.

*Present address: University of California, San Francisco, CA 94143, USA.
†Present address: Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA.
‡Present address: University of California, San Diego, La Jolla, CA 92093, USA.
§To whom correspondence should be addressed. E-mail: jill@eps.berkeley.edu

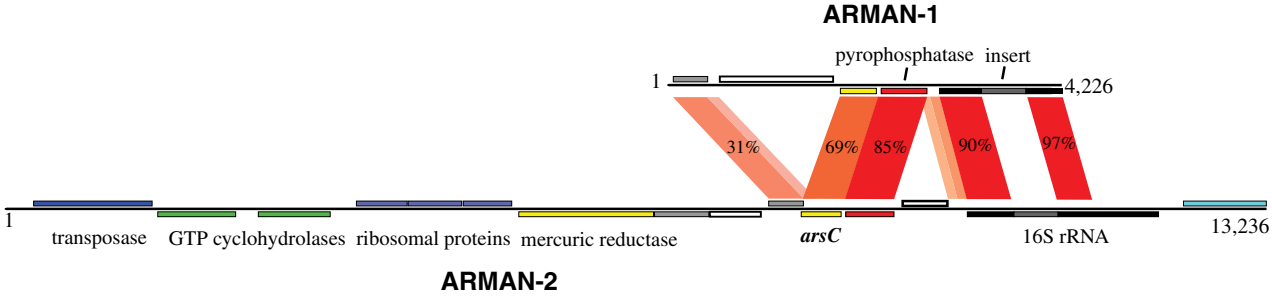

Fig. 1. Comparison of syntenous genomic regions of ARMAN-1 [from the "5-way" (CG) community (*4*)] and ARMAN-2 (from the UBA community). Orthologs and their protein identity are indicated by the red bands. The percentage similarity for the 16*S* rRNA gene sequences is also shown. Numbers at ends indicate length (number of nucleotides); predicted open reading frames for hypothetical proteins are indicated by boxes.

Supporting Online Material for

Human Catechol-*O*-Methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure

A. G. Nackley, S. A. Shabalina, I. E. Tchivileva, K. Satterfield, O. Korchynskyi, S. S. Makarov, W. Maixner, L Diatchenko*

*To whom correspondence should be addressed. E-mail: lbdiatch@email.unc.edu

Published 22 December 2006, *Science* **314**, 1930 (2006)
DOI: 10.1126/science.1131262

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 to S6
References

Supporting Online Material for

Common Human Catechol-O-methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure

AG Nackley, SA Shabalina, IE Tchivileva, K Satterfield, O Korchynskyi, SS Makarov, W Maixner, and L Diatchenko*

*To whom correspondence should be addressed. Email: lbdiatch@email.unc.edu

This PDF file includes:

Materials and Methods
SOM Text
figs. S1 to S6
References

Supporting Online Material

Supporting Materials and Methods

Association of Common COMT Variants with Pain Sensitivity

Please see (1) for a detailed description of the analyses of associations between SNPs, haplotypes, and pain responsiveness and related procedures.

Prediction of RNA Secondary Structure

Secondary structures of the full length LPS, APS, HPS, HPS Asm, HPS Lsm, and HPS Ldm mRNA transcripts were predicted using the RNA Mfold program (versions 3.1 and 3.2) (*1*) and the Afold program (*2*). Energy minimization was performed by a dynamic programming method that finds the secondary structure with the minimum free energy with sums comprised of stacking, loop length, etc. (*2*). The RNA folding parameters were developed and published by the Turner group (*3*). Suboptimal stem-loop structures were analyzed by the Hybrid program (*4*) for the full-length COMT transcripts, and for truncated transcript sequences of different lengths ranging from the rs4680 region to the end of the transcript with 100 nt window length.

Assessment of COMT Enzyme Activity, Protein Expression, and RNA Expression
Construction of equal length COMT plasmids. Full-length pCMV-SPORT6-based cDNA clones corresponding to the three COMT haplotypes were obtained from the IMAGE clone collection (Open Biosystems, Huntsville, AL, USA). S-COMT clones BG290167, CA489448, and BF037202 represent LPS, APS, and HPS haplotypes, respectively. Clones corresponding to all three haplotypes and including the transcriptional start site were constructed using the unique restriction enzyme BSPMI. The CDS gene region containing all three SNPs were cut and inserted into the plasmids using the S-COMT (BG290167) and MB-COMT (BI835796) clones containing the entire COMT 5' and 3' ends as a backbone vector. Plasmid DNA was purified using the EndoFree Plasmid Maxi purification kit (Qiagen, Germantown, MD, USA). Once plasmids were isolated, DNA sequences were confirmed by double sequencing at the UNC core sequencing facility.

Transient transfection of COMT cDNA clones. A rat adrenal cell line (PC12) was transiently transfected into six-well plates using SuperFect Reagent (Qiagen) in accordance in manufacture's recommendations. The amount of plasmid was kept at 2 µg/well. The amount of control cDNA plasmid for transfection efficiency (pSV-βGalactosidase vector; Promega, Madison, WI, USA and SEAP; Clonetech, Mountain View, CA, USA) was kept at 0.1 µg. Transfection with the vector lacking the insert was done for each experiment. Cells lysates were collected approximately 40 hours post-transfection.

Enzymatic assay. After removing the media, cells were washed twice with 0.9% saline solution (1 ml/35 mm well) and then covered with deionized water containing 10 mM CDTA (300 µl/35 mm well). The wells were freeze/thawed (-80°C/RT) five times and the lysate collected in 1.7 ml tubes. The tubes were centrifuged at 2000 g for 10 min and filtrate removed. The enzymatic COMT assay was based on the method described by Masuda's group (*5*). Purified lysates (8 µl) were incubated with 200 µM Sadenosyl-L-methionine (SAMe; ICN Chemicals, Aurora OH, USA), 7.5 mM L-norepinephrine (NE; Sigma Chemical Co., St. Louis MO, USA) and 2mM MgCl2 in 50 mM phosphate buffered saline for 60 min in the final volume of 22 µl. The reaction was terminated using 20 µl of 0.4 M hydrocholoric acid and 1 µl of 330 mM EDTA. The same reaction in the presence of 15 mM EDTA was carried out in pararlel for

each lysate to bind Mg^{+2} ions required for COMT activity. COMT activity was assessed as measurement of normetanephrine (NMN) by Normetanephrine ELISA kit (IBL, Hamburg, Germany) in accordance with manufacture's recommendations using 10 µl of above reaction mixture. COMT activity was determined after subtracting the amount of NMN produced by endogenous enzymatic activity (transfection with empty vector). The non-specific background was determined in parallel assays performed in the presence of EDTA and then subtracted from each reading. COMT activity was then normalized for transfection efficiency by measuring the β-galactosidase activity for each lysate. βgalactosidase activity was determined using β-galactosidase enzyme systems (Promega), according to the supplier protocol.

Western blot. Purified lysates, normalized for protein content using a BCA assay, were run on 10% Novex Tris-Glycine gels (Invitrogen) and transferred to nitrocellulose membranes (Whatman, Florham Park, NJ, USA). Blots containing COMT protein were blocked with 5% NF milk for 30 min at RT, incubated with COMT polyclonal 1° antibody (1:10,000; Chemicon, Temecula, CA, USA) o/n at 4°C, and then incubated with Goat Anti-Rabbit IgG HRP polyclonal 2° antibody (1:10,000; Chemicon) for 1 hr at RT. Blots were washed with PBST for 10 min at RT, exposed to chemiluminescence reagent (Pierce, Milwaukee, WI, USA), and developed. Blots were then stripped using Restore western stripping buffer (Pierce) and equal loading of samples verified by β-actin staining. Blots were incubated with β-actin polyclonal 1° antibody (1:10,000; Santa Cruz Biotechnology, Santa Cruz, CA, USA) for 1 hr at RT followed by Goat Anti-Rabbit IgG HRP polyclonal 2° antibody (1:10,000; Chemicon) for 1 hr at RT and chemiluminescent reagent.

Real time-PCR. Total RNA was isolated using the Trizol reagent (Invitrogen, Carlsbad, CA, USA). The isolated RNA was treated with RNase free-DNase I (Promega) and reverse transcribed by Thermo-X reverse transcriptase (Invitrogen). The cDNA for COMT and SEAP was amplified with DyNAmo-SYBRGreen qPCR kit (MJ Research) using forward and reverse PCR primers (TGAACGTGGGCGACAAGAAAGGCAAGAT and TGACCTTGTCCTTCACGCCAGCGAAAT, respectively, for COMT and GCCGACCACTCCCACGTCTT and CCCGCTCTCGCTCTCGGTAA, respectively, for SEAP). Opticon-2 Real Time Fluorescence Detection System (MJ Research, Reno, NV, USA) was used for measuring fluorescence.

Assessment of mRNA degradation rates. The time course of mRNA degradation was measured after actinomyicin D (actD; Sigma) treatment. 36 hours after transfection of COMT and SEAP plasmids, cells were treated with actD (10 µg/ml) and collected 0, 2, 4, or 8 hours post treatment. Total RNA was isolated using Trizol and then processed for real time-PCR.

Alteration of mRNA secondary structure. To make the HPS LPS-like single mutant and the double mutant, site-directed mutagenesis was performed using the QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene, LaJolla, CA, USA) according to the manufacturer's instructions. The LPS-like single mutant was generated using forward and reverse primers containing the 403C to G mutation (CACCATCGAGAT**G**AACCCCGACTGTG and CACAGTCGGGGTT**C**ATCTCGATGGTG, respectively; MWG, High Point, NC, USA). The second mutation was then introduced in the single nucleotide mutant using forward and reverse primers containing the 479G to C mutation (CACCCTTGTGGTT**C**GAGCGTCCCAGG and CCTGGGACGCTC**G**AACCACAAGGGTG, respectively; MWG). The 403C to G mutation results in an amino acid change from serine (ser) to threonine (thr). Although there were no synonymous substitutions that would destroy the HPS RNA secondary structure, ser and thr belong to the same class of small intermediately charged amino acids. Furthermore, the

double mutant RNA carries the same thr residue and thus serves as a control for any affect of amino acid substitution.

To make the HPS APS-like single nucleotide mutant, a 78 nucleotide double-stranded oligonucleotide containing the 533C to G substitution (AGGTCACCCTTGTGGTTGGAGCGTCCCAGGACATCAT CCCCCAGCTGAAGAAGAAGTATGATGTGGACACA**G**TGGACA; Integrated DNA Technologies, Coralville, IA, USA) and its reverse compliment were annealed and then inserted into the expression vector with the HPS construct into a PflFI- PflFI site, replacing the original sequence.

To construct cDNA clones with truncated 5' and 3' untranslated regions (UTRs) , the PCR amplified COMT fragments corresponding to the CDS region were inserted into the pcDNA3 vector using BamH1 (New England Biolabs, Ipswich, MA, USA) and Xba1 (New England Biolabs) sites. The forward primers used for amplification of *S-COMT* and *MB-COMT* were AAAGGATCCGGGCCTGGTGCTGCTGGT and AAAGGATCCGCCACCGCCATTGCCGCC, respectively. The reverse primer used was AAATCTAGATCACTTGTCATCATCGTCCTTGTAGTCGGGCCCTGCTTCGCTGCCT.

Statistical analyses. The standard curve for ELISA was determined using a 4 parameter sigmoidal dose-response model for one-site competitive binding systems. In order to include data from our 0 concentration sample, the value was approximated as 2 log units below the lowest non-0 concentration. COMT activity data were analyzed by ANOVA and Tukey's Multiple Comparison post hoc tests. For measures of RNA abundance and degradation, *COMT* RNA was normalized to SEAP and then analyzed by ANOVA and Tukey's Multiple Comparison post hoc tests. $P < 0.05$ was considered significant.

Supporting Notes

Association between COMT haplotypes and pain sensitivity

The HPS and LPS haplotypes that both code for the stable v*al*^{158} variant were associated with the two extreme pain phenotypes; thus, the effect of haplotype on pain sensitivity in our study cannot be explained by the sum of the effects of functional SNPs. Instead, the *val*^{158}*met* SNP interacts with other SNPs to determine phenotype. This allelelic interaction could lead to low enzyme activity by coding for reductions in mRNA transcription efficiency, mRNA stability, or protein translation efficiency. An alteration in mRNA expression is unlikely, however, because SNPs located in the *S-COMT* promoter region (rs6269 and rs4633) do not contribute independently to pain sensitivity (Fig. 1A; see also (*6*)). Furthermore, SNP rs4633 is not associated with strong alterations in RNA abundance (*7*). This implies that an interaction between allelic variants in the coding region of the *COMT* RNA affects gene function.

Prediction of mRNA secondary structures corresponding to COMT haplotypes

Secondary structures of the full length LPS, APS, and HPS mRNA transcripts were predicted using the RNA Mfold (*1, 3*) and Afold (*2*) programs. mRNA folding analyses demonstrated that the major *COMT* haplotypes differ with respect to mRNA secondary structure. The LPS haplotype codes for the shortest, least stable local stem-loop structure and the HPS haplotype codes for the longest, most stable local stem-loop structure in the *val*^{158} region for both *S-COMT* and *MB-COMT* (Fig. 1B and fig. S1A). These structures were found in 80%, 67% and 73% of all suboptimal S-COMT mRNA structures for LPS, APS and HPS haplotypes,

respectively.

Assessment of enzymatic activity, protein expression, and RNA abundance

Relative to the LPS haplotype, the HPS haplotype showed a 25-fold and 18-fold reduction in enzymatic activity for S- and MB-COMT constructs, respectively (Fig. 1C and fig. S1B). The HPS haplotype also exhibited dramatic reductions in S- and MB-COMT protein expression (Fig.1D and fig. S1C). The APS haplotype displayed a moderate 2.5-fold and 3-fold reduction in enzymatic activity for S- and MB-COMT constructs, respectively, while protein expression levels did not differ (Fig. 1C, D and fig. S1B, C). No differences in total RNA abundance were observed for the S- or MB-COMT clones (fig. S2A, B). These data illustrate that the reduced enzymatic activity corresponding to the HPS haplotype is paralleled by reduced protein levels, an effect that could be mediated by local mRNA secondary structure at the level of protein synthesis and/or mRNA degradation.

Assessment of mRNA degradation rates

While producing cumulative differences in protein levels, the efficient transcription driven by the CMV promoter likely masks the changes in degradation rates of exogenous haplotype-specific *COMT* RNA. Thus, we measured mRNA degradation in the presence of actinomycin D (actD), a nonspecific inhibitor of transcription. The rate of mRNA degradation was significantly different among *S*-and *MB-COMT* mRNA corresponding to the three haplotypes, however, in the opposite direction to that observed for activity and protein levels. The maximal differences occurred within 2-4 hours following actD treatment: ~70% mRNA remained for the HPS haplotype, while ~34% and ~45% mRNA remained for the APS and LPS haplotypes, respectively (fig. S2C, D). Consequently, the low protein levels observed for the HPS construct do not result from an increase in the rate of mRNA degradation. These data suggest that although the HPS haplotype codes for less efficient translation, it also codes for a more stable transcript, consistent with the hypothesis that the more stable local mRNA secondary structure prevents both ribosome binding/translocation and accessibility to exonucleases (*8, 9*).

Site-directed mutagenesis of the HPS haplotype

To directly assess if differences in protein translation efficiency result from differences in local secondary structure of the corresponding mRNAs, site-directed mutagenesis was performed. The stable stem-loop structure of *S*-and *MB-COMT* mRNA corresponding to the HPS haplotype is supported by base pairs between several critical nucleotides, including 403C and 479G in *S-COMT* and 625C and 701G in *MB-COMT* (Fig. 2A and fig. S3A). Mutation of 403C to G in *S-COMT* or 625 C to G in *MB-COMT* destroys the stable stem-loop structure and converts it into a LPS haplotype-like structure (HPS Lsm). Double mutation of mRNA in position 403C to G and 479G to C in *S-COMT* or 625 C to G and 701 G to C in *MB-COMT* reconstructs the original long stem-loop structure (HPS dm). The single- and double-nucleotide HPS mutants (HPS Lsm and HPS dm, respectively) were transiently transfected to PC-12 cells. As predicted by the mRNA secondary structure folding analyses, the HPS Lsm exhibited increased COMT enzymatic activity and protein levels equivalent to the LPS haplotype, while the HPS dm exhibited reduced enzymatic activity and protein levels equivalent to the original HPS haplotype (Fig. 2B, C and fig. S3B, C).

In contrast to the HPS haplotype, protein levels did not parallel COMT enzymatic activity for the APS haplotype. Thus, the moderate reduction in enzymatic activity exhibited by the APS haplotype, possessing the *met*^{158} allele, most likely results from the previously reported decrease in enzyme thermostability rather than altered secondary structure of the corresponding mRNA. To test this hypothesis, we employed site-directed mutagenesis of the

HPS clone. Mutation of 533C to G destroys the stable stem-loop structure and converts it into an APS haplotype-like structure (HPS Asm), without changing the amino acid (fig. S4A). The corresponding single-nucleotide mutant and the original LPS, APS and HPS clones were transiently transfected to PC-12 cells. Cells transfected with the HPS Asm exhibited enzymatic activity and protein levels that did not differ from cells transfected with the LPS clone (fig. S4B, C). Both the HPS Asm and the LPS clone carry the *val*^{158} allele. Therefore, these results confirm that the *met*^{158} allele, not a more stable mRNA secondary structure, drives the reduced enzymatic activity observed for the APS haplotype.

Assessment of mRNA secondary structure using an alternate approach

To verify the role of *val*^{158}*met* vs. haplotype in mediating the biological activity of COMT, an alternate approach of modifying mRNA secondary structure was employed. As 5' and 3' untranslated regions (UTRs) are a significant element of the *COMT* haplotype-specific stem-loop structure, S- and MBCOMT constructs corresponding to LPS, APS, and HPS haplotypes were truncated at their UTRs, while the coding region was conserved. Folding analyses show that mRNA lacking 5' and 3' UTRs form identical secondary structures for each *S*- and *MB-COMT* haplotype (fig. S5A and data not shown, respectively). The truncated APS haplotype exhibited COMT enzymatic activity that was reduced 2- to 3-fold in cells expressing S- and MB-COMT, respectively, as compared to the LPS and HPS haplotypes. (fig. S5B, C). These results are consistent with previous findings that the low activity ^{158}*met* allele leads to a 3- to 4-fold reduction in human COMT activity (*10*). Protein levels did not differ for the modified S- and MB-COMT constructs corresponding to the three haplotypes (fig. S5D, E). Thus, we can firmly conclude that the APS haplotype, containing the *met*^{158} allele, reduces COMT enzymatic activity by altering protein stability. It is important to note, however, that this difference is moderate relative to the mRNA structure-dependent difference coded by LPS and HPS haplotypes.

Putative molecular mechanism(s) whereby mRNA secondary structures affect protein levels

Although the precise molecular mechanism whereby the different mRNA secondary structures produce different protein levels should be further investigated, several putative pathways may be proposed. The most obvious putative mechanism is the one involving differential effects at the ribosomal level. That is, more stable RNA secondary structures may produce less protein because of impeded ribosome binding or translocation (9). Alternatively, the secondary structures may affect interactions with distinct mRNA binding proteins that provide intracellular mRNA transport to translation site. For example, it has been shown that Staufen (stau) and fragile X mental retardation (FMRP) proteins that facilitate RNA transport, require specific mRNA secondary structures to bind (10,11). These effects may be of particular relevance for the COMT mRNA in primary afferent neurons, where the number of mRNAs have to travel over very long distances to the sites of translation (12). Finally, we do not exclude that some other, yet to be identified mechanisms, may play a role.

Supporting Figures

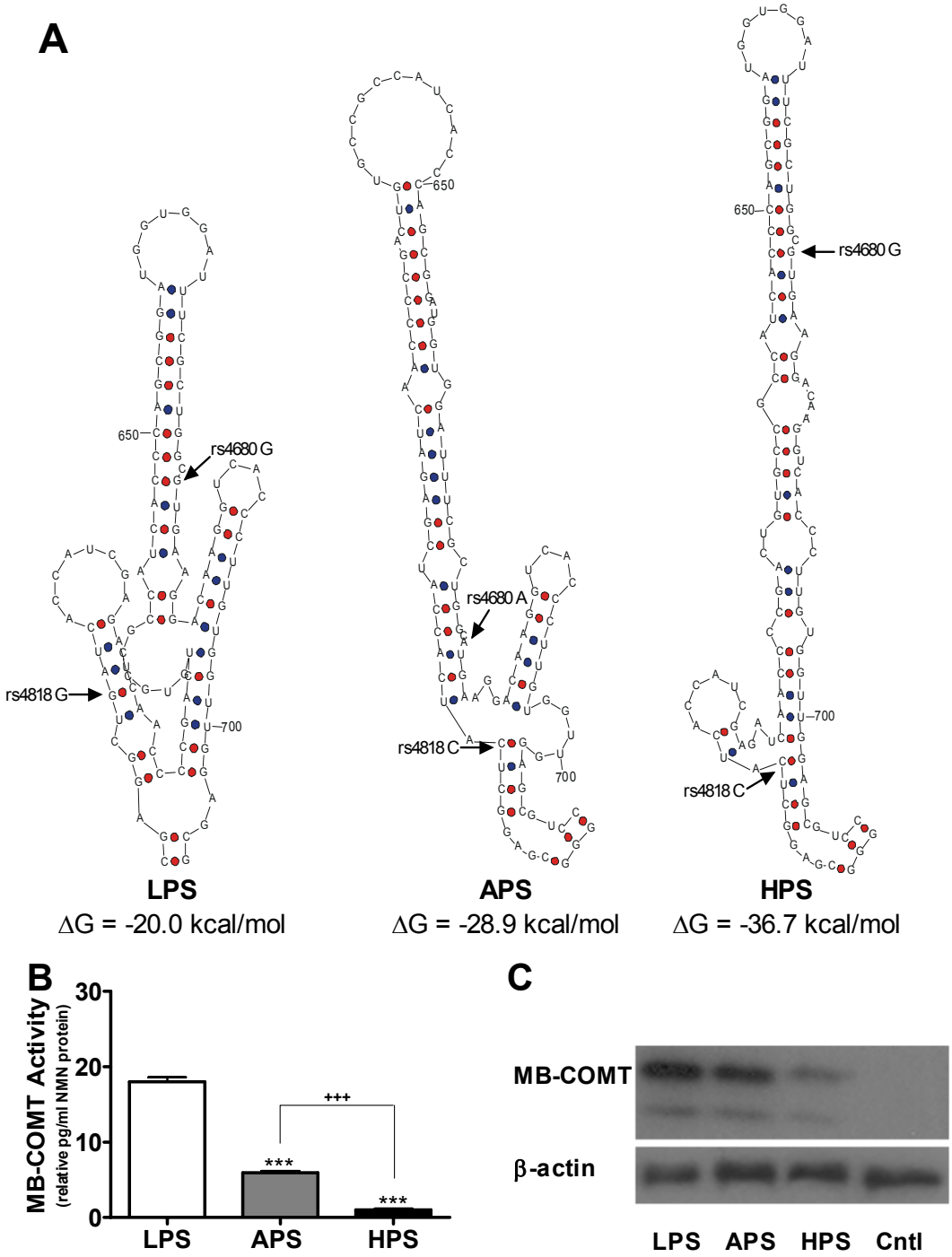

fig. S1 Common haplotypes of the human *COMT* gene that are associated with pain sensitivity differ with respect to mRNA secondary structure, enzymatic activity, and protein expression for *MB-COMT*. **(A)** RNA folding diagrams depict the calculated secondary structures of the *COMT* mRNAs corresponding to the three haplotypes. Arrows indicate the position of alternative alleles. ΔG = -20.0 kcal/mole, -28.9 kcal/mole, and -36.7 kcal/mole for local stem-loop structures corresponding to the LPS, APS, and HPS haplotype, respectively. **(B)** The LPS haplotype exhibited the highest, while the HPS haplotype exhibited the lowest enzymatic activity in cells expressing COMT. **(C)** Similarly, the HPS haplotype showed pronounced reductions in protein expression relative to the LPS haplotype. However, protein expression did not differ for the APS and LPS haplotypes. Data are expressed as Mean ± SEM. $^{***}P < 0.001$, ≠ LPS. $^{+++}P < 0.001$, ≠ APS.

A

```
Homo sapiens (LPS)   GI:113713957      GCUGGGGGCCUACUGUGGCUACUCAGCUGUGCGCAUGGCCCGCCUGCUGUCACCAGGGGC
Pan troglodytes      GI:55659344       GCUGGGGGCCUACUGUGGCUACUCAGCUGUGCGCAUGGCCCGCCUGCUGUCACCAGGGGC
Macaca mulatta       GI:109093289      GCUGGGGGCCUACUGUGGCUACUCAGCUGUGCGCAUGGCCCGCCUGCUGUCACCGGGUGC
Macaca fascicularis  GI:94977203       GCUGGGGGCCUACUGUGGCUACUCAGCUGUGCGCAUGGCCCGCCUGCUGUCACCGGGUGC
Equus caballus       GI:58531100       GCUGGGGGCCUACUGCGGCUACUCGGCCGUGCGCAUGGCCCGCCUGCUGCCGCCCGGCGC
Bos taurus           GI:76639808       GCUGGGCGCCUACUGCGGCUACUCGGCAGUGCGAAUGGCCCGCCUGCUGCUGCCCGGCGC
Sus scrofa           GI:89753956       GCUGGGGGCUUACUGUGGCUACUCGGCUGUGCGCAUGGCCCGCCUGCUGCUGCCCAGCGC
Canis familiaris     GI:51890220       GCUGGGAGCCUACUGCGGCUACUCGGCCGUGCGCAUGGCCCGCCUGCUGGAGCCCGGGGC
                                       ****** ** ***** ******** ** ***** ***************   **  * **

Homo sapiens (LPS)   GI:113713957      GAGGCUGAUCACCAUCGAGAUCAACCCCGACUGUGCCGCCAUCACCCAGCGGAUGGUGGA
Pan troglodytes      GI:55659344       GAGGCUGAUCACCAUCGAGAUCAACCCCGACUGUGCCGCCAUCACCCAGCGGAUGGUGGA
Macaca mulatta       GI:109093289      GAGGCUGCUCACCAUCGAGAUCAACCCUGACUACGCCGCCAUCACCCAGCGGAUGGUGGA
Macaca fascicularis  GI:94977203       GAGGCUGCUCACCAUCGAGAUCAACCCUGACUACGCCGCCAUCACCCAGCGGAUGGUGGA
Equus caballus       GI:58531100       CCGUCUGCUCACCAUCGAGAUCAACCCUGACUAUGCCGCCAUCACCCAGCGGAUGCUGGA
Bos taurus           GI:76639808       CCGGCUGCUCACCAUCGAGUUCAAUCCCGACUAUGCCGCCAUCACCCAGCGGAUGGUGGA
Sus scrofa           GI:89753956       CCGGCUGCUGACCAUCGAGCUUAACCCUGACAAUGCUGCCAUCGCCCAGCAGGUGGUGGA
Canis familiaris     GI:51890220       CCGCCUGAUCACCAUCGAGCUCAACCCUGACUUCGCCGCCAUCACCCAGCAGAUGCUGGA
                                         * *** * ********* * ** ** ***   ** ****** ****** * ** ****

Homo sapiens (LPS)   GI:113713957      UUUCGCUGGCGUGAAGGACAAGGUCACCCUUGUGGUUGGAGCGUCCCAGGACAUCAUCCC
Pan troglodytes      GI:55659344       UUUCGCUGGCGUGAAGGACAAGGUCACCCUUGUGGUUGGAGCGUCCCAGGACAUCAUCCC
Macaca mulatta       GI:109093289      UUUCGCUGGCAUGCAGGACAAGGUCACCGUUGUGGUUGGGGCGUCCCAGGACAUCAUCCC
Macaca fascicularis  GI:94977203       UUUCGCUGGCAUGCAGGACAAGGUCACCCUUGUGGUUGGAGCGUCCCAGGACAUCAUCCC
Equus caballus       GI:58531100       CUUCGCGGGCCUGCAGGACAGGGUAACCGUUGUCCUCGGGGCCUCCCAGGACAUCAUCCC
Bos taurus           GI:76639808       GUUUGCAGGCCUGCAGGACAAGGUCACCGUUGUUCUUGGAGCAUCCCAGGACAUCAUCCC
Sus scrofa           GI:89753956       CUUCGCAGGCCUGCAGGACAGGGUGACCGUUGUCGUCGGGGCAUCCCAGGACAUCAUCCC
Canis familiaris     GI:51890220       CUUCGCAGGCCUGCAGGACAGGGUCACCAUUCUGACCGGGGCAUCCCAGGACACCAUCCC
                                        ** ** *** ** ****** *** *** ** *    ** ** ********** ******
```

B **C**

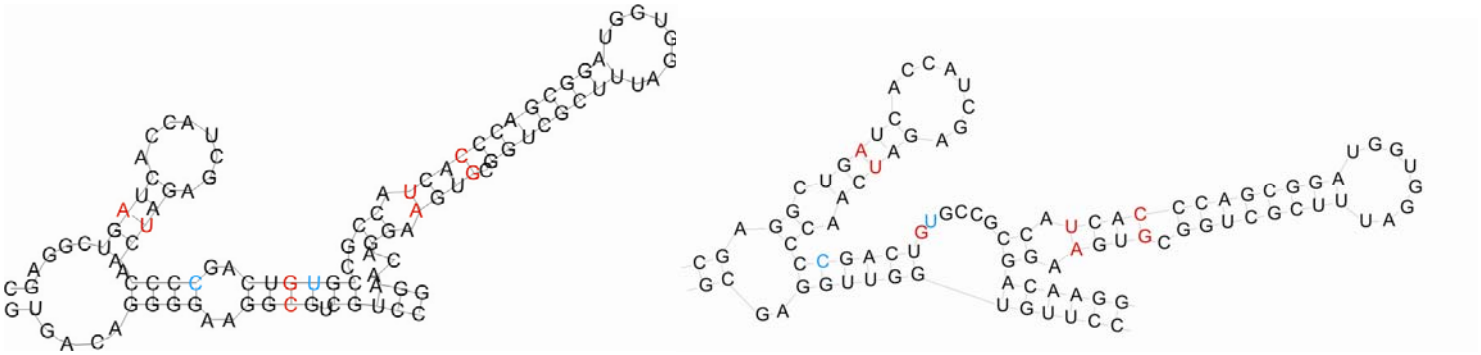

fig. S2 Multiple alignment (A) and c**onsensus *COMT* RNA secondary structures (B-C).** A fragment of multiple alignment of mammalian *COMT* sequences (A) was created by the program Muscle (*11*). The consensus RNA secondary structures were predicted using RNAalifold (B) and KNetFold programs (C) (*9, 10*). Both predicted structures for primates only (B) and mammalian (C) *COMT* sequences were LPS-like and did not contain highly stable local stem-loop structures observed for the human HPS haplotype. Multiple matches conserved for all sequences are indicated with asterisks. Nucleotide positions that are not paired in all aligned sequences are shown in red. Nucleotide positions that may be involved in Watson-Crick or wobble base pairing are shown in blue.

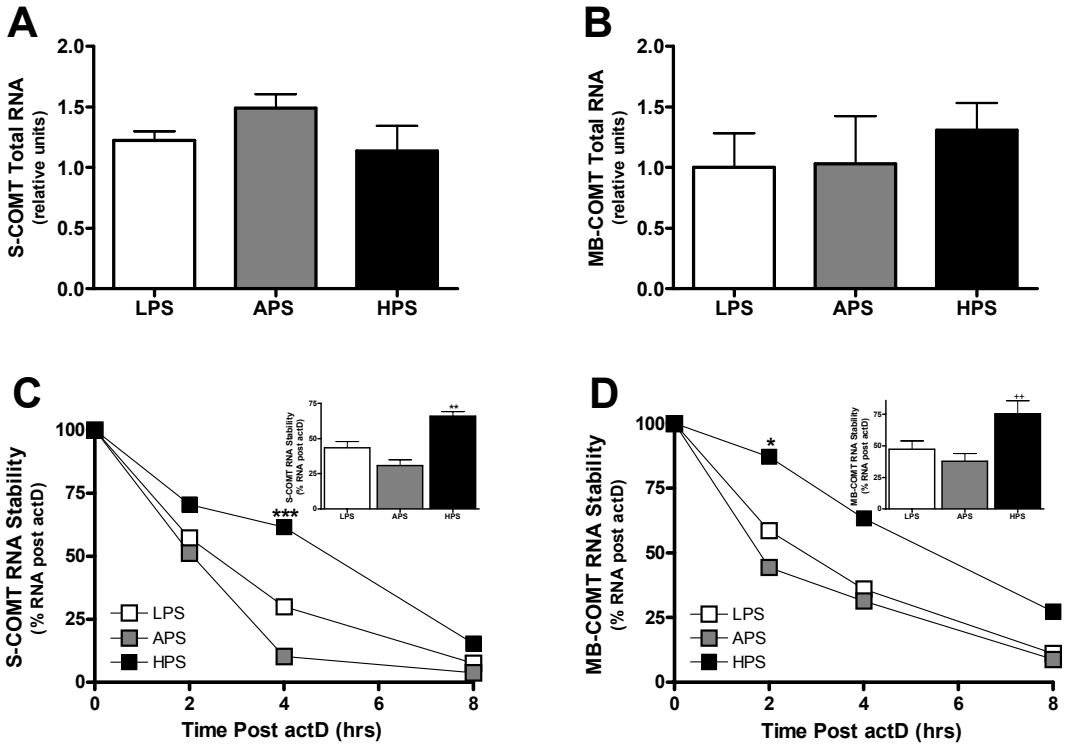

fig. S3 The HPS haplotype codes for reduced RNA degradation rates. **(A, B)** Total RNA levels were approximately the same for cells transfected with clones corresponding to the LPS, APS, and HPS haplotypes. **(C, D)** Following actD treatment, the rate of mRNA degradation was significantly less for the HPS haplotype in cells expressing S-COMT or MB-COMT. Maximal group differences were observed at 2 and 4 hours post actD treatment; mRNA corresponding to the HPS haplotype was nearly twice as stable as that corresponding to the LPS and APS haplotypes. Data are expressed as Mean $\pm$ SEM. $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05 \neq$ LPS. $^{++}P < 0.01 \neq$ APS.

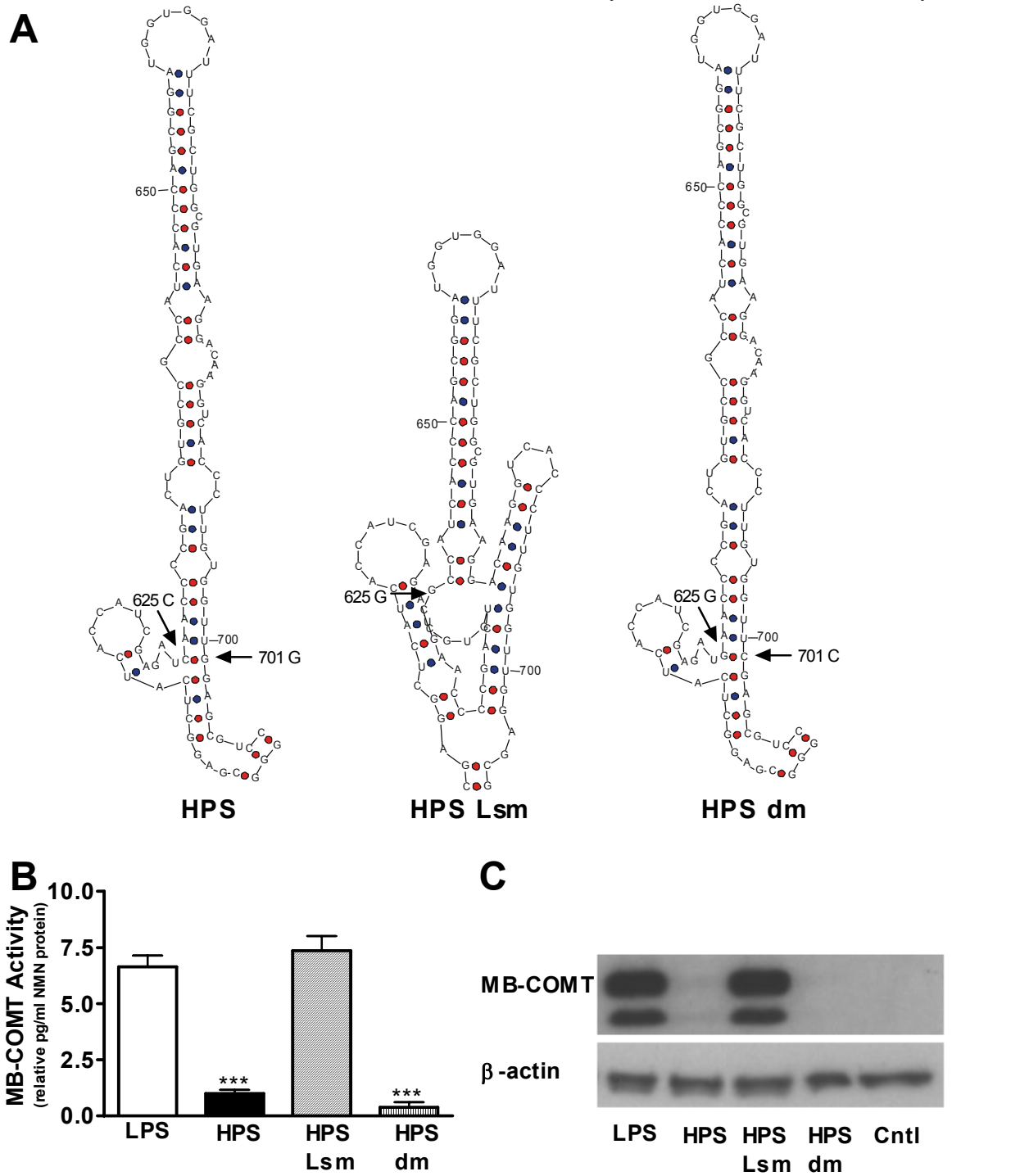

fig. S4 Site directed mutagenesis that destroys the stable stem loop structure corresponding to the HPS haplotype, restores MB-COMT enzymatic activity and protein expression to levels produced by the LPS clone. **(A)** mRNA folding diagrams depict the mRNA local stem-loop structures for *COMT* transcripts corresponding to the original and mutated HPS haplotypes. The mRNA structure corresponding to the HPS haplotype was changed to an LPS haplotype-like structure (HPS Lsm) by single mutation of 403C to G. The original HPS haplotype structure was reconstructed by double mutation of interacting nucleotides 403C to G and 479G to C (HPS dm). **(B)** The HPS Lsm exhibited COMT enzymatic activity equivalent to the LPS haplotype, while the HPS dm exhibited reduced enzymatic activity. **(C)** Protein expression profiles for the two mutants paralleled the enzymatic activity data. Converting the secondary structure corresponding to the HPS haplotype into a less stable LPS-like mRNA stem-loop structure with lower ΔG restored protein levels. No differences in protein expression were observed between the HPS haplotype and the HPS dm. Data are expressed as Mean ± SEM. ***$P < 0.001 \neq$ LPS.

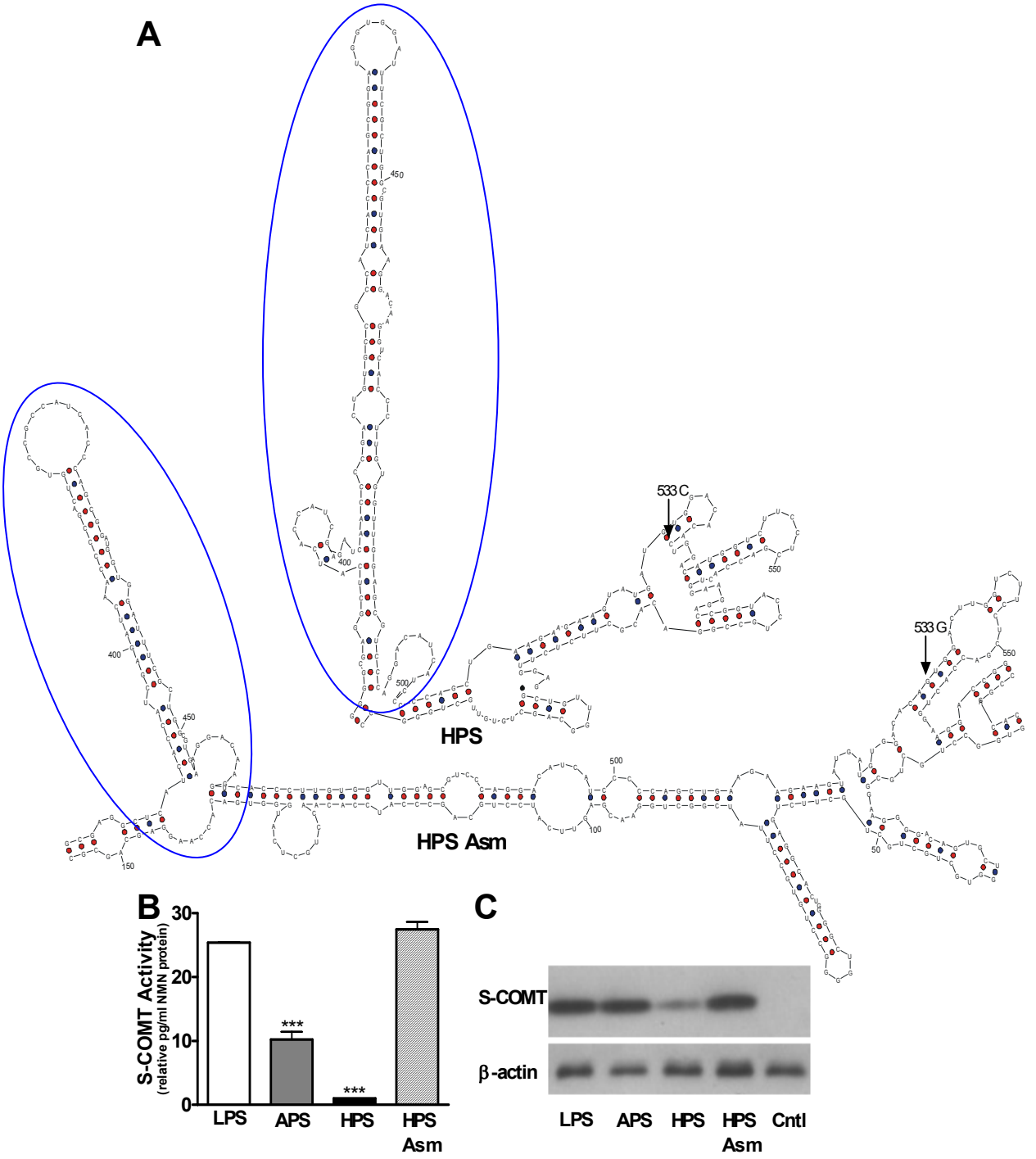

fig. S5 Cells transfected with the HPS Asm clone exhibited enzymatic activity and protein levels that do not differ from cells transfected with the LPS clone. **(A)** mRNA folding diagrams depict the mRNA local stem-loop structures for *COMT* transcripts corresponding to the original and mutated HPS haplotypes. The secondary structure corresponding to the HPS haplotype was converted into an APS haplotype-like structure (HPS Asm) by a mutation of 533C to G. The local stem-loop structures in the polymorphic region are circled. **(B)** In cells expressing S-COMT, the HPS Asm exhibited enzymatic activity that did not differ from the LPS haplotype. **(C)** Protein expression profiles for the HPS Asm mutant paralleled the enzymatic activity data. Converting the secondary structure corresponding to the HPS haplotype into a less stable APS-like mRNA stem-loop structure with lower ΔG restored protein levels. Data are expressed as Mean ± SEM. ***$P < 0.001$ ≠ LPS.

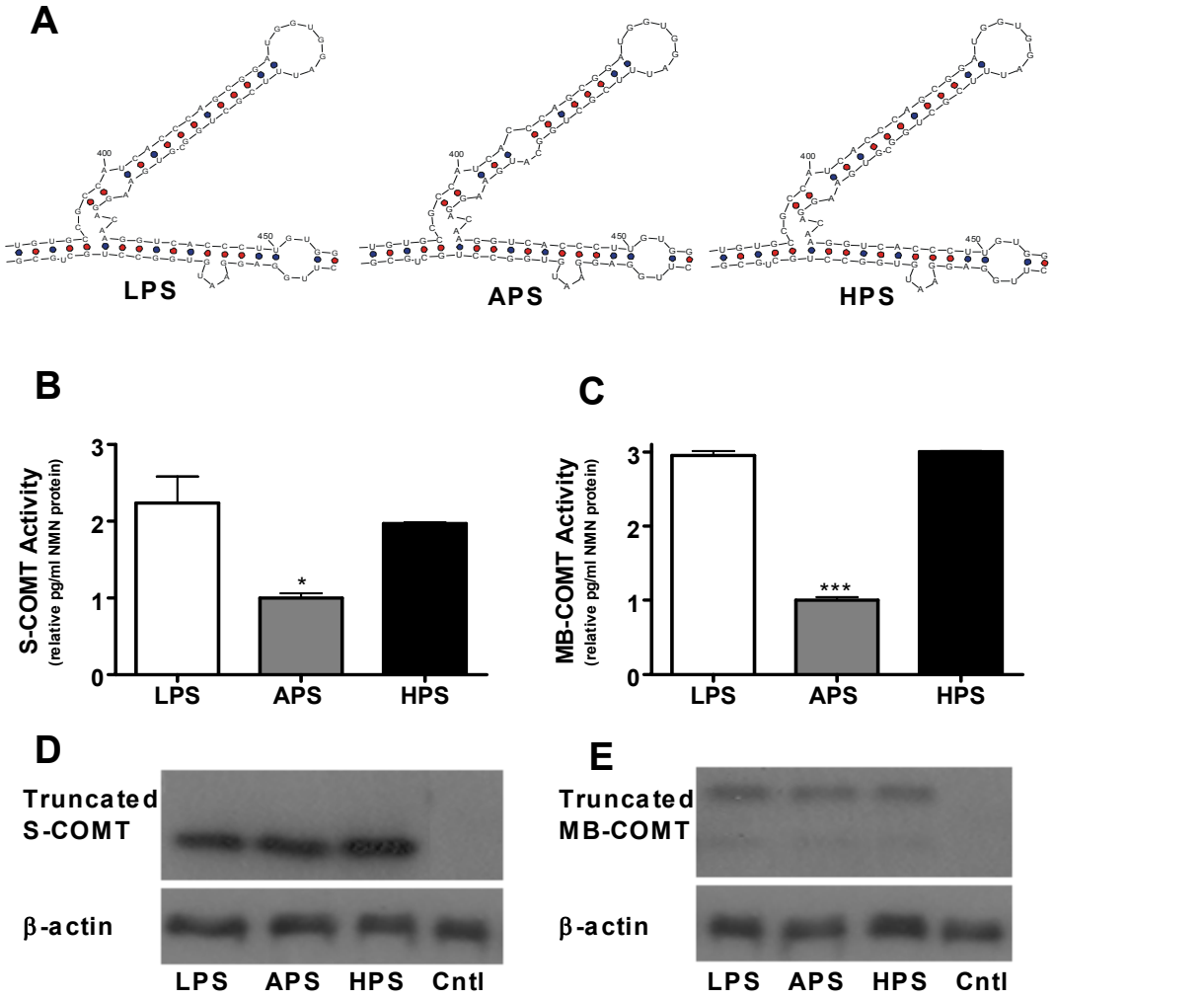

fig. S6 Eliminating the differences in thermodynamic stability of RNA coded by the HPS and LPS haplotypes eliminates differences in enzyme activity and protein expression. **(A)** Folding analyses show that *COMT* mRNA lacking 5' and 3' UTRs form similarly stable secondary structures for each haplotype. **(B, C)** Previously observed differences in COMT enzymatic activity between clones coded by the HPS and LPS haplotypes were eliminated in the modified constructs. Yet, enzymatic activity remained significantly lower for the S-COMT and MB-COMT truncated APS clone. **(D, E)** Protein levels did not differ for the three modified S- and MB-COMT constructs. Data are expressed as Mean ± SEM. ***$P < 0.001$, *$P < 0.05$ ≠LPS and HPS.

References

1. M. Zuker, *Nucleic Acids Res* **31**, 3406 (Jul 1, 2003).
2. A. Y. Ogurtsov, S. A. Shabalina, A. S. Kondrashov, M. A. Roytberg, *Bioinformatics* (Mar 16, 2006).
3. D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, *J Mol Biol* **288**, 911 (May 21, 1999).
4. N. N. Nazipova *et al.*, *Comput Appl Biosci* **11**, 423 (Aug, 1995).
5. M. Masuda, M. Tsunoda, Y. Yusa, S. Yamada, K. Imai, *Ann Clin Biochem* **39**, 589 (Nov, 2002).
6. L. Diatchenko *et al.*, *Hum Mol Genet* **14**, 135 (Jan 1, 2005).
7. N. J. Bray *et al.*, *Am J Hum Genet* **73**, 152 (Jul, 2003).
8. Y. Chen *et al.*, *Genes Genet Syst* **74**, 271 (Dec, 1999).
9. D. B. Carlini, Y. Chen, W. Stephan, *Genetics* **159**, 623 (Oct, 2001).
10. B. Boudikova, C. Szumlanski, B. Maidak, R. Weinshilboum, *Clin Pharmacol Ther* **48**, 381 (Oct, 1990).