Short Communication

Haplotype-Phenotype Relationships of Paraoxonase-1

Jia Chen,[1,2,6] **Wendy Chan,**[1] **Sylvan Wallenstein,**[3] **Gertrud Berkowitz,**[1] **and James G. Wetmur**[4,5]

Departments of [1]Community and Preventive Medicine, [2]Pediatrics, [3]Biomathematics, [4]Microbiology, and [5]Human Genetics, and [6]Derald H. Ruttenberg Cancer Center, Mount Sinai School of Medicine, New York, New York

Abstract

Paraoxonase 1 (PON1) is an enzyme with multiple activities, including detoxification of organophosphates. It is believed to be important in preventing neurotoxic damage and has also been implicated in atherosclerosis. The *PON1* gene contains five common polymorphisms, three in the promoter ($-909G > C$, $-162A > G$, $-108C > T$) and two in the coding region (M55L, Q192R) with varying but incomplete linkage disequilibrium. Our previous study showed that functional polymorphisms in *PON1* were strongly associated with enzymatic activity in both pregnant women [26-30 weeks of gestation] and neonates. However, there was substantial overlapping of enzyme activities between genotypes. In this study, we investigated whether haplotype (genotype + phase) information would strengthen the genotype-phenotype relationship for PON1. The study consisted of a multiethnic population of 402 mothers and 229 neonates. Haplotypes were imputed by two widely used programs, PHASE and tagSNPs, which yielded very similar results. There were seven haplotypes with a frequency of 5% or higher in at least one ethnic group of the study population. Haplotype composition varied substantially with respect to ethnicity. Haplotypes in Caucasians and African-Americans showed the largest difference, and Caribbean Hispanics seemed to be a mixture of Caucasian and African ancestry. Collectively, the genetic (genotype or haplotype) contribution to PON1 enzymatic activity (measured as phenyl-acetate hydrolysis) was greater in neonates compared with mothers. Specifically, 16.6% of PON1 variability was explained by genotypes in mothers compared with 30.9% in neonates. Haplotype information offered a slightly increased power in predicting PON1 activity; they explained 35.5% and 19.3% of PON1 variability in neonates and mothers, respectively. (Cancer Epidemiol Biomarkers Prev 2005;14(3):731–4)

Introduction

Paraoxonase-1 (PON1), found in high-density lipoprotein, is an enzyme that is capable of hydrolysis of a wide variety of substrates including thiolactones (e.g., homocysteine thiolactone; ref. 1), aryl esters (e.g., phenyl acetate), and organophosphates (e.g., paraoxon; ref. 2). PON1 detoxifies organophosphates by cleavage of active oxons that are potential cholinesterase inhibitors in the peripheral and central nervous system. PON1 is also thought to affect lipid metabolism. PON1 knockout mice are not only more sensitive to organophosphate toxicity but also are more prone to atherosclerosis when fed a high-fat, high-cholesterol diet (3). In humans, PON1 is believed to be important in preventing neurotoxic damage and atherosclerosis, and even non-Hodgkin's lymphoma (4). *PON1* is a moderate size gene with 25.8 kb downstream from the initiation codon encoding the 355 amino acids. It contains five common single nucleotide polymorphisms (SNP), three in the promoter ($-909G > C$, $-162A > G$, $-108C > T$) and two in the coding region (L55M, Q192R) with incomplete linkage disequilibrium (5). Our previous study has shown that these SNPs in *PON1* were strongly associated with enzymatic activity in both pregnant women and neonates (5).

Promoter variants may affect the level of expression by more than 2-fold (6). Q192R affects the relative rate of hydrolysis of certain organophosphate substrates, such as paraoxon, compared with phenylacetate by as much as an order of magnitude, but has only a small effect on the relative rates of hydrolysis of chlorpyrifos oxon and phenylacetate (7). The L55M polymorphism may affect PON1 protein stability (8), and in our study contributed significantly to plasma enzymatic activity. The phase information, in addition to genotypes of these SNPs (i.e., haplotypes), may be highly relevant to the PON1 activity; genotype information alone may distort the underlying contributions from combinations of SNPs to PON1 activity. It is possible that an apparent genotype-phenotype relationship could be due to linkage disequilibrium with the functional SNP. Brophy et al. (6) suggested that the apparent effect of the L55M polymorphism on enzyme concentration may be due to linkage disequilibrium with one of the promoter variants. Thus, haplotype information may help clarify the genetic influence on PON1 activity.

Recent developments in statistical methods have permitted the reconstruction of haplotypes and the estimation of haplotype frequencies in unrelated populations based on SNP data. The maximum likelihood–based expectation maximization algorithm and the Bayesian-based algorithm have been the most widely used approaches for haplotype inference. We used two recently developed programs, PHASE and tagSNPs, to infer haplotypes for the five common SNPs in the *PON1* gene and to examine their associations with PON1 activity.

Methods

The study population is from an on-going study at the Mount Sinai Children's Environmental Health Center to assess prospectively infant growth and neurodevelopment associated with pesticide exposure in urban New York City. The study

protocol was approved by the Institutional Review Board. A total of 402 maternal blood samples were obtained in heparin-treated vacutainers from pregnant women who were 26 to 30 weeks in gestational age and were self-identified as being Caucasian (n = 82), African-American (n = 117), or Hispanic of Caribbean origin (n = 203). At birth, a sample of umbilical cord blood was obtained for 223 infants (Caucasian n = 55, African-American n = 61, Hispanic n = 107) using the same anticoagulant. Plasma was separated immediately (within 24 hours) by two cycles of centrifugation. Three aliquots of maternal plasma and three aliquots of cord blood plasma were frozen at −70°C, one for determination of PON1 activity. The buffy coat was separated from the RBC, and DNA was extracted and purified using a QIAamp blood kit (Qiagen, Valencia, CA) as described by the manufacturer.

Methods for genotyping and measuring PON1 activity have been published previously (1). In brief, genotypes were determined by a clamp-dependent allele-specific PCR (9). The activity of PON1 was measured by phenylacetate hydrolysis. These previously reported values (1) were used as variables to infer haplotypes and to investigate haplotype-phenotype relationships among mothers and neonates.

Two approaches were used for inferring haplotypes from the genotype data. The tagSNPs method (10) uses a modified Excoffier-Slatkin expectation maximization algorithm (11), which involves maximum likelihood estimation to approximate haplotype frequencies among unrelated subjects assuming Hardy-Weinberg equilibrium. Another method, PHASE (12), uses a Bayesian-based approach; it uses a prior distribution and likelihood estimation to determine the posterior distribution of population haplotype frequencies. The method assumes that unknown haplotypes are random values and attempts to determine their conditional distribution based on known genotype and haplotype data. Both programs calculate conditional probability estimates for the haplotypes carried by each subject that can subsequently be used in a multivariate regression model to test for haplotype-disease associations.

Because of the diversity of the study population, haplotypes were imputed by ethnicity. Haplotypes with frequencies ≥5% in specific ethnic populations were considered ''common'' haplotypes whereas those with <5% frequency were combined into one group for subsequent analysis. Imputed haplotypes were used to assess associations of haplotypes with PON1 activity. Subjects were grouped according to their imputed diplotypes (haplotype pairs) to compare PON1 activity by diplotype within mothers and neonates. When more than one diplotype was inferred for a subject, only the pair with the highest certainty in imputed haplotypes was included in this analysis. To assess the contributions of genetic factors to PON1 activity, the coefficient of determination, R^2, was calculated to determine the variation in PON1 activity explained by the five SNP genotypes and by conditionally imputed haplotypes. Genotypes were entered into a regression model as ordinal variables, whereas for haplotypes, imputed probabilities were used as predictors. An examination of the distribution of PON1 activity levels in mothers and neonates did not reveal strong departures from normality. Therefore, original PON1 values were used for the analysis. Statistical analysis was done using SAS Software, version 8.1.

Results and Discussion

We employed two widely used statistical methods (i.e., PHASE and tagSNPs) to impute *PON1* haplotypes based on genotype information. The two algorithmic methods yielded quantitatively similar results (data not shown). As a result, PHASE was used for subsequent analyses because it has been shown to reduce error rates in haplotype reconstruction compared with the expectation maximization algorithm (12, 13). Table 1 shows the distribution of common haplotypes (≥5% in at least one ethnic group) of 402 mothers (804 chromosomes) and 223 neonates (446 chromosomes) imputed by PHASE. As our study population consisted of Caucasians, Hispanics, and African-Americans, we compared haplotype distribution with respect to ethnic groups (Table 1). The haplotype composition varied substantially by ethnicity; Caucasians and African-Americans showed the largest difference from each other whereas Hispanics seemed to fall in the middle of the two groups. For example, whereas the haplotype ''21221'' was the most prevalent (35%) haplotype among Caucasian mothers, it was much less prevalent (7%) among African-American mothers; its frequency was 17% among Hispanics.

The primary purpose of the study was to examine whether haplotype information (genotype + phase) would strengthen the genotype-phenotype relationship for PON1, which may help clarify genotype-disease relationships in the future. Although we previously reported that *PON1* genotypes were significantly correlated with PON1 activities, there was substantial overlapping in enzyme activities with respect to genotypes (1). Phase information (alignment information of multiple SNPs along one chromosome) may help explain some portion of the overlapping activity and strengthen the genotype-phenotype relationship. For example, whereas the

Table 1. Conditionally imputed PON1 haplotypes by ethnicity

Haplotype*	Mothers				Neonates			
	Frequency (%)				Frequency (%)			
	Caucasian	African-American	Hispanic	All	Caucasian	African-American	Hispanic	All
	($N^†$ = 164)	(N = 234)	(N = 406)	(N = 804)	(N = 110)	(N = 122)	(N = 214)	(N = 446)
21221	35.1	7.1	16.6	19.6	26.9	6.4	13.5	15.6
11112	6.5	25.2	18.1	16.6	4.7	31.8	20.9	19.2
12111	17.4	14.7	16.2	16.1	15.0	15.3	15.8	15.4
12112	1.7	22.1	8.7	10.8	3.7	22.4	7.8	11.3
21212	15.9	3.8	8.8	9.5	11.5	3.2	10.8	8.5
11111	2.1	8.3	8.3	6.2	4.3	5.2	10.7	6.7
21211	8.5	2.9	6.2	5.9	16.6	0.6	7.5	8.2
11121	8.2	0.8	3.4	4.1	11.5	2.2	0.9	4.9
11122	0.0	7.5	3.1	3.5	2.3	6.9	4.9	4.7
Other	4.7	7.6	10.7	7.7	3.6	5.9	7.2	5.6

*The SNP positions within a haplotype are the following: −909G > C, −162A > G, −108C > T, M55L, Q192R;. 1 = wild type; 2 = variant. A total of 26 and 29 haplotypes were inferred in mothers and neonates, respectively.
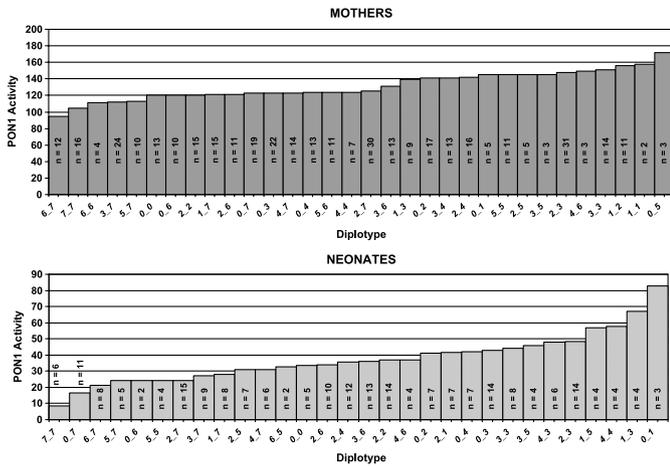†Number of chromosomes.

Figure 1. Associations between diplotypes (haplotype pairs) and PON1 activity in mothers and neonates. The SNP positions within a haplotype are the following: $-909G > C$, $-162A > G$, $-108C > T$, M55L, Q192R; ''1'' represents wild-type allele and ''2'' represents variant allele. Common haplotypes (>5%) are labeled as the following: 1,11111; 2,11112; 3,12111; 4,12112; 5,21211; 6,21212; 7,21221. Low frequency haplotypes (<5%) are labeled as ''0.''

promoter $-108C > T$ SNP of *PON1* may affect the level of expression (6), the L55M coding SNP may influence protein stability (8). It is conceivable that variant genotypes residing on the same chromosome may have synergistic adverse effects on PON1 activity.

Figure 1 depicts the relationships between the *PON1* diplotypes (haplotype pairs) and PON1 activity in mothers and neonates. First of all, there was a larger variation in PON1 activity with respect to diplotypes in neonates than in mothers. The plots indicated that diplotypes carrying variant alleles at multiple loci tended to be associated with lower PON1 activity (i.e., they have a tendency to fall into the left side of the plot). The trend stayed the same when we considered the variant alleles of -909 and -108 jointly because of the high degree of linkage disequilibrium ($D' > 0.90$ in all ethnic groups; ref. 1).

We compared the portion of variation in PON1 activity attributable to genotypes of five SNPs versus seven common haplotypes by using the coefficient of determination, R^2 (Table 2). The purpose for this analysis was to explore whether any gain in information would be achieved by using conditionally imputed haplotypes to explain variability in PON1 activity as opposed to genotype data alone. Genotypes alone do not take into account ambiguous haplotype phase when more than one heterozygous genotype is present. Independent of the approach, a much higher portion of the enzyme variability in neonates was explained by genetic factors (either genotypes or haplotypes) compared with mothers. For example, 35.5% of variability in PON1 activity in neonates was explained by the common haplotypes compared with 19.3% in mothers. When comparing overall enzyme variability explained by genotypes versus haplotypes, haplotypes offered a slightly increased power in explaining PON1 enzyme variability in both neonates and mothers. For example, haplotypes explained 35.5% PON1 variability in neonates compared with 30.9% explained by genotypes.

Table 2 also indicated that the genetic contribution varied significantly with ethnicity. Because the effect of the genotypes or haplotypes was much greater in neonates than adults, and because the frequencies of genotypes and haplotypes varied with ethnicity, the average PON1 activity might be strongly influenced by ethnicity. Specifically, the average PON1

activities in neonates were 29.2, 38.3, and 48.4 units/mL in Caucasians, Hispanics, and African-Americans, respectively. However, this trend was not observed in mothers. For both mothers and neonates, more of the variance in enzymatic activity was explained by genotype for Caucasians than African-Americans. Because of genetic admixture, Hispanics were expected to fall in between. The rule was violated for the mothers when analyzed using genotypes. However, when the data were analyzed with respect to haplotypes, the expected order was restored, with Hispanics falling between Caucasians and African-Americans in mothers as well as neonates. This observation suggests that haplotype-based analyses may improve uncertainties or abnormalities introduced by analyses based solely on genotype data.

Because one of the main focuses of the parent study in the Mount Sinai Children's Environmental Health Center was to assess health effects of chlorpyrifos exposure, phenylacetate was used as the substrate for measuring PON1 enzymatic activity. The Q192R polymorphism, which primarily affects PON1 activity with certain organophosphate substrates, such as paraoxon, has little effect when chlorpyrifos or phenylacetate is the substrate. The genetic contribution to the variance of PON1 activity would have been larger had paraoxon been used as the substrate (14). The genotype-haplotype comparison in this study is limited to the effect of the *PON1* polymorphisms on PON1 protein levels in blood.

It is worth pointing out that this study was not carried out to define haplotype blocks of *PON1*. Jarvik et al. (14) recently reported that substantial recombination has occurred within the *PON1* gene resulting in high haplotype diversity and broken haplotype blocks within the gene. Our results support this notion as at least 19 different haplotypes were present in our multiethnic population. Because $-909G > C$ and $-108C > T$ were in nearly complete linkage disequilibrium ($D' > 0.90$; ref. 1), Jarvik et al. (14) concluded that a large proportion of haplotype diversity can be captured by four SNPs (i.e., $-162A > G$, $-108C > T$, L55M, and Q192R), the same polymorphisms used for our analysis.

To date, our study is the largest to explore the haplotype-phenotype relationships of PON1. The population consists of 402 mothers and 229 neonates from a multiethnic population, providing a unique opportunity to examine ethnic-specific haplotype structures. It is not known why genotype-phenotype or haplotype-phenotype association is

Table 2. Proportion of variance (R^2) in PON1 activity explained by genotypes and haplotypes

Group	Arylesterase activity		R^2 for PON1 activity	
	Mean (units/mL)	SD	Genotypes (-909, -162, -108, 55, 192)*	Haplotypes (PON1$_{-909/-162/-108/55/192}$[†])
Mothers	(N = 396)			
All	132.4	33.4	16.6	19.3[‡]
Caucasian	133.1	31.9	20.4	33.4
Hispanic	135.8	35.8	28.8	24.9
African-American	126.0	29.3	9.6	7.9
Neonates	(N = 190)			
All	38.8	21.2	30.9	35.5[‡]
Caucasian	29.2	16.6	70.4	60.2
Hispanic	38.3	20.5	23.6	30.8
African-American	48.4	22.4	14.6	19.7

*Genotypes treated as ordinal variables (0,1,2) into the regression model.
[†] Common haplotypes (>5% frequency) entered into a regression model; homologous haplotypes recoded to 2; reference group = common haplotype with the highest PON1 activity by group.
[‡] Adjusted for ethnicity.

so much poorer in African-Americans than Caucasians, both for the mother and neonates. One might speculate that additional and important genetic polymorphisms in *PON1* or in regulation of PON1 have yet to be identified. Overall imputed *PON1* haplotypes contributed marginally to explaining phenotypic variances in PON1, but may improve uncertainties or abnormalities introduced by analyses based solely on genotype data within a multiethnic population.

Acknowledgments

References

1. Jakubowski H, Ambrosius WT, Pratt JH. Genetic determinants of homocysteine thiolactonase activity in humans: implications for atherosclerosis. FEBS Lett 2001;491:35–9.
2. Gan KN, Smolen A, Eckerson HW, Ladu BN. Purification of human serum paraoxonase arylesterase—evidence for one esterase catalyzing both activities. Drug Metabolism and Disposition 1991;19:100–6.
3. Shih DM, Gu LJ, Xia YR, Navab M, Li WF, Hama S, et al. Mice lacking serum paraoxonase are susceptible to organophosphate toxicity and atherosclerosis. Nature 1998;394:284–7.
4. Kerridge I, Lincz L, Scorgie F, Hickey D, Granter N, Spencer A. Association between xenobiotic gene polymorphisms and non-Hodgkin's lymphoma risk. Br J Haematol 2002;118:477–81.
5. Chen J, Kumar M, Chan W, Berkowitz G, Wetmur JG. Increased influence of genetic variation on PON1 activity in neonates. Environ Health Perspect 2003;111:1403–9.
6. Brophy VH, Jampsa RL, Clendenning JB, McKinstry LA, Jarvik GP, Furlong CE. Effects of 5′ regulatory-region polymorphisms on paraoxonase-gene (PON1) expression. Am J Hum Genet 2001;68:1428–36.
7. Davies HG, Richter RJ, Keifer M, Broomfield CA, Sowalla J, Furlong CE. The effect of the human serum paraoxonase polymorphism is reversed with diazoxon, soman and sarin. Nat Genet 1996;14:334–6.
8. Leviev I, Deakin S, James RW. Decreased stability of the M54 isoform of paraoxonase as a contributory factor to variations in human serum paraoxonase concentrations. J Lipid Res 2001;42:528–35.
9. Germer S, Holland MJ, Higuchi R. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. Genome Res 2000;10:258–66.
10. 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.
11. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 1995;12:921–7.
12. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–89.
13. Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 2003;73:1162–9.
14. Jarvik GP, Hatsukami TS, Carlson C, et al. Paraoxonase activity, but not haplotype utilizing the linkage disequilibrium structure, predicts vascular disease. Arterioscler Thromb Vasc Biol 2003;23:1465–71.

Haplotype-Phenotype Relationships of Paraoxonase-1

Jia Chen, Wendy Chan, Sylvan Wallenstein, et al.

Updated version	Access the most recent version of this article at: http://cebp.aacrjournals.org/content/14/3/731	

Cited articles	This article cites 14 articles, 4 of which you can access for free at: http://cebp.aacrjournals.org/content/14/3/731.full#ref-list-1	
Citing articles	This article has been cited by 1 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/14/3/731.full#related-urls	

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.	
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.	
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/14/3/731. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.	