Mutation-Positive and Mutation-Negative Patients with Cowden and Bannayan-Riley-Ruvalcaba Syndromes Associated with Distinct 10q Haplotypes

Marcus G. Pezzolesi, Yan Li, Xiao-Ping Zhou, Robert Pilarski, Lei Shen, and Charis Eng

Phosphatase and tensin homolog deleted on chromosome 10 (*PTEN*) encodes a tumor-suppressor phosphatase frequently mutated in both sporadic and heritable forms of human cancer. Germline mutations are associated with a number of heritable cancer syndromes that are jointly referred to as the "*PTEN* hamartoma tumor syndrome" (PHTS) and include Cowden syndrome, Bannayan-Riley-Ruvalcaba syndrome, Proteus syndrome, and Proteus-like syndrome. Germline *PTEN* mutations have been identified in a significant proportion of patients with PHTS; however, there are still many individuals with classic diagnostic features for whom mutations have yet to be identified. To address this, we took a haplotype-based approach and investigated the association of specific genomic regions of the *PTEN* locus with PHTS. We found this locus to be characterized by three distinct haplotype blocks 33 kb, 65 kb, and 43 kb in length. Comparisons of the haplotype distributions for all three blocks differed significantly among patients with PHTS and controls ($P = .0098$, $P < .0001$, and $P < .0001$ for blocks 1, 2, and 3, respectively). "Rare" haplotype blocks and extended haplotypes account for two-to-threefold more PHTS chromosomes than control chromosomes. *PTEN* mutation–negative patients are strongly associated with a haplotype block spanning a region upstream of *PTEN* and the gene's first intron ($P = .0027$). Furthermore, allelic combinations contribute to the phenotypic complexity of this syndrome. Taken together, these data suggest that specific haplotypes and rare alleles underlie the disease etiology in these sample populations; constitute low-penetrance, modifying loci; and, specifically in the case of patients with PHTS for whom traditional mutations have yet to be identified, may harbor pathogenic variant(s) that have escaped detection by standard *PTEN* mutation–scanning methodologies.

Phosphatase and tensin homolog deleted on chromosome 10 (*PTEN* [MIM 601728]) (also known as "mutated in multiple advanced cancers 1" [*MMAC1*] and "tensin-like phosphatase 1" [*TEP1*]) encodes a tumor-suppressor phosphatase that signals down the phosphoinositol-3-kinase (PI3K)/AKT pathway, effecting apoptosis and cell-cycle arrest.[1–3] Germline *PTEN* mutations are primarily associated with a number of apparently clinically distinct heritable cancer syndromes jointly referred to as the "*PTEN* hamartoma tumor syndrome" (PHTS).[4] These include Cowden syndrome (CS [MIM 158350]), Bannayan-Riley-Ruvalcaba syndrome (BRRS [MIM 153480]), Proteus syndrome (PS [MIM 176920]), and Proteus-like syndrome (PLS). All four syndromes are characterized by multiple hamartomatous lesions affecting derivatives of all three germ-cell layers. In CS, patients are also at an increased risk of developing breast, thyroid, and endometrial cancer.[5,6] To date, germline *PTEN* mutations have been identified in 85% of patients who received a diagnosis of CS and 65% of patients who received a diagnosis of BRRS.[4,7] Additionally, 20% and 50% of patients with PS and PLS, respectively, have also been shown to carry *PTEN* germline mutations.[8–10]

Mutation scanning of *PTEN* has primarily focused on the gene's nine exons and intron-exon boundaries, which span ~103 kb on chromosome subband 10q23.3. Germline mutations have been reported throughout *PTEN,* with the exception of exon 9, and the majority of these localize to its phosphatase catalytic core located in exon 5.[1,11] More recently, mutations in the core promoter region of *PTEN* have also been identified and have been found to be associated with CS and increased phosphorylated AKT levels.[7] However, despite the significant proportion of patients with known *PTEN* mutations, there are still many individuals with classic PHTS diagnostic features for whom mutations have yet to be identified. Notably, CS is believed to be linked to the *PTEN* region, without genetic heterogeneity.[12] In BRRS, on the other hand, the extent of genetic heterogeneity is unknown. Other mechanisms, such as modifiers of *PTEN* or another gene (or genes) that have yet to be identified, may be causal of this syndrome.[4,13] For individuals with PHTS, particularly those with CS, and without identifiable germline mutations, it is likely that the molecular mechanism(s) underlying their disease involves genetic alteration outside of the *PTEN* coding sequence, possibly involving elements associated

From the Genomic Medicine Institute (M.G.P.; X.-P.Z.; C.E.), Lerner Research Institute (M.G.P.; X.-P.Z.; C.E.), and Taussig Cancer Center (C.E.), Cleveland Clinic Foundation, and Department of Genetics (C.E.) and CASE Comprehensive Cancer Center (C.E.), Case Western Reserve University School of Medicine, Cleveland; and Division of Biostatistics, School of Public Health (Y.L.; L.S.), Clinical Cancer Genetics Program, Human Cancer Genetics Program, Comprehensive Cancer Center (X.-P.Z.; R.P.), and Division of Human Genetics, Department of Internal Medicine (R.P.), The Ohio State University, Columbus

Received May 30, 2006; accepted for publication September 5, 2006; electronically published September 29, 2006.

Address for correspondence and reprints: Dr. Charis Eng, Genomic Medicine Institute, Cleveland Clinic Lerner Research Institute, 9500 Euclid Avenue, NE-50, Cleveland, OH 44195. E-mail: engc@ccf.org

Am. J. Hum. Genet. 2006;79:923–934. 0002-9297/2006/7905-0014$15.00

in its transregulation or deregulation and which may lie upstream, downstream, or intronic of *PTEN*. Identifying the mechanism of PTEN dysfunction in these patients is critical and of significant importance to the practice of personalized genetic healthcare.

To aid in identifying these genetic alterations, we used a haplotype-based approach to investigate the association of specific genomic regions of the *PTEN* locus with disease. Through this approach, we demonstrate that specific haplotypes, perhaps acting as low-penetrance susceptibility loci, are associated with PHTS in *PTEN* mutation–negative samples. In addition to furthering our understanding of the role *PTEN* has in patients without detectable mutations, we have also identified specific haplotypes that may act as low-penetrance alleles, or modifying factors, which could influence phenotypic expression in a subset of CS/BRRS-affected patients with known germline *PTEN* mutations.

Material and Methods

Study Subjects

A total of 447 unrelated subjects were included in the current analysis—94 white control subjects, 148 white patients with PHTS without detectable germline *PTEN* mutations (i.e., *PTEN* mutation–negative patients), and 205 white patients with PHTS with previously identified germline *PTEN* mutations/variations (i.e., *PTEN* mutation/variation–positive patients). DNA for control subjects (Utah residents with ancestry from northern and western Europe) was acquired from the Coriell Institute for Medical Research (Camden, NJ). All subjects with PHTS were enrolled by referral from centers located throughout the United States, Canada, and Europe. Informed consent was acquired for all referred subjects in accordance with procedures approved by the human subjects protection committees of each respective institution.

Among the *PTEN* mutation–negative patients, 94 had classic CS, 10 had classic BRRS, 4 exhibited features of both CS and BRRS (termed "CS-BRRS overlap"), and 39 exhibited a CS-like phenotype (i.e., had some features of CS but did not meet operational diagnostic criteria). One *PTEN* mutation–negative patient could not be classified.

The cohort of *PTEN* mutation/variation–positive patients included 103 mutation-positive samples (i.e., samples with pathogenic heterozygous missense or nonsense mutations) and 102 variation-positive samples. This latter group consists primarily of individuals with identified variants of unknown significance located in the *PTEN* core promoter region or within potential splice donor/acceptor sites. Among the *PTEN* mutation-positive patients, 34 had classic CS, 18 had classic BRRS, 10 exhibited features of CS-BRRS overlap, and 40 were classified as CS-like. One *PTEN* mutation–positive patient could not be classified. The *PTEN* variation–positive patients included 39 with classic CS, 2 with classic BRRS, 6 with both CS and BRRS features, and 52 classified as having CS-like phenotype. Three *PTEN* variation–positive patients could not be classified.

All patients classified as having CS in the current study met operational criteria established by the International Cowden Consortium and curated by the National Comprehensive Cancer Network.[6]

SNP Genotyping

SNPs spanning the *PTEN* locus and located ~1 every 5 kb were selected from the dbSNP database for validation and estimation of minor-allele frequency in a 10-sample screening set consisting of 5 white control subjects and 5 white patient samples. Twenty-four screened SNPs were found to have a minor-allele frequency $\geq$0.10 and met our criteria for inclusion in this study. To achieve a uniformly spaced SNP map, six additional SNPs with a minor-allele frequency $\geq$0.10 were identified by DNA resequencing in our screening set. All 30 SNPs were genotyped in our 447-sample cohort. PCRs included 12.5 μl HotStarTaq Master Mix (Qiagen), 10 mM forward primer, 10 mM reverse primer, and 20 ng of template DNA and used the following thermal cycling conditions: 95°C for 15 min, 34 cycles at 95°C for 30 s, 50°C–58°C for 45 s, and 72°C for 1 min, followed by a final extension at 72°C for 10 min. Twenty-nine SNPs were genotyped using either RFLP, SNaPshot (Applied Biosystems), or fragment analysis. SNaPshot and fragment-analysis products were electrophoresed using an ABI 3730 DNA Analyzer (Applied Biosystems) and were analyzed using GeneMapper v3.5 software (Applied Biosystems). *rs12573787* was genotyped by direct DNA resequencing. Primer sequences and genotyping methodologies are provided in appendix A (online only).

Hemizygous PTEN *Deletion Analysis*

Real-time quantitative PCR was used to investigate potential microdeletions in both control (n = 4) and *PTEN* mutation–negative (n = 14) samples for which homozygosity was observed for all 30 SNPs. Fifteen *PTEN* mutation/variation–positive samples were also homozygous for SNPs assayed in this region; however, by virtue of their heterozygous mutations/variations, these samples are assumed to carry two copies of the *PTEN* allele. Copy-number determinations were made for exons 2 and 5 of our target gene, *PTEN,* and for exon 7 of a control reference gene, *GAPDH*. Four homozygous control samples and four homozygous *PTEN* mutation/variation–positive samples were used as negative controls. Additionally, two samples previously determined to have *PTEN* deletions (one spanning the entire *PTEN* locus, the other spanning both the *PTEN* and *BMPR1A* genes) were assayed as positive controls. PCR efficiencies for each amplicon were determined by standard curve analysis using serial dilutions of genomic DNA from a control sample (100 ng, 50 ng, 25 ng, and 12.5 ng). The calculated PCR efficiencies for these amplicons ranged from 76% to 81%.

Determination of gene copy number was assayed using 12.5 μl iQ SYBR Green Supermix (Bio-Rad Laboratories), 10 mM forward primer, 10 mM reverse primer, and 20 ng of template DNA. Thermal cycling conditions were 95°C for 3 min and 40 cycles at 95°C for 30 s followed by 58°C for 30 s and 72°C for 30 s, with the use of an ABI 7700 Sequence Detection System (Applied Biosystems). Target and reference genes were assayed in triplicate for each sample and were subject to melt-curve analysis to determine amplicon specificity. The relative quantification of gene copy number for both *PTEN* amplicons was determined using the comparative delta Ct method ($2^{-\Delta\Delta Ct}$) as described by Livak et al.[14]

Linkage Disequilibrium (LD) and Haplotype Analysis

Following assessment of Hardy-Weinberg equilibrium (HWE) at each polymorphic locus, pairwise LD coefficients (Lewontin's D') were estimated using the LDmax software program and were vi-

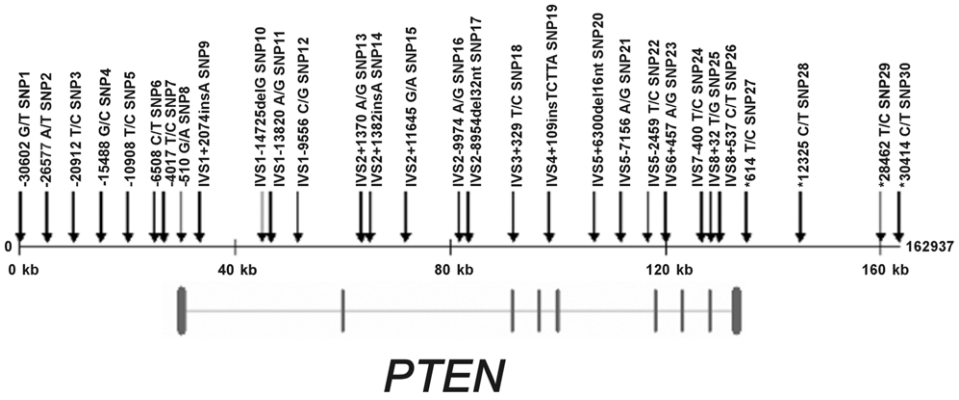

Figure 1. Schematic diagram of the *PTEN* locus and SNPs included in the current analysis

sualized using the GOLD graphical interface (Gonçalo Abecasis' Software Page).[15] D' was calculated and plotted separately for each sample population (control subjects, *PTEN* mutation–negative patients, and *PTEN* mutation/variation–positive patients). LD blocks were determined using data from the control population and the dynamic programming algorithms implemented in the HapBlock software program (empirical LD method; $D' > 0.90$ for strong LD).[16,17] After the block partitioning, haplotype phase was reconstructed for each block and all genotyped samples by use of the SNPHAP software program, based on pairwise LD measurements and the expectation-maximization (EM) algorithm, and the PHASE v2.1 software program, based on a Bayesian approach.[18,19] Additionally, haplotype phase was reconstructed for the extended 30-SNP haplotype for all samples.

Statistical Analysis

Allele and genotype frequencies were computed for each SNP. *P* values for HWE were obtained, and Bonferroni adjustment was applied, to control the overall type I error rate at 0.05. Each patient group (sharing the same mutation status) was compared with the controls in their allele and genotype distributions for each SNP. Following haplotype reconstruction, haplotypes from PHASE were selected for comparisons. For each block and the extended block, a number of tests were performed. First, haplotype frequencies in all phenotype groups with distinct mutation statuses were compared using a Pearson χ^2 test, where rare haplotypes (expected frequency <5 for any group) were pooled together to make the χ^2 approximation accurate as determined by the criterion of Cochran.[20] We applied Bonferroni adjustment to the four overall tests, using the significance level of 0.05/4 (0.0125) for each test. Each pair of groups was then compared using a Pearson χ^2 test with the same criterion of pooling the rare haplotypes. If the result of the overall test is statistically significant ($P < .0125$), the subsequent pairwise tests provide more-specific comparisons between groups. The first χ^2 test controls the overall type I error rate, but we further adjust for multiple tests between pairs of groups by using 0.05/6 (0.0083) as the significant level for each such test. Following this, we compared groups with different clinical features in terms of the haplotype frequencies, using the same approach of an overall Pearson χ^2 test and subsequent comparisons of each group (one at a time) with the controls, pooling rare haplotypes in each test as described above. The same set of tests was also performed for the controls and the subset of patients classified as mutation positive or mutation negative. As for the first group of tests, we use 0.0125 as the significance level for each overall test, to adjust for the total number of blocks (four, including three haplotype blocks and the extended block), and 0.0125 as the significance level for each subsequent pairwise comparison, to adjust for the number of groups being compared with the control group in turn.

Results

SNP Analysis and Identification of Hemizygous Deletions

We developed an informative marker set comprised of 30 relatively evenly spaced SNPs (1 SNP every 5.6 kb, with a minor-allele frequency >10%) across a 163-kb region spanning the entire *PTEN* locus and including 30 kb of flanking sequence (fig. 1 and table 1). The majority of identified SNPs are intronic (18 of 30), 11 are outside of the gene (7 upstream and 4 downstream), and 1 is located in the 5′ UTR of *PTEN*. They include 19 transitions, 5 transversions, and 6 insertion/deletion polymorphisms. Table 2 shows the allele frequencies for all 30 polymorphisms genotyped in the control and PHTS-affected patient populations. No significant departures from HWE were observed. Figure 2 summarizes the $-\log_{10}P$ values from comparisons of allele frequencies among *PTEN* mutation–negative, *PTEN* mutation–positive, and *PTEN* variation–positive groups versus the control population. Overall, results from 13 (14%) of the 90 comparisons were significant at the 0.05 level. Specifically, the allele frequency of SNP2 differed significantly among *PTEN* mutation–positive samples and control samples ($P = .0083$). More strikingly, the allele frequencies of SNPs 10, 12, 14, 19, 24, 25, and 27 were all significantly different from the control population among the *PTEN* variation–positive group (*P* values <.01). Additionally, SNP16 and SNP17 both achieved statistical significance for this same comparison ($P = .0127$ and $P = .0123$, respectively).

Of the 447 samples, 33 (7.4%) were found to be homozygous for all 30 SNPs in our panel, including 4 (4.3%) of 94 control samples, 14 (9.5%) of 148 *PTEN* mutation–negative samples, and 15 (7.3%) of 205 *PTEN* mutation/

Table 1. Characteristics of the 30-SNP Panel

SNP	dbSNP ID	Position[a]	Variation (Major/Minor Alleles)	Minor-Allele Frequency[b]	Location[c]
1	*rs7085791*	89583605	G/T	.12	−30602
2	*rs10887756*	89587630	A/T	.15	−26577
3	*rs10887758*	89593295	T/C	.20	−20912
4	*rs11202585*	89598759	G/C	.19	−15448
5	*ss52090924*[d]	89603299	T/C	.20	−10908
6	*rs11202590*	89607699	C/T	.14	−6508
7	*rs1903860*	89610190	T/C	.13	−4017
8	*rs12573787*	89613696	G/A	.14	−510
9	*rs3216482*	89616359	ins/del A	.20	IVS1+2074
10	*rs11355437*	89629037	del/ins G	.40	IVS1−14725
11	*rs2673836*	89629942	A/G	.29	IVS1−13820
12	*ss52090925*[d]	89634206	C/G	.21	IVS1−9556
13	*rs10887763*	89645216	A/G	.14	IVS2+1370
14	*rs3831732*	89645229	ins/del A	.39	IVS2+1382
15	*rs12569872*	89655492	G/A	.14	IVS2+11645
16	*rs1234224*	89665276	A/G	.32	IVS2−9974
17	*ss52090926*[d]	89666296	del/ins 32 nt	.39	IVS2−8954
18	*rs10490920*	89675623	T/C	.14	IVS3+329
19	*rs3830675*	89680936	ins/del TCTTA	.31	IVS4+109
20	*ss52090927*[d]	89689289	del/ins 16 nt	.15	IVS5+6300
21	*rs2299941*	89694699	A/G	.12	IVS5−7156
22	*ss52090928*[d]	89699396	T/C	.21	IVS5−2459
23	*rs2673832*	89702453	A/G	.14	IVS6+457
24	*ss52090929*[d]	89710231	T/C	.22	IVS7−400
25	*rs555895*	89710887	T/G	.31	IVS8+32
26	*rs926091*	89711392	C/T	.14	IVS8+537
27	*rs701848*	89716725	T/C	.39	*614
28	*rs10509532*	89727534	C/T	.14	*12325
29	*rs7908337*	89743671	T/C	.24	*28462
30	*rs11202614*	89745623	C/T	.14	*30414

[a] SNP position on chromosome 10 according to March 2006 human genome assembly, NCBI build 36.1 (hg18).

[b] Frequency in control population.

[c] Location relative to translation start codon (−), *PTEN* exons (IVS), or translation stop codon (*).

[d] SNPs identified by DNA resequencing in our screening set.

variation–positive samples. Because heterozygousity had been identified previously in the *PTEN* mutation/variation–positive samples, *PTEN* copy-number determinations were made only for the control and *PTEN* mutation–negative samples. Elsewhere, we reported that $2^{-\Delta\Delta Ct}$ values close to 1 indicate the presence of two *PTEN* alleles, whereas values close to 0.5 indicate hemizygous *PTEN* deletions.[7] As shown in figure 3, the control samples were found to have average $2^{-\Delta\Delta Ct}$ values of 1.09 ± 0.14 for *PTEN* exon 2 and 1.06 ± 0.20 for *PTEN* exon 5, confirming that these samples retain two copies of *PTEN*. Similarly, a subset of *PTEN* mutation/variation–positive samples had average $2^{-\Delta\Delta Ct}$ values of 0.94 ± 0.14 for *PTEN* exon 2 and 0.97 ± 0.12 for *PTEN* exon 5. Two samples known to harbor hemizygous germline deletions spanning the entire *PTEN* locus displayed average values of 0.67 and 0.53 for the two *PTEN* amplicons. Twelve homozygous *PTEN* mutation–negative samples exhibited $2^{-\Delta\Delta Ct}$ values similar to those observed in the control and *PTEN* mutation/variation–positive samples (1.14–1.66 for *PTEN* exon 2 and 0.95–1.51 for *PTEN* exon 5). Two samples, 1582-02 (0.46 for *PTEN* exon 2 and 0.21 for *PTEN* exon 5) and 2849-01 (0.72 for *PTEN* exon 2 and 0.57 for *PTEN* exon 5), had $2^{-\Delta\Delta Ct}$ values that were consistent with hemizygous deletions. Because of their hemizygous status at this locus, both 1582-02 and 2849-01 were excluded from the subsequent LD and haplotype analyses.

Table 2. Summary of SNP Allele-Frequency Data for Control Sample and PHTS-Affected Patient Populations

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

LD along the PTEN *Locus*

We found three distinct haplotype blocks characterized by strong LD in the control population (fig. 4*A*). Block 1 spans SNP1 (−30602 G/T) to SNP9 (IVS1+2074insA) (33 kb), block 2 spans SNP11 (IVS1−13820 A/G) to SNP21 (IVS5−7156 A/G) (65 kb), and block 3 spans SNP23 (IVS6+457 A/G) to SNP30 (*30414 C/T) (43 kb). Adjacent

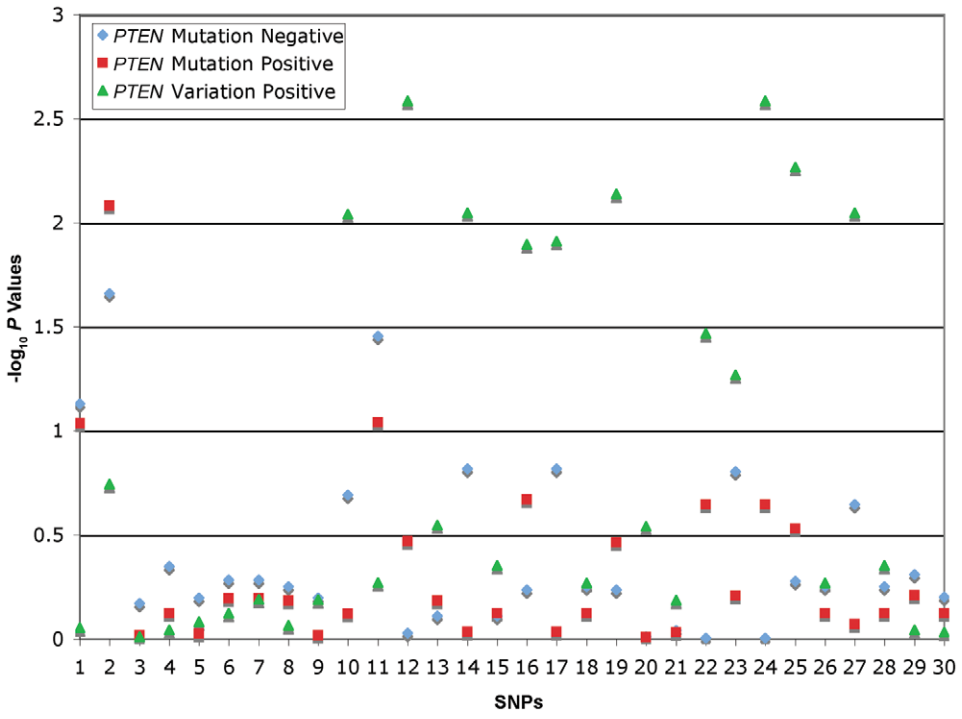

Figure 2. Summary of SNP allele-frequency *P* values for PHTS-affected patient populations versus the control population. Allele frequencies among three PHTS-affected patient populations (*PTEN* mutation–negative, *PTEN* mutation–positive, and *PTEN* variation–positive groups) were compared with the control population for all 30 SNPs by use of a Pearson χ^2 test. The $-\log_{10}$ of the *P* values was plotted for each comparison and for all SNPs. Note that $-\log_{10}$ *P* value of 1 = *P* value of .1, $-\log_{10}$ *P* value of 2 = *P* value of .01, and $-\log_{10}$ *P* value of 3 = *P* value of .001.

to each partitioned block, LD decays. SNP10 (IVS1−14725 delG) displayed average D' values of 0.75 and 0.85 with blocks 1 and 2, respectively, and could not be assigned to either block. Similarly, SNP22 (IVS5−2459 T/C) had an average $D' < 0.90$ and was not in strong LD with either adjacent block, suggesting that both SNPs lie in or near putative recombination hotspots. The *PTEN* haplotype structure in two PHTS-affected patient populations (146 unrelated *PTEN* mutation–negative and 205 unrelated *PTEN* mutation/variation–positive PHTS samples) is shown in figure 4*B* and 4*C*, respectively. Similar to the control population, significant LD was observed for the entire region. However, compared with in controls, the overall LD patterns observed in the PHTS samples appear to be distinct. LD in these samples suggests less recombination of the adjacent blocks and the presence of extended haplotypes across this locus.

Haplotype Association Analysis at the PTEN *Locus*

Having identified three regions of strong LD flanked by two apparent recombination hotspots, we next proceeded to investigate the haplotypes contained within each LD block. Haplotype phase was reconstructed using both the SNPHAP and PHASE software programs. The two algorithms performed similarly, and agreement was reached for 98.8% of the reconstructed haplotype blocks and for 96.5% of the reconstructed chromosomes (i.e., extended haplotypes) (data not shown). PHASE haplotype blocks and haplotype-block frequencies for all chromosomes are shown in table 3. The number of common haplotypes accounting for >80% of the observed chromosomes varied among the three blocks. We identified five common haplotypes for both block 1 and block 2 and a total of seven common haplotypes for block 3. For block 3, the number of common haplotypes also varied among sample groups. The haplotype distributions for each block differed significantly among the examined groups (table 3).

The distribution of the five block 1 haplotypes among controls, *PTEN* mutation–negative patients, *PTEN* mutation–positive patients, and *PTEN* variation–positive patients was significantly different ($\chi^2 = 30.66$; $P = .0098$). Haplotype 1 was found to be underrepresented in *PTEN* mutation–negative samples (49.7%) and overrepresented in the control population (63.8%). Haplotype 2 was overrepresented in *PTEN* mutation–negative and *PTEN* mutation–positive samples compared with both control and *PTEN* variation–positive samples (18.2% and 16.5% vs. 12.2% and 12.3%, respectively). Interestingly, the percentage of low-frequency haplotypes was also overrepresented among both *PTEN* mutation–negative and *PTEN* variation–positive samples (10.3% and 8.8%, respectively) compared with controls (2.7%).

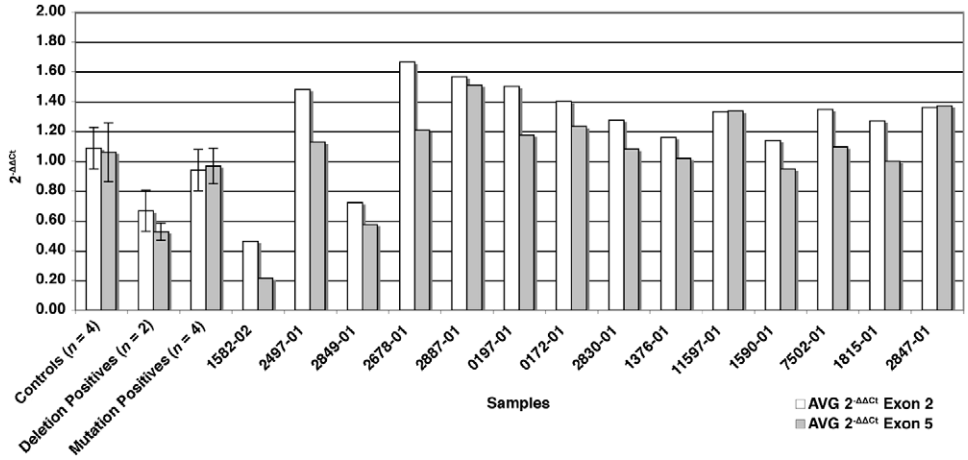

Figure 3. Hemizygous *PTEN* deletion analysis. *PTEN* copy number was estimated at exons 2 and 5 by use of the Livak method for control ($n = 4$), *PTEN* mutation/variation–positive ($n = 4$), and *PTEN* mutation–negative ($n = 14$) samples found to be homozygous for all 30 genotyped SNPs, as well as for known *PTEN* deletion–positive samples ($n = 2$). The $2^{-\Delta\Delta Ct}$ values for the control samples ranged from 0.87 to 1.38. *PTEN* mutation/variation–positive samples (known to have heterozygous *PTEN* mutations/variations) displayed values between 0.75 and 1.13. *PTEN* deletion–positive samples had average (AVG) $2^{-\Delta\Delta Ct}$ values of 0.67 and 0.53 for exons 2 and 5, respectively. Twelve *PTEN* mutation–negative samples had values similar to those of the control and *PTEN* mutation–positive samples (0.95 to 1.66). Two *PTEN* mutation–negative samples (1582-02 and 2849-01) displayed $2^{-\Delta\Delta Ct}$ values similar to those of the *PTEN* deletion–positive samples, ranging from 0.21 to 0.72.

Statistically significant differences were also observed for the haplotype distributions of blocks 2 and 3 between the examined sample populations ($\chi^2 = 45.31$ and 62.53, respectively; $P < .0001$ for both comparisons). For block 2, haplotype 1 was underrepresented in both the *PTEN* mutation–negative samples (19.2%) and the *PTEN* mutation–positive samples (21.4%) compared with control subjects (29.3%). Haplotype 2 was the most frequent haplotype among the *PTEN* variation–positive samples (32.4%) and was overrepresented in this group compared with both control and *PTEN* mutation–negative samples (15.4% and 16.4%, respectively). The converse was observed for haplotype 4; a 9.8% haplotype frequency was seen in the *PTEN* variation–positive samples, compared with 21.3% and 20.2% in the control and *PTEN* mutation–negative samples, respectively.

As observed for block 1, low-frequency haplotypes were also overrepresented in PHTS samples. These haplotypes were overrepresented in both *PTEN* mutation–negative and *PTEN* mutation–positive samples compared with controls for block 2 (8.9% and 9.2% vs. 3.7%, respectively). For block 3, low-frequency haplotypes were only represented in the three PHTS sample groups (2.7% in *PTEN* mutation–negative samples, 2.4% in *PTEN* mutation–positive samples, and 5.4% in *PTEN* variation–positive samples).

Block 3–haplotype 2 was underrepresented in *PTEN* variation–positive samples (9.8%) and overrepresented in the control (21.3%) and *PTEN* mutation–negative (20.5%) populations. As discussed above for block 2–haplotypes 2 and 4 among these same three sample populations, block 3–haplotype 6 also displayed an inverse relationship with block 3–haplotype 2 in *PTEN* variation–positive samples (19.1%) compared with the control (6.9%) and *PTEN* mutation–negative (6.5%) samples. This observation suggests that a founder haplotype is formed by the extended haplotype between blocks 2 and 3 (haplotypes 4 and 2, respectively). Furthermore, an extended haplotype may also exist between block 2–haplotype 2 and block 3–haplotype 6; however, the former appears to be associated with more haplotype diversity (table 4).

To explore genetic associations pertaining to extended haplotypes, we also reconstructed haplotypes spanning all 30 SNPs (table 4). Of all haplotypes observed in our cohort, 10 extended haplotypes represented 81.9%, and 71 additional "rare" extended haplotypes accounted for the remaining 18.1% (data not shown). Statistically significant differences were observed between the sample populations ($\chi^2 = 77.64$; $P = .0001$). Haplotype 2 was observed to be underrepresented in both *PTEN* mutation–negative (8.6%) and *PTEN* mutation–positive (8.7%) samples. This same haplotype was overrepresented in the *PTEN* variation–positive samples (18.6%). Haplotype 5 was overrepresented in the control population (13.8%) and underrepresented in both the *PTEN* mutation–negative and *PTEN* variation–positive groups (7.5% and 5.9%, respectively). Interestingly, extended haplotype 1, the most frequent haplotype observed in all sampled chromosomes (16.0%), was underrepresented in *PTEN* variation–positive samples (9.3%) compared with both control (18.6%) and *PTEN* mutation–negative (19.2%) samples. This haplotype is comprised of block 2–haplotype 4 and block 3–haplotype 2 as well as block 1–haplotype 1 (the most common haplotype observed in this block: ⩾50% in all sample pop-

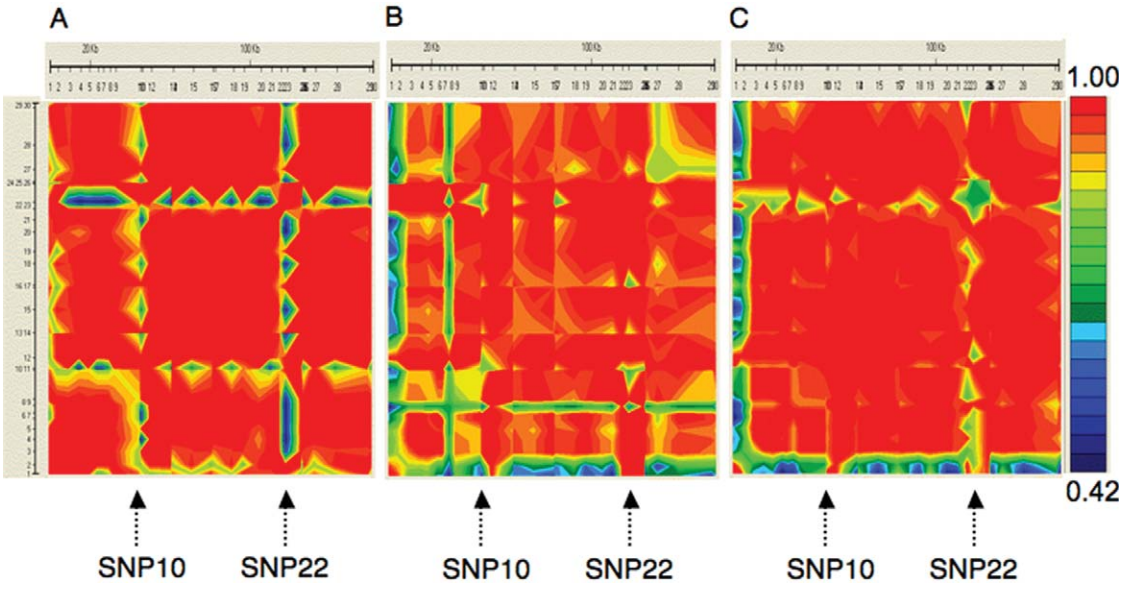

Figure 4. GOLD plot of pairwise LD between 30 SNPs. *D′* values are reported for all three sample groups: 94 control samples (*A*), 146 *PTEN* mutation–negative samples (*B*), and 205 *PTEN* mutation/variation–positive samples (*C*). The control samples display three distinct haplotype blocks: block 1 from SNP1 (−30602 G/T) to SNP9 (IVS1+2074insA), block 2 from SNP11 (IVS1−13820 A/G) to SNP21 (IVS5−7156 A/G), and block 3 from SNP23 (IVS6+457 A/G) to SNP30 (*30414 C/T). SNP10 (IVS1−14725delG) and SNP22 (IVS5−2459 T/C) appear to lie near or within areas of historical recombination. Both the *PTEN* mutation–negative and the *PTEN* mutation/variation–positive samples display varied LD patterns across this locus compared with the control population.

ulations). This strongly suggests that, despite the presence of two recombination hotspots, a founder haplotype likely exists for this region of 10q. Two additional extended haplotypes, 2 and 5, were also observed to be overrepresented in the control population (13.3% and 13.8%, respectively) compared with the *PTEN* mutation–negative group (8.6% and 7.5%, respectively). Haplotype 2 was also underrepresented in *PTEN* mutation–positive samples (8.7%).

Additionally, as observed for each of the three individual blocks, the frequencies of "rare" extended haplotypes were different among the different sample populations, accounting for only 12.8% of control chromosomes, compared with 22.6% and 18.6% of *PTEN* mutation–negative and *PTEN* variation–positive chromosomes, respectively. These data suggest that rare alleles may underlie the disease etiology in these sample populations and, more specifically in the case of the *PTEN* mutation–negative group, may harbor pathogenic variant(s) that escaped detection by "standard" *PTEN* mutation–scanning methodologies.

To examine these associations further, we performed a series of comparative haplotype analyses among PHTS and control samples for haplotype blocks and the extended haplotypes (table 5). A significant difference was observed for block 1 between the *PTEN* mutation–negative and control samples ($\chi^2 = 18.20$; $P = .0027$) (table 5). For *PTEN* variation–positive samples, block 2, block 3, and the extended haplotype all differed significantly from the control population ($\chi^2 = 22.06$, $P = .0005$; $\chi^2 = 37.96$, $P < .0001$; and $\chi^2 = 38.84$, $P < .0001$, respectively). Notably, the allele frequencies of several individual SNPs comprising these haplotype blocks were significantly different among these same two groups (table 2 and fig. 2). A comparison among *PTEN* mutation–negative and *PTEN* variation–positive samples revealed significant differences at these same genomic regions: block 2 ($\chi^2 = 28.65$; $P < .0001$), block 3 ($\chi^2 = 39.97$; $P < .0001$), and the extended haplotype ($\chi^2 = 44.13$; $P < .0001$). In a comparison based on stratification by clinical diagnoses (table 5), block 2, block 3, and the extended haplotype were also associated with patients with CS-like, reaching statistical significance for each of these comparisons ($\chi^2 = 18.46$, $P = .0024$; $\chi^2 = 24.35$, $P = .0010$; and $\chi^2 = 28.02$, $P = .0018$, respectively). A similar trend was observed for this phenotype when the *PTEN* mutation–negative and *PTEN* mutation–positive groups were combined (block 2: $\chi^2 = 13.60$, $P = .0587$; block 3: $\chi^2 = 12.61$, $P = .0273$; and the extended haplotype: $\chi^2 = 21.81$, $P = .0095$) (table 5). Although interesting, only the comparison of the extended haplotype was statistically significant. Additionally, among *PTEN* mutation–negative and *PTEN* mutation–positive patients with CS, block 1 appeared to show an association with this phenotype ($\chi^2 = 14.16$; $P = .0146$), although this result did not reach statistical significance following Bonferroni adjustment.

Discussion

PHTS represents an assemblage of phenotypically diverse syndromes manifested by germline pathogenic mutations in the *PTEN* gene. Standard germline mutation scanning

Table 3. Haplotype Blocks across the *PTEN* Locus

Block and Haplotypes	Controls ($n = 188$)[a]	*PTEN* Mutation Negative ($n = 292$)[a]	*PTEN* Mutation Positive ($n = 206$)[a]	*PTEN* Variation Positive ($n = 204$)[a]
Block 1 haplotypes[b]:				
1. GATGTCTGD	.638	.497	.549	.559
2. TTTGTCTGD	.122	.182	.165	.123
3. GACCCTCAI	.138	.120	.141	.108
4. GTTGTCTGD	.027	.055	.073	.074
5. GACCCCTGI	.048	.045	.044	.049
Low frequency	.027	.103	.029	.088
Block 2 haplotypes[c]:				
1. GCADGAITDIA	.293	.192	.214	.255
2. ACADGGITIIA	.154	.164	.199	.324
3. ACAIGADTDIA	.176	.240	.204	.162
4. AGAIGADTDIA	.213	.202	.165	.098
5. ACGDAGICIDG	.128	.113	.126	.103
Low frequency	.037	.089	.092	.059
Block 3 haplotypes[d]:				
1. ATTCCCTC	.176	.226	.214	.157
2. ACTCCCTC	.213	.205	.160	.098
3. ATTCTCTC	.160	.123	.136	.216
4. ATGTTTCT	.144	.154	.150	.098
5. ATGCTCCC	.101	.110	.107	.118
6. ATGCTCTC	.069	.065	.097	.191
7. GTTCTCTC	.138	.089	.117	.069
Low frequency	.000	.027	.024	.054

[a] n = Number of haplotypes.
[b] $\chi^2 = 30.66$; $P = .0098$.
[c] $\chi^2 = 45.31$; $P < .0001$.
[d] $\chi^2 = 62.53$; $P < .0001$.

has identified causal variants in a majority of patients diagnosed with this complex disorder, particularly in patients who received a diagnosis of CS or BRRS.[1,6] Despite extensive mutation scanning, however, the etiologic variant(s) have yet to be identified in 15% and 35% of patients with these two syndromes. To investigate genetic associations with *PTEN* in this subset of patients, as well as to characterize the haplotype architecture of this locus, we chose to use a case-control haplotype-based approach.

Similar approaches have been used to examine genetic associations at a growing number of candidate genes.[21–23] Haplotype-based approaches are of particular interest, since most reports of disease-associated mutations describe variants that directly alter the protein-coding sequence of a gene. These studies fail to consider other mechanisms that may alter gene function and, where mutations are not found, may overlook polymorphisms that reside outside of the coding region. Such mechanisms include alterations of gene regulation through the disruption of *trans*-acting factor(s) and *cis*-acting sequence-element interactions, resulting in a pathologic state.[24]

While the mutation spectrum of *PTEN* in PHTS has been well studied, its haplotype architecture has not. The extent of LD across this region has been examined in three previous studies.[25–27] Hamilton et al.[26] first reported the existence of two distinct four-marker haplotypes in the general population but found no association with prostate cancer and benign prostatic hyperplasia. A study by Zhang et al.[27] examined the association of this same locus with smoking initiation and nicotine addiction, by use of five haplotype-tagging SNPs (htSNPs) selected using the SNPbrowser software program (Applied Biosystems). In this study, three haplotype blocks were observed; block 1 spanned 41 kb (from nucleotide position 89606485 to 89647130), block 2 spanned 16 kb (from nucleotide position 89679301 to 89695409), and block 3 included a single SNP located at position 89716724. As the authors noted, this differed slightly from the *PTEN* haplotype structure observed by the International HapMap Project. Most recently, Haiman et al.[25] investigated the influence of common variations across this region and the risk of sporadic breast and prostate cancer. Also employing a htSNP approach, these authors identified nine common haplotypes representing >87% of all chromosomes across 123 kb of the *PTEN* locus. Among these common haplotypes, no strong association was found with either type of sporadic cancer.

For the present study, haplotype phase was reconstructed for our samples by use of the SNPHAP software program, based on pairwise LD measurements and the EM algorithm.[28,29] Previous studies have demonstrated the appropriateness of the EM algorithm for inference of haplotypes from data obtained from unrelated individuals.[28,30–32] Because our analysis relied on statistical inferences of haplotypes from unphased data, we chose to validate this reconstruction by using a second algorithm based on a Bayesian approach as implemented in the PHASE software program.[19,33] Although the two programs

Table 4. Extended Haplotypes for All 30 SNPs across the *PTEN* Locus

PTEN Extended Haplotypes	Total (n = 890)[a]	Controls (n = 188)[a]	*PTEN* Mutation Negative (n = 292)[a]	*PTEN* Mutation Positive (n = 206)[a]	*PTEN* Variation Positive (n = 204)[a]
1. GATGTCTGDDAGAIGADTDIACACTCCCTC	.160	.186	.192	.155	.093
2. GATGTCTGDIGCADGAITDIATATTCTCTC	.119	.133	.086	.087	.186
3. TTTGTCTGDDACAIGADTDIATATTCCCTC	.113	.101	.137	.121	.083
4. GACCCTCAIIACGDAGICIDGTATGTTTCT	.099	.117	.082	.117	.088
5. GATGTCTGDIGCADGAITDIATGTTCTCTC	.092	.138	.075	.107	.059
6. GATGTCTGDIACADGGITIIATATGCTCTC	.064	.027	.031	.073	.137
7. GATGTCTGDIACADGGITIIATATGCTCCC	.054	.048	.055	.063	.049
8. GACCCCTGIIACADGGITIIATATGCTCCC	.044	.048	.038	.044	.049
9. GATGTCTGDDACAIGADTDIATATTCCCTC	.039	.059	.048	.029	.020
10. GTTGTCTGDDACAIGADTDIATATTCCCTC	.035	.016	.031	.044	.049
Low Frequency	.181	.128	.226	.16	.186

NOTE.—$\chi^2 = 77.64$; $P < .0001$.

[a] n = Number of haplotypes.

rely on different mathematical approaches, both algorithms performed remarkably similarly.

Our analysis of the LD structure across this region of 10q revealed three distinct haplotype blocks; block 1 spans 33 kb (from nucleotide position 89583605 to 89616359), block 2 spans 65 kb (from nucleotide position 89629942 to 89694699), and block 3 spans 43 kb (from nucleotide position 89702453 to 89745623). Block 2 is flanked by regions of decreased LD, suggesting that SNPs at these sites lie within areas of chromosome recombination. Our block partitioning, based on the method by Gabriel et al.,[17] partially agreed with that described by Zhang et al.[16] However, on the basis of our data, block 1 described by Zhang et al. is actually made up of two distinct blocks.[16] As previously mentioned, these authors defined this region by using two htSNPs. To ensure the accurate characterization of this region, we chose to empirically assess its haplotype architecture, using a high-density set of polymorphic markers. Because the extent of LD is variable in this region, the htSNP approach failed to capture all pertinent information regarding the locus in question, specifically regarding the breakdown of LD observed at SNP10 (IVS1−14725delG) and SNP22 (IVS5−2459 T/C). Therefore, a more dense marker set is required. htSNP approaches are capable of capturing most haplotype diversity within a population—that is, ~90% of all chromosomes in a given population.[17] However, for uncommon haplotypes, particularly in cases where the causal allele is underrepresented, this approach is limited. Our finding that "rare" haplotype blocks account for two-to-threefold more PHTS chromosomes than control chromosomes and "rare" extended haplotypes account for nearly twofold more *PTEN* mutation–negative and *PTEN* variation–positive chromosomes suggests that, for rare diseases such as PHTS, low-frequency, or "rare," haplotypes are the ones associated with disease and may harbor pathogenic variants.

In our effort to characterize the haplotype architecture of the *PTEN* locus, we identified two PHTS-affected patients, 1582-02 and 2849-01, with hemizygous microdeletions. Each retained only a single copy of the *PTEN* allele; 1582-02 retained extended haplotype 4, and 2849-01 retained extended haplotype 5. These haplotypes had allele frequencies of 9.9% and 9.2%, respectively, in the entire sample population, resulting in a <1% chance of homozygosity for these alleles. By contrast, three of the four homozygous control samples were homozygous for the most frequent haplotype observed in our study. On the basis of the analysis of microsatellite markers, these deletions span <~312 kb to 390 kb, respectively (data not shown). Previously, we identified *PTEN* deletions in only three patients with PHTS, all of whom were clinically diagnosed with BRRS or CS/BRRS overlap.[7] The patients identified in the current study have diagnoses of classic CS (2849-01) and CS-like phenotype (1582-02). Implications from these data extend to the clinical realm, suggesting that *PTEN* deletion analysis is warranted in all PHTS-affected patients with CS, BRRS, CS/BRRS, and CS-like phenotypes who lack apparent germline mutations.

Interestingly, one *PTEN* mutation–negative sample was homozygous for a "rare" extended haplotype with an allele frequency <0.7% in the entire study population. Close inspection of this haplotype revealed that blocks 2 and 3 were relatively common, whereas block 1 consisted of a low-frequency block. This low-frequency haplotype block, GACCCTCGI, was only observed in eight samples: seven *PTEN* mutation–negative samples and one *PTEN* variation–positive sample. Carriers of this allele include four patients with CS, three patients with CS-like, and one patient with CS/BRRS. For our homozygous sample, this suggests that, because of the locations of our amplicons, our deletion analysis may have been unable to detect a possible deletion of the 5′ region of this locus. These data implicate the GACCCTCGI block as a low-frequency, highly penetrant PHTS-susceptibility allele. Furthermore, all eight samples have similar "rare" extended haplotypes; five (three CS and two CS-like) share the same haplotype, one (CS/BRRS) deviates from this haplotype by a single variation in block 2, and two (one CS and one CS-like) are variable for both block 2 and block 3. Although the SNPs that make up this block and extended haplotype are not causal (based on their frequency in the control population), they are likely in LD with an unknown functional

Table 5. Comparative Haplotype Analysis

Comparison	Block 1		Block 2		Block 3		Extended Haplotype	
	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*
Mutation status[a]:								
PTEN Mut− vs. control	18.20	**.0027**	12.03	.0614	10.44	.1649	17.27	.0447
PTEN Mut+ vs. control	6.78	.2376	9.66	.0854	8.67	.2771	13.34	.2054
PTEN Var+ vs. control	12.34	.0304	22.06	**.0005**	37.96	**<.0001**	38.84	**<.0001**
PTEN Mut− vs. *PTEN* Mut+	10.91	.0531	3.41	.7566	3.83	.7987	13.05	.2899
PTEN Mut− vs. *PTEN* Var+	5.02	.5415	28.65	**<.0001**	39.97	**<.0001**	44.13	**<.0001**
PTEN Mut+ vs. *PTEN* Var+	8.38	.1364	13.82	.0318	21.65	**.0029**	20.31	.0161
Clinical features[b]:								
Overall	9.32	.3162	29.76	.0193	26.42	.0484	7.98	.0924
CS vs. control	12.36	.0302	7.61	.1788	10.08	.1841	15.51	.1147
BRRS vs. control	1.57	.6667	9.87	.0789	10.03	.1233	5.07	.4065
CS/BRRS vs. control	1.87	.3932	9.19	.0564	1.31	.8600	.49	.4825
CS-like vs. control	12.94	.0240	18.46	**.0024**	24.35	**.0010**	28.02	**.0018**
Mutation status and clinical features[c]:								
Overall	8.82	.0659	13.41	.0984	23.70	.0220	3.58	.4700
PTEN Mut− and Mut+ CS vs. control	14.16	.0146	12.40	.0883	10.12	.0720	11.98	.1519
PTEN Mut− and Mut+ BRRS vs. control	.96	.8107	9.04	.1715	9.02	.1083	4.11	.5339
PTEN Mut− and Mut+ CS/BRRS vs. control	.70	.4027	.04	.8415	4.32	.1155	.29	.5890
PTEN Mut− and Mut+ CS-like vs. control	11.35	.0449	13.60	.0587	12.61	.0273	21.81	**.0095**

NOTE.—Significant results are indicated in bold type. Mut+ = mutation positive; Mut− = mutation negative; Var+ = variation positive.

[a] Patients with PHTS were stratified on the basis of their *PTEN* mutation status and were compared with controls as well as each other. The Bonferroni-adjusted nominal significance level used for this comparison was $P < .0083$.

[b] An overall comparison was made on the basis of stratification of clinical features followed by comparisons based on clinical diagnoses (CS, BRRS, CS/BRRS, or CS-like phenotype) for all patient samples, irrespective of mutation status, which were compared with controls. The Bonferroni-adjusted nominal significance level used for this comparison was $P < .0125$.

[c] Overall comparisons of patient clinical diagnoses among *PTEN* mutation–negative and *PTEN* mutation–positive samples were performed, followed by comparisons made among this group versus control samples on the basis of patient clinical diagnosis. The Bonferroni-adjusted nominal significance level used for this comparison was $P < .0125$.

variant conferring disease susceptibility. This further supports the notion that "rare," low-frequency alleles (LD blocks and/or extended haplotypes) may be associated with disease and should therefore be considered as candidate susceptibility alleles in rare disorders.

In addition to an association with rare haplotypes, our analysis of haplotype blocks and extended haplotypes revealed significant differences among the control group and various patient sample populations. The number and frequency of common haplotypes needed to cover >80% of the observed chromosomes varied for each of the three blocks and the extended haplotype. Similar to the association with rare alleles, these data suggest greater haplotype diversity among the PHTS-affected patient populations compared with the control group and are indicative of a higher degree of recombination of the "ancient haplotype." Interestingly, the overall LD pattern observed in our patient samples appears to indicate the presence of extended haplotypes. This effect seemed most apparent when *PTEN* variation–positive patients were compared with controls, revealing significant differences between these groups for blocks 2 and 3, as well as for the extended haplotype, and suggesting less recombination among patients with PHTS. Furthermore, pairwise comparisons between groups revealed that the *PTEN* mutation–negative and *PTEN* mutation–positive groups were most similar, suggesting that different pathogenic variants may have arisen from similar haplotypic backgrounds. Taken together, these data indicate that some patients with PHTS—that is, *PTEN* mutation–positive individuals and perhaps *PTEN* variation–positive individuals—exhibit a haplotype-founder effect, while others—that is, *PTEN* mutation–negative individuals—harbor rare extended haplotypes that have undergone extensive "shuffling" of the LD blocks across this region.

Interestingly, among *PTEN* mutation–negative samples, the strongest genetic effect appears to be associated with haplotypes forming block 1 (a block spanning at least 30 kb upstream of *PTEN* that includes several kilobasepairs of the gene's first intron). With the exception of *PTEN*'s core promoter and exon 1, this region has not been well characterized. Screening efforts that have failed to identify mutations/variations at these sites in this group of patients suggest that alterations in this region may have a role in the regulation of *PTEN*. These likely involve novel regulatory elements and contribute to its deregulation.

Various PHTSs, such as BRRS and CS, appear to be caused by the same *PTEN* mutations, despite clear differences in phenotypic presentation.[1] The R130X mutation in exon 5, for example, occurs in eight *PTEN* mutation–positive patients included in this study. Among these individuals, three have a clinical diagnosis of CS, two have a clinical

diagnosis of BRRS, and three have a clinical diagnosis of CS/BRRS. Both individuals with BRRS are carriers of extended haplotypes 3 and 10 and exhibit classic features of BRRS, including macrocephaly, lipomas, and pigmented macules of the penis. The probability of this genotype in the general population is <0.3%, suggesting that this infrequent allelic combination likely contributes to their phenotype and that low-penetrance functional variants reside on these loci. Furthermore, although stratification by clinical phenotype was only minimally associated with our haplotypes, correlations from these data become more apparent when the patient's mutation status is considered.

In addition to providing a panel of informative markers for testing genetic associations at the *PTEN* locus, our data strongly suggest that specific haplotypes along this region are associated with increased PHTS susceptibility. *PTEN* mutation–negative samples lacking traditional mutations in the *PTEN* coding sequence possess a significantly different haplotype architecture than that of control samples. Along with an association with block 1 of this locus, "rare" alleles comprise this architecture and may underlie the disease etiology in these patients. Furthermore, haplotype profiles in patients with PHTS with known mutations/variations contribute to the phenotypic complexity of this syndrome. Although the mechanisms underlying these relationships have yet to be elucidated, these data suggest that associated chromosomal segments likely harbor variants, potentially involved in the transcriptional regulation of *PTEN,* which are both pathogenic and/or modifying in nature and manifest as low-penetrance disease-susceptibility alleles.

Acknowledgments

This work was partially supported by the American Cancer Society (RSG-02-151-01-CCE [to C.E.]). C.E. is a recipient of the Doris Duke Distinguished Clinical Scientist Award. M.G.P. is a predoctoral fellow in the Cleveland Clinic Lerner Research Institute and is also a graduate student of the Integrated Biomedical Sciences Graduate Program of The Ohio State University. We thank Drs. Kevin Zbuk, Frank Weber, and Kristin Waite for helpful discussions in preparation of the manuscript.

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
Gonçalo Abecasis' Software Page, http://www.sph.umich.edu/csg/abecasis/software.html (for GOLD software)
HapBlock, http://www.cmb.usc.edu/msms/HapBlock/
International HapMap Project, http://www.hapmap.org/
National Comprehensive Cancer Network, http://www.nccn.org
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *PTEN,* CS, BRRS, and PS)
PHASE software, http://www.stat.washington.edu/stephens/software.html
SNPHAP, http://www-gene.cimr.cam.ac.uk/clayton/software/snphap.txt

References

1. Eng C (2003) PTEN: one gene, many syndromes. Hum Mutat 22:183–198
2. Maehama T, Dixon JE (1998) The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. J Biol Chem 273: 13375–13378
3. Stambolic V, Suzuki A, de la Pompa JL, Brothers GM, Mirtsos C, Sasaki T, Ruland J, Penninger JM, Siderovski DP, Mak TW (1998) Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. Cell 95:29–39
4. Marsh DJ, Kum JB, Lunetta KL, Bennett MJ, Gorlin RJ, Ahmed SF, Bodurtha J, et al (1999) PTEN mutation spectrum and genotype-phenotype correlations in Bannayan-Riley-Ruvalcaba syndrome suggest a single entity with Cowden syndrome. Hum Mol Genet 8:1461–1472
5. Eng C (2000) Will the real Cowden syndrome please stand up: revised diagnostic criteria. J Med Genet 37:828–830
6. Pilarski R, Eng C (2004) Will the real Cowden syndrome please stand up (again)? Expanding mutational and clinical spectra of the PTEN hamartoma tumour syndrome. J Med Genet 41:323–326
7. Zhou XP, Waite KA, Pilarski R, Hampel H, Fernandez MJ, Bos C, Dasouki M, Feldman GL, Greenberg LA, Ivanovich J, Matloff E, Patterson A, Pierpont ME, Russo D, Nassif NT, Eng C (2003) Germline PTEN promoter mutations and deletions in Cowden/Bannayan-Riley-Ruvalcaba syndrome result in aberrant PTEN protein and dysregulation of the phosphoinositol-3-kinase/Akt pathway. Am J Hum Genet 73:404–411
8. Smith JM, Kirk EP, Theodosopoulos G, Marshall GM, Walker J, Rogers M, Field M, Brereton JJ, Marsh DJ (2002) Germline mutation of the tumour suppressor PTEN in Proteus syndrome. J Med Genet 39:937–940
9. Zhou X, Hampel H, Thiele H, Gorlin RJ, Hennekam RC, Parisi M, Winter RM, Eng C (2001) Association of germline mutation in the PTEN tumour suppressor gene and Proteus and Proteus-like syndromes. Lancet 358:210–211
10. Loffeld A, McLellan NJ, Cole T, Payne SJ, Fricker D, Moss C (2006) Epidermal naevus in Proteus syndrome showing loss of heterozygosity for an inherited PTEN mutation. Br J Dermatol 154:1194–1198
11. Bonneau D, Longy M (2000) Mutations of the human PTEN gene. Hum Mutat 16:109–122
12. Nelen MR, Padberg GW, Peeters EA, Lin AY, van den Helm B, Frants RR, Coulon V, Goldstein AM, van Reen MM, Easton DF, Eeles RA, Hodgsen S, Mulvihill JJ, Murday VA, Tucker MA, Mariman EC, Starink TM, Ponder BA, Ropers HH, Kremer H, Longy M, Eng C (1996) Localization of the gene for Cowden disease to chromosome 10q22-23. Nat Genet 13:114–116
13. Carethers JM, Furnari FB, Zigman AF, Lavine JE, Jones MC, Graham GE, Teebi AS, Huang HJ, Ha HT, Chauhan DP, Chang CL, Cavenee WK, Boland CR (1998) Absence of PTEN/MMAC1 germ-line mutations in sporadic Bannayan-Riley-Ruvalcaba syndrome. Cancer Res 58:2724–2726
14. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. Methods 25:402–408
15. Abecasis GR, Cookson WO (2000) GOLD—graphical overview of linkage disequilibrium. Bioinformatics 16:182–183
16. Zhang K, Deng M, Chen T, Waterman MS, Sun F (2002) A

dynamic programming algorithm for haplotype block partitioning. Proc Natl Acad Sci USA 99:7335–7339
17. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229
18. Clayton D, Chapman J, Cooper J (2004) Use of unphased multilocus genotype data in indirect association studies. Genet Epidemiol 27:415–428
19. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989
20. Cochran W (1954) Some methods of strengthening the common chi-square tests. Biometrics 10:417–451
21. Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB (2000) Complex promoter and coding region β_2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci USA 97:10483–10488
22. Greenwood TA, Kelsoe JR (2003) Promoter and intronic variants affect the transcriptional regulation of the human dopamine transporter gene. Genomics 82:511–520
23. Yu C, Zhou Y, Miao X, Xiong P, Tan W, Lin D (2004) Functional haplotypes in the promoter of matrix metalloproteinase-2 predict risk of the occurrence and metastasis of esophageal cancer. Cancer Res 64:7622–7628
24. Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76:8–32
25. Haiman CA, Stram DO, Cheng I, Giorgi EE, Pooler L, Penney K, Le Marchand L, Henderson BE, Freedman ML (2006) Common genetic variation at PTEN and risk of sporadic breast and prostate cancer. Cancer Epidemiol Biomarkers Prev 15: 1021–1025
26. Hamilton JA, Stewart LM, Ajayi L, Gray IC, Gray NE, Roberts KG, Watson GJ, Kaisary AV, Snary D (2000) The expression profile for the tumour suppressor gene PTEN and associated polymorphic markers. Br J Cancer 82:1671–1676
27. Zhang L, Kendler KS, Chen X (2006) Association of the phosphatase and tensin homolog gene (PTEN) with smoking initiation and nicotine dependence. Am J Med Genet B Neuropsychiatr Genet 141:10–14
28. Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927
29. Lewontin RC (1964) The interaction of selection and linkage. II. Optimum models. Genetics 50:757–782
30. Bonnen PE, Wang PJ, Kimmel M, Chakraborty R, Nelson DL (2002) Haplotype and linkage disequilibrium architecture for human cancer-associated genes. Genome Res 12:1846–1853
31. Niu T, Qin ZS, Xu X, Liu JS (2002) Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. Am J Hum Genet 70:157–169
32. Tishkoff SA, Pakstis AJ, Ruano G, Kidd KK (2000) The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus. Am J Hum Genet 67: 518–522
33. Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–1169