Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Biomarkers in Medicine

Ethnogeographic prevalence and implications of the 677C>T and 1298A>C *MTHFR* polymorphisms in US primary care populations

James S Graydon[1], Karla Claudio[2], Seth Baker[3], Mohan Kocherla[1], Mark Ferreira[1], Abiel Roche-Lima[4], Jovaniel Rodríguez-Maldonado[4], Jorge Duconge[2] & Gualberto Ruaño*[,1]

[1]Laboratory of Personalized Health, Genomas, Inc., Hartford, CT 06102, USA

[2]Pharmaceutical Sciences department, University of Puerto Rico School of Pharmacy, San Juan, PR 00936, USA

[3]Clinical Laboratory Partners, Hartford Healthcare, Hartford, CT 06102, USA

[4]Center for Collaborative Research in Health Disparities, University of Puerto Rico School of Medicine, San Juan, PR 00936, USA

*Author for correspondence: Tel.: +860 972 4574; Fax: +860 545 4575; gualberto.ruano@hhchealth.org

Aim: Variants of the *MTHFR* gene have been associated with a wide range of diseases. **Materials & methods:** The present study analyzed data from clinical genotyping of *MTHFR 677C>T* and *1298A>C* in 1405 patients in urban primary care settings. **Results:** Striking differences in ethnogeographic frequencies of *MTHFR* polymorphisms were observed. African–Americans appear to be protected from MTHFR deficiency. Hispanics and Caucasians may be at elevated risk due to increased frequencies of *677C>T* and *1298A>C*, respectively. **Conclusion:** Individuals carrying mutations for both genes were rare and doubly homozygous mutants were absent, suggesting the *TTcc* is extremely rare in the greater population. The results suggest multilocus *MTHFR* genotyping may yield deeper insight into the ethnogeographic association between *MTHFR* variants and disease.

First draft submitted: 26 October 2018; Accepted for publication: 25 March 2019; Published online: 3 June 2019

Keywords: ethnic distributions • genome evolution • *MTHFR* haplotypes

Methylenetetrahydrofolate reductase (MTHFR) is the rate-limiting enzyme in the methyl cycle and encoded by the *MTHFR* gene. MTHFR catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, a co-substrate for the formation of tetrahydrofolate and methionine from homocysteine. Tetrahydrofolate formation is essential for the synthesis of macromolecules (DNA and RNA) required for cell division, and methionine is necessary for the synthesis of amino acids and proteins.

MTHFR deficiency may predispose individuals to hyperhomocysteinemia (high blood homocysteine levels) in a dynamic interaction between genetics and nutrition [1]. Given the essential role of folate in the human body, supplementation is traditionally recommended when MTHFR is deficient. Low dietary intake of folic acid can cause mild hyperhomocysteinemia, mainly in individuals with *MTHFR* mutations [2]. MTHFR deficiency has been associated with higher susceptibility to coronary heart disease [3], neural tube defects [4,5], dementia [6], mental disorders [7] and several forms of cancer [8], though many of these associations are only apparent when observed alongside low dietary intake of B vitamins or folate or with elevated homocysteine levels. Additionally, diabetic peripheral neuropathy and retinopathy were observed to occur more frequently among individuals with hyperhomocysteinemia [9]. Furthermore, MTHFR deficiency has also been associated with an increased risk of drug-induced toxicity caused by drugs that affect folate homeostasis (e.g., methotrexate) [10].

The *MTHFR* genomic locus is located on the short arm of chromosome 1 (*1p36.22*) and spans 20,373 base pairs (Supplementary Figure 1). Polymorphisms in the *MTHFR* gene are associated with MTHFR deficiency [2]. The *MTHFR* nucleotide at position *677* has two possible alleles: *C* (wild type) or *T* (variant allele, identified as rs1801133). Another *MTHFR* polymorphism at position *1298* has been described and contains two possible alleles: *A* (wild type) or *C* (variant allele, identified as rs1801131).

Future Medicine

MTHFR 677C>T encodes for a valine substitution of alanine at amino acid 222 (Ala222Val) in exon 4, which results in a deficient, thermolabile enzyme (as determined by residual activity after heat inactivation). The degree of enzyme thermolability is much greater in individuals homozygous for the variant allele (*677 TT*) who have 20% residual function compared with heterozygous individuals (*677 CT*) who have 56% and homozygous individuals for the wild-type allele (*677 CC*) who have 66% residual function [11]. Individuals homozygous for the variant allele (*677 TT*) are predisposed to mild hyperhomocysteinemia because they have less active MTHFR available to produce 5-methyltetrahydrofolate resulting in homocysteine accumulation. *MTHFR 677C>T* affects the binding site of the cofactor FAD [12,13]. As a consequence, the protein encoded by *677 T* (222 Val) loses its FAD cofactor three-times faster than the functional protein.

MTHFR 1298A>C encodes for an alanine substitution of glutamic acid at amino acid 429 (Glu429Ala) of exon 7. The polymorphism has not been observed to affect the thermolability of MTHFR [14]. Additionally, these studies show that neither FAD release nor the protective effect of 5-methyl-THF or homocysteine levels was affected. However, a study in a Canadian population showed reduced enzyme activity in homozygous individuals for *MTHFR 1298A>C* [15]. A simulation model suggested that this variation lies in the SAM regulation domain. Consequently, the binding of SAM induces conformational changes to MTHFR that blocks its activity [13,16].

According to the 1000 Genomes Project, approximately 25% of the global population are carriers of *MTHFR 677C>T*, Hispanics being the population with the highest frequency (47%), followed by Europeans (36%), East Asians (30%), South Asians (12%) and Africans (9%) [17]. Approximately 13.5% of Europeans are homozygous for the variant allele. *MTHFR 1298A>C* is also present in about 25% of the global population and occurs more frequently in South East Asians (42%) and Europeans (31%). The frequency of the *1298A>C* polymorphism in Hispanics and Africans is 15%. Approximately 11% of Europeans are homozygous for the *1298A>C* variant allele.

While the frequencies of individual polymorphisms at loci *677* and *1298* have been examined, the frequencies of their configurations, diplotypes and haplotypes, have not been well ascertained, especially across ethnicities. A study performed by Kałużna *et al.* found that haplotypes, *Ca*, *Cc* and *Ta*, were present in approximately equal proportions in individuals with and without acute lymphoblastic leukemia, though the *Tc* haplotype was found only in patients with the disease [18]. A study by Stangler Herodež *et al.* observed a wider distribution in haplotype frequencies and found that the *Tc* haplotype was associated with spontaneous abortion [19]. In both studies, the subject population was entirely Caucasian, limiting the applicability of the results in other ethnicities. A third study conducted in a wide range of Chinese ethnicities found an even greater distribution of haplotype frequencies, indicating a strong association between ethnicity and *MTHFR* haplotypes [20]. However, while studies have begun to examine multiple loci of *MTHFR*, many studies have and still focus only on single mutations and do not consider haplotypes. Additionally, considering haplotypes between ethnicities may yield greater insight into the risk of *MTHFR*-associated diseases in diverse populations.

The complex and somewhat ambiguous relationship of *MTHFR* polymorphisms with various disease states suggests that a combinatorial approach may yield greater clarity. Understanding the association between *MTHFR* combinatory polymorphisms and ethnicity may help reveal at-risk populations and maximize the efficacy of preventative efforts. Therefore, the present study analyzes data collected from *MTHFR* genotyping of loci *677* and *1298* in various ethnic populations and employs diplotype and haplotype analysis to provide greater insight into the complex landscape of *MTHFR* polymorphisms and ethnicity.

Materials & methods

Participants

The study cohort was comprised of 1405 patients referred to the Genomas Laboratory of Personalized Health (LPH) from February 2014 to October 2015. *MTHFR* diagnostic genotyping was part of their clinical care and disease management. They were referred by healthcare providers in primary care, general internal medicine and psychiatry from across the USA. Patient ethnicity was self-reported as African–American, Caucasian or Hispanic. In the present study, 'Hispanic' refers to all individuals who self-identified as Hispanic or Latino regardless of their familial lineage or familial country of origin. The majority of Hispanics (76%) were referred from the Northeast of the USA, suggesting that the majority of Hispanics tested were Caribbean descendants (Puerto Rico, Dominican Republic and Cuba). Additional ethnicities were reported but were excluded from subsequent analysis due to insufficient sample sizes.

Genotyping

DNA extraction and analysis were performed at the LPH. DNA was extracted from epithelial tissue obtained with buccal swabs using the Qiagen EZ-1 DNA Tissue Kit and EZ-1 bio-robotic instrument (Qiagen, MD, USA). The eSensor® Genotyping Test was utilized as an *in vitro* diagnostic for the detection and genotyping of point polymorphisms (*C* to *T* at position *677*) and (*A* to *C* at position *1298*) of the human 5, 10 *MTHFR* gene [21]. The GenMark eSensor® Thrombosis Risk Test was performed on the XT-8 instrument (GenMark Diagnostics, Inc., CA, USA) to genotype the samples using a solid-phase electrochemical reaction. In the XT-8 instrument, simultaneous hybridization of a capture probe, target DNA sequence and ferrocene-bearing signal probe complex occurs on an electrode housed in a cartridge which causes the ferrous ion in each ferrocene group to undergo cyclic oxidation and reduction, which leads to the generation of an alternating current at the electrode surface. The alternating current is measured using voltammetry. Signals from the ferrocene labels are detected and measured by instrument software and the ratio of signals allows for automated identification of the genotype. Genotyping boundaries and signal threshold for each polymorphism are preprogrammed into instrument software. LPH further ensures accuracy in genotyping through participation in College of American Pathologists quality control testing.

Haplotype phasing

The Beagle phasing algorithm was used for the statistical haplotype phasing [22]. The patients were grouped by three ethnic groups (African–American, Caucasian and Hispanic). For phasing implementation with the Beagle software, a reference file with phased individuals for each ethnic group was required as a parameter. The references were obtained from the 1000 Genomes Project [17]. The reference population and data obtained from the 1000 Genomes Project used as a parameter for statistical phasing with Beagle software were as follows, organized by ethnic group: for the Caucasian cohort, Northern Europeans from UT, USA *(code CEU*; European Superpopulation); for the African cohort, African Ancestry in Southwest USA (code ASW; African Superpopulation); for the Hispanic cohort, Mexican Ancestry from Los Angeles, CA, USA (code MXL), Puerto Ricans from Puerto Rico *(code PUR)*, Colombians from Medellin, Colombia (code CLM) and Peruvians from Lima, Peru (code PEL; all in the Americas Superpopulation).

For Hispanics, all the populations were selected (e.g., MXL, PUR, CLM and PEL) 'Americas' was the superpopulation. This criterion was justified because the Hispanic ethnic group in the present study originated from different backgrounds, which included individuals of Mexican, Puerto Rican and Peruvian descent – among others. CEU and ASW references were selected because these represent the most frequent Caucasian and African populations in the USA [23]. As a result, variant call format files were obtained for each ethnic group with the phased haplotype. These files were processed and the final phased haplotypes were determined.

Frequency analysis

Frequency values were obtained for each allele within each cohort and normalized for cohort size across the entire sample. The standard and normalized frequencies were calculated to quantify the contribution of individual alleles to both a specific ethnicity and to the total cohort. Within a given ethnic group, the standard frequency of an allele was obtained by simply dividing the count for a given allele by the number of individuals in the ethnic group. In the case of haplotypes, the sample size was twice the number of patients. For the normalized frequency values for a given allele across the entire sample, the ethnic frequency of the allele in a given ethnic group was divided by the sum of each of the frequencies for that allele in the three ethnic groups (African–American, Caucasian and Hispanic).

F_{ST} calculations

Wright's Fixation Index (F_{ST}) values were calculated by performing pairwise comparisons of the corresponding variant (i.e., genotype, diplotype or haplotype) frequencies in all subgroups. Accordingly, the Wright's F_{ST} metric to variant frequencies was applied in this study to characterize differentiation at two specific loci of clinical and pharmacological relevance in the *MTHFR* gene, *677C>T* and *1298A>C*, as well as to identify substructures within the study cohort. The formula to calculate F_{ST} was as follows [24,25]:

$$F_{ST} = \frac{(\mathrm{POP}_1 - \mathrm{POP}_2)^2}{(\mathrm{POP}_1 + \mathrm{POP}_2)(2 - \mathrm{POP}_1 - \mathrm{POP}_2)}$$

Table 1. Demographics for study cohort (n = 1405).

Age (years)	Ethnicity			Total, n (%)
	African–American	Caucasion	Hispanic	
Female				
10–19	1	10	4	15 (2%)
20–29	3	16	5	24 (3%)
30–39	6	25	15	46 (6%)
40–49	12	35	12	59 (7%)
50–59	26	78	24	128 (16%)
60–69	59	98	31	189 (23%)
70–79	69	109	37	215 (26%)
80–89	33	53	26	112 (14%)
>90	3	18	2	23 (3%)
Total	212	442	156	810 (58%)
Male				
10–19	0	6	1	7 (1%)
20–29	2	11	5	18 (3%)
30–39	6	16	8	30 (5%)
40–49	7	44	8	59 (10%)
50–59	20	73	11	104 (18%)
60–69	38	88	25	151 (25%)
70–79	31	93	21	145 (24%)
80–89	17	46	7	70 (12%)
>90	1	8	2	11 (2%)
Total	122	385	88	595 (42%)
Total cohort	334 (24%)	827 (59%)	244 (17%)	1405 (100%)

POP_1 and POP_2 denote the observed frequency of a given allele, diplotype or haplotype in population 1 and population 2, respectively. F_{ST} values range from a lower bound of 0, where the variant frequency is the same in the two populations, and an upper value of 1 where the variant frequency is 0 in one population and 1 in the other. For the purposes of the present study, the 'populations' refer to self-identified African–Americans, Caucasians and Hispanics in the overall study cohort of patients from the US East Coast. The genotype, diplotype and haplotype frequencies in a given subpopulation reflect the relative proportion of individuals self-identified as part of such subpopulation among all participants.

The extent of genetic divergence between any two populations is denoted according to Wright's qualitative guidelines. A variant-specific pairwise F_{ST} value less than 0.050 indicates low genetic divergence, more than 0.050 but less than 0.150 corresponds to moderate divergence, more than 0.150 but less than 0.250 corresponds to high divergence, and more than 0.250 corresponds to very high divergence [21].

Ethical consent

Patients signed an informed consent form agreeing to DNA testing and to the use of their deidentified information of a statistical nature for validation, research and accreditation purposes.

Results

Demographics

Our cohort consisted of 1405 primary care patients with a distribution of ethnicities enriched for minorities (Table 1). African–Americans and Hispanics constituted, respectively, 24 and 17% of the sample. The age distribution was predominantly 60–79 years of age, which is consistent with the demographics of primary care. At 58% of the total sample, the female representation was higher than the male, particularly in the African–American and Hispanic cohorts, where females accounted for nearly 2/3 of the patients.

Table 2. Distribution of *MTHFR* genotypes for loci *677* and *1298* by ethnicity (n = 1405).

Locus	Genotype	African–American (n = 334)	Caucasian (n = 827)	Hispanic (n = 244)	Total (n = 1405)
677 C/T	CC	238	378	89	705 (50%)
	CT	87	369	104	560 (40%)
	TT	9	80	51	140 (10%)
	%C	84%	68%	58%	70%
	%T	16%	32%	42%	30%
1298 a/c	aa	234	329	144	707 (50%)
	ac	88	390	93	571 (41%)
	cc	12	108	7	127 (9%)
	%a	83%	68%	78%	73%
	%c	17%	32%	22%	27%

The percentage of polymorphisms observed in each ethnicity is included.

Table 3. Distribution of *MTHFR* diplotypes for loci *677* and *1298* by ethnicity (n = 1405).

Diplotype	African–American (n = 334)	Caucasian (n = 827)	Hispanic (n = 244)	Total (n = 1405)
CCaa	158	88	35	281 (20%)
CCac	68	163	47	278 (20%)
CCcc	12	78	7	97 (7%)
CTaa	68	183	58	309 (22%)
CTac	19	205	46	270 (19%)
CTcc	0	2	0	2 (0%)
TTaa	8	107	51	166 (12%)
TTac	1	1	0	2 (0%)
TTcc	0	0	0	0 (0%)

Genotypes

For the *677* locus, the frequency of the *T* allele was 30% overall, with the highest frequency in the Hispanic cohort (42%; Table 2). For the *1298* locus, the frequency of the *c* allele was 27% overall, with the highest frequency in the Caucasian cohort (32%). For both loci, the African–American cohort evidenced the least relative frequency of the respective minor allele (16 and 17%).

The cohort-normalized distribution of the genotypes reveals that the Hispanic cohort has the highest prevalence of *MTHFR* functionally deficient *TT* genotypes at the *677* locus (63%), and Caucasians, the highest prevalence of functionally deficient *cc* alleles at the *1298* locus (Figure 1A). The African–American cohort has the highest prevalence of normal genotypes at both loci, *677 CC* (46%) and *1298 aa* (42%).

Diplotypes

There are nine possible diplotypes for the *677*/*1298* allelic combinations. In the total sample (Table 3), there were four diplotypes at approximately 20% frequency each: *CCaa*, *CCac*, *CTaa*, *CTac*. Two diplotypes were found at approximately 10% each, *CCcc* and *TTaa*. It should be noted that, despite the large sample size, two diplotypes were rare, *CTcc* and *TTac*, and one diplotype was not observed at all, *TTcc*.

The cohort-normalized frequency of the doubly wild-type diplotype, *CCaa*, was 65% in the African–American patients. Caucasians evidenced the lowest prevalence of this wild-type diplotype at 15%. The singly heterozygous diplotypes with a single dysfunctional allele, *CCac* and *CTaa*, were in equilibrium and evenly distributed across the three subgroups with nearly equal normalized frequencies at 31–36% each (Figure 1B).

Haplotypes

Haplotypes are not actual observations but were modeled and tallied statistically by the Beagle phasing algorithm (Table 4). In the total cohort, the wild-type haplotype *Ca* was most prevalent at 41%, and the singly mutated

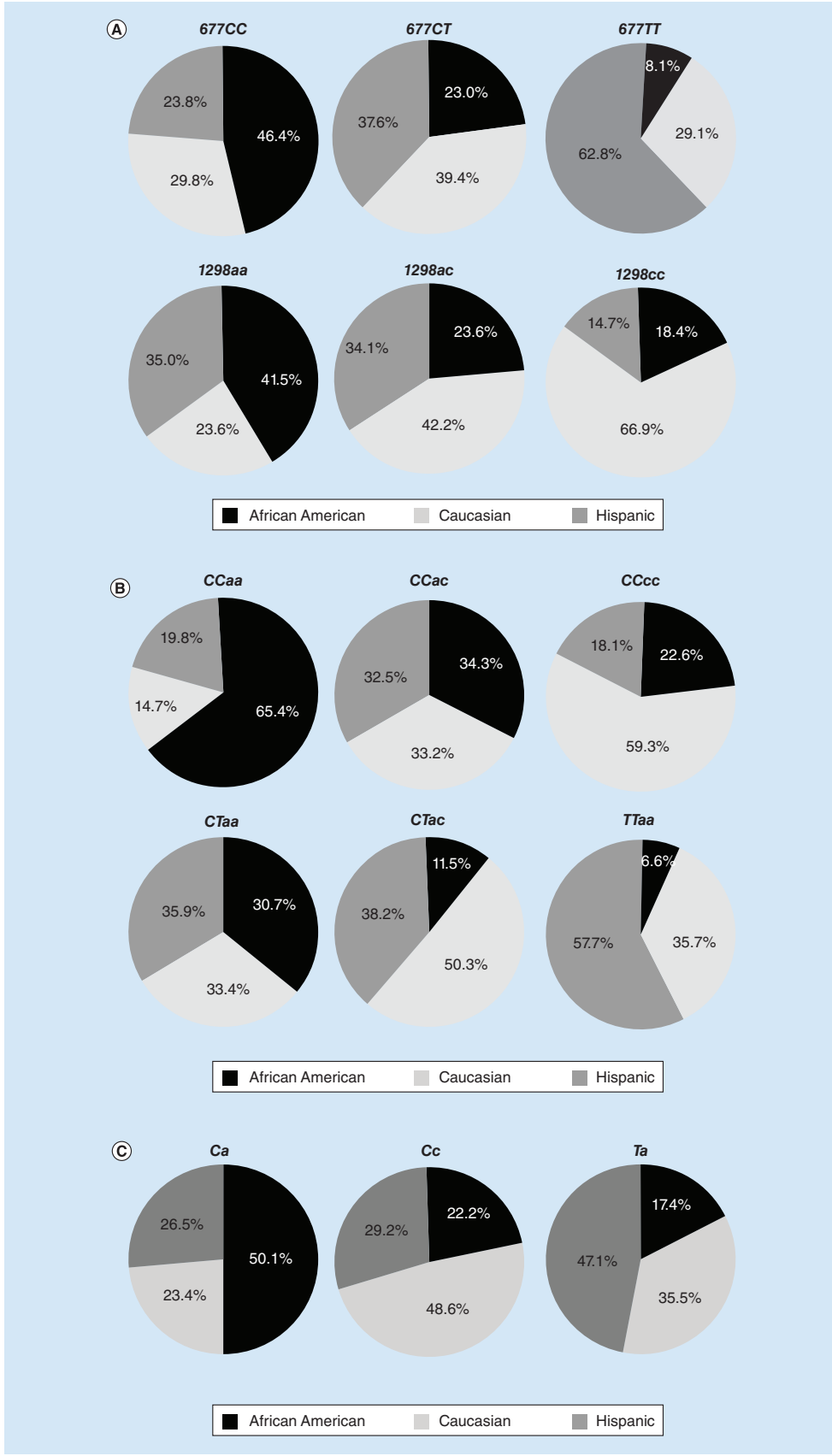

Figure 1. **(A)** Normalized frequency distribution of genotypes by ethnicity (n = 1405). **(B)** Normalized frequency distribution of diplotypes by ethnicity (n = 1409). Diplotypes *CTcc*, *TTac* and *TTcc* were excluded from analysis because they were rare or not observed. **(C)** Normalized frequency distribution of haplotypes by ethnicity (n = 1405), corresponding to 2810 haplotypes. Haplotype *Tc* was excluded from analysis because it was rare or not observed.

Table 4. Distribution of *MTHFR* haplotypes for loci *677* and *1298* by ethnicity (n = 1405), corresponding to 2810 haplotypes.

Haplotype	African–American (n = 668)	Caucasian (n = 1654)	Hispanic (n = 488)	Total (n = 2810)
Ca	452	522	175	1149 (40.9%)
Cc	111	603	107	821 (29.2%)
Ta	104	526	206	836 (29.8%)
Tc	1	3	0	4 (0.1%)

haplotypes *Cc* and *Ta* were equally represented at 29–30%. The doubly dysfunctional haplotype *Tc* was very rare at 0.1%.

The wild-type haplotype *Ca* was most common in African–Americans but not in Caucasians and Hispanics (Figure 1C). On a cohort normalized basis, the frequency of doubly wild-type *Ca* was highest in the African–American subcohort at 50%. For the singly mutated haplotypes, Caucasians had nearly half the *Cc* haplotype (49%) and Hispanics nearly half the *Ta* (47%).

Quantification of genetic divergence

Data for Wright's F_{ST} values examining the various polymorphisms and ethnic pairwise comparisons are shown in Supplementary Table 1. The *677C>T* and *1298A>C* loci evidenced low divergence across most pairwise ethnic comparisons of genotypes, except for moderate divergence between African–Americans and Hispanics for *677C>T* (F_{ST} = 0.096), and between African–Americans and Caucasians for *1298A>C* (0.050). The lowest divergence of *1298A>C* occurred between African–Americans and Hispanics (0.004).

Low divergence was observed for diplotypes, except for *CCaa* and *TTaa*. High divergence was observed for *CCaa* between Caucasians and African–Americans (F_{ST} = 0.163) and moderate divergence, between African–Americans and Hispanics (0.127). Moderate divergence was observed for *TTaa* between African–Americans and Hispanics (0.083) and between Caucasians and Hispanics (0.112). A divergence of nearly zero was observed for *CCac* and *CTaa* for all pairwise comparisons. Diplotypes *CTCC, TTac* and *TTcc* appeared too infrequently to perform F_{ST} calculations or were not observe.

Low to moderate divergence was observed for haplotypes. Moderate divergence was observed between Caucasians and African–Americans for *Ca* (F_{ST} = 0.130) and *Cc* (0.050). Moderate divergence was also observed between African–Americans and Hispanics for *Ca* (0.101) and *Ta* (0.086). Haplotype *Tc* was too infrequent to perform F_{ST} calculations.

Altogether, Hispanics are genetically closer to Caucasians than African–Americans at the evaluated *MTHFR* loci (*677C>T* and *1298A>C*) as measured by Wright's F_{ST} values of genotypes, diplotypes and haplotypes. Genetic divergence was found lowest when Hispanics were compared with Caucasians than when Caucasians were compared with African–Americans or African–Americans to Hispanics. Average F_{ST} values were lowest when Hispanics were compared with Caucasians (0.019 for genotypes, 0.024 for diplotypes and 0.013 for haplotypes).

Discussion

Ethnogeographic distributions

The present study surveys a primary care population in urban centers of the US with substantial minority representation. It is thus a practical representation of the groups in which disease prevention efforts should be enhanced. The objective of this study was to determine the ethnogeographic distribution of *MTHFR* alleles, diplotypes and haplotypes at the positions *677C>T* and *1298A>C*.

Previous studies have found that the prevalence of *MTHFR* polymorphisms, particularly the *677C>T* polymorphism, is strongly associated with ethnic and geographic distributions. The highest frequencies of *677C>T* polymorphisms are found in Mexican Amerindian populations (65%) [26] and in northern Han Chinese populations (45%) [27]. Lower *677C>T* frequencies are observed in non-Italian European populations (26.9–33.9% in Russian and Spanish populations, respectively) [28] and the lowest are observed in African populations (8.6% in Zambian populations) [29]. Interestingly, southern Italian populations display unusually high *677C>T* frequencies for European populations, up to 46% [28].

The present study found *677C>T* frequencies similar to those reported for the represented ethnic groups (Table 2), with the highest *677 T* allelic frequency observed in Hispanics (42%), the lowest in African–Americans

(16%) and with Caucasians in between (32%). The *677 T* frequency observed in Caucasians was the same previously reported (31.7% in Caucasians in Atlanta, GA) [28]. This genetic difference may be due to the greater prevalence of Hispanics in the Northeastern US or the greater prevalence of African–Americans in the Southern US.

The cohort normalized frequency of *677 TT* homozygotes was found to be much higher in Hispanics (63%) than in African–Americans (8%) or Caucasians (29%; Figure 1A). Similarly, the *677/1298* haplotype *Ta* was observed more frequently in Hispanics (47%) compared with African–Americans (17%) or Caucasians (36%; Figure 1C).

In contrast, the highest *1298 c* allele frequency was observed in Caucasians (32%), with lower frequencies in Hispanics (22%) and African–Americans (17%; Table 2). The cohort normalized frequency of *1298 cc* homozygotes was found to be much higher in Caucasians (67%) than in African–Americans (18%) or Hispanics (15%; Figure 1A). Similarly, the *677/1298* haplotype *Cc* was observed more frequently in Caucasians (49%) compared with African–Americans (22%) or Hispanics (29%; Figure 1C).

Overall, the genomic divergence between these populations was low to moderate. This analysis of the F_{ST} values highlights the problem faced by researchers and clinicians when assessing genomic data derived from multiethnic populations, with distinct ancestries and admixture patterns. Our findings confirm the notion that genomic variability is highly sensitive to within-population diversity and substructures. For example, the terms 'Hispanic' and 'Latino', may obscure underlying genetic differences between Hispanics ethnic groups; while Mexican Amerindian and Chilean populations display large *677C>T* frequencies (65 and 52%, respectively) [26,30], Brazilian Amerindian and Colombian populations display frequencies of only 24 and 13.7%, respectively [31]. Hispanics are an admixed population with ancestral contribution from Europeans, Native Americans and Africans; however, the relative contribution of each parental population may differ across ethnogeographic groups [32,33]. Consequently, relevant variants may distribute differently thus generating unique haplotypic blocks characteristic of each group [25,34].

Evolutionary implications

A previous bioinformatics analysis performed by Khan and Jamil showed that *MTHFR* genetic sequence is conserved across archaebacteria and eukaryotes [13]. They also found that the active site of MTHFR (specifically, the catalytic residues) as well as the function and structure are well conserved across species from the three different groups. However, homozygosity of *677C>T* has a relatively high frequency in some populations despite the important biological role of this enzyme in the cell. Meadows *et al.* proposed that the alteration of *677C>T* may result from natural selection to confer protection against malaria in Mediterranean populations where the disease has been historically endemic and where the homozygous genotype is relatively high (15–20% in southern Europe) [35]. This study observed that MTHFR-deficient mice survived longer than those overexpressing the enzyme (who died earlier than mice expressing the wild-type enzyme). Taken altogether, it seems that these regions in the *MTHFR* gene exhibit separately some degree of evolutionary divergence, but together they have strong evolutionary constraint given the important biological function of this enzyme in gene regulation and in amino acid and DNA synthesis.

Three diplotypes were rare or not observed (*CTcc, TTcc, TTac*; Table 3 & Figure 1C), suggesting that those diplotypes are strongly selected against. The low prevalence of the *TTcc* diplotype could be interpreted as a result of very low fitness of this haplotype, or of the effect of chance and environment in the form of genetic drift events, or a combination of both forces. A recent study performed by Stangler Herodež *et al.* found that these haplotypes occur more frequently in spontaneously aborted fetal tissue in Slovenian populations, corroborating the hypothesis that *CTcc*, *TTcc* and *TTac* diplotypes preclude normal development [19]. Similarly, Karimian and Colagar found that the *677C>T* polymorphism may increase the risk of infertility in men [36]. While beyond the scope of this study, the striking absence of these three diplotypes in healthy adults but elevated prevalence in spontaneously aborted fetal tissue suggests that *MTHFR* may play a role in fetal development and pregnancy.

Clinical implications

Neuropathy

According to early reports, there seems to be a genetic component that may play a role in the development of diabetic peripheral neuropathy (DPN) [37,38]. Jimenez *et al.* found an association between the *MTHFR 1298A>C* variant and susceptibility to DPN in Puerto Ricans with Type 2 diabetes mellitus treated with metformin [39]. Results suggest that *1298A>C* but not *677C>T* was associated with DPN susceptibility in this cohort. The authors suggested that a differential haplotype structure and LD pattern in the study cohort might explain the observed association of *1298A>C*, but not of *677C>T*, with DPN susceptibility in these patients [39]. It has also

been found by others that individuals with *MTHFR 1298A>C*, but not *677C>T*, have been associated with metabolic syndrome [40]. However, after sorting individuals in the Jimenez *et al.* study by their carrier status, no significant associations were observed between these genetic variants (independently or combined) and any of the biochemical markers (HbA1c, folate, vitamin B12, homocysteine) [40]. Notably, 67% of participants in this study carried at least one of these *MTHFR* polymorphisms.

Furthermore, Wu *et al.* suggested that the *677C>T* variant could be a risk factor for DPN in Turkish but not in Japanese and Pakistani individuals [41]. A significant association between *677C>T* and DPN was indeed found in their meta-analysis. In addition, Yigit *et al.* found a significant association between *677C>T* and DPN in Turkish with Type 2 diabetes mellitus [9].

In this study, two diplotypes were rarely observed, *CTcc* and *TTac*, and one diplotype was not observed at all, *TTcc*. From biochemical and structural studies, it is apparent that the protein encoded by *677C>T* (222 Val) has lesser affinity for flavin adenine dinucleotide (FAD) cofactor [11] and the protein encoded by *1298A>C* (429 Ala) has an altered SAM regulatory domain [15]. The presence of both alterations in the *Tc* haplotype is therefore highly dysfunctional and likely selected against evolutionarily. When it can be inferred from the diplotype that the patient is a carrier of the *Tc* haplotype, the clinical guidance may warn of greater risk with greater accuracy than can be garnered from a single variant at loci *677* or *1298* alone.

Thrombophilia

In a previous report of a retrospective review study by Morales-Borges, the frequency of *MTHFR* polymorphisms among Puerto Ricans patients in an ambulatory practice of community-based mixed hematology and medical oncology was 40% [42]. This was in good concordance with published data that states the *MTHFR* variants among Hispanics may be higher than in other reported groups [43]. The prevalence for the *677C>T* polymorphism was 57%, though it was more frequent among women who also presented with venous thrombosis, such as recurrent miscarriage syndrome and deep venous thrombosis. This cross-sectional study of medical records review was aimed at determining the prevalence of *MTHFR* polymorphisms in this Puerto Rican cohort and their relationship with thrombophilia. Strikingly, 70% of cases who carried a polymorphism on the *MTHFR* locus also had another genetic disorder such as factor V leiden mutation, lupus anticoagulant and deficiencies of antithrombin III and/or protein C and S [42].

Clinical utility & recommendations

There is conflicting guidance in the literature and among clinical associations regarding interpretation of the compound heterozygote *677C>T* and *1298A>C*. According to the 2013 practice guidelines established by the American College of Medical genetics, the heterozygous variant 677 CT, the homozygous variant 1298 cc and the compound 677/1298 heterozygotes *CTac* are all unlikely to be independent risk factors for venous thrombosis or hyperhomocysteinemia [44]. The American Association for Clinical Chemistry risk guidance for the *MTHFR* compound heterozygote is contrary to that offered by American College of Medical genetics, which does not indicate increased risk of hyperhomocysteinemia. The American Association for Clinical Chemistry, in its database, Lab Tests Online, states: 'If a person has two copies (homozygous) of MTHFR 677 TT, or has one copy of 677 T and one of 1298 c, then it is likely that elevated homocysteine levels are due to these inherited mutations, or that the mutations are contributing to them' [45].

Limitations

In the present study, 'Hispanic' refers to all individuals who self-identified as Hispanic or Latino regardless of their familial lineage or familial country of origin. The majority of Hispanics (76%) originated from the Northeast of the USA, suggesting that the majority of Hispanics tested were of Caribbean descent (Puerto Rico, Dominican Republic and Cuba). However, the generality of the terms, 'Hispanic' and 'Latino', may obscure underlying genetic differences between various ethnic groups as evidenced by the discrepancies in *677C>T* frequency between various Hispanic/Latino groups. Further research is required to better elucidate the distribution of *MTHFR* polymorphisms in Hispanic and Latino populations thromboembolism [46,47].

Similarly, there is a need for more research on the admixture of African–American populations, as our sample is regionally limited. It is evident, however, that the *MTHFR* gene is most functional in this cohort, which could represent an ancestral evolutionary or migratory selection in folate metabolism.

Due to limitations on the clinical data made available when a patient or clinician requested genetic testing, as well as restrictions imposed by the ethical consent each patient provided, the present study cannot provide insight into the diseases, symptoms or at-risk statuses associated with each patient which may have prompted genetic testing of *MTHFR*. While the focus of this study is the prevalence of specific *MTHFR* polymorphisms associated with various ethnicities, future studies may wish to incorporate disease data for patients to better elucidate any possible associations between ethnicity, *MTHFR* polymorphisms and disease prevalence.

Currently, there remains lack of consensus about the interpretation and clinical implications of *MTHFR* genotypes at loci *677* and *1298* and *677/1298* allele combinations. Recent publications have disproven an association between hyperhomocysteinemia and risk for coronary heart disease and between *MTHFR* polymorphism status and risk for venous thromboembolism [48]. The paradoxical prevalence of studies associating MTHFR levels with disease states and clinical associations recommending against the routine use of *MTHFR* genotyping suggests that the association between *MTHFR* and disease is complex and remains poorly understood. Further research is required to establish the clinical utility of *MTHFR* genotyping and to elucidate its benefit to specific ethnic groups.

Conclusion

This study examined the prevalence of *MTHFR 677C>T* and *1298A>C* polymorphisms in 1405 patients referred from primary care centers in the USA. Considerable differences in prevalence were observed among the three most common self-reported ethnicities (African–Americans, Caucasians and Hispanics) for *MTHFR* genotypes, diplotypes and haplotypes, suggesting that MTHFR sufficiency and deficiency is strongly associated with ethnicity. However, the genetic divergence quantified by F_{ST} value was low to moderate. Our findings also suggested that *MTHFR* genetic arrangement in Hispanics is more similar to Caucasians than to African–Americans. The diplotype frequencies suggested a negative selection of multiply deficient alleles. The diplotypes with three dysfunctional alleles (homozygous mutant for one allele and heterozygous for the other), *TTac* and *CTcc*, were rare. The doubly homozygous mutant diplotype with four dysfunctional alleles, *TTcc*, was not observed at all. African–Americans evidenced the highest prevalence of functional alleles. The results support the hypothesis that *MTHFR* single-locus genotyping may be insufficient to understand risk of MTHFR deficiency and that future analyses should focus on diplotypes and haplotypes. These results were based on ethnically diverse populations and may be of clinical utility to healthcare centers that serve these populations.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/bmm-2018-0392

Author contributions

JS Graydon and K Claudio contributed equally as first authors, providing data analysis and critical interpretation, performing literature reviews, and writing the article. S Baker and M Ferreira assisted with data acquisition and analysis. M Kocherla contributed technical insight into the methods used. A Roche-Lima and J Rodríguez-Maldonado contributed statistical analysis and interpretation. J Duconge provided insight into the literature and contributed to the clinical interpretation of the data and the writing of the article. G Ruaño provided critical insight in the subject from a biological as well as clinical perspective and contributed to the writing of the article.

Financial & competing interests disclosure

G Ruaño is founder and President of Genomas, Inc. JS Graydon, M Kocherla and M Ferreira were employees of Genomas, Inc. This research has been supported by Genomas internal research funds and in part by NIH (grant number SC1 HL123911), from the National Heart, Lung and Blood Institute (NHLBI), as part of the SCORE MBRS Program supported by the National Institute of General Medical Sciences (NIGMS), and U54 MD007600 from the National Institute on Minority Health and Health Disparities (NIMHD), as part of the Research centers in Minority Institutions (RCMI) Program. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Summary points

- *MTHFR* gene have been associated with a number of diseases, including coronary heart disease, psychological disorders and some cancers; however, there is currently no consensus on the actual risk conveyed by specific *MTHFR* mutations.
- While two polymorphisms for deficient variants, *MTHFR 677C>T* and *1298A>C*, are commonly tested, the literature is lacking in detailed comparisons of the combined frequencies of the two polymorphisms, potentially contributing to the lack of consensus between *MTHFR* mutations and disease risk.
- The present study analyzed results from 1405 patients genotyped for *MTHFR 677C>T* and *1298A>C* mutations referred from primary care centers in the USA.

Results

- The results indicated that there exist ethnogeographic differences in *677C>T* and *1298A>C* polymorphisms, with African–Americans carrying the lowest prevalence of *677C>T* (16%) and *1298A>C* (17%), Caucasians intermediate prevalence of *677C>T* (32%) and highest prevalence of *1298A>C* (32%), and Hispanics the highest prevalence of *677C>T* (42%) and intermediate prevalence of *1298A>C* (22%).
- Analysis with Wright's Fixation Index (F_{ST}) revealed low to moderate genetic divergence overall, with Caucasians and Hispanics displaying the least divergence of any pairwise comparison.

Discussion

- Importantly, diplotype analysis revealed striking insights into ethnogeographic prevalence of *MTHFR* mutations, such as the elevated prevalence of doubly homozygous wild-type alleles (*677CC* and *1298aa*, i.e. *CCaa*) in African–Americans (65.4% of *CCaa* population) which may convey protection from MTHFR deficiency.
- The diplotypes *CTcc* and *TTac* were exceedingly rare and the diplotype *TTcc* was absent from the study, suggesting negative selection against multiple dysfunctional polymorphisms.

Conclusion

- Overall, these results indicate a strong correlation between ethnicity and *MTHFR* diplotypes, suggesting that single-locus genotyping may obscure disease associations, especially in genetically homogeneous populations, and that future studies should focus on multiple loci.

Ethical conduct of research

The authors state that appropriate consent was obtained prior to the testing of patient samples and that all subsequent analysis was permitted by the consent provided by each patient or patient guardian agreeing to DNA testing and to the use of their deidentified information of a statistical nature for validation, research and accreditation purposes.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Rozen R. Molecular genetics of methylenetetrahydrofolate reductase deficiency. *J. Inherit. Metab. Dis.* 19(5), 589–594 (1996).
2. Engbersen AM, Franken DG, Boers GH, Stevens EM, Trijbels FJ, Blom HJ. Thermolabile 5,10-methylenetetrahydrofolate reductase as a cause of mild hyperhomocysteinemia. *Am. J. Hum. Genet.* 56(1), 142–150 (1995).
3. Klerk M, Verhoef P, Clarke R *et al.* MTHFR 677C→T polymorphism and risk of coronary heart disease. *J. Am. Med. Assoc.* 288(16), 2023–2031 (2002).
4. Yan L, Zhao L, Long Y *et al.* Association of the maternal *MTHFR* C677T polymorphism with susceptibility to neural tube defects in offsprings: evidence from 25 case-control studies. *PLoS ONE* 7(10), e41689 (2012).
5. Zhang T, Lou J, Zhong R *et al.* Genetic variants in the folate pathway and the risk of neural tube defects: a meta-analysis of the published literature. *PLoS ONE* 8(4), e59570 (2013).
6. Porter K, Hughes CF, Hoey L *et al.* Investigation of the role of riboflavin, vitamin B6 and *MTHFR* genotype as determinants of cognitive health in ageing. *Proc. Nutr. Soc.* 75(OCE3), E114 (2016).
7. Gilbody S, Lewis S, Lightfoot T. Methylenetetrahydrofolate reductase (*MTHFR*) genetic polymorphisms and psychiatric disorders: a HuGE review. *Am. J. Epidemiol.* 165(1), 1–13 (2007).
8. Xie SZ, Liu ZZ, Yu JH *et al.* Association between the *MTHFR* C677T polymorphism and risk of cancer: evidence from 446 case–control studies. *Tumor Biol.* 36(11), 8953–8972 (2015).
9. Yigit S, Karakus N, Inanir A. Association of *MTHFR* gene C677T mutation with diabetic peripheral neuropathy and diabetic retinopathy. *Mol. Vis.* 19 1626–1630 (2013).
10. Fisher MC, Cronstein BN. Metaanalysis of methylenetetrahydrofolate reductase (*MTHFR*) polymorphisms affecting methotrexate toxicity. *J. Rheumatol.* 36(3), 539–545 (2009).

11. Frosst P, Blom HJ, Milos R *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat. Genet.* 10(1), 111–113 (1995).

• **Classical biochemical study describing the *MTHFR* 677C>T (Ala222Val) polymorphism and its functional implications.**

12. Goyette P, Pai A, Milos R *et al.* Gene structure of human and mouse methylenetetrahydrofolate reductase (*MTHFR*). *Mamm. Genome.* 9(8), 652–656 (1998).

13. Khan M, Jamil K. Study on the conserved and the polymorphic sites of *MTHFR* using bioinformatices approaches. *Trends Bioinforma.* 1(1), 7–17 (2008).

14. Weisberg IS, Jacques PF, Selhub J *et al.* The 1298A→C polymorphism in methylenetetrahydrofolate reductase (*MTHFR*): *in vitro* expression and association with homocysteine. *Atherosclerosis* 156(2), 409–415 (2001).

15. Weisberg I, Tran P, Christensen B, Sibani S, Rozen R. A second genetic polymorphism in methylenetetrahydrofolate reductase (*MTHFR*) associated with decreased enzyme activity. *Mol. Genet. Metab.* 64(3), 169–72 (1998).

• **Classical biochemical study describing the *MTHFR* 1298A>C (Glu429Ala) polymorphism and its functional implications.**

16. Jencks DA, Mathews RG. Allosteric inhibition of methylenetetrahydrofolate reductase by adenosylmethionine. Effects of adenosylmethionine and NADPH on the equilibrium between active and inactive forms of the enzyme and on the kinetics of approach to equilibrium. *J. Biol. Chem.* 262(6), 2485–2493 (1987).

17. 1000 Genomes Project Consortium, Auton A, Brooks LD *et al.* A global reference for human genetic variation. *Nature* 526(7571), 68–74 (2015).

18. Kałuzna EM, Strauss E, Światek-Kościelna B *et al.* The methylenetetrahydrofolate reductase 677T-1298C haplotype is a risk factor for acute lymphoblastic leukemia in children. *Medicine* 96(51), 18–19 (2017).

19. Stangler Herodež Š, Zagradišnik B, Erjavec Škerget A *et al. MTHFR* C677T and A1298C genotypes and haplotypes in Slovenian couples with unexplained infertility problems and in embryonic tissues from spontaneous abortions. *Balk. J. Med. Genet.* 16(1), 31–40 (2013).

•• **An important look into the role of *MTHFR* genotypes in infertility using a multiloci view (*MTHFR* 677 and 1298).**

20. Fan S, Yang B, Zhi X, Wang Y, Zheng Q, Sun G. Combined genotype and haplotype distributions of *MTHFR C677T* and *A1298C* polymorphisms: a cross-sectional descriptive study of 13,473 Chinese adult women. *Medicine* 95(48), e5355 (2016).

21. Diagnostics OM. 510(k) Substantial equivalence determination decision summary. 1–11 (2010). www.accessdata.fda.gov/cdrh_docs/reviews/K093974.pdf

22. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81(5), 1084–1097 (2007).

23. Fesinmeyer MD, North KE, Ritchie MD *et al.* Genetic risk factors for body mass index and obesity in an ethnically diverse population: results from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *Obesity* 21(4), 835–846 (2013).

24. Wright S. Evolution and the genetics of populations, volume 4. University of Chicago Press, Chicago, IL, USA (1984).

25. Suarez-Kurtz G, Pena SDJ, Hutz MH. Application of the $F_{(ST)}$ statistics to explore pharmacogenomic diversity in the Brazilian population. *Pharmacogenomics* 13(7), 771–7 (2012).

• **A relevant example that show how the F_{ST} statistic may be applied to an extremely heterogeneous population to better understand its internal composition and relation to external populations.**

26. Contreras-Cubas C, Sanchez-Hernandez BE, Garcia-Ortiz H *et al.* Heterogenous distribution of *MTHFR* gene variants among mestizos and diverse amerindian groups from Mexico. *PLoS ONE* 11(9), 26–27 e0163248 (2016).

27. Wang X, Fu J, Li Q, Zeng D. Geographical and ethnic distributions of the *MTHFR C677T*, *A1298C* and *MTRR A66G* gene polymorphisms in Chinese populations: a meta-analysis. *PLoS ONE* 11(4), e0152414 (2016).

28. Wilcken B. Geographical and ethnic variation of the *677C>T* allele of 5,10 methylenetetrahydrofolate reductase (*MTHFR*): findings from over 7000 newborns from 16 areas world wide. *J. Med. Genet.* 40(8), 619–625 (2003).

•• **A large, multinational study that helps demonstrate the strong association between *MTHFR 677C>T* and various geographic ethnicities.**

29. Atadzhanov M, Mwaba MH, Mukomena PN *et al.* Frequency of *APOE*, *MTHFR* and *ACE* polymorphisms in the Zambian population. *BMC Res. Notes* 7(1), 194 (2014).

30. Roco Á, Quiñones L, Agúndez JAG *et al.* Frequencies of 23 functionally significant variant alleles related with metabolism of antineoplastic drugs in the Chilean population: comparison with Caucasian and Asian populations. *Front. Genet.* 3(NOV), 1–9 (2012).

31. Romero-Sánchez C, Gómez-Gutierrez A, Gómez PE, Casas-Gomez MC, Briceño I. *C677T* (RS1801133) *MTHFR* gene polymorphism frequency in a Colombian population. *Colomb. Med.* 46(2), 75–79 (2015).

32. Suarez-Kurtz G. Pharmacogenomics in admixed populations. *Trends Pharmacol. Sci.* 26(4), 196–201 (2005).

33. Via M, Gignoux CR, Roth LA *et al.* History shaped the geographic distribution of genomic admixture on the island of Puerto Rico. *PLoS ONE* 6(1), e16513 (2011).

34. Bryc K, Velez C, Karafet T *et al.* Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci. USA* 107(Suppl. 2), 8954–8961 (2010).

35. Meadows DN, Pyzik M, Wu Q *et al.* Increased resistance to malaria in mice with methylenetetrahydrofolate reductase (MTHFR) deficiency suggests a mechanism for selection of the *MTHFR 677C>T* (c.665C>T) variant. *Hum. Mutat.* 35(5), 594–600 (2014).
36. Karimian M, Colagar AH. Association of *C677T* transition of the human methylenetetrahydrofolate reductase (*MTHFR*) gene with male infertility. *Reprod. Fertil. Dev.* 28(6), 785 (2016).
37. Ukinc K, Ersoz HO, Karahan C *et al.* Methyltetrahydrofolate reductase *C677T* gene mutation and hyperhomocysteinemia as a novel risk factor for diabetic nephropathy. *Endocrine* 36(2), 255–261 (2009).
38. Schwahn B, Rozen R. Polymorphisms in the methylenetetrahydrofolate reductase gene: clinical consequences. *Am. J. Pharmacogenomics.* 1(3), 189–201 (2001).
39. Jiménez-Ramírez FJ, Castro LM, Ortiz C *et al.* Role of treatment-modifying *MTHFR 677C>*T and *1298A>C* polymorphisms in metformin-treated Puerto Rican patients with Type 2 diabetes mellitus and peripheral neuropathy. *Drug Metab. Pers. Ther.* 32(1), 23–32 (2017).
40. van Winkel R, Rutten BP, Peerbooms O, Peuskens J, van Os J, De Hert M. *MTHFR* and risk of metabolic syndrome in patients with schizophrenia. *Schizophr. Res.* 121(1–3), 193–198 (2010).
41. Wu S, Han Y, Hu Q *et al.* Effects of common polymorphisms in the *MTHFR* and *ACE* genes on diabetic peripheral neuropathy progression: a meta-analysis. *Mol. Neurobiol.* 54(4), 2435–2444 (2017).
42. Morales-Borges RH. Prevalence of *MTHFR C677T* and *A1298C* mutations and thrombophilia in Puerto Rico. *J. Blood Disord. Transfus.* 05(05), 213 (2014).
43. Chowdary D, Streck D, Schwalb MN. High incidence of two methylenetetrahydrofolate reductase mutations (*C677T* and *A1298C*) in Hispanics. *Genet. Test.* 7(3), 255–257 (2003).
44. Hickey SE, Curry CJ, Toriello HV. ACMG Practice Guideline: lack of evidence for *MTHFR* polymorphism testing. *Genet. Med.* 15(2), 153–156 (2013).

•• **Extensive review of clinical correlations and controversies on the diagnostic guidance afforded by methylenetetrahydrofolate reductase polymorphisms.**

45. American Association for Clinical Chemistry. *MTHFR* mutation: the test. (2016). https://labtestsonline.org/understanding/analytes/mthfr/tab/test/
46. Franco RF, Araújo AG, Guerreiro JF, Elion J, Zago MA. Analysis of the 677 C→T mutation of the methylenetetrahydrofolate reductase gene in different ethnic groups. *Thromb. Haemost.* 79(1), 119–121 (1998).
47. Chen J, Teo YY, Toh DSL, Sung C. Interethnic comparisons of important pharmacology genes using SNP databases: potential application to drug regulatory assessments. *Pharmacogenomics* 11(8), 1077–94 (2010).
48. Tsai AW, Cushman M, Tsai MY *et al.* Serum homocysteine, thermolabile variant of methylene tetrahydrofolate reductase (*MTHFR*), and venous thromboembolism: Longitudinal Investigation of Thromboembolism Etiology (LITE). *Am. J. Hematol.* 72(3), 192–200 (2003).