High-throughput multiplex single-nucleotide polymorphism (SNP) analysis in genes involved in methionine metabolism

Biochem Genet
DOI 10.1007/s10528-008-9159-5

High-Throughput Multiplex Single-Nucleotide Polymorphism (SNP) Analysis in Genes Involved in Methionine Metabolism

Betti Giusti · Ilaria Sestini · Claudia Saracini · Elena Sticchi · Paola Bolli · Alberto Magi · Anna Maria Gori · Rossella Marcucci · Gian Franco Gensini · Rosanna Abbate

Received: 20 June 2007 / Accepted: 24 January 2008

Abstract Hyperhomocysteinemia is a well-known independent marker factor for atherothrombotic diseases and may result from acquired and genetic influences. Several polymorphisms are suspected to be associated with hyperhomocysteinemia, but data are limited and inconsistent. High-throughput genotyping technologies, such as GenomeLab SNPStream, are now available. Moreover, an appropriate selection of SNPs to be analyzed could represent a strong resource to define the role of genetic risk factors. We developed a multiplex PCR-oligonucleotide extension approach by GenomeLab platform. We selected 72 SNPs based on their putative function and frequency in the candidate genes AHCY, BHMT, BHMT2, CBS, ENOSF1, FOLH1, MTHFD1, MTHFR, MTR, MTRR, NNMT, PON1, PON2, SLC19A1, SHMT1, TCN2, and TYMS. We were able to analyze 57 of the SNPs (79%). For MTHFR C677T and A1298C and MTR A2756G SNPs, we compared data obtained with an electronic microchip technology and found 99.2% concordance. We also performed a haplotype analysis. This approach could represent a useful tool to investigate the genotype–phenotype correlation and the association of these genes with hyperhomocysteinemia and correlated diseases.

Keywords Hyperhomocysteinemia · High-throughput technologies · Genotyping · Primer extension · Single-nucleotide polymorphism (SNP) · Homocysteine

Electronic supplementary material The online version of this article (doi:10.1007/s10528-008-9159-5) contains supplementary material, which is available to authorized users.

B. Giusti (✉) · I. Sestini · C. Saracini · A. Magi · A. M. Gori · R. Marcucci · R. Abbate
Department of Medical and Surgical Critical Care, Center of Research, Transfer, and High Education (DENOTHE), University of Florence, Viale Morgagni 85, 50134 Florence, Italy
e-mail: betti.giusti@unifi.it

E. Sticchi · P. Bolli · G. F. Gensini
Centro S. Maria agli Ulivi, Fondazione Don Carlo Gnocchi Onlus IRCCS, Impruneta, Florence, Italy

Introduction

Hyperhomocysteinemia (hyperHcy) has been implicated as an independent marker factor for cardiovascular diseases (McCully 1969; Eikelboom et al. 1999) and is associated with various other diseases and/or clinical conditions including Alzheimer's disease (Clarke et al. 1998), neural tube defects (Blom et al. 2006), schizophrenia (Applebaum et al. 2004), end-stage renal disease (Van Guldener and Stehouwer 2003), osteoporosis (Villadsen et al. 2005) and non-insulin-dependent diabetes (De Luis et al. 2005; Rudy et al. 2005). Moreover, perturbations in methionine metabolism and methylation reactions may be involved in cancer processes, such as colorectal cancer (Singh et al. 2003; Kim 2005; Sharp and Little 2004). Homocysteine (Hcy) is a highly reactive, sulfur-containing amino acid formed during methionine metabolism in the cell. It is a key branch-point intermediate in the ubiquitous methionine cycle, the function of which is to generate one-carbon methyl groups for transmethylation reactions that are essential for several biological processes. It is estimated that 5 to 7% of the general population have mild to moderate hyperHcy. Mild hyperHcy may result from both acquired (e.g., vitamin deficiencies) and genetic influences (Fowler 2005; Gellekink et al. 2005). Although hyperHcy has been associated with several diseases, the mechanism of Hcy-induced deleterious effects is not fully elucidated. Among the various mechanisms proposed to explain the harmful effects of Hcy, one is that it modulates the expression of certain genes that may lead to several pathological conditions (Castro et al. 2005).

Hcy-induced regulation of the expression of genes through the interference of DNA epigenetic phenotype modification (alteration of DNA methylation) (Castro et al. 2006) or by hitherto unknown mechanisms is predicted to lead to pathological conditions either directly or indirectly. Data from the literature provide evidence of 135 genes that either modulate the levels of Hcy (23 genes) or are modulated (112 genes) by elevated levels of Hcy (Sharma et al. 2006).

Several genetic polymorphisms in genes coding for enzymes involved in Hcy metabolism are demonstrated or suspected to be associated with hyperHcy. Some of these polymorphisms have a large number of data in the literature, even if not always consistent [e.g., C677T and A1298C in 5,10-methylenetetrahydrofolate reductase (MTHFR), A2756G in methionine synthase (MTR), G80A in solute carrier family 19 (folate transporter), member 1 (SLC19A1), C776G (TCN2)]. In some cases (e.g., C677T and A1298C MTHFR polymorphisms), the rare allele is associated with reduced enzyme activity in vitro (Frosst et al. 1995; Weisberg et al. 1998). At present, for a large number of genes involved in methionine metabolism, comprising those most studied, a deeper evaluation of SNPs might be necessary to determine whether they represent a genetic risk for Hcy metabolism perturbation and correlated diseases.

Besides the putative role in influencing homocysteine and homocysteine-thiolactone levels (Perla-Kajan et al. 2007), polymorphisms in enzymes involved in methionine metabolism may also play a role in altering DNA methylation reactions and thus genetic stability and gene expression (Waterland 2006). Therefore, the pivotal role of methionine metabolism in determining levels of highly reactive homocysteine and in altering genetic stability and gene expression leads us to

assume that a study of several polymorphisms in different genes involved in this pathway might be significant.

In the present study, the SNP selection was to permit construction of informative haplotypes in different selected candidate genes, independent of their known role in influencing Hcy levels, to use them in future susceptibility and association studies.

Thus, we evaluated a high-throughput analytical strategy using the GenomeLab SNPStream technology (Beckman Coulter, Fullerton, CA), based on multiplex primer extension reactions. It can be applied in the evaluation of genotype-phenotype correlation and association studies aimed at investigating the role of several genetic polymorphisms in different genes involved in methionine metabolism.

Materials and Methods

Selection of Polymorphisms

We selected SNPs in candidate genes according to their demonstrated or putative function on the basis of literature data, localization in the promoter or regulatory region or exons (SNP causing amino acid substitution), and/or with heterozygosity values greater than 0.300, extracted from the dbSNP database (NCBI, National Center for Biotechnology Information, Bethesda, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp&cmd=search&term=). The annotation of each SNP and its frequency in Caucasian populations were assessed in dbSNP NCBI and ENSEMBL (http://www.ensembl.org/index.html) databases.

Blood Collection and DNA Extraction

Peripheral blood samples were collected from 376 consecutive, Italian, apparently healthy subjects (24–98 years old, median age: 72; male/female: 207/169). To enroll healthy subjects, we excluded those with a history of cardiovascular disease or venous thromboembolic events, evaluated by a questionnaire structured to identify symptom-free subjects and to exclude those who were suspected of having any form of vascular disease. No subject had abnormal liver or renal function.

Genomic DNA was isolated from peripheral whole blood cells using the Flexigene DNA kit (Qiagen, Germany).

Genotyping of 72 Polymorphisms by GenomeLab SNPStream DNA Microarray Technology

All the SNPs were analyzed with GenomeLab SNPStream DNA microarray technology.

Primer Design

Genomic DNA sequences encompassing the polymorphic sites were subjected to a Blast search (Basic Local Alignment Sequence Tool; http://www.ncbi.nlm.nih.gov/BLAST/) in order to demonstrate highly homologous and repetitive elements.

Therefore, a specific software (http://www.autoprimer.com, Beckman Coulter, Fullerton, CA) was used for the design of the 12 pairs of primers for the multiplex PCR and of the 12 tagged extension primers for each panel. PCR primers were designed to amplify a short sequence (90–175 bp) surrounding the polymorphic site for each SNP of interest, while the single-strand tagged extension primers, complementary to the 3′ sequence flanking the SNP site, permitted the discrimination of the various alleles (Supplementary Table A; data available from the corresponding author on request).

Automation System

The liquid handling in all the reaction protocols is supported by Biomek FX Workstation (Beckman Coulter, Fullerton, CA).

Multiplex PCR and Clean-up Reactions

Multiplex PCR was performed in an MJ thermocycler (MJ Research, Waltham, MA) according to the recommended PCR condition: one denaturation cycle at 94°C for 1 min, 40 cycles with denaturation at 94°C for 1 min, annealing at 55°C for 30 s, and extension at 72°C for 1 min, followed by a final extension at 72°C for 1 min. The reaction was performed in a final volume of 5 µl with 4 ng genomic DNA, 75 µM dNTP mix, 50 nM primer pool, 5 mM $MgCl_2$, and 0.5 U *Taq* polymerase (GoTaq, Promega, Milan, Italy) in PCR 1 × buffer. PCR products were purified by Exo-Sap reaction (0.67 U exonuclease I and 0.33 U shrimp alkaline phosphatase in 3 µl mix volume), performed in an MJ thermocycler to degrade unincorporated PCR primers and dNTPs (USB Corp., Cleveland), following the recommended conditions: first step at 37°C for 30 min and final incubation at 96°C for 10 min.

Primer Extension Reaction

Extension protocol was performed using an appropriate mix containing TAMRA and Bodipy-fluorescein labeled dideoxynucleotides (ddNTPs) and a pool of 12 allele-specific tagged extension primers. The tagged extension primers can be directed toward a specific location in each well (containing a 4 × 4 oligonucleotide array) of the 384 SNPware plate, because of the presence of a tag at the 5′-terminus of the extension primer, complementary to the arrayed oligonucleotides. The 3′-terminus of the tagged extension primers is complementary to the sequence strictly adjacent to the SNP site (Denomme and Van Oene 2005) (Fig. 1). The extension reaction permits the incorporation of dye-labeled dideoxynucleotides at the 3′-terminus of the tagged extension primer. In each 4 × 4 array, three positive controls and one negative control were included to ensure accuracy in the genotype determination (Fig. 1). The extension reaction was performed in an MJ thermocycler according to the recommended extension conditions: a cycle at 96°C for 3 min, 44 cycles with a first step at 94°C for 20 sec and a second step at 40°C for 11 sec,

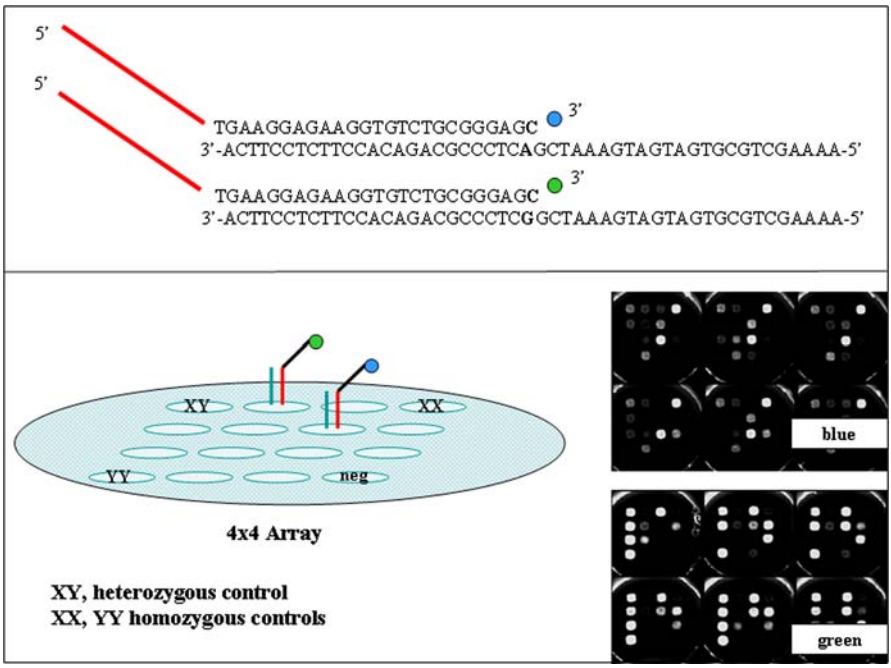

Fig. 1 (**a**) Extension primers complementary to the sequence adjacent to the SNP site, characterized by the presence of a tag at the 5′ end. The tagged extension primers can be recognized by oligonucleotides located in specific positions in the 4 × 4 array. (**b**) The 4 × 4 array has 4 control locations and 12 probe locations for 12 SNPs. Scanning images are shown on the right (blue/green laser detection)

and final hold at 4°C. The reaction was performed in a final volume of 15 µl with 0.04 µl SNP extension mix, 5.6 µl SNPware Extension Dilution Buffer, 0.3 µl SNP 20 × extension mix, and 0.03 µl SNPware DNA polymerase (Beckman Coulter, Fullerton, CA). The SNP 20 × extension mix contains the dye-labeled dideoxynucleotides, used in the primer extension reaction, and self-annealing oligonucleotides that are extended with dye-labeled terminators; these probes serve as positive controls in each well.

Hybridization Reaction

The primer extension products were transferred to the 384 SNPware plate, previously washed three times with 1 × SNPware Wash Buffer I, so that each extended primer would hybridize in a specific way to one of the unique probes arrayed in each well. A 12 µl volume of hybridization mix (Beckman Coulter, Fullerton, CA) was added to each well of the plate containing the primer extension products; thus, 8 µl of the previous mixture was transferred to the corresponding well of the SNPware plate. The SNPware plate with the hybridization mix was incubated at 42°C for 2 h (±15 min) in a humid chamber. Thereafter, the SNPware plate was washed three times, adding 1 × SNPware Wash Buffer II to each well to remove the probes not arrayed.

Data Analysis

The SNPware plate was analyzed with the GenomeLab SNPStream array imager. Blue and green lasers (488 and 532 nm, respectively) detected the fluorescent color of the extended base for each spot. GetGenos software performed the image analysis, assigning the genotype to each SNP spot on the basis of the relative fluorescence intensity for any SNP. Self-annealing oligonucleotides, extended with dye-labeled ddNTPs as described previously (extension reaction) and hybridized to specific probes in each 4 × 4 array, are used to get a correct grid alignment. The top-left location of the 4 × 4 array contained a mixture of two probes hybridizing to self-extending oligonucleotides that can incorporate both blue and green dye-labeled ddNTPs (heterozygous control), while the top-right and the bottom-left positions had probes that hybridized to self-extending oligonucleotides incorporating just blue or green dye-labeled ddNTPs, respectively (homozygous controls) (Fig. 1). Moreover, the bottom-right location had probes that were not self-extending and lacked complementarity to any DNA in the reaction (negative control).

In the genotype analysis, three possible calls to each spot in the 384 SNPware plate can be obtained and collected into three clusters, one representing the homozygous XX genotype (as shown in the rightmost side of the graph), one representing the heterozygous XY genotype (central cluster), and one representing the homozygous YY genotype (as shown in the leftmost side of the graph) (Fig. 2) (Huang et al. 2004). Plot data can be finally converted into an Excel format by the software itself, registering the genotype observed for each SNP in the samples analyzed.

Genotyping Three Polymorphisms by Nanogen Technology

The *MTHFR* C677T (rs1801133), *MTHFR* A1298C (rs1801131), and *MTR* A2756G (rs1805087) gene polymorphisms were evaluated by the NanoChip Molecular Biology Workstation (NanoChip cartridge, Nanogen, San Diego, CA), using the protocol extensively described in a previous paper (Giusti et al. 2004). *MTHFR* and *MTR* gene sequences were obtained from GeneBank (http://www.ncbi.nlm.nih.gov, accession nos. AL953897.6 and AC013268.5). The oligonucleotides used for PCR and for hybridization, along with the melting, annealing, and stringency temperatures, are reported in Supplementary Table B (data available from the corresponding author on request). The software of the system directly assigned the genotype to each sample.

Linkage Disequilibrium and Haplotype Analysis

To prepare data files for linkage disequilibrium by SNPalyze software and for haplotype reconstruction analysis by Phase version 2.1 software, data files obtained by GetGenos SNPStream software were processed in the R environment (http://www.r-project.org).

Pairwise linkage disequilibrium was evaluated using SNPalyze software (Dynacom, Chiba, Japan).

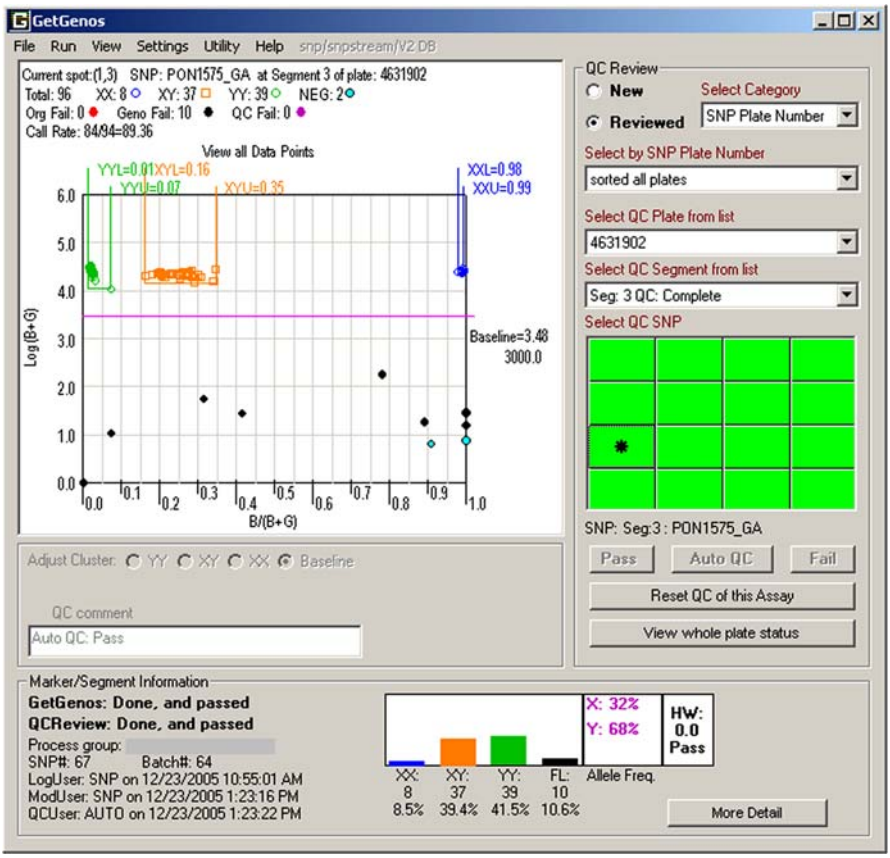

Fig. 2 Computer-displayed scatter plot of the TAMRA-Bodipy fluorescence associated with PON1 575G > A SNP (rs662). The three clusters of genotypes are represented by green (YY), orange (XY), and blue (XX)

The Phase version 2.1 software (Scheet and Stephens 2006) is an implementation of the Bayesian method of haplotype reconstruction, allowing use of a priori expectations to assign haplotypes correctly to individuals in a dataset. Phase uses a Gibb's sampling algorithm, a type of Markov chain-Monte Carlo algorithm, to obtain an approximate sample from the posterior distribution of Pr(H|G), where *H* is the unknown haplotype pair and *G* is the known genotype.

Results

We selected 72 SNPs in 17 candidate genes: 7 in *MTHFD1* [methylenetetrahydrofolate dehydrogenase (NADP + dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase], 5 in *NNMT* (nicotinamide *N*-methyltransferase), 3 in *PON1* (paraoxonase 1), 2 in *PON2* (paraoxonase 2), 5 in

TCN2 (transcobalamin II), 4 in *AHCY* (S-adenosylhomocysteine hydrolase), 7 in *MTRR* (5-methyltetrahydrofolate-homocysteine methyltransferase reductase), 5 in *BHMT* (betaine-homocysteine methyltransferase), 3 in *FOLH1* [folate hydrolase (prostate-specific membrane antigen) 1], 5 in *MTHFR* [5,10-methylenetetrahydrofolate reductase (NADPH)], 6 in *MTR* (5-methyltetrahydrofolate-homocysteine methyltransferase), 2 in *TYMS* (thymidylate synthetase), 1 in *ENOSF1* (enolase superfamily member 1), 3 in *SHMT1* [serine hydroxymethyltransferase 1 (soluble)], 5 in *SLC19A1* [solute carrier family 19 (folate transporter), member 1], 7 in *CBS* (cystathionine-beta-synthase), and 2 in *BHMT2* (betaine-homocysteine methyltransferase 2).

They were analyzed in six panels of 12 SNPs each, according to their nucleotide substitution (1 AT, 1 CA, 1 GC, 1 GA, and 2 CT panels).

The conversion rate among the six panels ranged from 67 to 100%, with 12/12 SNPs working in 1 panel, 10/12 in 1 panel, 9/12 in 3 panels, and 8/12 in 1 panel. Therefore, we could analyze 57 out of 72 (79.2%) SNPs. If we excluded the panel with the higher number of failed SNPs, we observed a conversion rate of 81.7% for oligonucleotides designed by Autoprimer software.

We genotyped 376 subjects. The genotype distribution and allele frequency of the 57 SNPs are reported in Table 1. At the first analysis, the median number of called samples was 340 (range: 136–368), with a median percentage of called samples of 90.4 (range: 36.2–97.9%).

All the passed SNPs under study resulted in Hardy-Weinberg equilibrium.

For the polymorphisms C677T and A1298C in the MTHFR gene and A2756G in the MTR gene, we compared data obtained with the high-throughput GenomeLab SNPStream technology with those obtained with the low-to-medium throughput Nanogen electronic microchip technology in 288 subjects. We found a 99.2% concordance between the results obtained with the two technologies. This concordance was observed for all of the samples independent of the genotype and the SNPs studied. For the seven discordant samples, 6/7 discordant genotypes had low fluorescence signals in the Nanogen analysis, and the repetition of the PCR and Nanogen analysis confirmed the genotype identified by the SNPStream technology. The direct sequencing of the 1/7 discordant genotype with apparently the correct genotype with the two platforms demonstrated that the correct genotype was that identified by SNPStream technology (data not shown).

Pairwise Linkage Disequilibrium Measurement and Haplotype Reconstruction

The pairwise linkage disequilibrium between all pairs of SNP markers of a gene in the subjects studied is reported in Supplementary Table C (data available from the corresponding author on request), as indicated by the D′, r^2, Akaike's information criterion, and *P* value for the chi-squared test (Stephens et al. 2001). Several genes among those investigated showed a high percentage of SNP, with D′ values greater than 0.5. These data suggested that sequence variants within the considered gene are in strong disequilibrium.

Table 1 Genotype distribution and allele frequency of all selected and analyzed SNPs

Gene	Accession No.	Ensemble gene	SNP ID	Region	Position[a]	Panel	Passed\ Failed[b]	Number of called genotype	Genotype distribution *n* (%)			Allele frequency	Validation status
AHCY	AL356299	ENSG00000101444	rs819146	5′utr	1	2 CA	Passed	316(84.1)	AA(66.7)	AC(27.0)	CC(6.3)	C = 0.200	yes
20cen-q13.1			rs7271501	intron	11481	3 CG	Passed	368(97.9)	CC(99.7)	CG(0.3)	GG(0.0)	G = 0.001	yes
			rs819159	intron	12720	1 AT	FAILED						yes
			rs4239	intron	23040	1 AT	FAILED						yes
BHMT	AC008502.8	ENSG00000145692	rs651852	intron	1	4 CT	Passed	357(94.9)	CC(29.8)	CT(51.5)	TT(18.7)	T = 0.450	yes
5q13.1-q15			rs567754	intron	7356	5 CT	Passed	337(89.6)	CC(43.4)	CT(44.5)	TT(12.1)	T = 0.343	yes
			rs3733890	exon	12899	4 CT[c]	Passed	332(88.3)	CC(46.5)	CC(39.0)	TT(14.5)	T = 0.340	yes
			rs10037045	intron	13510	1 AT	FAILED						yes
			rs585800	3′utr	18148	1 AT	Passed	356(94.7)	AA(59.0)	AT(36.0)	TT(5.0)	T = 0.180	yes
BHMT2	AC008502.8	ENSG00000132840	rs644191	intron	1	6 GA	Passed	336(89.4)	AA(68.0)	AG(28.0)	GG(4.0)	G = 0.177	yes
5q13			rs682985	exon	8344	4 CT	Passed	352(93.6)	TT(38.9)	TC(57.1)	CC(20.0)	C = 0.420	yes
CBS	AP001630.1	ENSG00000148450	Pro88Ser	exon	1542	4 CT	Passed	364(96.8)	CC(91.0)	CT(9.0)	TT(0.0)	T = 0.040	no
21q22.3			Ala114Val	exon	3753	4 CT	FAILED						no
			Gly116Arg	exon	3758	6 GA	Passed	352(93.6)	GG(100)	GA(0.0)	AA(0.0)	A = 0.000	no
			rs234706	exon	4866	4 CT	FAILED						yes
			rs12329790	exon	7032	5 CT	FAILED						yes
			rs1801181	exon	9600	5 CT	FAILED						no
			rs1051319	3′utr	16349	3 CG[c]	Passed	356(94.7)	GG(77.0)	GC(22.0)	CC(1.0)	C = 0.117	yes
ENOSF1	AP001178.4	ENSG00000132199	rs8423	3′utr	14776[d]	2 CA	Passed	128(34.0)	AA(28.0)	AC(52.0)	CC(20.0)	A = 0.540	yes
18p11.32													
FOLH1	AC110742.4	ENSG00000086205	rs202676	exon	1	6 GA[c]	Passed	344(91.5)	AA(67.0)	AG(31.0)	GG(2.0)	G = 0.170	yes
11p11.2			rs202680	exon	5735	1 AT	Passed	352(93.6)	AA(5.7)	AT(33.8)	TT(60.5)	T = 0.220	yes
			H475Y	exon	35616	5 CT	FAILED						

Table 1 continued

Gene	Accession No.	Ensemble gene	SNP ID	Region	Position[a]	Panel	Passed\ Failed[b]	Number of called genotype	Genotype distribution *n* (%)			Allele frequency	Validation status
MTHFD1	AL122035.6	ENSG00000100714	rs2357481	5′utr	1	1 AT	Passed	336(89.4)	AA(99.0)	AT(1.0)	TT(0.0)	T = 0.003	no
14q24			rs1076991	5′utr	7040	5 CT[c]	Passed	360(95.7)	TT(30.1)	TC(48.4)	CC(21.5)	C = 0.457	yes
			rs3783732	Intron	7793	1 AT	Passed	136(36.2)	AA(100)	AT(0.0)	TT(0.0)	T = 0.000	yes
			rs8003379	Intron	25598	2 CA	Passed	264(70.2)	AA(44.5)	AC(44.5)	CC(1.0)	C = 0.334	yes
			rs1950902	Exon	32638	5 CT	Passed	306(81.4)	CC(78.4)	CT(20.3)	TT(1.3)	T = 0.114	yes
			rs4902283	Exon	34870	4 CT	Passed	332(88.3)	CC(98.0)	CT(2.0)	TT(0.0)	T = 0.009	yes
			rs2236225	Exon	59199	6 GA	Passed	348(92.5)	GG(37.0)	GA(44.0)	AA(19.0)	A = 0.415	yes
MTHFR	AL953897.6	ENSG00000177000	rs1801133	Exon	1	2 CA	Passed	316(84.0)	CC(11.0)	CA(40.0)	AA(49.0)	A = 0.456	yes
1p36.3			rs1801131	Exon	1902	4 CT	Passed	336(89.4)	CC(27.0)	CT(52.0)	TT(21.0)	T = 0.473	yes
			rs2274976	Exon	5451	5 CT[c]	Passed	360(95.7)	CC(89.8)	CT(10.2)	TT(0.0)	T = 0.052	yes
			rs4846049	3′utr	6013	2 CA	Passed	324(86.2)	CC(55.0)	CA(34.0)	AA(11.0)	A = 0.286	yes
			rs1537514	3′utr	8310	3 CG	Passed	212(56.4)	CC(95.0)	CG(5.0)	GG(0.0)	G = 0.002	yes
MTR	AC013268.5	ENSG00000116984	rs4659725	Intron	1	3 CG	Passed	352(93.6)	CC(30.0)	CG(48.0)	GG(22.0)	G = 0.456	yes
1q43			rs180508	Exon	55039	6 GA	Passed	344(91.5)	AA(72.0)	AG(25.0)	GG(3.0)	G = 0.152	no
			rs2275566	Intron	55101	5 CT	Passed	289(76.9)	TT(30.1)	TC(21)	CC(48.9)	C = 0.455	yes
			rs2853522	3′utr	67595	2 CA	FAILED						yes
			rs2853523	3′utr	68737	2 CA	Passed	336(89.4)	CC(37.0)	CA(42.0)	AA(21.0)	A = 0.419	yes
			rs6676866	3′utr	71165	2 CA[c]	Passed	340(90.4)	CC(30.0)	CA(46.0)	AA(24.0)	A = 0.470	yes
MTRR	AC010346.6	ENSG00000124275	rs326118	5′utr	1	2 CA[c]	Passed	340(90.4)	AA(77.0)	AC(20.0)	CC(3.0)	C = 0.132	yes
5p15.3-p15.2			rs1801394	Exon	2438	6 GA	Passed	352(93.6)	AA(32.0)	AG(44.0)	GG(24.0)	G = 0.463	yes
			rs1532268	Exon	9644	4 CT	Passed	348(92.5)	CC(41..4)	CT(46..2)	TT(12.4)	T = 0.356	yes
			rs2303080	Exon	9889	1 AT	Passed	324(86.2)	TT(99.0)	TA(1.0)	AA(0.0)	A = 0.005	yes

Table 1 continued

Gene	Accession No.	Ensemble gene	SNP ID	Region	Position[a]	Panel	Passed\ Failed[b]	Number of called genotype	Genotype distribution *n* (%)			Allele frequency	Validation status
			rs10064631	Exon	17420	3 CG[c]	Passed	360(95.7)	GG(98.0)	GC(2.0)	CC(0.0)	C = 0.010	yes
			rs16879334	Exon	23020	3 CG	Passed	332(88.3)	CC(100)	CG(0.0)	GG(0.0)	G = 0.00	no
			rs8659	3′utr	32347	1 AT	Passed	348(92.5)	AA(36.0)	AT(50.0)	TT(14.0)	T = 0.390	yes
NNMT	AP002518.3	ENSG00000166741	rs566775	5′utr	1	4 CT	Passed	324(86.2)	CC(62.5)	CT(35.5)	TT(2)	T = 0.200	yes
11q23.1			rs511430	5′utr	3193	3 CG	FAILED						no
			rs10891639	5′utr	8374	2 CA[c]	Passed	336(89.4)	CC(70.0)	CA(26.0)	AA(4.0)	A = 0.165	no
			rs4646335	exon	41683	1 AT	Passed	368(97.9)	TT(70.0)	TA(27.0)	AA(3.0)	A = 0.168	yes
			rs3819100	Intron	42217	4 CT	Passed	360(95.7)	TT(60.0)	TC(34.0)	CC(6.0)	C = 0.230	yes
PON1	AC004022.1	ENSG00000005421	rs854560	exon	1	1 AT	Passed	352(93.6)	TT(41.0)	TA(49.0)	AA(10.0)	A = 0.345	yes
7q21.3			rs662	exon	8638	6 GA	Passed	344(91.5)	AA(51.0)	AG(39.0)	GG(10.0)	G = 0.294	yes
			rs3917594	exon	9645	6 GA	Passed	356(94.7)	GG(100)	GA(0.0)	AA(0.0)	A = 0.000	yes
PON2	AC005021.1	ENSG00000105854	rs11545941	exon	1	3 CG	Passed	356(94.7)	CC(63.8)	CG(34.6)	GG(1.6)	G = 0.190	yes
7q21.3			rs6954345	exon	6240	3 CG	FAILED						yes
SLC19A1	AL133493.3	ENSG00000035928	rs3788205	5′utr	1	6 GA	Passed	304(80.8)	GG(57.0)	GA(39.0)	AA(4.0)	A = 0.231	yes
21q22.3			rs3177999	5′utr	6462	5 CT	Passed	304(80.8)	TT(36.3)	TC(48.9)	CC(14.8)	C = 0.393	yes
			rs1051266	5′utr	6584	5 CT[c]	Passed	256(68.1)	CC(34.3)	CT(51.2)	TT(14.5)	T = 0.401	yes
			rs12659	exon	12822	5 CT	FAILED						yes
			rs1051296	3′utr	29551	2 CA[c]	FAILED						yes
SHMT1	AC127537.7	ENSG00000176974	rs638416	intron	1	3 CG[c]	Passed	340(90.4)	CC(32.0)	CG(50.0)	GG(18.0)	G = 0.433	yes
17p11.2			rs1979277	exon	34893	6 GA[c]	Passed	328(87.2)	GG(99.0)	GA(1.0)	AA(0.0)	A = 0.003	yes
			rs3783	exon	34972	3 CG	FAILED						yes
TCN2	AC005006.2	ENSG00000185339	rs5749131	5′utr	1	6 GA	Passed	324(86.2)	GG(43.0)	GA(44.0)	AA(13.0)	A = 0.350	yes
22q12.2			rs9606756	exon	5038	4 CT[c]	FAILED						yes

Table 1 continued

Gene	Accession No.	Ensemble gene	SNP ID	Region	Position[a]	Panel	Passed\ Failed[b]	Number of called genotype	Genotype distribution *n* (%)			Allele frequency	Validation status
			rs1801198	exon	9788	3 CG	Passed	348(92.5)	CC(42.0)	CG(45.0)	GG(13.0)	G = 0.354	yes
			rs2072195	3′utr	20911	1 AT	Passed	156(41.5)	TT(95.0)	TA(5.0)	AA(0.0)	A = 0.030	yes
			rs10418	3′utr	21130	6 GA[c]	Passed	344(91.5)	GG(58.0)	GA(38.0)	AA(4.0)	A = 0.233	yes
TYMS	AP001178.4	ENSG00000176890	rs502396	intron	1	5 CT	Passed	302(80.3)	CC(28.1)	CT(49.0)	TT(22.9)	T = 0.473	yes
18p11.32			rs16430	3′utr	14208	2 CA[c]	Passed	140(37.2)	CC(26.0)	CA(54.0)	AA(20.0)	A = 0.465	yes

[a] Polymorphisms are numbered downstream from the first SNP at its 5′ of the genomic sequence

[b] Spots with low fluorescence intensity or ambiguous genotype calls are classified as failed

[c] Primer designed on complementary sequence

[d] Nucleotide is numbered from the first SNP localized in the TYMS gene sequence

Table 2 shows the results of the haplotype reconstruction analysis that we used in association analyses. All haplotypes with an estimated frequency higher than 0.01 are reported.

Comparative Cost Analysis

Table 3 compares the costs of the GenomeLab SNPStream technology with respect to a low-to-medium throughput technology, the electronic microchip technology, and a classical manual approach such as RFLP analysis. The cost per SNP of the electronic microchip technology was 3.85 times that of the SNPStream, and RFLP analysis cost 4.26 to 9.92 times as much.

Discussion

The emerging role of single-nucleotide polymorphism in clinical association and pharmacogenetic studies has created a need for high-throughput genotyping technologies.

In this study, we describe the development, optimization, and validation of a microarray approach using a commercial system, GenomeLab SNPStream technology, based on the minisequencing single-nucleotide primer extension principle, for large-scale SNP detection in genes encoding molecules involved in Hcy metabolism.

Mild hyperHcy is an established independent marker factor for atherothrombotic diseases (Eikelboom et al. 1999; McCully 1996). The multifactorial pathogenesis of this disorder includes environmental and genetic factors interacting to produce high plasma Hcy levels (Fowler 2005; Gellekink et al. 2005).

Polymorphic variants in genes encoding enzymes involved in Hcy metabolism (MTHFR, MTR, MTRR, CBS) (Frosst et al. 1995; Weisberg et al. 1998) influence their expression and functional activities and have been proposed as susceptibility factors for hyperHcy. Several studies are available in the literature that investigate the genetic factors involved in hyperHcy, but they did not involve sufficient numbers of patients and/or SNPs to assure statistically valid conclusions (Lewis et al. 2005).

The heterogeneous nature of this complex trait highlights the need to study a larger number of samples and SNPs to perform genetic association studies useful in the evaluation of the underlying genetic causes of hyperHcy and of the mechanisms at the basis of the beginning and of the progression of cardiovascular diseases. Moreover, understanding the genetic architecture of the complex polygenic trait, namely hyperHcy, or in general the misregulation of methionine metabolism is a fundamental goal of the biological and medical research involved in other fields such as cancer and neurological disorders (Gellekink et al. 2005; Sharma et al 2006; Kwok 2001).

Association studies need to genotype a dense set of SNPs in a large panel of individuals and to test each SNP, or set of local haplotypes constructed from the SNP data, in the genotype-phenotype association. Even with some a priori

Table 2 Haplotype frequency

Haplotype[a,b]	Frequency	Haplotype[a,b]	Frequency	Haplotype[a,b]	Frequency
MTHFR		GCP2		CBS	
taccc	0.430	aa	0.514	cgtg	0.428
cccac	0.222	at	0.314	cgcg	0.412
caccc	0.219	ga	0.165	cgcc	0.078
cctac	0.029	Others	0.007	cgtc	0.037
cccag	0.024			tgcg	0.024
cacac	0.019	MTR		tgcc	0.012
tcccc	0.017	gacaa	0.419	Others	0.009
ccccc	0.012	catcc	0.381		
tatcc	0.011	cgtcc	0.131	MTHFD1	
others	0.017	catca	0.024	ataacca	0.236
		ggcaa	0.015	acacccg	0.186
NNMT		gaccc	0.011	ataaccg	0.171
cctt	0.518	Others	0.019	acaaccg	0.100
tctt	0.153			acaacca	0.062
ccac	0.091	RFC1		acaccca	0.062
catt	0.064	gtc	0.402	ataatcg	0.046
caac	0.039	gct	0.359	atacccg	0.035
cctc	0.032	atc	0.198	ataccca	0.024
catc	0.031	act	0.037	acaatcg	0.021
tcac	0.019	Others	0.004	ataatca	0.019
tctc	0.010			acactcg	0.011
tatc	0.010	TCN2		Others	0.027
taac	0.010	gctg	0.356		
Others	0.023	agtg	0.288	MTRR	
		gcta	0.210	aactgct	0.242
BHMT2		ggtg	0.056	agctgca	0.241
at	0.568	actg	0.038	agttgca	0.178
ac	0.249	acta	0.021	aattgca	0.158
gc	0.165	gcag	0.014	cactgct	0.090
gt	0.017	Others	0.017	aactgca	0.015
				aattgct	0.010
BHMT		TYMS		Others	
		tca	0.383		
ttca	0.287	cac	0.346	SHMT	
ccta	0.254	cca	0.138	cg	0.563
ccct	0.161	tac	0.074	gg	0.433
ccca	0.084	ccc	0.042	Others	0.004
tcta	0.079	tcc	0.010		
tcct	0.051	Others	0.007	AHCY	
ctca	0.036			ac	0.823

Table 2 continued

Haplotype[a,b]	Frequency	Haplotype[a,b]	Frequency	Haplotype[a,b]	Frequency
tcca	0.019			cc	0.175
others	0.029			Others	0.002

[a] All SNPs in haplotypes are reported from 5′ to 3′ end of the relative gene, excluding failed SNPs

[b] All haplotypes with an estimated frequency lower than 0.01 are reported as "others"

Table 3 Cost of SNP detection by three methods

Method	Supplies[a] €	Technician			Total running cost per SNP €
		Total time consumed[b] hours	Hands-on time hours	Wages €	
RFLP	2–5	216	96	0.26	2.26–5.26
Microelectronic Chip[c]	2	31	14	0.04	2.04
GenomeLab SNPStream	0.50	26	12	0.03	0.53

[a] "Supplies" includes all materials common to the three genotyping methods (Thermowell plates, DNA polymerase, deoxynucleotides, primers, Thermowell seals, tips) and materials specific to a method: for RFLP, agarose gel, Tris-acetate-EDTA buffer, DNA molecular weight markers, Polaroid film to photograph DNA fragments and restriction endonucleases; for microelectronic chip, Macherey-Nagel 96-well plates, Nanogen cartridges, universal reporters, histidine buffer, high and low salt buffers; for GenomeLab SNPStream, SNPware Core Reagent kit, SNPware two dideoxynucleotide extension mix kit, SBE clean-up reagent, extension oligonucleotides

[b] Time consumed when genotyping 384 samples for 12 SNPs

[c] According to our optimized protocol previously published (Giusti et al. 2004)

knowledge of a candidate gene region contributing to a disease phenotype, a large number of SNPs need to be genotyped in an association study to ensure one of the genotyped SNPs is causative or is in strong linkage disequilibrium with the causative site. It is also important that SNPs are genotyped in a very large panel of individuals to provide sufficient power to detect variants that may have only subtle phenotypic effects.

During the last decade the wide-ranging application of SNP in research and molecular diagnostics has required further development of robust, flexible, cost-effective technology platforms for scoring genotypes in large numbers of samples to evaluate the genotype–phenotype correlation. Moreover, after the pathogenetic role of a polymorphism/mutation is identified, clinical diagnosis (diagnostic test) for analysis of many SNPs in large sample sets will be required, which must satisfy the quality features needed in the routine of a molecular genetic laboratory. A plethora of SNP genotyping platforms is currently available (Kwok 2001; Syvanen 2001, 2005), offering different advantages and disadvantages with respect to the characteristics of the technology (principle, cost, automation) and to the type of study to be performed.

Among the available advanced technologies, first we chose low-throughput Nanogen technology (100–1400 SNPs for chip) (Giusti et al. 2004) and more recently high-throughput GenomeLab SNPStream technology, based on the primer extension principle (Denomme and Van Oene 2005). The latter system uses fluorescent multiplex SNP-IT primer extension assays for allele determination of previously discovered SNPs and allows genotyping of up to 890,000 SNP genotypes per day, with only 4 ng of genomic DNA (Huang et al. 2004).

In the current study we selected and analyzed SNPs in 17 genes that may play a role in modulating Hcy plasma levels and/or DNA methylation reactions, so altering genetic stability and gene expression. In each gene we identified several polymorphisms independent from their known role in influencing Hcy levels and DNA methylation reactions. We chose SNPs according to their demonstrated or putative function evaluated on the basis of literature data, their localization in the promoter or other regulatory regions or exons, and their potential utility in haplotype reconstruction.

Among the 72 selected SNPs, 57 passed all the steps in the software- and technician-surveilled quality check and were thus suitable for the genotyping analysis. The conversion rate of 79.2% rose to 81.7% if we excluded the panel with the higher number of failed SNPs among the six designed panels.

Adherence to Hardy-Weinberg equilibrium (HWE) is a common criterion to assess the quality of a genotyping assay, as a deviation can suggest incorrect genotype assignments (Hosking et al. 2004). However, selection, mutation, or migration can also cause deviation from HWE, and the power to detect these processes increases with the sample size. Of our 57 converted assays, no SNP showed a deviation from HWE.

In the first analysis, our data showed a call rate higher than 90%, meaning that less than 10% of the samples needed to be reanalyzed. This is an important issue, as the higher the call rate, the lower the need for further analysis, which reduces costs and time of analysis and increases the benefit of automation.

The analysis of linkage disequilibrium of the SNPs and of the haplotype reconstruction indicates that the selected SNPs could be suitable to investigate and could illuminate the role of genes involved in Hcy metabolism in determining plasma Hcy levels and clinical phenotypes.

In this study we compared data from two technologies available in our laboratory, the Nanogen technology and the GenomeLab SNPStream technology. Our data demonstrated 99.2% concordance between the results obtained by the two technologies. We observed that six of the seven discordant samples were failed by the Nanogen technology because of their low signal-to-noise ratio. The repetition of PCR and analysis by Nanogen technology of the six samples confirmed the results obtained by the GenomeLab SNPStream technology. Concerning one sample, the repetition of the analysis by the two technologies confirmed the discordant result. The direct sequencing analysis showed that the correct genotype was that obtained using the GenomeLab SNPStream technology. This datum confirms the expected higher accuracy of primer extension based technologies in allele discrimination.

Genomelab SNPStream is a cost-effective and time-saving alternative. In fact, the cost of this high-throughput technology is a quarter that of the Nanogen

technology and can be as little as a tenth the cost of a classical RFLP genotyping assay. Moreover, this technology permits shorter analysis time compared with Nanogen and RFLP approaches (16% and 88%, respectively), resulting in the possibility of increasing the number of genotyped SNPs over which to amortize the cost of the instrument.

In conclusion, our data highlight that the described approach using the commercially available high-throughput GenomeLab SNPStream technology offers considerable advantages for SNP genotyping on a large scale, representing a useful tool to investigate the genotype-phenotype correlation and the association between the methionine cycle genes and hyperHcy and/or diseases correlated to methionine metabolism alterations, as well as other complex multifactorial traits.

Acknowledgments This work was supported by grants from Genopolis government FIRB project (RBLA038RMA_008) and from the Ente Cassa di Risparmio di Firenze to CeRA (Multidisciplinary Center of Research on Food Sciences), Florence, Italy.

References

Applebaum J, Shimon H, Sela BA, Belmaker RH, Levine J (2004) Homocysteine levels in newly admitted schizophrenic patients. J Psychiatr Res 38:413–416

Blom HJ, Shaw GM, den Heijer M, Finnell RH (2006) Neural tube defects and folate: case far from closed. Nat Rev Neurosci 7:724–731

Castro R, Rivera I, Martins C, Struys EA, Jansen EE, Clode N, Graca LM, Blom HJ, Jakobs C, de Almeida IT (2005) Intracellular S-adenosylhomocysteine increased levels are associated with DNA hypomethylation in HUVEC. J Mol Med 83:831–836

Castro R, Rivera I, Blom HJ, Jakobs C, Tavares de Almeida I (2006) Homocysteine metabolism, hyperhomocysteinaemia and vascular disease: an overview. J Inherit Metab Dis 29:3–20

Clarke R, Smith AD, Jobst KA, Refsum H, Sutton L, Ueland PM (1998) Folate, vitamin B12, and serum total homocysteine levels in confirmed Alzheimer disease. Arch Neurol 55:1449–1455

De Luis DA, Fernandez N, Arranz ML, Aller R, Izaola O, Romero E (2005) Total homocysteine levels relation with chronic complications of diabetes, body composition, and other cardiovascular risk factors in a population of patients with diabetes mellitus type 2. J Diabetes Complications 19:42–46

Denomme GA, Van Oene M (2005) High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. Transfusion 45:660–666

Eikelboom JW, Lonn E, Genest J Jr Hankey G, Yusuf S (1999) Homocyst(e)ine and cardiovascular disease: a critical review of the epidemiologic evidence. Ann Intern Med 131:363–375

Fowler B (2005) Homocysteine: overview of biochemistry, molecular biology, and role in disease processes. Semin Vasc Med 5:77–86

Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJH, den Heijer M, Kluijtmans LAJ, van den Heuve LP, Rozen R (1995) A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet 10:111–113

Gellekink H, den Heijer M, Heil SG, Blom HJ (2005) Genetic determinants of plasma total homocysteine. Semin Vasc Med 5:98–109

Giusti B, Frusconi S, Rossi L, Bernabini S, Poggi F, Giotti I, Abbate R, Pepe G, Torricelli F (2004) Improvement of low-density microelectronic array technology to characterize 14 mutations/single-nucleotide polymorphisms from several human genes on a large scale. Clin Chem 50:775–777

Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, Riley J, Purvis I, Xu CF (2004) Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J Hum Genet 12: 395–399

Huang CY, Studebaker J, Yuryev A, Huang J, Scott KE, Kuebler J, Varde S, Alfisi S, Gelfand CA, Pohl M, Boyce-Jacino MT (2004) Auto-validation of fluorescent primer extension genotyping assay using signal clustering and neural networks. BMC Bioinformatics 2:36

Kim YI (2005) Nutritional epigenetics: impact of folate deficiency on DNA methylation and colon cancer susceptibility. J Nutr 135:2703–2709

Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet 2:235–258

Lewis SJ, Ebrahim S, Davey Smith G (2005) Meta-analysis of MTHFR 677C->T polymorphism and coronary heart disease: does totality of evidence support causal role for homocysteine and preventive potential of folate? BMJ 331:1053

McCully KS (1969) Vascular pathology of homocysteinemia: implications for the pathogenesis of arteriosclerosis. Am J Pathol 56:111–128

McCully KS (1996) Homocysteine and vascular disease. Nat Med 2:386–389

Perla-Kajan J, Twardowski T, Jakubowski H (2007) Mechanisms of homocysteine toxicity in humans. Amino Acids 32:561–572

Rudy A, Kowalska I, Straczkowski M, Kinalska I (2005) Homocysteine concentrations and vascular complications in patients with type 2 diabetes. Diabetes Metab 31:112–117

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 78: 629–644

Sharma P, Senthilkumar RD, Brahmachari V, Sundaramoorthy E, Mahajan A, Sharma A, Sengupta S (2006) Mining literature for a comprehensive pathway analysis: a case study for retrieval of homocysteine related genes for genetic and epigenetic studies. Lipids Health Dis. 23

Sharp L, Little J (2004) Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review. Am J Epidemiol 159:423–443

Singh SM, Murphy B, O'Reilly RL (2003) Involvement of gene-diet/drug interaction in DNA methylation and its contribution to complex diseases: from cancer to schizophrenia. Clin Genet 64:451–460

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978–989

Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930–942

Syvänen AC (2005) Toward genome-wide SNP genotyping. Nat Genet Suppl 37:S5-S10

Van Guldener C, Stehouwer CD (2003) Homocysteine metabolism in renal disease. Clin Chem Lab Med 41:1412–1417

Villadsen MM, Bunger MH, Carstens M, Stenkjaer L, Langdahl BL (2005) Methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism is associated with osteoporotic vertebral fractures, but is a weak predictor of BMD. Osteoporos Int 16:411–416

Waterland RA (2006) Assessing the effects of high methionine intake on DNA methylation. J Nutr 136:1706S–1710S

Weisberg I, Tran P, Christensen B, Sibani S, Rozen R (1998) A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. Mol Genet Metab 64:169–172