Available online at www.sciencedirect.com
ScienceDirect
Clinical Biochemistry 42 (2009) 201–209
CLINICAL BIOCHEMISTRY

High-throughput genotyping of copy number variation in *Glutathione S-Transferases M1* and *T1* using real-time PCR in 20,687 individuals

Marianne S. Nørskov [a], Ruth Frikke-Schmidt [a], Steffen Loft [b], Anne Tybjærg-Hansen [a,c,*]

[a] *Department of Clinical Biochemistry, KB 3011, Section for Molecular Genetics, Rigshospitalet, Copenhagen University Hospital Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark*
[b] *Department of Environmental Health, Institute of Public Health, Faculty of Health Sciences, University of Copenhagen, Denmark*
[c] *Faculty of Health Sciences, University of Copenhagen, Denmark*

Received 11 June 2008; received in revised form 3 October 2008; accepted 20 October 2008
Available online 7 November 2008

Abstract

Objectives: Characteristic for the genes encoding glutathione S-transferase (GST) M1 and GSTT1 is a null allele, suggested to increase susceptibility to chronic diseases. We report an optimized method for the determination of copy number variation (CNV) in *GST* genes.

Design and methods: Real-time multiplex PCR reactions were optimized for quantification of *GSTM1* and *GSTT1* CNV using the ΔC_t method, a fixed volume of diluted DNA, a total volume of 10 μL, 384-well formats, and single determinations of each sample.

Results: Consistent genotyping was obtained using DNA in a range of 0.41 ng to 100 ng. In a general population sample of 20,687 individuals the genotype frequencies were concordant with other methods used as standards. Throughput was 4600 genotypes per day at a reagent price of 0.5 euros per sample.

Conclusions: This high-throughput, low cost method accurately determines CNV in the *GST* genes enabling reliable estimates of disease prediction in large epidemiological samples.

Keywords: Copy number variation; Glutathione S-transferase; High-throughput genotyping; Real-time PCR

Introduction

Exposure to environmental factors such as smoking and air pollution causes inflammation, oxidative stress and increased risk of lung cancer [1,2], bladder cancer [3], ischemic heart disease [4–7] and airway disease [8,9].

Glutathione S-transferases (GSTs) are a superfamily of phase II drug-metabolizing enzymes catalyzing the conjugation of reduced glutathione with a variety of electrophilic compounds, including carcinogens and environmental toxins, thereby protecting the cell against xenobiotics and oxidative stress [10]. A *GSTM1*0* null allele is thought to result from homologous unequal crossing over between two highly identical 4.2 kb repeated sequences flanking the *GSTM1* gene, resulting in a 15 kb deletion including the entire *GSTM1* gene [11]. A similar mechanism results in the *GSTT1*0* null allele [12]. A gene dosage effect between gene copy number and enzyme activity has been reported for both *GSTM1* and *GSTT1* [12,13]. The null genotypes *GSTM1*0/0* and *GSTT1*0/0* are associated with complete loss of catalytic activity [14,15], and have been suggested to be associated with increased risk of ischemic heart disease in smokers [16], with asthma [17,18], and with cancer [19–22]. *GSTM1*0/0* individuals also appear to have enhanced allergic responses in the presence of diesel exhaust particles [23]. The frequencies of the *GSTM1*0/0* and *GSTT1*0/0* genotypes in Caucasians are approximately 53% and 20% [24].

To perform large population based, molecular epidemiological studies of *GST* genes and other copy number variation (CNV) genes of potential importance for human disease, high sample throughput in molecular analyses combined with low cost is in high demand. In TaqMan real-time PCR reactions fluorogenic probes hybridize to the target gene sequence, and 5′

* Corresponding author. Department of Clinical Biochemistry, KB 3011, Section for Molecular Genetics, Rigshospitalet, Copenhagen University Hospital Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark. Fax: +45 3545 4160.
E-mail address: at-h@rh.regionh.dk (A. Tybjærg-Hansen).

0009-9120/$ - see front matter © 2008 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.clinbiochem.2008.10.020

nuclease cleavage of the probe leads to an increase in fluorescence proportional to the concentration of target sequences in the initial sample. The signal attributable to the 5′ nuclease reaction is expressed as the normalized change in fluorescence, ΔR_n. The ΔR_n value increases as a function of the number of cycles during the exponential phase of the PCR, because the amplicon copy number increases until the reaction approaches a plateau. The threshold cycle (C_t) is defined as the fractional cycle number at which an increase in ΔR_n above a baseline signal is detected, and is inversely correlated to the initial DNA concentration in a sample on a logarithmic scale.

The exponential amplification in a PCR reaction is described by $Y=X*(1+E)^n$, where X denotes the initial number of copies of the target gene, Y denotes the number of copies after n cycles, and E is the amplification efficiency. An amplification efficiency of 100% means that the amount of copies of the target gene is doubled per cycle in the exponential phase of the amplification reaction ($Y=X*(1+1)^1$).

The first improved real-time PCR methods to determine CNV in *GSTM1* and *GSTT1* in whole blood [17,25,26] and in paraffin embedded tissue [26] were published recently. Because these methods relied on the use of fixed concentrations of sample DNA and triple determinations of each sample in a 96-well format, they were both too time-consuming and too expensive for the determination of CNV in large epidemiological samples. This was also reflected in the limited number of samples genotyped ($n=29$–1032) in these studies.

The aim of the present study was therefore to establish a robust high-throughput, low cost method for the determination of CNV for *GST* genes also amenable to other CNV genes. This was achieved using a fixed volume of DNA extracted by a standard commercial method (QIAamp DNA Mini Kit, Qiagen, Hilden, Germany), but without prior determination of DNA concentration, and by multiplexing target and endogenous control genes in a 10 μL total volume in a 384-well format using single instead of multiple determinations of each sample. We used the ΔC_t method, obviating the need for standard curves, to determine CNV in a sample, where $\Delta C_t=(C_{t\text{Target gene}}-C_{t\text{Endogenous control gene}})$. The endogenous control is a single copy gene present in all samples and serves the purpose of normalizing for differences in input DNA. For the determination of CNV in large samples, the ΔC_t method will group individuals on the same plate into clusters by number of gene copies. Thus, the determination of CNV genotype is dependent on the ΔC_t-value of the unknown samples relative to the ΔC_t-value of the control samples on that specific plate.

We present the development, optimization and validation of this method, which easily generated 4600 genotypes per day at a cost of 0.5 euros per sample, and was used to genotype 20,687 individuals from the Danish general population.

Methods

Identification of controls for determination of copy number variation

Positive controls i.e. individuals carrying at least one copy number of the *GST* genes were identified among 20 participants in the Copenhagen City Heart Study (CCHS, see Subjects): DNA was extracted from 200 μL whole blood using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), and tested with *GSTM1*, *GSTT1* and the endogenous control assay (RNaseP) in singleplex realtime PCR reactions using 2 μL of DNA (diluted 1:10), 250 nmol/L probe, and 36 μmol/L primers (Table 1). Samples with amplification of the *GST* genes were characterized as non-nulls (*GST*1/0* and *GST*1/1),* and samples without amplification as *GST*0/0*. In non-null samples, *GSTT1*0* and *GSTM1*0* alleles were subsequently amplified by long-range PCR using primer pairs as previously described to discriminate between *GST*1/0* and *GST*1/1* [12,27]. The genotypes of the two controls used in all further optimization were: *GSTM1*1/0*, *GSTT1*1/0* (Control 1) and *GSTM1*1/0*, *GSTT1*1/1* (Control 2).

Optimization of method for determination of CNV

Real-time TaqMan assay conditions

Primers and 6-carboxyfluorescein (6-FAM) labeled probes used to amplify *GSTT1* and *GSTM1* in all real-time PCR reactions are shown in Table 1. TaqMan RNaseP Control Reagents Kit containing 20× concentrated VIC-labeled probe and gene specific primers was used as an endogenous control (Applied Biosystems, Foster City, CA, USA). PCR conditions: following an initial step at 50 °C for 2 min and a denaturation step at 95 °C for 10 min, amplification was performed for 40 cycles at 95 °C for 15 s, and at 60 °C for 1 min. All PCR reactions used to optimize the method for determination of CNV were performed in triplicate in 384-well formats, in a 10 μL final volume with 1× TaqMan Universal PCR Master Mix and 250 nmol/L probe unless otherwise noted, using an ABI 7900HT instrument with a plate stacker (Applied Biosystems). Data were analyzed using the Absolute Quantification (Standard Curve)

Table 1
Primers and probes used for determination of copy number variation in *GSTM1* and *GSTT1*

Name	Sequence (5′–3′)	5′ reporter fluorochrome (probes only)	3′ modification (probes only)
*GSTM1*_forward	CTGAGCCCTGCTCGGTTTAG		
*GSTM1*_reverse	ATGGGCATGGTGCTGGTT		
*GSTT1*_ forward	CGGTCGGTCCCCACTATG		
*GSTT1*_reverse	CGAAGGGAATGTCGTTCTTCTT		
*GSTM1*_probe	CTGTCTGCGGAATC	6-FAM [a]	MGB [a]
*GSTT1*_probe	TACCTGGACCTGCTGTC	6-FAM [a]	MGB [a]

[a] FAM: 6-carboxyfluorescein. MGB: minor groove binder. MGB probes contain a non-fluorescent quencher.

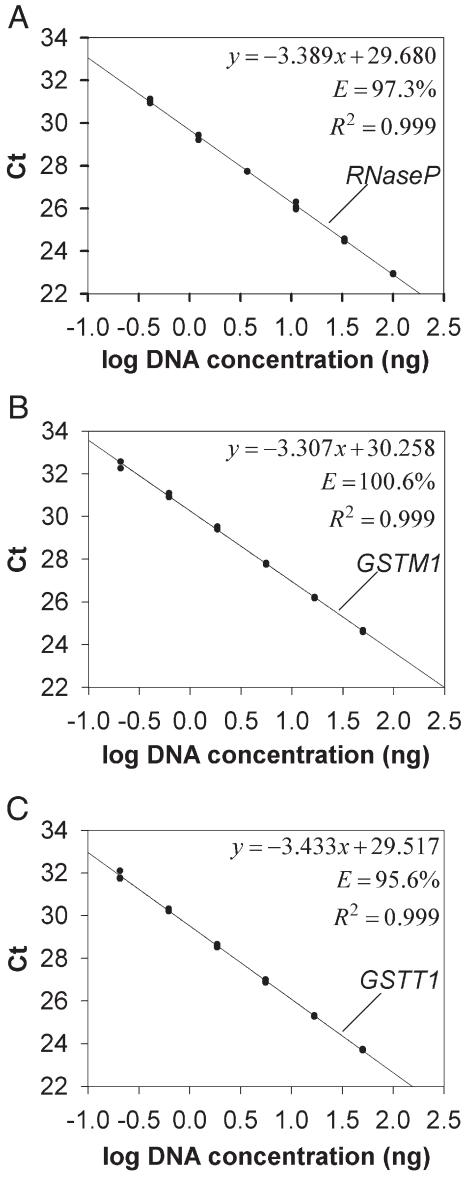

Fig. 1. Determination of efficiency with non-limited primers. Standard curves of genes assayed in singleplex were generated by plotting C_t values versus log DNA concentration for *RNaseP* (Panel A), *GSTM1* (Panel B), *GSTT1* (Panel C). Each DNA concentration was assayed in triplicate, and each point represents one of three C_t determinations. PCR efficiency is calculated as: $E=(10^{(-1/\text{slope})}-1)$. Results for Control 1 are shown in all panels.

document in the SDS software (Version 2.2). The "automatic C_t" option in the analysis setting was used, whereby the optimal baseline and threshold values are set automatically for each 384-well plate.

Requirements for multiplexing using the ΔC_t method

Simultaneous amplification i.e. multiplexing of *RNaseP* and the *GST* gene in the same sample serves the purpose of normalizing for difference in input DNA (concentration) between samples, and ensures that no false positive *GST*0/0* genotypes due to PCR or pipetting failure, or insufficient DNA concentration in the original sample are generated. Determination of CNV in a sample was based on the ΔC_t value, calculated as $C_{tGST}-C_{tRNaseP}$. The use of multiplexing and the ΔC_t method requires that: 1) reactions are primer-limited to ensure that the amplification of the target gene does not influence the amplification of the endogenous control or vice versa; 2) the amplification efficiencies of the *GST* genes and *RNaseP* gene must be similar and close to 100% for singleplex reactions before and after primer limitation and for multiplex reactions; 3) the amplification efficiencies of the target gene and the endogenous control must be reproducible and approximately equal [28]. The fulfilment of these criteria is described below using control DNA from two individuals with known genotypes (see Identification of controls).

Determining the linear dynamic range and defining limiting primer concentrations. The linear dynamic ranges for *GST* and *RNaseP* genes were determined in singleplex reactions by plotting C_t values versus log DNA concentrations for six standard DNA concentrations with 3-fold dilutions in between, starting at 50 ng genomic DNA per well (Fig. 1). Each standard dilution was assayed with 250 nmol/L probe and a default primer concentration of 900 nmol/L was used as non-limiting for all assays. All reactions were performed in triplicate. Based on the linear dynamic range, control DNA with a fixed DNA concentration of 3 ng per well which represented the lowest concentration of input DNA expected for any sample, was used for defining limiting primer concentrations. Primer limitation was performed in singleplex reactions to identify the optimal primer concentration for each gene to be used in multiplex reactions (Fig. 2). The RNaseP mix was primer limited in singleplex reactions using 1x mix (900 nmol/L of each primer and 250 nmol/L probe) and 1/2× mix (450 nmol/L of each primer and 125 nmol/L probe). For GST assays the concentration of each primer in singleplex reactions varied as follows: 300 nmol/L, 200 nmol/L, 150 nmol/L, 100 nmol/L, 75 nmol/L, 50 nmol/L, and 25 nmol/L. The normalized change in fluorescence (ΔR_n) was plotted versus cycle numbers and the lowest primer concentration which did not increase the threshold cycle number (C_t value, i.e. the fractional cycle number at which an increase in ΔR_n above a baseline signal was detected) was selected for further optimization. All reactions were performed in triplicate.

Amplification efficiency and validation of the ΔC_t method. After defining limiting primer concentrations, these should be verified experimentally. This is accomplished by comparing amplification efficiency in singleplex reactions using non-limited primers (Fig. 1) with amplification efficiency in multiplex reactions using limited primers (Figs. 3A, C). Amplification efficiency is calculated from the slope of a standard curve of C_t values plotted against log DNA concentrations, using the following equation: $E=(10^{(-1/\text{slope})}-1)$ [29]. Six DNA standard concentrations were measured with 3-fold dilution in between starting at 100 ng genomic DNA per well. Each standard dilution was assayed with 250 nmol/L probe and limited primer concentrations in multiplex reactions. All reactions were performed in triplicate.

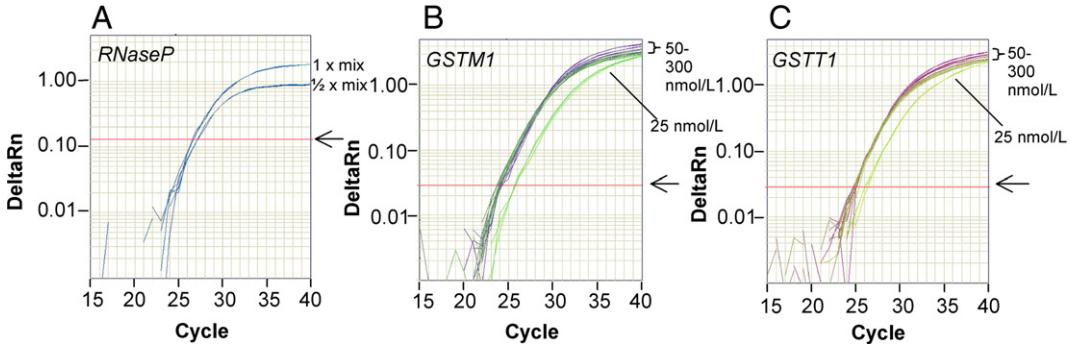

Fig. 2. Primer limitation of GST assays and the RNaseP assay. The RNaseP assay was optimized in primer limitation experiments using 1× mix (900 nmol/L of primers and 250 nmol/L of probe) and 1/2× mix (450 nmol/L of primers and 125 nmol/L of probe). The normalized change in fluorescence (ΔR_n) was plotted on a log-axis versus cycle number (Panel A). The (red) horizontal line indicates the threshold (marked with an arrow), which was set automatically using the "automatic C_t" option, as recommended by the manufacturer. The GSTM1 (Panel B) and GSTT1 (Panel C) assays were optimized with a dilution series of primers (as indicated) and 250 nmol/L of probe. Results for Control 1 are shown in all panels.

To validate the use of the ΔC_t method, the slope of ΔC_t ($C_{tGST} - C_{tRNaseP}$) versus log DNA concentration in multiplex reactions was evaluated (Figs. 3B, D). A slope of 0 signifies equal amplification efficiencies of the two multiplexed genes, and slopes less than +/−0.1 are accepted for using the ΔC_t method [30].

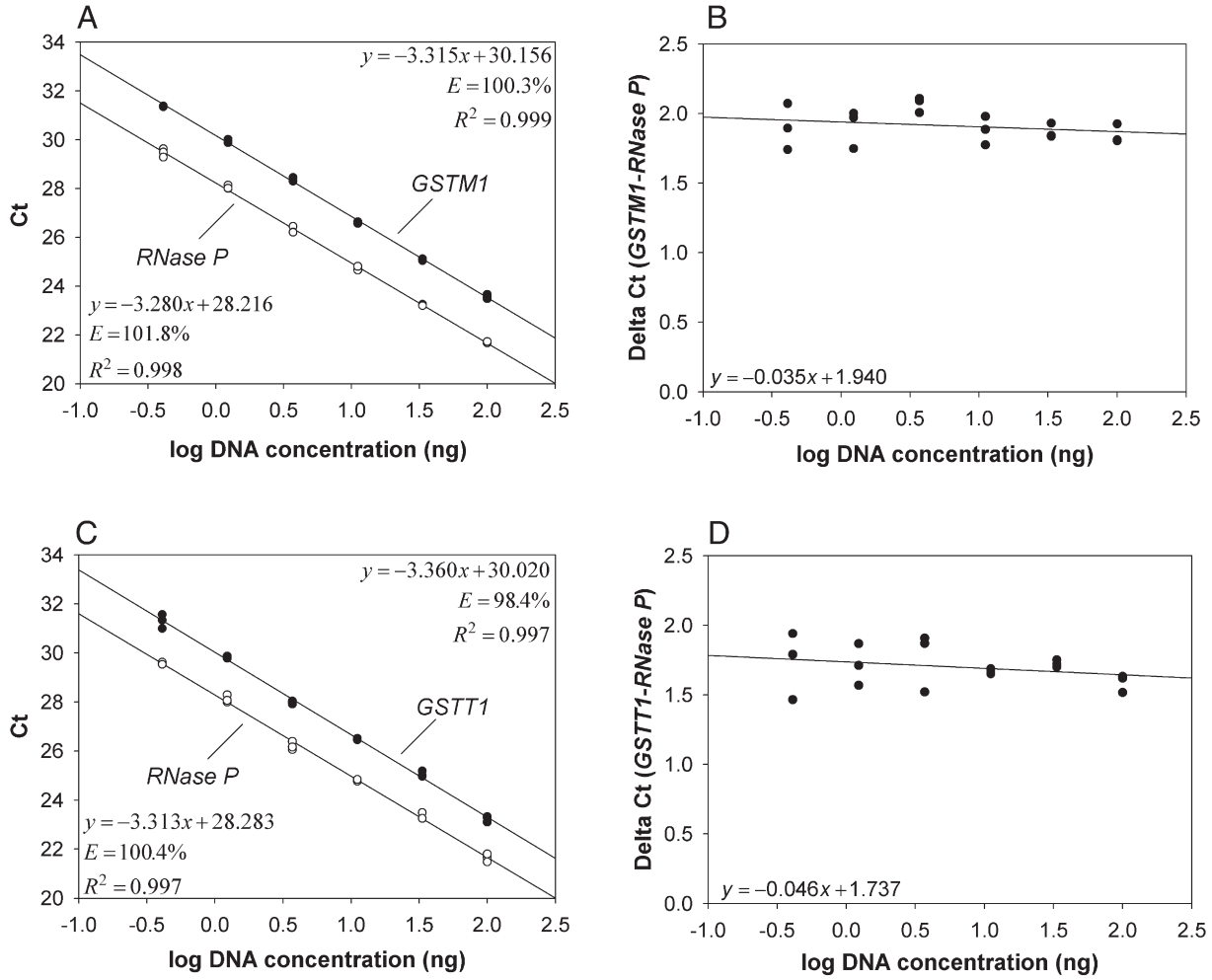

Fig. 3. Determination of efficiency and validation of ΔC_t method. Standard curves of genes assayed in multiplex were generated by plotting C_t values versus log DNA concentration (0.41 to 100 ng) for *GSTM1* (•) and *RNaseP* (o) (Panel A) and *GSTT1* (•) and *RNaseP* (o) (Panel C). The figure shows a typical example of 1 of 10 independent runs. Each DNA concentration was assayed in triplicate, and each point represents one of three C_t determinations. PCR efficiency is calculated as: $E = (10^{(-1/\text{slope})} - 1)$. C_t values in panels A and C were subtracted and plotted versus log DNA concentration (Panels B and D, respectively). Results for Control 1 are shown in all panels.

Reproducibility. To test the reproducibility of the multiplexed reactions, the mean slope of standard curves (plots of C_t values versus log DNA concentrations in multiplex reactions as described above), and the amplification efficiencies for target and control genes were determined for both control samples in 10 consecutive runs. To test the reproducibility of the ΔC_t method the mean differences in amplification efficiencies between target and control genes at different DNA concentrations (ΔC_t versus log DNA concentration from multiplex reactions as described above) were determined for both control samples in 10 consecutive runs.

Genotyping of 20,687 individuals from the general population

Subjects

Studies were approved by institutional review boards and Danish ethical committees (KF)V.100.2039/91 and (KF)01-144/01, Copenhagen and Frederiksberg Committee. Written informed consent was obtained from all participants. All participants were white and of Danish descent.

The Copenhagen City Heart Study. The Copenhagen City Heart Study (CCHS) is a prospective population study of individuals selected based on the national Danish Civil Registration System to reflect the Danish general population aged 20–80+ years. At the 1991–1994 and 2001–2003 examinations, 10,632 participants gave blood for DNA analyses [31,32].

The Copenhagen General Population Study. The Copenhagen General Population Study is a cross-sectional study of the Danish general population initiated in 2003 and still recruiting; the aim is to total 100,000 participants ascertained exactly as in the CCHS [33]. We genotyped the first 10,055 individuals from this study.

Genotyping

DNA was extracted from 200 μL whole blood using QIAamp DNA Mini Kit (Qiagen), eluted in a total volume of 400 μL, diluted ×5, and stored in 96-well microtiter plates. DNA concentration was not determined before use, but the yield from 200 μL whole blood is typically between 4 and 12 μg. DNA from the 20,687 samples from the CCHS and the Copenhagen

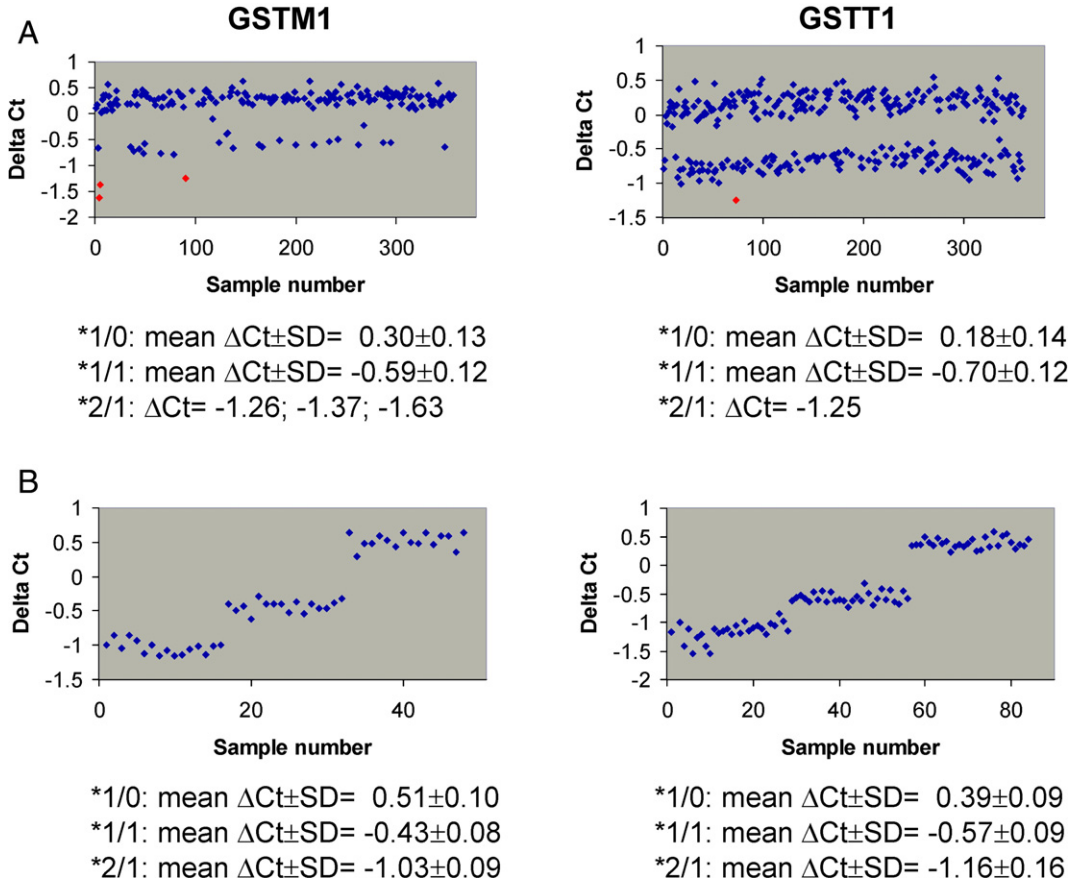

Fig. 4. Scatterplot of ΔC_t values ($C_{tGST}-C_{tRNaseP}$) for *GSTM1* and *GSTT1* non-null samples from the general population sample. Panel A shows two typical plots, where each plot represents non-null samples from 380 samples analyzed on a 384-well plate. Red squares indicate *2/1 samples, the ΔC_t values of which are indicated below the plot. Panel B shows plots where all *2/1 samples found in the entire study among the 20,687 samples ($n=16$ for *GSTM1*, $n=27$ for *GSTT1*) are assayed together with the same number of randomly selected *1/0 and *1/1 samples. On all plots, each square represents one sample. *2/1 samples cluster at the lowest ΔC_t values and *1/0 samples at the highest ΔC_t values, and *1/1 samples in between. The mean ΔC_t value is shown below each plot. *GST*0/0* samples are not plotted, because they are not amplified.

General Population Study were transferred to 384-well microtiter plates using a Biomek 2000 robot for automatic dispensing (Ramcon, Birkerød, Denmark). Aliquots of 2 μL diluted DNA (approximately 4–12 ng) were added to 8 μL PCR mix (1× TaqMan Universal PCR Master Mix, 1× TaqMan RNaseP Control Reagents Kit (900 nmol/L of each primer and 250 nmol/L probe), 100 nmol/L of each *GSTM1* or *GSTT1* primers and 250 nmol/L of *GSTM1* or *GSTT1* probe). In each 384-well plate, samples were tested as single determinations, and two no template controls and two controls with *1/0 and *1/1 genotypes, respectively, were included. Samples were assayed using the Relative Quantification ($\Delta\Delta C_t$) document in the SDS software and analyzed using the Relative Quantification ($\Delta\Delta C_t$) Study document. The "automatic C_t" option in the analysis setting was used, whereby the optimal baseline and threshold values are set automatically for each run. For each 384-well plate, ΔC_t values $=(C_{tGST}-C_{tRNaseP})$ were plotted versus sample number, and samples were assigned a genotype based on the position of the ΔC_t value in genotype clusters (Fig. 4). The two control samples defined the *1/0 and *1/1 genotype clusters. An arbitrary line midway between the *1/0 and *1/1 genotype clusters, was used for the preparation of

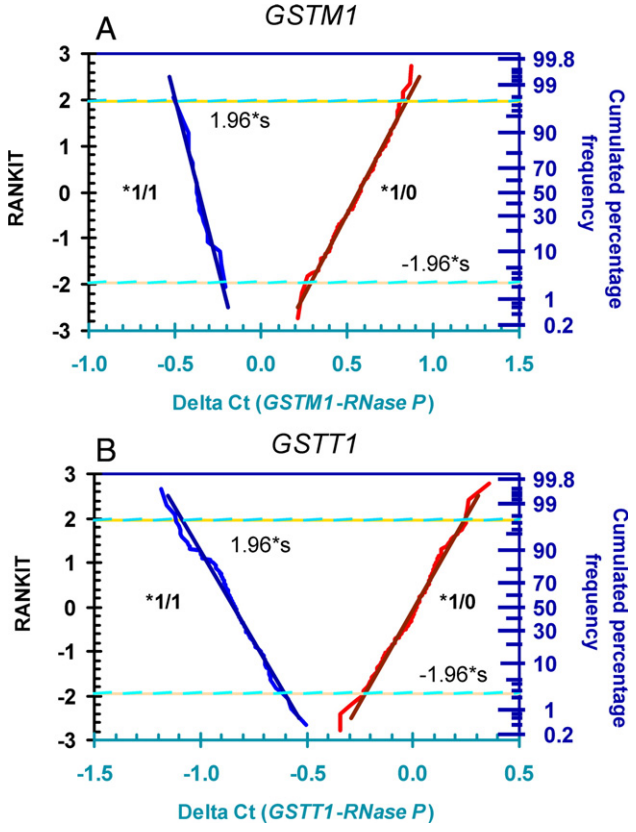

Fig. 5. Identification of cutoff intervals for samples with *1/0 and *1/1 genotypes. Samples were assigned a genotype (*1/0 or *1/1) based on ΔC_t values. ΔC_t values of samples designated *1/0 and *1/1 were sorted ascending and descending, respectively. The cumulated frequency of the two groups of sorted samples was plotted versus ΔC_t values. A Rankit *y*-axis, representing the inverse normal function of the cumulated frequency, is shown to the left. Limits for 2 SD are indicated at Rankit values −1.96 and 1.96 respectively. Results for one run with samples assayed for *GSTM1* (Panel A) and *GSTT1* (Panel B) are shown.

Table 2
GSTM1 and *GSTT1* genotype data of a general population sample ($n=20{,}687$)

Gene	Call	First run [a]	After rerun [b]
GSTM1	*0/0	10,137 (49%)	10,726 (52%)
	*1/0	7804 (38%)	8225 (40%)
	*1/1	1463 (7%)	1598 (8%)
	*2/1	21 (0.1%)	16 (0.1%)
	Undetermined [c]	1262 (6.1%)	122 (0.6%)
GSTT1	*0/0	2793 (14%)	3030 (15%)
	*1/0	9239 (47%)	9764 (47%)
	*1/1	7313 (35%)	7786 (38%)
	*2/1	51 (0.2%)	27 (0.1%)
	Undetermined [c]	1291 (6.2%)	80 (0.4%)

[a] Data are presented as absolute numbers and percentage in parentheses.
[b] Rerun samples are samples that are identified as undetermined after the first run.
[c] The undetermined group includes samples with atypical amplification plots, samples with no amplification of *RNaseP* and samples identified as at risk of misclassification by the Rankit plot.

Rankit plots (Fig. 5, see Statistical analysis). Samples identified as at risk of misclassification on Rankit plots were undetermined samples. Altogether, "undetermined samples" were clear outliers on amplification plots, samples with no amplification of *RNaseP*, and Rankit reruns (Table 2).

Samples were assigned the *2/1 genotype when clearly separated from the *1/1 cluster (Fig. 4A). *2/1 samples were verified in re-amplification runs. Finally, all *2/1 samples identified in the study were assayed together on the same 384-plate together with a comparable number of *1/1 and *1/0 samples (Fig. 4B).

The expected difference in ΔC_t values between genotype clusters (the $\Delta\Delta C_t$ value) can be derived from the equation: fold change in gene copies $=2^{-\Delta\Delta Ct}$, where $\Delta\Delta C_t=\Delta C_{tunknown\ genotype}-\Delta C_{treference\ genotype}$, provided the amplification efficiency of both control and target gene is close to 100%. Thus, with 100% amplification efficiency, the expected $\Delta\Delta C_t$ value for *1/1 samples versus *1/0 (where fold change equals 2) is −1.00, and the expected $\Delta\Delta C_t$ value for *2/1 samples versus *1/0 samples (where fold change equals 3) is −1.58.

Finally, to determine the amount of misclassification, we repeated the genotyping of the first 1139 individuals from CCHS for both GST genes.

Validation

To validate the high-throughput genotyping method, we genotyped the first 40 individuals with the genotypes *GSTM1*0/0*, **1/0*, **1/1*, *GSTT1*0/0*, **1/0*, **1/1*, and all individuals with the genotypes *GSTM1*2/1* or *GSTT1*2/1* by long-range PCR as previously described [12,34].

Statistical analysis

To identify *1/0 samples at risk of being misclassified as *1/1 and vice versa, Rankit plots were constructed for each 384-well plate [35]. ΔC_t values of samples designated *1/1 were sorted descending and ΔC_t values of samples designated *1/0 were sorted ascending. The cumulated frequencies of the sorted samples were plotted versus ΔC_t value resulting in a V formed plot. The cumulated frequency was converted to Rankit values

by inverse normal transformation, and introduced on a new *y*-axis. Undetermined Rankit reruns were samples with ΔC_t values between the intersection points of the Rankit plots with the 2 SD limit (=Rankit −1.96). These samples were reamplified regardless of whether they could reliably be assigned a genotype using the ΔC_t method above.

Results

Requirements for multiplexing using the ΔC_t method

For simplicity only data obtained with Control 1 are shown. Similar results were obtained with Control 2.

Determining the linear dynamic range and defining limiting primer concentrations

The linear dynamic range for assays using non-limited primers in singleplex was at least 0.21–50 ng, and the amplification efficiency was close to 100% (95.6%–100.6%) for all three genes (Fig. 1).

In primer limitation studies in singleplex using a fixed DNA concentration of 3 ng per well, 1/2× RNaseP mix (450 nmol/L each primer, 125 nmol/L probe) increased the C_t value compared to 1× RNaseP mix (Fig. 2, Panel A). For *GSTM1* and *GSTT1*, C_t values increased for primer concentrations lower than 50 nmol/L (Fig. 2, Panels B and C). Based on these results we used 1X RNaseP mix (900 nmol/L each primer, 250 nmol/L probe), and 100 nmol/L of each GST primer as limiting primers and 250 nmol/L probe for *GSTM1* and *GSTT1* in all multiplex reactions.

Amplification efficiency and validation of the ΔC_t method

Fig. 3 (Panels A and C) shows examples of 1 of 10 standard curves for *GSTM1* and *GSTT1* run with limited primer concentrations, as determined above in multiplex with *RNaseP*. For both genes, reactions were performed in triplicate and the linear dynamic range was at least 0.41–100 ng per well; for comparison the amount of DNA in 2 μL 10× diluted DNA (the amount used for genotyping in this study) isolated as described is typically between 4 and 12 ng. For *GSTM1* and *RNaseP* in multiplex the amplification efficiencies were 100.3% and 101.8%, respectively (Panel A). The efficiencies were 98.4% and 100.4%, respectively, for *GSTT1* and *RNaseP* in multiplex (Panel C). The corresponding plots of ΔC_t values versus log DNA concentration are shown in Fig. 3, Panels B and D. The slope was −0.035 for *GSTM1* in multiplex, and −0.046 for *GSTT1* in multiplex, well within the +/−0.1 accepted for using the ΔC_t method.

Reproducibility

The reproducibility of the multiplexed assays was tested in ten consecutive runs for each target gene. The mean amplification efficiencies (+/−SD) of ten standard curves of C_t versus log DNA concentration were 100.5% (+/−1.22%) and 100.5% (+/−2.38%) for *GSTM1* and *RNaseP* in multiplex, and 99.9% (+/−1.34%) and 100.7% (+/−1.97%) for *GSTT1* and *RNaseP* in multiplex. The mean slopes (+/−SD) of ΔC_t versus log DNA concentration from 10 runs were 0.003 (+/−0.05) for *GSTM1*, and −0.018 (+/−0.05) for *GSTT1* in multiplex reactions well within the acceptable +/−0.1.

Taken together, primer limitation validated the use of 1× RNaseP mix and 100 nmol/L GST primers in multiplex reactions. The use of multiplexing was verified, as no reduction in amplification efficiency was observed for any assay when shifting from singleplex (Fig. 1) to multiplex reactions (Fig. 3, Panels A and C). High reproducibility and a slope of ΔC_t versus log DNA concentration close to zero in 10 consecutive runs validated the use of the ΔC_t method within a linear dynamic spanning the range of DNA concentrations expected in samples isolated by a standard commercial method (QIAamp DNA Mini Kit).

Genotyping of 20,687 individuals from the general population

To validate the performance of the genotyping method, the optimized assays were used to genotype 20,687 DNA samples from two large population studies, the Copenhagen City Heart Study and the Copenhagen General Population Study. ΔC_t values clearly segregated into four distinct groups, representing *0/0 (no amplification), *1/0, *1/1 and *2/1 samples (Fig. 4). The obtained $\Delta\Delta C_t$ values were very close to those expected assuming 100% amplification efficiency (see "Genotyping" in Methods): $\Delta\Delta C_t$ (±SD) for *1/1 versus *1/0 samples on single determinations was −0.89±0.18 for a typical plate for *GSTM1,* and −0.88±0.18 for a typical plate for *GSTT1* (expected $\Delta\Delta C_t$=−1.0, Fig. 4A); $\Delta\Delta C_t$ (±SD) for *2/1 versus *1/0 samples (n=16, the total number of *2/1 samples identified in the study versus n=16, randomly selected from the population sample) on single determinations was −1.54±0.13 for *GSTM1*, and −1.55±0.18 for *GSTT1* (n=27, the total number of *2/1 samples identified in the study versus n=28, randomly selected from the population sample; expected $\Delta\Delta C_t$=−1.58, Fig. 4B).

In order not to misclassify any *1/0 or *1/1 samples, with approximately 1 ΔC_t value in between, a Rankit plot was applied. Fig. 5 shows examples of Rankit plots for *GSTM1* (Panel A) and *GSTT1* (Panel B). As the Rankit plot transforms Gaussian distributed values into straight lines, Fig. 5 shows that the ΔC_t values for *1/0 and *1/1 samples were normally distributed. The samples at risk of misclassification are located below a Rankit value of −1.96 and between the two Rankit plots.

The genotyping results for the general population sample obtained before and after rerun are given in Table 2. The rerun frequency was approximately 6% for both assays and was equally distributed between genotypes. After rerun, 0.4–0.6% of samples could not be assigned a genotype. The genotype distribution for *GSTM1* was 52% *0/0, 40% *1/0 and 8% *1/1. Sixteen individuals (0.1%) had three *GSTM1* copies. For *GSTT1* the genotype distribution was 15% *0/0, 47% *1/0 and 38% *1/1 (Table 2). Twenty-seven individuals carried three *GSTT1* copies (0.1%). Genotype frequencies did not differ from those predicted by the Hardy–Weinberg equilibrium (χ^2: P (*GSTM1*)=0.15; P (*GSTT1*)=0.12).

When comparing the first and second round of genotyping of the first 1139 individuals, we found 0.2% discrepancy between

genotypes for *GSTT1* and no discrepancy for *GSTM1*. The true genotype of these samples was verified by long-range PCR.

Finally, genotyping by long-range PCR of 283 samples – 40 for each of the genotypes *0/0, *1/0 and *1/1 for both genes and all *2/1 samples (n=16 for *GSTM1* and n=27 for *GSTT1*; Table 2) – was in 100% agreement with the genotypes determined by high-throughput real-time PCR for the *0/0 and *1/0 genotypes. As expected, long-range PCR could not discriminate clearly between *1/1 and *2/1 genotypes, and typed both genotypes as *1/1, although the separation of these genotypes was quite clear using real-time PCR (Fig. 4B).

Discussion

This paper presents a real-time PCR method to determine CNV in *GSTM1* and *GSTT1* in large epidemiological samples with the endogenous control *RNaseP* as a reference. The principal findings are: 1) The method provides high sample throughput; 2) The linear dynamic range spans from 0.41 to 100 ng of DNA, as a minimum. As a standard purification from whole blood contains DNA at concentrations within this range, any given sample can be genotyped without the need for time-consuming concentration determinations; 3) By multiplexing target and endogenous control genes in 10 μL total volume in a 384-well format using single instead of multiple determinations of each sample, cost per sample was considerably reduced. 4) With single determinations, this method provides reproducible genotypes with less than 0.2% misclassification; 5) The method was used to genotype 20,687 individuals from the general population with less than 0.6% undetermined for both genes after rerun; 6) In contrast to long-range PCR, the method can discriminate *2/1 from *1/1 genotypes.

Discrimination between *GST*1/0* and *GST*1/1* individuals was previously performed with time-consuming long-range PCR unsuitable for high-throughput, partly because to determine the genotype it is necessary to run each individual DNA on an agarose gel following PCR [12,27,34,36]. Another problem with long-range PCR, as demonstrated in this study, is misclassification of *2/1 genotypes as *1/1, because the diagnosis in this case relies on differences in intensity of a single band on a gel. Genotyping has also been performed by expensive and time-consuming fluorescent-based fragment analysis [36]. Characteristically therefore, these methods have been used to genotype relatively few individuals (n=29–1032).

The first high-throughput methods to determine *GSTM1* and *GSTT1* CNV [17,25,26] used triple determinations and 96-well format, resulting in a throughput of approximately 384 genotype determinations per day. The new method presented here uses single determinations, 10 μL volume and 384-well formats, and includes a stack-holder that enables automatic loading of plates for analysis. With this method approximately 4600 genotypes can be determined per day. With the highly optimized assays, more than 99% were assigned a genotype in a clinical sample of 20,687 individuals. Using the method with lower throughput, due to individual determinations of DNA concentration prior to analysis and triplicate runs, Brasch-Andersen et al. [17] obtained 98% and 97% genotype assignment for *GSTM1* and *GSTT1*, respectively. In conclusion, despite single determinations, we obtain slightly better genotyping assignment than the previously described method, without the need for time-consuming DNA concentration determinations, and with much higher throughput.

The optimized assays were highly sensitive in discriminating between *GST*1/0* and *GST*1/1* samples which segregated into two separate groups with no overlap in between. In the paper by Brasch-Andersen et al. [17] an overlap between *GST*1/0* and *GST*1/1* samples resulted in approximately 2% and 1% misclassification of *GSTM1* and *GSTT1* samples, respectively. In the present study, we used a very conservative approach and reanalyzed all samples that were within upper and lower 2 SD limits for *GST*1/1* and *GST*1/0* samples on Rankit plots, respectively, although no overlap occurred between the two genotypes.

The *GST*0* allele is thought to result from homologous unequal crossing over between two highly identical sequences flanking the GST gene [11,12]. The same mechanism can result in gene duplication (two gene copies on the same chromosome) which has been reported for *GSTM1* [13], but to our knowledge never previously for *GSTT1*. The high sensitivity of the real-time method allowed identification of individuals with three *GSTM1* gene copies and three *GSTT1* gene copies. This is not possible with the long-range PCR method which does not discriminate properly between one or more gene copies on the same allele. No samples with three gene copies were identified in any of the former studies using real-time PCR [17,25], most likely due to the small sample sizes tested. Finally, in validation experiments there was complete concordance between the *0/0 and *1/0 genotypes measured by real-time PCR and validated by long-range PCR, but as expected all 43 *2/1 genotypes were misclassified as *1/1 using long-range PCR, due to the inability of this method to discriminate between one or more gene copies when present on the same allele.

A gene dosage effect between gene copy number and enzyme activity has been reported for both *GSTM1* [13] and *GSTT1* [12]. Therefore, as GSTs are involved in detoxification of for example carcinogens and environmental toxins, there are potential links between CNV in these genes and the development of chronic diseases. In a review by Bolt and Thier [37], the connection between *GSTM1* and *GSTT1* genotypes and chronic diseases in humans was presented. Bladder cancer is to a large extent caused by smoking and the risk associated with the null genotype of *GSTM1* was approximately 1.5 in two large meta-analyses [21,22], and a possible gene dosage effect was suggested [22]. Although also lung cancer is to a large extent caused by smoking and environmental factors potentially detoxified by GSTs, the risk associated with the null genotypes of *GSTM1* and *GSTT1* was only marginally above 1 in a large scale meta-analysis including >20,000 cases [20]. However, these latter studies have not discriminated between *1/0 and *1/1 genotypes and *1/1 and *2/1 genotypes due to lack of appropriate methods, and possible gene dosage effects are very likely to have weakened the associations. High-throughput CNV assays are thus in high demand.

In conclusion, the optimized ΔC_t method provides precise, low cost and high-throughput determination of CNV in *GSTM1*

and *GSTT1*, and is anticipated to be a strong tool for the determination of gene-dosage effect of *GSTM1* and *GSTT1* on risk of chronic diseases in large populations. The design, optimization, and validation approach described here can be used to determine CNV in other genes as well.

Acknowledgments

We thank Mette Refstrup for her persistent attention to the details of the large-scale genotyping. We are indebted to the staff and participants of The Copenhagen City Heart Study and The Copenhagen General Population Study for their important contributions.

References

[1] Katsouyanni K, Pershagen G. Ambient air pollution exposure and cancer. Cancer Causes Control 1997;8:284–91.
[2] Nyberg F, Gustavsson P, Jarup L, et al. Urban air pollution and lung cancer in Stockholm. Epidemiology 2000;11:487–95.
[3] Brennan P, Bogillot O, Cordier S, et al. Cigarette smoking and bladder cancer in men: a pooled analysis of 11 case-control studies. Int J Cancer 2000;86:289–94.
[4] Dockery DW, Pope CA, Xu X, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med 1993;329:1753–9.
[5] Peters A, Dockery DW, Muller JE, Mittleman MA. Increased particulate air pollution and the triggering of myocardial infarction. Circulation 2001; 103:2810–5.
[6] Czernin J, Waldherr C. Cigarette smoking and coronary blood flow. Prog Cardiovasc Dis 2003;45:395–404.
[7] Peters A, von Klot S, Heier M, et al. Exposure to traffic and the onset of myocardial infarction. N Engl J Med 2004;351:1721–30.
[8] Salam MT, Li YF, Langholz B, Gilliland FD. Early-life environmental risk factors for asthma: findings from the Children's Health Study. Environ Health Perspect 2004;112:760–5.
[9] Radon K, Busching K, Heinrich J, et al. Passive smoking exposure: a risk factor for chronic bronchitis and asthma in adults? Chest 2002;122: 1086–90.
[10] Hayes JD, Strange RC. Glutathione S-transferase polymorphisms and their biological consequences. Pharmacology 2000;61:154–66.
[11] Xu S, Wang Y, Roe B, Pearson WR. Characterization of the human class Mu glutathione S-transferase gene cluster and the GSTM1 deletion. J Biol Chem 1998;273:3517–27.
[12] Sprenger R, Schlagenhaufer R, Kerb R, et al. Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype–phenotype correlation. Pharmacogenetics 2000;10:557–65.
[13] McLellan RA, Oscarson M, Alexandrie AK, et al. Characterization of a human glutathione S-transferase mu cluster containing a duplicated GSTM1 gene that causes ultrarapid enzyme activity. Mol Pharmacol 1997;52:958–65.
[14] Seidegard J, Vorachek WR, Pero RW, Pearson WR. Hereditary differences in the expression of the human glutathione transferase active on trans-stilbene oxide are due to a gene deletion. Proc Natl Acad Sci U S A 1988;85:7293–7.
[15] Pemble S, Schroeder KR, Spencer SR, et al. Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. Biochem J 1994;300:271–6.
[16] Masetti S, Botto N, Manfredi S, et al. Interactive effect of the glutathione S-transferase genes and cigarette smoking on occurrence and severity of coronary artery risk. J Mol Med 2003;81:488–94.
[17] Brasch-Andersen C, Christiansen L, Tan Q, Haagerup A, Vestbo J, Kruse TA. Possible gene dosage effect of glutathione-S-transferases on atopic asthma: using real-time PCR for quantification of GSTM1 and GSTT1 gene copy numbers. Hum Mutat 2004;24:208–14.
[18] Ivaschenko TE, Sideleva OG, Baranov VS. Glutathione- S-transferase micro and theta gene polymorphisms as new risk factors of atopic bronchial asthma. J Mol Med 2002;80:39–43.
[19] Rebbeck TR. Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. Cancer Epidemiol Biomark Prev 1997;6:733–43.
[20] Ye Z, Song H, Higgins JP, Pharoah P, Danesh J. Five glutathione s-transferase gene variants in 23,452 cases of lung cancer and 30,397 controls: meta-analysis of 130 studies. PLoS Med 2006;3:524–34.
[21] Engel LS, Taioli E, Pfeiffer R, et al. Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: a HuGE review. Am J Epidemiol 2002;156:95–109.
[22] Garcia-Closas M, Malats N, Silverman D, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. Lancet 2005;366:649–59.
[23] Gilliland FD, Li YF, Saxon A, Diaz-Sanchez D. Effect of glutathione-S-transferase M1 and P1 genotypes on xenobiotic enhancement of allergic responses: randomised, placebo-controlled crossover study. Lancet 2004; 363:119–25.
[24] Garte S, Gaspari L, Alexandrie AK, et al. Metabolic gene polymorphism frequencies in control populations. Cancer Epidemiol Biomark Prev 2001; 10:1239–48.
[25] Girault I, Lidereau R, Bieche I. Trimodal GSTT1 and GSTM1 genotyping assay by real-time PCR. Int J Biol Markers 2005;20:81–6.
[26] Bediaga NG, Alfonso-Sanchez MA, de Renobales M, Rocandio AM, Arroyo M, de Pancorbo MM. GSTT1 and GSTM1 gene copy number analysis in paraffin-embedded tissue using quantitative real-time PCR. Anal Biochem 2008;378:221–3.
[27] Barnette P, Scholl R, Blandford M, et al. High-throughput detection of glutathione s-transferase polymorphic alleles in a pediatric cancer population. Cancer Epidemiol Biomark Prev 2004;13:304–13.
[28] Applied Biosystems, User Bulletin #5: ABI Prism 7700 Sequence Detection System 2005; http://www.appliedbiosystems.com (Accessed September 2007).
[29] Applied Biosystems, Amplification Efficiency of TaqMan® Gene Expression Assays: Application Note 2005; http://www.appliedbiosystems.com (Accessed September 2007).
[30] Applied Biosystems, Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR 2005; http://www.appliedbiosystems.com (Accessed September 2007).
[31] Appleyard M, Hansen AT, Jensen G, Schnohr P, Nyboe J, The Copenhagen CIty Heart Study. Østerbroundersøgelsen. A book of tables with data from the first examination (1976–78) and a five year follow-up (1981–83) The Copenhagen City Heart Study Group. Scand J Soc Med 1989;41 (Suppl):1–160.
[32] Schnohr P, Jensen GB, Lange P, Scharling H, Appleyard M. The Copenhagen City Heart Study – Østerbroundersøgelsen, Tables with Data from the third examination 1991–1994. Eur Heart J 2001;3(Suppl H):H1–H83.
[33] Nordestgaard BG, Benn M, Schnohr P, Tybjaerg-Hansen A. Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. JAMA 2007;298:299–308.
[34] Buchard A, Sanchez JJ, Dalhoff K, Morling N. Multiplex PCR detection of GSTM1, GSTT1, and GSTP1 gene variants: simultaneously detecting GSTM1 and GSTT1 gene copy number and the allelic status of the GSTP1 Ile105Val genetic variant. J Mol Diagnostics 2007;9:612–7.
[35] Hyltoft PP, Blaabjerg O, Andersen M, Jorgensen LG, Schousboe K, Jensen E. Graphical interpretation of confidence curves in rankit plots. Clin Chem Lab Med 2004;42:715–24.
[36] Moyer AM, Salavaggione OE, Hebbring SJ, et al. Glutathione S-transferase T1 and M1: gene sequence variation and functional genomics. Clin Cancer Res 2007;13:7207–16.
[37] Bolt HM, Thier R. Relevance of the deletion polymorphisms of the glutathione S-transferases GSTT1 and GSTM1 in pharmacology and toxicology. Curr Drug Metab 2006;7:613–28.