

### Association between CYP2E1 polymorphisms and risk of differentiated thyroid carcinoma

Lucia Pellé 1, Monica Cipollini 1, Roman Tremmel 2,3 , Cristina Romei 4, Gisella Figlioli 1, Federica Gemignani 1, Ombretta Melaiu 1, Chiara De Santi 1, Elisa Barone 1, Rossella Elisei 4, Eric Seiser 5, Federico Innocenti 5, Ulrich M. Zanger 2,3 , Stefano Landi 1

1.Department of Biology University of Pisa, Pisa, Italy; 2.Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany; 3.University of Tuebingen, Tuebingen, Germany; 4.Department of Endocrinology and Metabolism, Orthopedics and Traumatology, Occupational Medicine, University of Pisa, Pisa, Italy; 5.University of North Carolina-Chapel Hill, Chapel Hill, USA

---

#### Abstract

Differentiated thyroid carcinoma (DTC) results from complex interactions between genetic and environmental factors. Known etiological factors include exposure to ionizing radiations, previous thyroid diseases, and hormone factors. It has been speculated that dietary acrylamide (AA) formed in diverse foods following the Maillard's reaction could be a contributing factor for DTC in humans. Upon absorption, AA is biotransformed mainly by cytochrome P450 2E1 (CYP2E1) to glycidamide (GA). Considering that polymorphisms within CYP2E1 were found associated with endogenous levels of AA-Valine and GA-Valine hemoglobin adducts in humans, we raised the hypothesis that specific CYP2E1 genotypes could be associated with the risk of DTC. Analysis of four haplotype tagging SNPs (ht-SNPs) within the locus in a discovery case-control study (N = 350/350) indicated an association between rs2480258 and DTC risk. This ht-SNP resides within a linkage disequilibrium block spanning intron VIII and the 3'-untranslated region. Extended analysis in a large replication set (2429 controls and 767 cases) confirmed the association, with odds ratios for GA and AA genotypes of 1.24 (95 % confidence interval (CI) 1.03–1.48) and 1.56 (95 % CI, 1.06–2.30), respectively. Functionally, the minor allele was associated with low levels of CYP2E1 mRNA and protein expression as well as lower enzymatic activity in a series of 149 human liver samples. Our data support the hypothesis that inter-individual differences in CYP2E1 activity could modulate the risk of developing DTC suggesting that the exposure to specific xenobiotics, such as AA, could play a role in this process.

#### Keywords

Differentiated thyroid carcinoma Case-control association study Polymorphisms Cytochrome P450 2E1 Genetic susceptibility

## Introduction

Thyroid cancer constitutes approximately 1 % of all human cancer and is the most common endocrine malignancy (Sipos and Mazzaferri [2010](#)). The papillary (PTC) and the follicular thyroid carcinoma (FTC), overall defined as differentiated thyroid carcinoma (DTC), cover about 95 % of all types of thyroid neoplasms (Grande et al. [2012](#)). The main etiological factors responsible for DTC are inheritance (Malchoff and Malchoff [2004](#)), exposure to ionizing radiations (Ron [2007](#)), insufficient iodine intake (Cardis et al. [2005](#)), previous thyroid diseases such as goiter, and hormonal factors (Surks et al. [1990](#)).

Fioretti and collaborators showed that a history of benign thyroid disease accounted for 18.9 % of cases, radiotherapy for 1.2 %, residence for  $\geq 20$  years in endemic goiter areas for 2.4 % of cases, and their combination for 21.7 % of thyroid cancer cases; selected indicators of a poor diet accounted for 40.9 % of thyroid cancer cases in this population (Fioretti et al. [1999](#)). The combination of all factors considered explained over 57 % of thyroid cancer cases in both sexes. However, other environmental factors are suspected to play a role, because an increasing incidence is observed in recent years. A possible role of acrylamide (AA) for thyroid carcinogenesis has been debated since the first observations in animals (Johnson et al. [1986](#)). Acute in vivo treatments of rats with AA resulted in morphological changes of thyroid follicular cells, thyroid stimulation, increased levels of serum T4, decreased TSH levels (Khan et al. [1999](#)), increased rates of DTC in carcinogenicity assays (Friedman et al. [1995](#); Beland et al. [2013](#)), and even increased rates of FTC (Maronpot et al. [2015](#)).

Humans are exposed to AA through their diet. The formation of AA in food occurs via the Maillard's reaction between amino acids and sugars at high temperatures during browning and frying (Tareke et al. [2002](#); Mottram et al. [2002](#); Stadler et al. [2002](#); Yaylayan et al. [2003](#); Zyzak et al. [2003](#)). The exposure of humans to AA should not be underestimated, representing a real concern for public health (Mottram et al. [2002](#); Stadler et al. [2002](#); Mei et al. [2008](#)). In fact, potato chips and French fries contain elevated concentrations of AA (240  $\mu\text{g}/\text{kg}$ ). However, significant exposure to AA comes also from bread (9–24  $\mu\text{g}/\text{kg}$ ), coffee (200–520  $\mu\text{g}/\text{kg}$ ), and other foods, including cooked meat (82–84  $\mu\text{g}/\text{kg}$ ) (EFSA [2012](#); Friedman [2003](#); Freisling et al. [2013](#)), resulting in an average intake of 0.31–1.1  $\mu\text{g}/\text{kg}$  body weight per day (EFSA [2011](#); Pedreschi et al. [2014](#)). A positive association between AA dietary intake and risk of endometrial (Hogervorst et al. [2007](#); Obón-Santacana et al. [2014](#)), ovarian (Hogervorst et al. [2013](#)), and renal cancers (Pelucchi et al. [2011](#)) was found in humans. An increased risk was also reported for DTC in a French study (Colonna et al. [2010](#)). After a revision of published studies, the International Agency for Research on Cancer (IARC) classified AA as a probable human carcinogen (group 2A) (Iarc [1994](#)).

It was shown that upon absorption, AA is metabolically activated mainly by CYP2E1 to glycidamide (GA) (Sumner et al. [1999](#); Ghanayem et al. [2005](#); Settels et al. [2008](#)). In humans, it was shown that inhibition of CYP2E1 profoundly affects the hematologic and urinary GA/AA ratio in in vivo experiments

(Doroshenko et al. [2009](#)). CYP2E1 is a polymorphic enzyme and *CYP2E1* gene variants are related to the activity of the encoded enzyme, although systematic investigations are missing (Neafsey et al. [2009](#); Zanger and Schwab [2013](#); <http://www.cypalleles.ki.se/>). Moreover, in a recent work, Huang et al. ([2012](#)) found that genotypes of *CYP2E1* contribute to the inter-individual variability in endogenous levels of AA-Valine and GA-Valine hemoglobin adducts (Huang et al. [2012](#)).

Following all these considerations, we hypothesized that exposure to AA could increase the risk of developing DTC and that *CYP2E1* polymorphisms may modulate this risk by metabolic interference. For this reason, in the present work, we performed a case–control association study on DTC by genotyping haplotype tagging single nucleotide polymorphisms (ht-SNPs) within *CYP2E1*. Corroborating our hypothesis, we found a statistically significant association between a SNP tagging the linkage disequilibrium (LD) block at the 3' end of the gene and risk of DTC. Moreover, using human liver tissues we showed that variants in this genetic region affect the expression and function of the enzyme. Our findings extend the current knowledge about *CYP2E1* functional polymorphisms, in particular within the 3' region, and their potential involvement in cancer risk.

## **Materials and methods**

### Case–control study

DTC patients were volunteers followed up at the Department of Endocrinology of the University Hospital of Pisa (Italy). Details of the ongoing study were already given elsewhere (Cancemi et al. [2011](#)). Briefly, controls were blood donors collected at Meyer Hospital of Florence (Italy) and healthy subjects working at the University Hospital of Pisa recruited during their routine visits in the context of a program of surveillance performed by the Occupational Medicine. The eligibility criteria were the same for cases and controls, and it included a minimal age of 18 years. As exclusion criteria, volunteers should not have been affected by any malignancy, chronic inflammatory disease, or related diseases in the past. A further exclusion criterion for controls was the presence of thyroid nodules, when known. The interview of both cases and controls was carried out via self-administered questionnaires during blood withdrawal. For each subject, the following data were collected: ethnicity, gender, age, height, body weight, birthplace and years of residence in the birthplace, place of residence, smoking habits, smoking history, and any positive family history of thyroid disease. In the present study, 1140 consecutive cases and 2800 controls were analyzed in two independent sample sets. The first (the “discovery set,” DS) consisted of 350 cases and 350 controls, while the remaining set of volunteers were included as “replication set” (RS). Peripheral blood samples, otherwise disposed following the analyses of clinical routine, were used for genetic analyses. The study was approved by the ethics committee of the University of Pisa and conducted in accordance with the Helsinki Declaration. All volunteers gave written informed consent to participate in the study and to allow the use of their biological samples for research purposes.

### SNP selection

*CYP2E1* is localized on chromosome 10 from 133,527 kb to 133,539 kb (Build GRCh38: December 2013, hg38). Three blocks of LD were identified in Caucasians with the tool Haploview (Barrett et al. [2005](#)) and the method of confidence intervals (Gabriel et al. [2002](#)). Four ht-SNPs were chosen with MAF (minor allele frequency) >10 % using the tagger function (SNP ID in the dbSNP at <http://www.ncbi.nlm.nih.gov/snp/> and HGVS, the Human Genome Variation Society, names are given below): rs9418982 (NC\_000010.10:g.133507118C>T) belongs to the first block of LD, within the promoter region (133,507,118–133,527,063 bp), and it is in high LD with the polymorphisms PstI and RsaI, repeatedly studied in previous works and known as \*5A and \*5B alleles; rs7092584 (NC\_000010.10:g.133,534,753 C>T) and rs743534 (NC\_000010.10:g. 133,535,722 C>A) are located in the second block (133,531,207–133,535,722 bp); rs2480258 (NC\_000010.10:g. 133,538,596 T>C) is located in the third block (133,535,863–133,546,944 bp).

#### DNA extraction and SNP genotyping

DNA extraction was performed on peripheral blood using the PureGene Blood Kit (Gentra Systems, Inc., Minneapolis, MN). Samples from cases and controls were randomized and mixed on PCR plates, so that an equal number of cases and controls could be analyzed simultaneously by personnel blinded for the case/control status. Genotyping for rs9418982, rs7092584, rs743534, and rs2480258 was carried out using pre-designed TaqMan® SNP Genotyping Assays, according to protocol specified by the manufacturer (Life Technologies Inc., Grand Island, New York, USA). Four percent of DNA samples were repeated as quality control.

Genotypes of rs9418982, rs7092584, rs743534, and rs2480258 were in agreement with Hardy–Weinberg equilibrium. The quality control of genotypes was satisfactory (repetitions gave >99.5 % of identities).

#### Human liver samples

Liver tissues and corresponding blood samples for genomic DNA extraction were previously collected from Caucasian patients undergoing liver surgery at the Department of General, Visceral, and Transplantation Surgery at the Charité (Campus Virchow, University Medical Center Charité, Humboldt University Berlin, Germany). The use of human liver tissues was approved by the local ethical committees the Charité, Humboldt University (Berlin, Germany), and the University Clinic Tuebingen, Germany. Written informed consent was obtained from all liver donors, and the study was conducted in accordance with the Declaration of Helsinki. All tissue samples had been examined by a pathologist, and only histologically normal liver tissue was collected and stored at –80 °C. For this study, we used 149 liver samples for which the following information was available (Klein et al. [2010](#)): gender [male ( $n = 71$ ) vs. female ( $n = 78$ )], age [median age = 59 years], smoking habits [smokers ( $n = 29$ ) vs. non-smokers ( $n = 116$ )], alcohol consumption [yes ( $n = 48$ ) vs. no ( $n = 95$ )], presurgical drug exposure [yes ( $n = 109$ ) vs. no ( $n = 40$ )], cholestasis [yes ( $n = 25$ ) vs. no ( $n = 120$ )], CRP [elevated levels ( $n = 7$ ) vs. normal levels ( $n = 139$ )] and the diagnosis leading to liver resection [primary liver tumors ( $n = 66$ ) vs. metastases

( $n = 82$ )). Sample numbers do not always sum up to 149 because of missing information. The genotypes for the SNP rs2515642 were extracted from SNP microarray analysis (Schröder et al. [2013](#)). We used the SNP microarray data to impute the genotypes of rs2480258 based on the CEU HapMap3 and 1000G reference panel (release June 2010, b36) using SHAPEIT v2 and IMPUTE v2.3.1 pipeline (Howie et al. [2009](#), Delaneau et al. [2012](#)). Prior to imputation, low-quality variants (MAF <1 %, call rate <95 % and HWE  $P$ value < $10^{-6}$ ) were filtered out. The software was run using default options, and for each liver sample the corresponding genotype for rs2480258 was assigned to the highest maximum posterior probability output except that for a value <0.4 a missing genotype was assigned instead. For the SNP rs2480258 average certainty of best-guess genotype and info score, which were referred to as imputation quality scores, was 0.99 for both metrics.

#### Hepatic CYP2E1 expression phenotypes

High-quality total RNA was isolated from liver tissue using Trizol<sup>®</sup> (Invitrogen, CA, USA)/Qiagen RNeasy<sup>®</sup> protocol as described previously (Gomes et al. [2009](#)). Expression of CYP2E1 mRNA was quantified using pre-developed TaqMan Gene Expression Assay and a TaqMan 7500 system (Applied Biosystems, Darmstadt, Germany) as described (Rieger et al. [2013](#)). CYP2E1 protein expression was analyzed by Western blot in liver microsomes (Tremmel et al. [2015](#)). CYP2E1 enzyme activity was quantified by measuring liver microsomal chlorzoxazone 6-hydroxylation by LC-MSMS analysis as described (Gomes et al. [2009](#)).

#### In silico analyses

For the analysis of LD in cases and controls, HapMap project (Data release 28, Phase II + III, August 10, on NCBI B36 assembly, dbSNP b126; <http://www.hapmap.org>) was browsed. For intronic SNPs, the effect on splicing was evaluated with two online prediction tools. The first, “Splice Site Prediction by Neural Network,” is a software developed within Berkeley Drosophila Genome Project (freely available at the URL: [http://www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html); Reese et al. [1997](#)), whereas “SpliceAid” is a database of strictly experimentally assessed target RNA sequences in humans (available at the URL: <http://www.introni.it/splicing.html>; Piva et al. [2009](#)). Moreover, in order to estimate the effect of SNPs falling within miRNA binding sites (the so-called miRSNPs) of mRNAs, the tools PolymiRTS Database 3.0 (available at the URL <http://compbio.uthsc.edu/miRSNP>; Ziebarth et al. [2012](#); Bhattacharya et al. [2014](#)), RNAhybrid (available at the URL: <http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/>; Rehmsmeier et al. [2004](#)), and MicroSnipers (available at the URL: <http://epicenter.ie-freiburg.mpg.de/services/microsniper/>) were used with modalities employed in previous studies (Naccarati et al. [2012](#)).

#### Statistical analyses

Hardy–Weinberg equilibrium in controls was tested for each polymorphism by the Chi-squared test ( $\chi^2$ ) test (1 degree of freedom). A logistic regression analysis was used to examine the associations between

the genotypes and the considered risk factors. The association analyses were based on the estimation of the odds ratios (ORs) and of their 95 % confidence intervals (CIs). Genotypes were analyzed with a multivariate logistic regression (MLR) model, allowing the ORs to be adjusted for covariates as linear variables (the OR<sub>adj</sub>). Each genotype category was compared using the common homozygotes as reference category, and the *P* value of the association (*P*<sub>ass</sub>) was calculated separately for heterozygotes and homozygotes. In the additive model the SNP was coded as the number of risk alleles: 0 = homozygous reference, 1 = heterozygous, 2 = homozygous variant, and the MLR was performed providing the per-allele risk. Moreover, the MLR analysis was performed on the dominant or recessive models according to the most likely mode of inheritance, evaluated with an extended MAX test (Freidlin et al. [2002](#)). Finally, the total number of alleles was analyzed by counting the total number of risk and reference alleles among cases and controls in a 2 × 2 table. The correction of probabilities for multiple testing was performed according to the calculation of the false discovery rate following the Benjamini and Hochberg procedure (Benjamini et al. [2001](#)). The software Statgraphics Centurion software (StatPoint Inc., USA) was used.

Statistical analysis of the liver samples was performed using the software R 3.0.1 (R Core Team [2014](#)) with the additional package SNPAssoc-1.9-2 (González et al. [2007](#)). The influence of the polymorphisms on CYP2E1 mRNA, protein, and activity phenotypes was investigated using pairwise *t* tests. Multivariate association analysis between the SNP rs2480258 under three different genetic models (additive, dominant, and recessive) and each CYP2E1 phenotype measurement, corrected for the eight non-genetic factors, was performed using the generalized linear model function of the SNPAssoc package. For all analyses CYP2E1 phenotype measurements were power-transformed (mRNA:  $\lambda = 0.45$ ; protein:  $\lambda = 0.3$ ) or log-transformed (CYP2E1 enzyme activity) to fulfill Gaussian distribution assumption which was confirmed using normal quantile–quantile plots and Shapiro–Wilk tests (all *P* values >0.57). The transformed phenotypes showed homogeneity of variances which was verified using Levene’s tests (all *P* values >0.4). The statistical threshold of significance was 5 %. All statistical tests were two-sided.

## Results

### General patient’s characteristics and discovery set analysis

The characteristics of the populations employed in the DS and in the RS are reported in Table [1](#). Females were more frequent among cases than controls (OR = 3.28, 95 % CI = 2.39–4.50; OR = 2.97, 95 % CI = 2.55–3.46, in DS and RS, respectively), with a sex ratio typical of the disease. Moreover, it should be also noted that cases had a slightly lower age than controls. Finally, a positive association was also observed between smoking habit and risk of DTC with an OR of 1.37 (95 % CI = 1.18–1.57) in the RS.

The OR<sub>adj</sub>, the 95 % CIs, and their *P*<sub>ass</sub> following MLR analysis of the additive and the best inheritance models are reported in Table [2](#) for the DS. After analysis of 350 cases and 350 controls, no association

was observed between rs9418982, rs7092584, rs743534, and risk of DTC. On the other hand, uncommon homozygotes for rs2480258 showed a statistically significantly increased risk (OR = 2.63, 95 % CI = 1.15–6.03;  $P_{\text{ass}} = 0.02$ ). The Cochran–Armitage trend test suggested the recessive model as the best genetic model of inheritance ( $P_{\text{trend}} = 0.02$ ). Under this model, the AA-genotype was related to an increased predisposition to DTC (OR = 2.57, 95 % CI = 1.14–5.83;  $P_{\text{ass}} = 0.02$ ; Table 2).

Considering the low number of homozygotes found in DS, a RS was used and 767 DTC cases and 2429 controls were genotyped (Table 3). The A-allele of rs2480258 was associated in a statistically significant way with the risk (OR = 1.17, 95 % CI = 1.02–1.33;  $P_{\text{ass}} = 0.02$ ). Under the hypothesis of the additive model ( $P_{\text{trend}} = 0.023$ ), heterozygotes and homozygotes showed a trend of increased risk (OR = 1.24, 95 % CI = 1.03–1.48,  $P_{\text{ass}} = 0.02$  and OR = 1.56, 95 % CI = 1.06–2.30,  $P_{\text{ass}} = 0.02$ , respectively). When the groups of DS and RS were pooled (1117 DTC and 2779 controls) the results were confirmed, reinforcing the statistical power of the study. The per-allele risk (A-allele vs. G-allele) was OR = 1.17 (95 % CI = 1.04–1.31;  $P_{\text{ass}} = 0.008$ ), and heterozygotes and homozygotes showed an OR = 1.20 (95 % CI = 1.02–1.40;  $P_{\text{ass}} = 0.02$ ) and OR = 1.68 (95 % CI = 1.20–2.35;  $P_{\text{ass}} = 0.002$ ), respectively. The per-allele and rare homozygotes statistical significance held also following corrections for multiple testing ( $q$ -values of 0.04 and 0.01, respectively).

#### Functional assessment of CYP2E1 polymorphisms in human liver

Following these positive results, we performed a genotype–phenotype correlation analysis in 149 human liver samples with phenotypic data for the mRNA expression, protein levels, and enzymatic activity of microsomal CYP2E1. We found that rs2515642, a SNP with available genotype information assessed by microarray (Schröder et al. 2013) and in strong LD with rs2480258 ( $r^2 = 0.96$ ), was significantly associated with a reduced CYP2E1 expression phenotype, at the level of mRNA, protein, and enzymatic activity (Fig. 1; upper panel). When rs2480258 genotypes were imputed, we observed similar genotype frequencies (GG: 68 %; GA: 29 %; AA: 3 %) in human liver subjects compared to the frequencies in the DS and RS cohorts. The association was confirmed, in that compared to the homozygous G-allele, the A-allele was significantly associated with lower levels of mRNA (GA: 0.83-fold,  $P = 0.04$ ; AA: 0.33-fold,  $P = 0.002$ ), protein (GA: 0.74-fold,  $P =$  not significant; AA: 0.47-fold,  $P = 0.03$ ), and chlorzoxazone 6-hydroxylation (GA: 0.79-fold,  $P = 0.02$ ; AA: 0.8-fold,  $P =$  not significant; Fig. 1; lower panel). Multivariate analysis using the same three genetic models and correction for covariates (sex, age, nicotine and alcohol consumption, presurgical drug exposure, C-reactive protein, cholestasis, and diagnosis leading to liver resection) confirmed the significant associations of the A-allele with lower mRNA (additive model;  $P = 0.01$ ), lower protein levels (additive model;  $P = 0.01$ ), and decreased enzyme activity (additive model;  $P = 0.04$ ).

As shown in Table 4, only two main haplotypes, #1 and #2, are prevalent among Caucasians in the third LD block, marked by the low-expression allele rs2480258, with frequencies of 0.759 and 0.234, respectively. Within this block the SNPs rs2515642, rs2480259, rs2480258, rs2249694, and rs2249695 were evaluated for altering the splicing efficiency of the intron VIII. However, no differences were noted (data not reported for brevity). The SNPs located within the 3'UTR (rs24080257 and rs2480256) were evaluated as putative miRSNPs (i.e., polymorphisms located within miRNA-binding target sites), and the A-allele of rs2480257 was predicted to affect a conserved binding site for hsa-miR-5582-3p, hsa-miR-570-3p, and hsa-miR-203a. On the other hand, the A-allele of rs2480256 was predicted to affect two miRNA binding sites: one for hsa-miR-510-3p and another for hsa-miR-570-3p (results presented in supplementary Table 1). Thus, hsa-miR-570-3p was predicted to be affected by both polymorphisms.

## Discussion

Over a hundred case-control association studies were published on SNPs within *CYP2E1* in relation to the risk of various diseases. As showed in Fig. 2 (and supplementary Table 2), most of them focused on the alleles \*5A, \*5B, or \*6 (i.e., rs2031920, RsaI, and rs3813867, PstI, at the 5' region, or rs6413432, DraI, within intron VI, respectively).

To our knowledge, this is the first study to evaluate multiple regions of the *CYP2E1* locus in relation to the risk of DTC.

A novel association for rs2480258, a ht-SNP within a LD block spanning intron VIII and the 3' UTR of the gene, was found, in that the minor allele was associated with increased risk for DTC (Table 3). Interestingly, we found that the minor allele of rs2480258 was also associated with a reduced *CYP2E1* expression phenotype at three levels, i.e., mRNA, protein, and *CYP2E1* enzyme activity, in a panel of 149 liver tissues. In order to better understand the biology of *CYP2E1*, all the intron VIII and 3'UTR SNPs of the LD block marked by rs2480258 were evaluated *in silico* for predicting their biological effects. While no consequences were noted for the splicing of intron VIII, the two haplotypes showed differences in regard to the binding of four microRNAs (miRNAs): hsa-miR-5582-3p, hsa-miR-570-3p, hsa-miR-203a, and hsa-miR-510-3p. It could thus be speculated that carriers of haplotype #2 have an altered miRNA-dependent *CYP2E1* regulation, leading to reduced *CYP2E1* protein levels and activity (i.e., a "slow metabolizer" phenotype). Although limited information is available for the predicted miRNAs hsa-miR-203a, which was found down-regulated in central nervous system (Gaur et al. 2007) and up-regulated in breast (Iorio et al. 2005), and colorectal cancer (Bandrés et al. 2006), varied expression was observed for hsa-miR-570-3p in hepatoma-HepG2, osteosarcoma-U2OS, lung adenocarcinoma A549-I, renal carcinoma-DH, neuroblastoma SHSYSY, and Burkitt-DG-75 cancer cell lines (<http://www.microrna.org>). No robust information is available for thyroid carcinoma. Of interest and in agreement with our *in silico* prediction, miR-570 was recently shown to be expressed in healthy human liver and to negatively

regulate CYP2E1 protein expression by translational repression in cell culture and tissue depending on the genotype for the 3'UTR SNPs rs2480256 and rs2480257 (which are linked to the DTC risk allele rs2480258, see above) (Nakano et al. [2015](#)). We acknowledge that our results seem to be counterintuitive with respect to the starting hypothesis. Thus, metabolic activation of AA to GA by CYP2E1 should be reduced in individuals with slow metabolizer phenotype, e.g., carriers of haplotype #2, resulting in decreased cancer risk, if GA is considered to be the ultimate carcinogen. As the present findings are in contrast to this expectation, they may point at a more differentiated role of dietary AA in the development of DTC. In slow metabolizers, circulating levels of AA should be elevated for prolonged times, so that target organs such as the thyroid may be reached more easily, where conversion into GA may take place in organo. However, available data do not support substantial expression of CYP2E1 in the thyroid. Alternatively, AA itself could exert its toxic effects directly. For example, AA was found to be clastogenic and toxic for neurological and reproductive systems (Besaratina and Pfeifer [2003](#); Koyama et al. [2006](#)). AA may also act indirectly as an endocrine disruptor, as it was found to be associated with various sex hormone levels (Hogervorst et al. [2013](#)). The finding that tissue and organ distribution of GA-DNA adducts does not correlate with AA-induced tumors in rat organs and reinforces the role of a direct effect of AA (Segeberäck et al. [1995](#)).

Of course we also acknowledge that there are some limitations. For example, information collected for diet and smoking habits was limited and cases and controls were not frequency-matched for gender or age, although these covariates were taken into account by the statistical model allowing the adjustment of ORs. However, we believe that the association described in this study is not a chance finding. In fact, the study had an adequate statistical power to detect small effect sizes, given the large sample size of cases and controls genotyped. In addition, the results were replicated in two independent sample sets (DS and RS).

Furthermore, the significance of CYP2E1 mRNA for functional phenotype remains mysterious. While miR-570 offers a mechanistic explanation for the lower expression of the rs2480258-tagged DTC risk allele on the protein level, there exist inconsistent data regarding the mRNA level. The study of Nakano et al. ([2015](#)) as well as preliminary results from additional eQTL datasets seems not to replicate the effect of the A-allele of rs2480258 on mRNA expression (Eric Schadt, personal communication; ES and FI, unpublished data). Such differences could be ascribed to the different origins of the liver samples, i.e., normal hepatic tissues withdrawn from patients undergone surgical resection for hepatocellular carcinoma (Schröder et al. [2013](#)) or normal livers from postmortem resections (Seiser et al. [2014](#); Schadt et al. [2008](#)). Further functional assessment of these liver tissues by comparable methods and assays should therefore be performed on this allele in future studies.

In conclusion, we are aware that present results do not provide direct proofs on a problematic role of AA for human health. The departing hypothesis suggested *CYP2E1* as a promising candidate gene to be

evaluated, but our positive findings could be explained also by the biotransformation of other xenobiotic or endogenous molecules, in a more complex gene–environment interaction. In any case, the role of CYP2E1 for thyroid carcinogenesis is suggested here. Considering the large amount of previous positive studies (Fig. 2), CYP2E1 appears as a key xenobiotic metabolizing enzyme involved in the genetic susceptibility of several types of cancer and the biotransformation of AA may play a role for most of them. In the future, in vivo studies investigating AA metabolism in relation to *CYP2E1* haplotype #2 should be undertaken, in order to provide more direct evidences on the role of CYP2E1 and AA in human health.

### **Acknowledgements**

This work was funded by a grant from Istituto Toscano Tumori (Florence, Italy), the Robert Bosch Foundation, Stuttgart, Germany (to U.M.Z), and by NIH/NIDDK R21DK081157-01A2 and NIH/NCI K07CA140390-01. The authors thank Uwe Fuhr and Roberto Barale for helpful suggestions.

Conflict of interest. The authors declare that they have no conflict of interest.

## References

1. Andrés E, Cubedo E, Agirre X et al (2006) Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. *Mol Cancer* 5:29. doi: 10.1186/1476-4598-5-29
2. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265. doi: 10.1093/bioinformatics/bth457
3. Beland FA, Mellick PW, Olson GR et al (2013) Carcinogenicity of acrylamide in B6C3F(1) mice and F344/N rats from a 2-year drinking water exposure. *Food Chem Toxicol* 51:149–159. doi: 10.1016/j.fct.2012.09.017
4. Benjamini Y, Drai D, Elmer G et al (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125:279–284. doi: 10.1016/S0166-4328(01)00297-2
5. Besaratinia A, Pfeifer GP (2003) Weak yet distinct mutagenicity of acrylamide in mammalian cells. *J Natl Cancer Inst* 95:889–896
6. Bhattacharya A, Ziebarth JD, Cui Y (2014) PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res* 42:D86–D91. doi: 10.1093/nar/gkt1028
7. Cancemi L, Romei C, Bertocchi S et al (2011) Evidences that the polymorphism Pro-282-Ala within the tumor suppressor gene WWOX is a new risk factor for differentiated thyroid carcinoma. *Int J Cancer* 129:2816–2824. doi: 10.1002/ijc.25937
8. Cardis E, Kesminiene A, Ivanov V et al (2005) Risk of thyroid cancer after exposure to 131I in childhood. *J Natl Cancer Inst* 97:724–732. doi: 10.1093/jnci/dji129
9. Colonna M, Bossard N, Guizard A-V et al (2010) Descriptive epidemiology of thyroid cancer in France: incidence, mortality and survival. *Ann Endocrinol* 71:95–101. doi: 10.1016/j.ando.2009.11.006 (Paris)
10. Delaneau O, Marchini J, Zagury J-F (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9:179–181. doi: 10.1038/nmeth.1785
11. Doroshenko O, Fuhr U, Kunz D et al (2009) In vivo role of cytochrome P450 2E1 and glutathione-S-transferase activity for acrylamide toxicokinetics in humans. *Cancer Epidemiol Biomarkers Prev* 18:433–443. doi: 10.1158/1055-9965.EPI-08-0832
12. EFSA (European Food Safety Authority) (2011) Results on acrylamide levels in food from monitoring years 2007–2009 and exposure assessment. <http://www.efsa.europa.eu/it/efsajournal/doc/2133.pdf>
13. EFSA (European Food Safety Authority) (2012) Update on acrylamide levels in food from monitoring years 2007 to 2010. <http://www.efsa.europa.eu/en/efsajournal/doc/2938.pdf>

14. Fioretti F, Tavani A, Gallus S, Franceschi S, Negri E, La Vecchia C (1999) Case-control study of thyroid cancer in Northern Italy: attributable risk. *Int J Epidemiol* 28:626–630
15. Freidlin B, Zheng G, Li Z, Gastwirth JL (2002) Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Hum Hered* 53:146–152
16. Freisling H, Moskal A, Ferrari P et al (2013) Dietary acrylamide intake of adults in the European Prospective Investigation into Cancer and Nutrition differs greatly according to geographical region. *Eur J Nutr* 52:1369–1380. doi: 10.1007/s00394-012-0446-x
17. Friedman M (2003) Chemistry, biochemistry, and safety of acrylamide. A review. *J Agric Food Chem* 51:4504–4526. doi: 10.1021/jf030204+
18. Friedman MA, Dulak LH, Stedham MA (1995) A lifetime oncogenicity study in rats with acrylamide. *Fundam Appl Toxicol* 27:95–105
19. Gabriel SB, Schaffner SF, Nguyen H et al (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229. doi: 10.1126/science.1069424
20. Gaur A, Jewell DA, Liang Y et al (2007) Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Res* 67:2456–2468. doi: 10.1158/0008-5472.CAN-06-2698
21. Ghanayem BI, McDaniel LP, Churchwell MI et al (2005) Role of CYP2E1 in the epoxidation of acrylamide to glycidamide and formation of DNA and hemoglobin adducts. *Toxicol Sci* 88:311–318. doi: 10.1093/toxsci/kfi307
22. Gomes AM, Winter S, Klein K et al (2009) Pharmacogenomics of human liver cytochrome P450 oxidoreductase: multifactorial analysis and impact on microsomal drug oxidation. *Pharmacogenomics* 10:579–599. doi: 10.2217/pgs.09.7
23. González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, Moreno V (2007) SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics* 23:644–645
24. Grande E, Díez JJ, Zafon C, Capdevila J (2012) Thyroid cancer: molecular aspects and new therapeutic strategies. *J Thyroid Res*. doi: 10.1155/2012/847108Central
25. Hogervorst JG, Schouten LJ, Konings EJ et al (2007) A prospective study of dietary acrylamide intake and the risk of endometrial, ovarian, and breast cancer. *Cancer Epidemiol Biomarkers Prev* 16:2304–2313. doi: 10.1158/1055-9965.EPI-07-0581
26. Hogervorst JG, Fortner RT, Mucci LA et al (2013) Associations between dietary acrylamide intake and plasma sex hormone levels. *Cancer Epidemiol Biomarkers Prev* 22:2024–2036. doi: 10.1158/1055-9965.EPI-13-0509
27. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5:e1000529. doi: 10.1371/journal.pgen.1000529Central

28. Huang Y-F, Chiang S-Y, Liou S-H et al (2012) The modifying effect of CYP2E1, GST, and mEH genotypes on the formation of hemoglobin adducts of acrylamide and glycidamide in workers exposed to acrylamide. *Toxicol Lett* 215:92–99. doi: 10.1016/j.toxlet.2012.10.003
29. IARC (1994) Monographs on the evaluation of carcinogenic risks to humans <http://monographs.iarc.fr/ENG/Monographs/vol60/mono60.pdf>
30. Iorio MV, Ferracin M, Liu C-G et al (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 65:7065–7070. doi: 10.1158/0008-5472.CAN-05-1783
31. Johnson KA, Gorzinski SJ, Bodner KM et al (1986) Chronic toxicity and oncogenicity study on acrylamide incorporated in the drinking water of Fischer 344 rats. *Toxicol Appl Pharmacol* 85:154–168
32. Khan MA, Davis CA, Foley GL et al (1999) Changes in thyroid gland morphology after acute acrylamide exposure. *Toxicol Sci* 47:151–157
33. Klein K, Winter S, Turpeinen M et al (2010) Pathway-targeted pharmacogenomics of CYP1A2 in human liver. *Front Pharmacol*. doi: 10.3389/fphar.2010.00129Central
34. Koyama N, Sakamoto H, Sakuraba M et al (2006) Genotoxicity of acrylamide and glycidamide in human lymphoblastoid TK6 cells. *Mutat Res* 603:151–158. doi: 10.1016/j.mrgentox.2005.11.006
35. Malchoff CD, Malchoff DM (2004) Familial papillary thyroid carcinoma. *Cancer Treat Res* 122:381–387
36. Maronpot RR, Thoolen RJMM, Hansen B (2015) Two-year carcinogenicity study of acrylamide in Wistar Han rats with in utero exposure. *Exp Toxicol Pathol* 67:189–195. doi: 10.1016/j.etp.2014.11.009
37. Mei N, Guo L, Tseng J et al (2008) Gene expression changes associated with xenobiotic metabolism pathways in mice exposed to acrylamide. *Environ Mol Mutagen* 49:741–745. doi: 10.1002/em.20429
38. Mottram DS, Wedzicha BL, Dodson AT (2002) Acrylamide is formed in the Maillard reaction. *Nature* 419:448–449. doi: 10.1038/419448a
39. Naccarati A, Pardini B, Stefano L et al (2012) Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk. *Carcinogenesis* 33:1346–1351. doi: 10.1093/carcin/bgs172
40. Nakano M, Mohri T, Fukami T et al (2015) Single-nucleotide polymorphisms in cytochrome P450 2E1 (CYP2E1) 3'-untranslated region affect the regulation of CYP2E1 by miR-570. *Drug Metab Dispos* 43:1450–1457. doi: 10.1124/dmd.115.065664
41. Neafsey P, Ginsberg G, Hattis D et al (2009) Genetic polymorphism in CYP2E1: Population distribution of CYP2E1 activity. *J Toxicol Environ Health B Crit Rev* 12:362–388. doi: 10.1080/10937400903158359
42. Obón-Santacana M, Kaaks R, Slimani N et al (2014) Dietary intake of acrylamide and endometrial cancer risk in the European Prospective Investigation into Cancer and Nutrition cohort. *Br J Cancer* 111:987–997. doi: 10.1038/bjc.2014.328Central

43. Pedreschi F, Mariotti MS, Granby K (2014) Current issues in dietary acrylamide: formation, mitigation and risk assessment. *J Sci Food Agric* 94:9–20. doi: 10.1002/jsfa.6349
44. Pelucchi C, La Vecchia C, Bosetti C et al (2011) Exposure to acrylamide and human cancer—a review and meta-analysis of epidemiologic studies. *Ann Oncol* 22:1487–1499. doi: 10.1093/annonc/mdq610
45. Piva F, Giulietti M, Nocchi L, Principato G (2009) SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinformatics* 25:1211–1213. doi: 10.1093/bioinformatics/btp124
46. R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>
47. Reese MG, Eeckman FH, Kulp D, Haussler D (1997) Improved splice site detection in Genie. *J Comput Biol* 4:311–323
48. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. *RNA* 10:1507–1517. doi: 10.1261/rna.5248604Central
49. Rieger JK, Klein K, Winter S, Zanger UM (2013) Expression variability of absorption, distribution, metabolism, excretion-related microRNAs in human liver: influence of nongenetic factors and association with gene expression. *Drug Metab Dispos* 41:1752–1762. doi: 10.1124/dmd.113.052126
50. Ron E (2007) Thyroid cancer incidence among people living in areas contaminated by radiation from the chernobyl accident. *Health Phys* 93:502–511. doi: 10.1097/01.HP.0000279018.93081.29
51. Schadt EE, Molony C, Chudin E et al (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 6:e107. doi: 10.1371/journal.pbio.0060107Central
52. Schröder A, Klein K, Winter S et al (2013) Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. *Pharmacogenomics J* 13:12–20. doi: 10.1038/tpj.2011.44
53. Segerbäck D, Calleman CJ, Schroeder JL et al (1995) Formation of N-7-(2-carbamoyl-2-hydroxyethyl)guanine in DNA of the mouse and the rat following intraperitoneal administration of [<sup>14</sup>C]acrylamide. *Carcinogenesis* 16:1161–1165
54. Seiser EL et al (2014) Meta-analysis of liver eQTL studies and cross-tissue eQTL comparison using GTEx Data. <http://www.ashg.org/2014meeting/abstracts/fulltext/f140123474.htm>. <http://www.ashg.org/2014meeting/abstracts/fulltext/f140123474.htm>. Accessed 24 Jul 2015
55. Settels E, Bernauer U, Palavinkas R et al (2008) Human CYP2E1 mediates the formation of glycidamide from acrylamide. *Arch Toxicol* 82:717–727. doi: 10.1007/s00204-008-0296-8
56. Sipos JA, Mazzaferri EL (2010) Thyroid cancer epidemiology and prognostic variables. *Clin Oncol* 22:395–404. doi: 10.1016/j.clon.2010.05.004
57. Stadler RH, Blank I, Varga N et al (2002) Acrylamide from Maillard reaction products. *Nature* 419:449–450. doi: 10.1038/419449a

58. Sumner SC, Fennell TR, Moore TA et al (1999) Role of cytochrome P450 2E1 in the metabolism of acrylamide and acrylonitrile in mice. *Chem Res Toxicol* 12:1110–1116
59. Surks MI, Chopra IJ, Mariash CN et al (1990) American Thyroid Association guidelines for use of laboratory tests in thyroid disorders. *JAMA* 263:1529–1532
60. Tareke E, Rydberg P, Karlsson P et al (2002) Analysis of acrylamide, a carcinogen formed in heated foodstuffs. *J Agric Food Chem* 50:4998–5006
61. Tremmel R, Klein K, Winter S, Schaeffeler E, Zanger UM (2015) Gene copy number variation analysis reveals dosage-insensitive expression of CYP2E1. *Pharmacogenomics J*. doi: 10.1038/tpj.2015.69
62. Yaylayan VA, Machiels D, Istasse L (2003) Thermal decomposition of specifically phosphorylated D-glucoses and their role in the control of the Maillard reaction. *J Agric Food Chem* 51:3358–3366. doi: 10.1021/jf034037p
63. Zanger UM, Schwab M (2013) Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* 138:103–141. doi: 10.1016/j.pharmthera.2012.12.007
64. Ziebarth JD, Bhattacharya A, Chen A, Cui Y (2012) PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res* 40:D216–D221. doi: 10.1093/nar/gkr1026
65. Zyzak DV, Sanders RA, Stojanovic M et al (2003) Acrylamide formation mechanism in heated foods. *J Agric Food Chem* 51:4782–4787. doi: 10.1021/jf034180i

**Table 1**

Characteristics of controls and DTC patients

	Discovery set				Replication set			
	Controls <i>N</i> (%)	Cases <i>N</i> (%)	OR <sup>a</sup> (95 % CI)	<i>P</i> value	Controls <i>N</i> (%)	Cases <i>N</i> (%)	OR (95 % CI)	<i>P</i> value
Sex								
Male	195 (56.0 %)	97 (27.7 %)	Reference		1441 (51.8 %)	296 (26.5 %)	Reference	
Female	155 (44.0 %)	253 (72.3 %)	3.28 (2.39–4.50)	<10 <sup>-6a</sup>	1338 (48.2 %)	821 (73.5 %)	2.97 (2.55–3.46)	<10 <sup>-6a</sup>
Age (years)								
Average	49.26 ± 6.82	45.62 ± 13.72		8.7 × 10 <sup>-5b</sup>	49.30 ± 14.13	45.85 ± 13.06		<10 <sup>-6b</sup>
Median	48.0	45.0			48.0	45.0		
Body mass index								
Average	25.54 ± 3.29	26.17 ± 4.93		0.047 <sup>b</sup>	25.70 ± 4.28	25.90 ± 4.73		0.08 <sup>b</sup>
Median	25.26	25.57			25.00	25.20		
Smoking								
Non-smokers	240 (68.6 %)	212 (60.6 %)	Reference		1844 (66.4 %)	660 (59.2 %)	Reference	
Smokers + Ex-smokers	110 (31.4 %)	138 (39.4 %)	1.42 (1.04–1.94)	0.027 <sup>a</sup>	935 (33.6 %)	457 (40.8 %)	1.37 (1.18–1.57)	1.71 × 10 <sup>-5a</sup>
Total	350	350			2779	1117		
Diagnosis								
PTC <sup>c</sup>		322 (92.0 %)				1030 (92.2 %)		
FTC <sup>d</sup>		28 (8.0 %)				87 (7.8 %)		

Smoker quitters were grouped with smokers, and volunteers exposed to passive cigarette smoke are grouped with non-smokers

<sup>a</sup>Crude odds ratio from logistic regression analysis

<sup>b</sup>*P* values from *t* Student's test

<sup>c</sup>Papillary

<sup>d</sup>Follicular thyroid carcinoma

**Table 2**Statistical analyses for SNPs within *CYP2E1* gene on discovery set

	Controls <i>N</i> (%)	DTC <i>N</i> (%)	Or <sup>b</sup> (95 % CI)	<i>P</i> value
rs9418982				
C-allele <sup>a</sup>	154 (22.1 %)	149 (21.2 %)	0.95 (0.74–1.22)	0.67 <sup>c</sup>
Genotypes				
T/T	213 (60.8 %)	217 (62.2 %)	Reference	
T/C	120 (34.2 %)	117 (33.3 %)	1.00 (0.69–1.45)	1.00 <sup>d</sup>
C/C	17 (5.0 %)	16 (4.5 %)	0.80 (0.35–1.83)	0.60 <sup>d</sup>
Best model <sup>e</sup>	Additive; <i>P</i> -trend = 0.675			
rs7092584				
T-allele <sup>a</sup>	67 (18.7 %)	67 (19.3 %)	1.04 (0.72–1.49)	0.84 <sup>c</sup>
Genotypes				
C/C	288 (82.2 %)	285 (81.3 %)	Reference	
C/T	59 (16.9 %)	63 (18.1 %)	1.37 (0.85–2.21)	0.20 <sup>d</sup>
T/T	3 (0.9 %)	2 (0.6 %)	0.69 (0.06–7.41)	0.76 <sup>d</sup>
Best model <sup>e</sup>	Additive; <i>P</i> -trend = 0.656			
rs743534				
G-allele <sup>a</sup>	114 (16.2 %)	73 (16.2 %)	0.95 (0.70–1.30)	0.71 <sup>c</sup>
Genotypes				
T/T	241 (69.0 %)	276 (78.9 %)	Reference	
T/G	104 (29.6 %)	67 (19.2 %)	1.00 (0.67–1.50)	1.00 <sup>d</sup>

	Controls <i>N</i> (%)	DTC <i>N</i> (%)	Or <sup>b</sup> (95 % CI)	<i>P</i> value
G/G	5 (1.4 %)	7 (1.9 %)	0.83 (0.18–3.70)	0.81 <sup>d</sup>
Best model <sup>e</sup>	Dominant; <i>P</i> -trend = 0.369			
rs2480258				
A-allele <sup>a</sup>	151 (21.6 %)	173 (24.7 %)	1.19 (0.93–1.53)	0.16 <sup>c</sup>
Genotypes				
G/G	209 (59.7 %)	200 (57.1 %)	Reference	
G/A	131 (37.4 %)	127 (36.3 %)	1.06 (0.76–1.48)	0.70 <sup>d</sup>
A/A	10 (2.9 %)	23 (6.6 %)	<b>2.63 (1.15–6.03)</b>	<b>0.02<sup>d</sup></b>
Best model <sup>e</sup>	Recessive; <i>P</i> -trend = <b>0.020</b>			
A/A vs. (G/G + G/A)	340 (97.1 %)	327 (93.4 %)	<b>2.57 (1.14–5.83)</b>	<b>0.02<sup>d</sup></b>

Significant *P* values are in bold

<sup>a</sup>Minor allele frequency

<sup>b</sup>Adjusted by age, gender, BMI, and smoke

<sup>c</sup>Two-sided Chi-square test for distribution of allelic frequencies

<sup>d</sup>*P*-association of the odds ratio from multivariate logistic regression analysis adjusted for covariates

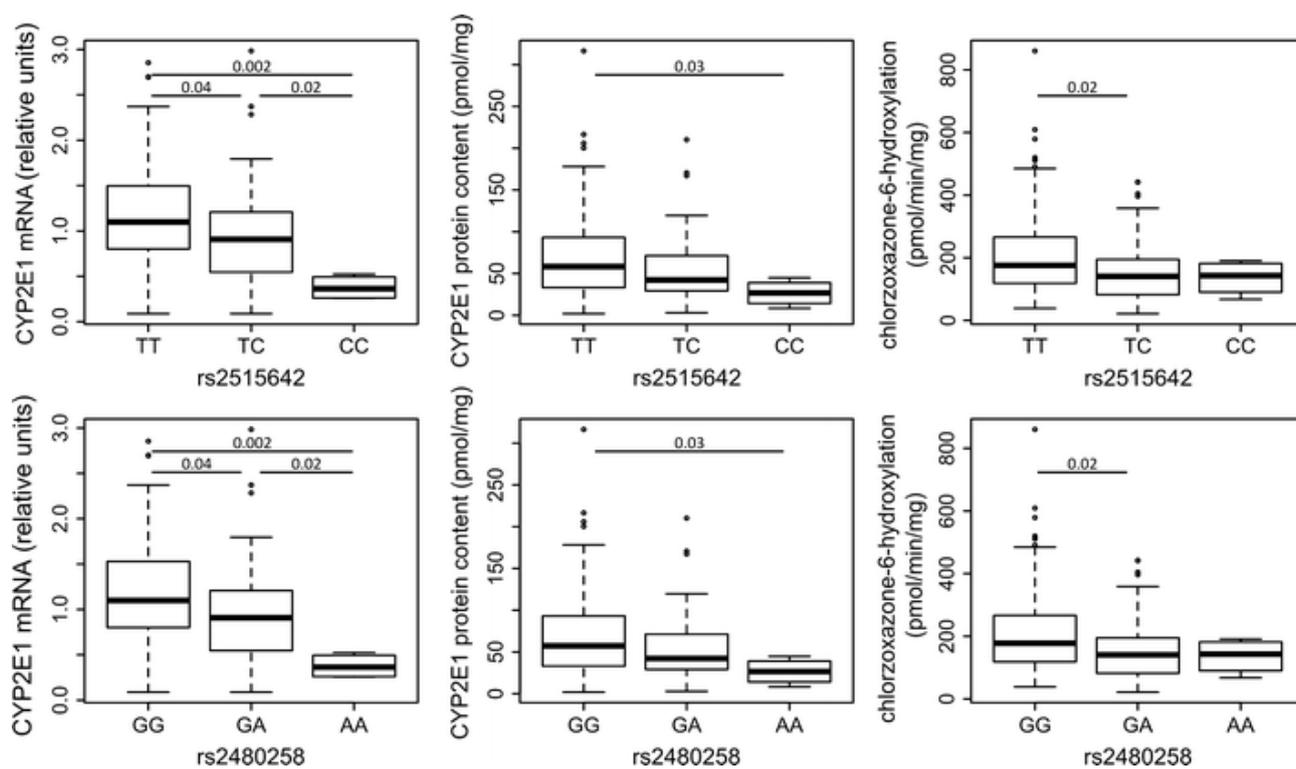
<sup>e</sup>Best model for NMTC (Non-medullary thyroid carcinomas), according to the Cochran–Armitage trend test Analyses in the replication set

**Table 3**

Statistical analyses for rs2480258 on replication set and all sample set

	Replication set				Pooled sets			
	Controls <i>N</i> (%)	DTC <i>N</i> (%)	OR <sup>b</sup> (95 % CI)	<i>P</i> value	Controls <i>N</i> (%)	DTC <i>N</i> (%)	OR <sup>b</sup> (95 % CI)	<i>P</i> value
rs2480258								
A-allele <sup>a</sup>	1078 (22.2 %)	383 (25.0 %)	<b>1.17</b> ( <b>1.02–1.33</b> )	<b>0.02<sup>c</sup></b>	1229 (22.1 %)	556 (24.9 %)	<b>1.17</b> ( <b>1.04–1.31</b> )	<b>0.008<sup>c</sup></b>
Genotypes								
G/G	1461 (60.2 %)	431 (56.2 %)	Reference		1670 (60.1 %)	631 (56.9 %)	Reference	
G/A	858 (35.3 %)	289 (37.7 %)	<b>1.24</b> ( <b>1.03–1.48</b> )	<b>0.02<sup>d</sup></b>	989 (35.6 %)	416 (36.9 %)	<b>1.20</b> ( <b>1.02–1.40</b> )	<b>0.02<sup>d</sup></b>
A/A	110 (4.5 %)	47 (6.1 %)	<b>1.56</b> ( <b>1.06–2.30</b> )	<b>0.02<sup>d</sup></b>	120 (4.3 %)	70 (6.2 %)	<b>1.68</b> ( <b>1.20–2.35</b> )	<b>0.002<sup>d</sup></b>
Best model <sup>e</sup>	Additive; <i>P</i> -trend = 0.023	Additive; <i>P</i> -trend = 0.008						
Total	2429	767			2779	1117		

Significant *P* values are in bold<sup>a</sup>Minor allele frequency<sup>b</sup>Adjusted by age, gender, BMI, and smoke<sup>c</sup>Two-sided Chi-square test for distribution of allelic frequencies<sup>d</sup>*P*-association of the odd ratio from multivariate logistic regression analysis adjusted for covariates<sup>e</sup>best model for NMTC, according to the Cochran–Armitage trend test



**Fig. 1**

CYP2E1 genotype–phenotype correlation analysis in 149 human liver subjects. Relative CYP2E1 mRNA expression, measured using qPCR, protein levels, determined by Western blot analysis, and enzyme activity, quantified by LC-MSMS analysis of chlorzoxazone-6-hydroxylation (mg indicates mg of microsomal protein), are plotted against the genotypes of **a** rs2515642 and **b** rs2480258 as box-and-whisker plots with outliers. Statistically significant *P* values <0.05 are displayed (pairwise *t* tests)

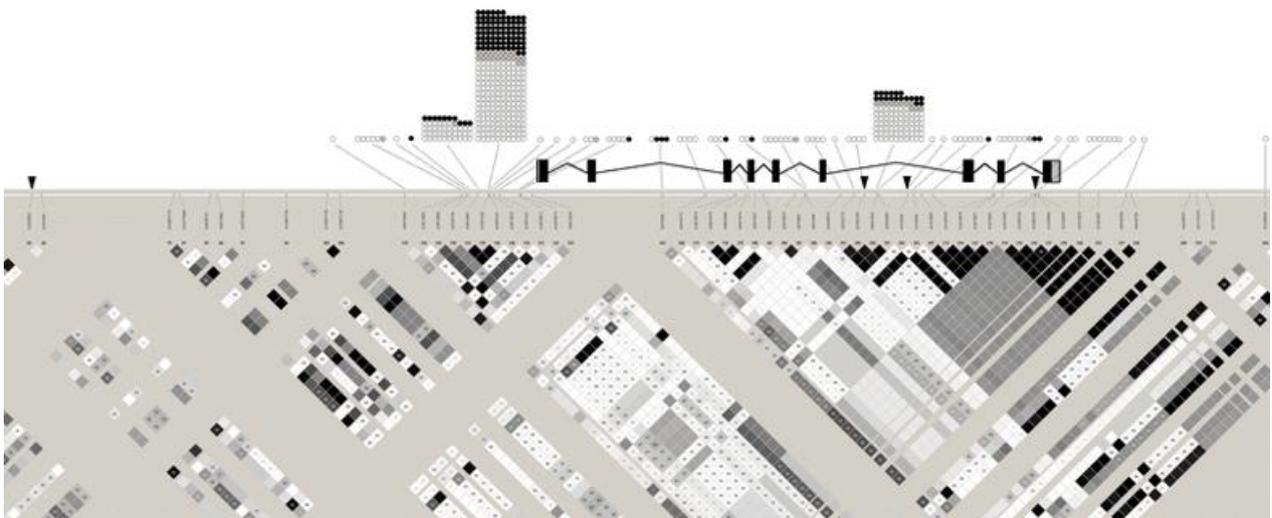
**Table 4**

SNPs within the 3'UTR of *CYP2E1* gene are in high linkage disequilibrium with various other SNPs in the nearby region leading to only two prevalent haplotypes (1 and 2) surrounding the 3'UTR, in the Caucasian population (data from [www.1000Genomes.org](http://www.1000Genomes.org))

SNP ID	Chromosome 10	Gene region	Haplotype					
			#1	#2	#3	#4	#5	#6
rs2515642	133,538,509	IVS8	T	C	T	T	T	T
rs2480259	133,538,572	IVS8	G	A	G	G	G	G
rs2480258 <sup>a</sup>	133,538,596	IVS8	C	T	C	T	C	C
rs2249694	133,538,649	IVS8	G	A	G	G	A	G
rs2249695	133,538,664	IVS8	C	T	C	C	C	C
rs2480257	133,539,005	3'UTR	A	T	T	A	A	A
rs2480256	133,539,010	3'UTR	G	A	A	G	G	G
rs2480255	133,539,351	3'flank	C	T	C	C	C	T
rs2480254	133,539,365	3'flank	C	T	C	C	C	T
rs2515644	133,539,575	3'flank	C	A	C	C	C	C
		Frequency	0.759	0.234	0.004	0.001	0.001	0.001

All SNPs are mapped to the forward strand on the reference assembly (GRCh38)

<sup>a</sup>ht-SNP



**Fig. 2**

Graph representing published case–control association studies on specific polymorphisms of *CYP2E1* in relation with the risk of various types of diseases (not limited to cancer) in humans. Black dots represent significant associations at 0.05 level, gray dots represent associations with *P* values ranging between 0.05 and 0.10, white dots represent negative studies. The structure of the gene is over-imposed on the top of the LD structure. The LD measure is the  $r^2$  (*black*, LD = 1; *white*, LD = 0). See supplementary Table 2 for further details on studies