Etude des facteurs de risque génétiques et des interactions gène-environnement dans les cancers différenciés de la thyroïde

Catherine Ines Tcheandjieu Gueliatcha

▶ **To cite this version:**

Catherine Ines Tcheandjieu Gueliatcha. Etude des facteurs de risque génétiques et des interactions gène-environnement dans les cancers différenciés de la thyroïde. Cancer. Université Paris Saclay (COmUE), 2017. Français. NNT : 2017SACLS064 . tel-02509609

HAL Id: tel-02509609
https://theses.hal.science/tel-02509609

Submitted on 17 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2017SACLS064

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
UNIVERSITE DE PARIS SUD

ECOLE DOCTORALE N° 570
EDSP Santé publique

Spécialité de doctorat : Santé publique - épidémiologie

Par

Madame Catherine Ines Tcheandjieu Gueliatcha

ETUDE DES FACTEURS DE RISQUE GENETIQUES ET DES INTERACTIONS GENE-ENVIRONNEMENT DANS LES CANCERS DIFFERENCIES DE LA THYROÏDE

Thèse présentée et soutenue à Villejuif, le 16 Mars 2017

Composition du Jury :

Monsieur Philippe Broët	Professeur, Université de paris Sud Université de paris Saclay	Président
Monsieur James McKay	Chercheur CIRC, Lyon	Rapporteur
Madame Emmanuelle Génin	Directrice de recherche, Inserm, Université de Bretagne Occidentale	Rapporteur
Madame Simone Benhamou	Directrice de recherche, Inserm, Fondation Jean Dausset (CEPH)	Examinatrice
Madame Sabina Rinaldi	Chercheure, CIRC, Lyon	Examinatrice
Monsieur Pascal Guénel	Directeur de recherche, Inserm Université de Paris Sud- Paris Saclay	Directeur de thèse
Madame Thérèse Truong	Chargée de recherche, Inserm Université de Paris Sud- Paris Saclay	Co-directrice de thèse

A André-Mathurin Tcheandjieu, mon père

Remerciements

Tout d'abord, je souhaite remercier mes directeurs de thèse, Pascal Guénel et Thérèse Truong, de m'avoir accueillie dans leur équipe et de m'avoir permis de réaliser cette thèse. Je vous remercie pour votre encadrement et la confiance que vous m'avez accordée. Je remercie tout particulièrement Thérèse qui a principalement encadré cette thèse. Merci pour son soutien tant sur le plan personnel que professionnel, pour toute les opportunités professionnelles qu'elle m'a permis d'obtenir et pour tous ses conseils. Merci pour l'autonomie qu'elle m'a laissée, et pour m'avoir fait bénéficier de ses connaissances et de son expérience. Merci pour tout le temps et la patience qu'elle m'a accordée, pour sa générosité et tous les moments de rire aux éclats que nous avons eu… C'était un réel plaisir de travailler avec Thérèse et je suis très reconnaissante pour tout ce qu'elle m'a donné et transmis au cours de ces années de thèse.

Je remercie tous les membres du jury d'avoir accepté d'évaluer mon travail de thèse et d'être venus pour en discuter le jour de ma soutenance. Merci à Philippe Broët de me faire l'honneur de présider le jury. Merci à mes deux rapporteurs Emmanuelle Genin et James McKay pour le temps qu'ils ont dédié à la relecture de mon manuscrit et pour leurs commentaires très enrichissants. Merci à Simone Benhamou et Sabina Rinaldi d'avoir accepté d'être membre du jury et d'apporter leur expertise.

Je remercie l'Ecole des Hautes Etudes en Santé Publique (EHESP) pour avoir financé ma thèse ainsi que mes formations et séjours collaboratifs à l'étranger. Je tiens également à remercier le directeur de l'Ecole Doctorale de Santé publique (EDSP) et un grand merci à Audrey Bourgeois, notre ange gardien de l'EDSP pour son aide, sa patience, sa bonne humeur et sa disponibilité inconditionnelle.

Mes remerciements vont également à Alice Sigurdsson et Alina Brenner de l'équipe d'épidémiologie des radiations au département d'épidémiologie et génétique des cancers du NCI (REB/DCEG/NCI) d'avoir accepté de monter un projet collaboratif avec nous. Merci tout particulièrement à Alina Brenner pour son encadrement, sa disponibilité et sa bienveillance qui a rendu mon séjour au NCI riche et agréable. Merci aussi à Ruth Pfeiffer pour m'avoir fait bénéficier de son savoir et de ses connaissances statistiques.

Merci aux membres des groupes des études CATHY, Nouvelle Calédonie, USRT/UTMDACC et YoungThyr pour m'avoir permis d'utiliser leurs données.

Merci à Fabienne Lesueur pour sa disponibilité, ses conseils, sa relecture de mon premier travail de thèse et de m'avoir initiée à la bio-informatique.

Merci aux membres anciens et nouveaux de l'équipe cancer environnement du CESP pour la bonne humeur et la bonne ambiance au sein de l'équipe. Un merci tout particulièrement à Harinakshi d'avoir été une très bonne amie et une co-bureau parfaite ces années. Merci à Marie pour son aide, sa disponibilité pour les données et sa bonne humeur, je me souviens encore de ma première macro SAS réalisé avec son aide. Merci à Corinne pour sa disponibilité et sa patience. Merci à Yves pour sa joie de vivre, les cours de musique et les supers soirées…☺ . Merci à Gaëlle pour son grand cœur, le partage d'idées et d'astuces cheveux… Merci à Agnès pour m'avoir soutenue lors de l'écriture de mon manuscrit, pour sa relecture et pour toutes les discussions très génétique et autres… ☺.

Merci à mes amis actuels ou anciens thésards : Sophie, Astrid, Manik, Pauline, Eve, Jonathan, Wen Lun, Guy Marino, Fanny, Fabienne… avec qui ces années ont été agréables notamment à travers nos activités de l'asso ED420 ou hors asso. Merci à Astrid et Sophie pour nos soirées appart, et un merci particulier à Astrid pour son soutien, sa présence et ces années d'amitié. Merci à Sophie qui a su être présente et me soutenir même à distance.

Merci à mes amis du CESP Julie et Sandra ; merci à Samuel, Jeanne, Zhen, Josefin, Alex et Marie Michelle pour les années de bonheur passées à la fac du Kremlin Bicêtre ; à mes amis et famille « Djazair » et particulièrement Arsène et Guy avec qui j'ai fait le parcours Cameroun-Algérie-France et surtout à Anicet pour tous les coups de fil de désespoir…. Merci à Alex How Kit pour la relecture de ma thèse, ses suggestions et nos débats scientifiques ; Merci à Essi mon coloc et ami de Washington DC pour son accueil chaleureux, sa gentillesse, son soutien, sa présence… « you are the best ☺ »

Un grand merci à Yasmine et Fred qui sont plus que des amis et grâce à qui j'ai gagné une nouvelle famille, merci pour les weekends à Ticheville, les séjours Ski, musique et autres…

Merci à Monette et Bernard Bathany ma famille bretonne pour avoir cru en moi et pour leur soutien…

Merci à ma famille qui m'a toujours soutenue, supportée et encouragée. Un merci particulier à mon oncle François qui m'a accueillie et épaulée à Paris, merci à Ameny et Christelle, mes sœurs et amies. Merci à ma maman d'avoir toujours été là pour moi et spéciale dédicace à mon père pour qui j'ai réalisé ce rêve. J'espère que depuis là-haut tu verras ce travail que je te dédie.

Titre : Etude des facteurs de risque génétiques et des interactions gène-environnement dans les cancers différenciés de la thyroïde

Mots Clés : cancers de la thyroïde, épidémiologie, génétique

Résumé

Contexte : L'incidence des cancers différenciés de la thyroïde (CDT) est caractérisée par de fortes variations géographiques et ethniques. L'un des plus forts taux d'incidence a été rapporté en Nouvelle-Calédonie et particulièrement dans la population mélanésienne. En dehors de l'exposition aux radiations ionisantes dans l'enfance et l'obésité, les facteurs de risque de CDT restent très mal connus. Bien qu'il ait été rapporté que ce cancer constitue une des localisations cancéreuses ayant la composante héréditaire la plus élevée, le rôle des facteurs génétiques a été jusqu'à présent peu étudié. Des études de cas familiaux et des études gène-candidat ont identifié des polymorphismes dans plusieurs gènes mais très peu de ces associations ont été répliquées. Des études d'association génomes entier (GWAS) ont également mis en évidence des régions de susceptibilité génétiques de CDT, les associations les plus robustes ont été rapportés sur les loci 9q22 et 14q13. Au final, l'ensemble des variants identifiés jusqu'à présent ne permettent d'expliquer qu'une faible part de l'héritabilité génétique dans l'étiologie des CDT, suggérant que d'autres facteurs de risque génétiques restent à découvrir.

Objectifs : L'objectif général de ce travail était d'étudier les facteurs de risque génétique et leurs interactions avec les facteurs environnementaux. Plus spécifiquement, il s'agissait 1) d'étudier le rôle de gènes candidats ou de loci identifiés dans des études GWAS dans deux études cas-témoins menées en Nouvelle-Calédonie et en France métropolitaine 2) d'explorer plus en détail dans ces populations, les loci GWAS 9q22 et 14q13 pour l'identification des variants causaux ou d'autres variants candidats ; 3) d'identifier de nouveaux facteurs de risque génétiques dans les CDT chez les femmes en utilisant une approche dite par « pathway candidat » en combinant les données de deux études françaises et américaines.

Matériels et méthodes : Les analyses portant sur les gènes candidats et les loci GWAS reposent sur deux populations d'études : une population européenne de 508 cas et 621 témoins issus des études cas-témoins conduites en France métropolitaine (étude CATHY) et en Nouvelle-Calédonie (NC) et une population mélanésienne de 156 cas et 114 témoins de l'étude de NC. Les approches pathway candidat ont porté dans un premier temps sur les femmes d'origine européenne de l'étude CATHY (365 cas/376 témoins) et d'une partie de l'étude cas-témoins Young-Thyr (83 cas/93 témoins) conduites en France métropolitaine. Dans un second temps, nous avons combiné ces sujets avec les femmes de l'étude cas-témoins USRT/UTMDACC (332 cas/443 témoins) conduite aux Etats-Unis.

Résultats : Nous n'avons pas observé d'association entre les CDT et les polymorphismes des gènes identifiés précédemment par les études gène-candidat dans les populations européennes et mélanésiennes étudiées. Toutefois, nous avons montré que la délétion des gènes *GSTM1* et *GSTT1* pouvait moduler les associations rapportées entre l'obésité ou la consommation d'alcool et le risque de CDT. Nous avons répliqué chez les européens et chez les mélanésiens les associations avec des variants des loci GWAS 9q22, 14q13 et 2q35. Nous avons également mis en évidence, dans les régions GWAS 9q22 et 14q13 de nouveaux variants candidats à risque de CDT et rapporté

des interactions entre ces variants et la parité ou la consommation de tabac. Les analyses par pathway portant sur une population de femmes d'origine européenne suggèrent des interactions entre la consommation d'alcool ou la consommation de tabac et les gènes impliqués dans le métabolisme de ces substances et des interactions entre l'âge aux premières règles, la prise de contraceptifs oraux et les gènes impliqués dans la biosynthèse et le métabolisme des hormones stéroïdiennes.

Title: Study of genetics risk factors and gene-environment interaction in differentiated thyroid cancers

Key words: thyroid cancers, epidemiology, genetic

Abstract

Context: Differentiated thyroid cancer (DTC) incidence is characterized by considerable geographic and ethnic variations. Particularly high incidence rates were observed in Melanesian women of New Caledonia. Except for the exposure to ionizing radiation in childhood and obesity, the role of other DTC risk factors is not clearly established. Genetic factors have been suggested to play an important role in DTC risk since epidemiological studies have shown that DTC has a higher familial relative risk than any other cancers. Linkage studies in multiple-case DTC families and candidate gene studies have identified polymorphisms in several genes but very few have been replicated so far. Genome-wide association studies (GWAS) also identified several DTC susceptibility loci with the most robust associations reported on the loci 9q22 and 14q13. Only few susceptibility loci were highlighted by GWAS and the identified variants were shown to account less than 10% of the DTC familial risk, emphasizing that much remains to be discovered

Objectives: The main objective of this work was to study the role of genetic risk factors and their interaction with environmental factors in DTC risk. More specifically, we aimed to: 1) replicate the association between DTC risk and polymorphisms reported in candidate gene and GWAS studies in 2 case-control studies conducted in Metropolitan France and New Caledonia; 2) identify potential causal variants of DTC risk in GWAS loci 9q22 and 14q13 using fine-mapping approach; 3) identify new genetic risk factors for DTC in women using pathway approach by pooling data from 2 case-control studies conducted in France and USA..

Materials and methods: The analysis of the candidate genes and of the GWAS loci were based on a European population of 508 cases and 621 controls from 2 case-control studies conducted in metropolitan France (CATHY study) and in New Caledonia (NC study) and, a Melanesian population of 156 cases and 114 controls from the NC study. The pathway analysis was conducted in a first step on European women from the CATHY study (365 cases/376 controls) and from the Young-Thyr study (83 cases /93 controls) both conducted in metropolitan France. In a second step, we pooled the data from CATHY/Young-Thyr study and USRT/UTMDACC study (332 cases/443 controls) conducted in the United States.

Results: In Europeans and Melanesians, we found no association with polymorphisms reported previously by candidate genes studies. However, we observed that among these genes, *GSTM1* and *GSTT1* may modulate the associations between DTC risk and obesity or alcohol consumption. Some polymorphisms identified in GWAS studies at loci 9q22, 14q13 and 2q35 were replicated in Europeans and in Melanesians. In the GWAS loci 9q22 and 14q13, we identify new variants that can be functionally related to DTC pathogenesis in Europeans and Melanesians. We also reported interactions between some of these variants and parity or tobacco smoking. The analysis of candidate pathways in European women showed interactions between alcohol consumption or tobacco smoking and genes involved in

the metabolism of these compounds and, between age at first menarche or oral contraception and genes involved in biosynthesis and metabolism of sex steroid hormones.

Production scientifique

Publications

Tcheandjieu C, Lesueur F, Sanchez M, Baron-Dubourdieu D, Guizard AV, Mulot C, Puig PL, Schvartz C, Truong T and Guenel P. Fine-mapping of two differentiated thyroid carcinoma susceptibility loci at 9q22.33 and 14q13.3 detects novel candidate functional SNPs in Europeans from metropolitan France and Melanesians from New Caledonia. **Int. J. Cancer. 2016;139.**

Tcheandjieu C, Cordina-Duverger E**,** Sanchez M, Baron-Dubourdieu D, Guizard AV, Mulot C, Puig PL, Schvartz C, Guenel P and Truong T. Association between *GSTM1* and *GSTT1* Copy number variation and differentiated thyroid Risk in Europeans from metropolitan France and Melanesians from New Caledonia (**soumis**)

Tcheandjieu C, Cordina-Duverger E**,** Sanchez M ,…, Schvartz C, Guizard AV, Mulot C, Puig PL, de Vathaire F, Truong T and Guenel P. Study of differentiated thyroid cancer genetic susceptibility risk using a candidate pathway approach in a French population-based case-control study (**en préparation**)

Tcheandjieu C, …,Pfeiffer R, Alice Sigurdsson,…, deVathaire F, Guenel P, Brenner A and Truong T. Genetic variants in metabolism of steroid hormones and metabolism of xenobiotics pathways and the risk of differentiated thyroid cancer in women: a pooled analysis of 2 case-control studies (**en preparation**)

Congrès internationaux

Communication Orale

Association des Epidémiologiste de langue Française (ADELF)
"Rôle des polymorphismes de *GSTM1* et *GSTT1* dans le risque de cancers différenciés de la thyroïde en France Métropolitaine et en Nouvelle-Calédonie"
Tcheandjieu C, Cordina-Duverger E, Sanchez M, Guenel P et Truong T ; **Sept 2016**

Communications affichées

American Association of Cancer Research Annual meeting (AACR), Philadelphia, US
"Fine Mapping of differentiated thyroid cancer risk GWAS loci at 9q22 and 14q13 in Metropolitan France and in New Caledonia"
Tcheandjieu C, Lesueur F, Sanchez M, Truong T and Guenel P ; **Avril 2015**

American society of Human genetics Annual meeting (ASHG), Baltimore, US
"Fine Mapping of differentiated thyroid cancer risk GWAS loci at 9q22 and 14q13 in Metropolitan France and in New Caledonia"
Tcheandjieu C, Lesueur F, Sanchez M, Truong T and Guenel P, **Octobre 2015**

IARC's 50Th Anniversary Conference, Lyon, France
"Role of *GSTM1* and *GSTT1* copy number variations in differentiated thyroid cancer risk among Europeans from metropolitan France and Melanesians from New-Caledonia" **Tcheandjieu C**, Cordina-Duverger E, Sanchez M, Truong T and Guenel P, **Juin 2016**

American society of Human genetics Annual meeting (ASHG), Vancouver, Canada
"Role of *GSTM1* and *GSTT1* copy number variations in differentiated thyroid cancer risk among Europeans from metropolitan France and Melanesians from New-Caledonia" **Tcheandjieu C**, Cordina-Duverger E, Sanchez M, Truong T and Guenel P, **Octobre 2016**

Table de matière

Introduction générale 16

1. La glande thyroïdienne 16
2. Les cancers de la thyroïde 17
3. Epidémiologie descriptive des cancers de la thyroïde. 17
4. Les facteurs de risque des cancers de la thyroïde 20
   - 4.1. Les antécédents personnels de pathologie bénigne de la thyroïde 20
   - 4.2. Expositions aux radiations ionisantes 20
   - 4.3. Les facteurs anthropométriques 20
   - 4.4. Les facteurs hormonaux et reproductifs 21
   - 4.5. La consommation d'alcool et de tabac 24
   - 4.6. Apport alimentaire en iode 25
   - 4.7. Antécédent familiaux des CDT 25
5. Facteurs génétiques 27
   - 5.1. Les gènes identifiés dans les cas familiaux 27
   - 5.2. Les gènes identifiés selon les approches gène candidat 27
   - 5.3. Les études GWAS dans les CDT 29
   - 5.4. Les interactions gènes-environnement dans les CDT 31

Objectifs de la thèse 32

Partie 1 Etude des gènes candidats et des régions identifiées par les études GWAS dans les cancers de la thyroïde 34

- 1.1. Introduction 34
- 1.2. Matériel et méthodes 35
  - 1.2.1. Populations d'étude 35
  - 1.2.1.a. Etude cas-témoins en Nouvelle-Calédonie 35
  - 1.2.1.b. Etude cas-témoins en France-métropolitaine : étude CATHY 35
  - 1.2.2. Recueil des données épidémiologique et définitions des variables 36

Définition des variables 37

- 1.2.3. Données génétiques 39
- 1.2.3.a. Génotypage de SNP de gènes candidats et des régions 9q22 et 14q13 39
- 1.2.3.b. Génotypage du de nombre de copies de *GSTM1* et *GSTT1* 43
- 1.2.3.c. Données génétiques obtenues par imputation 45
- 1.2.4. Analyses statistiques 49
- 1.2.4.a. Association entre les SNP et le risque de CDT 49

1.2.4.b. Interaction avec les facteurs de risque non génétiques 50
1.2.4.c. Calcul de puissance statistique..... 50
1.2.5. Annotation fonctionnelle des variants 51
1.3. Résultats 53
1.3.1. Caractéristiques des populations d'étude des facteurs de risque environnementaux ou liés au mode de vie des CDT 53
1.3.2. Association entre les polymorphismes candidats et le risque de CDT..... 56
1.3.2.a. Introduction..... 56
1.3.2.b. Résultats 56
1.3.2.c. Synthèse des résultats et comparaison avec la littérature 59
1.3.3. Rôle de la variation du nombre de copie de *GSTM1* et *GSTT1* dans les CDT 61
1.3.3.a. Introduction..... 61
1.3.3.b. Résultats 62
1.3.3.c. Synthèse des résultats et comparaison avec la littérature 68
1.3.4. Fine-mapping des loci 9q22 et 14q13 dans les cancers différenciés de la thyroïde chez les Européens et Mélanésiens de France Métropolitaine et de Nouvelle-Calédonie..... 72
1.3.4.a. Introduction..... 72
1.3.4.b. Résultats 74
1.3.4.c. Synthèse des résultats et comparaison à la littérature..... 84
1.4. Discussion et conclusion..... 88
Partie 2 Etude des facteurs de risque génétique dans les cancers différenciés de la thyroïde selon une approche par pathway candidat 92
2.1. Contexte et objectif..... 92
2.2. Matériel et méthodes..... 95
2.2.1. Populations d'étude 95
2.2.1.a. Etude Young-Thyr 95
2.2.1.b. Etude cas-témoins USRT/UTMDACC 96
2.2.1.c. Standardisation des données épidémiologiques 98
2.2.2. Données génétiques..... 99
2.2.2.a. Données génétiques des études CATHY/Young-Thyr – Puce Beluhca 99
2.2.2.b. Données génétiques étude USRT/UTMDACC..... 104
2.2.3. Analyse statistique 108
2.2.3.a. Analyse par SNP..... 108
2.2.3.b. Analyse par gène et par pathway..... 108
2.3. Résultats 113
2.3.1. Résultats d'analyse des données CATHY/Young-Thyr..... 113
2.3.1.a. Caractéristique de la population d'étude et association avec les facteurs de risque non génétiques 113

2.3.1.b. Association entre les facteurs de risque génétique et les CDT 116
2.3.2. Résultat de l'analyses poolée CATHY/Young-Thyr et USRT/UTMDACC 122
2.3.2.a. Description population d'études et association avec les facteurs de risque non-génétique 122
2.3.2.b. Association entre les facteurs de risque génétique et les CDT 124
2.4. Synthèse des résultats et comparaison avec la littérature 129
Conclusion générale……140
Perspectives……142

Liste des tableaux

Tableau 1: Polymorphismes associés au risque de CDT dans les études GWAS 30

Tableau 2: Liste des SNP sélectionnés pour réplication dans les études CATHY et Nouvelle Calédonie. 40

Tableau 3 : tableau de comparaison des fréquences du nombre de copies de GSTM1 et GSTT1 entre les européens et mélanésiens de notre études et celles disponible dans la littérature. 44

Tableau 4 : Puissance statistique pour mettre en évidence un OR de 0,60 ; 0,80 ; 1,20; 1,30 ou 1,50 avec $\alpha=10^{-3}$ dans nos deux population d'études 51

Tableau 5 : Distribution des cas et témoins des études CATHY et NC selon l'âge, le sexe, la région de résidence et le type histologique de cancer 53

Tableau 6: Facteurs de risque liés au mode de vie et risque de CDT en population générale et chez les femmes dans les populations européennes et mélanésiennes. 55

Tableau 7: Résultats ORs d'association entre les SNP et le risque de CDT chez les européens et les mélanésiens 58

Tableau 8 : Distribution des fréquences des génotypes de GSTM1 et GSTT1 et de la combinaison des deux gènes chez les européens et les mélanésiens et association avec les CDT. 63

Tableau 9: IMC, consommation d'alcool et de tabac et risque de CDT : analyses stratifiées sur les génotypes de GSTM1, GSTT1 et la combinaison des deux gènes 65

Tableau 10 : Facteurs hormonaux et reproductifs et risque de cancer de la thyroïde : analyses stratifiées sur les génotypes de GSTM1, GSTT1 et la combinaison des deux gènes. 67

Tableau 11: OR et IC de l'association entre les tops SNP et SNP GWAS des loci 9q22 et 14q13 chez les européens et les mélanésiens 75

Tableau 12: Résultats d'analyse des combinaisons d'haplotypes les plus fréquents pour les SNP rs944289, rs116909374 et rs999640 dans le locus 14q13 chez les Européens. 78

Tableau 13 : tableau des gènes inclut dans les pathways d'intérêt dans notre étude. 104

Tableau 14 : Tableau résumant le nombre de SNP et gènes des pathways d'intérêt dans l'étude CATHY/Young-Thyr et le pool de données des études CATHY/ Young-Thyr et USRT/UTMDACC et nombre de SNP manquants dans l'étude USRT/UTMDACC. 106

Tableau 15: Puissance pour mettre en évidence des OR entre 1,10 et 1,40 dans l'étude CATHY/Young-Thyr et l'analyse poolée CATHY/Young-Thyr et USRT/UTMDACC pour des SNP ayant un une fréquence allélique entre 0,05 et 0,45 et pour un risque d'erreur α de 10^{-3}. . 112

Tableau 16 : Caractéristiques sociodémographiques des cas et témoins des études CATHY, Young-Thyr et le pool des deux études. 113

Tableau 17: Associations entre les facteurs de risque suspectés et les CDT dans l'étude CATHY, Young-Thyr et l'étude CATHY/Young-Thyr (les ORs ont été ajustés sur l'âge en classe de 10 et le département de résidence) 115

Tableau 18 : top 30 des SNP significativement associés au risque de CDT au seuil 5% avant correction FDR 117

Tableau 19 : p-values d'association entre quelques gènes et le risque de CDT et interaction avec les facteurs de risque non génétique (les gènes présentés sont ceux pour lesquels des interactions avec les facteurs de risque non génétiques sont significatives au seuil 5%) 119

Tableau 20: p-values d'association entre les pathways et le risque de CDT et leurs interactions avec les facteurs de risque non génétiques 120

Tableau 21 : P value résultats d'association entre les pathways et le risque de CDT stratifié sur le statut tabagique, l'âge aux premières règles et la prise de contraception orale dans l'étude CATHY/Young-Thyr 121

Tableau 22: Caractéristiques de la population d'étude et OR d'association avec l'IMC, la consommation d'alcool et de tabac. 123

Tableau 23 : OR et IC à 95% de l'association entre les 30 SNP les plus associés (sur 1031 SNP étudiés) au risque de CDT dans l'analyse poolée et leurs associations correspondantes dans chacune des études CATHY/YOUNG-THYR et USRT/UTMDACC séparément 125

Tableau 24 : p-value d'association entre les gènes les plus associés au risque de CDT dans l'analyse poolée, et leurs associations dans les études CATHY/ Young-Thyr et USRT/UTMDACC séparément. Les gènes présentés sont ceux significatif au seuil 5% avant correction FDR dans l'analyse poolée. 126

Tableau 25: p-value d'association entre les pathways et le risque de CDT dans l'analyse poolée et dans les études CATHY/ Young-Thyr et USRT/UTMDACC 126

Tableau 26 : P-value d'interaction avec l'IMC, la consommation d'alcool et le statut tabagique au niveau des gènes dans l'analyse poolée des études CATHY/Young-Thyr et USRT/UTMDACC. 127

Tableau 27 : Résultats des p-values d'interaction avec l'IMC, le tabac et l'alcool dans l'analyse poolée au niveau pathway 128

Liste des figures

Figure 1 : La glande thyroïdienne ; anatomie (1A) et histologie (1B) 16

Figure 2 : Incidence standardisée cancer de la thyroïde 19

Figure 3 : Principe de l'imputation ; figure tirée de l'article de Marchini et al, Nature Genetic, 2010 (Marchini & Howie, 2010).. 45

Figure 4 : Manhattan plot de l'association entre les SNP du locus 9q22 et le risque de CDT chez les européens (A) et les mélanésiens (B) ainsi que les corrélations entre SNP dans chacune des populations (C et D). 74

Figure 5 : plot des log10 p-values d'association entre les SNP de la région 9q22 et le risque de CDT chez les européens (A) ainsi que les résultats des analyses conditionnelles le SNP rs13295081 (B) et rs1867277 (C). 75

Figure 6: Manhattan plot de l'association entre les SNP du locus 14q13 et le risque de CDT chez les européens (A) et les mélanésiens (B) et les corrélations entre SNP dans chacune des populations (C et D)..... 77

Figure 7 : plot des p-values des résultats d'association entre les SNP de la région 14q13 et les CDT chez les Européens (A) et des résultats d'analyse conditionnelle sur le SNP rs13295081 (B) et rs1867277 (C). 78

Figure 8 : plot des p-values des résultats d'association entre les SNP de la région 14q13 et les CDT chez les mélanésiens (A) ainsi que les résultats d'analyse conditionnelle sur le SNP rs1755774 (B).. 79

Figure 9 : Snapshot de UCSC genome browser (build 37/hg19) indiquant la région du locus 14q13 où sont localisés les SNP rs1755774 82

Figure 10 : Forest plot des interactions gènes-environnement des SNP rs116909374 et rs1755774 chez les européens et les mélanésiens 83

Figure 11: Projection des individus des études CATHY/YOUNG-THYRsur les 2 premières composantes principales. 101

Figure 12 : Projection des individus des études USRT/UTMDACC sur les 2 premières composantes principales..... 106

Figure 13 indique le QQ-plot des p-values d'association avec les CDT pour les 1280 SNP analysés. 116

Figure 14 : QQplot des résultats de p-values d'associations entre les SNP et le risque de CDT dans l'étude CATHY/Young-Thyr. 116

Figure 15 : QQ plot des résultats d'association entre les SNP et le risque de CDT dans les études CATHY/YOUNG-THYR, USRT/UTMADCC et dans l'analyse poolées des études.. 124

Liste des annexes :

Annexe 1 : Méthode d'estimation du nombre de SNP indépendant pour la correction sur les tests multiples. 162

Annexe 2: Annexes des figures 163

Annexe 3: Résultats d'association entre les facteurs environnementaux et le risque de CDT dans les populations initiales des études CATHY et NC 172

Annexe 4: P-value d'interaction entre les variants candidats et les facteurs de risques environnementaux 175

Annexe 5: Association entre le risque de CDT et les acteurs de risque environnementaux stratifié sur les gènes GSTM1 et GSTT1 176

Annexe 6 : ORs et IC à 95% des SNP significativement associés au risque de CDT seuil 5% dans le locus 9q22 chez les européens et les mélanésiens au 178

Annexe 7: OR et IC à 95% des SNP significativement associés au risque de CDT seuil 5% dans le locus 14q13 chez les européens et les mélanésiens. 180

Annexes 8 : P value d'association entre les gènes et le risque de CDT dans l'étude CATHY/Young-thyr et interaction avec les facteurs de risque d'intérêt. Les gènes sont classé selon la plus petite p-value observée pour l'effet propre des gènes 187

Annexes 9 : Resultats des Pvalue d'interaction dans l'analyse poolée..... 191

Liste des Abréviations

ADN : Acide Désoxyribonucléique

CDT : Cancer Différenciés de la Thyroïde

CEPH : Centre d'étude du polymorphisme Humain

CIRC : Centre International de Recherche sur les Cancers

CNMT : Cancer Non Médullaire de la Thyroïde

CNV : Copy Number Variation

CPT : Cancer Papillaire de la thyroïde

eQTL : Expression quantitative trait loci

EST : Expressed Sequence Tag

FT : Facteur de Transcription

GWAS: Genome Wide Association Study

HW : Hardy-Weinberg

IMC : Indice de Masse Corporelle

LincRNA: Long intergenic non-coding RNA

MAF : Minor Allele Frequency

NC : Nouvelle Caledonie

SNP : Single Number Polymorphism

T3 : Triiodothyronine

T4 : Tétraiodothyronine

THM : Traitement Hormonal de la Ménopause

TRH : Thyroid Releasing Hormone

TSH : Thyroid Stimulating Hormone

UTMDACC : University of Texas MD Anderson Case Control Study

USRT : United State Radio Technologist Cohort study

Introduction générale et Objectifs

Introduction générale

1. La glande thyroïdienne

La thyroïde est une glande endocrine située dans la région cervicale médiane basse, devant les cordes vocales et le larynx et en avant de la partie supérieure de la trachée (figure 1A). Elle est formée de deux lobes reliés par un isthme. La glande thyroïde est responsable de la sécrétion des hormones thyroïdiennes qui interviennent dans le développement du système nerveux et du squelette chez le fœtus et l'enfant, et dans la régulation du métabolisme systémique chez l'adulte.

La thyroïde est constituée de lobules eux-mêmes formés de la coalescence de plusieurs follicules (figure 1B). Le follicule est formé : d'un épithélium reposant sur une membrane basale et regroupant différents types de cellules et d'une cavité centrale appelée espace folliculaire ; cette espace est remplie d'une substance semi-visqueuse nommée colloïde. L'épithélium comporte 2 types de cellules :

- Les cellules folliculaires ou thyréocytes (figure 1B) qui incorporent l'iode et synthétisent la triiodothyronine (T3) et la tétraiodothyronine ou thyroxine (T4)) sous le contrôle des hormones de l'axe hypothalamo-hypophysaire (thyréostimuline (TSH) et la thyréolibérine (TRH)).
- Les cellules parafolliculaires ou cellules C (figure 1B) qui sécrètent la calcitonine.

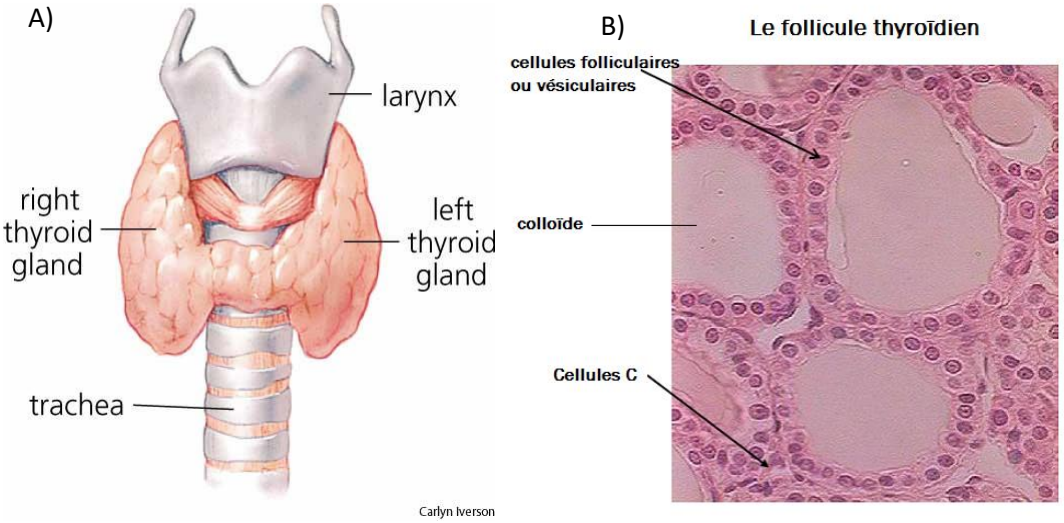


Figure 1 : La glande thyroïdienne ; anatomie (1A) et histologie (1B)

2. Les cancers de la thyroïde

On distingue deux grands types histologiques de cancer de la thyroïde :

- Les cancers des cellules médullaires qui représentent 5% des cancers de la thyroïde, dérivent des cellules C
- Les cancers des cellules folliculaires encore appelés cancers non médullaires de la thyroïde (CNMT) qui représentent 95% des cancers de la thyroïde et sont classés en « cancers non différenciés » ou anaplasiques (10%) et en « cancers différenciés » (90%), qui se subdivisent eux-mêmes en cancers papillaires (80-85%) et cancers vésiculaires (15-20%).

Contrairement aux cancers anaplasiques qui sont très agressifs et associés à un très mauvais pronostic, les cancers différenciés de la thyroïde (CDT) sont des cancers à bon pronostic avec une survie relative globale à 5 ans de plus de 90%. Malgré leur bon pronostic, les CDT sont responsables d'une altération de la qualité de vie des patients et d'un coût financier pour la société qui ne doivent pas être négligés (Singer et al. 2012).

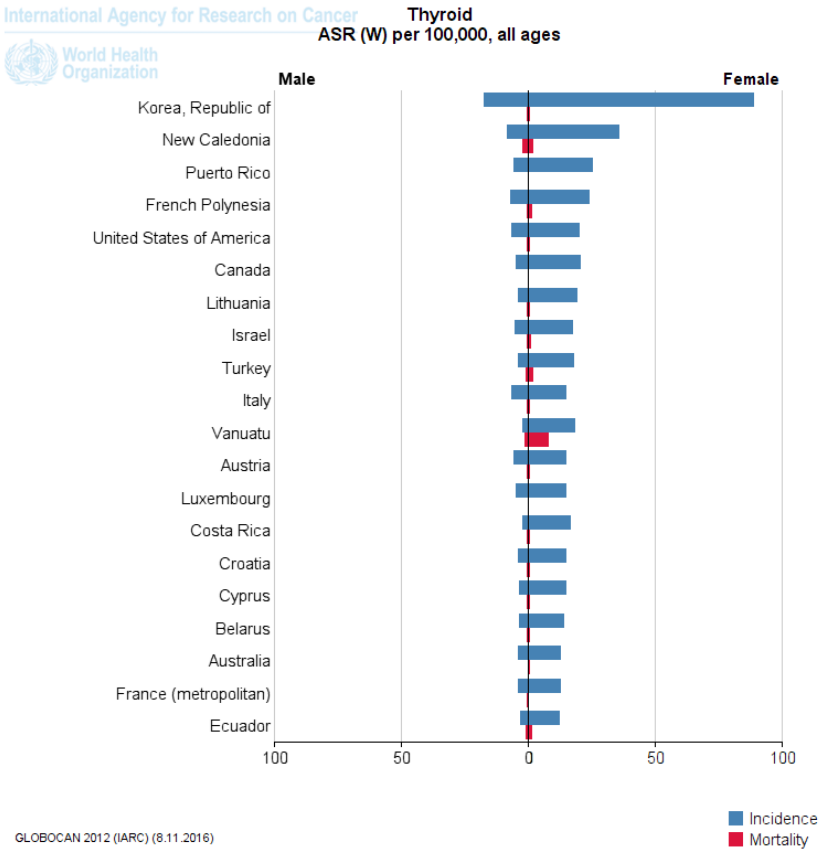
Le diagnostic de cancer de la thyroïde peut reposer sur l'apparition de signes cliniques, par exemple la palpation d'une masse au niveau de la glande thyroïde. Le diagnostic peut aussi être fortuit, lors d'une échographie thyroïdienne ou bien lors de l'examen anatomo-pathologique d'une thyroïde réséquée à cause d'un goitre.

3. Epidémiologie descriptive des cancers de la thyroïde.

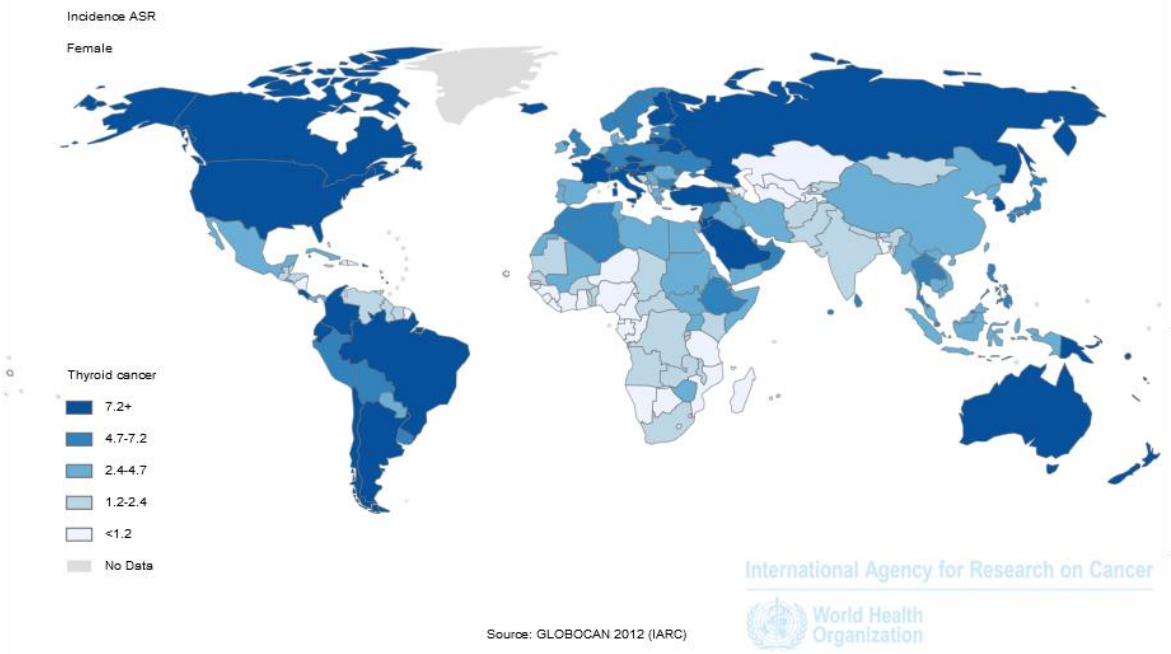
Les cancers de la thyroïde sont les cancers les plus fréquents des glandes endocrines bien qu'ils soient des cancers relativement rares. Ils sont plus fréquents chez les femmes que les hommes avec un sex-ratio entre 3 et 4 pour les formes papillaires et vésiculaires (Kilfoy et al. 2009). Ce ratio est particulièrement plus élevé dans l'adolescence (entre 10 et 19 ans) et décroit ensuite avec l'âge pour attendre un ratio d'environ 1,1 au-delà de 80 ans (Kilfoy et al. 2009). Dans les deux sexes, la plus forte incidence de CT est observée à l'âge adulte au-delà de 30 ans avec un pic d'incidence en général entre 40 et 49 ans chez les femmes et entre 50 et 59 ans chez les hommes (Kilfoy et al. 2009).

L'incidence des CT était d'environ 8,5 pour 100 000 personnes année (PA) en 2012 dans le monde ; soit près de 300 000 nouveaux cas diagnostiqués qui correspondent à environ 2% de l'ensemble des cancers (GLOBOCAN 2012). En France, l'incidence des cancers de la thyroïde était de 13,8 pour 100 000 PA chez les femmes et 5,2 pour 100 000 PA chez les hommes en 2012 (Binder-Foucard et al. 2014).

L'incidence des cancers de la thyroïde est caractérisée par de fortes variations géographiques (figure 2) et ethniques dans le monde. En 2012, les taux d'incidences les plus élevés ont été rapportés en Corée du Sud (88,6/100 000 PA chez les femmes et 17,3/100 000 PA chez les hommes) et en Nouvelle-Calédonie (35,8/100 000 chez les femmes et 8,3/100 000 chez les hommes) (GLOBOCAN 2012) (figure 2a). En Nouvelle-Calédonie, de très forts taux d'incidence ont été observés en particulier dans la population mélanésienne. Une étude a ainsi estimé l'incidence des cancers de la thyroïde à 71,4/100 000 PA chez les femmes mélanésiennes et à 10,4/100 000 PA chez les hommes mélanésiens sur la période 1985-1999 (Truong et al. 2007).



a)



b)

Figure 2 : Incidence standardisée cancer de la thyroïde; a) 20 pays ayant les plus fortes incidences des cancers de la thyroïde en 2012 ; b) cartographie de l'incidence des cancers de la thyroïde chez les femmes dans le monde en 2012

Ces dernières décennies, on note une augmentation des incidences de cancers de la thyroïde dans le monde. En France, l'incidence de CT est passée de 2,8 pour 100 000 PA chez les femmes en 1980 à 13,8 pour 100 000 PA en 2012 avec une augmentation annuelle de 5,1% entre 1980 et 2012 et de 2,7% entre 2005 et 2012 (Binder-Foucard et al. 2014). Chez les hommes, elle était de 1,1 pour 100 000 PA en 1980 et de 5,2 pour 100 000 PA en 2012 avec une augmentation annuelle quasi constante entre 1980 et 2012 soit 5,2 % par an (Binder-Foucard et al. 2014). Cette élévation d'incidence est en partie expliquée par l'évolution des pratiques médicales et de prise en charge des patients notamment avec un taux élevé de « screening » et de diagnostic précoce des micro carcinomes (Kitahara & Sosa 2016; Vigneri et al. 2014; Wiltshire et al. 2016; Vaccarella et al. 2015). Certaines études suggèrent que les changements dans les prévalences des facteurs de risque environnementaux ou liés au mode de vie pourraient également contribuer à l'augmentation de l'incidence des cancers de la thyroïde au moins dans certaines populations (Kitahara & Sosa 2016; Wiltshire et al. 2016).

4. Les facteurs de risque des cancers de la thyroïde

4.1. Les antécédents personnels de pathologie bénigne de la thyroïde

Une forte association a été mise en évidence dans plusieurs études entre les antécédents personnels de goitre multinodulaire ou de nodules thyroïdiens bénins et le risque de cancer de la thyroïde. Les risques relatifs de cancer de la thyroïde associés aux antécédents de goitre sont proches de 5 dans la littérature (Franceschi et al. 1999; Horn-Ross et al. 2001; A Memon et al. 2002; Truong et al. 2005). Les goitres multi-nodulaires pourraient toutefois constituer une pathologie précancéreuse plutôt qu'un facteur de risque à proprement parler du cancer thyroïdien, dont il partagerait les mêmes facteurs étiologiques.

4.2. Expositions aux radiations ionisantes

L'exposition aux radiations ionisantes durant l'enfance est le seul facteur de risque clairement établi dans les CDT (UNSCEAR 2000 Charles, 2001). Les sources d'exposition aux radiations ionisantes incluent les rayonnements ionisants à visée diagnostique et thérapeutique et les retombées nucléaires de l'accident de Tchernobyl ou de la bombe atomique d'Hiroshima et Nagasaki au Japon. Un pic d'incidence des CDT a en effet été observé en Ukraine durant les années suivant l'accident de Tchernobyl (Zablotska et al. 2008). De même, une très forte incidence de CDT a été rapportée au Japon chez les survivants de la bombe atomique d'Hiroshima et Nagasaki et tout particulièrement chez les personnes ayant moins de 30 ans au moment de l'irradiation (Prentice et al. 1982). L'utilisation des radiations à visée thérapeutique répandue jusque dans les années 1960 serait également responsable d'une augmentation d'incidence des CDT observée entre les années 1950 et les années 1980 chez les personnes ayant été exposées pendant l'enfance (Ron et al. 1995).

4.3. Les facteurs anthropométriques

Les caractéristiques anthropométriques les plus fréquemment étudiées dans les cancers de la thyroïde sont la taille, le poids et l'indice de masse corporelle (IMC). Certaines études se sont également intéressées à la surface corporelle, au rapport taille-hanche.

Poids et taille

Le poids et la taille au moment du diagnostic apparaissent indépendamment associés à une augmentation du risque de CDT dans des études de cohortes, les études cas-témoins ainsi que dans des analyses « poolées » de plusieurs études (Guignard et al. 2007; Suzuki et al. 2008; Brindel et al. 2009; Cléro et al. 2010; Xhaard et al. 2015; Rinaldi et al. 2012)

Indice de masse corporelle (IMC)

Plusieurs études cas-témoins et de cohortes ont montré une association entre un indice de masse corporelle élevé au diagnostic et le risque de CDT avec un niveau de risque souvent plus élevé chez les femmes que les hommes (Guignard et al. 2007; Suzuki et al. 2008; Brindel et al. 2009; Cléro et al. 2010; Kitahara et al. 2011; Kitahara et al. 2014; Rinaldi et al. 2012; Xhaard et al. 2015).

Dans une analyse poolée de 5 études prospectives, une relation linéaire entre l'IMC et le risque de CDT a été rapportée avec un HR de 1,17 [1,10 – 1,24] pour une augmentation de 5 kg/m² de l'IMC (Kitahara et al. 2011). Plus récemment, une méta-analyse incluant 32 études cas-témoins et de cohorte a également conclu à une relation positive entre l'obésité et le risque de CDT avec un RR global de l'obésité vs poids normal estimé à 1,33 [1,24-1,42] (Ma et al. 2015). Enfin, un rapport récent du Centre International de Recherche sur les Cancers (CIRC) considère qu'il existe un niveau de preuve suffisant pour conclure à l'existence d'une relation entre le surpoids et le risque de survenue de CDT chez l'Homme, avec un RR estimé à 1,1 [1,0 – 1,1] pour une augmentation de 5 unité d'IMC (Secretan et al. 2016).

4.4. Les facteurs hormonaux et reproductifs

L'incidence plus élevée des CDT chez les femmes que chez les hommes et le pic d'incidence observée chez les femmes au cours de la période reproductive ont conduit à suspecter un rôle des changements hormonaux liés par exemple aux cycles menstruels, à la grossesse, à la ménopause ou encore à l'utilisation de traitements hormonaux dans l'étiologie des CDT. On connaît par ailleurs le rôle des œstrogènes sur la prolifération, la migration et l'invasion des cellules cancéreuses de la thyroïde.

i. Les facteurs menstruels

Les facteurs menstruels le plus souvent étudiés dans la littérature sont la régularité des menstruations, l'âge aux premières règles, le statut ménopausique, l'âge à la ménopause, et les antécédents de chirurgie gynécologique. La régularité des règles, l'âge aux premières règles et l'âge à la ménopause sont des indicateurs de la durée d'exposition aux hormones ovariennes. Le statut ménopausique et les antécédents d'ovariectomie correspondent à une modification des conditions hormonales endogènes.

Age aux premières règles

L'âge tardif aux premières règles a été associé à une faible augmentation de risque de CDT dans la plupart des études cas-témoins (Negri et al., 1999; Truong et al., 2005) alors que les études de cohortes ne rapportent généralement pas d'association (Meinhold et al., 2010; Zamora-Ros et al., 2015) à l'exception d'une étude de cohorte réalisée en 2011 chez des enseignantes de Californie aux Etats-Unis (Horn-Ross et al., 2011). Deux méta-analyses récentes n'ont cependant pas mis en évidence une association entre les CDT et un âge tardif aux premières règles (Cao et al. 2015; Caini et al. 2015).

Irrégularité des règles

Très peu d'études se sont intéressées à l'association entre l'irrégularité des règles et les CDT avec des résultats peu concluants. Alors que certains auteurs ont rapporté une augmentation du risque de CDT chez les femmes ayant déclaré avoir toujours eu des règles irrégulières (Rossing et al. 2000; Truong et al. 2005), d'autres auteurs n'ont montré aucune association significative (Brindel et al. 2008; Horn-Ross et al. 2011).

Le statut ménopausique et l'âge à la ménopause

Une analyse poolée de 14 études cas témoins publiée en 1999, avait montré une association positive entre la ménopause naturelle et le risque de CDT comparé aux femmes toujours réglés (Negri et al. 1999). Cette association n'a cependant pas été retrouvé dans d'autres études (Zamora-Ros et al. 2015; Horn-Ross et al. 2011; Meinhold et al. 2010) ; deux méta-analyses récente n'ont également pas montré une association entre la ménopause naturelle et le risque de CDT (Cao et al. 2015; Caini et al. 2015). La ménopause artificielle suite

à une ovariectomie bilatérale a par contre été associée à une augmentation du risque de CDT dans plusieurs études (Negri et al. 1999; Brindel et al. 2009; Zamora-Ros et al. 2015).

Les résultats sur l'âge à la ménopause et le risque de CDT rapportés dans la littérature sont discordants. Dans une méta-analyse récente portant sur 24 études incluant des études cas-témoins et des études de cohortes, un âge tardif à la ménopause était associé à une augmentation de risque de CDT (Cao et al. 2015). Cette association n'était pas significative dans une autre méta-analyse (Caini et al. 2015) incluant uniquement des études de cohortes prospectives (21 études).

ii. Les facteurs reproductifs

Parmi les facteurs reproductifs, la plupart des études se sont intéressées au nombre de grossesses à terme et à l'âge à la première grossesse. D'autres facteurs tel que l'allaitement, le délai depuis la dernière grossesse, les fausses couches et les interruptions volontaires de grossesses ont également été étudiés mais les résultats ne sont pas concluants.

La parité a été le plus souvent associée à une augmentation de risque de CDT dans les études cas-témoins (Truong et al. 2005; Xhaard et al. 2014; Brindel et al. 2008; Negri et al. 1999) mais pas dans les études de cohortes (Horn-Ross et al., 2011; Meinhold et al., 2010; Zamora-Ros et al., 2015). Une parité élevée a été associée à un risque de CDT avec des associations qui devenaient significatives au-delà de 4 grossesses à terme (Truong et al. 2005; Brindel et al. 2008; Anjum Memon et al. 2002). Une méta-analyse récente a montré une association positive entre la parité et le risque de CDT mais aucune relation dose-effet n'a été mise en évidence (Zhu et al. 2016).

iii. Les traitements hormonaux

Les contraceptifs oraux ont été intensivement investigués mais les résultats ne sont pas concluants. Une analyse poolée de 13 études cas-témoins rapporte une augmentation du risque de CDT chez les femmes ayant pris une contraception orale (La Vecchia, 1999). Des études plus récentes ne montrent pas d'association (Brindel et al., 2008; Horn-Ross et al., 2011; Truong et al., 2005; Xhaard et al., 2014; Zamora-Ros et al., 2015), alors qu'une méta-

analyse portant sur 9 études de cohortes prospectives suggère une association inverse entre la prise de contraception et le risque de CDT (Wu & Zhu 2015).

Concernant les traitements hormonaux de la ménopause (THM), plusieurs études ont montré une association inverse entre la prise de THM et le risque de CDT (Negri et al. 1999; Truong et al. 2005) alors que d'autres n'ont pas trouvé d'association (Meinhold et al., 2010; Xhaard et al., 2014; Zamora-Ros et al., 2015). Deux méta-analyses récentes n'ont cependant pas conclu à une association entre les THM et le risque de CDT (Caini et al. 2015; Cao et al. 2015).

4.5. La consommation d'alcool et de tabac

Plusieurs études ont examiné la relation entre le tabagisme et la consommation d'alcool et le risque du cancer de la thyroïde. La plupart de ces études rapportent une association inverse avec la consommation d'alcool (Cho & Kim, 2014; Kabat, Kim, Wactawski-Wende, & Rohan, 2012; Kitahara et al., 2012; Meinhold et al., 2009; Sen et al., 2015). Récemment, une méta-analyse basée sur 33 études a montré une association inverse entre la consommation d'alcool et le risque de CDT avec un RR de 0,75 [0,69–0,83] pour les plus fortes consommations par rapport aux plus faibles consommations (Hong et al. 2016). Cette association persistait lorsque les analyses étaient stratifiées sur le type d'étude, le type histologique de cancer et le sexe (Hong et al. 2016).

La relation entre la consommation de tabac et le risque de CDT reste peu claire et les résultats d'études épidémiologiques sont hétérogènes. Plusieurs études ont rapporté une association inverse entre la consommation de tabac et le risque de CDT (Kabat et al. 2012; Belin et al. 2004; Soldin et al. 2009; Guignard et al. 2007; Meinhold et al. 2010) ainsi qu'une relation dose-effet inverse avec l'intensité du tabagisme (Kitahara, Linet, et al. 2012; Kreiger & Parkes 2000; Nagano et al. 2007), la durée (Rossing et al. 2000; Kreiger & Parkes 2000) et le nombre de paquets années consommés (Rossing et al. 2000; Kitahara, Linet, et al. 2012). D'autres études n'ont pas rapporté d'association entre le tabac et le risque de CDT (Iribarren, Haselkorn et al. 2001; Navarro Silvera, Miller et al. 2005). Une méta-analyse publiée en 2014 à partir de 31 études observationnelles incluant des études cas-témoins et de cohortes, a montré une association inverse entre les fumeurs actuels et le risque de CDT comparé aux

non-fumeurs (Cho & Kim 2014). Cette association inverse était particulièrement marquée lorsque l'analyse était restreinte aux études cas-témoins (Cho & Kim 2014).

4.6. Apport alimentaire en iode

Le rôle de l'iode alimentaire dans la carcinogénèse du cancer de la thyroïde a été évoqué dans plusieurs études (Truong et al. 2010; Horn-Ross et al. 2011; Zimmermann & Galetti 2015; Ren, Juan Jesus Lence-Anta, et al. 2016) sans que cette relation n'ait été clairement mise en évidence. Il a été suggéré qu'un excès d'iode alimentaire pouvait être associé à une augmentation de risque de cancer papillaire (Kolonel et al. 1990) alors qu'un déficit en iode pouvait plutôt être associé au risque de cancer vésiculaire (Zimmermann & Galetti 2015). Une revue récente souligne que les études expérimentales confirment l'association positive entre le déficit en iode et le risque de cancers vésiculaire, mais ne confirment pas la relation avec l'excès d'iode alimentaire (Zimmermann & Galetti, 2015).

4.7. Antécédent familiaux des CDT

Des études de cohorte historiques à partir de bases de données nationales dans l'Utah, en Islande et en Norvège ont montré que l'existence d'un antécédent familial de cancer de la thyroïde chez les apparentés de premier degré était associé à un risque relatif entre 3 et 8 (Goldgar, et al. 1994; Kari Hemminki, et al, 2005; K Hemminki & Vaittinen, 1999; Weber & Eng, 2005). Le cancer de la thyroïde constitue une des localisations cancéreuses dont le risque familial est le plus élevé (Czene et al. 2002). L'augmentation du risque de cancer associée aux antécédents familiaux peut refléter l'existence de facteurs de risque génétiques communs, un même mode de vie et/ou des expositions environnementales communes entre les membres d'une même famille. Dans une étude menée en Suède à partir du registre national des familles et du registre national des cancers, et reposant sur l'utilisation de modèles structuraux pour expliquer les associations entre apparentés, la part des cancers de la thyroïde attribuable à des facteurs génétiques était estimée à 53%, et était beaucoup plus importante que pour les autres localisations (Czene et al. 2002). Ces résultats suggèrent un rôle important des facteurs de risque génétiques.

Synthèse sur les facteurs de risque non génétique

Les résultats des études épidémiologiques et expérimentales ont montré que l'exposition aux radiations ionisantes particulièrement durant l'enfance est un facteur de risque établi de CDT. Des résultats consistants ont également été rapportés pour les antécédents familiaux de CDT. L'obésité est également considérée comme un facteur de risque établi des CDT.

Un déficit alimentaire en iode semble être associé à une augmentation du risque de CDT. Les études épidémiologiques vont plutôt dans le sens d'une association inverse avec la consommation d'alcool et de tabac.

Les résultats concernant les facteurs hormonaux et reproductifs sont plus hétérogènes. La plupart des études s'accordent sur une augmentation de risque de CDT chez les femmes ayant une parité élevée. Un âge tardif aux premières règles et le statut ménopausique ont été positivement associés au risque de CDT dans certaines études alors que des associations inverses ont été rapportées dans d'autres. Pour les traitements hormonaux exogènes plusieurs études épidémiologiques ont montré une association inverse entre la prise de contraceptifs oraux ou de traitements hormonaux de la ménopause (THM) et le risque de CDT ; ces résultats n'ont cependant pas été confirmés dans des méta-analyses récentes. D'autres études sont nécessaires pour mieux comprendre le rôle des hormones stéroïdiennes dans l'étiologie des CDT et les prédispositions plus fortes observées chez les femmes par rapport aux hommes.

5. Facteurs génétiques

Les facteurs génétiques sont surtout connus pour les cancers médullaires de la thyroïde dont la plupart font partie des néoplasies endocriniennes multiples de type 2 (NEM 2A/B), des affections héréditaires particulièrement rares liées à des mutations germinales du proto-oncogène RET (Weber & Eng 2005). La composante génétique dans les cancers non médullaires est encore peu connue.

5.1. Les gènes identifiés dans les cas familiaux

Seuls 5 à 10% des CDT sont d'origine familiale; la grande majorité des cancers étant sporadiques. Les études de liaisons génétiques portant sur des familles avec plusieurs cas de CDT ont mis en évidence quelques loci candidats de susceptibilité, notamment les loci 14q31 (Bignell et al. 1997), 19q13.2 (Canzian et al. 1998), 1q21 (Malchoff et al. 2000) et 2q21 (McKay et al. 2001) qui pourraient contenir des gènes de prédisposition. Ensuite d'autres études de liaison pan-génomique portant sur de plus larges familles de cas ont identifié 3 autres loci associés au risque de CDT : 8p23-p22, 8q24 et 12q14 (He et al. 2009; Cavaco et al. 2008). Ces résultats n'ont cependant pas été uniformément répliqués et aucun gène à forte pénétrance n'a été identifié dans ces études.

5.2. Les gènes identifiés selon les approches gène candidat

La plupart des études cas-témoins qui se sont intéressées aux facteurs génétiques dans le risque de CDT ont utilisé une approche gène candidat, en ciblant les gènes ayant une fonction *a priori* pertinente dans l'étiologie des CDT ou impliqués dans des voies biologiques d'intérêt. Les fonctions biologiques les plus étudiées sont : le maintien de la stabilité du génome incluant la réparation de l'ADN et la régulation du cycle cellulaire, du fait de leurs rôles dans la réparation des dommages cellulaires induits par les radiations ionisantes (Ron et al. 1995; Williams 2002); la régulation des signaux kinase-dépendants ; le métabolisme et de la détoxification des substances endogènes et exogènes et le métabolisme des hormones stéroïdiennes. D'autres études se sont plutôt intéressées aux polymorphismes qui ont été retrouvés dans des mutations tumorales de CDT ou dans d'autres types de cancers. Les résultats de ces études sont hétérogènes et très peu de polymorphismes ont été répliqués dans des études indépendantes.

Parmi les polymorphismes les plus étudiés et répliqués dans au moins deux études, on peut citer les polymorphismes rs861539 dans *XRRC3* (Sturgis et al. 2005; Bastos et al. 2009) ; rs25487 et rs1799782 dans *XRRC1* (Chiang et al. 2008; Ho et al. 2009; Akulevich et al. 2009) ; rs1042522 du gène suppresseur de tumeurs *TP53 ;* plusieurs polymorphismes dans *ATM* dont rs1801516 initialement associé au risque de CPT radio-induit (Akulevich et al. 2009) ; le polymorphisme rs1867277 ainsi que l'expansion en poly-alanine dans l'unique exon de *FOXE1* (Landa et al. 2009; Kallel et al. 2010) et enfin, les polymorphismes rs732609 et rs2048722 dans le gène *TPO* (Cipollini et al. 2013). Ces polymorphismes sont situés pour la plupart dans les parties codantes des gènes et sont associés à un changement d'acide aminé ou de site de fixation de facteurs de transcription pouvant être à l'origine d'une modification de l'expression des gènes.

XRRC1, *XRRC3*, *TP53* et *ATM* sont des gènes de la réparation de l'ADN et de la régulation du cycle cellulaire ; ils jouent un rôle important dans la réparation des cassures double brin de l'ADN dues à une exposition aux radiations ionisantes. *FOXE1* est un facteur de transcription spécifique à la thyroïde ; il joue un rôle important dans le développement et la différenciation de la glande thyroïdienne ainsi que dans la régulation de la transcription d'autres gènes spécifique à la thyroïde comme *TPO* et *TG* (Landa & Robledo 2011). *TPO* est une peroxydase thyroïdienne ayant un rôle essentiel dans la biosynthèse des hormones thyroïdiennes. Des mutations dans ce gène sont les causes les plus fréquentes d'hypothyroïdie congénitale associée à une déficience en synthèse des hormones thyroïdiennes (Avbelj et al. 2007).

Plusieurs études se sont également intéressées à la délétion complète des gènes *GSTM1* et *GSTT1* et les génotypes du gène *NAT2* dans les CDT mais les résultats des études sont très hétérogènes et non concluants (Hernández et al. 2003; Canbay et al. 2003; Gaspar et al. 2004; Bufalo et al. 2006; Kweon et al. 2014; Siraj et al. 2008; Landa & Robledo 2011; Guilhen et al. 2009; Morari et al. 2002). *GSTM1, GSTT1* et *NAT2* sont des gènes impliqués dans le métabolisme de phase 2 des composés endogènes et exogènes incluant les carcinogènes ; ces gènes, du fait de leurs fonctions, ont été suspectés dans l'étiologie de plusieurs cancers.

Parmi les polymorphismes non répliqués, on peut notamment citer : les réarrangements du proto-oncogène RET fréquemment retrouvés dans les mutations tumorales de CPT d'origine sporadiques ou radio-induites (Lesueur et al. 2002; Lönn et al. 2007) ; les

polymorphismes des gènes *CYP1A1* et *CYP2D6* qui sont des gènes impliqués dans la phase 1 du métabolisme (Siraj et al. 2008; Bufalo et al. 2006). Plusieurs autres variants des gènes *ERCC2, MTF1, GSTP1, TSHR, RAD52* et *EGFR* identifiés dans d'autres cancers n'ont pas été associés au risque de CDT.

Ces résultats ont été décrits dans une revue de littérature publiée en 2011 (Landa & Robledo 2011).

5.3. Les études GWAS dans les CDT

Les études pan génomiques ou GWAS (« Genome Wide association study ») ont permis d'identifier des régions chromosomiques associées aux CDT.

Depuis 2009, cinq GWAS ont été réalisées sur les CDT. La première (Gudmundsson et al. 2009), menée en Islande en 2009 chez 378 cas des cancers papillaires de la thyroïde (CPT) et 37 196 témoins avait mis en évidence deux variants : rs965513 (OR = 1.75 (IC à 95% : 1,59 – 2,05), $p = 1,7 \times 10^{-27}$) situé sur le locus 9q22.33 en amont de *FOXE1* et rs944289 (OR = 1,37 (IC à 95% : 1,24 – 1,52), $p=1,0x10^{-9}$) situé sur le locus 14q13.3 près de *NKX2-1*. Comme *FOXE1, NKX2*-1 est un facteur de transcription spécifique à la thyroïde qui joue un rôle important dans le développement et la différenciation de la thyroïde ainsi que la régulation de la fonction thyroïdienne.

La deuxième étude GWAS menée en 2010 sur une population biélorusse ayant été exposée aux radiations ionisantes lors de l'accident de Tchernobyl (187 cas et 172 témoins âgés de moins de 18 ans lors de l'accident) a également montré une association entre le risque de CDT et le polymorphisme rs965513 (9q22) (OR = 1,65 (IC à 95% : 1,43 – 1,91), $p = 5,9 \times 10^{-9}$) mais pas le polymorphisme rs944289 (14q13) (Takahashi et al. 2010). Ce GWAS avait également répliqué le variant rs1867277 dans *FOXE1* sur locus 9q22 et précédemment mis en évidence dans une étude gène candidat (Landa et al. 2009).

La troisième étude GWAS menée en 2012 chez 27 758 individus islandais avait identifié 22 loci associés aux taux circulant de TSH (Thyroid Stimulating Hormone). Ces 22 loci ont ensuite été testés en association avec le risque de CPT (sur environ 530 cas et 3000 témoins), confirmant l'association avec rs965513 (locus 9q22) et rs944289 (locus 14q13) et révélant 3 nouveaux polymorphismes : rs966423 sur le locus 2q35 proche de *DIRC3* (OR = 1,26 (IC à 95% : 1,11 – 1,43), $p = 3,8 \times 10^{-4}$); rs2439302 sur le locus 8p12 près de *NRG1* (OR (95% IC) = 1.41 (1,23

– 1,62), p = 1,3 x 10^{-6}) et rs116909374 sur le locus 14q13.3 près de *NKX2-1* (OR (95% IC) = 2,03 (1,54 – 2,67), p = 5,4 x 10^{-7}) (Gudmundsson et al. 2012).

La quatrième étude GWAS (Köhler et al. 2013), menée en 2013 sur une population italienne à forte incidence de CDT incluaient 690 cas et 497 témoins. En plus de la réplication des polymorphismes de 9q22, cette étude a suggéré de nouveaux polymorphismes à risque de CDT près des gènes *DICR3* (2q35), *IMMP2L* (7q21), *RARRES1* (3q25) et SNAPC4/CARD9 (9q34) (Tableau 1).

Enfin, la cinquième étude GWAS, réalisée sur une population espagnole de 398 cas de et 502 témoins (Mancikova et al. 2015), avec un set de réplication comprenant 1422 cas et 1908 témoins provenant de 3 études espagnoles et 1 étude italienne. Cette étude GWAS a identifié de nouveaux locus dans les chromosomes 9q22, 6q14 et 10q26 (tableau 1) en plus de la réplication des variants rs965513 et rs1867277 du locus 9q22 et rs944289 du locus 14q13 précédemment mis en évidence par les GWAS (Mancikova et al. 2015).

Tableau 1: Polymorphismes associés au risque de CDT dans les études GWAS

Loci	SNP	Gène(s)	Etude[a]	N Ca/N Co[b]	Al[c]	OR[a]	IC à 95 % [a]	p-val[a]	Rep[c]
9q22	rs965513	*FOXE1*	Gudmundsson et al, 2009 (1)	962/38923	A	1,75	(1,59 - 1,94)	1,10E-27	2, 3, 4, 5
14q13	rs944289	*NKX2-1*	Gudmundsson et al, 2009 (1)	1006/38345	T	1,37	(1,24 - 1,52)	2,00E-09	4, 5
2q35	rs966423*	*DIRC3*	Gudmundsson et al, 2012 (3)	1150/41448	C	1,34	(1,22 - 1,47)	1,30E-09	
8p12	rs2439302	*NRG1*	Gudmundsson et al, 2012 (3)	1134/5625	G	1,36	(1,23 - 1,50)	2,00E-09	
14q13	rs116909374	*MBIP,NKX2-1*	Gudmundsson et al, 2012 (3)	1138/5340	T	2,09	(1,68 - 2,60)	4,60E-11	
2q35	Rs6759952*	*DIRC3*	Köhler et al, 2013 (4)	2131/2068	T	1,30	(1,18 - 1,43)	7,30E-08	
3q35	rs7617304	*RARRES1*	Köhler et al, 2013 (4)	2139/2068	A	1,25	(1,12 - 1,39)	4,00E-05	
7q21	rs7800391	*IMMP2*	Köhler et al, 2013 (4)	2134/2043	T	1,25	(1,14 - 1,38)	5,7E-06	
7q21	rs10238549	*IMMP2*	Köhler et al, 2013 (4)	2142/2059	C	1,27	(1,15 - 1,40)	4,10E-6	
9q34	rs1078500	*SNAPC4*	Köhler et al, 2013 (4)	2126/2053	T	1,23	(1,12 - 1,36)	3,5E-05	
6q14	rs4075570	*HTR1B*	Mancikova et al, 2015 (5)	502/398	A	0,68	(0,51 - 0,80)	9,60E-05	
9q22	rs7028661	*FOXE1*	Mancikova et al, 2015 (5)	2409/1819	A	1,64	(1,49 - 1,81)	1,00E-22	
9q22	rs10122541	*FOXE1*	Mancikova et al, 2015 (5)	2409/1819	C	1,54	(1,40 - 1,70)	1,10E-17	
9q22	rs7037324	*FOXE1*	Mancikova et al, 2015 (5)	2409/1819	A	1,54	(1,39 - 1,70)	1,20E-17	
10q26	rs2997312	*WDR11-AS1*	Mancikova et al, 2015 (5)	1998/1464	A	1,35	(1,16 - 1,57)	1,20E-04	
10q26	rs125467	*WDR11-AS1*	Mancikova et al, 2015 (5)	2409/1817	C	1,38	(1,18 - 1,62)	5,90E-05	

(1) Gudmundsson et al, 2009 ; (2)Takahashi et al, 2009 ; (3)Gudmunsson et al, 2012 ; (4)Köhler et al, 2013 ; (5)Mancikova et al, 2015. [a] étude ayant identifié le SNP ; [b] N ca/N Co : nombre de cas/nombre de témoins dans l'étude initiale ; [c]Al: allèle à risque ; [d]Rep : réplication (études ayant répliqué l'association avec le SNP) ; *rs966423 et rs6759952 sont situé le même locus et sont en déséquilibre de liaison.

Au final, près de 18 polymorphismes dans 12 loci différents ont été associés au risque de CDT par des études GWAS (tableau 1). Ces résultats supportent l'hypothèse selon laquelle les prédispositions génétiques dans les CDT résulteraient d'une combinaison de plusieurs allèles à faible pénétrance. Une étude publiée en 2013 a montré que l'effet combiné des polymorphismes rs965513 (locus 9q22), rs944289 et rs116909374 (locus 14q13), rs966423 (locus 2q35), rs2439302 (locus 8p12) identifiés dans les études GWAS jusqu'en 2013 expliquerait 11% du risque familial de CPT dans une cohorte de l'Ohio et 6% du risque dans

une cohorte polonaise (Liyanarachchi et al. 2013). De même, une étude publiée en 2015 a montré que, l'ensemble des 12 SNP rapportés dans des études GWAS ou de réplication rapportés à partir de populations italiennes (Köhler et al. 2013; Figlioli et al. 2014) expliquerait environ 5% du risque de CDT dans cette population (Figlioli et al. 2015). Ces résultats suggèrent que d'autres variants de susceptibilité génétique dans les CDT restent à découvrir. Nous pouvons également noter que les études GWAS ont identifié des polymorphismes dans des régions de susceptibilité génétique, mais leur rôle causal n'est pas démontré. Il est possible que ces signaux soient corrélés à des variants causaux qu'il reste à découvrir.

5.4. Les interactions gènes-environnement dans les CDT

Les études d'interactions gène-environnement dans les cancers de la thyroïde ont surtout porté sur l'exposition aux radiations ionisantes. Une étude GWAS (Takahashi et al. 2010) avait évalué si l'association entre les CDT et les polymorphismes de *FOXE1* était modulé selon l'exposition aux radiations ionisantes mais n'avait pas observé d'interaction significative (Takahashi et al. 2010). Dans une autre étude, menée dans une population exposée aux retombées d'essais nucléaires, l'exposition aux rayonnements ionisants était associé au risque de cancer thyroïdien seulement en présence de l'allèle commun du polymorphisme rs1799782 situé sur *XRCC1* (Sigurdson et al. 2009). De même dans une étude menée sur une population exposée aux radiations ionisantes lors de l'accident de Tchernobyl (Akulevich et al. 2009), le variant rs1042522 du gène *TP53* avait été associé aux cancers de la thyroïde radio-induits. Plus récemment, une étude réalisée sur la population polynésienne a rapporté une interaction entre le polymorphisme GWAS rs944289 (locus 14q13) et la dose de radiations ionisantes reçues avant l'âge de 15 ans (Maillard et al. 2015).

Les interactions entre les génotypes nul/non-nul (délétions) des gènes *GSTM1* et *GSTT1* et l'âge, le sexe, l'IMC et la consommation de tabac ont été étudiées mais les résultats rapportés ne sont pas concluants (Canbay et al. 2003; Kweon et al. 2014). Plus récemment, d'autres auteurs ont évalué les interactions avec l'IMC, la consommation d'iode et la parité dans différentes populations pour les polymorphismes rs1801516 dans *ATM*, rs1867277 dans *FOXE1* et les SNP GWAS rs944289 et rs965513 (Ren, Juan J. Lence-Anta, et al. 2016; Pereda et al. 2015; Maillard et al. 2015) et ont rapporté une interaction significative entre le polymorphisme rs965513 et la consommation d'iode dans la population européenne (Ren,

Juan Jesus Lence-Anta, et al. 2016). Ces résultats n'ont pour le moment pas été répliqués dans d'autres études.

Objectifs de la thèse

L'objectif de ce travail de thèse était d'approfondir les connaissances sur l'étiologie des CDT en étudiant le rôle des facteurs de risque génétiques et de leurs interactions avec les facteurs de risque environnementaux ou liés au mode de vie dans les CDT.

La première partie de ma thèse vise à répliquer les associations entre le CDT et les polymorphismes identifiés dans les études gènes candidats ou dans des études GWAS à partir de deux études cas-témoins menées dans des contextes génétiques et environnementaux très contrastés : la Nouvelle-Calédonie et la France métropolitaine. Dans cette partie, l'analyse porte sur une population d'origine européenne d'une part, et sur une population d'origine mélanésienne d'autre part dans laquelle les taux d'incidence sont beaucoup plus élevés. Les gènes étudiés içi sont sélectionnés parmi les variants GWAS et les gènes candidats associés au cancer de la thyroïde dans au moins deux études jusqu'en 2012. Nous nous sommes intéressés plus particulièrement aux régions 9q22 et 14q13 identifiés par plusieurs études GWAS, pour lesquelles nous avons essayé d'identifier des polymorphismes qui pourraient être de meilleurs candidats à la causalité dans les cancers de la thyroïde en utilisant une approche dite par « fine-mapping ». Cette analyse vise également à rechercher si des facteurs génétiques peuvent expliquer la différence d'incidence observée entre ces deux populations

La deuxième partie de ma thèse vise à identifier de nouveaux facteurs de risque génétiques dans les CDT en étudiant des ensembles de gènes impliqués dans des voies biologiques d'intérêt, telles que celles liées au métabolisme et à la détoxification des xénobiotiques et des hormones stéroïdiennes. Ces analyses, dites « pathways candidats », sont menées dans des populations d'origine européenne de France Métropolitaine (études CATHY et YOUNG-THYR) et des Etats-Unis (études USRT et UTMDACC).

Partie 1 : Etude des gènes candidats et des régions identifiées par les études GWAS dans les cancers de la thyroïde

Partie 1 Etude des gènes candidats et des régions identifiées par les études GWAS dans les cancers de la thyroïde

1.1. Introduction

Dans cette partie, nous nous intéressons au rôle de gènes candidats dans l'étiologie des CDT dans les populations d'origine européenne et mélanésienne.

Dans un premier temps, nous avons étudiés les polymorphismes associés au risque de CDT dans au moins deux études, les polymorphismes identifiés dans des études fonctionnelles ou liés à l'obésité qui constitue un facteur de risque reconnu des CDT. Dans un second temps, nous nous sommes intéressés à la relation entre le nombre de copies des gènes *GSTT1* et *GSTM1* et les CDT. Enfin, nous avons exploré les loci 9q22 et 14q13 identifiés par les GWAS, en utilisant une approche dite par « fine-mapping » afin d'identifier des variants pouvant être des meilleurs candidats à risque de CDT.

La plupart des études qui se sont intéressées aux facteurs de risque génétiques dans les CDT ont porté sur des populations d'origine européenne, alors qu'aucune n'a porté jusqu'à présent sur la population mélanésienne de NC, où l'incidence des CDT est particulièrement élevée. Dans ce travail, nous avons également cherché à savoir si les polymorphismes précédemment associés au CDT dans des populations européennes étaient aussi associés au CDT chez les mélanésiens.

1.2. Matériel et méthodes

1.2.1. Populations d'étude

1.2.1.a. Etude cas-témoins en Nouvelle-Calédonie

Sélection des cas

Les cas sont des individus ayant un cancer papillaire ou vésiculaire de la thyroïde diagnostiqué entre le 1er janvier 1993 et le 31 décembre 1999, et résidant en Nouvelle-Calédonie depuis au moins 5 ans au moment du diagnostic. Les cas ont été identifiés de manière exhaustive, sur la base du diagnostic anatomopathologique, par les deux laboratoires d'anatomopathologie et le registre des cancers de Nouvelle-Calédonie. Sur 369 cas de cancer de la thyroïde éligibles pour cette étude, 37 (10%) n'ont pas pu être interrogés : 9 ont refusé de participer, 21 étaient décédés au moment des entretiens et 5 n'ont pas été contactés. Au final, l'étude a porté sur 332 cas dont 244 cas d'origine mélanésienne et 42 cas d'origine européenne.

Sélection des témoins

Les témoins ont été sélectionnés à partir des listes électorales de Nouvelle-Calédonie et été appariés par fréquence aux cas sur le sexe et l'âge (par classes de 5 ans). Une année de référence, équivalente de l'année de diagnostic chez les cas, a été attribuée aux témoins. Sur 473 témoins éligibles, 61 (13 %) n'ont pas été inclus dans l'étude : 13 étaient décédés, 24 ont refusé de participer et 24 n'ont pas pu être contactés. Au final l'étude a porté sur 412 témoins dont 189 témoins d'origine mélanésienne et 133 témoins d'origine européenne.

1.2.1.b. Etude cas-témoins en France-métropolitaine : étude CATHY

L'étude CATHY est une étude cas-témoins en population générale conduite entre 2005 et 2007 dans trois départements: le Calvados, la Marne et les Ardennes.

Sélection des cas

Les cas éligibles pour l'étude étaient l'ensemble des patients avec un CDT diagnostiqué entre le 1er janvier 2002 et le 31 décembre 2007, résidant dans l'un des trois départements concerné et âgés d'au moins 25 ans au moment du diagnostic. Ils ont été identifiés par le registre général des cancers du Calvados et le registre spécialisé des cancers de la thyroïde de

Marne-Ardennes. Sur 853 cas éligibles, 232 (27%) n'ont pas pu être interrogés : 148 ont refusé de participer, 29 étaient décédés au moment de l'entretien, 52 n'ont pas pu être contactés, et 3 étaient trop malades pour répondre. L'étude a porté sur 621 cas dont 594 cas d'origine européenne.

Sélection des témoins

Les témoins ont été sélectionnés par tirage au sort par l'institut de sondage CSA parmi les sujets âgés de 25 ans ou plus l'année du diagnostic des cas et résident dans les départements du Calvados, de la Marne et des Ardennes. Le tirage au sort était réalisé à partir des annuaires téléphoniques, sur la base de numéros de téléphone incrémentés d'une unité afin de pouvoir contacter également les personnes sur liste rouge. Les témoins ont été appariés par fréquence aux cas sur le sexe et l'âge quinquennal. Le tirage au sort des témoins était stratifié sur la catégorie socioprofessionnelle de manière à ce que la répartition des catégories socioprofessionnelles parmi les témoins soit représentative de celle de leur département de résidence. Une année de référence entre 2002 et 2007, correspondant à l'année de diagnostic des cas appariés, était attribuée aux témoins. Sur 943 cas éligibles, 236 (25%) n'ont pas pu être interrogés : 232 ont refusé de participer et 4 n'ont pas pu être contactés. L'étude a porté sur 706 témoins dont 690 témoins d'origine européenne.

1.2.2. Recueil des données épidémiologique et définitions des variables

Dans chacune des études, les sujets étaient interviewés à l'aide d'un questionnaire standardisé lors d'entretiens en face-à-face à domicile réalisés par des enquêtrices spécialement formées pour l'étude. Les données recueillies concernaient les informations sociodémographiques, les antécédents personnels et familiaux de pathologies de la thyroïde, les caractéristiques de la vie menstruelle et reproductive, les traitements hormonaux exogènes, les caractéristiques anthropométriques, les habitudes alimentaires, la consommation de tabac et d'alcool, les loisirs, les intitulés de la profession et du secteur d'activité, et l'historique résidentiel ainsi que l'origine ethnique. Dans les études de NC et CATHY, l'origine ethnique était déclarée par le sujet lui-même. Seules les expositions antérieures au diagnostic pour les cas, et à l'année de référence pour les témoins, ont été considérées dans les analyses.

Dans le cadre de l'étude des facteurs de risque génétiques dans les CDT, une banque d'ADN a été créée. La collecte d'ADN a été effectuée au moment de l'enquête pour l'étude CATHY et a posteriori pour l'étude de NC, en procédant soit à un prélèvement de salive (KIT Oragène®), soit à un recueil de cellules buccales à partir de brossettes. L'ADN a été obtenu chez 231 cas et 270 témoins de l'étude de NC, dont 169 cas et 126 témoins d'origine mélanésienne, et chez 517 cas et 568 témoins de l'étude CATHY.

Définition des variables

Pour l'étude des interactions gène-environnement, nous nous sommes intéressés aux facteurs suivants : l'indice de masse corporelle (IMC), la consommation d'alcool et de tabac et certain facteurs hormonaux et reproductifs tels que l'âge aux premières règles, la parité, le nombre de grossesses à terme, la prise de contraceptifs oraux, le statut ménopausique et la prise de traitement hormonal de la ménopause. Les informations concernant ces variables ont été standardisées entre les études CATHY et NC. Nous présentons ici en détail les variables utilisées dans les analyses.

i. Indice de masse corporelle

L'IMC était défini comme le rapport du poids (kg) à l'année de référence sur le carré de la taille (m). Nous l'avons ensuite regroupé en plusieurs catégories selon la définition de l'OMS: IMC<18.5 kg/m² (maigreur) ; IMC de 18,5 à 24,99 kg/m² (normal) ; IMC de 25 à 29,99 kg/m² (surpoids) et IMC≥30 kg/m² (obésité). Pour les tests d'interaction, nous avons regroupé les catégories d'IMC en 2 classes (<25 kg/m² et ≥25 kg/m²) ou en 3 classes (<25 kg/m² ; entre 25 – 29,99 kg/m² et ≥30 kg/m²).

ii. Consommation d'alcool

Dans les questionnaires, les individus étaient interrogés sur leur consommation de vin, champagne, cidre, bière, whisky et pastis, apéritifs (boissons à 20°) et digestifs (boissons à 40°). Pour chaque type d'alcool, la fréquence (« jamais ou presque », nombre de fois par jour, semaine ou mois) et la quantité consommée (en nombre de verres par consommation) ont été recueillies. Le nombre de verres d'alcool a été calculé sans distinction du type d'alcool en considérant que chaque verre de ces différents types d'alcool contient approximativement le même nombre de grammes d'alcool.

iii. Consommation de tabac

Les informations recueillies étaient le statut tabagique, l'âge au début du tabac, la quantité moyenne consommée, le temps depuis l'arrêt, et le nombre total d'années d'arrêt. D'autre part, le type de tabac consommé (cigarette, tabac à rouler, cigares, pipes ou autres) était également renseigné. A partir de ces informations, nous avons défini :

- Le statut tabagique codé en non-fumeur, ancien fumeur et fumeur actuel. Un individu était considéré ancien fumeur s'il avait arrêté de fumer depuis au moins 1 an à l'année de référence.
- Le nombre de paquets-années de cigarettes fumées calculé à partir des informations sur l'âge de début du tabagisme, sa durée, et le délai écoulé depuis l'arrêt du tabagisme. Le nombre de paquets-années a été estimé pour chaque individu en multipliant la durée de sa consommation en années par le nombre de paquets de cigarettes fumées par jour.

iv. Facteurs hormonaux et reproductifs

Les caractéristiques de la vie menstruelle et reproductive qui ont été étudiées sont : l'âge aux premières règles, le statut ménopausique, la parité et le nombre de grossesses à terme, la prise de contraceptifs oraux et la prise de traitements hormonaux de la ménopause (THM).

Une femme était considérée comme ménopausée si ses dernières règles dataient de 12 mois ou plus à la date de référence ; si elle avait subi une ovariectomie bilatérale totale ; si elle prenait un THM ; si elle avait subi une hystérectomie avant arrêt naturelle des règles et était âgée de 50 ans ou plus si la date des dernières règles était inconnue et qu'elle était âgée de 50 ans ou plus.

Une femme était considérée comme pare si elle avait eu une grossesse d'aux moins 24 semaines (6 mois) quel que soit l'issue de la grossesse. Le nombre de grossesse à terme (>6 mois) était également renseigné.

Pour la prise de contraception orale ou de THM, nous avons étudié le fait d'avoir déjà ou jamais utilisé chacun des traitements avant la date de diagnostic ou de référence.

1.2.3. Données génétiques

Les sujets de l'étude CATHY et de NC chez qui, un prélèvement salivaire était disponible, ont été génotypés à l'aide de deux puces : une puce ciblant des SNP de gènes candidats et des SNP dans les régions 9q22 et 14q13 ; et une puce permettant de déterminer le nombre de copies des gènes *GSTT1* et *GSTM1*. Nous avons ensuite imputé des SNP dans les régions 9q22 et 14q13 afin d'avoir une couverture plus fine de ces régions.

Je décris ci-dessous le génotype et le contrôle qualité pour chacune des deux puces. Ensuite je détaillerai la stratégie d'imputation dans les régions 9q22 et 14q13 qui a permis de générer des données génétiques supplémentaires.

1.2.3.a. Génotypage de SNP de gènes candidats et des régions 9q22 et 14q13

Nous avons conçu une puce génétique de 96 SNP regroupant 38 SNP de gènes candidats, 11 tags-SNP (incluant les SNP rs965513 et rs1867277) du locus 9q22 (entre le SNP GWAS rs965513 et le gène *FOXE1*) et 47 tags SNP (incluant les SNP rs944289 et rs116909374) du locus 14q13 (entre les gènes *PTCSC3* et *NKX2-1*).

Les SNP des gènes candidats sélectionnés pour réplication dans notre étude sont rapportés dans le Tableau 2; ils ont été sélectionnés de la manière suivante :

- Polymorphismes associés au risque de CDT dans au moins deux études (Landa & Robledo 2011), en 2012 soit 10 SNP incluant les SNP GWAS rs2439302 du locus 8q12 et rs966423 du locus 2q35 ainsi que 7 SNP identifiés dans les études de liaisons pan-génomique.
- Polymorphismes associés au risque de CDT dans une seule étude mais pouvant être impliqués dans la survenue de CDT selon des études fonctionnelles (10 SNP dont 5 SNP dans le gène ATM).
- Polymorphismes des gènes de l'obésité mis en évidence dans les études du consortium PAGE (Fesinmeyer et al. 2013) (« Population Architecture using Genomics and Epidemiology ») et répliqués dans d'autres études (9 SNP). Nous nous sommes intéressés à ces polymorphismes du fait du rôle avéré de l'obésité dans le risque de CDT.

Tableau 2: Liste des SNP sélectionnés pour réplication dans les études CATHY et Nouvelle Calédonie.

SNP	Gene	Position	Fonction	Données littérature
Variants répliqués dans au moins 2 études				
rs6983267	*gene desert*	8q24		Landa et al. 2011
rs1042522	*TP53*	17p13	p.Arg72Pro	Landa et al. 2011
rs2910164	*pre-miRNA-146a*	5q24		Landa et al. 2011
rs861539	*XRCC3*	14q32	p.Thr241Met	Landa et al. 2011
rs25487	*XRCC1*	19q13	p.Arg399Gln	Landa et al. 2011
rs1799782	*XRCC1*	19q13	p.Arg194Trp	Landa et al. 2011
rs966423		2q35		SNP GWAS (Gudmunson et al. 2012)
rs2439302		8p12		SNP GWAS (Gudmunson et al. 2012)
rs732609	*TPO*	2	exon12	Cipollini et al 2013
rs2048722	*TPO*	2		Cipollini et al 2013
Variants non répliqués mais avec fonction biologique pertinente dans les CDT				
rs1133076	*TG*	8q24	p.Arg2530Gln	Landa et al. 2011
rs4752904	*PTPRJ*	11p11.2	p.Asp872Gln	Landa et al 2011
rs4903957	*TSHR*			Associé au taux circulant de TSH (Arnaud-Lopez et al, 2008)
rs2252696	*TG*	8q24		TSH(Arnaud-Lopez et al 2008) + familial PTC (He et al 2009)
rs1801516	*ATM*	11q22	Asp1853Asn	Landa et al. 2011
rs664677	*ATM*	11q22	IVS22-77 T>C	Landa et al. 2011
rs1800057	*ATM*	11q22	P1054R	Xu et al 2012
rs189037	*ATM*	11q22	promoteur	Xu et al 2012
rs3092993	*ATM*	11q22		Damiola et al. 2014
rs3764340	*WWOX*	16q23		Cancemi et al. 2011
SNP des cas familiaux de PTC				
rs6885099	*PDE8B*	5q13		TSH (Arnaud-Lopez et al. 2008)
rs1445493		8q23		Cas familiaux (Cavaco et. al 2008)
rs822318		8p22		Cas familiaux
rs9942754	*SLA/TG*	8q24		Genome-wide linkage analysis (He et al. 2009)
rs2702968	*SLA/TG*	8q24		Genome-wide linkage analysis (He et al, 2009
rs2702967	*SLA/TG*	8q24		Genome-wide linkage analysis (He et al. 2009)
rs17695552	*SLA/TG*	8q24		Genome-wide linkage analysis (He et al. 2009)
rs2741205	*SLA/TG*	8q24		Genome-wide linkage analysis (He et al. 2009)
Gènes obésité (Fesinmeger et al 2013)				
rs9939609	*FTO*	16		
rs10767664	*BDNF*	11		
rs7498665	*SH2B1*	16		
rs6548238	*TMEM18*	2		
rs2815752	*NEGR1*	1		
rs17782313	*MC4R*	18		
rs12970134	*MC4R*	18		
rs11084753	*KCTD15*	19		
rs13130484	*GNPDA2*	4		
rs10838738	*MTCH2*	11		

Afin de mieux couvrir les régions 9q22 et 14q13 identifiés par les études GWAS, des tags SNP ont été sélectionnés de manière à représenter au moins 80% de la variabilité génétique dans chacune des régions ; les SNP sélectionnés devaient avoir une fréquence de l'allèle mineur d'au moins 0,05 et une corrélation d'au moins 0,8 avec les autres tags SNP du locus

dans la population européenne de HapMap. Le design et la sélection des SNP a été effectué avec le logiciel Tagger software (Howie et al. 2011)).

Le génotypage de la puce (puce « Fluidigm Dynamic array » à 96 SNP) a été effectué par la société Integragen.

i. Contrôle qualité des SNP

Les amorces (« oligonucleotide primers ») utilisées pour le génotypage des SNP ont dans été vérifiées sur un échantillon de 90 sujets CEPH pour lesquels le génotypage est connu et publié dans HapMap et dans le projet 1000 génomes. Les génotypes observés avec ces amorces avaient un taux de concordance d'au moins 98 % avec les génotypes publiés pour ces SNP chez les 90 sujets CEPH.

Le taux de génotypage ou « call rate » de tous les SNP génotypés était supérieur à 95%. Nous avons également testé la déviation à l'équilibre d'Hardy Weinberg (HW) pour chaque SNP séparément chez les européens et chez les mélanésiens.

La loi d'équilibre d'HW stipule qu'au sein d'une population théorique idéale, les fréquences alléliques et génotypiques d'un polymorphisme sont distribuées de manière équilibrée d'une génération à l'autre. Les fréquences génotypiques sont déterminées à partir des fréquences alléliques par la relation suivante : dans le cas d'un polymorphisme à deux allèles A de fréquence p et deux allèles a de fréquence (1-p), nous avons une fréquence attendue de p^2 pour le génotype AA, 2p(1-p) pour Aa et $(1-p)^2$ pour le génotype aa. Un écart trop important à l'équilibre d'Hardy-Weinberg peut être la conséquence de problème de génotypage. Le test d'équilibre de HW correspond à par un test de chi2 à 1 degré de liberté comparant les génotypes observés aux attendus sous l'équilibre de HW.

Nous avons corrigé le test de l'équilibre d'HW sur le nombre de tests effectués en appliquant la méthode de Bonferroni. Les SNP ayant une p-value inférieure à $5x10^{-4}$ (sur la base de 96 tests effectués) pour le test d'équilibre d'HW étaient ainsi exclus de l'analyse. Dans nos deux populations, tous les SNP génotypés respectaient l'équilibre d'HW. Aucun des 96 SNP, n'a donc été exclu de l'analyse.

ii. Contrôle qualité des sujets

Parmi les 231 cas et 270 témoins de l'étude Nouvelle-Calédonie, et 517 cas et 568 témoins de l'étude CATHY pour lesquels nous avions des prélèvements ADN, 64 cas et 65 témoins de l'étude CATHY n'ont pas été génotypés du fait d'une trop faible concentration d'ADN.

Les sujets ayant un taux de génotypage ou « call rate » < 90% (10 cas et 2 témoins de l'étude CATHY et 9 cas ainsi que 9 témoins de l'étude de NC) ainsi que les sujets s'étant déclarés d'origine non européenne ou non mélanésienne ont été exclu de l'analyse (28 cas et 12 témoins dans l'étude CATHY, 32 cas et 59 témoins dans l'étude de NC).

Au total, 481 cas et 544 témoins d'origine européenne de l'étude CATHY ainsi que 183 cas et 194 témoins de l'étude de NC (dont 156 cas /112 témoins d'origine mélanésienne et 27 cas /82 témoins d'origine européenne) sont inclus dans les analyses. Les fréquences alléliques des SNP des témoins d'origine européenne de l'étude CATHY et de l'étude de NC étant similaires (annexe 10), nous avons regroupé tous les sujets européens de ces études.

Au final, les analyses ont donc porté sur 508 cas et 629 témoins d'origine européenne provenant des études CATHY et NC, et 156 cas et 112 témoins d'origine mélanésienne provenant de l'étude de NC (Figure 3).

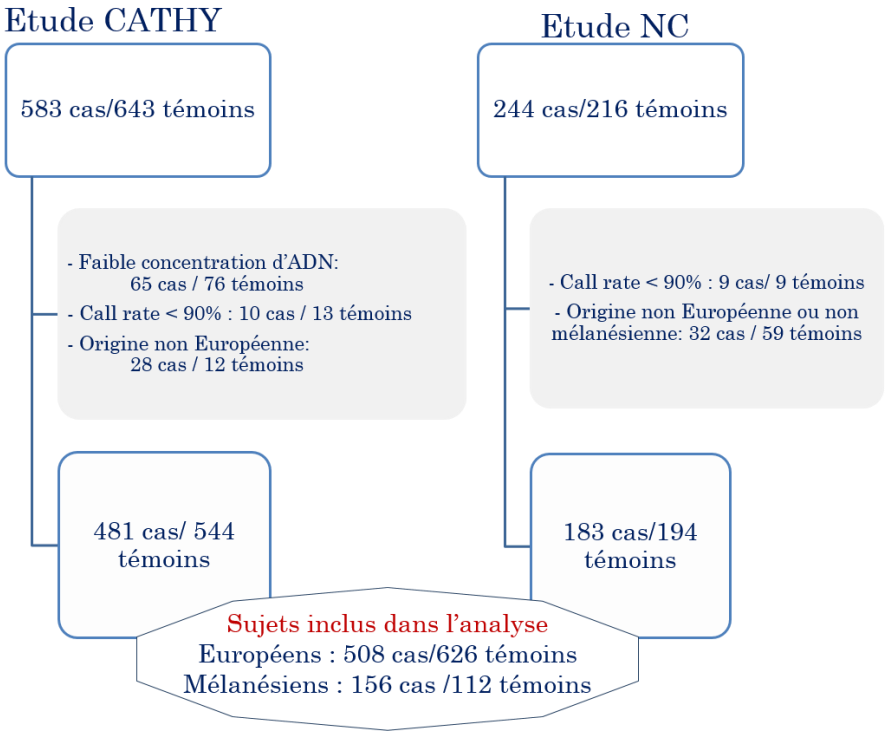


Figure 3 : Flow chart du contrôle qualité des sujets de l'étude CATHY et NC.

1.2.3.b. Génotypage du de nombre de copies de *GSTM1* et *GSTT1*

Le nombre de copies des gènes *GSTM1* et *GSTT1* a été déterminé par la méthode de PCR quantitative à temps réel ou « real-time PCR » qui permet de quantifier l'ADN présent dans une réaction grâce aux fluorescences émises par la sonde d'amplification. La fluorescence est utilisée pour mesurer l'accumulation des produits de la PCR après chacun de ses cycles. Les mesures de fluorescence permettent d'extrapoler la quantité d'ADN cible présent dans l'échantillon avant l'amplification ; la détection et la quantification de l'ADN sont ainsi obtenues dans la même réaction. La PCR quantitative a été réalisée sur le système rapide 7900HT conçu par Applied Biosystem avec comme gène de référence l'RNase P et les amorces spécifiques à *GSTM1* (sonde Hs02575461_CN) et *GSTT1* (sonde Hs00010004_CN). L'utilisation d'un gène de référence permet de définir la concentration du gène cible dans l'échantillon.

Chaque échantillon a été exécuté en triplicat avec 50ng d'ADN génomique. Avant génotypage, les amorces spécifiques de *GSTM1* et *GSTT1* avaient d'abord été validées sur 90 sujets CEPH pour lesquels le nombre de copies de ces gènes est connu. La quantification du nombre de copies de chaque gène a été effectuée avec le logiciel de CopyCaller V1 d'Applied Biosystems.

i. Contrôle qualité

65 cas et 76 témoins de l'étude CATHY n'ont pas pu être génotypés du fait d'une faible concentration de l'ADN. Après le génotypage, nous avons effectué un contrôle qualité et exclu 68 sujets de l'étude CATHY (30 cas et 38 témoins) ainsi que 15 sujets de l'étude de NC (9 cas et 6 témoins) pour les raisons suivantes :

- Sujets ayant un génotype indéterminé pour *GSTM1* et *GSTT1* (51 sujets dans l'étude CATHY et 17 sujets dans l'étude de NC).
- Sujets ayant un nombre de copie des deux gènes > 4 (17 sujets dans l'étude CATHY et 1 sujet dans l'étude de NC) car les duplications de *GSTM1* et *GSTT1* sont rare dans la population humaine et un nombre de copie du gène > 4 est généralement considéré comme aberrant dans la littérature (Rose-Zerilli et al. 2009; Rudolph et al. 2012).
- Sujets ayant un nombre de copie de *GSTM1* ou de *GSTT1* >4 ont été codées en valeur manquante pour le gène correspondant

Les sujets d'origine non européenne et non mélanésienne ont également été exclus des analyses soit 3 cas et 4 témoins dans l'étude CATHY ainsi que 35 cas et 59 témoins de l'étude de NC.

Les fréquences du nombre de copies de *GSTM1* et *GSTT1* chez les témoins de notre population européenne étaient concordantes avec celles de la littérature (Tableau 3) (Rose-Zerilli et al. 2009; Rudolph et al. 2012; Huang et al. 2009). Aucune étude n'avait jusque-là été menée dans la population mélanésienne. Cependant, la fréquence des sujets de génotype nul pour *GSTM1* dans notre population mélanésienne était également comparable à celle précédemment décrite dans une population océanienne (Huang et al. 2009; Rebbeck 1997) (Tableau 3).

Tableau 3 : tableau de comparaison des fréquences du nombre de copies de GSTM1 et GSTT1 entre les européens et mélanésiens de notre études et celles disponible dans la littérature.

	Européens								**Mélanésiens**							
	GSTM1				***GSTT1***				***GSTM1***				***GSTT1***			
	Karageorgi et al., 2011		Européens CATHY		Karageorgi et al., 2011		Européens CATHY		Rebbek et al., 1997		Mélanésiens NC		Rebbek et al., 1997		Mélanésiens NC	
CNV	count	freq	count	freq	count	freq	count	freq	count	freq	count	freq	count	freq	count	freq
0	621	0,53	252	0,50	239	0,20	104	0,21	-	0,64-1,00	94	0,80	-		39	0,39
1	442	0,38	205	0,41	526	0,45	225	0,45	-	-	16	0,14	-		44	0,44
2/3	114	0,10	40	0,09	401	0,34	155	0,32	-		4	0,04	-		16	0,17

Etant donné les fréquences similaires du nombre de copies de *GSTM1* et *GSTT1* chez les témoins d'origine européenne de l'étude CATHY et de NC, Nous avons combiné les deux populations d'études (annexe 11). Au final, les analyses ont porté sur 508 cas et 626 témoins d'origine européenne ainsi que 156 cas et 114 témoins d'origine mélanésienne.

Pour les analyses statistiques, nous avons modélisé *GSTM1* et *GSTT1* sous forme de :

- nombre de copies du gène (CNV) en 3 classe : les individus ayant au moins deux copies du gène (homozygote non-nul) ; ceux ayant une copie du gène (hétérozygote) et ceux ayant 0 copie du gène (homozygote nul)
- présence ou non du gène (génotype nul/non-nul) où nous avons comparé les individus ayant au moins une copie du gène (génotype non-nul) à ceux n'ayant aucune copie du gène (génotype nul).

- combinaison des génotypes des deux gènes (*GSTM1*/*GSTT1*) où nous avons comparé les individus ayant au moins une copie des deux gènes à ceux ayant soit l'un des deux gènes, soit aucun des deux gènes.

1.2.3.c. Données génétiques obtenues par imputation

L'imputation est une technique d'inférence des génotypes manquants d'un ou de plusieurs SNP en utilisant : les informations génétiques de ces SNP ; les échantillons d'haplotypes de populations de référence (telles que HapMap ou 1000 génomes), et les génotypes obtenus sur les SNP voisins dans l'échantillon. L'imputation permet d'augmenter la couverture génétique dans l' échantillon étudié et ainsi permettre de caractériser plus finement des régions chromosomiques d'intérêt ; elle permet également de combiner les données générées sur des puces différentes et de faciliter les méta-analyses (Marchini & Howie 2010).

Principe et méthodologie de l'imputation

Dans un échantillon de données de personnes apparentées, les haplotypes des individus sont liés les uns aux autres sur des séquences d'ADN plus ou moins longues du fait de l'identité par descendance. Les méthodes d'imputation sont basées sur l'identification des haplotypes dans ces régions partagées entre les individus de l'échantillon d'étude et les individus de la population de référence pour prédire les génotypes manquants. Le principe de l'imputation est représenté sur la figure ci-dessous (Figure 4).

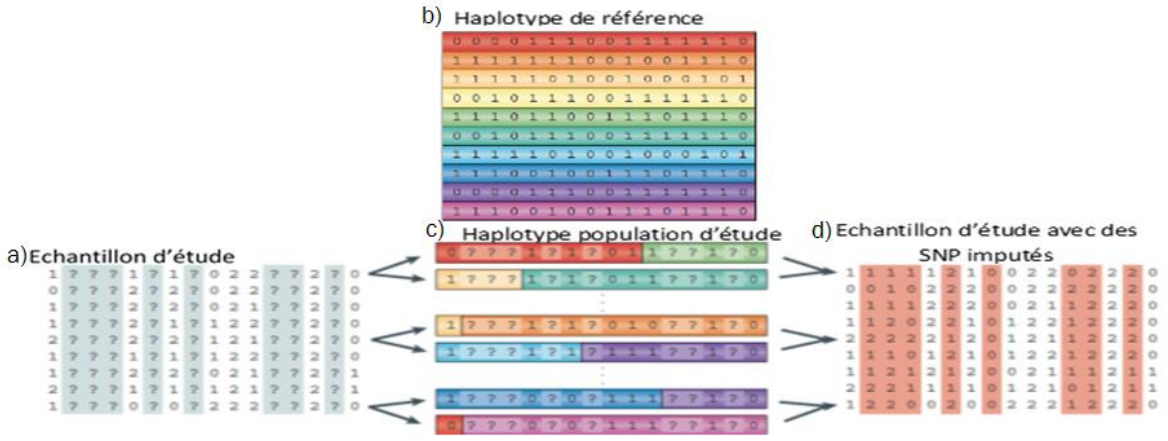


Figure 4: Principe de l'imputation ; figure tirée de l'article de Marchini et al, Nature Genetic, 2010 (Marchini & Howie 2010). a) l'échantillon d'étude représente la base de donnée génétique intérêt avec

des génotypes manquants, b) set d'halotype de la population de référence, c) haplotype de la population d'étude est estimé à partir de l'échantillon d'étude selon la même mosaïque que celle de la population de référence, d) l'haplotype de référence est utilisé pour imputer les allèles manquants dans l'échantillon d'étude et créer ainsi des génotypes (rouge foncé).

Plusieurs logiciels d'imputations ont été proposés comme MACH (Li et al. 2009), BEAGLE et IMPUTE2 (Howie et al. 2009) ; ces logiciels utilisent les méthodes principalement basées sur des chaines de Markov cachées développées pour la modélisation des déséquilibres de liaisons et des taux de recombinaisons. Le génotype manquant G_{ij} du SNP i chez un individu j est prédit à partir d'un modèle probabiliste qui s'exprime comme suit :

$$P_{ijk} = \Pr(G_{ij} = k), k \in \{0,1,2\}, \sum_j P_{ijk} = 1$$

K représente le nombre de copies des deux allèles de chaque SNP codés en 0 et 1. Le génotype final prédit pour le SNP i chez l'individu j est donné par l'estimation du nombre de copie de l'allèle de référence comme suit :

$$D_{ij} = 2 \times P_{ij2} + 1 \times P_{ij1} + 0 \times P_{ij0}$$

Les valeurs D_{ij} sont appelées « doses alléliques » et sont comprises entre 0 et 2. Une mesure de l'évaluation de la certitude de l'imputation est donnée après imputation par les logiciels. Elle correspond au rapport de la variance des probabilités de génotypes obtenues à l'imputation sur la variance des génotypes attendues sous l'équilibre d'Hardy Weinberg et varie entre 0 et 1. Cet indicateur est en général utilisé pour le contrôle qualité des génotypes imputés. Par exemple, pour le logiciel IMPUTE2, il est appelé « INFO ».

Pour le fine-mapping des loci 9q22 et 14q13, nous avons imputé les SNP non génotypés dans chacune des régions en utilisant le logiciel IMPUTE2 avec comme panel de référence, les génotypes de toutes les populations du projet 1000 génomes (« 1000 genomes integrated phase 1 version 3 released September 2013 »). Le panel cosmopolitain est recommandé car il permet au logiciel d'utiliser le panel de référence le plus adapté pour chaque individu permettant ainsi d'améliorer la précision de l'imputation (Howie et al. 2011), et est particulièrement recommandée pour l'imputation dans une population non référencée comme la population mélanésienne (Guan & Stephens 2008; Howie et al. 2011).

L'imputation a porté sur un intervalle de 70 kb dans le locus 9q22 entre rs965513 et *FOXE1* (bornes 100,55x10^6 – 100,65x10^6 ; build37) et sur un intervalle de 400kb dans le locus 14q13 entre *PTCSC3* et *NKX2-1* (bornes 36,6x10^6 – 37,0x10^6; build37). Seuls les SNP ayant une MAF>0,02 dans les populations de référence étaient imputés. Nous avons effectué l'imputation séparément chez les européens et mélanésiens et sur chaque locus.

Contrôle qualité de l'imputation

Le contrôle qualité était stratifié sur l'origine ethnique comme l'imputation. Pour évaluer la qualité de l'imputation, nous avons estimé la concordance entre les génotypes prédit par l'imputation et les génotypes réels. Pour ce faire, nous avons masqué les génotypes de SNP pris au hasard, soit 2 SNP dans le locus 9q22 et 4 SNP dans le locus 14q13 et imputé ces génotypes dans notre échantillon. Nous avons ensuite estimé le taux de concordance de l'imputation pour chacun des SNP par le rapport de la somme des génotypes correctement prédit sur la somme des génotypes initiaux. Un génotype était considéré comme correctement prédit lorsque la probabilité à postériori de l'imputation pour le SNP i chez un individu j (P_{ij}) était >0.90. Pour les 6 SNP, le taux de concordance était >90% dans les deux populations d'études.

Après imputation, nous avons obtenu 70 SNP supplémentaires dans la région 9q22 et 514 SNP supplémentaires dans la région 14q13.

Nous avons ensuite effectué un contrôle qualité post-imputation qui consistait à supprimer les SNP dont le score de qualité d'imputation était inférieur à 0.70 (score info<0.7). Cela correspondait pour le locus 9q22, à 12 SNP chez les européens et 8 SNP chez les mélanésiens ; et pour le locus 14q13, à 35 SNP chez les européens et 49 SNP chez les mélanésiens. Nous avons également vérifié le respect de l'équilibre d'HW pour tous les SNP imputés par un test de chi2 à 2 ddl. Pour la correction sur le nombre de tests effectués, nous avons appliqué la méthode de Bonferroni sur la base de 115 tests indépendants estimés par le logiciel GEC sur l'ensemble des 724 SNP imputés et génotypés (M.-X. Li et al. 2012) ; la méthode d'estimation du nombre de test indépendant est décrite dans l'annexe 1. Le seuil de significativité considéré après prise en compte de tests multiples était de 4,4x10^{-4} ; tous les SNP respectaient l'équilibre d'HW.

Au final, le nombre de SNP imputés inclus dans les analyses était:

- pour les locus 9q22 : 58 SNP dans la population européenne et 62 SNP dans la population mélanésienne
- pour le locus 14q13 : 479 SNP dans la population européenne et 465 SNP dans la population mélanésienne

Synthèse des données génétiques

Au final, les analyses de cette partie de ma thèse portent sur :

- 36 SNP candidats génotypés chez 508 cas et 626 témoins d'origine européenne et chez 156 cas et 112 témoins d'origine mélanésienne.
- Le nombre de copies de *GSTM1* et *GSTT1* disponible chez 573 cas et 584 témoins d'origine européenne et chez 156 cas et 115 témoins d'origine mélanésienne.
- Les SNP génotypés et imputés dans les loci 9q22 et 14q13 pour le fine mapping de ces régions chez les sujets d'origine européenne (508 cas et 626 témoins) et mélanésienne (156 cas et 112 témoins), soit :
    - ✓ 69 SNP dans la population européenne et 75 SNP dans la population mélanésienne pour le locus 9q22.
    - ✓ 525 SNP dans la population européenne et 511 SNP dans population mélanésienne pour le locus 14q13.

1.2.4. Analyses statistiques

1.2.4.a. Association entre les SNP et le risque de CDT

i. Effet propre de chaque polymorphisme (« single SNP association »)

Les OR et intervalles de confiance de l'association entre chaque polymorphisme et le risque de CDT ont été estimés par régression logistique non-conditionnelle avec ajustement sur l'âge, le sexe et la région de résidence (Marne, Ardennes, Calvados et Nouvelle Calédonie) pour l'analyse chez les européens des études CATHY et NC, et province de résidence (soit Nord, Sud et îles Loyauté) pour la population mélanésienne de l'étude de NC) en utilisant un modèle log-additif. Pour les données d'imputation, les SNP sont codés en « dose allélique » avec les valeurs continue compris entre 0 et 2. Toutes les analyses ont été stratifiées sur l'origine ethnique déclarée.

ii. Régression conditionnelle

Dans les analyses de fine-mapping, on recherche le meilleur candidat dans une région où les SNP sont en fort déséquilibre de liaison. Pour identifier les signaux d'association indépendants dans les loci 9q22 et 14q13, nous avons effectué des analyses conditionnelles qui consiste à analyser les signaux secondaires d'association en prenant en compte le signal prédominant ou le meilleur candidat dans la région. Les SNP significatifs lors de l'analyse conditionnelle sur un SNP particulier sont considérés comme des signaux indépendants de celui-ci

iii. Analyse Haplotypique

A partir des SNP indépendamment associés au risque de CDT dans chaque locus et dans chacune des populations, nous avons effectué des combinaisons d'haplotype et testé leurs associations avec le risque de CDT avec le package Haplo.Stat du logiciel R.

iv. Correction sur les tests multiples

Les p-values d'associations entre les SNP et le risque de CDT ont été corrigées sur les tests multiples. Des seuils différents ont été appliqués pour les différentes parties des résultats. Pour les analyses portant sur les variants candidats, nous avons appliqué le seuil de $1x10^{-3}$ correspondant à une correction de Bonferroni basée sur le nombre de tests effectués soit 36 tests. Concernant l'analyse des variants génotypés et imputés des régions 9q22 et 14q13, le seuil de significativité considéré était $p<4,4x10^{-4}$ en considérant une correction de Bonferroni basée sur une estimation de 116 tests indépendants (M.-X. Li et al. 2012).

1.2.4.b. Interaction avec les facteurs de risque non génétiques

Pour tester les interactions entre les facteurs de risque génétique et les facteurs environnementaux, nous avons utilisé la méthode du rapport de vraisemblance comparant le modèle avec interaction et le modèle sans interaction. Les facteurs de risque d'intérêt pour l'étude des interactions étaient l'indice de masse corporelle, la consommation d'alcool et de tabac et les facteurs hormonaux et reproductifs dont l'âge aux premières règles, la prise de contraceptifs oraux, la parité, le statut ménopausique et la prise de traitement hormonaux. Nous ne nous sommes pas intéressés aux radiations ionisantes dans notre étude car les populations étudiées n'ont pas été exposées à de fortes doses de radiations ionisantes d'origine environnementale.

1.2.4.c. Calcul de puissance statistique

Nous avons calculé la puissance statistique que nous disposions pour mettre en évidence une association entre le risque de CDT et un SNP compte tenue de la taille de mes échantillons d'étude (tableau 4). La puissance du test dépend de l'amplitude de l'OR, de la fréquence allélique du SNP (MAF), de la taille de l'échantillon et du risque d'erreur α. Dans la population européenne, la puissance statistique était raisonnable pour des OR supérieurs ou égal à 1,30 avec des SNP de fréquence ≥0.15. Dans la population mélanésienne, la puissance statistique était globalement limitée pour des OR ≤1,4.

Tableau 4 : Puissance statistique pour mettre en évidence un OR de 1,10, 1,20; 1,30 ou 1,40 avec $\alpha=10^{-3}$ dans nos deux population d'études pour l'effet propre des SNP et les interactions gènes-environnement. Pour les interactions gènes-environnement, nous prenons l'exemple de l'obésité en considérant une prévalence de 15% chez les européens et 25% chez les mélanésiens, avec un OR=1.30.

		OR (gènes)				OR (interaction GE)			
MAF chez les témoins	**N cas/ N témoins**	**1,10**	**1,20**	**1,30**	**1,40**	**1,10**	**1,20**	**1,30**	**1,40**
Européens	**508/629**								
0,05		0,4%	3%	9%	24%	0.2%	0.4%	0.9%	1%
0,15		1%	13%	45%	80%	0.3%	1%	0.3%	1%
0,25		2%	24%	68%	94%	0.4%	2%	5%	14%
0,35		3%	32%	78%	97%	0.4%	2%	7%	18%
0,45		4%	35%	81%	98%	0.4%	2%	7%	18%
Mélanésiens	**156/112**								
0,05		0,1%	0,5%	1%	3%	0.1%	0.2%	0.4%	0.6%
0,15		0,3%	2%	6%	16%	0.2%	0.4%	1%	2%
0,25		0,5%	3%	11%	28%	0.2%	0.6%	1%	3%
0,35		0,6%	4%	15%	35%	0.2%	0.6%	2%	3%
0,45		0,7%	5%	16%	37%	0.2%	0.6%	2%	3%

1.2.5. Annotation fonctionnelle des variants

L'annotation fonctionnelle in silico permet de prioriser les variants d'intérêt à partir à partir d'informations sur leur fonction biologique. Cela peut être particulièrement utile dans les analyses de fine-mapping lorsque l'on étudie de nombreux variants fortement corrélés entre eux. Pour ce faire, nous avons utilisé différentes bases de données bio-informatiques.

Nous avons utilisé RegulomeDB (http://regulome.stanford.edu) (Boyle et al. 2012) et Haploreg (www.broadinstitute.org/mammals/haploreg) (Ward & Kellis 2012) pour prédire si les SNP sont situés dans les zones de régulations. Il s'agit de régions du génome impliquées dans la régulation de la transcription des gènes. Elles sont généralement denses en séquence de régulation qui peuvent être des séquences de fixation des facteurs de transcription (ie. protéines nécessaires à l'initiation ou à la régulation de la transcription d'un gène), des séquences régulatrices ou « enhancers » (généralement situés en amont d'un gène et capable d'activer le promoteur de ce gène pour la transcription) et des zones de modifications d'histones. Une altération dans ces régions peut entrainer une modification de l'expression du gène cible.

Les prédictions obtenues avec regulomeDB et Haploreg sont basées sur les informations issues, des données de Chip-Seq (données d'expression des protéines et facteurs de transcription issus d'études expérimentales) ou de Position Weight Matrix (PWM). Les PWM

sont des scores obtenus à partir d'un modèle probabiliste caractérisant la préférence d'un facteur de transcription à se fixer sur une séquence d'ADN (Chen et al. 2007). La statistique de test permettant de définir si la variation d'un nucléotide altère significativement le score de fixation d'un FT sur une séquence d'ADN a été obtenue à partir de l'outil is-rSNP (http://bioinformatics.research.nicta.com.au/software/is-rsnp).

Nous avons utilisé USCS genomes Browser pour l'identification des lincRNA non caractérisés exprimés dans la thyroïde ou des ESTs (ou « Expressed Sequence Tag ») dans les régions où étaient situées les SNP candidats. Les LincRNA sont des ARN inter-génique longs et non codante dont l'expression est souvent modifiée dans les cancers. Les EST sont des séquences courtes d'ADN codante, elles permettent d'identifier les transcrits de gènes ou encore servent à la détermination de nouveaux gènes ou séquences codantes.

Enfin, à partir des données d'expression des gènes mesurés dans les cellules lymphoblastoïdes chez 210 individus de la base de données de phase 1 de HAPMAP et qui sont disponibles dans la base eQTL (http://app3.titan.uio.no/biotools). Nous avons évalué la corrélation entre le génotype de SNP d'intérêt et le niveau d'expression de gènes d'intérêt comme par exemple les gènes *FOXE1* sur le locus 9q22 ou *PTCSC3* et *NKX2-1* sur le locus 14q13.

1.3. Résultats

1.3.1. Caractéristiques des populations d'étude des facteurs de risque environnementaux ou liés au mode de vie des CDT

Le Tableau 5 présente la distribution par groupe ethnique des cas et témoins selon l'âge, le sexe, la région de résidence et le type histologique. Du fait de l'appariement par fréquence sur l'âge et sur le sexe, la distribution de ces variables était très proche chez les cas et les témoins. On note toutefois quelques différences qui s'expliquent par le fait que les cas et les témoins n'étaient pas appariés sur le groupe ethnique. L'âge moyen au diagnostic des cas était d'environ 51 ans chez les européens et 46 ans chez les mélanésiens. 20% des cas étaient des hommes dans la population européenne contre 5% seulement dans la population mélanésienne

Tableau 5 : Distribution des cas et témoins des études CATHY et NC selon l'âge, le sexe, la région de résidence et le type histologique de cancer

	Européens		**Mélanésiens**	
Variables	Cas n=505 (%)	Témoins n= 622 (%)	Cas n= 156 (%)	Témoins n= 114 (%)
Age en classe de 5 ans				
<30	25 (4,95)	36 (5,79)	14 (8,98)	7(6,14)
30-34	23 (4,55)	34 (5,47)	3 (1,92)	4 (3,51)
35-39	15 (2,97)	25 (4,02)	14 (8,97)	8 (7,02)
40-44	25 (4,95)	41 (6,59)	22 (14,10)	14 (12,28)
45-49	48 (9,50)	60 (9,65)	19 (12,18)	16 (14,04)
50-54	44 (8,71)	74 (11,90)	14 (8,97)	16 (14,04)
55-59	77 (15,25)	69 (11,09)	13 (8,33)	12 (10,53)
60-64	93 (18,42)	87 (13,99)	20 (12,82)	15 (13,16)
65-69	68 (13,47)	77 (12,38)	24 (15,38)	8 (7,02)
70-74	52 (10,30)	54 (8,68)	13 (7,45)	14 (11,63)
>=75	35 (6,93)	65 (10,45)	1 (0,87)	1(0,64)
Sexe				
Femmes	403 (79,80)	463 (74,44)	148 (94,87)	105 (92,11)
Hommes	102 (20,20)	159 (25,56)	8 (5,13)	9 (7,89)
Etudes				
CATHY	478 (94,65)	541 (86,08)		
New-Calédonie (NC)	27 (5,35)	81 (13,02)	156 (100)	114 (100)
Département de résidence étude CATHY				
Calvados	152 (32,80)	194 (35,86)		
Marne	217 (45,40)	250 (46,21)		
Ardennes	109 (22 ,80)	97 (17,93)		
Province de résidence étude NC				
Nord	3 (11,11)	6 (7,41)	35 (22,44)	29 (25,44)
Sud	23 (85,19)	74 (91,36)	54 (34,62)	46 (40,35)
Iles loyautés	1 (3,70)	1 (1,23)	67 (42,95)	39 (34,21)
Types histologique des cancers				
Papillaires	445 (88,12)		132 (84,62)	
Vésiculaires	60 (11,98)		24 (15,38)	

Dans l'étude CATHY, 45% des cas résidaient dans le département de la Marne, 32% dans le Calvados et 22% dans les Ardennes ; dans l'étude de NC les individus d'origine européenne résidaient à plus de 90% dans la province Sud, alors que les mélanésiens étaient répartis entre les 3 provinces de NC. Plus de 85% des cas de cancers étaient de type papillaire.

Le tableau 6 présente les résultats d'association avec les facteurs de risque des cancers thyroïdiens. L'obésité (IMC $\geq$30 kg/m²) est associée à une augmentation de l'OR statistiquement significative chez les mélanésiens, et à la limite de la significativité chez les européens. On note que la prévalence de l'obésité est plus importante dans la population mélanésienne (25%) que dans la population européenne (17%). Une association inverse est observée avec la consommation d'alcool chez les européens, mais pas chez les mélanésiens chez qui on observe une plus forte prévalence de non buveurs. Le statut tabagique et la consommation de tabac en paquets-année ne sont pas associés au risque de CDT dans aucune des deux populations.

Chez les femmes européennes, les OR augmentent avec l'âge tardif aux premières règles (>13 ans) et le statut ménopausique et diminuent avec la prise de traitements hormonaux exogènes (pilule ou traitement hormonal de la ménopause (THM)). Ces associations ne sont pas observées chez les femmes mélanésiennes.

De manière général, les résultats observés pour l'ensemble des facteurs environnementaux étudiés dans cette sous-population d'étude étaient comparables à ceux précédemment rapportés dans la population initiale de l'étude CATHY (621 cas et 706 témoins) et dans la population mélanésienne de l'étude NC (332 cas et 412 témoins) (annexe 3 : tableau 1 à 3).

Tableau 6: Facteurs de risque liés au mode de vie et risque de CDT en population générale et chez les femmes dans les populations européennes et mélanésiennes.

	Européens				Mélanésiens			
Variables	**Cas n= 504(%)**	**Témoins n= 620(%)**	**OR**	**(95% CI)**	**Cas n= 156(%)**	**Témoins n= 114(%)**	**OR**	**(95% CI)**
Population générale								
IMC (kg/m^2) [a]								
<25	237 (47,02)	323 (52,10)	1,00		38 (24,36)	38 (33,33)	1,00	
25-29,9	162 (32,14)	188 (30,32)	1,37	(1,01-1,87)	52 (33,33)	41 (35,96)	1,28	(0,65-2,50)
≥30	104 (20,63)	105 (16,94)	1,33	(0,93-1,91)	63 (40,38)	29 (25,44)	2,70	(1,34-5,44)
DM	1 (0,20)	4 (0,65)			3 (1,92)	6 (5,26)		
Consommation d'alcool (verres/jour) [a]								
Jamais	141 (27,98)	130 (20,97)	1,00		69 (44,23)	45 (39,47)	1,00	
<1	249 (49,40)	333 (53,71)	0,65	(0,47-0,89)	66 (42,31)	53 (46,49)	0,68	(0,37-1,26)
≥1	111 (22,02)	152 (24,52)	0,64	(0,42-0,96)	19 (12,18)	14 (12,28)	1,08	(0,42-2,75)
DM	3 (0,60)	5 (0,81)			2 (1,28)	2 (1,75)		
Statut tabagique [b]								
Non-fumeur	269 (53,37)	322 (51,94)	1,00		95 (60,9)	70 (61,40)	1,00	
Ancien fumeur	147 (29,17)	159 (25,65)	1,09	(0,81-1,47)	25 (16,03)	15 (13,16)	0,86	(0,38-1,95)
Fumeur actuel	88 (17,46)	139 (22,42)	0,80	(0,57-1,12)	36 (23,08)	29 (25,44)	1,07	(0,55-2,06)
DM	0 (0)	0 (0)			0 (0)	0 (0)		
Consommation de tabac en paquet-années [a]								
Non-fumeur	269 (53,37)	322 (51,94)	1,00		95 (60,90)	70 (61,40)	1,00	
<20	165 (32,74)	191 (30,81)	1,14	(0,84-1,55)	42 (26,92)	26 (22,81)	1,27	(0,59-2,25)
≥20	68 (13,49)	99 (15,97)	0,82	(0,54-1,26)	19 (12,18)	17 (14,91)	0,82	(0,36-1,85)
DM	2 (0,40)	8 (1,29)			0 (0)	1 (0,88)		
Femmes uniquement								
Age aux 1ères règles (années) [c]								
≤13	329 (81,84)	398 (85,96)	1,00		63 (42,57)	56 (53,33)	1,00	
>13	71 (17,66)	63 (13,61)	1,47	(0,99-2,17)	78 (52,70)	43 (40,95)	1,31	(0,72-2,41)
DM	2 (0,50)	2 (0,43)			7 (4,73)	6 (5,71)		
Prise de contraceptifs oraux [c]								
Jamais pris	109 (27,11)	122 (26,35)	1,00		110 (74,32)	77 (73,33)	1,00	
Déjà pris	288 (71,64)	338 (73)	0,72	(0,49-1,06)	38 (25,68)	27(25,71)	1,27	(0,62-2,61)
DM	5 (1,24)	3 (0,65)			0 (0)	1 (0,95)		
Nombre de grossesse à terme [c]								
0	36 (8,96)	53 (11,45)	1,00		18 (12,16)	13 (12,38)	1,00	
1-2	225 (55,97)	251 (54,21)	1,20	(0,72-1,99)	28 (18,92)	29 (27,62)	0,52	(0,17-1,59)
≥3	141 (35,07)	159 (34,34)	1,23	(0,72-2,12)	95 (64,19)	60 (57,14)	0,67	(0,23-1,91)
DM	0 (0)	0 (0)			7 (4,73)	3 (2,86)		
Statut ménopausique [c]								
Pré-ménopause	140(41,90)	170(50,40)	1,00		84 (63,10)	58 (64,4)	1,00	
Post-ménopause	194(58,10)	167(49,60)	1,30	(0,94-1,80)	49 (36,80)	167 (35,60)	1,04	(0,34-3,63)
Prise de THM[d]								
Jamais pris	91 (48,15)	123 (56,94)	1,00					
A Déjà pris	97 (51,32)	87 (40,27)	0.63	(0,40-0,98)				

[a] Le modèle multivarié inclut l'IMC, la consommation de tabac en verres/jour et la quantité de tabac en paquets année consommé. [b] Le modèle multivarié inclut l'IMC, la consommation de tabac en verres/jour et le statut tabagique. [c] Le modèle multivarié inclut l'IMC, la consommation de tabac en verres/jour, la quantité de tabac en paquets année ainsi que les facteurs hormonaux et reproducteurs. [d] Traitements hormonaux de la ménopause.

1.3.2. Association entre les polymorphismes candidats et le risque de CDT

1.3.2.a. Introduction

Dans cette partie, nous nous sommes intéressés à la relation entre le risque de CDT les SNP candidats identifiés dans la littérature jusqu'en 2012 (Landa & Robledo 2011) et des polymorphismes des gènes de l'obésité (Fesinmeyer et al. 2013).

Notre objectifs était d'étudier le risque de CDT associé aux SNP candidats sélectionnés, chez les européens dans les études CATHY et NC (réplication) et chez les mélanésiens de NC et tester leurs interactions avec les facteurs de risque établis ou suspectés des CDT (IMC, consommation d'alcool et de tabac, facteurs hormonaux et reproductifs.

1.3.2.b. Résultats

Dans le Tableau 7, sont indiqués les ORs et les IC à 95% des associations entre les SNP candidats et le risque de CDT. Lorsqu'on compare les fréquences alléliques des variants dans les deux populations, on observe qu'elles sont très différentes entre européens et mélanésiens. Pour certains SNP comme par exemple : rs1042522 (*TP53*) ; rs4752904 (*PTPRJ*) ; rs966423 (*DIRC3*) et rs11084753, l'allèle rare chez les européens était plutôt très fréquent chez les mélanésiens. D'autre SNP comme rs1801516; rs3092993 ou rs1800057 dans ATM et rs9942754 (*TG,SLA*) étaient au contraire monomorphiques chez les mélanésiens Cependant, l'hétérogénéité des ORs d'associations entre les deux études n'est pas significative (tableau 7).

Parmi les 37 SNP candidats analysés, aucun SNP n'était significativement associé au risque de CDT dans les deux populations après correction sur les tests multiples ($p<10^{-3}$) (Tableau 7). Chez les européens, nous avons néanmoins observé une association à la limite de la significativité pour le SNP GWAS rs966423 du locus 2q35 (OR=1,31 (IC à 95% : 1,10 – 1,56), $p=2x10^{-3}$), qui n'était pas répliquée chez les mélanésiens (OR=0,94 (IC à 95% : 0,31 – 2,84), p=0,90) avec des fréquences alléliques très différentes. Une association limite significative avec le CDT a été retrouvé pour le SNP rs3764340 (*WWOX*) chez les européens (OR=1.55 (IC à 95% :1,07 – 2,25) ; p=0.02) mais pas chez les mélanésiens (OR=1.08 (IC à 95% : 0,55 – 2,13 ; p=0.82). Parmi les SNP liés à l'obésité, rs10838738 (*MTCH2*) et rs11084753 (*KCTD15*) étaient

aussi associés au risque de CDT au seuil de 5% chez les européens, avec des OR du même ordre de grandeur mais non significatifs chez les mélanésiens (Tableau 7).

Interaction avec les facteurs environnementaux

Aucune interaction significative après prise en compte des tests multiples n'était observée ; toutefois, des interactions à la limite de la significativité ont été observées avec les facteurs de risque environnementaux suivants:

- l'IMC avec le SNP rs1042522 dans *TP53* ($p=4.6x10^{-03}$) ;
- le statut tabagique avec le SNP rs9942754 sur le locus 8q24 près des gènes *TG,SLA* ($p=3.7x10^{-03}$)
- la parité avec le SNP rs9442754 ($p=10^{-2}$)
- l'âge aux premières règles avec les SNP rs11084753 (dans *KCTD15*), rs17782313 (dans *MC4R*) et rs6548238 (dans *TMEM18*) (annexe 4).

Chez les mélanésiens, nous avons également observé des interactions à la limite significative de la significativité entre l'âge aux premières règles et le SNP rs10838738 (dans *MTCH2) ($p=2x10^{-3}$)* et entre la consommation d'alcool et le SNP rs2048722 (dans *TPO)* (p=0,01) (annexe 4).

Tableau 7: Résultats ORs d'association entre les SNP et le risque de CDT chez les européens et les mélanésiens

					Européens (508 cas/629 témoins)					Mélanésiens (156 cas/112 témoins)					
					MAF (%)					MAF (%)					
Chr	**SNP**	**A risque**	**A ref**	**Gènes**	**Cas**	**Témoins**	**OR**	**IC à 95%**	**p value**	**Cas**	**Témoins**	**OR**	**IC à 95%**	**p value**	**p-het**
Variants candidats répliqués dans 2 études															
17	rs1042522	C	G	*TP53*	25,39	25,84	0,96	(0,79 -1,16)	0,66	66,23	68,02	0,95	(0,67 -1,34)	0,75	0,87
19	rs25487	T	C	*XRCC1*	34,55	35,82	0,95	(0,80 -1,14)	0,60	28,53	34,38	0,76	(0,52 -1,10)	0,14	0,28
	rs1799782	A	G		7,10	7,20	1,00	(0,72 -1,40)	0,99	16,35	16,52	0,93	(0,57 -1,51)	0,76	0,93
5q24	rs2910164	C	G	*pre-miR-146a*	25,39	22,76	1,20	(0,99 -1,46)	0,07	35,58	37,95	0,85	(0,58 -1,26)	0,42	0,18
14	rs861539	T	C	*XRRC3*	33,98	37,30	1,04	(0,87- 1,23)	0,66	5,73	6,76	0,77	(0,39- 1,51)	0,44	0,43
8q24	rs6983267	G	T	*Gene desert*	47,05	48,80	0,95	(0,80 -1,13)	0,56	86,04	86,94	0,88	(0,53 -1,46)	0,62	0,92
2q35	**rs966423**	**C**	**T**	***DIRC3***	**47,83**	**42,25**	**1,31**	**(1,10 – 1,56)**	**$2x10^{-3}$**	97,42	97,32	0,94	(0,31 -2,84)	0,91	0,69
8p12	rs2439302	C	G	*NRG1*	48,92	45,85	1,15	(0,90 -1,36)	0,09	39,68	41,07	0,95	(0,67 -1,35)	0,76	0,33
2	rs2048722	C	T	*TPO*	44,08	42,87	1,04	(0,88 -1,24)	0,61	88,14	81,70	1,53	(0,98 -2,54)	0,06	0,09
	rs732609	C	A	*TPO*	41,04	38,98	1,07	(0,91 -1,27)	0,40	44,55	38,18	1,28	(0,88 -1,85)	0,19	0,32
Variants fonctionnel retrouvés dans 1 étude															
	rs4752904	G	A	*PTPRJ*	44,77	44,41	1,04	(0,88 -1,24)	0,61	61,22	62,50	0,90	(0,63 -1,31)	0,59	0,66
11	rs3092993	A	C	*ATM*	13,19	12,64	1,06	(0,83 -1,35)	0,64	0,64	0,00	-	-	-	
	rs1801516	A	G		13,19	12,70	1,06	(0,83 -1,34)	0,66	0,96	0,00	-	-	-	
	rs664677	C	T		40,14	40,42	0,99	(0,83 -1,17)	0,88	64,10	62,05	1,07	(0,75 -1,53)	0,71	0,64
	rs1800057	C	C		3,54	3,52	1,07	(0,68 -1,70)	0,76	0,00	0,00	-	-	-	
	rs189037	A	G		40,81	41,29	0,98	(0,82 -1,16)	0,77	62,90	64,41	0,91	(0,62 -1,32)	0,60	0,81
14	rs4903957	A	G	*TSHR*	34,00	35,30	0,93	(0,78- 1,10)	0,39	5,45	5,36	1,04	(0,48-2,27)	0,91	0,90
8	rs1133076	A	G	*TG*	48,72	48,16	1,04	(0,88 -1,23)	0,61	34,62	33,93	0,99	(0,69 -1,44)	0,97	0,98
	rs2252696	C	A	*TG*	46,06	48,88	0,91	(0,78 -1,08)	0,27	80,13	79,91	1,06	(0,69 -1,65)	0,78	0,68
16	**rs3764340**	**G**	**C**	***WWOX***	**6,69**	**4,47**	**1,55**	**(1,07 -2,25)**	**0,02**	7,05	6,70	1,08	(0,55 -2,13)	0,81	0,33
Variants candidats des études des cas familiaux															
5q13	rs6885099	G	A	*PDE8B*	47,62	43,44	0,97	(0,82 -1,14)	0,69	81,41	79,02	1,13	(0,74 -1,73)	0,57	0,47
8q23	rs1445493	A	G	_	34,65	36,00	0,92	(0,77 -1,10)	0,34	0,96	2,23	0,48	(0,11 -2,11)	0,33	0,29
8p22	rs822318	C	T	_	46,56	47,68	0,99	(0,83 -1,17)	0,86	86,54	80,80	1,55	(0,96 -2,51)	0,07	0,07
8q24	rs9942754	T	C	*TG,SLA*	1,48	0,96	1,56	(0,74 -3,31)	0,24	0,00	0,00	-	-	-	
	rs2702968	T	A		49,51	49,28	1,03	(0,87 -1,22)	0,73	20,83	21,43	0,91	(0,60 -1,40)	0,67	0,80
	rs2702967	T	C		49,51	49,28	1,03	(0,87 -1,22)	0,73	20,83	21,43	0,91	(0,60 -1,40)	0,67	0,80
	rs17695552	C	T		24,21	22,63	1,09	(0,89 -1,33)	0,40	0,96	0,89	1,05	(0,17 -6,54)	0,95	0,97
	rs2741205	A	G		25,49	27,40	0,89	(0,74 -1,07)	0,21	64,10	63,51	1,05	(0,73 -1,51)	0,80	0,52
Variants des gènes de l'obésité															
1	rs2815752	C	T	*NEGR1*	34,45	32,69	1,06	(0,89 -1,27)	0,50	13,46	16,07	0,81	(0,50 -1,30)	0,38	0,30
11	rs10767664	T	A	*BDNF*	22,88	22,84	0,99	(0,81 -1,21)	0,92	25,32	28,57	0,79	(0,53 -1,17)	0,24	0,43
11	**rs10838738**	**G**	**A**	***MTCH2***	**31,00**	**36,02**	**0,80**	**(0,67 -0,96)**	**0,01**	49,68	53,13	0,89	(0,63 -1,26)	0,52	0,65
16	rs9939609	A	T	*FTO*	39,76	43,60	0,86	(0,73 -1,02)	0,09	19,55	20,54	0,93	(0,61 -1,41)	0,72	0,72
	rs7498665	G	A	*SH2B1*	37,48	38,12	1,01	(0,84 -1,20)	0,94	83,01	78,57	1,40	(0,88 -2,21)	0,15	0,21
18	rs17782313	C	T	*MC4R*	23,87	22,35	1,10	(0,90 -1,34)	0,34	20,65	19,20	1,15	(0,74 -1,79)	0,54	0,98
	rs12970134	A	G		27,18	25,49	1,11	(0,91 -1,34)	0,31	20,83	18,92	1,17	(0,75 -1,84)	0,48	0,88
19	**rs11084753**	**A**	**G**	***KCTD15***	**36,59**	**33,25**	**1,23**	**(1,03 -1,47)**	**0,02**	65,71	62,05	1,16	(0,80 -1,68)	0,42	0,87
2	rs6548238	T	C	*TMEM18*	17,52	17,01	1,04	(0,84 -1,29)	0,71	0,32	0,89	0,35	(0,03 -4,01)	0,39	0,37
4	rs13130484	T	c	*GNPDA2*	43,10	43,05	1,00	(0,84 -1,18)	0,97	5,77	10,27	0,51	(0,26 -1,00)	0,05	0,05

A_risque : Allèle à risque ; A_ref : A de référence, phet : p-value hétérogénéité entre les études

1.3.2.c. Synthèse des résultats et comparaison avec la littérature

Parmi les polymorphismes précédemment associés au risque de CDT, seuls les SNP rs966423 (locus 2q35) et rs3764340 (WWOX) avaient des associations à la limite de la significativité chez les européens. Aucun des SNP n'était associé au risque de CDT dans la population mélanésienne. Bien qu'il s'agisse de la première étude ayant analysé le rôle de facteurs de risque génétique chez les mélanésiens, nous disposions dans cette population que d'une puissance limitée de moins de 50% pour mettre en évidence des associations significatives.

Le SNP rs966423 avait initialement été mis en évidence dans une étude GWAS qui avait étudié la relation entre les SNP associés aux taux circulants de TSH avant d'être associé au risque de CDT avec un OR de 1,34 (Tableau 1) (Gudmundsson et al. 2012). Cet association est proche de celle que nous observons dans notre étude chez les européens (OR=1.31) mais pas chez les mélanésiens (OR=0.94). De telle association ont également été rapportés dans d'autres études gène-candidat (Wang et al. 2013; Figlioli et al. 2014) ainsi qu'une récente méta-analyse sur des populations européennes (Figlioli et al. 2016).

Le second SNP rs3764340 répliqué dans la population européenne de notre étude, avait été associé au risque de CDT dans une étude gène-candidat ayant inclus 1741 cas et 1042 témoins d'origine européenne (Cancemi et al. 2011). Ce SNP entraine un changement de la proline en alanine dans le codon 282 du gène *WWOX*. Les auteurs avaient prédit in silico que ce changement d'acide aminé dans WWOX entrainerait une modification de son expression et par conséquent une altération de la fonction du gène. *WWOX* est un gène suppresseur de tumeur dont la sous-expression pourrait contribuer à la carcinogénèse de la thyroïde (Cancemi et al. 2011). L'OR rapporté dans cette étude (OR=1.48) est similaire à celle que nous observons dans notre population européenne (OR=1,55), mais pas dans notre population mélanésienne (OR=1.08). A notre connaissance, aucune autre étude n'a rapporté une association entre le polymorphisme rs3764340 et les CDT

Parmi les variants des gènes de l'obésité, nous avons rapporté des associations à la limite de la significativité pour les variants rs10838738 (*MTCH2*) et rs11084753 (*KCTD15*), avec des OR du même ordre dans les 2 groupes ethniques étudiés. Hormis notre étude, aucune autre étude n'a exploré le rôle de ces polymorphismes dans les CDT.

Il est à noter que parmi les différents variants candidats que nous avons étudiés, une récente méta-analyse des SNP associés au risque de CDT dans la littérature jusqu'en 2015 (Figlioli et al. 2016) a rapporté une association significative entre les CDT et les variants rs25487 dans *XRRC1* (OR=0,76 (IC à 95% : 0,59 – 0,99)) et rs9939609 dans FTO (OR=0,77 (IC à 95% : 0,66 – 0,87)). Dans notre étude, nous retrouvons respectivement pour ces variants, un OR de 0,95 (IC à 95% : 0,80 – 1,14) et de 0,73 (IC à 95% : 0,73 – 1,02) dans la population européenne.

Interaction gène environnement

Nous n'avons pas mis en évidence d'interaction significative entre les polymorphismes sélectionnés et les facteurs de risque étudiés (qui sont la consommation d'alcool et de tabac, l'IMC et les facteurs hormonaux et reproductifs chez les femmes). Très peu d'étude se sont intéressée aux interactions gènes-environnement dans les CDT. Parmi les polymorphisme d'intérêt dans notre étude, seul le SNP rs1801516 avait fait l'objet d'étude d'interaction avec la parité, l'IMC et la consommation de tabac (Pereda et al. 2015; Maillard et al. 2015). Une interaction avec la parité avait été rapportée dans la population cubaine (Pereda et al. 2015), avec un OR de CDT chez les femmes porteuses de l'allèle rs1801516[A] qui était 3,5 fois plus élevé chez celles ayant eu au moins deux grossesses à terme que chez celles ayant eu moins de 2 grossesses à terme (Pereda et al. 2015). Le SNP rs1801516 n'a pas pu être étudié dans la population mélanésienne du fait de sa très faible fréquence allélique (MAF<0,001) ; chez les européens, nous n'avons pas rapportés d'interaction entre ce SNP et le nombre de grossesse à terme.

En conclusion, nos résultats ne confirment le rôle que d'un nombre réduit de polymorphismes candidats chez les européens et aucun chez les mélanésiens. De plus, nous n'avons pas observé d'interaction significative avec les facteurs de risque suspectés dans les CDT.

1.3.3. Rôle de la variation du nombre de copie de *GSTM1* et *GSTT1* dans les CDT

1.3.3.a. Introduction

GSTM1 et *GSTT1* sont des gènes de la famille des glutathion S-transférases qui jouent un rôle important dans la détoxification des composés issus du métabolisme des substances endogènes et exogènes incluant les carcinogènes en les rendant hydrosolubles par conjugaison avec le glutathion.

GSTM1 et *GSTT1* sont caractérisés par une variation du nombre de copies dans le génome, chaque gène pouvant être présent en plus de 2 exemplaires ou avoir une délétion complète. Une relation dose-effet entre le nombre de copies de *GSTT1* et *GSTM1* et leur activité enzymatique a été décrite avec une perte totale de l'activité lors de la délétion complète du gène (Butler et al. 2011; Schlattl et al. 2011). De ce fait, ils sont suspectés de jouer un rôle dans l'apparition de plusieurs cancers, la diminution ou la perte totale de l'activité enzymatique pouvant diminuer les capacités de détoxification.

Une méta-analyse publiée en 2012 et incluant les résultats de 12 études cas-témoins ne conclut pas à une association entre le risque de CDT et la délétion complète (ou génotype nul) des gènes *GSTM1* ou *GSTT1* (J. Li et al. 2012). Cependant, une hétérogénéité significative entre les études a été rapportée. Il est a noté que les études incluses dans cette méta-analyse portaient sur des populations européennes (Stankov et al. 2006; Canbay et al. 2003; Hernández et al. 2003; Gaspar et al. 2004) (Ho et al. 2006), saoudiennes (Siraj et al. 2008) et brésiliennes (Morari et al. 2002; Guilhen et al. 2009; Reis et al. 2010). Ces populations ont, non seulement, des fréquences différentes de génotypes *GSTM1*-nul et *GSTT1*-nul mais sont aussi exposées de manières différentes à certains facteurs liés au mode de vie tels que la consommation d'alcool ou l'obésité, qui peuvent interagir avec *GSTM1* et *GSTT1* pour moduler la susceptibilité individuelle au risque de CDT. L'effet propre des habitudes de vie et de leurs interactions avec les gènes sur le risque de cancer pourraient ainsi expliquer les différences observées entre les études.

L'existence d'une interaction entre les gènes *GSTM1*/*GSTT1* et la consommation d'alcool n'a jamais été étudiée dans les CDT. Une étude sur la consommation de tabac (Hernández et

al. 2003), et une seconde sur l'IMC (Kweon et al. 2014), n'ont pas mis en évidence l'existence d'une interaction avec les gènes *GSTM1*/*GSTT1*.

L'objectif de notre étude était d'étudier la relation entre le nombre de copies de *GSTM1* et *GSTT1* et le risque de CDT chez les européens de France métropolitaine et les mélanésiens de Nouvelle-Calédonie. Nous avons également étudié si les génotypes pouvaient moduler l'association entre les CDT et certains facteurs de risque suspectés tels que l'obésité, la consommation de tabac et d'alcool des facteurs hormonaux et reproductifs (âge aux premières règles, parité, prise de contraceptifs oraux et le statut ménopausique).

1.3.3.b. Résultats

i. Distribution des génotypes de *GSTM1* et *GTT1* et association avec les CDT.

Les distributions des génotypes de *GSTM1* et *GSTT1* chez les cas et témoins ainsi que leurs ORs d'association avec les CDT sont présentés dans le *Tableau 8*. On observe que la fréquence des génotypes nuls de *GSTM1* et *GSTT1* était plus faible dans la population européenne que dans la population mélanésienne. Les proportions de génotypes nuls pour *GSTM1* étaient de 51,2% chez les témoins d'origine européenne contre 82,5% chez les témoins d'origine mélanésienne ; les proportions correspondantes pour les génotypes de *GSTT1* étaient de 20,5% et 39,4% 10,1% des témoins européens avaient un génotype nul pour les deux gènes simultanément contre 36,5% chez les mélanésiens.

Nous n'avons pas observé d'association significative entre *GSTM1* et/ou *GSTT1* et le risque de CDT (tableau 8).

Tableau 8 : Distribution des fréquences des génotypes de GSTM1 et GSTT1 et de la combinaison des deux gènes chez les européens et les mélanésiens et association avec les CDT.

Variables	Européens				Mélanésiens			
	Cas 504(%)	Témoins 620(%)	OR	(95% CI)	Cas 156(%)	Témoins 114(%)	OR	(95% CI)
GSTM1 (nombre de copies)								
0	248 (52,65)	298 (51,20)	0,92	(0,60-1,41)	124 (80)	94 (82,46)	0,80	(0,21-3,08)
1	172 (36,52)	231 (39,69)	0,80	(0,52-1,25)	25 (16,13)	16 (14,04)	0,82	(0,19-3,54)
≥2	51 (10,83)	53 (9,11)	1,00		6 (3,87)	4 (3,51)	1,00	
DM	33	38			1	0		
Génotype GSTM1								
Nul	248 (52,65)	298 (51,2)	1,09	(0,85, 1,40)	124 (80)	94 (82,46)	0,94	(0,49-1,82)
Non-nul	223 (47,35)	284 (48,8)	1,00		31 (20)	20 (17,54)	1,00	
DM	33	38			1	0		
GSTT1 (nombre de copies)								
0	90 (18,99)	116 (20,53)	0,81	(0,57-1,15)	53 (40,46)	39 (39,39)	0,80	(0,36-1,75)
1	196 (41,35)	254 (44,96)	0,80	(0,61-1,06)	48 (36,64)	44 (44,44)	0,61	(0,28-1,33)
≥2	188 (39,66)	195 (34,51)	1,00		30 (22,9)	16 (16,16)	1,00	
DM	30	55			25	15		
Génotype *GSTT1*								
Nul	90 (18,99)	116 (20,53)	0,92	(0,67-1,26)	53 (40,46)	39 (39,39)	1,12	(0,64-1,97)
Non-nul	384 (81,01)	449 (79,47)	1,00		78 (59,54)	60 (60,61)	1,00	
DM	30	55			25	15		
Combinaison de *GSTM1* et *GSTT1*								
GSTM1/GSTT1 (nombre de copies)								
0 copie pour *GSTM1* et *GSTT1*	41 (9,30)	53 (10,06)	0,97	(0,60-1,55)	42 (32,31)	36 (36,36)	0,97	(0,38-2,50)
Au moins 1e copie pour *GSTM1* ou *GSTT1*	229 (51,93)	274 (51,99)	1,01	(0,76-1,31)	72 (55,38)	52 (52,53)	1,18	(0,48-2,91)
Au moins 1e copie pour *GSTM1* et *GSTT1*	171 (38,78)	200 (37,95)	1,00		16 (12,31)	11 (11,11)	1,00	
DM	63	93			26	15		
Génotype *GSTM1/GSTT1*								
Nul/nul	41 (9,30)	53 (10,06)	0,97	(0,60-1,55)	42 (32,31)	36 (36,36)	0,97	(0,38-2,50)
Nul/non-nul	189 (42,86)	222 (42,13)	1,02	(0,77-1,36)	62 (47,69)	49 (49,49)	1,06	(0,43-2,63)
Non-nul/nul	40 (9,07)	52 (9,87)	0,90	(0,56-1,44)	10 (7,69)	3 (3,03)	3,44	(0,69-17,19)
Non-nul/non-nul	171 (38,78)	200 (37,95)	1,00		16 (12,31)	11 (11,11)	1,00	
DM	63	93			26	15		

Les ORs ont été ajustés sur l'âge, le sexe et la région de résidence, DM : Données manquante

ii. Interactions gène-environnement et analyses stratifiées

Le Tableau 9 présente les résultats d'association entre l'IMC, la consommation de tabac et d'alcool et le risque de CDT, stratifiés sur les génotypes nuls et non-nuls de *GSTM1*, *GSTT1* et la combinaison des deux gènes (*GSTM1*/*GSTT1*) chez les européens et mélanésiens.

Interaction avec l'IMC

Dans les deux populations, l'association entre obésité et les CDT est plus élevée chez les individus de génotype nul pour *GSTM1* ou *GSTT1* que pour ceux de génotype non-nul. Ces associations sont plus marquées chez les individus ayant une délétion complète des deux gènes, avec des OR de 4,09 [0.98 – 17,13] chez les européens et 13,28 [2,79 – 63,17] chez les mélanésiens. En considérant le nombre de copies des gènes, on constate que l'OR associé à l'obésité augmente lorsque le nombre de copies du gène *GSTT1* (annexe 5 : tableau 1).

L'interaction entre l'IMC et *GSTM1* ou *GSTT1* ou la combinaison des deux gènes n'était cependant pas significative (Tableau 9).

Interaction avec l'alcool

Dans le Tableau *6* présenté plus haut, nous avons observé que la consommation d'alcool est inversement associée au risque de CDT chez les européens. Cette association inverse était plus marquée chez les individus ayant le génotype nul pour *GSTM1* ou *GSTT1* et encore plus prononcée chez les sujets de génotype nul pour les deux gènes (Tableau 9) ; les *p-values* d'interactions n'étaient cependant pas significatives. Par ailleurs, des interactions significatives entre la quantité moyenne d'alcool consommé (en nombre de verres/jour) et les génotypes nul/non-nul de *GSTM1* (p=0,048) et *GSTT1* (p=0,03) sont observés chez les européens.

Interaction avec le tabac

Pour la consommation de tabac, aucune différence n'était observée après stratification des analyses sur les génotypes nul/non-nul des gènes (Tableau *9*).

Interaction avec les facteurs hormonaux et reproductifs chez les femmes

Les OR pour les facteurs hormonaux et reproductifs stratifiés sur les génotypes des gènes *GSTM1* et *GSTT1* sont présentés dans le Tableau *10 GSTM1* ne modifie pas les associations du CDT avec l'âge aux premières règles, la prise de contraceptifs oraux, la parité ou le statut ménopausique. Concernant *GSTT1*, des interactions significatives avec l'âge aux premières règles (*p*=0,02) et la parité (*p*=0,02) sont observées chez les femmes européennes. En effet, un âge tardif aux premières règles était positivement associé au risque de CDT chez les femmes *GSTT1*-nul (OR = 1.79 [1.14-2.82]) et inversement associé au risque de CDT chez les femmes de génotype non-nul (OR=0.57 [0.21-1.54]). De même la parité était positivement associée au risque de CDT chez femmes de génotype non-nul pour *GSTT1* (OR=5.71 [1.29-25.2]) mais pas chez les femmes de génotype nul (OR=0.86 [0.49-1.51]). Ces résultats ne sont pas confirmés chez les femmes mélanésiennes.

Tableau 9: IMC, consommation d'alcool et de tabac et risque de CDT : analyses stratifiées sur les génotypes de *GSTM1*, *GSTT1* et la combinaison des deux gènes

	Européens							Mélanésiens						
	Non-nul			Nul				Non-nul			Nul			
Variables	**Ca/Te**	**OR**	**IC à 95%**	**Ca/Te**	**OR**	**IC à 95%**	***p***	**Ca/Te**	**OR**	**IC à 95%**	**Ca/Te**	**OR**	**IC à 95%**	***p***
GSTM1														
IMC (kg/m²) [a]							0,27							0,29
<25	105/157	1,00		120/154	1,00			10/6	1,00	1,00	27/31	1,00		
25-29,9	72/74	1,48	(0,97-2,26)	75/91	1,03	(0,68-1,55)		6/6	0,26	(0,03- 2,28)	45/34	1,63	(0,79- 3,39)	
≥30	44/52	1,14	(0,69-1,89)	51/44	1,40	(0,86-2,29)		14/7	1,39	(0,25- 7,82)	48/21	3,22	(1,47- 7,04)	
Consommation d'alcool [a]							0,08							0,71
Jamais	63/70	1,00	1,00	75/51	1,00			13/8	1,00	1,00	55/34	1,00	1,00	
Déja	158/213	0,83	(0,55-1,27)	171/238	0,50	(0,3-0,76)		17/11	0,85	(0,17- 4,10)	65/52	0,70	(0,37- 1,32)	
Consommation d'alcool (verre/jour)[a]							0,048							
Jamais	63/70	1,00		75/51	1,00			-	-	-	-	-	-	
<1	117/141	0,89	(0,5-1,38)	119/174	0,47	(0,30-0,74)		-	-	-	-	-	-	
≥1	41/72	0,67	(0,38-1,19)	52/64	0,59	(0,34-1,04)		-	-	-	-	-	-	
Statut tabagique [b]							0,32							0,85
Jamais	116/154	1,00		140/155	1,00			73/52	1,00	1,00	17/12	1,00		
Déjà	105/129	0,76	(0,42-1,38)	106/134	0,85	(0,49-1,47)		47/34	1,19	(0,64- 2,20)	13/7	0,70	(0,14- 3,44)	
Consommation de tabac (paquet-année)[c]							0,30							
Jamais	116/154	1,00		140/155	1,00			-	-	-	-	-	-	
<20	81/86	1,36	(0,89-2,08)	74/90	0,92	(0,61-1,39)		-	-	-	-	-	-	
≥20	24/43	0,82	(0,40-1,49)	32/44	0,84	(0,48-1,47)		-	-	-	-	-	-	
GSTT1														
IMC (kg/m²) [a]							0,59							0,17
<25	201/253	1,00		37/62	1,00			19/12	1,00	1,00	14/19	1,00		
25-29,9	124/138	1,14	(0,83-1,56)	35/38	1,86	(0,90-3,87)		29/26	0,64	(0,24- 1,72)	13/11	2,14	(0,65- 7,03)	
≥30	86/84	1,25	(0,87-1,81)	18/14	2,17	(0,86-5,47)		28/16	1,30	(0,48- 3,54)	23/7	7,63	(2,1- 27,64)	
Consommation d'alcool [a]														
Jamais	114/106	1,00		25/20	1,00			30/19			26/13	1,00	1,00	
Déjà	297/369	0,79	(0,57-1,09)	65/94	0,61	(0,29-1,28)		46/35	0,81	(0,36- 1,80	24/24	0,36	(0,13- 1,04)	
Consommation d'alcool (verre/jour)[a]							0,03							
Jamais	114/106	1,00		25/20	1,00			-	-	-	-	-	-	

<1	200/260	0,74	(0,53-1,03)	51/58	0,76	(0,35-1,64)		-	-	-	-	-	-	
≥1	97/109	0,96	(0,63-1,47)	14/36	0,33	(0,12-0,87)		-	-	-	-	-	-	
Statut tabagique [b]							0,53							0,80
Jamais	220/256	1,00		48/55	1,00	1,00		45/35	1,00	1,00	29/20	1,00		
Déjà	191/219	0,89	(0,59-1,36)	42/59	0,88	(0,34-2,23)		31/19	1,30	(0,58- 2,91)	21/17	1,08	(0,39- 2,98)	
Consommation de tabac (paquet-année)[c]							0,71							
Jamais	220/256	1,00		48/55	1,00			-	-	-	-	-	-	
<20	136/145	1,16	(0,80-1,58)	29/39	0,90	(0,45-1,81)		-	-	-	-	-	-	
≥20	55/74	0,87	(0,57-1,34)	13/20	1,00	(0,38-2,62)		-	-	-	-	-	-	
***GSTM1* et *GSTT1*[c]**														
IMC (kg/m²) [a]							0.75							0.06
<25	190/252	1,00		15/31	1,00			23/13	1,00	1,00	10/18	1,00		
25-29,9	125/125	1,34	(0,97-1,86)	15/16	1,64	(0,54-4,99)		31/27	0,61	(0,24- 1,54)	10/10	2,48	(0,64- 9,63)	
≥30	79/84	1,20	(0,82-1,75)	11/5	4,09	(0,98-17,13)		32/17	1,26	(0,49- 3,23)	19/6	13,28	(2,79- 63,17)	
Consommation d'alcool [a]							0.22							0,44
Jamais	119/102	1,00		12/6	1,00			33/20	1,00	1,00	23/12	1,00		
Déjà	275/359	0,69	(0,50-0,96)	29/46	0,28	(0,08-1,02)		53/37	0,80	(0,37- 1,72)	16/22	0,24	(0,07- 0,81)	
Consommation d'alcool (verres/jour)[a]							0,32							
Jamais	119/102	1,00		12/6	1,00			-	-	-	-	-	-	
<1	198/254	0,68	(0,48-0,95)	22/31	0,37	(0,09-1,46)		-	-	-	-	-	-	
≥1	77/105	0,75	(0,48-1,16)	7/15	0,16	(0,03-0,77)		-	-	-	-	-	-	
Statut tabagique [b]							0,64							0,44
Jamais	215/253	1,00		24/27	1,00			48/37	1,00	1,00	25/18	1,00		
A déjà fumé	179/208	0,77	(0,49-1,21)	17/25	0,94	(0,21-4,10)		38/20	1,45	(0,68- 3,11)	14/16	1,13	(0,35- 3,61)	
Consommation de tabac (paquets-année)[c]														
Jamais	215/253	1,00		24/27	1,00			-	-	-	-	-	-	
<20	137/139	1,25	(0,91-1,73)	10/17	0,58	(0,18-1,80)		-	-	-	-	-	-	
≥20	42/69	0,77	(0,49-1,22)	7/8	1,36	(0,29-6,45)		-	-	-	-	-	-	

[a] le modèle multivarié inclut l'IMC, la consommation de tabac en verres/jour et la quantité de tabac en paquets année consommé [b] le modèle multivarié inclut l'IMC, la consommation de tabac en verres/jour et le statut tabagique. [c] le modèle multivarié inclut l'IMC, la consommation de tabac en verres/jour, la quantité de tabac en paquets année ainsi que les facteurs hormonaux et reproducteurs. Ca/Te: Cas/Témoins, OR: Odd ratios, IC à 95% IC: Intervalle de confiance à 95%, *p*: *p-value* of interaction test

Tableau 10 : Facteurs hormonaux et reproductifs et risque de cancer de la thyroïde : analyses stratifiées sur les génotypes de *GSTM1, GSTT1* et la *co*mbinaison des deux gènes.

	Européens							Mélanésiens						
	Non-nul			Nul				Non-nul			Nul			
Variable	**Ca/Te**	**OR**	**95%IC**	**Ca/Te**	**OR**	**IC à 95%**	***P***	**Ca/Te**	**OR**	**IC à 95%**	**Ca/Te**	**OR**	**IC à 95%**	***P***
GSTM1														
Age aux 1ères règles [a](année)							0,81							0,71
≤13	148/186	1,00		168/196	1,00			11/12	1,00		50/41	1,00		
>13	35/33	1,53	(0,88-2,65)	32/28	1,41	(0,79-2,52)		16/6	3,16	(0,55-18,08)	59/34	1,39	(0,74-2,63)	
Prise de contraceptifs Oraux							0,22							0,86
Jamais	50/47	1,00		55/69	1,00			21/14	1,00		79/54	1,00		
A déjà pris	133/172	0,55	(0,31-0,97)	145/155	1,01	(0,59-1,74)		6/4	0,84	(0,10-7,42)	30/21	1,16	(0,54-2,52)	
Parité [a]							0,36							0,90
Nullipare	16/28	1,00		18/22	1,00			2/1	1,00		14/9	1,00		
Pare	167/191	1,51	(0,75-3,05)	182/202	0,87	(0,42-1,83)		25/17	0,92	(0,03-26,45)	95/66	0,79	(0,29-2,19)	
Statut ménopausique [a]							0,84							0,79
Pré-ménopause	72/81	1,00		65/84	1,00			13/9	1,00		17/11	1,00		
Post-ménopause	99/83	1,12	(0,23-5,35)	91/81	3,43	(1,36-8,66)		92/64	0,75	(0,25-2,23)	10/6	7E-10	(0-inf)	
GSTT1														
Age aux 1er règles [a](année)							0,02							0,73
≤13	262/311	1,00		59/63	1,00			29/26	1,00		25/19	1,00		
>13	59/38	1,79	(1,14-2,82)	9/20	0,57	(0,21-1,54)		39/21	1,54	(0,67-3,54)	24/15	1,47	(0,50-4,36)	
Prise de contraceptifs Oraux							0,42							0,87
Jamais	88/84	1,00		20/30	1,00			49/33	1,00		36/25	1,00		
A déjà pris	233/265	0,73	(0,48-1,13)	48/53	0,8	(0,28-2,33)		19/14	1,28	(0,47-3,52)	13/9	0,79	(0,21-2,99)	
Parité [a]							0,02							0,67
Nullipare	33/33	1,00		3/13	1,00			5/2	1,00		8/6			
Pare	288/316	0,86	(0,49-1,51)	65/70	5,71	(1,29-25,2)		63/45	0,44	(0,06-3,04)	41/28	0,84	(0,20-3,63)	
Statut ménopausique [a]							0,89							0,50
Pré-ménopause	23/29	1,00		109/130	1,00			32/21	1,00		44/30	1,00		
Post-ménopause	35/32	1,19	(0,25-5,62)	156/125	1,86	(0,95-3,64)		14/13	1,24	(0,18-8,84)	24/15	0,55	(0,09-3,24)	
***GSTM1* and *GSTT1*[b]**														
Age aux 1ères règles [a] (année)							0,14							0,58
≤13	268/311	1,00		27/34	1,00			31/27	1,00		23/18	1,00		
>13	61/44	1,66	(1,07-2,58)	4/10	0,47	(0,10-2,23)		46/23	1,65	(0,75-3,63)	16/13	1,28	(0,39-4,27)	
Prise de contraceptifs Oraux							0,92							0,92
Jamais	240/266	0,77	(0,50-1,18)	20/27	0,56	(0,12-2,72)		22/14	1,18	(0,45-3,08)	10/9	0,91	(0,20-4,10)	
A déjà pris	89/89	1,00		11/17	1,00			55/36	1,00		29/22	1,00		
Parité [a]														0,65
Nullipare	32/37	1,00		0/7	-	-		5/2	1,00		7/6	1,00		
Pare	297/318	1,01	(0,59-1,74)	31/37	-	-		72/48	0,52	(0,08-3,40)	32/25	0,71	(0,13-3,97)	
Statut ménopausique [a]							0,40							0,49
Pré-ménopause	11/12	1,00		118/141	1,00			23/21	1,00		52/30	1,00		
Post-ménopause	15/19	0,43	(0,03-5,44)	172/135	1,90	(1,00-3,63)		13/10	3,30	(0,34-31,81)	25/18	0,43	(0,07-2,46	

[a] le modèle multivarié inclut l'IMC, la consommation de tabac en verres/jour, la quantité de tabac en paquets année ainsi que les facteurs hormonaux et reproducteur

1.3.3.c. Synthèse des résultats et comparaison avec la littérature

Nous n'avons pas mis en évidence d'effet propre de *GSTM1* et *GSTT1* sur le risque de CDT. Nos résultats sont concordants avec une méta-analyse récente de 12 études, qui n'avait pas montré d'association entre les CDT et les génotype de *GSTM1*, *GSTT1* ou à la combinaison des deux (*GSTM1/GSTT1*) (J. Li et al. 2012).

Il est a noté que les résultats des études inclus dans cette méta-analyse étaient hétérogènes. Certains auteurs avaient rapporté des associations positives entre la délétion complète de *GSTM1* (génotype nul) et le risque de CDT (Canbay et al. 2003; Reis et al. 2010), d'autre auteurs ont rapportés des associations inverses (Siraj et al. 2008; Guilhen et al. 2009) ou, n'ont pas observé d'associations (Hernández et al. 2003; Gaspar et al. 2004; Bufalo et al. 2006; Ho et al. 2006; Stankov et al. 2006; Lemos et al. 2008). De tel résultats sont également rapportés concernant *GSTT1* (Siraj et al. 2008; Reis et al. 2010; Morari et al. 2002; Hernández et al. 2003; Gaspar et al. 2004; Bufalo et al. 2006; Ho et al. 2006; Stankov et al. 2006; Lemos et al. 2008; Guilhen et al. 2009). L'hétérogénéité observée dans ces études pourrait être expliquée par une modification de l'effet des gènes en rapport avec les facteurs de risque environnementaux des CDT, non pris en compte dans les études, mais variables d'une population à l'autre.

Notre étude, menée dans deux populations ayant des structures génétiques différentes et des expositions aux facteurs environnementaux variables, a permis d'étudier les interactions gène-environnement. Nos résultats suggèrent l'existence d'interactions entre *GSTT1* et *GSTM1* et certains facteurs de risque en rapport avec le mode de vie. Nous avons ainsi observé que l'association entre les CDT et l'obésité ou la consommation d'alcool est plus marquée chez les individus ayant un génotype nul pour *GSTM1* et/ou *GSTM1*. Des interactions significatives entre *GSTT1* et l'âge aux premières règles ou la parité ont également été observées chez les femmes européennes.

Compte tenu du rôle de *GSTM1* et *GSTT1* dans la détoxification de divers composés, l'étude des interactions de ces gènes avec les facteurs de risque environnementaux semble essentielle pour mieux comprendre leur rôle dans l'étiologie des CDT.

Interaction avec obésité

L'obésité a été positivement associée au risque de CDT dans des études menées dans des populations variées (Kitahara et al. 2014; Zhang et al. 2014; Marcello et al. 2014; Ma et al. 2015; Secretan et al. 2016; Guignard et al. 2007). Elle est considérée par le CIRC comme un facteur de risque avéré de CDT (Secretan et al. 2016). Cette association est plus marquée chez les sujets ayant un génotype nul pour *GSTM1* et/ou *GSTT1* dans les deux populations étudiées. A notre connaissance, la seule étude ayant évalué l'interaction entre l'IMC et les GSTs n'a pas montré de modification d'effet selon le génotype (Kweon et al. 2014). Dans cette étude menée sur une population coréenne de 1372 cas et 1669 témoins, la proportion de sujets en surpoids était plus faible (34,8%) que dans les populations étudiées ici (46,94% chez les européens et 64,76% chez les mélanésiens) ; ce qui pourrait expliquer une plus faible capacité à mettre en évidence une interaction. Des études de plus grande taille et sur les populations ayant une plus forte prévalence de l'obésité sont néanmoins nécessaires pour confirmer nos résultats.

Les mécanismes physiopathologiques associant l'obésité au risque de CDT sont encore mal connus. L'augmentation du tissu adipeux et l'insulino-résistance sont à l'origine d'un stress oxydatif et de la formation de dérivés réactifs de l'oxygène (ROS « Reactive Oxygene Species ») (Pazaitou-Panayiotou et al. 2013). *GSTM1* et *GSTT1* contribuent à la réduction du stress oxydatif en favorisant l'élimination des ROS. Les individus obèses ayant une délétion de *GSTM1* et *GSTT1* pourraient ainsi avoir un risque de CDT plus élevé que les individus obèses de génotype non-nul pour ces gènes.

Interaction avec la consommation d'alcool et de tabac

La consommation d'alcool a été associée négativement au risque de CDT dans plusieurs études (Meinhold et al. 2009; Kitahara, Linet, et al. 2012; Sen et al. 2015; Kabat et al. 2012; Cho & Kim 2014) ainsi que dans une méta-analyse récente (Hong et al. 2016). Cette association inverse est confirmée dans nos données, et est plus marquée chez les sujets *GSTM1*-nul et/ou *GSTT1*-nul que chez les sujets sans délétion du gène. Chez les européens l'interaction entre ces gènes et la consommation d'alcool est significative. Des interactions entre GSTs et l'alcool ont été décrites dans les cancers du sein (Zheng et al. 2003), du poumon (Mota et al. 2015) et de l'estomac (Lao et al. 2014). Notre étude est la première à rapporter de tels résultats dans les CDT.

Le mécanisme physiopathologique à l'origine de cette interaction n'est pas élucidé. L'alcool ou les radicaux libres issus du métabolisme de l'alcool peuvent entraîner un dérèglement de l'axe hypothalamo-hypophysaire (Hong et al. 2016; Valeix et al. 2008), et avoir un impact sur la fonction thyroïdienne (Valeix et al. 2008; Balhara & Deb 2013). Cependant, ce mécanisme ne permet pas d'expliquer la diminution de risque de CDT liée à la consommation d'alcool ou l'effet synergique de l'alcool et des génotypes nul des GSTs.

Contrairement à l'alcool, la consommation de tabac n'est pas associée dans nos données au risque de CDT, et aucune modification d'effet selon les génotypes nul ou non-nul des gènes GSTs n'est observée. Ce résultat est concordant avec ceux d'une étude précédente ayant testé les interactions entre les GSTs et la consommation de tabac (Huang et al. 2009)

Interaction avec les facteurs hormonaux et reproductifs

Du fait de l'incidence des CDT plus élevée chez les femmes que chez les hommes, on a depuis longtemps suspecté l'existence d'un rôle majeur des hormones sexuelles féminines dans la carcinogenèse thyroïdienne, sans toutefois parvenir à éclaircir le rôle des facteurs hormonaux et reproductifs. L'âge tardif aux premières règles était positivement associé au risque de CDT dans nos données. Ce résultat a été rapporté dans plusieurs études (Negri et al. 1999; Horn-Ross et al. 2011; Truong et al. 2005) mais pas dans une méta-analyse récente (Cao et al. 2015). Nous avons également rapporté une association inverse entre la prise de contraceptifs oraux et le risque de CDT, comme chez d'autres auteurs (Zamora-Ros et al. 2015; Xhaard et al. 2014). Le risque de CDT n'était pas associé à la parité et était légèrement augmenté chez les femmes ménopausées par rapport aux non ménopausées. Ces résultats sont détaillés par ailleurs pour les femmes de l'étude CATHY (Cordina-Duverger et al., en révision) et de l'étude de NC (Truong et al. 2005).

Notre étude est la première à étudier l'interaction entre le génotype GSTs et les facteurs hormonaux et reproductifs dans les CDT. Aucune interaction entre le *GSTM1* et l'âge aux premières règles, la prise de contraceptifs oraux, la parité ou la ménopause n'a été observée dans les populations étudiées. Une interaction significative de *GSTT1* avec l'âge aux premières règles et la parité a été observée chez les femmes d'origine européenne.

Les mécanismes biologiques pouvant expliquer le rôle des facteurs hormonaux et reproductifs dans la pathogénèse des CDT sont là encore très mal connus. Il a été suggéré que des taux élevés d'œstrogènes pendant la grossesse pourraient influencer la prolifération des cellules thyroïdiennes malignes (Xu et al. 2013; Derwahl & Nicula 2014; Di Vito et al. 2011). Le métabolisme des œstrogènes entraîne la production de quinones et d'autres molécules cytotoxiques qui sont éliminées par *GSTM1* et *GSTT1* (Mitrunen & Hirvonen 2003). Il est de ce fait possible que la capacité des gènes *GSTM1* et *GSTT1* à éliminer les métabolites des œstrogènes soit modifié par les délétions de ces gènes. Une interaction entre la parité et les génotypes GSTs a été rapportée dans une étude sur le cancer du sein (Park et al. 2003).

En conclusion, nos résultats suggèrent que les génotypes de *GSTM1* et *GSTT1* peuvent modifier les associations entre le risque de CDT et l'obésité et la consommation d'alcool. Les disparités dans les fréquences de délétion de *GSTM1* et *GSTT1,* et dans la prévalence de l'obésité plus fréquente chez les mélanésiens que chez les européens, peuvent expliquer une partie de la plus forte incidence observée chez les mélanésiens.

1.3.4. Fine-mapping des loci 9q22 et 14q13 dans les cancers différenciés de la thyroïde chez les Européens et Mélanésiens de France Métropolitaine et de Nouvelle-Calédonie

1.3.4.a. Introduction

Les polymorphismes associés au risque de maladie dans les GWAS ne sont pas nécessairement causaux et peuvent être en déséquilibre de liaison avec un variant causal qui reste à déterminer. Ceci peut s'expliquer par exemple par la non inclusion du variant causal dans la puce ou parce que le polymorphisme causal est d'une nature différente des marqueurs étudiés. Il convient alors de rechercher tous les SNP en déséquilibre de liaison avec les SNP GWAS et de les prioriser selon leur fonction biologique.

Les études GWAS sur les CDT ont mis en avant de façon robuste les variants rs1867277 et rs965513 du locus 9q22 et les variants rs944289 et rs116909374 du locus 14q13. Ces deux loci seront analysés dans cette partie.

Locus 9q22

Le SNP rs965513 dans le locus 9q22 près de *FOXE1* (à 57 kb en amont du gène) a été mis en évidence dans la première étude GWAS sur les CDT (Gudmundsson et al. 2009). La même année, une étude gène candidat suivie d'une approche fonctionnelle avait mis en évidence le polymorphisme rs1867277 situé dans le promoteur de *FOXE1* (Landa et al. 2009). Les auteurs de cette étude avaient ensuite montré que l'allèle à risque A de rs1867277 pouvait altérer la régulation de l'expression de *FOXE1* en modifiant le site de fixation des facteurs de transcription USF1/USF2 qui sont impliqués dans la transcription et la régulation de l'expression du gène (Landa et al. 2009). Par ailleurs, les variants rs965513 et rs1867277 avaient été répliqués par la deuxième étude GWAS publiée en 2010 sur une population de Biélorussie exposée aux radiations ionisantes lors de l'accident de Tchernobyl (Takahashi et al. 2010). De ces études, a émergé la question de savoir si les deux variants sont indépendamment associés au risque de CDT ou s'ils représentent un même signal ; et s'il existe d'autres SNP associés au CDT dans le locus pouvant être liés à des fonctions biologiques d'intérêt.

Locus 14q13

Le SNP rs944289 situé en amont de *NKX2-1* dans le locus 14q13 a avait également été mis en évidence par la première étude GWAS (Gudmundsson et al. 2009). Ce SNP est situé

dans une région inter génique et aucun rôle fonctionnel n'était connu pour ce polymorphisme. Pour mieux comprendre le rôle du SNP rs944289 dans le risque de CDT, une étude fonctionnelle publiée en 2012 a caractérisé la région contenant ce SNP et a mis en évidence un lincRNA (« long intergenique non coding RNA ») exprimé dans les tissus de la thyroide dénommé *PTCSC3 pour « Papillary thyroid carcinoma suceptibility candidats 3 »* (Jendrzejewski et al. 2012). *PTCSC3* est un lincRNA, long de 40 kb et contenant 3 exons ; le promoteur de ce gène est situé à 3.2 kb en amont du SNP GWAS rs944289. Les auteurs avaient montré que ce lincRNA est surexprimé dans les tissus thyroïdiens sains et sous exprimé dans les tissus cancéreux (Jendrzejewski et al. 2012). Les auteurs avaient également montré que l'allèle à risque T de rs944289 modifie l'expression du lincRNA dans les cellules cancéreuses de la thyroïde. En effet, rs944289[T] diminue l'affinité des facteurs de transcriptions p42, CEBP/α et CEBP/β à leurs sites de fixation ce qui entrainerait une diminution de l'activation du promoteur de *PTCSC3* et par conséquent une réduction significative de l'expression du gène (Jendrzejewski et al. 2012).

Un autre variant du locus 14q13, rs116909374, avait été mis en évidence dans une étude GWAS (Gudmundsson et al. 2012). Ce SNP est également situé dans la région intergénique en amont de *NKX2-1* et en aval du premier SNP GWAS rs944289 du locus. De même que dans le locus 9q22, se pose la question de savoir si les deux SNP (rs944289 et rs116909374) sont indépendamment associés au risque de CDT ou s'ils sont la résultante d'un seul et même signal ; et s'il existe sur le locus, d'autres SNP pouvant expliquer la relation avec les CDT observée.

Objectif

L'objectif général de cette partie était d'étudier plus en détail, le rôle des polymorphismes des loci 9q22 et 14q13 dans le risque de CDT dans une population européenne de France métropolitaine et mélanésienne de Nouvelle-Calédonie. Plus spécifiquement, les objectifs étaient de : 1) Répliquer dans un premier temps les SNP rs965513 et rs1867277 du locus 9q22 et les SNP rs944289 et rs116909374 du locus 14q13 dans les populations européenne et mélanésienne de France métropolitaine et de NC ; deux populations dans lesquels ces polymorphismes n'avaient jamais été investigués. 2) Rechercher s'il existe dans les loci 9q22 (entre rs965513 et FOXE1) et 14q13 (entre *PTCSC3* et *NKX2-1*) d'autres variants associés au risque de DTC pouvant être de meilleurs candidats d'un point de

vue fonctionnel et/ou statistique que ceux précédemment identifiés. 3) Tester les interactions entre les variants d'intérêts identifiés et les facteurs de risque suspectés de CDT tels que l'IMC, la consommation de tabac et d'alcool et les facteurs hormonaux et reproductifs.

1.3.4.b. Résultats

i. Locus 9q22

Dans le locus 9q22, rs965513 et rs1867277 sont situés dans des blocs de LD différents, chez les européens comme chez les mélanésiens (figure 5C et 5D). Toutefois, on observe que le bloc de LD contenant le SNP rs1867277 est plus large chez les mélanésiens que les européens (figure 5C et 5D). Globalement, les fréquences alléliques étaient légèrement plus faibles chez les mélanésiens que chez les européens mais les OR d'association entre les SNP et le risque de CDT étaient du même ordre de grandeur (Tableau 11 et annexe 6).

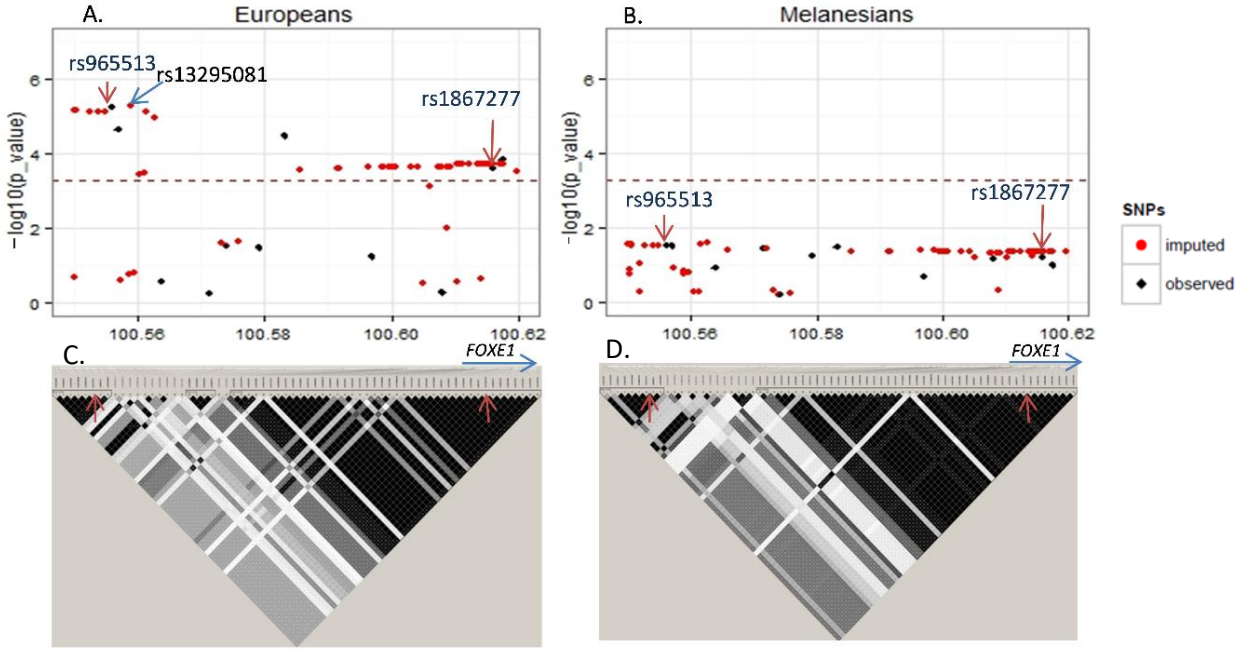


Figure 5 : Manhattan plot de l'association entre les SNP du locus 9q22 et le risque de CDT chez les européens (A) et les mélanésiens (B) ainsi que les corrélations entre SNP dans chacune des populations (C et D). Les SNP génotypés et imputés sont représentés selon leurs positions sur le chromosome. L'axe des abscisses correspond à la position du SNP sur le chromosome et l'axe des ordonnées représente la transformation logarithmique des p-values d'association. Les flèches rouges représentent les SNP GWAS et les flèches bleu, les SNP ayant les signaux les plus robustes.

Tableau 11: OR et IC de l'association entre les tops SNP et SNP GWAS des loci 9q22 et 14q13 chez les européens et les mélanésiens

					européens				mélanésiens				
			Allèle		MAF[4]				MAF[4]				
Locus	SNPs	Position	Ref[3]	risk	Cas	Témoins	OR (95% IC)	p-values	Cas	Témoins	OR(95% IC)	p-values	phet
9q22,33	rs13295081[1]	100559011	C	T	0,49	0,40	1,44 (1,22 - 1,70)	$5,46 \times 10^{-6}$	-	-	-	-	
	rs965513[2]	100556109	G	A	0,43	0,33	1,52 (1,28 - 1,81)	$5,87 \times 10^{-6}$	0,38	0,30	1,43 (1,00 - 2,07)	0,03	0.98
	rs10759944	100556972	G	A	0,43	0,34	1,48 (1,25 - 1,76)	$2,25 \times 10^{-5}$	0,39	0,31	1,42 (0,95 – 2,05)	0,03	0.90
	rs1867277[2]	100615914	G	A	0,46	0,38	1,38 (1,17 - 1,63)	$2,59 \times 10^{-4}$	0,36	0,29	1,34 (0,93 - 1,94)	0,06	0.92
14q13,3	rs1755774	36688516	A	G	0,29	0,34	0,76 (0,63- 0,92)	$7,29 \times 10^{-3}$	0,53	0,70	0,50 (0,35 - 0,71)	$1,62 \times 10^{-5}$	0.09
	rs2787423	36687054	G	A	0,29	0,34	0,78 (0,65 - 0,94)	$8,78 \times 10^{-3}$	0,54	0,69	0,51 (0,35 - 0,73)	$3,69 \times 10^{-5}$	0.09
	rs7494749	36615427	G	A	0,38	0,45	0,73 (0,62 - 0,87)	$4,34 \times 10^{-4}$	0,64	0,76	0,57 (0,39 - 0,83)	$7,94 \times 10^{-4}$	0.15
	rs944289[2]	36649246	T	C	0,34	0,41	0,74 (0,61 - 0,89)	$4,78 \times 10^{-4}$	0,62	0,74	0,57 (0,39 - 0,83)	$1,13 \times 10^{-3}$	0.15
	rs116909374[2]	36738362	C	T	0,05	0,03	1,97 (1,26 - 3,10)	$3,25 \times 10^{-3}$	0,01	0,004	1,44 (0,13 - 15,96)	0,76	0.84
	rs999460	36984511	C	T	0,34	0,39	0,8 (0,68 - 0,96)	$5,18 \times 10^{-3}$	0,21	0,19	1,14 (0,74 - 1,76)	0,59	0.11

[1]rs13295081 n'a pas passé le contrôle qualité dans la population mélanésienne.[2] SNP GWAS, [3]ref : allèle de référence.[4]MAF : fréquence de l'allèle mineur ou de l'allèle à risque. Les ORs sont ajustés sur l'âge, le sexe et le département de résidence.

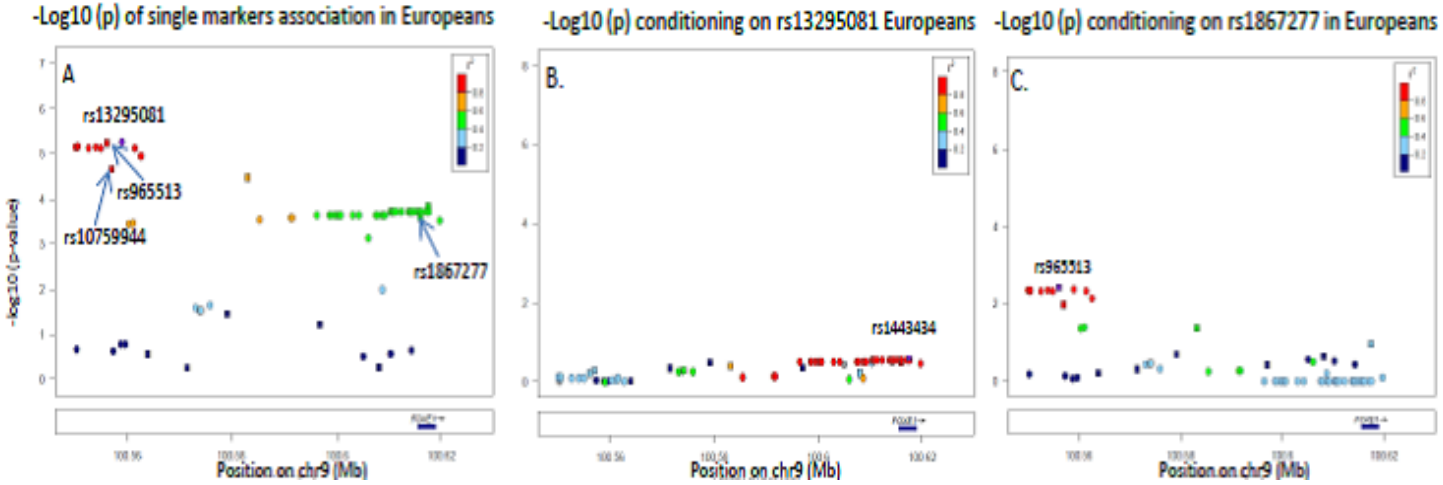


Figure 6 : plot des log10 p-values d'association entre les SNP de la région 9q22 et le risque de CDT chez les européens (A) ainsi que les résultats des analyses conditionnelles le SNP rs13295081 (B) et rs1867277 (C). Les ronds représentent les SNP imputés et les carrés représentent les SNP génotypés. Les couleurs représentent les corrélations (r^2) entre les SNP, elles vont des plus corrélés (en violet) au moins corrélé (bleu foncés).

Population européenne

Chez les européens, sur 69 SNP analysés, 52 SNP étaient significativement associés au risque de CDT ($p< 4 \times 10^{-4}$) parmi lesquels les SNP rs965513 (OR (rs965513[A]) = 1.52, $p=6\times10^{-6}$) et rs1867277 (OR(rs187277[A]) = 1.38, $p=3\times10^{-4}$) (figure 5A,Tableau 11). Le signal le plus significatif était observé pour le SNP imputé rs13295081 (OR (rs13295081[T] = 1.44, $p=5\times10^{-6}$). Ce SNP est situé dans la région intergénique près du SNP GWAS rs965513 et est fortement corrélé à ce dernier ($r^2=1$, figure 5C) ce qui suggère que les associations observées avec ces deux SNP résultent d'un même signal. Parmi les 48 autres SNP significativement associés au risque de CDT, 14 SNP étaient fortement ou modérément corrélés aux SNP rs13295081 et rs965513 ($r^2>0.60$) et 36 SNP étaient fortement corrélés à rs1867277 ($r^2>0.80$).

Aucun signal d'association ne persiste lorsque l'on conditionne les analyses sur rs13295081 (ayant montré le plus fort signal) ou rs965513 (figure 6B). Par contre, des associations à la limite de la significativité sont observées pour ces SNP et d'autres SNP du même bloc de LD lorsqu'on prend en compte l'effet de rs1867277 (figure 6C). Ces résultats indiquent que les SNP rs965513 et rs1867277 ne sont pas indépendamment associés au risque de CDT. Les signaux prédominants sont dans le bloc de LD contenant rs965513.

Population mélanésienne

Chez les mélanésiens, aucun des 75 SNP analysés n'étaient significativement associés au risque de CDT après correction sur les tests multiples (figure 3B, Tableau 11 et annexe 5). Toutefois, pour les SNP rs965513 (OR (rs965513[A]) = 1.43, p=0,03) et rs1867277 (OR (rs187277[A]) = 1.34, p=0,06), des ORs comparables à ceux observés dans la population européenne ont été observés. Le SNP rs13295081 n'a pas été analysé dans cette population car il n'a pas passé le contrôle qualité de l'imputation (score INFO=0,4).

ii. Locus 14q13

Dans le locus 14q13, on observe que la taille des blocs de LD est relativement plus large chez les mélanésiens que chez les européens. Les fréquences alléliques sont très différentes chez les européens et chez les mélanésiens. En revanche les OR d'association avec les CDT

allaient dans le même sens dans les deux populations et étaient du même ordre de grandeur pour la plupart des SNP (Tableau 11 et annexe 6).

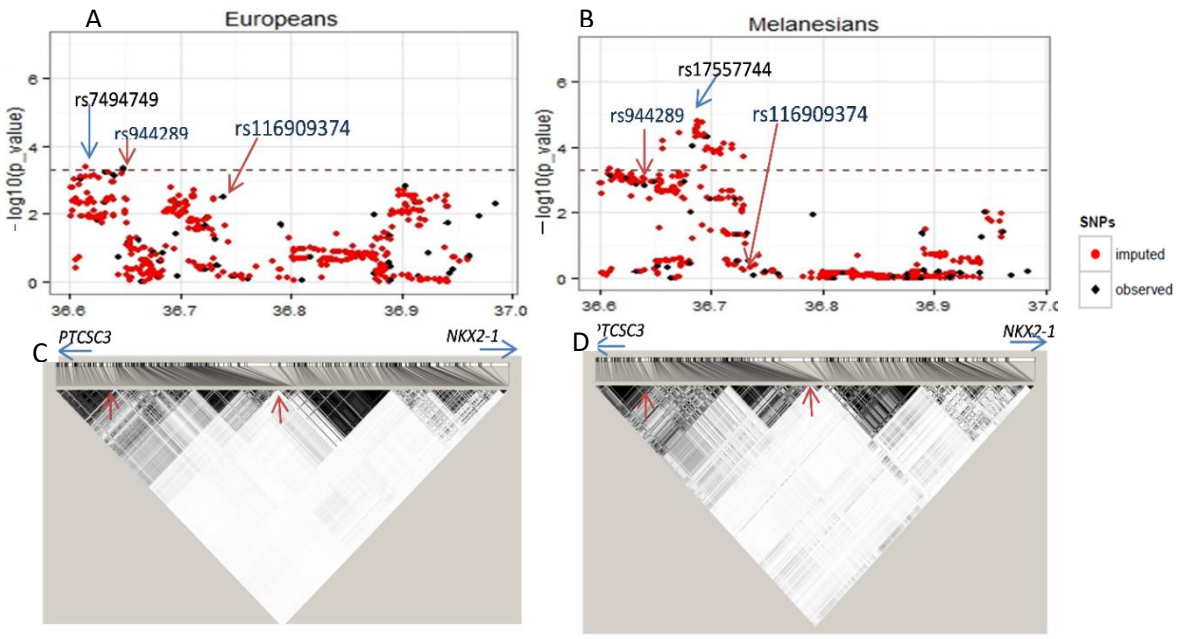


Figure 7: Manhattan plot de l'association entre les SNP du locus 14q13 et le risque de CDT chez les européens (A) et les mélanésiens (B) et les corrélations (r^2) entre SNP dans chacune des populations (C et D). Les SNP génotypés et imputés sont représentés selon leurs positions sur le chromosome. L'axe des abscisses correspond à la position du SNP sur le chromosome et l'axe des ordonnées représente la transformation logarithmique des *p-values* d'association. Les flèches rouges représentent les SNP GWAS et les flèches bleu, les SNP ayant les signaux les plus robustes dans le locus.

Population européenne

Sur 525 SNP analysés dans cette population, seul le SNP imputé rs7494749 était significativement associé au risque de CDT (OR (rs7494749[A]) = 0,73, p=4,3x10^{-4}, figure 7A). Ce SNP est situé dans le 3ème intron de *PTCSC3*, en amont du SNP GWAS rs944289 et est fortement corrélé à ce dernier (r^2=0.98, figure 5C). Rs944289 était le second SNP ayant le plus fort signal (OR (rs944289[C]) = 0,74, p=4,7x10^{-4}). Une association à la limite de la significativité a également été observée pour le SNP GWAS rs116909374 (OR (rs116909374[T]) = 1,92 ; p=3,3x10^{-3}).

Les analyses conditionnelles ont montré que lorsqu'on prend en compte l'effet des SNP rs7494749 ou rs944289, l'association avec le SNP GWAS rs116909374 persiste (figure 8B). En conditionnant sur les SNP rs944289 et rs116909374, une association persiste pour le SNP rs999460 ($p<4x10^{-3}$) (figure 8C). Ces résultats suggèrent que dans le locus 14q13 les SNP rs944289, rs116909374 et rs999460, sont indépendamment associés au risque de CDT chez les européens.

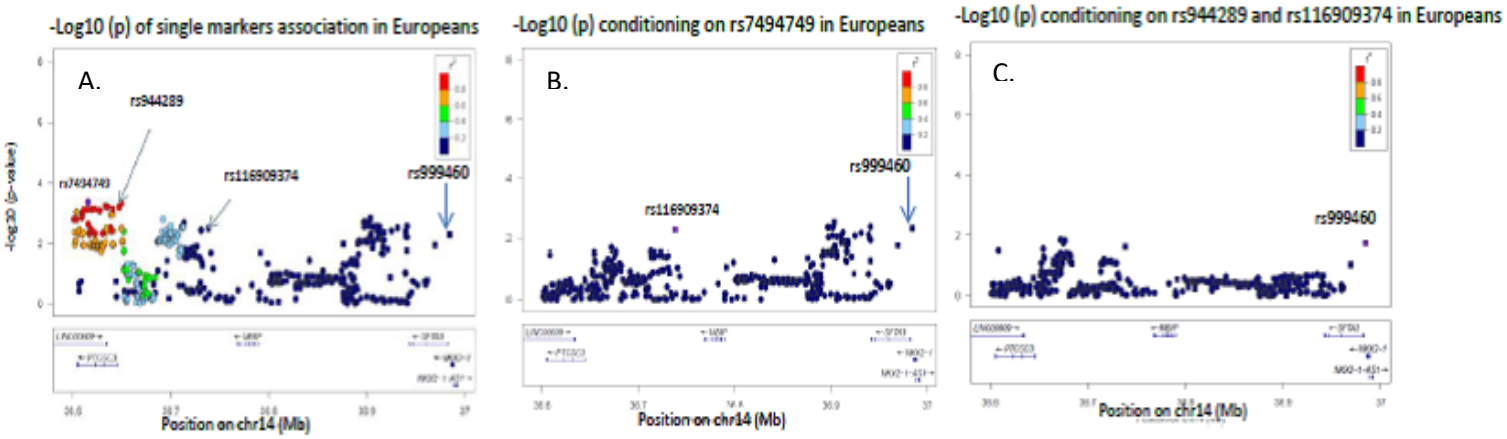


Figure 8 : plot chez les Européens des p-values des résultats d'association entre les SNP de la région 14q13 et les CDT chez les Européens (A) et de l'analyse conditionnelle sur le SNP rs13295081 (B) et rs1867277 (C). Les ronds représentent les SNPs imputés et les carrés représentent les SNP génotypés. Les couleurs représentent les corrélations (r^2) entre les SNP, depuis les plus fortes (en violet) au moins fortes (bleu foncé).

L'analyse des différentes combinaisons d'haplotypes constituées par les SNP rs944289, rs116909374 et rs999460 a montré une diminution significative du risque de CDT chez les individus portant l'haplotype rs944289[C]-rs116909374[C]-rs999460[T] (OR=0,51, $p=3,11x10^{-6}$) par rapport à l'haplotype de référence rs944289[T]-rs116909374[C]-rs999460[C] (Tableau 12).

Tableau 12: Résultats d'analyse des combinaisons d'haplotypes les plus fréquents pour les SNP rs944289, rs116909374 et rs999640 dans le locus 14q13 chez les Européens.

Haplotype			Fréquences				
rs944289	**rs116909374**	**rs999460**	**Cas**	**Témoins**	**OR**	**IC (95%)**	**p-value**
T	C	C	0,39	0,36	1,00	ref	1,00
C	C	C	0,23	0,23	0,93	(0,74-1,15)	0,48
T	C	T	0,22	0,21	0,93	(0,74-1,16)	0,52
C	**C**	**T**	**0,10**	**0,17**	**0,51**	**(0,38-0,68)**	**$3,1 \times 10^{-6}$**
T	**T**	**C**	**0,03**	**0,01**	**2,47**	**(1,25-4,86)**	**$8,9 \times 10^{-3}$**
C	T	T	0,02	0,006	3,16	(1,27-7,86)	0,01

Population mélanésienne

Sur 511 SNP analysés, 38 SNP étaient significativement associés au risque de CDT (annexe 7, figure 7B). Ces SNP sont tous situés dans la région intergénique près de *PTCSC3* et la plupart d'entre eux étaient fortement corrélés au SNP rs1755774 ayant montré le plus fort signal (OR (rs1755774[G]) = 0.50 p=1.6x10^{-4}, Tableau 11 et figure 7B). On observe une association à la limite de la significativité pour le SNP GWAS rs944289 et les CDT (rs944289[C]) = 0.57, p=1.1x10^{-3}) alors que rs116909374 n'était pas associé au risque de CDT (OR (rs116909374[T]) = 1.44, p=0.76) (Tableau *11* et figure 7B). On observe toutefois que ce SNP a une fréquence allélique très faible dans cette population mais que l'OR observé chez les mélanésiens va dans le même sens que celui des européens (Tableau 11). En revanche, le SNP rs944690 retrouvé chez les européens et formant avec rs944289 et rs116909374 l'haplotype à risque n'a pas été répliqué chez les mélanésiens (Tableau 11).

L'analyse conditionnelle sur le SNP rs1755774 n'a révélé aucun autre signal persistant (figure 9B) suggérant ainsi une non-indépendance des signaux d'association observés sur le locus chez les mélanésiens.

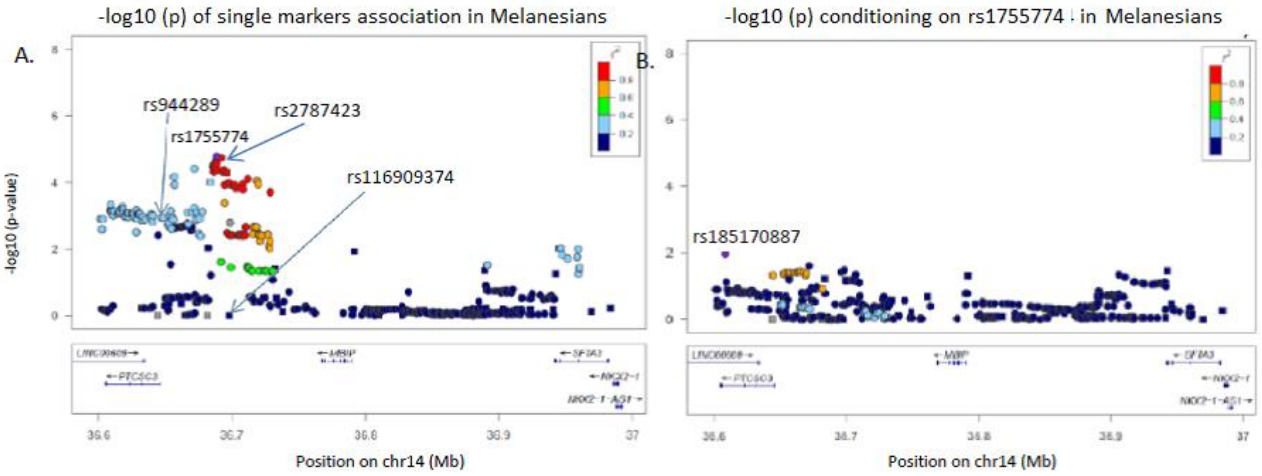


Figure 9 : plot chez les mélanésiens des p-values des résultats d'association entre les SNP de la région 14q13 et les CDT (A) ainsi que les résultats d'analyse conditionnelle sur le SNP rs1755774 (B). Les ronds représentent les SNP imputés et les carrés représentent les SNP génotypés. Les couleurs représentent les corrélations (r^2) entre les SNP, depuis les plus fortes (en violet) au moins fortes (bleu foncé.

iii. Prédiction fonctionnelle des SNP a risque dans les loci 9q22 et 14q13

Locus 9q22

L'annotation fonctionnelle des SNP significativement associés au risque de CDT chez les européens nous a permis de prédire *in silico*, des modifications fonctionnelles pouvant être en rapport avec les CDT pour le SNP rs10759944 situé près de rs965513 et fortement corrélé à ce dernier (r^2=0.99, figure 6A). En effet, selon les données Chip-seq, rs10759944 est situé dans le site de site fixation de la protéine régulatrice ESR1 et son allèle à risque modifierait le site de fixation des facteurs de transcription (FT) tels qu'E2F1 et AP-2α selon les scores de prédiction de PWM.

ESR1, E2F1 et AP-2α sont des facteurs de transcription suspectés dans la pathogénèse des CDT (Di Vito et al. 2011; Onda et al. 2004; Chiefari et al. 2002).

ESR1 est un récepteur nucléaire oestrogénique activé par les hormones stéroïdiennes. Ce récepteur est impliqué dans la régulation et l'expression des gènes et affectent la prolifération et la différenciation cellulaire des tissus cibles. ESR1 est exprimé dans la thyroïde et sa surexpression a été rapporté dans les tissus tumoraux de CPT (Di Vito et al. 2011; Xu et al. 2013).

E2F1 est un FT contrôlant le cycle cellulaire et agissant comme une protéine suppresseur de tumeur (Onda et al. 2004) et AP2-α est une protéine interagissant au niveau cellulaire avec les enhancers pour la régulation de la transcription des gènes cibles (Chiefari et al. 2002). La surexpression d'E2F1 et AP2-α a également été montrée dans les CDT mais le mécanisme biologique liant ces FT au risque de CDT n'est pas clairement connu/défini (Onda et al. 2004; Chiefari et al. 2002).

Locus 14q13

Dans le locus 14q13, nous avons effectué l'annotation fonctionnelle des SNP rs116909374 et rs999640 indépendamment associés au risque des CDT dans la population européenne ainsi que tous les SNP significativement associé au risque de CDT chez les mélanésiens. Nous avons observé in silico que le SNP rs116909374 est situé sur le site de fixation de la protéine régulatrice ESR1 ; ce SNP est également situé un enhancer selon l'analyse de données Chip-Seq des cellules H1 ainsi que dans le promoteur d'un lincRNA non

caractérisé (TCONS_00022711) exprimé en abondance dans la thyroïde. Le SNP rs999640 est localisé dans un promoteur peu actif selon les données Chip-Seq des cellules H1 et sur le site de fixation de la protéine EZH2. Le score d'analyse des PWM montre une modification du site de fixation des FT ZNF143 et NKX2-1 par l'allèle à risque T de ce SNP.

EZH2 est la sous-unité catalytique du complexe protéique PRC2 (Polycomp Repressive Complex 2). Cette protéine intervient dans la méthylation de l'ADN ainsi que la répression de la transcription des gènes impliqués dans le développement et la différenciation cellulaire (Simon & Lange 2008; Zhou et al. 2016). La surexpression de la protéine EZH2 a été retrouvée dans plusieurs cancers incluant les CDT. De plus, EZH2 semble impliqué dans la régulation de la transcription de *NKX2-1* (Morey & Helin 2010) suggérant ainsi qu'une altération du site de fixation de cette protéine par l'allèle a risque T de rs999460 pourrait modifier la transcription et l'expression de *NKX2-1*. Toutefois, l'analyse des données eQTL n'a pas montré de corrélation entre le niveau d'expression de *NKX2-1* et rs999460.

Parmi les SNP significativement associés au risque de CDT dans le locus 14q13 chez les mélanésiens, rs2787423 et rs67263215 ont des fonctions prédites pouvant être liées au risque de CDT. Selon les données Chip-Seq, ces SNP sont situés dans le site de fixation des protéines STAT3 et FOS (figure 10). L'analyse des PWM suggère que l'allèle à risque G du SNP rs2787423 modifie le site de fixation de p300, CEBP/α et CEBP/β qui sont des protéines normalement exprimées dans les tissus sains de la thyroïde et supprimées dans les tissues tumoraux. Rs67263215[T] elle semble aussi modifier le site de fixation des protéines BDP1, TATA box et YY1 qui ensemble, forment un complexe multi-protéique impliqué dans l'initiation et la transcription de plusieurs gènes (Klug & Beato 1996; Austen et al. 1997). Rs67263215 et rs2787423 sont aussi situés dans un enhancer selon les données chip-seq des cellules HMEC issues de tissu épithélial mammaire humain et en amont d'un lincRNA caractérisé nommée TCONS_0002269 ayant une expression abondante dans les tissues thyroïdiens (figure 8).

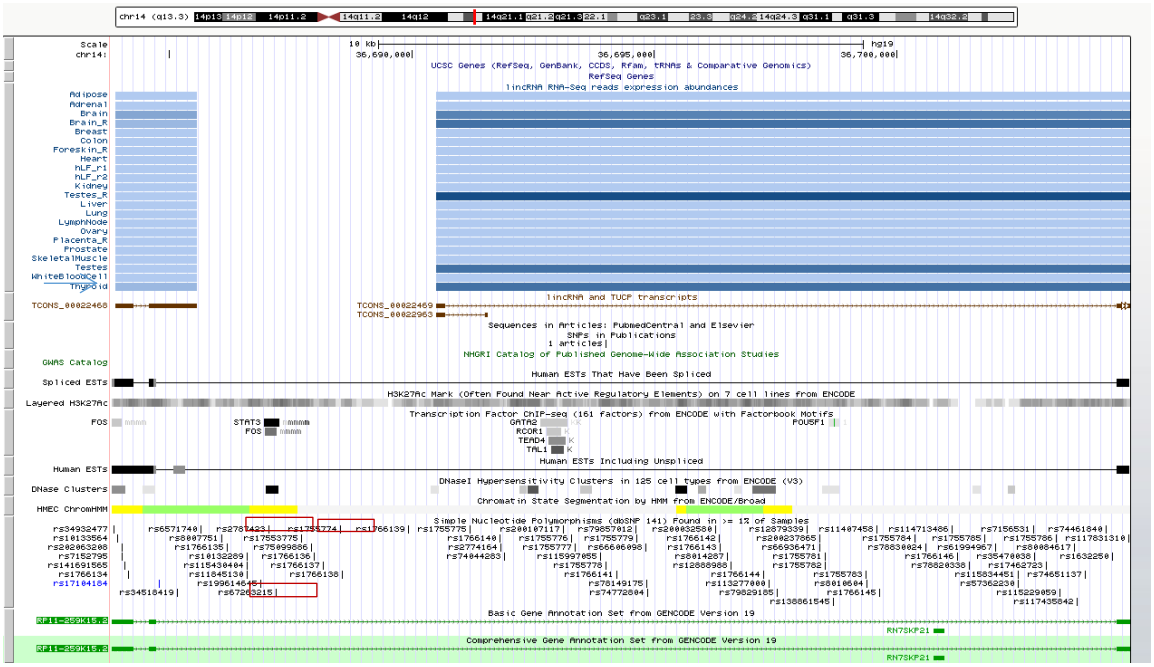


Figure 10 : Snapshot de UCSC genome browser (build 37/hg19) indiquant la région du locus 14q13 où sont localisés les SNP rs1755774, les sites de fixation des protéines régulatrices observés dans des données Chip-Seq ; les lincRNA et leur niveau d'expression dans différents tissues. Les carrés rouges indiquent les SNP rs1755774, rs2787423 et rs67263215. La flèche bleue indique l'abondance de l'expression des lincRNA dans les tissues de la thyroïde (données de RNA_seq).

iv. Interaction entre les SNP candidats des loci 9q22 et 14q13 et les facteurs de risque non génétique

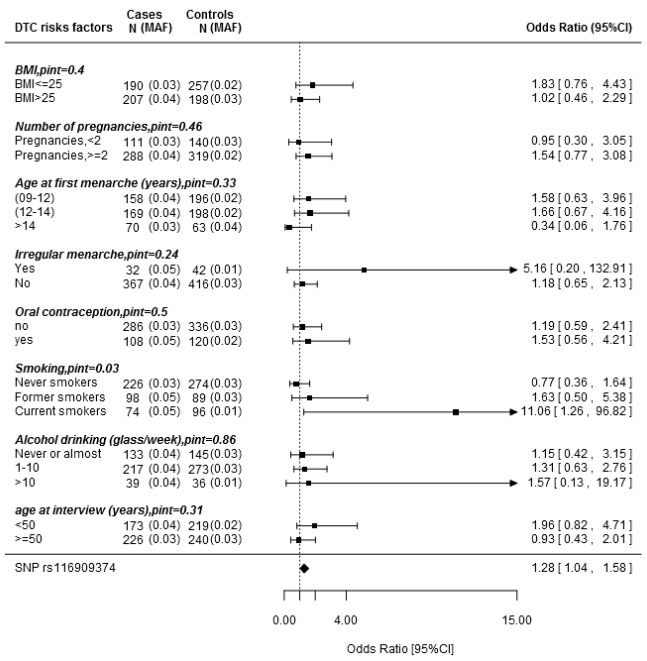
Nous avons étudié les interactions entre les SNP des loci 9q22 et 14q13 et les facteurs de risque environnementaux ou liés au mode de vie.

Les figures 1 et 2 de l'annexe 2 montrent les « forest plot » des associations entre les CDT et les SNP candidats des loci 9q22 (rs965513, rs10759944 et rs1867277) et 14q13 (rs944289, rs1755774, rs2787423, rs116909374 et rs999460) stratifiés sur l'IMC, la consommation d'alcool, le statut tabagique, l'âge aux premières règles, l'irrégularité des règles et le nombre de grossesses à terme chez les européens et les mélanésiens. Dans le locus 9q22, nous n'avons pas trouvé d'interaction entre les SNP candidats et les facteurs de risque précités (annexe 2 : figure 1).

Dans le locus 14q13, nous avons observé une interaction significative entre le SNP rs116909374 et le statut tabagique chez les européennes (p-interaction = 0.03 ; figure 11). En effet, l'association entre l'allèle T du SNP rs116909374 et le risque de CDT était plus marquée chez les fumeurs actuels (OR= 11,06 (IC à 95% 1,26 – 96,82] ; p=0,01) que chez les anciens

fumeurs (OR = 1,63 (IC à 95% : 0,50 – 5,38) ; p=0,13 et les non-fumeurs (OR = 0,77 (IC à 95% : 0.36 – 1.64) ; p=0,70). Cette interaction n'a pas été étudiée chez les mélanésiens du fait de la très fiable fréquence de l'allèle à risque (MAF=0,01).

A) Forest plot SNP rs116909374 chez les européens



B) Forest plot SNP rs175574 chez les mélanésiennes

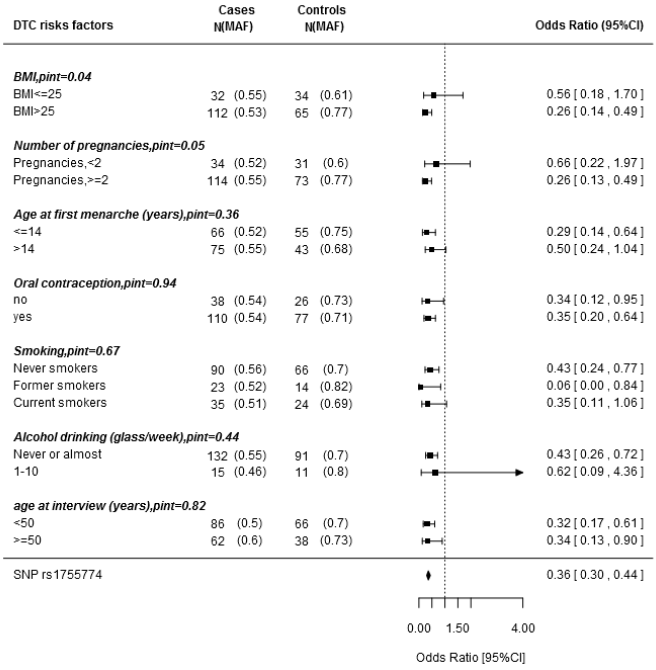


Figure 11 : Forest plot des interactions gènes-environnement des SNP rs116909374 et rs1755774 chez les européennes et les mélanésiennes

Chez les femmes mélanésiennes, nous avons observé une interaction significative entre le SNP rs1755774 et l'IMC (p-interaction = 0,04) ou la parité (p-interaction = 0,04) (figure 11). La diminution du risque de CDT observé avec l'allèle G du SNP rs1755774 était plus marquée chez les femmes ayant eu deux grossesses à terme et plus (OR = 0,26 (IC à 95% : 0,13 – 0,49); $p=3x10^{-5}$) que chez les femmes ayant eu moins de deux grossesses à termes (OR = 0,66 (IC à 95% : 0,22 – 1,97) ; p=0.41). De même, la diminution du risque de CDT observée avec ce SNP était également plus marquée chez les femmes ayant un BMI>25 kg/m² (OR = 0,26 (IC à 95% : 0,14 – 0,49) ; $p=3x10^{-5}$) alors que l'association n'était pas significative chez les femmes ayant un BMI≤25 kg/m² (OR = 0.56 (IC à 95% : 0,18 – 1,70) ; p=0.31). Cependant, ces interactions n'ont pas été répliquées chez les femmes européennes (annexe 2 : figure 2).

1.3.4.c. Synthèse des résultats et comparaison à la littérature

Locus 9q22

Dans le locus 9q22, nous avons répliqué dans les populations européenne et mélanésienne les associations avec le CDT précédemment rapportées pour les SNP rs965513 et rs1867277 (Gudmundsson et al. 2009; Landa et al. 2009). Nous avons montré que ces SNP, localisés dans 2 blocs de LD différents, ne sont pas indépendamment associés au risque de CDT mais que le signal prédominant dans le locus est celui du bloc de LD contenant le SNP rs965513. Ce résultat est concordant avec une précédente étude (Jones et al. 2012) qui avait montré une diminution des signaux d'associations des SNP rs965513 et rs1867277 lorsqu'ils étaient ajustés l'un sur l'autre.

Nous avons également montré que les associations les plus fortes étaient situées dans le bloc de LD contenant le SNP GWAS rs965513. L'annotation fonctionnelle de ce bloc de LD montre que rs10759944 situé près de rs965513 et fortement corrélé à ce dernier, modifierait le site de fixation de la protéine régulatrice ESR1 ou des facteurs de transcription comme E2F1 et AP-2α. ESR1, E2F1 et AP-2 α sont des protéines dont l'expression est modifiée dans les tissus tumoraux de CDT et dont le rôle a été suggéré dans la pathogénèse des CDT (Di Vito et al. 2011; Onda et al. 2004; Chiefari et al. 2002). Rs10759944 constitue donc un candidat potentiellement causal dans le CDT. Des études fonctionnelles in vitro sont nécessaires pour confirmer la relation entre rs1075944 et ces protéines.

La fonction biologique du SNP rs10759944 a également été décrite dans une étude récente ayant exploré le rôle des SNP candidats dans le locus 9q22 sur un intervalle plus large que celui de notre étude (He et al. 2015). Les auteurs ont rapporté la présence d'au moins trois enhancers dans lesquels les SNP rs7864322, rs13252658, rs7847449 et rs10759944 semblent avoir un rôle fonctionnel (He et al. 2015). Parmi ces 4 SNP, seul rs10759944 a pu être analysé dans notre étude. Le SNP rs7864322 est situé au-delà de la région que nous avions définie. Rs12352658 et rs7847449 ont été imputés mais avec une faible qualité d'imputation (r-info<0.7). L'étude d'He et al. montre, en particulier, que l'enhancer contenant le SNP rs10759944 interagirait avec les promoteurs de *PTCSC2* et *FOXE1* dans les tissus thyroïdiens indiquant que cet enhancer et rs10759944 pourraient être actifs dans la thyroïde et que l'altération des sites de régulation par l'allèle à risque du SNP pourrait modifier la

susceptibilité individuelle au risque CDT (He et al. 2015). *PTCSC2* (« Papillary Thyroid Cancer Susceptibility Candidate 2 ») est le LincRNA identifié dans le locus 9q22 en 2015 par les mêmes auteurs. Il comporte 8 isoformes non transcrites dont isoforme C long de plus de 67kb qui chevauche le SNP GWAS rs965513 et le promoteur de *FOXE1* (He et al. 2015). Ce LincRNA, comme *PTCSC3* semble impliqué dans la pathogénèse des CDT car, il est surexprimé dans les tissus de la thyroïde et sousexprimé dans le tissues cancéreux (He et al. 2015).

L'association avec le risque de CDT observé pour le SNP rs10759944 dans notre étude avait également été rapportée dans la première étude GWAS sur la population islandaise (Gudmundsson et al. 2009), dans l'étude GWAS sur des cas radio-induits de cancer papillaires de la thyroïde en Biélorussie(Takahashi et al. 2010) et dans une étude des cas familiaux de CDT (Bonora et al. 2013)

Locus 14q13

Dans le locus 14q13, nous avons répliqué l'association entre le risque de CDT et le SNP GWAS rs944289 chez les européens et les mélanésiens alors que rs116909374 n'a été répliqué que chez les européens. Il est toutefois à noter que l'OR observé pour ce SNP était proche de celui observé chez les européens mais la fréquence de l'allèle mineur de ce SNP est très faible dans la population mélanésienne (MAF=0,004). Nous avons montré que rs944289 et rs116909374 étaient indépendamment associés au risque CDT dans la population européenne. Ce résultat est concordant avec les résultats de Gudmundsson et al (Gudmundsson et al. 2012) montrant une persistance des signaux d'associations des SNP rs944289 et rs116909374 lorsqu'ils étaient ajustés l'un sur l'autre. En plus des SNP rs944289 et rs116909374, nous avons observé que rs999640 était indépendamment associé au risque de CDT chez les européens, l'haplotype rs944289[C]-rs116909374[C]-rs999460[T] constitué par ces trois SNP était significativement associé à une diminution de risque de CDT (OR=0,51, $p=3,11x10^{-6}$) comparé à l'haplotype de référence rs944289[T]-rs116909374[C]-rs999460[C]. Ces résultats suggèrent que cet haplotype pourrait tagguer un variant causal qui reste à identifier ou encore qu'il existe un réel effet combiné de ces SNP sur le risque de CDT.

L'association entre rs944289 et les CDT a été retrouvée dans plusieurs études et dans différentes populations alors que rs116909374 a pu être analysé que dans la population européenne. Ce SNP est le plus souvent monoallélique dans d'autres populations comme les

populations japonaise (Moriya et al. 2014) et polynésienne (Maillard et al. 2015). Une association avec rs999640 avait également été rapportée dans une étude portant sur les cas familiaux de CDT (Bonora et al. 2013).

Concernant le rôle biologique de ces SNP, comme mentionné plus haut, une étude fonctionnelle avait montré que l'allèle T du SNP rs944289 réduisait l'affinité des FT p300, CEBP/α et CEBP/β à leur site de fixation induisant ainsi une réduction de l'activation et de l'expression de *PTCSC3* dans les tissus cancéreux de la thyroïde (Jendrzejewski et al. 2012). Aucune étude n'avait jusqu'à présent exploré le rôle fonctionnel des SNP rs116909374 et rs999460. Nous avons prédit *in silico* que l'allèle à risque T du SNP rs116909374 modifierait le site de fixation de la protéine régulatrice ESR1 ; ce SNP est également situé dans le promoteur d'un petit lincRNA (TCONS_0022711) non exprimé dans la thyroïde. Rs999460 est situé dans un enhancer et son allèle à risque semble altérer le site de fixation de la protéine EZH2 dont la surexpression a été corrélée à la progression des cancers incluant les cancers de la thyroïde par divers mécanismes (Simon & Lange 2008; Zhou et al. 2016). Des études expérimentales sur les SNP rs116909374 et rs999460 seraient nécessaires pour mieux comprendre leurs rôles dans la pathogénèse des CDT.

Dans la population mélanésienne, nous avons mis en évidence de nouveaux SNP candidats à risque de CDT dont le SNP rs1755774 (OR=0.50, $p=1.6x10^{-5}$) qui avait montré le plus fort signal et les SNP rs2787423 et rs67263215 pour lesquels nous avons observé des fonctions prédites *in silico* qui sont potentiellement intéressantes. Ces trois SNP sont situés dans un même bloc de LD et ne sont pas corrélés aux SNP GWAS rapportés dans le locus 14q13. Par ailleurs, nous avons rapporté la présence d'un LincRNA TCONS_0002269 non caractérisé et exprimé dans la thyroïde dans cette région. Ces éléments pourraient agir ensemble ou séparément pour modifier la susceptibilité individuelle au CDT. D'autres études sont nécessaires pour mieux comprendre la relation entre ces variants et le risque de CDT.

Interaction gène-environnement

Dans cette étude nous avons exploré les interactions entre les SNP candidats et : l'IMC, la consommation d'alcool, la consommation de tabac et les facteurs hormonaux et reproductifs. Nous avons mis en évidence, pour la première fois, des interactions entre les SNP rs1755774 et rs2787423 et la parité chez les femmes mélanésiennes ainsi qu'une interaction entre le

statut tabagique et le SNP rs116909374 chez les européennes. Les interactions gène-environnement dans les CDT ont été explorées dans deux études (Pereda et al. 2015; Maillard et al. 2015) entre les SNP rs965513 et rs1867277 du locus 9q22 et les SNP rs944289 et rs116909374 du locus 14q13 et divers facteurs tels que l'IMC, la parité, le BSA (body surface area), la consommation d'iode et la dose de radiation ionisante , mais avec des résultats négatifs. Ces résultats doivent être interprétés avec précaution étant donné le faible nombre d'individus inclus dans certaines strates, du fait de l'absence de mécanismes biologiques connus permettant d'expliquer les interactions avec la parité ou le statut tabagique.

Conclusion

Nous avons confirmé dans la population européenne, les associations avec les SNP GWAS précédemment associés au risque de CDT dans les loci 9q22 et 14q13. Dans le locus 9q22, nous avons mis en évidence un nouveau variant candidat potentiellement intéressant d'un point de vue fonctionnel. Dans le locus 14q13 ; un haplotype à risque chez les européens a été mis en évidence, ainsi que de nouveaux variants de susceptibilité génétique de CDT chez les mélanésiens. Ces résultats doivent être confirmés par d'autres études indépendantes et par des études expérimentales in vivo. On note également que les associations et les fréquences alléliques observées dans les 2 groupes ethniques ne permettent pas d'expliquer la plus forte incidence de CDT chez les mélanésiens par rapport aux européens.

1.4. Discussion et conclusion

Dans cette première partie, l'objectif était d'analyser, dans les populations européenne et mélanésienne, les polymorphismes précédemment associés au risque de CDT dans les études gènes candidats et les études GWAS; d'explorer plus en détail les loci GWAS 9q22 et 14q13 ; et de tester les interactions avec les facteurs de risque environnementaux ou liés au mode de vie.

Au final, notre étude ne confirme pas les associations avec les polymorphismes précédemment associés au CDT dans les études gènes candidats. En revanche, nos résultats suggèrent que les gènes *GSTM1* et *GSTT1* modulent l'association entre le CDT et l'obésité ou la consommation d'alcool. Nous avons confirmé, dans la population européenne, les associations avec les SNP GWAS rs966423 (2q35), rs1867277 et rs965513 (9q22), rs944289 et rs116909374 (14q13). Chez les mélanésiens, les associations pour ces SNP étaient à la limite de la significativité sauf rs966423 qui n'était pas associé au risque de CDT. Dans les loci 9q22 et 14q13, nous avons mis en évidence de nouveaux variants associés au risque de CDT dans les deux populations étudiées ; ces résultats doivent être confirmés dans d'autres études.

Forces et limites de l'étude

Nos résultats sont basés sur les données de deux études cas-témoins en population générale disposant d'un questionnaire détaillé sur les caractéristiques socio-économiques des sujets, les expositions environnementales ainsi que leur mode de vie, nous permettant d'explorer l'existence d'interactions gène-environnement dans les CDT.

Ces données détaillées sont d'autant plus importantes que les deux populations sont contrastées tant au niveau génétique qu'au niveau de l'environnement ou du mode de vie. De plus, la Nouvelle-Calédonie se caractérise par une incidence du cancer de la thyroïde parmi les plus élevées au monde, en particulier dans l'ethnie mélanésienne. L'étude NC avait déjà permis de mettre en évidence plusieurs caractéristiques individuelles associées au risque de cancer de la thyroïde, notamment la forte prévalence de l'obésité et la parité élevée (Truong et al. 2005; Guignard et al. 2007). Nos résultats montrent que l'obésité serait plus fortement associée au risque de DTC chez les sujets *GSTM1*/*GSTT1* nuls. Si ces résultats sont confirmés par d'autres études, la prévalence plus élevée de génotypes nuls des gènes *GSTM1* et *GSTT1*

combinée à une plus forte prévalence de l'obésité chez les mélanésiens pourrait ainsi contribuer à expliquer en partie la différence d'incidence observée entre européens et mélanésiens.

L'une des principales limites de notre étude est le manque de puissance statistique, en particulier dans la population mélanésienne. Pour laquelle nous avions un plus faible nombre de sujets inclus et de plus faibles fréquences alléliques pour les polymorphismes d'intérêt notamment les SNP du locus 9q22 et des SNP GWAS rs966423 (locus 2q35) et rs116909374 (locus 14q13). Nos effectifs ont aussi limité la possibilité d'analyses stratifiées sur certaines classes d'exposition ou sur le type histologique des CDT.

Nous ne pouvons également pas exclure un possible biais de stratification sur les populations dans notre étude du fait de l'utilisation de l'ethnie auto-déclaré pour classer les individus selon les origines mélanésiennes ou européennes. L'utilisation des données génétiques aurait pût permettre d'identifier les clusters plus homogènes de sujets appartenant à chaque groupe ethnique, toutefois nous ne disposions pas d'assez de données pour effectuer de telles analyses.

La possibilité d'un biais de détection ne peut pas être écartée pour le cancer de la thyroïde. Ce biais de classement sur la maladie est rendu possible par la prévalence élevée de cancers micro papillaires existant à l'état latent et l'existence d'une surveillance médicale rapprochée de certains groupes d'individus amenant à la découverte fortuite de micro-cancers. Les conséquences possibles du biais de détection sont l'existence d'une incidence artificiellement élevée de cancers chez les individus possédant les caractéristiques à l'origine d'un surdiagnostic, et la mise en évidence de fausses associations. Un biais de détection ne peut toutefois pas être évoqué pour expliquer les associations avec les facteurs génétiques étudiés.

Comme dans toutes études cas-témoins, le recueil d'informations par questionnaire peut être à l'origine d'un biais de mémoire à l'origine d'une distorsion de l'information entre les cas et les témoins. Ce biais est peu vraisemblable pour les facteurs comme la parité, l'âge aux premières règles ou le statut ménopausique, mais a pu survenir lors du recueil d'informations sur les caractéristiques anthropométriques, la consommation d'alcool ou de tabac. Des biais de classement non différentiels sont également possibles. Plusieurs études

ont ainsi évalué la fiabilité du poids et de la taille déclarée par les individus en les comparants aux valeurs mesurées. Les résultats montrent une tendance des individus à surestimer leur taille et à sous-estimer leur poids, en particulier chez les sujets en surpoids (Engstrom, Paterson et al. 2003). Toutefois, la consistance des informations d'une étude à l'autre ou d'une population à l'autre pour les facteurs comme le poids, la consommation d'alcool ou de tabac permettent de minimiser l'importance de tels biais dans les résultats observés.

Conclusion

Comme pour la plupart des cancers, le modèle le plus plausible et le plus communément admis pour expliquer la composante génétique totale des CDT est un modèle polygénique impliquant un grand nombre de gènes, chacun d'entre eux conférant individuellement un risque faible ou modéré. Globalement, les études gènes candidats ont été décevantes pour identifier des gènes de susceptibilité dans les CDT, alors que les études GWAS ont permis de mettre en évidence quelques polymorphismes de susceptibilité. Toutefois, on estime que l'ensemble des polymorphismes identifiés à ce jour ne permettent d'expliquer qu'environ 4% de l'héritabilité génétique dans les CDT (Figlioli et al. 2015).

Les principales raisons avancées pour expliquer la faible héritabilité expliquée par les études GWAS dans les maladies complexes comme les cancers incluent le manque de puissance statistique pour mettre en évidence des variants rares, ou l'absence de prise en compte des interactions gène-environnement ou gène-gène (Manolio et al. 2009). La stratégie d'analyse couramment utilisée dans les GWAS consistant à tester chaque SNP individuellement pourrait également ne pas convenir pour détecter des associations avec de multiples variants ayant des effets faibles ou les interactions gène-environnement. Des approches complémentaires sont donc nécessaires pour identifier de nouveaux facteurs de risque génétiques des CDT.

Partie 2 : Etude des facteurs de risque génétique dans les cancers différenciés de la thyroïde selon une approche par pathway candidat

Partie 2 Etude des facteurs de risque génétique dans les cancers différenciés de la thyroïde selon une approche par pathway candidat

2.1. Contexte et objectif

L'analyse par pathway (ou gene-set analysis) consiste à considérer globalement les gènes contribuant à une même fonction biologique et non plus individuellement comme dans les approches plus classiques. Une approche possible est de combiner les tests d'association de l'ensemble des SNP des gènes du pathway avec le phénotype d'intérêt afin de proposer un test global (Jin et al. 2014; Ramanan et al. 2012; Evangelou et al. 2012; Wu et al. 2011).

Les avantages de cette approche par rapport à une analyse basée sur des puces GWAS couvrant l'ensemble du génome sont : 1) on dispose d'une meilleure couverture des SNP des gènes étudiés ; 2) le nombre de tests statistiques est réduit permettant ainsi un contrôle moins sévère des faux positifs et l'utilisation de méthodes moins conservatrices ; 3) un gain de puissance en cumulant plusieurs signaux faibles dans un gène ou groupe de gènes ; 4) pour les études pangénomiques, l'information est structurée par les connaissances biologiques et facilite l'interprétation des résultats. Cette approche permet également de limiter l'étude des interactions gène-environnement aux gènes et aux facteurs environnementaux impliqués dans une même fonction biologique.

Dans les CDT, à notre connaissance il existe 5 études ayant utilisé une approche par pathway candidat pour identifier de nouveaux facteurs de risque génétiques. Toutes ont été menées à partir de la même étude cas-témoins américaine incluant 344 cas et 452 témoins (étude USRT/UTMDACC) (Neta et al. 2011; Schonfeld et al. 2012; Aschebrook-Kilfoy et al. 2012; Kitahara, Neta, et al. 2012; Brenner et al. 2013). Les groupes de gènes étudiés sont ceux liés à la réparation de l'ADN et au maintien de l'intégrité du génome (Neta et al. 2011), à la biosynthèse et au métabolisme des hormones sexuelles (Schonfeld et al. 2012), au métabolisme et à la détoxification de substances exogènes et endogènes (Aschebrook-Kilfoy et al. 2012), à l'obésité (Kitahara, Neta, et al. 2012) et à la réponse immunitaire et au processus inflammatoire (Brenner et al. 2013).

Les auteurs ont mis en évidence des associations avec les gènes *ALKBH3, HDAC4, BAK1, FAF1_CDKN2C, DACT3* et *FZD6* qui font partie des gènes de la réparation de l'ADN et du maintien de l'intégrité du génome (Neta et al. 2011), et le gène *SERPINA5* impliqué dans la réponse immunitaire (Brenner et al. 2013). Des associations avec les pathways de « l'apoptose » et de la « réponse immunitaire » ont aussi été observées. L'étude sur les gènes impliqués dans le métabolisme et la détoxification rapporte également des interactions entre les gènes *UGT2B7* et *NAT1* et la consommation d'alcool ainsi qu'entre le gène *CYP26B1* et la consommation de tabac (Aschebrook-Kilfoy et al. 2012). Les auteurs insistent toutefois sur la nécessité de confirmer leurs résultats par des études indépendantes.

Dans cette partie de ma thèse, nous nous sommes intéressés aux gènes impliqués dans :

1) Le métabolisme des xénobiotiques : les xénobiotiques sont des substances étrangères à l'organisme tel que l'alcool, le tabac, les médicaments et les polluants de l'environnement. Ces substances ou leurs métabolites sont capables de causer des dommages cellulaires si elles ne sont pas éliminées. L'alcool et le tabac par exemple ont été suspectés dans les CDT selon des mécanismes biologique impliquant une action directe de leurs métabolites sur les cellules thyroïdiennes ou indirect à travers un dérèglement de l'axe hypothalamo-pituitaire-thyroïde (Cho & Kim 2014; Hong et al. 2016). Le métabolisme et l'élimination de ces composéssont assurés par une cascade de réactions enzymatiques dont les principaux acteurs sont les cytochromes p450, les glutathions S-transférases et les UDP-glucuronosyltransférase (UGT). Il a été suggéré que les variations génétiques dans les gènes codant pour ces enzymes pourraient provoquer un déséquilibre métabolique responsable de la susceptibilité individuelle aux carcinogènes (Adjadj et al. 2009; Hirvonen 1999).

2) La biosynthèse et le métabolisme des hormones stéroïdiennes : Ces hormones, les œstrogènes en particulier, sont suspectés de jouer un rôle dans l'étiologie des CDT du fait de l'incidence plus élevée de ces cancers chez les femmes. Des résultats d'études expérimentales montrent également que les œstrogènes stimulent la prolifération des cellules thyroïdiennes (Derwahl & Nicula 2014; Xu et al. 2013). Les polymorphismes dans les gènes codant pour les enzymes impliqués dans la synthèse

et le métabolisme de ces hormones, peuvent influencer la réponse individuelle aux expositions hormonales et jouer un rôle central dans le développement tumoral en modifiant le risque de CDT lié à ces expositions.

Les objectifs de ce travail étaient : (1) d'explorer le rôle des gènes du « métabolisme des xénobiotiques » et des gènes de la « biosynthèse et du métabolisme des hormones stéroïdiennes » dans les CDT en utilisant une approche par pathway ; (2) de tester les interactions entre ces gènes et les facteurs de risque tels que l'obésité, la consommation d'alcool, la consommation de tabac et les facteurs hormonaux et reproductifs (âge aux premières règles, parité et prise de contraceptifs oraux).

L'ensemble des analyses seront conduites chez des femmes d'origine européenne. Dans un premier temps, nous explorerons le rôle des pathways d'intérêt dans les études CATHY/Young-Thyr conduites en France et génotypées par la puce Beluhca. Dans un second temps, nous effectuerons une analyse combiné des données des études françaises (CATHY/Young-Thyr) et des données d'études menées aux Etats-Unis (USRT/UTMDACC) afin d'avoir une meilleure puissance statistique. Le génotypage des données françaises et américaines ayant été effectué à l'aide de deux puces différentes, des imputations sur les données génétiques ont été nécessaires pour pooler les études. Cette analyse poolée a été faite en collaboration avec Alina Brenner et Ruth Pfeiffer du National Cancer Institute (NCI).

2.2. Matériel et méthodes

2.2.1. Populations d'étude

Les analyses ont porté sur 810 femmes (407 cas et 403 témoins) de l'étude CATHY (décrite dans la partie 1), ainsi que sur 200 femmes de l'étude Young-Thyr (100 cas et 100 témoins). Nous avons ensuite inclus pour l'analyse poolée, les femmes de l'étude USRT/UTMDACC soit 332 cas et 443 témoins. Nous décrivons ici les données de l'étude Young-Thyr et de l'étude USRT/UTMDACC.

2.2.1.a. Etude Young-Thyr

L'étude Young-Thyr est une étude cas-témoins menée dans l'Est de la France entre juillet 2005 et octobre 2010 dans les régions ayant potentiellement été exposées aux retombées nucléaires après l'accident de Tchernobyl en 1986.

Sélection des cas

Les cas étaient les patients diagnostiqués d'un cancer différencié de la thyroïde entre le 1er janvier 2002 et le 31 décembre 2006, âgés de moins de 35 ans au moment du diagnostic et ayant leur résidence principale dans l'une des régions de l'est de la France (Alsace, Champagne-Ardenne, Corse, Franche-Comté, Lorraine, Rhône-Alpes ou Provence-Alpes-Côte d'Azur). Les cas ont été identifiés par trois sources principales selon la région : 1) registre général du cancer en Champagne-Ardenne, Alsace et Rhône-Alpes ; 2) le registre national du cancer de la petite enfance (qui contient des informations sur les enfants de <15 ans) dans toutes les régions; et 3) les hôpitaux privés et publics en Lorraine, Franche-Comté, Corse et Provence-Alpes-Côte d'Azur. Dans ces 4 dernières régions, qui ne sont pas couvertes par un registre général de cancer, l'identification des cas de cancer a été réalisée en collaboration avec les structures locales de santé publique. Sur 1049 cas éligibles, 223 (21%) n'ont pas pu être interrogés et 21 n'ont pas été appariés. L'étude a porté sur 805 cas.

Sélection des témoins

Les témoins ont été sélectionnés parmi la population générale par tirage au sort à partir des annuaires téléphoniques et appariés individuellement aux cas sur le sexe, la date de naissance (plus ou moins 1 an) et la région de résidence. Sur 1045 cas éligibles, 164 (16%)

n'ont pas pu être interrogés et 5 n'ont pas été inclus car ils n'avaient pas de témoin apparié. L'étude a porté sur 876 témoins, certains cas étant appariés à 2 témoins.

Recueil des données

Le recueil des données avait été fait lors des entretiens en face-à-face à domicile par un enquêteur spécialement formé pour cela et à l'aide d'un questionnaire standardisé envoyé aux sujets au moins une semaine avant l'entretien afin qu'ils puissent se munir de documents tels que des résultats de radiologie ou encore demander des informations relatives à leur jeunesse à leurs proches. Les données recueillies concernaient les informations sociodémographiques, anthropométriques (poids et taille), antécédents personnels et familiaux de pathologie thyroïdienne, les lieux de résidence, l'éducation et de la profession, les facteurs gynécologiques et reproductifs, l'exposition médicale aux rayons X, et le régime alimentaire, tant au moment de l'interview que pendant l'enfance. Une grande partie du questionnaire est commun à celui utilisé dans l'étude CATHY.

Un prélèvement salivaire (Kit Oragène) a été effectué chez 727 cas et 712 témoins pour constituer une banque d'ADN. 200 femmes (100 cas et 100 témoins) de cette étude ont été sélectionnées et génotypées avec les femmes de l'étude CATHY dans le cadre du projet Beluhca que je présente dans la suite.

2.2.1.b. Etude cas-témoins USRT/UTMDACC

L'étude USRT/UTMDACC est une étude cas-témoins sur les CPT menée aux Etats Unis ; elle est constituée d'une étude cas-témoin nichée dans la cohorte USRT et d'une cohorte de cas incidents de CPT diagnostiqués par le Centre de Cancérologie MD Anderson de l'Université du TEXAS (UTMDACC).

La cohorte USRT est une cohorte nationale de techniciens radiologistes initiée en 1980 aux Etats-Unis. L'objectif initial était d'étudier les conséquences de l'exposition aux radiations ionisantes chez des techniciens radiologistes sur le risque de survenue des cancers et autres pathologies. La cohorte initiale incluait 146 022 participants certifiés par le registre national des techniciens radiologistes comme occupant ce poste depuis au moins 2 ans entre 1926 et 1982. Le suivi des participants avait démarré en 1983 et un suivi actif de la cohorte est effectué au travers d'un renouvellement annuel du registre. Par ailleurs, un questionnaire est envoyé

par courrier aux participants tous les 7 ans environ pour le recueil et l'actualisation des informations. Entre 1983 et 2014, 4 questionnaires avaient été envoyés aux participants : le premier entre 1983 et 1989, le second entre 1994 et 1998, le troisième entre 2003 et 2005 et le dernier en 2012. Une étude cas-témoins nichée ayant pour objectif d'évaluer les facteurs de risque génétique dans les CPT a été initiée en 2005 et actualisée en 2008 sur la base des informations recueillies au travers des 3 questionnaires administrés entre 1983 et 2005 aux sujets de la cohorte USRT (Lönn et al. 2007; Neta et al. 2011).

Sélection des cas

Les cas de cancers de la thyroïde reportés dans l'un des 3 questionnaires entre 1983 et 2005 étaient confirmés par consultation du dossier médical du patient, des résultats anatomopathologiques, des registres des cancers ou autres sources d'informations médicales et recrutés pour un prélèvement sanguin. Sur les 415 cas de cancers de la thyroïde vivants en 2005 et éligibles pour recrutement, 183 cas n'ont pas été inclus dans l'étude pour plusieurs raisons : ils avaient refusé l'accès à leur dossier médical (66 cas); les résultats anatomopathologiques n'avaient pas pût être obtenus (25 cas) ; ils avaient refusé de donner un prélèvement sanguin (54 cas) ; ils n'avaient pas répondu à l'invitation pour la collecte d'ADN (19 cas) ou avaient répondu à l'invitation après la date limite du recrutement (19 cas) (Neta et al. 2011). 17 cas étaient des cancers folliculaires de la thyroïde et ont été également exclus. Au total, 215 cas pour lesquels un prélèvement sanguin a été obtenu ont été inclus dans l'étude.

Les cas de l'étude avait ensuite été complétés par une cohorte de 243 cas incident de CPT diagnostiqués entre 1999 et 2005 par le Centre de Cancérologie MD Anderson de l'Université du TEXAS (UTMDACC). Ces cas étaient des patients admis à l'hôpital pour une résection de leur tumeur (Ho et al. 2009).

Au final, l'étude cas-témoins USRT/UTMDACC inclut 458 cas dont 332 cas femmes.

Sélection des témoins

Les témoins ont été sélectionnés parmi un pool de témoins initialement recrutés au sein de la cohorte USRT pour des études cas-témoins portant sur les cancers du sein, de la prostate

et des testicules. Ces témoins étaient des sujets chez qui aucun cancer n'avait été diagnostiqué au moment de leur inclusion dans l'étude. Ils avaient été appariés aux cas sur le sexe et l'année de naissance (par intervalle de 5 ans) ; 3 témoins étaient sélectionnés pour un cas dans la cohorte USRT. Au total, 491 témoins pour lesquels un prélèvement sanguin était disponible ont été inclus dans l'étude (Lönn et al. 2007; Neta et al. 2011), dont 443 femmes. Il est a noté qu'aucun témoin n'avait été apparié individuellement aux cas issus du Centre de Cancérologie MD Anderson (UTMDACC).

Recueil des données

Dans la cohorte USRT, les informations recueillis par les différents questionnaires concernent les antécédents médicaux (incluant l'auto-déclaration d'un CDT), l'historique professionnel, les antécédents familiaux pour différentes pathologies notamment les pathologies thyroïdienne, les caractéristiques anthropométriques, les facteurs gynécologiques et reproductifs et divers facteurs environnementaux incluant la consommation d'alcool et de tabac (Sigurdson et al. 2003).

Concernant les cas incidents de CPT de UTMDACC, un questionnaire auto administré à ces patients avait permis de recueillir les informations concernant leurs caractéristiques sociodémographiques, les antécédents médicaux, l'exposition médicale aux radiations ionisantes, les facteurs hormonaux et reproductifs, les facteurs anthropométriques ainsi que la consommation d'alcool et de tabac (Ho et al. 2009).

2.2.1.c. Standardisation des données épidémiologiques

Pour l'étude des interactions gène-environnement, nous nous sommes intéressés comme dans la partie précédente, aux facteurs de risque suspectés dans les CDT qui sont : l'indice de masse corporelle (IMC), la consommation d'alcool et de tabac et certain facteurs hormonaux et reproductifs tels que l'âge aux premières règles, le nombre de grossesses à terme et la prise de contraceptifs oraux. Nous avons standardisé ces variables entre les études CATHY et Young-Thyr qui utilisaient un questionnaire quasiment identique.

L'analyse combinée avec les données de l'étude USRT/UTMDACC incluait uniquement les variables IMC, statut tabagique et consommation d'alcool (oui/non).

2.2.2. Données génétiques

2.2.2.a. Données génétiques des études CATHY/Young-Thyr – Puce Beluhca

Le projet Beluhca développé par l'équipe avait pour objectif d'explorer les interactions gène-environnement dans trois cancers chez la femme (sein, voies aérodigestives supérieures et poumon). Une puce génétique, nommée Beluhca, a ainsi été conçue afin d'inclure des gènes de pathways d'intérêt pour ces trois cancers, afin d'étudier les interactions entre les facteurs génétiques et environnementaux impliqués dans un même pathway. Les pathways sélectionnés portent ainsi sur des fonctions biologiques pertinentes par rapport à l'étiologie des cancers étudiés : métabolisme des hormones stéroïdiennes, rythme circadien, gènes de l'obésité, gènes impliqués dans la dépendance à la nicotine, etc. Cette puce n'était initialement pas prévue pour inclure les cancers de la thyroïde. Cependant, étant données les associations entre le risque de cancer thyroïdien et certains facteurs hormonaux, reproductifs et anthropométriques, les cancers de la thyroïde se prêtaient également aux objectifs initiaux de la puce Belhuca et ont été inclus à postériori dans le projet Belhuca.

Ainsi, 22 pathways candidat ont été sélectionnés en janvier 2011 à partir des données de KEGG (Kyoto Encyclopedia of Genes and Genomes), de Gene Ontology et d'une revue de la littérature. Pour chacun des 652 gènes appartenant à ces pathways, des SNP candidats et des tags SNP autour du gène sur un intervalle de +/- 5kb ont été sélectionnés de telles sorte qu'ils soient corrélés (r^2>0.8) à chacun des SNP de fréquence allélique>5% de cet intervalle (dans la population CEU de HapMap, 21 release Phase II, NCBI build 36.1, assembly dbSNPb126). La sélection des SNP a été effectuée à partir du logiciel Tagger Software. Un total de 8716 SNP ont ainsi été sélectionnés, dont 7729 ont pu être inclus sur la puce.
Le design de la puce a été réalisé par la société Illumina et le génotypage, effectué par la société Integragen.

i. Génotypage et contrôle qualité

La puce Beluhca a été utilisée pour le génotypage de 810 femmes de l'étude CATHY et 200 femmes de l'étude Young-Thyr. Après génotypage un contrôle qualité des SNP et des sujets a été effectué.

Contrôle qualité des SNP

Parmi les 7729 SNP ayant pu être génotypés : 728 SNP ont été exclus car leurs taux de génotypage était inférieurs à 95%, 143 SNP avaient une fréquence allélique <0.01, 1 SNP a été exclu par le test d'HW ($p<7x10^{-06}$ après correction de Bonferroni pour les tests) ; 16 SNP avaient des génotypes observés chez les 10 sujets CEPH inclus (sujets du projet HapMap) discordants avec leurs génotypes publiés. Au final, 6861 SNP dans 648 gènes ont été retenus.

Contrôle qualité des sujets

Parmi les 810 sujets de l'étude CATHY et les 200 sujets de l'étude Young-Thyr dont les prélèvements salivaires ont été envoyés pour génotypage, nous avons exclu des analyses 32 sujets pour les raisons suivantes : 29 sujets (23 de l'étude CATHY et 6 de l'étude Young-Thyr) n'avaient pas pu être génotypés du fait d'une faible concentration d'ADN ; 3 sujets (2 de l'étude CATHY et 1 de l'étude de Young-Thyr) ont été identifiés comme ayant un génotype masculin et enfin, 1 sujet de l'étude Young-Thyr a retiré son consentement de participation à l'étude.

Nous disposons donc de 977 sujets génotypés et validés (785 dans l'étude CATHY et 192 dans l'étude Young-Thyr).

Stratification de population - Analyse en composante principale

La différence de fréquence des allèles entre les groupes et sous-groupes ethniques constituant un échantillon d'analyse peut entraîner des résultats faussement positifs ou peut réduire la puissance dans les études génétiques. Dans les études génomiques cas-témoins en population, il est supposé que les cas et les témoins participants sont échantillonnés à partir de la même population, de sorte que les différences dans la fréquence des allèles soient liées uniquement au phénotype d'intérêt, plutôt qu'à des différences dans l'origine de population entre les cas et les témoins. Toutefois, si cette hypothèse n'est pas remplie, de fausses associations peuvent être observées (Tian et al. 2008). Il est judicieux d'appliquer les informations sur la structure de la population d'étude pour sélectionner de sous-groupes de sujets ayant une origine ancestrale très similaire ou presque homogène. Une telle approche permet de limiter la diversité ancestrale dans l'échantillon d'analyse et se rapprocher ainsi à l'étude d'une maladie dans un seul groupe ethnique.

Pour respecter ce cadre, nous avons réalisé une analyse en composantes principales (ACP), à l'aide du package FactoMineR, en me basant sur les données de 1342 SNP indépendants (approche pairwise avec $r^2<0.1$), sélectionnés à partir des 6861 SNP. La méthode ACP identifie les composantes principales qui représentent la structure de la population sur la base de corrélations génétiques entre les individus. 32 sujets « outliers » (19 sujets de l'étude CATHY et 13 sujets de l'étude YOUNGTHYR) ont été identifiés et exclus de l'analyse (Figure 12). Le seuil d'exclusion est défini par le cercle dont les coordonnées du centre sont les moyennes des deux axes (composantes principales 1 et 2) et le rayon du cercle est le vecteur géométrique construit à partir de 1.96 fois l'écart-type des composantes principales 1 et 2. Au final, 945 femmes dont 463 cas et 482 témoins ont été inclus dans les analyses.

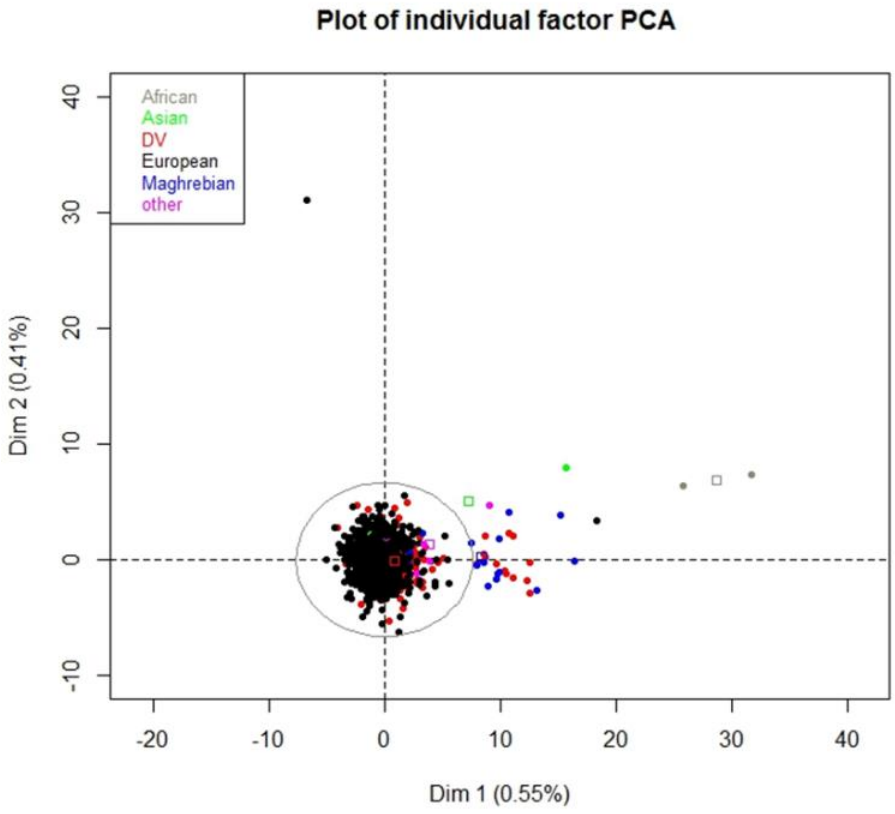


Figure 12: Projection des individus des études CATHY/YOUNG-THYRsur les 2 premières composantes principales. Les sujets en dehors du cercle sont exclus de l'analyse (outliers). Sur l'axe des abscisses est représentée la première dimension et l'axe des ordonnées, la deuxième dimension des composantes principales. Le rayon du cercle est le vecteur géométrique construit à partir de 1.96 fois l'écart-type des composantes principales 1 et 2. Les points représentent les individus et les couleurs, les origines ethniques déclarées des individus de l'étude CATHY. African : individus d'origine africaine, Asian : individus d'origine asiatique, European : individus d'origine européenne, other : autres origines. DV : individus de l'étude YOUNGTHYR pour lesquels nous n'avions pas l'information sur l'origine ethnique.

ii. Description des pathways d'intérêt dans la puce Beluhca

Dans le cadre de ces analyses, nous nous sommes intéressés à deux pathways de la puce Beluhca qui sont : le pathway du « métabolisme des xénobiotiques » et le pathway de la « biosynthèse et métabolisme des hormones stéroïdiennes »

Pathway du « métabolisme des xénobiotiques »

Les xénobiotiques sont métabolisés et éliminés de l'organisme par une cascade de réactions enzymatiques regroupés en 2 phases (annexe 2 : figure 5).

- La phase 1 qui comprend des réactions de fonctionnalisation catalysées majoritairement par les mono-oxygénases à cytochromes P450 (CYP). Dans cette phase, on y trouve aussi les déshydrogénases comme les ADH (alcool déshydrogénase) et les réductases comme les NADH responsable du métabolisme de l'alcool (Hirvonen, 1999).
- La phase 2 qui consiste essentiellement en des réactions de conjugaison où les métabolites issus de phase 1 sont hydrolysés et éliminés. Cette phase est assurée par des enzymes tels que les glutathion S-transférases (GST), les UDP-glucuronosyltransférases (UGT), les N-acétyltransférase (NAT) et les sulfotransférases (SULT) (Hirvonen 1999)

Dans la puce Belhuca, ce pathway regroupe 82 gènes dont 71 gènes ont été recensés par la base KEGG (hsa00980 : Metabolism of xenobiotics by cytochrome P450) et 11 gènes (*COMT, NAT1, NAT2, NQO1, SOD2, SULT1A1, SULT1A2, AHR, AKR1A1, CYP2A6, CYPD6*) ont été identifiés dans la littérature (Gemignani et al. 2007). Après contrôle qualité, ce pathway inclut au total 77 gènes et 574 SNP (Tableau 13).

Pathway de la « biosynthèse et du métabolisme des hormones stéroïdiennes »

Les hormones stéroïdiennes sont des hormones de croissance secrétées par les glandes endocrines. On distingue 3 grands groupes : les glucocorticoïdes, les minéralocorticoïdes, les stéroïdes sexuels (androgènes, œstrogènes et progestatifs). Nous nous sommes intéressés aux gènes principalement impliqués dans la synthèse et métabolismes des stéroïdes sexuels

du fait du rôle suspectés de ces hormones en particulier les œstrogènes dans l'étiologie des CDT.

Les œstrogènes sont un produit métabolique dont le niveau circulant est déterminé par la synthèse de novo, la conversion à partir d'autres hormones stéroïdiennes et les mécanismes d'élimination des œstrogènes (Low et al. 2010). Ces processus métaboliques sont régulés par un réseau d'enzymes codées par différents gènes tel que les cytochromes P450 (CYP) responsables de la conversion du cholestérol en androgènes (Low et al. 2010); les sulfate déshydrogénases (SULT) et les hydro stéroïdes déshydrogénases (HSD) qui assurent la conversion des androgènes en œstrogènes (Low et al. 2010).

Les œstrogènes sont ensuite métabolisés et éliminés de l'organisme par les enzymes de phase 1 et 2 du métabolisme qui sont d'ailleurs, les mêmes enzymes responsables du métabolisme des xénobiotiques expliquant les quelques gènes en commun entre les deux pathway qui nous intéresse (« du métabolisme des xénobiotiques » et de la « biosynthèse et métabolisme des hormones stéroïdiennes »). La figure 6 de l'annexe 2, résume les phases de biosynthèse et métabolisme des hormones stéroïdiennes et certains gènes impliqués dans ces phases.

Le pathway de la « biosynthèse et métabolisme des hormones stéroïdiennes » de la puce Belhuca regroupe des gènes qui ont été recensés par : la base de données KEGG (hsa00140 : steroid hormone biosynthesis ; 55 gènes), la base de données Gene Ontology (12 GO sélectionnés ; 65 gènes), et 37 gènes identifiés dans la littérature (He et al. 2010). Ce pathway est constitué de 96 gènes et 581 SNP après contrôle qualité (Tableau 13).

Les 2 pathways incluent au total après contrôle qualité 138 gènes et 1280 SNP ; ils ont en commun 36 gènes (277 SNP) qui sont principalement les gènes de la famille des : cytochromes p450 (CYP) ; UDP-glucuronosyltransférases (UGT) ; aldo-ketone réductases (AKR) et sulfotransférases (SULT) (Tableau 13).

Il est a noté que les gènes *UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8 et UGT1A9* appartenant à la famille des UDP-glucuronosyltransférases (UGT) sont tous situé sur le locus 2q37 et sont tagué par les mêmes SNP. Pour éviter la surestimation de leurs effets au niveau des pathways, nous les avons regroupé en une seule

région appelée locus UGT1A qui regroupe tous les tags-SNP de ces gènes. De même, les gènes UGT2A1 et UGT2A2 situé sur le locus 4q13 et tagués par les même SNP ont été regroupés en une seule région appelé UGT2A pour les analyse au niveau des pathways.

Tableau 13 : tableau des gènes inclut dans les pathways d'intérêt dans notre étude.

Pathways	Gènes inclus	Nombre de gènes	Nombre de SNP
Métabolisme des xénobiotiques	*ADH1A, ADH1B, ADH1C ,ADH4, ADH5, ADH6, ADH7, AHR, AKR1A1,* ***AKR1C1, AKR1C2, AKR1C3, AKR1C4****, ALDH1A3, ALDH3A1, ALDH3B1, ALDH3B2,* ***COMT, CYP1A1****,* ***CYP1A2, CYP1B1, CYP2A6, CYP2B6,*** *CYP2C18,* ***CYP2C19****, CYP2C8,* ***CYP2C9****,* ***CYP2D6, CYP2E1****, CYP2F1, CYP2S1,* ***CYP3A4, CYP3A43, CYP3A5, CYP3A7****, DHDH, EPHX1, GSTA1, GSTA2, GSTA3, GSTA4, GSTA5, GSTK1, GSTM2, GSTM3, GSTM4, GSTM5, GSTO1, GSTO2, GSTP1, GSTT2, GSTZ1, MGST1, MGST2, MGST3, NAT1, NAT2, NQO1, SOD2,* ***SULT1A1, SULT1A2, UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, UGT2A1, UGT2A2, UGT2A3, UGT2B10, UGT2B11, UGT2B4, UGT2B7***	77	581
Métabolisme des hormones stéroïdiennes	*AKR1B1, AKR1B10,* ***AKR1C1****,* ***AKR1C2****,* ***AKR1C3****,* ***AKR1C4****, AKR1D1, APOA2, APOA4, APOB, APOC3, AR, ASMT, CGA,* ***COMT****, CYP11A1, CYP11B1, CYP11B2, CYP17A1, CYP19A1,* ***CYP1A1****,* ***CYP1A2****,* ***CYP1B1****, CYP24A1, CYP27A1, CYP27B1,* ***CYP2A6, CYP2B6, CYP2C19, CYP2C9, CYP2D6, CYP2E1,*** *CYP2R1, CYP39A1, CYP3A4, CYP3A43,* ***CYP3A5****,* ***CYP3A7****, CYP46A1, CYP7A1, CYP7B1, ESR1, ESR2, FSHB, FSHR, GNRH1, GNRHR, HSD11B1, HSD17B1, HSD17B12, HSD17B13, HSD17B2, HSD17B3, HSD17B4, HSD17B6, HSD17B7, HSD17B8, HSD3B2, IL4, LDLR, LHB, LHCGR, MC2R, NR3C1, PGR, POMC, PPARD, SHBG, SRD5A1, SRD5A2, SRD5A3, SREBF2, STS,* ***SULT1A1, SULT1A2,*** *SULT1B1,* *SULT1E1, SULT2A1,* *SULT2B1, SULT4A1,* ***UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, UGT2A1, UGT2A2, UGT2A3, UGT2B10, UGT2B11, UGT2B4, UGT2B7****,*	96	974

Les gènes en gras sont ceux en communs entre les deux pathways.

2.2.2.b. Données génétiques étude USRT/UTMDACC

Dans le cadre de l'étude des facteurs de risque génétique dans différents cancers notamment les cancers du sein, de la prostate, des testicules et de la thyroïde, une puce génétique a été conçue par le NCI. Cette puce constituée de 27 904 SNP inclut :

- Des SNP dans 1316 gènes candidats des pathways dont les fonctions sont pertinentes dans l'étiologie des cancers étudiés. Ces pathways ont été définit à partir des informations biologiques et/ou fonctionnelles des gènes disponible dans diffèrent

bases de données notamment les bases de données « NCBI Entrez Gene » (www.ncbi.nlm.nih.gov/sites/entrez) et « Gene Cards » (www.genecards.org) et d'une revue de littérature (Loza et al. 2007). Parmi les pathways d'intérêt on peut citer ceux : de la réparation de l'ADN et du maintien de l'intégrité du génome ; du métabolisme et la détoxification des xénobiotiques ; de l'obésité ; de la réponse immunitaire et l'inflammation et de la biosynthèse, métabolisme et régulation des hormones sexuelles féminines.

- Des SNP couvrant les régions intergéniques précédemment associés au risque de plusieurs cancers

Les tags SNP des gènes candidats inclus dans la puce ont été sélectionnés dans les gènes candidats et autour, sur une distance de 20 kb en amont du début de la transcription du gène et de 10 kb au-delà du dernier exon du gène. Les tags-SNP étaient sélectionnés de manière à capturer tous les SNP de la région ayant une MAF>0.05 dans la population CEU de HapMap avec une corrélation d'au moins 0.8 (Neta et al. 2011). La puce avait été utilisée pour génotyper les sujets de l'étude USRT/UTMDACC par la plateforme de génotypage du NCI (National Cancer institut). Le génotypage des sujets a été effectué à partir du prélèvement sanguin obtenu chez les cas et témoins de l'étude.

Contrôle qualité

Après génotypage, un contrôle qualité des sujets et des SNP avait été effectué par le NCI. Parmi les 27 904 SNP initialement inclus dans la puce, 722 SNP n'ont pu être inclus dans le design de la puce (absence de sonde). Après génotypage 2344 SNP ont été exclus pour plusieurs raisons : 208 SNP étaient mono-morphique, 656 SNP avaient un taux de concordance avec les sujets de HapMap <95% ; 740 SNP avaient un taux de génotypage<90% et 740 SNP ne respectaient pas l'équilibre d'HW. De plus 1607 SNP intergéniques initialement inclus dans la puce pour réplication dans d'autre cancers et 5706 tags SNP qui avaient une fréquence de l'allèle mineur <0,10 ont aussi été exclus. Au final, l'étude incluait 17 525 tags SNP dans 1129 gènes ou régions candidates.

Sur les 458 cas (332 cas femmes) et 495 témoins (443 femmes) génotypés, 314 cas (dont 274 femmes) et 463 témoins (dont 423 femmes) ont passé le contrôle qualité.

Dans le cadre de l'analyse poolée, nous avons exclu les hommes et les sujets non européens identifiés par l'ACP (*Figure 13*). Notre population finale d'étude était constituée de 271 cas et 421 témoins.

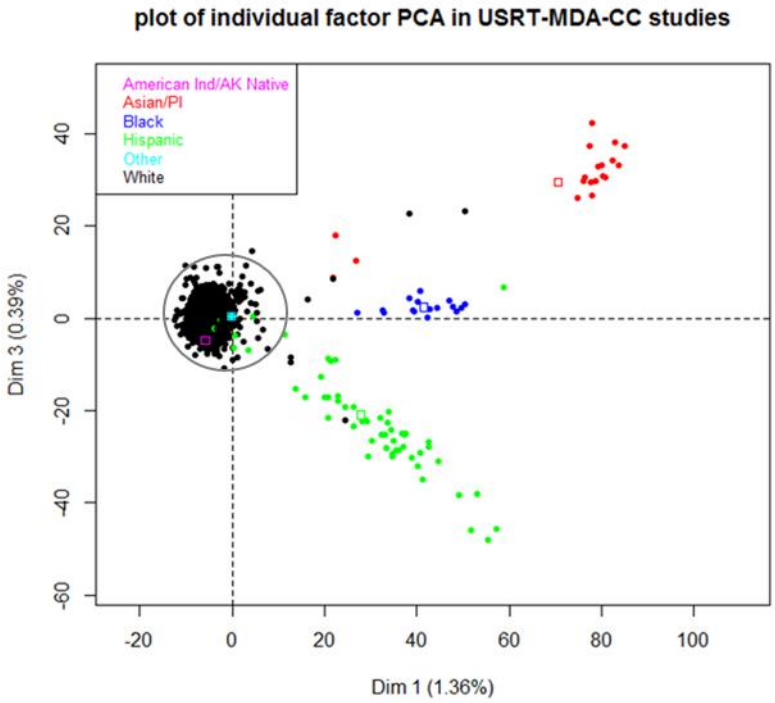


Figure 13 : Projection des individus des études USRT/UTMDACC sur les 2 premières composantes principales. Les sujets en dehors du cercle représentent les points en dehors du cercle. Sur l'axe des abscisses est représentée la première dimension et l'axe des ordonnées, la deuxième dimension des composantes principales. Les points représentent les individus et les couleurs, les origines ethniques déclarées par les sujets de l'étude USRT/UTMDACC. American Ind/AKT : origine Amérindienne ; Asian/PI : origine asiatique ; Black : afro-Américain ; Hispanic : origine hispanique ; White : origine européenne ; Other : autres origines

Imputation des données génétiques pour l'analyse poolée

Lorsque l'on combine les données des études CATHY/Young-Thyr et USRT/UTMDACC, seuls 460 SNP (de 118 gènes) sur les 1280 SNP (de 138 gènes des pathways d'intérêt étudiés dans CATHY/Young-Thyr) étaient génotypés dans l'étude USRT/UTMDACC. Le *Tableau 14* donne la répartition de ces SNP sur les 2 pathways candidats auxquels on s'intéresse.

Tableau 14 : Tableau résumant le nombre de SNP et gènes des pathways d'intérêt dans l'étude CATHY/Young-Thyr et le pool de données des études CATHY/ Young-Thyr et USRT/UTMDACC et nombre de SNP manquants dans l'étude USRT/UTMDACC.

Pathways	Données étude CATHY/YOUNGTHYR		En commun dans le pool de données		SNP manquants dans USRT/UTMDACC
	N gènes	N SNP	N gènes	N SNP	N SNP
Ensemble des pathways	138	1280	118	460	820
Métabolisme des xénobiotiques	77	581	73	232	343
Biosynthèse et métabolisme des hormones stéroïdiennes	96	974	70	333	641

Afin de pouvoir utiliser les données de l'étude USRT/UTMDACC pour l'analyse des 2 pathways de Beluhca, nous avons imputé les SNP génotypés dans l'étude CATHY/Young-Thyr et manquants dans l'étude USRT/UTMDACC Tableau 14). Pour cela, nous avons utilisé le logiciel IMPUTE2 en utilisant comme référence l'ensemble des populations disponibles dans le projet 1000 génomes. Nous avons effectué l'imputation séparément pour chaque chromosome.

Dans un premier temps, nous avons identifié pour l'ensemble des pathways d'intérêt, le nombre de SNP génotypés dans chaque chromosome dans l'étude CATHY/Young-Thyr. Nous avons ensuite créé, pour chaque chromosome, la liste de SNP manquant dans l'étude USRT/UTMDACC mais présents dans l'étude CATHY/Young-Thyr. Pour améliorer la qualité des données imputées et réduire le temps d'imputation par le logiciel, nous avons également stratifié l'imputation sur des fragments de 6MB au maximum selon les recommandations du logiciel IMPUTE2. Par exemple, l'imputation des SNP du chromosome 1 a été stratifiée sur 8 fragments d'ADN correspondant aux régions d'ADN où sont localisées les SNP à imputer. Pour chaque région de chromosome, nous avons spécifié la liste de SNP qui était à imputer afin de ne capturer que les SNP d'intérêt pour mes analyses.

Après imputation, nous avons effectué un contrôle qualité qui consistait à supprimer les SNP ayant un paramètre INFO<0.6 ou qui ne respectaient pas l'équilibre d'HW (p-value du test d'HW< 10^{-03}).

L'imputation a porté sur 820 SNP parmi lesquels nous avons imputé 775 SNP. Après imputation, nous avons exclu 204 SNP car la qualité de l'imputation était insuffisante (paramètre (INFO < 0.6 pour 201 SNP) ou l'équilibre d'HW n'était pas respecté (3 SNP). Au total 571 SNP imputés ont été inclus dans les analyses.

Au final, sur les 2 pathways d'intérêt définis sur la puce Beluhca, je disposais de 1031 SNP (460 SNP génotypés et 571 SNP imputés) dans 118 gènes pour l'analyse poolée.

2.2.3. Analyse statistique

2.2.3.a. Analyse par SNP

Les OR et intervalles de confiances de l'association entre le risque de CDT et les SNP dans les gènes de pathways d'intérêt ont été estimés par régression logistique non-conditionnelle avec ajustement sur l'âge, le sexe, les 2 premières composantes principales, la région de résidence (pour l'analyse des données CATHY/Young-Thyr) et l'étude (pour l'analyse poolée des données CATHY/Young-Thyr et USRT/UTMDACC). Le nombre de composante principale à inclure dans le modèle a été défini selon le critère du coude qui consiste à sélectionner les axes dont les valeurs propres sont situées avant le décrochement (coude) observé lorsqu'on représente les valeurs propres de l'ensemble des composantes sur un graphe (annexe 2 ; figure 7). Pour chaque SNP, nous avons utilisé le modèle log-additif où les SNP sont codés entre 0 et 2 selon le nombre de copies de l'allèle rare. Pour les tests d'interaction entre les SNP et les facteurs de risque suspectés, nous avons utilisé la méthode du rapport du maximum de vraisemblance.

2.2.3.b. Analyse par gène et par pathway

Plusieurs méthodes existent pour l'analyse des associations au niveau des gènes ou des pathways (définis comme un groupe de gènes) (Mooney & Wilmot 2015; Wang et al. 2007). On distingue les méthodes dites « self-contained » des méthodes dites « compétitives » qui se différencient sur l'hypothèse nulle testée. Les méthodes self-contained testent l'hypothèse nulle que les gènes du pathway ne sont pas associés au phénotype d'intérêt alors que les méthodes compétitives testent l'hypothèse nulle que les gènes du pathway d'intérêt sont aussi associés au phénotype que les gènes en dehors du pathway. Ces dernières méthodes ne sont utilisées que dans un cadre pangénomique alors que les méthodes self-contained sont utilisées à la fois dans une approche pathway candidat et pangénomique. Des études ont comparé les performances de quelques méthodes mais il n'y a pas de méthode de référence identifiée à ce jour (Mooney & Wilmot 2015; Evangelou et al. 2012; Mitchell 2015).

Pour nos analyses, , nous avons choisi la méthode ARTP « Adaptive Rank Truncated Product » (Yu et al. 2009) qui est une méthode « self-contained ».

Il a été montré que, la méthode ARTP est plus puissante que les méthodes suivantes «Tail Strength Measure», «Fisher's combination of p-values», «Hypergeometric Test» et «Gene-Set Enrichment Analysis (GSEA)» (Evangelou et al., 2012). Cette méthode a également l'avantage d'être facile à adapter pour étudier les interactions gène-environnement (package PIGE).

Méthode ARTP

Une approche pour définir l'association entre un phénotype et un gène est de combiner les p-values d'association entre le phénotype et chaque SNP appartenant à ce gène. Ainsi, la méthode de Fisher calcule le produit des p-values de tous les SNP du gène afin d'obtenir une statistique suivant une loi du Chi2 sous les hypothèses de non association de chacun des SNP (loi uniforme sous H0) et d'indépendance des SNP (Fisher, 1932). La méthode « truncated product » (Zaykin et al. 2002) se base sur le produit de toutes les p-values inférieures à un certain seuil et la méthode « rank truncated product » ou RTP (Dubridge and Koleman, 2003) fait le produits des K plus petites p-values d'association de SNP pour définir une statistique au niveau du gène. La méthode ARTP est une adaptation de la méthode RTP permettant d'optimiser la valeur de K parmi un ensemble de valeurs candidates. La p-value d'association finale au niveau du gène étant la p-value minimum de la statistique observée pour ces différentes valeurs de K (Yu et al. 2009). Une démarche similaire permet de déterminer la p-value d'un pathway à partir des p-value des gènes.

Principe de la méthode ARTP

Supposons que l'on ait L SNP dans un gène pour lesquels nous souhaitons tester l'association à un phénotype. On conduit ainsi L tests sous l'hypothèse nulle H0 et on définit la p-value de chaque test $\{p_1;\ldots;p_L\}$. On définit les statistiques d'ordre de ces p-values : $p_{(1)}<p_{(2)}<\ldots<p_{(L)}$, $p_{(l)}$ définissant la l-ième plus petite p-value. La méthode RTP propose de définir la statistique W(k) suivante pour tester l'hypothèse nulle globale au niveau du gène :

$$W(k) = \prod_{i=1}^{k} p(i),$$

Avec k, entier défini à priori dans l'intervalle [1, L]. Si les SNP sont indépendants, la distribution de W(k) et donc sa p-value associée peut s'obtenir de façon analytique. Si les SNP sont corrélés, ce qui est le plus souvent le cas lorsque l'on considère plusieurs SNP dans une même région génomique, une procédure de permutation est utilisée pour simuler la distribution de W(k) sous l'hypothèse nulle et permet ainsi d'obtenir le degré de significativité de W(k). Dans la procédure de permutation, des bases de données sont générée sous l'hypothèse nulle d'absence d'association entre le phénotype et les facteurs génétiques en permutant aléatoirement le phénotype. La statistique W(k) est alors calculée pour chaque base permutée afin d'obtenir sa distribution sous l'hypothèse nulle. La statistique observée dans la base initiale est alors comparée à cette distribution pour le calcul de son degré de significativité.

Pour remédier au choix arbitraire de la valeur k, la méthode ARTP (Yu et al. 2009) propose d'estimer la p-value de W(k) pour J valeurs de k : $k_1 < \ldots < k_J$. On note $s(k_j)$ la p-value estimée de $W(k_j)$, avec $1 \leq j \leq J$. J représente le nombre de points de troncature k choisi et peut varier selon le nombre de SNP dans un gène. Pour limiter le problème de tests multiples, il est recommandé de choisir un maximum de 10 points de troncature.

La statistique définie par l'ARTP est alors la statistique MinP correspondant à la plus petite valeur de s(kj).

$$MinP = min_{1 \leq j \leq J}\, s(k_j)$$

La significativité pour $MinP$ est évaluée à l'aide d'une procédure de permutation du statut cas/témoin dans la base de données initial.

Pour nos analyses, nous avons choisi 5 points de troncature (J=5) définis à chaque 5% des meilleurs SNP en terme de p-value pour l'association au niveau du gène. De la même façon, des p-values par pathway sont définies à partir des p-values estimés au niveau des gènes. Pour déterminer le niveau de significativité de $MinP$, nous avons appliqué 10 000 permutations du statut cas/témoins de notre base de données initiale. Les variables d'ajustement ont été permutées en même temps que le statut cas/témoins afin de préserver l'ajustement.

Pour l'étude des interactions gène-environnement au niveau des gènes et au niveau des pathways, la méthode ARTP a été légèrement modifiée pour utiliser les p-values d'interaction « SNP x Environnement » comme données d'entrée au lieu de la p-value de chaque SNP (package PIGE dans R développé dans l'équipe). La procédure de permutation a également été remplacée par un bootstrap paramétrique afin de fixer les facteurs d'ajustement ainsi que les effets propres des gènes et des facteurs environnementaux (Bůžková et al. 2011)

Correction sur les tests multiples

Pour prendre en compte le risque d'erreur lié à la multiplicité des tests effectués au niveau des SNP, gène et pathways, nous avons utilisé la méthode FDR « False Discovery Rate » qui est une méthode moins stricte que la méthode de Bonferroni. La méthode FDR contrôle l'espérance du taux de faux positifs, ou le nombre de faux positifs parmi les différences déclarées significatives. En pratique, elle consiste à ordonner par ordre croissant les m p-values des m tests (des SNP ou des gènes) et rechercher le plus haut rang k des p-values te que : $p - value(k) \geq \alpha k/m$ avec α correspondant au risque d'erreur de première espèce.

Test d'hétérogénéité

Avant de pooler les études CATHY/Young-Thyr et USRT/UTMDACC, nous avons vérifié l'absence d'hétérogénéité entre les deux groupes d'études, en comparant la statistique de test de l'association entre le CDT et chaque SNP entre les deux études avec le test de Wald.

Calcul de puissance

Nous avons effectué un calcul de puissance pour l'analyse par SNP. En considérant un risque d'erreur de première espèce à 10^{-3}, nous avons estimé la puissance pour mettre en évidence des OR entre 1,10 et 1,40 étant données notre taille d'échantillons pour des SNP ayant des fréquences alléliques entre 0,05 et 0,45 dans l'étude CATHY/Young-Thyr et dans l'analyse combinant les études CATHY/Young-Thyr et USRT/UTMDACC (. Il a été montré que les analyses par pathways avec la méthode ARTP sont généralement plus puissantes que les analyses SNP par SNP (Yu et al. 2009). Le calcul de puissance pour les analyses par pathway pour notre cas particulier nécessiterait des analyses sur données simulées.

Tableau 15). Globalement, nous observons que nous avons une puissance limitée pour mettre en évidence des OR≤1,20 et des SNP de fréquence allélique inférieure à 0,15 dans les deux groupes d'études. Il a été montré que les analyses par pathways avec la méthode ARTP sont généralement plus puissantes que les analyses SNP par SNP (Yu et al. 2009). Le calcul de puissance pour les analyses par pathway pour notre cas particulier nécessiterait des analyses sur données simulées.

Tableau 15: Puissance pour mettre en évidence des OR entre 1,10 et 1,40 dans l'étude CATHY/Young-Thyr et l'analyse poolée CATHY/Young-Thyr et USRT/UTMDACC pour des SNP ayant un une fréquence allélique entre 0,05 et 0,45 et pour un risque d'erreur α de 10^{-3} pour l'effet propre des SNP et les interactions gènes-environnement. Pour les interactions gènes-environnement, nous faisons l'hypothèse d'un facteur E telle que l'obésité avec une prévalence de 15% chez les témoins et un OR=1.34)

		OR				OR(GE) interaction			
MAF chez les témoins		**1,10**	**1,20**	**1,30**	**1,40**	**1,10**	**1,20**	**1,30**	**1,40**
CATHY/Young-Thyr	**463/482**								
0,05		0.4%	2%	8%	20%	0.1%	0.3%	0.7%	1%
0,15		1%	11%	39%	74%	0.2%	0.8%	2%	6%
0,25		2%	21%	62 %	91%	0.3%	1%	4%	11%
0,35		3%	27%	73%	96%	0.3%	2%	5%	13%
0,45		3%	30%	76%	96%	0.4%	2%	5%	14%
Analyse poolée	**743/893**								
0,05		0.6%	5%	18%	43%	0.1%	0.5%	1%	2%
0,15		3%	25%	77%	95%	0.3%	2%	5%	14%
0,25		5%	44%	90%	99%	0.4%	3%	1%	24%
0,35		6%	55%	95%	99%	0.5%	3%	12%	30%
0,45		7%	59%	96%	99%	0.5%	3%	13%	31%

2.3. Résultats

2.3.1. Résultats d'analyse des données CATHY/Young-Thyr

2.3.1.a. Caractéristique de la population d'étude et association avec les facteurs de risque non génétiques

Le Tableau 16 présente les caractéristiques sociodémographiques des cas et témoins des études CATHY/Young-Thyr. La moyenne d'âge est de 50,7 ans dans l'étude CATHY et 28 ans dans l'étude Young-Thyr. Cette différence d'âge est attendue puisque l'étude Young-Thyr inclut des cas qui avaient moins de 15 ans au moment de l'accident de Tchernobyl en 1986.

Tableau 16 : Caractéristiques sociodémographiques des cas et témoins des études CATHY, Young-Thyr et le pool des deux études.

	CATHY		Young-Thyr		CATHY/ Young-Thyr	
Variable	**Cas (%) N=377**	**Témoins (%) N=389**	**Cas (%) N=86**	**Témoins (%) N=93**	**Cas (%) N=463**	**Témoins (%) N=482**
Age en classe de 10						
<30	16 (4,24)	26 (6,68)	43 (50,00)	49 (52,69)	59 (12,74)	75 (15,56)
30-39	54 (14,32)	66 (16,97)	43 (50,00)	44 (47,31)	97 (20,95)	110 (22,82)
40-49	86 (22,81)	96 (24,68)	-	-	86 (18,57)	96 (19,92)
50-59	127 (33,69)	102 (26,22)	-	-	127 (27,43)	102 (21,16)
60-69	65 (17,24)	64 (16,45)	-	-	65 (14,04)	64 (13,28)
>=70	29 (7,69)	35 (9)	-	-	29 (6,26)	35 (7,26)
Etude						
CATHY	377 (81,43)	389 (80,71)			377 (81,43)	389 (80,71)
YOUNGTHYR			86 (18,57)	93 (19,29)	86 (18,57)	93 (19,29)
Département de résidence						
Ardenne	83 (17,93)	69 (14,32)	-	-	83 (17,93)	69 (14,32)
Calvados	124 (26,78)	137 (28,42)	-	-	124 (26,78)	137 (28,42)
Marne	170 (36,72)	183 (37,97)	-	-	170 (36,72)	183 (37,97)
Autre	-	-	86 (18,57)	93 (19,29)	86 (18,57)	93 (19,29)
Type Histologique						
Papillaire	334 (88,59)	-	78 (91,76)	-	412 (89,18)	-
Vésiculaire	43 (11,41)	-	7 (8,24)	-	50 (10,82)	-

Le Tableau 17 présente l'association entre les facteurs de risque non génétiques et les CDT chez les femmes dans chacune des études CATHY et Young-Thyr, et en combinant les deux études. On observe qu'un IMC≥30 est positivement associé au risque de CDT dans l'étude CATHY (OR=1.43 [0,94 – 2,17]) alors que dans l'étude Young-thyr, il y a très peu de sujets en surpoids. La consommation d'au moins 1 verre d'alcool par jour est inversement associée au risque de CDT dans l'étude CATHY (OR=0,54 [0,31 – 0,97]). Dans l'étude Young-thyr, la consommation d'alcool reste très faible et ne permet pas de mettre en évidence d'association. La consommation de tabac en nombre de paquet-année et le statut tabagique ne sont pas associés au risque de CDT dans les deux études.

Concernant les facteurs hormonaux et reproductifs, nous observons un OR>1 lié à un âge tardif aux premières règles dans l'étude CATHY alors que l'OR est inférieur à 1 pour l'étude Young-Thyr. Des associations négatives sont retrouvées pour la prise de contraceptifs oraux dans les deux études avec un OR= 0,72 [0,49 – 1 ,06] lorsque les 2 études sont combinées. Concernant le nombre de grossesse à terme, des associations positives mais non significatives sont observées dans l'étude CATHY. Chez les femmes en post-ménopause dans l'étude CATHY, la prise de THM est inversement associée au risque de CDT (OR=0,61 [0,39 – 0,94]) (Tableau 17)

Tableau 17: Associations entre les facteurs de risque suspectés et les CDT dans l'étude CATHY, Young-Thyr et l'étude CATHY/Young-Thyr (les ORs ont été ajustés sur l'âge en classe de 10 et le département de résidence)

	CATHY				Young-Thyr				CATHY/YOUNG-THYR)			
Variables	**Cas n(%)**	**Témoins n(%)**	**OR**	**IC à 95%**	**Cas n(%)**	**Témoins n(%)**	**OR**	**IC à 95%**	**Cas n(%)**	**Témoins n(%)**	**OR**	**IC à 95%**
	N=365	**N=376**			**N=83**	**N=93**			**N=448**	**N=469**		
IMC (classification OMS)												
<18,5	9 (2,47)	12 (3,19)	0,78	(0,31-1,96)	8 (9,64)	9 (9,68)	0,72	(0,24-2,19)	17 (3,79)	21 (4,48)	0,89	(0,45-1,78)
18,5-24,9	172 (47,12)	205 (54,52)	Réf		52 (62,65)	64 (68,82)	Réf		224 (50)	269 (57,36)	Réf	
25-29,9	110 (30,14)	101 (26,86)	1,21	(0,85-1,72)	14 (16,87)	12 (12,90)	1,62	(0,63-4,21)	124 (27,68)	113 (24,09)	1,22	(0,88-1,69)
≥30	74 (20,27)	58 (15,43)	**1,43**	**(0,94-2,17)**	9 (10,84)	8 (8,60)	1,43	(0,45-4,54)	83 (18,53)	66 (14,07)	**1,38**	**(0,94-2,04)**
Consommation d'alcool en verres/jours												
Jamais ou presque	112 (30,68)	101 (26,86)	Réf		69 (83,13)	68 (73,12)	Réf		181 (40,4)	169 (36,03)	Réf	
0-1	201 (55,07)	218 (57,98)	0,82	(0,58-1,17)	14 (16,87)	20 (21,51)	0,81	(0,36-1,83)	215 (47,99)	238 (50,75)	0,79	(0,58-1,09)
1-2	26 (7,12)	38 (10,11)	**0,59**	**(0,33-1,08)**	0 (0)	3 (3,23)	-	-	26 (5,80)	41 (8,74)	**0,54**	**(0,31-0,97)**
>2	26 (7,12)	19 (5,05)	1,28	(0,65-2,51)	0 (0)	2 (2,15)	-	-	26 (5,80)	21 (4,48)	1,13	(0,59-2,16)
Statut tabagique												
Non-fumeur	209 (57,26)	219 (58,24)	Réf		38 (45,78)	42 (45,16)	Réf		247 (55,13)	261 (55,65)	Réf	
Ancien fumeur	91 (24,93)	79 (21,01)	1,26	(0,86-1,83)	16 (19,28)	17 (18,28)	1,24	(0,52-2,97)	107 (23,88)	96 (20,47)	1,23	(0,87-1,73)
Fumeur actuel	65 (17,81)	78 (20,74)	0,97	(0,64-1,48)	29 (34,94)	34 (36,56)	1,08	(0,53-2,20)	94 (20,98)	112 (23,88)	0,99	(0,70-1,42)
Consommation de tabac en paquets-année												
Non-fumeur	209 (57,26)	219 (58,24)	Réf		38 (45,78)	42 (45,16)	Réf		247 (55,13)	261 (55,65)	Réf	
0-10	42 (11,51)	40 (10,64)	1,20	(0,73-1,98)	4 (4,82)	8 (8,60)	0,60	(0,14-2,55)	46 (10,27)	48 (10,23)	1,08	(0,68-1,72)
10-20	81 (22,19)	78 (20,74)	1,26	(0,85-1,86)	41 (49,40)	41 (44,09)	1,27	(0,65-2,47)	122 (27,23)	119 (25,37)	1,25	(0,89-1,74)
>20	33 (9,04)	39 (10,37)	0,82	(0,48-1,41)	0 (0)	2 (2,15)			33 (7,37)	41 (8,74)	0,80	(0,47-1,34)
Age aux premières règles												
<=12	146 (40,00)	161 (42,82)	0,89	(0,61-1,31)	43 (51,81)	44 (47,31)	0,97	(0,45-2,09)	189 (42,19)	205 (43,71)	0,95	(0,67-1,33)
13	86 (23,56)	93 (24,73)	Réf		21 (25,30)	25 (26,88)	Réf		107 (23,88)	118 (25,16)	Réf	
14	73 (20,00)	71 (18,88)	1,06	(0,67-1,67)	13 (15,66)	14 (15,05)	1,00	(0,35-2,84)	86 (19,20)	85 (18,12)	1,08	(0,72-1,64)
>=15	60 (16,44)	51 (13,56)	1,32	(0,81-2,15)	6 (7,23)	10 (10,75)	0,56	(0,16-1,98)	66 (14,73)	61 (13,01)	1,22	(0,78-1,91)
Prise de contraceptifs oraux												
Jamais	98 (26,85)	85 (22,61)	Réf		14 (16,87)	8 (8,60)	Réf		112 (25,00)	93 (19,83)	Réf	
A déjà pris	267 (73,15)	291 (77,39)	0,79	(0,52-1,21)	69 (83,13)	85 (91,4)	0,43	(0,15-1,17)	336 (75,00)	376 (80,17)	**0,72**	**(0,49-1,06)**
Nombre de grossesses à terme												
0	29 (7,95)	40 (10,64)	Réf		49 (59,04)	54 (58,06)	Réf		78 (17,41)	94 (20,04)	Réf	
1	70 (19,18)	68 (18,09)	1,40	(0,76-2,58)	17 (20,48)	24 (25,81)	0,70	(0,30-1,65)	87 (19,42)	92 (19,62)	1,13	(0,71-1,82)
2	137 (37,53)	138 (36,70)	1,25	(0,71-2,21)	13 (15,66)	14 (15,05)	0,90	(0,33-2,45)	150 (33,48)	152 (32,41)	1,12	(0,71-1,78)
3	81 (22,19)	93 (24,73)	1,10	(0,60-2,03)	3 (3,61)	1 (1,08)	3,38	(0,31-36,99)	84 (18,75)	94 (20,04)	1,03	(0,61-1,74)
>=4	48 (13,15)	37 (9,84)	1,58	(0,79-3,17)	1 (1,20)	0 (0)			49 (10,94)	37 (7,89)	1,42	(0,76-2,64)

2.3.1.b. Association entre les facteurs de risque génétique et les CDT

i. Effet propre des SNP

Les associations entre les 1280 SNP des 2 pathways candidats de Beluhca et le risque de CDT ont été analysées dans l'étude CATHY/Young-Thyr sur 448 cas et 469 témoins. Le Tableau 18 présente les 30 SNP les plus significativement associés au risque de CDT (p<0,05) et la Figure 14 correspond au Q-Q-plot des p-values d'association avec les CDT pour les 1280 SNP analysés comparant leur distribution à celle d'une loi uniforme. Le QQ plot ne montre pas de déviation à la loi uniforme (inflation lambda=1,006). Parmi les 1280 SNP analysés, 69 SNP ont présenté une p-value< 5% (Figure 14) mais aucune association significative ne persistait après correction sur les tests multiples. Le meilleur signal observé est celui du SNP rs17868334 situé sur les gènes *UGT1A5/UGT1A4* (p=2,9x10^{-3}).

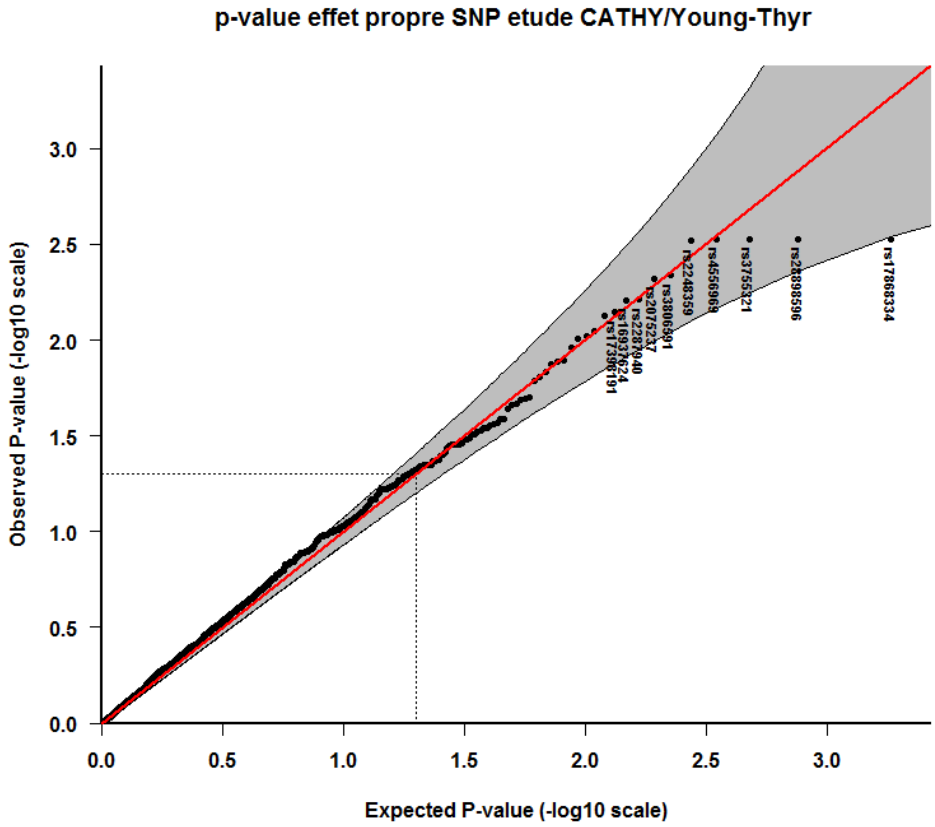


Figure 14 : Q-Qplot des résultats de p-values d'associations entre les SNP et le risque de CDT dans l'étude CATHY/Young-Thyr L'axe des abscisses représente les p-values attendues sous l'hypothèse d'une distribution uniforme et l'axe des ordonnées représente les p-values observées. Les traits en pointillés représentent le seuil de significativité à 5%.

Tableau 18 : top 30 des SNP significativement associés au risque de CDT au seuil 5% avant correction FDR

SNP	CHR[1]	position	Gène	Allèle		MAF[4]		OR	95%CI	pvalue	FDR[6]
				Risk[2]	Ref[3]	Cas	Co[5]				
rs17868334	2	234624817	*UGT1A5/UGT1A4*	G	A	0,11	0,07	1,68	(1,19-2,36)	2,96E-03	0,78
rs28898596	2	234620640	*UGT1A5/UGT1A6*	A	G	0,11	0,07	1,68	(1,19-2,36)	2,96E-03	0,78
rs3755321	2	234621825	*UGT1A5*	C	T	0,11	0,07	1,68	(1,19-2,36)	2,96E-03	0,78
rs4556969	2	234621404	*UGT1A6/UGT1A5*	T	C	0,11	0,07	1,68	(1,19-2,36)	2,96E-03	0,78
rs2248359	20	52791518	*CYP24A1/PFDN4*	T	C	0,37	0,44	0,75	(0,62-0,91)	3,03E-03	0,78
rs3806591	2	234626199	*UGT1A5/UGT1A4*	C	T	0,10	0,07	1,64	(1,17-2,31)	4,59E-03	0,85
rs2075237	12	16510079	*MGST1*	G	T	0,23	0,18	1,39	(1,11-1,75)	4,77E-03	0,85
rs2287940	19	41620419	*CYP2F1*	A	G	0,14	0,18	0,71	(0,55-0,91)	6,16E-03	0,85
rs16937624	11	43780395	*HSD17B12*	G	A	0,07	0,04	1,84	(1,19-2,85)	6,18E-03	0,85
rs17398191	2	48924876	*LHCGR*	T	C	0,06	0,04	1,88	(1,19-2,97)	7,18E-03	0,85
rs706350	4	140600437	*MGST2*	T	C	0,40	0,45	0,76	(0,63-0,93)	7,46E-03	0,85
rs305975	19	41628677	*CYP2F1*	G	T	0,18	0,23	0,74	(0,59-0,93)	8,98E-03	0,85
rs795589	4	140600125	*MGST2*	C	T	0,39	0,45	0,78	(0,64-0,94)	9,55E-03	0,85
rs6413421	10	135345811	*CYP2E1*	C	T	0,07	0,04	1,73	(1,14-2,62)	9,85E-03	0,85
rs4143630	10	5063944	*AKR1C2/AKR1C3*	T	C	0,45	0,52	1,28	(1,06-1,54)	1,09E-02	0,85
rs1587265	4	140601791	*MGST2*	G	A	0,44	0,38	1,29	(1,06-1,57)	1,27E-02	0,85
rs6413420	10	135340829	*LOC619207/CYP2E1*	T	G	0,07	0,04	1,69	(1,12-2,55)	1,29E-02	0,85
rs1937888	10	5060633	*AKR1C2/AKR1C3*	C	T	0,53	0,46	1,27	(1,05-1,54)	1,33E-02	0,85
rs1007540	2	49209108	*FSHR*	C	T	0,13	0,18	0,73	(0,57-0,94)	1,47E-02	0,85
rs795593	4	140608002	*MGST2*	G	A	0,13	0,09	1,46	(1,07-1,99)	1,55E-02	0,85
rs8190512	9	99050100	*HSD17B3*	C	A	0,50	0,45	1,27	(1,04-1,53)	1,63E-02	0,85
rs1404057	2	48924091	*LHCGR*	C	T	0,14	0,10	1,40	(1,05-1,86)	1,99E-02	0,85
rs2279357	15	74630623	*CYP11A1*	T	C	0,34	0,30	1,27	(1,04-1,56)	2,04E-02	0,85
rs2445762	15	51617708	*CYP19A1*	C	T	0,29	0,25	1,29	(1,04-1,59)	2,04E-02	0,85
rs2608624	6	52622475	*GSTA2*	G	T	0,34	0,39	0,79	(0,65-0,97)	2,15E-02	0,85
rs2608632	6	52627698	*GSTA2*	T	G	0,31	0,35	0,79	(0,64-0,97)	2,18E-02	0,85
rs4699733	4	100137534	*ADH6*	G	C	0,25	0,30	0,79	(0,65-0,97)	2,29E-02	0,85
rs1604144	2	234605835	*UGT1A5/UGT1A6*	T	C	0,30	0,26	1,26	(1,03-1,55)	2,60E-02	0,85
rs7562567	2	48953256	*LHCGR*	T	C	0,17	0,21	0,76	(0,60-0,97)	2,60E-02	0,85

Les ORs sont ajustés sur l'âge en classe de 5 ans, la région de résidence et les 2 premières composantes principales, [1]CHR : Chromosome ; [2]Risk: Allèle à risque (allèle mineur) ; [3]ref : Allèle de référence (Allèle majeur) ; [4]MAF (fréquence de l'allèle mineur ou allèle à risque) ; [5]Co : Témoins ; [6]FDR : correction False Discovery Rate.

ii. Effet propre des gènes et des pathways

Parmi les 138 gènes étudiés, seuls 2 gènes (*CYP2F1* et *UGT1A8*) ont présenté une p-value< 5% mais aucune association ne persistait après correction sur les tests multiples (Tableau 19). Aucune association significative n'était également observée au niveau des pathways (Tableau 19)

Interactions gène-environnement

Nous avons analysé les interactions au niveau des SNP, des gènes et des pathways avec les facteurs suivants : consommation d'alcool moyenne sur la vie entière (en 3 classes : jamais ; <1 verre par jour et ≥1 verre par jour), statut tabagique (en 3 classes : non-fumeur, ancien fumeur, fumeur actuel), quantité de tabac consommée sur la vie entière (en 3 classes en : non-fumeur, 1-10 paquets-années et >10 paquets-années), l'IMC (en 3 classes : <25 kg/m² ; entre 25 et 29,9 kg/m² et ≥30 kg/m²), l'âge aux premières règles (en 3 classes : <13

ans ; 13-14 ans et >14 ans) ; le nombre de grossesses à terme (classé en : 0, 1, 2 et ≥3) et la prise de contraceptifs oraux (classé en : jamais, a déjà été utilisatrice).

Les p-values des résultats d'interaction au niveau des SNP et des gènes sont présentées respectivement dans les annexes 10 et 11.

Interaction avec l'IMC

Parmi les 1280 SNP analysés, nous avons rapportés 68 SNP ayant une p-value d'interaction inférieure à 5%. Après correction sur les tests multiples, aucune interaction n'était significative. De même, au niveau des 138 gènes, aucune interaction n'a été retrouvée après correction sur les tests multiple (Tableau 19). L'analyse au niveau des pathways ne montre pas non plus d'interaction avec l'IMC (Tableau 20).

Interaction avec la consommation d'alcool

30 SNP sur les 1280 SNP analysés avaient une interaction significative au seuil 5 % avec la consommation d'alcool mais aucune des interactions ne persistait après correction sur les tests multiples.

Au niveau des gènes, seule l'interaction avec le gène *CYP24A1* persistait après correction sur les tests multiples (p-FDR=0,01; Tableau 19). Ce gène fait partie du pathway de la « biosynthèse et métabolisme des hormones stéroïdiennes » ; il contient 19 SNP parmi lesquels 4 SNP (rs2585428, rs2296241, rs4809959 et rs2181874) avaient des p-values d'interaction inférieures à 5%.

Au niveau des pathways, nous n'avons observé aucune interaction avec la consommation de d'alcool.

Tableau 19 : p-values d'association entre quelques gènes et le risque de CDT et interaction avec les facteurs de risque non génétique (les gènes présentés sont ceux pour lesquels des interactions avec les facteurs de risque non génétiques sont significatives au seuil 5%)

		Effet propre		Alcool		tabac en paquet année		IMC		contraceptifs oraux		Age première règles		nb de grossesse à terme	
Gènes	**N,SNP**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**
CYP2F1	7	2,37E-02	0,92	8,74E-01	1,00	**2,80E-03**	**0,09**	7,28E-01	1,00	2,94E-01	1,00	8,24E-01	1,00	3,50E-01	1,00
UGT1A8	69	4,77E-02	0,92	1,00E+00	1,00	2,16E-02	0,22	8,08E-01	1,00	9,95E-01	1,00	9,00E-01	1,00	1,00E+00	1,00
CYP24A1	27	7,74E-02	0,92	**1,00E-04**	**0,01**	2,97E-01	0,76	9,99E-01	1,00	3,94E-01	1,00	6,07E-02	0,78	1,00E+00	1,00
GSTK1	1	5,32E-01	0,94	1,05E-02	0,67	7,33E-01	0,99	8,18E-01	1,00	5,51E-01	1,00	2,80E-01	1,00	9,13E-01	1,00
NAT2	11	1,50E-01	0,92	2,23E-02	0,92	3,56E-01	0,82	4,40E-02	0,68	7,80E-01	1,00	9,00E-01	1,00	9,99E-01	1,00
ALDH3B1	7	1,96E-01	0,92	3,33E-02	0,92	3,65E-01	0,82	1,13E-01	0,87	5,59E-01	1,00	9,10E-01	1,00	9,23E-01	1,00
COMT	13	7,52E-01	0,96	3,78E-02	0,92	7,56E-01	0,99	3,50E-01	1,00	8,21E-01	1,00	9,56E-01	1,00	9,38E-01	1,00
SRD5A2	7	3,51E-01	0,94	9,98E-01	1,00	**3,10E-03**	**0,09**	9,40E-01	1,00	5,72E-02	0,58	3,34E-01	1,00	9,72E-01	1,00
ADH4	7	1,64E-01	0,92	9,59E-01	1,00	6,30E-03	0,13	9,66E-01	1,00	1,25E-01	0,70	5,37E-01	1,00	9,77E-01	1,00
CGA	5	8,05E-01	0,96	5,27E-01	1,00	**2,10E-03**	**0,09**	9,57E-01	1,00	8,11E-01	1,00	9,98E-01	1,00	9,97E-01	1,00
GSTA5	2	7,44E-01	0,96	9,13E-02	1,00	**5,00E-04**	**0,06**	3,29E-01	1,00	5,91E-01	1,00	6,23E-01	1,00	4,51E-01	1,00
SOD2	6	8,95E-01	0,96	9,42E-01	1,00	2,62E-01	0,76	9,64E-01	1,00	4,07E-02	0,58	7,11E-01	1,00	5,26E-01	1,00
ADH6	8	8,47E-02	0,92	9,91E-01	1,00	2,43E-02	0,22	8,03E-01	1,00	3,32E-01	1,00	1,71E-01	1,00	1,00E+00	1,00
ALDH3B2	3	3,03E-01	0,92	3,87E-01	1,00	7,64E-02	0,44	2,43E-02	0,52	8,36E-01	1,00	2,84E-01	1,00	6,09E-01	1,00
CYP2F1	7	2,37E-02	0,92	8,74E-01	1,00	**2,80E-03**	**0,09**	7,28E-01	1,00	2,94E-01	1,00	8,24E-01	1,00	3,50E-01	1,00
APOC3	3	1,56E-01	0,92	9,72E-01	1,00	**3,50E-03**	**0,09**	9,41E-01	1,00	9,17E-01	1,00	1,23E-01	1,00	4,54E-01	1,00
APOA4	3	1,54E-01	0,92	7,11E-01	1,00	7,10E-03	0,13	1,35E-01	0,87	9,29E-01	1,00	1,38E-01	1,00	6,01E-01	1,00
FSHR	47	4,51E-01	0,94	9,99E-01	1,00	9,10E-03	0,14	2,65E-01	0,99	8,83E-01	1,00	1,00E+00	1,00	8,06E-01	1,00
CYP2C9	18	1,91E-01	0,92	9,52E-01	1,00	9,80E-03	0,14	9,93E-01	1,00	6,39E-02	0,58	8,03E-01	1,00	5,88E-01	1,00
DHDH	6	3,81E-01	0,94	5,41E-01	1,00	1,15E-02	0,15	2,10E-01	0,97	3,22E-01	1,00	7,82E-01	1,00	5,78E-01	1,00
CYP46A1	12	4,37E-01	0,94	7,16E-01	1,00	1,82E-02	0,21	1,28E-01	0,87	8,05E-01	1,00	2,99E-02	0,64	7,46E-01	1,00
GNRHR	9	6,61E-01	0,96	8,26E-01	1,00	2,41E-02	0,22	5,41E-01	1,00	2,70E-01	1,00	7,57E-01	1,00	6,14E-01	1,00
FSHB	2	9,63E-01	0,97	9,85E-01	1,00	4,16E-02	0,35	2,99E-01	1,00	6,77E-01	1,00	8,36E-01	1,00	7,11E-01	1,00
CYP3A7	3	5,36E-01	0,94	8,40E-01	1,00	8,62E-01	0,99	3,80E-03	0,38	2,21E-01	0,94	2,08E-01	1,00	6,83E-01	1,00
MGST1	10	5,13E-02	0,92	5,62E-02	0,92	2,12E-01	0,71	6,00E-03	0,38	1,00E+00	1,00	2,96E-02	0,64	4,61E-01	1,00
POMC	7	2,75E-01	0,92	5,74E-01	1,00	9,78E-01	0,99	1,51E-02	0,52	9,19E-01	1,00	1,94E-01	1,00	7,45E-01	1,00
CYP3A4	3	4,54E-01	0,94	9,39E-01	1,00	6,26E-01	0,99	1,72E-02	0,52	1,26E-01	0,70	3,24E-01	1,00	7,46E-01	1,00
AKR1A1	3	6,61E-01	0,96	8,36E-01	1,00	6,50E-01	0,99	2,10E-02	0,52	8,88E-01	1,00	1,48E-01	1,00	5,92E-01	1,00
UGT2B7	5	4,46E-01	0,94	7,13E-01	1,00	6,18E-01	0,99	3,36E-02	0,61	7,45E-02	0,60	8,98E-01	1,00	2,24E-01	1,00
UGT2A3	3	9,99E-01	1,00	6,69E-01	1,00	3,93E-01	0,84	4,76E-02	0,68	2,66E-01	1,00	9,57E-01	1,00	9,26E-01	1,00
UGT2B11	3	8,35E-01	0,96	9,57E-01	1,00	2,95E-01	0,76	5,65E-02	0,71	1,40E-03	0,15	6,64E-01	1,00	8,37E-01	1,00
LHB	1	2,85E-01	0,92	3,35E-01	1,00	6,80E-02	0,44	7,40E-01	1,00	2,90E-03	0,15	8,67E-01	1,00	8,45E-01	1,00
AKR1D1	8	6,60E-01	0,96	9,44E-01	1,00	6,97E-02	0,44	7,33E-01	1,00	3,50E-03	0,15	6,91E-02	0,80	8,75E-01	1,00
CYP27B1	2	5,77E-01	0,95	4,49E-01	1,00	1,02E-01	0,46	8,24E-01	1,00	1,07E-02	0,34	9,07E-01	1,00	4,45E-01	1,00
CYP2C19	14	3,50E-01	0,94	3,49E-01	1,00	5,82E-01	0,99	9,14E-01	1,00	1,32E-02	0,34	1,91E-01	1,00	9,97E-01	1,00
CYP3A5	2	9,15E-01	0,96	8,36E-01	1,00	2,04E-01	0,71	6,80E-02	0,71	2,16E-02	0,46	5,49E-01	1,00	8,47E-01	1,00
SULT1A2	1	9,28E-01	0,96	1,21E-01	1,00	5,11E-02	0,41	4,93E-01	1,00	3,01E-02	0,55	9,15E-01	1,00	1,86E-01	1,00
UGT2B4	8	8,26E-01	0,96	9,89E-01	1,00	3,95E-01	0,84	9,53E-01	1,00	4,33E-02	0,58	6,81E-01	1,00	9,93E-01	1,00
CYP11A1	8	8,36E-02	0,92	1,42E-01	1,00	4,46E-01	0,88	1,00E+00	1,00	1,00E+00	1,00	**4,00E-04**	**0,05**	6,87E-02	1,00
HSD17B8	2	2,45E-01	0,92	7,44E-01	1,00	8,14E-01	0,99	7,21E-02	0,71	7,15E-01	1,00	**8,00E-04**	**0,05**	8,33E-01	1,00
CYP39A1	25	6,36E-01	0,96	9,72E-01	1,00	1,56E-01	0,59	7,89E-01	1,00	6,79E-01	1,00	2,97E-02	0,64	8,77E-01	1,00
PPARD	21	9,01E-01	0,96	7,39E-01	1,00	1,00E+00	1,00	1,00E+00	1,00	1,07E-01	0,70	4,05E-02	0,74	9,99E-01	1,00
SULT2A1	7	2,86E-01	0,92	2,14E-01	1,00	1,23E-01	0,49	9,90E-01	1,00	6,78E-01	1,00	4,63E-02	0,74	9,49E-01	1,00

Tableau 20: p-values d'association entre les pathways et le risque de CDT et leurs interactions avec les facteurs de risque non génétiques

		p-interaction					
Pathways	**Effet propre**	**Alcool**	**Tabac (pq-an)[1]**	**IMC**	**âge prem regle[2]**	**Contraceptifs oraux**	**nb grossesse à terme[3]**
Métabolisme des xénobiotiquew	0,22	0,56	**0,02**	0,11	0,96	0,11	1,00
Métabolisme et biosynthèse des hormones stéroïdiennes	0,87	0.48	**7,2E-3**	0,42	**0.05**	**0,03**	1 ,00

[1] Consommation de tabac en paquet-année ; [2]age aux premières règles ; [3]nombre de grossesse à terme,

Interaction avec la consommation de tabac

L'analyse par SNP ne montre pas d'interaction significative avec la consommation de tabac en paquets-années alors qu'au niveau des gènes une interaction limite significative avec le gène *GSTA5* (p-FDR=0,06) a été observée après correction sur les tests multiples. Des p-FDR < 0,10 ont également été observées pour les gènes *CYP2F1, SRD5A2, CGA et APOC3* (Tableau 19). *GSTA5* et *CYP2F1* appartiennent au pathway du métabolisme des xénobiotiques alors que *SRD5A2*, *CGA* et *APOC3* interviennent dans le métabolisme des hormones stéroïdiennes.

Au niveau des pathway, nous avons rapportés des interactions significatives avec les deux pathways étudiés (Tableau 20).

Interaction avec l'âge aux premières règles

Parmi les 1280 SNP analysés, 57 SNP avaient des interactions significatives au seuil 5 % mais qui ne persistaient pas après correction sur les tests multiples.

Au niveau des gènes, nous avons observé des interactions limites significatives après prise en compte des tests multiples avec les gènes *CYP11A1* et *HSD17B8 (p-FDR=0,05).* Ces deux gènes font partie du pathway de la « biosynthèse et du métabolisme des hormones stéroïdiennes ». Le gène *CYP11A1* est tagué par 5 SNP parmi lesquels le SNP rs2279357 pour lequel nous avons observé une p-value d'interaction inférieure à 5% avec l'âge aux premières règles. Le gène *HSD17B8* est tagué dans nos études par 2 SNP (rs107822 et rs1547387) qui avaient tous deux des p-values d'interactions inférieurs à 5%.

Au niveau des pathways, nous avons rapporté une interaction limite significative avec le pathway de la « biosynthèse et du métabolisme des hormones stéroïdiennes » (p=0,05).

Interaction avec la prise de contraceptifs oraux

Nous n'avons pas observé d'interactions significatives après prise en compte des tests multiples pour les analyses par SNP et par gène. Toutefois, nous avons noté une interaction avec le pathway de la « biosynthèse et métabolisme des hormones stéroïdiennes » (p=0,03). Les gènes qui contribuent le plus à cette interaction sont les gènes *LHB, UGT2B11, AKR1D1, CYP27B1 et CYP2C19.*

Interaction avec le nombre de grossesse à terme

Nous n'avons pas rapportés d'interaction avec le nombre de grossesse à terme au niveau des SNP, des gènes et des pathways.

iii. Analyse stratifiée

En stratifiant les analyses par pathway sur les facteurs pour lesquels nous avons observé des interactions (la consommation de tabac sur la vie entière, l'âge aux premières règles et la prise de contraceptifs oraux), nous n'avons pas observé d'association significative (Tableau 21).

Tableau 21 : P value résultats d'association entre les pathways et le risque de CDT stratifié sur le statut tabagique, l'âge aux premières règles et la prise de contraception orale dans l'étude CATHY/Young-Thyr.

	Tabac en paquets année			**Age aux premières règles**			**pilule**	
Pathway	**Non-fumeur**	**1-10**	**> 10**	**<13 ans**	**13-14 ans**	**>14 ans**	**Jamais**	**Déjà pris**
Métabolisme des xénobiotiques	0,11	0,15	0,84	0,54	0,36	0,98	0,42	0,06
Biosynthèse et métabolisme des hormones stéroïdiennes	0,94	0,87	0,28	0,50	0,35	0,55	0,62	0,84

2.3.2. Résultat de l'analyses poolée CATHY/Young-Thyr et USRT/UTMDACC

2.3.2.a. Description population d'études et association avec les facteurs de risque non-génétique

Le Tableau 22 présente les caractéristiques des populations finales d'études et les OR d'association avec l'IMC, la consommation d'alcool et de tabac. L'âge moyen est d'environ 46 ans dans l'étude CATHY/Young-Thyr, 40 ans dans l'étude USRT/UTMDACC et 45 ans en combinant les études. 89% cancers sont des cancers papillaires dans les études CATHY/Young-Thyr alors que tous les cancers sont de types papillaires dans l'étude USRT/UTMDACC. On retrouve bien l'association attendue avec l'obésité dans les 2 groupes d'études et dans l'analyse combinée. Des associations inverses avec la consommation d'alcool sont rapportées par l'étude CATHY/Young-Thyr et l'étude USRT/UTMDACC, ainsi qu'avec la consommation de tabac.

Tableau 22: Caractéristiques de la population d'étude et OR d'association avec l'IMC, la consommation d'alcool et de tabac.

	Données poolées (743/903)				CATHY/YOUNG-THYR (463/482)				USRT/UTMDACC (271/421)			
variables	**Cas (%)**	**Témoins (%)**	**OR**	**IC à 95%**	**Cas (%)**	**Témoins (%)**	**OR**	**IC à 95%**	**Cas (%)**	**Témoins (%)**	**OR**	**IC à 95%**
Age (année)												
<30	103 (14,03)	132 (14,62)	-	-	59 (12,74)	75 (15,56)	-	-	44 (16,24)	57 (13,54)	-	-
30-39	162 (22,07)	224 (24,81)	-	-	97 (20,95)	110 (22,82)	-	-	65 (23,99)	114 (27,08)	-	-
40-49	179 (24,39)	247 (27,35)	-	-	86 (18,57)	96 (19,92)	-	-	93 (34,32)	151 (35,87)	-	-
50-59	173 (23,57)	170 (18,83)	-	-	127 (27,43)	102 (21,16)	-	-	46 (16,97)	68 (16,15)	-	-
60-69	84 (11,44)	95 (10,52)	-	-	65 (14,04)	64 (13,28)	-	-	19 (7,01)	31 (7,36)	-	-
>=70	33 (4,5)	35 (3,88)	-	-	29 (6,26)	35 (7,26)	-	-	4 (1,48)	0 (0)	-	-
Etude												
CATHY	377 (51,36)	389 (43,08)	-	-	377 (81,43)	389 (80,71)	-	-	-	-	-	-
YOUNG-THYR	86 (11,72)	93 (10,3)	-	-	86 (18,57)	93 (19,29)	-	-	-	-	-	-
UTMDACC	90 (12,26)	0 (0)	-	-	-	-	-	-	90 (33,21)	0 (0)	-	-
USRT	181 (24,66)	421 (46,62)	-	-	-	-	-	-	181 (66,79)	421 (100)	-	-
Type histologique												
Papillaire	693(93,27)	-	-	-	412 (89,18)	-	-	-	281(100)	-	-	-
Vésiculaire	50 (6,73)	-	-	-	50 (10,82)	-	-	-	-	-	-	-
IMC (kg/m²)												
<25	403 (55,21)	576 (64,5)	1,00		246 (53,36)	293 (61,17)	1,00		157 (58,36)	283 (68,36)	1,00	
25-29,9	197 (26,99)	202 (22,62)	1,29	(1,00-1,65)	131 (28,42)	116 (24,22)	1,22	(0,89-1,68)	66 (24,54)	86 (20,77)	1,33	(0,89-1,98)
≥30	130 (17,81)	115 (12,88)	**1,45**	**(1,08-1,96)**	84 (18,22)	70 (14,61)	**1,34**	**(0,92-1,95)**	46 (17,1)	45 (10,87)	**1,64**	**(1,01-2,65)**
Consommation d'alcool												
Jamais	338 (46,62)	369 (41,88)	1,00		188 (40,78)	172 (35,98)	1,00		150 (56,82)	197 (48,88)	1,00	
Déjà pris	387 (53,38)	512 (58,12)	**0,76**	**(0,60-0,94)**	273 (59,22)	306 (64,02)	**0,75**	**(0,56-1,02)**	114 (43,18)	206 (51,12)	**0,83**	**(0,59-1,16)**
Statut tabagique												
Non-fumeur	420 (57,3)	488 (54,34)	1,00		253 (54,64)	267 (55,39)	1,00		167 (61,85)	221 (53,12)		
Ancien fumeur	180 (24,56)	188 (20,94)	1,11	(0,86-1,43)	112 (24,19)	100 (20,75)	1,16	(0,83-1,62)	68 (25,19)	88 (21,15)	1,07	(0,71-1,59)
Fumeur actuel	133 (18,14)	222 (24,72)	**0,74**	**(0,56-0,97)**	98 (21,17)	115 (23,86)	0,96	(0,68-1,35)	35 (12,96)	107 (25,72)	**0,47**	**(0,30-0,74)**

Les ORs sont ajustés sur l'âge, l'étude ou la région de résidence

2.3.2.b. Association entre les facteurs de risque génétique et les CDT

i. Effet propre des SNP

Au total, nous avons analysé 1031 SNP de 118 gènes en combinant les études CATHY/Young-Thyr et USRT/UTMDACC.

La Figure 15 présente le Q-Q plot de l'association entre ces SNP et le risque de CDT dans l'analyse poolée ainsi que dans chacune des études CATHY/Young-Thyr et USRT/UTMDACC séparément. Les QQ-plot entre les p-values observées et attendues ne montrent pas de déviation à la distribution uniforme indiquant qu'il ne semble pas y avoir de biais de stratification de population majeure, ni d'association significative entre les SNP et le CDT.

Le Tableau 23 présente le résultat des 30 SNP les plus associés au risque de CDT dans l'analyse poolée ainsi que dans les études CATHY/Young-Thyr et USRT/UTMDACC séparément.

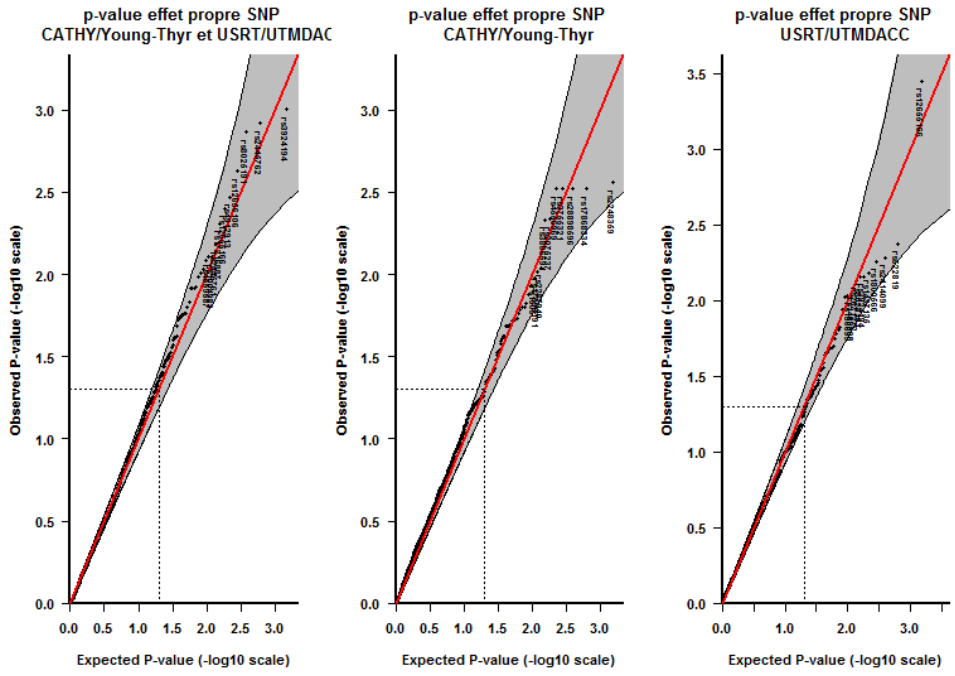


Figure 15 : Q-Q plot des résultats d'association entre les SNP et le risque de CDT dans les études CATHY/YOUNG-THYR, USRT/UTMADCC et dans l'analyse poolée des études. L'axe des abscisses représente les p-values attendues sous l'hypothèse d'une distribution uniforme et l'axe des ordonnées représente les p-values observées. Les traits en pointillés représentent le seuil de significativité à 5%.

Tableau 23 : OR et IC à 95% de l'association entre les 30 SNP les plus associés (sur 1031 SNP étudiés) au risque de CDT dans l'analyse poolée et leurs associations correspondantes dans chacune des études CATHY/YOUNG-THYR et USRT/UTMDACC séparément

		données poolées				**CATHY/Young-Thyr**				**USRT/UTMDACC**			
SNP	Gene	**OR**	**IC à 95%**	**p-value**	**FDR**	**OR**	**IC à 95%**	**p-value**	**FDR**	**OR**	**IC à 95%**	**p-value**	**FDR**
rs3924194	*UGT2B7*	0,71	(0,58-0,87)	9,92E-04	0,47	0,79	(0,60-1,05)	1,01E-01	0,87	0,64	(0,46-0,89)	6,93E-03	0,81
rs2445762	*CYP19A1*	0,77	(0,65-0,90)	1,21E-03	0,47	0,78	(0,63-0,96)	2,05E-02	0,81	0,74	(0,57-0,96)	2,14E-02	0,95
rs8025191	*CYP19A1*	1,33	(1,12-1,58)	1,36E-03	0,47	1,25	(0,99-1,57)	6,19E-02	0,81	1,47	(1,11-1,94)	7,03E-03	0,81
rs12656106	*NR3C1*	1,25	(1,08-1,44)	2,36E-03	0,61	1,23	(1,02-1,49)	3,11E-02	0,81	1,26	(1,01-1,58)	4,10E-02	0,95
rs4912913	*NR3C1*	1,23	(1,07-1,42)	3,38E-03	0,68	1,25	(1,04-1,51)	1,88E-02	0,81	1,23	(0,99-1,53)	6,65E-02	0,95
rs12655166	*NR3C1*	0,79	(0,68-0,93)	3,97E-03	0,68	0,96	(0,77-1,19)	6,78E-01	0,94	0,65	(0,51-0,82)	3,58E-04	0,37
rs17216887	*ADH5*	0,57	(0,38-0,84)	4,92E-03	0,69	0,61	(0,38-0,99)	4,43E-02	0,81	0,47	(0,22-1,00)	4,96E-02	0,95
rs2445761	*CYP19A1*	1,22	(1,06-1,40)	6,54E-03	0,69	1,15	(0,95-1,39)	1,51E-01	0,88	1,30	(1,04-1,63)	2,02E-02	0,95
rs1004983	*CYP19A1*	1,22	(1,05-1,40)	7,83E-03	0,69	1,12	(0,92-1,35)	2,59E-01	0,89	1,34	(1,07-1,68)	1,14E-02	0,90
rs4559587	*AKR1C3*	0,65	(0,47-0,89)	7,86E-03	0,69	0,67	(0,45-0,98)	4,13E-02	0,81	0,56	(0,30-1,04)	6,67E-02	0,95
rs8192047	*MGST2*	0,81	(0,69-0,95)	8,29E-03	0,69	0,81	(0,65-1,01)	6,17E-02	0,81	0,77	(0,60-0,99)	3,91E-02	0,95
rs1154415	*ADH5*	0,83	(0,72-0,95)	9,24E-03	0,69	0,83	(0,69-1,01)	5,88E-02	0,81	0,79	(0,63-0,99)	4,42E-02	0,95
rs28898596	*UGT1A8/UGT1A10/UGT1A9 /UGT1A7/UGT1A6/UGT1A5*	1,38	(1,08-1,77)	9,79E-03	0,69	1,68	(1,19-2,36)	3,03E-03	0,62	1,11	(0,75-1,64)	5,99E-01	0,98
rs915907	*CYP2E1*	1,31	(1,07-1,62)	1,03E-02	0,69	1,25	(0,94-1,64)	1,20E-01	0,87	1,32	(0,95-1,84)	9,64E-02	0,95
rs8190512	*HSD17B3*	1,21	(1,04-1,40)	1,20E-02	0,69	1,28	(1,05-1,56)	1,57E-02	0,81	1,16	(0,92-1,45)	2,09E-01	0,95
rs17868334	*UGT1A8/UGT1A10/ UGT1A9/UGT1A7/UGT1A6/ UGT1A5/UGT1A4*	1,37	(1,07-1,76)	1,21E-02	0,69	1,68	(1,19-2,36)	3,03E-03	0,62	1,09	(0,74-1,60)	6,76E-01	0,98
rs3755321	*UGT1A8/UGT1A10/UGT1A9 /UGT1A7/UGT1A6/UGT1A5*	1,37	(1,07-1,76)	1,21E-02	0,69	1,68	(1,19-2,36)	3,03E-03	0,62	1,09	(0,74-1,60)	6,78E-01	0,98
rs4556969	*UGT1A8/UGT1A10/UGT1A9 /UGT1A7/UGT1A6/UGT1A5*	1,37	(1,07-1,76)	1,21E-02	0,69	1,68	(1,19-2,36)	3,03E-03	0,62	1,09	(0,74-1,60)	6,78E-01	0,98
rs795589	*MGST2*	1,20	(1,04-1,38)	1,47E-02	0,72	1,26	(1,04-1,53)	2,06E-02	0,81	1,17	(0,93-1,46)	1,82E-01	0,95
rs3806591	*UGT1A8/UGT1A10/UGT1A9 /UGT1A7/UGT1A6/UGT1A5 /UGT1A4*	1,36	(1,06-1,74)	1,59E-02	0,72	1,64	(1,16-2,31)	4,66E-03	0,69	1,09	(0,74-1,60)	6,76E-01	0,98
rs2248359	*CYP24A1*	0,84	(0,73-0,97)	1,72E-02	0,72	0,75	(0,62-0,90)	2,76E-03	0,62	0,98	(0,79-1,23)	8,92E-01	0,98
rs9332238	*CYP2C9*	1,24	(1,04-1,48)	1,72E-02	0,72	1,30	(1,02-1,66)	3,29E-02	0,81	1,15	(0,87-1,51)	3,30E-01	0,98
rs17134588	*AKR1C4*	1,27	(1,04-1,55)	1,75E-02	0,72	1,19	(0,91-1,56)	1,99E-01	0,88	1,41	(1,04-1,92)	2,79E-02	0,95
rs3763941	*ALDH3B1*	0,66	(0,47-0,93)	1,75E-02	0,72	0,71	(0,44-1,15)	1,67E-01	0,88	0,64	(0,39-1,05)	7,55E-02	0,95
rs4293599	*LHCGR*	0,81	(0,68-0,96)	1,82E-02	0,72	0,85	(0,67-1,09)	1,96E-01	0,88	0,78	(0,59-1,03)	7,50E-02	0,95
rs706350	*MGST2*	0,84	(0,73-0,97)	1,83E-02	0,72	0,78	(0,64-0,95)	1,57E-02	0,81	0,87	(0,70-1,09)	2,33E-01	0,95
rs7586280	*LHCGR*	0,84	(0,72-0,97)	1,89E-02	0,72	0,90	(0,73-1,10)	3,07E-01	0,89	0,78	(0,62-0,99)	4,15E-02	0,95
rs1799853	*CYP2C9*	0,77	(0,62-0,96)	2,06E-02	0,76	0,71	(0,53-0,96)	2,64E-02	0,81	0,86	(0,63-1,19)	3,72E-01	0,98

Les ORs sont ajustés sur l'âge en classe de 10 ans, l'étude ou région de résidence et les 2 premières composantes principales

i. Effet propre des gènes et pathways

Le Tableau 24 présente les résultats pour les gènes associés au risque de CDT au seuil 5% dans l'analyse poolée ou dans les études CATHY/Young-Thyr et USRT/UTMDACC séparément.

Parmi les 118 gènes analysés, aucune association significative n'a été observée après correction sur les tests multiples (*Tableau 24*). De même, aucune association n'est observée au niveau des pathways (*Tableau 25*).

Tableau 24 : p-value d'association entre les gènes les plus associés au risque de CDT dans l'analyse poolée, et leurs associations dans les études CATHY/ Young-Thyr et USRT/UTMDACC séparément. Les gènes présentés sont ceux significatif au seuil 5% avant correction FDR dans l'analyse poolée.

		Analyse poolée		**URST/UTMDACC**		**CATHY/YOUNG-THYR**	
Gene	**N.SNP**	**p-value**	**FDR**	**p-value**	**FDR**	**p-value**	**FDR**
CYP19A1	44	**1.27E-02**	0.47	3.75E-01	0.91	7.95E-02	0.86
NR3C1	15	**1.38E-02**	0.47	2.08E-01	0.85	**8.20E-03**	0.54
UGT2B7	5	**1.58E-02**	0.47	4.15E-01	0.91	**4.79E-02**	0.86
ADH5	7	**1.61E-02**	0.47	1.28E-01	0.84	1.22E-01	0.86
MGST2	22	**4.38E-02**	0.70	1.61E-01	0.84	4.98E-01	0.94
NQO1	5	8.42E-02	0.70	9.25E-01	0.96	**9.10E-03**	0.54
UGT1A5	35	8.57E-02	0.70	**2.27E-02**	0.66	9.99E-01	1.00
UGT1A6	48	1.17E-01	0.70	**3.25E-02**	0.66	9.50E-01	1.00
GSTO2	4	1.25E-01	0.70	6.57E-01	0.91	**3.62E-02**	0.86
UGT1A7	55	1.30E-01	0.70	**3.65E-02**	0.66	8.44E-01	1.00
UGT1A9	60	1.54E-01	0.70	**4.35E-02**	0.66	8.48E-01	1.00
UGT1A10	65	1.65E-01	0.70	**4.52E-02**	0.66	8.71E-01	1.00
UGT1A8	68	1.73E-01	0.70	**4.65E-02**	0.66	8.94E-01	1.00
CYP2F1	5	2.84E-01	0.78	**2.77E-02**	0.66	**4.84E-01**	0.94
ALDH1A3	3	5.13E-01	0.88	1.14E-01	0.84	**2.10E-02**	0.83

Les p-values <0,05 dans l'analyse poolée ou dans chacune des études séparément sont en gras

Tableau 25: p-value d'association entre les pathways et le risque de CDT dans l'analyse poolée et dans les études CATHY/ Young-Thyr et USRT/UTMDACC

Pathways étudiés	**Analyse poolée**	**URST/UTMDACC**	**CATHY/YOUNGTHYR**
Métabolisme des xénobiotique	0.33	0.35	0.55
Biosynthèse et métabolisme des hormones stéroïdiennes	0.43	0.67	0.96

ii. Interaction gène-environnement

Nous avons testé l'interaction avec l'IMC (modélisé en <25 ; entre 25 et 29,9 et ≥30) ; la consommation d'alcool (classée en jamais/déjà pris) et le statut tabagique (classé en non-fumeur/ancien fumeur/fumeur actuel).

Interaction avec l'IMC

Parmi les 1031 SNP analysés, l'interaction avec l'IMC était significative au seuil 5% pour 87 SNP ; cependant, aucune de ces interactions ne persistait après correction sur les tests multiples.

Au niveau des gènes, nous avons observé des interactions significatives après prise en compte des tests multiples pour les gènes *UGT2A2, UGT2A1, UGT1A6, UGT1A10, UGT1A7, UGT1A8 et UGT1A9* (p<0.3 ; Tableau 26).

Tableau 26 : P-value d'interaction avec l'IMC, la consommation d'alcool et le statut tabagique au niveau des gènes dans l'analyse poolée des études CATHY/Young-Thyr et USRT/UTMDACC.

		IMC		Consommation d'alcool		Statut tabagique	
Gene	**N.SNP**	**p-value**	**p-FDR**	**p-value**	**p-FDR**	**p-value**	**p-FDR**
UGT2A2 *(1)*	25	**1.00E-04**	**0.01**	9.16E-01	1.00	9.94E-01	1.00
UGT2A1 *(1)*	30	**2.00E-04**	**0.01**	9.54E-01	1.00	9.82E-01	1.00
UGT1A6 *(2)*	48	**3.00E-04**	**0.01**	5.24E-01	0.96	1.00E+00	1.00
UGT1A10 *(2)*	65	**9.00E-04**	**0.02**	6.72E-01	0.99	9.96E-01	1.00
UGT1A7 *(2)*	55	**9.00E-04**	**0.02**	6.27E-01	0.96	9.99E-01	1.00
UGT1A8 *(2)*	68	**1.00E-03**	**0.02**	4.74E-01	0.91	9.98E-01	1.00
UGT1A9 *(2)*	60	**1.00E-03**	**0.02**	6.64E-01	0.99	9.97E-01	1.00
AKR1C2	16	1.15E-02	0.12	**5.00E-04**	**0.03**	6.21E-01	1.00
COMT	12	6.78E-01	0.95	**3.00E-04**	**0.03**	5.37E-01	1.00
CYP11B2	5	6.29E-01	0.95	7.52E-02	0.87	**1.00E-04**	**3.93E-03**
CYP11B1	6	8.33E-01	0.95	1.15E-01	0.87	**1.00E-04**	**3.93E-03**
SRD5A2	7	6.98E-01	0.95	9.22E-01	1.00	**1.00E-04**	**3.93E-03**

Seuls les gènes pour lesquels on observe une interaction significative après correction sur les tests multiple sont présentés.*(1) gènes situé sur le locus 4q13 et tagué par les même SNP ; (2) gènes situés sur le locus 2q37 et tagué par les mêmes SNP.*

Les gènes *UGT1A6, UGT1A10, UGT1A7, UGT1A8 et UGT1A9* font partie du même locus (2q37) et sont en partie tagué par les mêmes SNP par conséquent, les résultats observés pour ces gènes sont corrélés et représentent le même signal d'association. De même, *UGT2A1* et *UGT2A2* sont situé sur le même locus (4q13) et sont en partie tagué par les mêmes. Pour éviter une surestimation de l'effet de ces gènes au niveau des pathways, nous avons considéré dans les analyses par pathway, le gène *UGT1A8* uniquement qui couvre l'ensemble des tags SNP de la région *UGT1A* couverte par les gènes *UGT1A6, UGT1A10, UGT1A7, UGT1A8 et UGT1A9* et le gène *UGT2A1* qui couvrent tous les tags SNP de la région couverte par *UGT2A1* et *UGT2A*.

Nous avons observé des interactions significatives avec les 2 pathways étudiés (Tableau *27*). Les gènes ayant contribués à ces interactions sont : *UGT2A1, UGT1A8*, *UGT2B7, AKR1C2 et CYP2D6* qui sont des gènes en commun entre les deux pathways.

Tableau 27 : Résultats des p-values d'interaction avec l'IMC, le tabac et l'alcool dans l'analyse poolée au niveau pathway

		Interaction		
Pathways	**Effet propre**	**IMC**	**Alcool**	**Tabac**
Métabolisme des xénobiotique	0,33	**2E-03**	**3E-04**	0,70
Biosynthèse et métabolisme des hormones stéroïdiennes	0,44	**3E-03**	**3E-03**	**2E-04**

Interaction avec l'alcool

Aucun des 57 SNP pour lesquels l'interaction avec la consommation d'alcool était significative au seuil 5% ne persistaient après correction sur les tests multiples.

Au niveau des gènes, seuls les interactions avec *AKR1C2* (qui contient 16 SNP) et *COMT* (contenant 12 SNP) persistaient après prise en compte des tests multiples (Tableau 26 ; p-FDR=0,03). Ces 2 gènes appartiennent aux 2 pathways étudiés, pour lesquels nous avons également observé des interactions significatives.

Interaction avec le tabac

58 SNP sur 1031 SNP analysés avaient des p-values <0,05 mais aucune des interactions ne persistaient après la correction sur les tests multiples.

Au niveau des gènes, nous avons observé des interactions significatives avec les gènes *CYP11B2*, *CYP11B1* et *SRD5A2* (p=0.01). Ces gènes qui appartiennent au pathway de la « biosynthèse et du métabolisme des hormones stéroïdiennes » contribuent à l'association significative observée pour ce pathway.

2.4. Synthèse des résultats et comparaison avec la littérature

Nous avons analysé les associations entre les CDT et les SNP et gènes impliqués dans le « métabolisme des xénobiotiques » et dans la « biosynthèse et le métabolisme des hormones stéroïdiennes ». Les analyses ont été conduites dans l'étude CATHY/Young-Thyr avec une bonne couverture des pathways d'intérêt par la puce Beluhca (1280 SNP, 138 gènes), puis en combinant les données des études CATHY/Young-Thyr et USRT/UTMDACC sur un sous-ensemble de ces SNP (1031 SNP, 118 gènes).

Nos résultats sur l'étude CATHY/Young-Thyr ne montrent pas d'effet propre des SNP ou des gènes dans les 2 pathways d'intérêt sur le risque de CDT. Les interactions avec l'IMC, la consommation d'alcool, la consommation de tabac, la prise de contraceptifs oraux et la parité ont été étudiées. Nos résultats montrent une interaction significative entre le pathway du métabolisme des xénobiotiques et la consommation de tabac, en particulier avec les gènes *GSTA5* et *CYP2F1*. Nous avons également mis en évidence des interactions entre le pathway des hormones stéroïdiennes et la consommation de tabac, la prise de contraceptifs oraux ou l'âge aux premières règles. Les gènes qui contribuaient le plus à ces interactions sont *SRD5A2*, *CGA* et *APOC3* pour la consommation de tabac ; *LHB*, *UGT2B11* et *AKR1D1* pour la prise de contraceptifs oraux ; et *CYP11A1* et *HSD17B8* pour l'âge aux 1ères règles. Dans ce pathway, nous avons également noté une interaction significative entre le gène *CYP24A1* et la consommation d'alcool.

Les résultats des analyses portant sur l'ensemble des études CATHY/Young-Thyr et USRT/UTMDACC ne montrent pas d'effet propre des gènes dans les pathways d'intérêt. Les analyses d'interactions ont été conduites avec l'IMC, la consommation d'alcool (buveur/non buveur) et le statut tabagique. Des interactions significatives ont été rapportées entre les 2 pathways d'intérêt et l'IMC ou la consommation d'alcool. Les gènes qui contribuaient le plus à ces interactions appartiennent aux 2 pathways. Il s'agit des gènes des loci *UGT1A* (2q37) et *UGT2A* (4q13) pour l'interaction avec l'IMC ; *COMT* et *AKR1C2* pour l'interaction avec l'alcool.

Nous avons également rapporté une interaction entre le statut tabagique et le pathway des hormones stéroïdiennes et plus particulièrement pour les gènes *CYP11B2*, *CYP11B1* et *SRD5A2*.

Nous notons des différences dans les résultats sur les 2 groupes d'études. En effet, dans l'analyse poolée, le nombre de SNP est plus faible que dans l'analyse de CATHY/Young-Thyr. De plus, les données sur la quantité de tabac et/ou d'alcool consommée sur la vie entière n'étaient pas disponibles au moment des analyses. Les variables utilisées pour la consommation de tabac et la consommation d'alcool sont donc définies de façon plus grossière que dans l'analyse portant sur l'étude CATHY/Young-Thyr seule.

Nous discuterons les résultats pour chacun des pathways séparément.

i. Pathway du métabolisme des xénobiotiques

Plusieurs études se sont intéressées au rôle des polymorphismes des gènes du métabolisme des xénobiotiques dans le CDT. Les polymorphismes des gènes *CY1PA1* (rs1048943, rs4646903 et rs1799814) (Siraj et al. 2008; Figlioli et al. 2016; Bufalo et al. 2006; Reis et al. 2010) et *CYP2D6* (rs3892097) (Lemos et al. 2007) impliqués dans le métabolisme de phase 1 ont été associés au CDT. Dans notre étude, *CYP1A1* et *CYP2D6* sont respectivement tagués par 4 SNP (dont rs1799814) et 2 SNP (rs28371717 et rs28371725). Aucun de ces SNP n'était associé au risque de CDT dans nos analyses.

Une analyse menée sur l'étude cas-témoins USRT/UTMDACC menée aux états Unis s'est également intéressée aux polymorphismes des gènes de ces voies métaboliques dans les CDT selon une approche par pathway (Aschebrook-Kilfoy et al. 2012). Il est à noter que la définition du pathway dans cette étude est différente de la nôtre. En effet, les auteurs se sont intéressés aux gènes impliqués dans le métabolisme et la détoxification de manière générale et ont sélectionné les gènes à partir d'une revue de littérature. Dans notre étude, les gènes impliqués dans le métabolisme des xénobiotiques ont été recensés dans la base de données KEGG et complétés par des gènes identifiés dans la littérature.

Sur les 71 gènes du métabolisme de phase 1 et 2 inclus dans les données américaines, 64 gènes étaient également inclus dans les études françaises. Les données de cette étude sont celles inclus dans notre analyse poolée des études CATHY/Young-Thyr et USRT/UTMDACC.

Dans l'étude américaine, les auteurs n'ont pas mis en évidence d'association significative entre les pathways étudiés et le risque de CDT (Aschebrook-Kilfoy et al. 2012). Ils

ont cependant suggéré des associations avec les SNP des gènes *CYP8B1* et *FMO3* du métabolisme de phase 1 et le SNP rs3924194 du gène *UGT2B7* impliqué dans le métabolisme de phase 2. Seul *UGT2B7* était inclus dans la puce Belhuca et l'association avec le SNP rs3924194 avait été répliqué dans notre analyse poolée ($p=9x10^{-4}$) mais pas dans l'étude CATHY/Young-Thyr seule (p=0,10).

Interactions avec les facteurs de risque non génétiques

Interaction avec le tabac

Dans l'étude CATHY/Young-Thyr, nous avons observé une interaction entre la consommation de tabac en paquets-années et le pathway du « métabolisme des xénobiotiques » et plus particulièrement avec les gènes *GSTA5* et *CYP2F1*. Dans l'analyse poolée, seul le statut tabagique était disponible et les analyses d'interaction ne montrent pas d'association significative avec les gènes de ce pathway.

GSTA5 est un gène de la famille des glutathion S-transférase qui joue un rôle important dans la détoxification des radicaux libres issus du métabolisme de différentes substances incluant les composés du tabac. *CYP2F1* est un gène de la famille des cytochromes p450. Il est responsable du métabolisme de plusieurs composés pneumotoxiques ayant des propriétés carcinogènes comme les toxines du tabac (G. Tournel et al. 2007; Leclerc et al. 2011). *CYP2F1* est exprimé dans les tissus pulmonaires (G. Tournel et al. 2007; Leclerc et al. 2011), et est suspecté de jouer un rôle dans l'étiologie des cancers du poumon (Cruzan et al. 2013; Leclerc et al. 2011), mais aucune association entre les polymorphismes de ce gène et le risque de cancer du poumon n'a été rapportée (Gilles Tournel et al. 2007). Les mécanismes pouvant expliquer les interactions entre les gènes *GSTA5* et *CYP2F1* et le tabac restent à élucider.

D'autres études sur les CDT ont rapporté des interactions entre la consommation de tabac et les gènes du métabolisme et la détoxification notamment *GSTM1* (Canbay et al. 2003) et *CYP26B7* (Aschebrook-Kilfoy et al. 2012). Ces gènes n'étaient pas inclus dans notre étude.

Interaction avec l'IMC

Dans l'analyse poolée des études CATHY/Young-Thyr et USRT/UTMDACC, nous avons observé une interaction significative pour les gènes des loci *UGT2A* et *UGT1A (p-FDR=0.01)*.

Seule l'interaction avec le locus UGT1A était à la limite de la significativité dans l'étude CATHY/Young-Thyr (p=0,02). Les gènes *UGT1A* et *UGT2A* font partie de la famille des UDP glucuronosyltransférases et sont impliqués dans la phase 2 du métabolisme des composés endogènes, incluant les métabolites des œstrogènes, et exogènes. Ils appartiennent aux deux pathways étudiés (« métabolisme des xénobiotiques » et « biostynthèse et métabolisme des hormones stéroïdiennes ») d'où les interactions observées avec ces pathways.

Une relation entre le niveau d'expression des gènes de la famille des UGT dans les cellules juvéniles et l'obésité a été suggérée dans la littérature (Miyauchi et al. 2016). Les polymorphismes des gènes UGT1A et UGT2A ont cependant été très peu étudiés dans les CDT. De plus, notre étude est la première à avoir exploré l'interaction entre l'IMC et les gènes de ce pathways dans les CDT.

Interaction avec l'alcool

Nous avons observé une interaction significative avec le gène *COMT* et *AKR1C2* dans l'analyse poolée (p-FDR=0.03 pour les 2 gènes). Dans l'étude CATHY/Young-Thyr, seul le gène *COMT* avait montré une interaction à la limite de la significativité avec la consommation d'alcool (p=0,04).

COMT code pour la catechol O-methyltransférase, une enzyme impliquée dans la phase 2 du métabolisme des catécholamines comme la dopamine, l'épinéphrine et le catéchol-œstrogène (Männistö & Kaakkola 1999). Certains polymorphismes de *COMT* ont été associés avec le risque de cancer du sein (Williams & Phillips 2000), de schizophrénie (Zai et al. 2017), de dépendance à l'alcool (Klimkiewicz et al. 2016) et d'élévation de la pression sanguine chez les grands consommateurs d'alcool (Stewart et al. 2009).

AKR1C2 est un gène de la superfamille aldo/céto reductase, qui comprend plus de 40 enzymes et protéines connues. Ces enzymes catalysent la conversion d'aldéhydes et de cétones en leurs alcools correspondants en utilisant NADH et / ou NADPH comme cofacteurs ; ce gène est également impliqué dans la régulation et la biosynthèse des hormones stéroïdiennes (Penning et al. 2000). Les polymorphismes dans ces gènes ont été associés au risque de cancers du sein et de la prostate (Penning & Byrns 2009).

L'interaction entre la consommation d'alcool et les gènes *COMT* et *AKR1C2* dans les CDT n'a pas été rapporté dans la littérature à notre connaissance. Des interactions avec la consommation d'alcool ont en revanche été rapportés pour les gènes *UGT2B7* et *NAT1* (Aschebrook-Kilfoy et al. 2012). Ces interactions n'ont pas été retrouvées dans notre étude.

ii. Pathway de la biosynthèse et du métabolisme des hormones stéroïdiennes

Les polymorphismes des gènes impliqués dans la biosynthèse et le métabolisme des hormones stéroïdiennes ont été associés au risque de cancers du sein et de l'endomètre, dans lesquels les œstrogènes jouent un rôle majeur (Henderson & Feigelson 2000; Low et al. 2010; Mitrunen & Hirvonen 2003). Nous avons fait l'hypothèse que ces polymorphismes modifient également le risque de CDT. Dans une étude sur le rôle des polymorphismes du gène *ESR1* dans le risque de CDT une association avec rs2228480 a été mise en évidence (Rebaï et al. 2009). Ce SNP n'est cependant pas associé au risque de CDT dans nos données.

Le gène *CYP1A1* joue également un rôle essentiel dans le métabolisme des œstrogènes en assurant la conversion de l'œstradiol en hydroxy-oetradiol puis en semiquinone, métabolites actifs ayant des propriétés carcinogènes. Comme mentionné plus haut, les polymorphismes du gène *CYP1A1* ont été associés au risque de CDT dans plusieurs études (Siraj et al. 2008; Figlioli et al. 2016; Bufalo et al. 2006; Reis et al. 2010) ; ces résultats n'ont cependant pas été confirmés dans notre étude.

Les gènes impliqués dans la biosynthèse, le métabolisme et la régulation des hormones sexuelles féminines ont également été analysés précédemment dans l'étude USRT/UTMDACC avec une approche par pathway (Schonfeld et al. 2010). Cette étude incluait 58 gènes impliqués dans la biosynthèse et le métabolisme des hormones stéroïdiennes ; le développement des cellules germinales et des gonadotrophines et la synthèse de la prolactine. La sélection des gènes inclus dans cette étude, était basée sur une revue de littérature (Schonfeld et al. 2010).

Parmi les 58 gènes inclus dans cette précédente étude, 38 gènes étaient en commun avec l'étude CATHY/Young-Thyr, ces données ont fait l'objet de l'analyse poolée des études CATHY/Young-Thyr et USRT/UTMDACC.

Les auteurs de l'étude USRT/UTMDACC n'avaient pas mis en évidences d'association significative entre les gènes impliqués dans le métabolisme et la régulation des hormones sexuelles féminines et le risque de CDT, ils ont en revanche rapporté des associations limites significatives avec les SNP des gènes *CYP19A1, ESR1, HSD17B3* et *SULF1* (Schonfeld et al. 2010). Parmi ces gènes, seuls *CYP19A1* (tagué par 34 SNP) et *ESR1* (tagué par 54 SNP) sont inclus dans notre étude et ; seul le SNP rs2445762 (p=0,02) dans le gène *CYP19A1* était associé au risque de CDT au seuil 5% ; aucune association n'a été observée avec les SNP du gène *ESR1*.

Interactions avec les facteurs de risque non génétiques

Interaction avec l'alcool

Parmi les gènes du métabolisme des hormones stéroïdiennes, nous avons rapporté une interaction entre la consommation d'alcool et le gène *CYP24A1* dans l'étude CATHY/Young-Thyr (p-FDR=0.02) ; ce gène n'était cependant pas inclus dans l'analyse poolée des études USRT/MDACC et CATHY/Young-thyr.

CYP24A1 fait partir de la famille des Cytochrome P450 responsable du métabolisme de phase 1 ; il est particulièrement impliqué dans le métabolisme et la régulation la vitamine D (Omdahl et al. 2002). Des haplotypes (rs927650C/rs2296241G et rs927650C/rs2248137C/rs2296241G) dans ce gène ont été associés au risque de CDT (Penna-Martinez et al. 2012). Parmi les SNP de ces haplotypes, les SNP rs2296241 et rs927650 ont été analysés dans notre étude et l'interaction avec l'alcool était significative au seuil 5% pour le SNP rs2296241 ($p=9X10^{-4}$).

Une étude a également suggéré que *CYP24A1* semble jouer un rôle important dans la pathogénèse des cancer de la thyroïde par le biais de la régulation et l'inactivation du $1.25(OH)_2D3$, le composé actif de la vitamine D qui possède des propriétés anti-tumorales (Clinckspoor et al. 2013). Les polymorphismes de ce gène ont été associés au risque de cancer colorectal (Dou et al. 2016) mais leurs interactions avec l'alcool n'ont pas été étudiées. Le mécanisme par lequel ce gène interagirait avec l'alcool dans les CDT reste inconnu ; d'autres études sont nécessaires pour confirmer nos résultats.

Interaction avec la consommation de tabac

Parmi les gènes du métabolisme des hormones stéroïdiennes, nous avons mis en évidence des interactions entre le statut tabagique et le gène *SRD5A2* à la fois dans l'analyse poolée et dans l'étude CATHY/Young-thyr. Nous avons également rapporté des interactions avec les gènes *CYP11B1* et *CYP11B2* (p-FDR=0.001) dans l'analyse poolée mais pas dans l'étude CATHY/Young-Thyr.

SRD5A2 est un récepteur androgénique impliqué dans la conversion des testostérones en un métabolite plus actif ; la di-hydro testostérone. Les polymorphismes dans ce gène ont été associés avec consistance au risque de cancer de la prostate (Henderson & Feigelson 2000; Neuhouser et al. 2004). Cependant le rôle de ce gène dans les CDT n'est pas connu.

CYP11B1 et *CYP11B2* sont des gènes de la famille des cytochromes P450 responsables de la synthèse du cortisol et de l'aldostérone respectivement, hormones précurseurs des stéroïdes (Pilon et al. 1999). Des polymorphismes dans ces gènes ont été associés à plusieurs cancers incluant les cancers du sein (Fan et al. 2016; Mitrunen & Hirvonen 2003). Les interactions entre ces gènes et la consommation de tabac n'ont cependant pas été décrites dans la littérature. Notre étude est également la seule étude ayant exploré les interactions avec les polymorphismes des gènes impliqués dans la biosynthèse et le métabolisme des hormones stéroïdiennes dans les CDT.

Interaction avec les facteurs hormonaux et reproductifs

Les interactions avec les facteurs hormonaux et reproductifs n'ont pu être analysées que dans l'étude CATHY/Young-Thyr. Nous avons rapporté des interactions limites significatives avec l'âge aux premières règles, et plus spécifiquement pour les gènes *CYP11A1* et *HSD17B8*.

CYP11A1 est un gène de la famille des Cytochromes p450 essentiel dans la transformation du cholestérol en pregnénolone ; la première étape de la biosynthèse des hormones stéroïdes (Kristensen & Børresen-Dale 2000). Les mutations dans ce gène ont été associées au risque de plusieurs cancers incluant les cancers de la prostate et de l'endomètre (Fan et al. 2016).

HSD17B8 est une hydroxy stéroïde (17-beta) déshydrogénase, il intervient dans la régulation de la concentration biologique active des androgènes et œstrogènes (Rotinen et al. 2009). Les polymorphismes dans ce gène ont été associés au risque de cancers hormonaux dépendants incluant le cancer du sein (Mitrunen & Hirvonen 2003; Henderson & Feigelson 2000).

CYP11A1 et *HSD17B8* n'ont cependant pas été étudiés dans les CDT et les mécanismes pouvant expliqués les interactions avec l'âge aux premières règles ne sont pas clairs.

Force et limite de l'étude

Nous avons choisi d'étudier les gènes des pathways du « métabolisme des xénobiotiques » et de la « biosynthèse et métabolisme des hormones stéroïdiennes » avec une sélection exhaustive des gènes à partir de base de données existantes, complétée par une revue de la littérature. Le génotypage par la puce Beluhca a permis d'avoir une meilleure couverture des régions d'intérêt par génotypage que dans les études GWAS.

Cette analyse menée sur les études CATHY/Young-Thyr intégrait des données de qualité pour les études d'interaction gène-environnement mais avait une puissance limitée. Afin d'augmenter la puissance statistique, nous avons choisi d'utiliser une approche par pathway permettant de mettre en évidence une association globale au niveau des gènes et au niveau des pathways d'intérêt. Cette approche est utile pour détecter les effets combinés de polymorphismes génétiques qui sont individuellement faiblement associés à la maladie et peut fournir des hypothèses supplémentaires sur les mécanismes sous-jacents. Il existe toutefois de nombreuses façons de définir un gène à partir des SNP et de définir un pathway à partir des gènes, rendant difficile les comparaisons entre études.

Il existe également de nombreuses méthodes pour les analyses par pathway. Nous avons choisi la méthode ARTP car il s'agit de la méthode utilisée par les précédentes études sur les CDT ce qui permettait de comparer plus facilement nos résultats avec ceux de la littérature. De plus, cette méthode est facilement modifiable pour l'étendre aux analyses d'interaction gène-environnement. On note toutefois des limites à cette méthode : elle permet d'estimer des p-values associées aux gènes et aux pathways , mais ne permet pas de quantifier la force de l'association ; les résultats dépendent du nombre de SNP génotypés par

gène et du nombre de gènes constituant le pathway ; la puissance diminue avec le nombre de gènes considérés dans l'analyse ; elle ne prend pas en compte la corrélation entre SNP ce qui peut entrainer une surestimation ou sous-estimation de l'effet du gène si plusieurs SNP corrélés sont pris en compte dans l'analyse au niveau du gène. Un autre désavantage de la méthode ARTP est le temps nécessaire pour réaliser les analyses, en particulier dans le traitement de grandes bases de données car la méthode utilise un algorithme de permutation (ou de bootstrap pour les interactions gène-environnement) qui peut se révéler demandeur en temps de calcul. Il serait utile dans un second temps, de comparer les résultats que nous avons obtenus avec ceux donnés par d'autres méthodes telles que SKAT par exemple (Ionita-Laza et al. 2013).

Afin d'augmenter la puissance de notre étude, nous avons combiné nos données avec celles des études USRT/UTMDACC. Il s'agit ainsi de l'étude la plus importante en termes d'effectif pour l'analyse de ces pathways candidats dans le risque de CDT. Le fait de pooler nos données a toutefois des limites. Les deux groupes d'études n'ayant pas génotypé les mêmes SNP, nous avons dû imputer les SNP manquants dans l'étude USRT/MDACC et présente dans la puce Belhuca. Tous les SNP n'ont pas pu être imputés et nous avons de ce fait travaillé sur un sous ensemble de SNP par rapport aux SNP définis dans Beluhca. Il est également possible d'imputer l'ensemble des SNP génotypés par les deux études afin d'augmenter la couverture des gènes et des pathways d'intérêt. Cette analyse pourrait être envisagée dans un second temps.

De plus, dans l'étude USRT/UTMDACC nous ne disposions pas encore des données épidémiologiques sur les facteurs hormonaux et reproductifs et les quantités d'alcool et/ou de tabac consommés sur la vie entière, ce qui a limité les analyses d'interaction gène-environnement.

Conclusion

Globalement, nous n'avons pas montré d'association entre les CDT et les SNP, gènes ou pathways du « métabolisme des xénobiotiques » et de la « biosynthèse et métabolisme des hormones stéroïdiennes ». Nous avons cependant rapporté des interactions entre les gènes du métabolisme des xénobiotiques et l'IMC, la consommation d'alcool et la consommation de tabac. Nous avons également rapporté des interactions entre les gènes de la biosynthèse et

du métabolisme des hormones stéroïdiennes, et la consommation de tabac, la prise de contraceptifs oraux et l'âge aux premières règles. Des analyses complémentaires sur nos données telles que les analyses d'interaction avec les facteurs hormonaux et reproductifs, la consommation d'alcool et de tabac moyenne sur la vie entière dans l'étude poolée doivent être effectuées. Par ailleurs, nos résultats doivent être confirmés par des études indépendantes.

Conclusion générale et perspectives

Conclusion générale

Cette thèse nous a permis d'approfondir les connaissances sur le rôle des facteurs de risque génétiques dans les CDT et leurs interactions avec les facteurs de risque environnementaux ou liés au mode de vie.

Dans une première partie nous avons étudié le rôle de gènes candidats ou de loci GWAS dans deux populations très contrastées, une population mélanésienne caractérisée par une forte incidence de CDT et une population européenne. Nous avons également étudié en détail les loci GWAS 9q22 et 14q13. Dans une deuxième partie, plus exploratoire, l'objectif était d'identifier de nouveaux facteurs de risque génétiques par une approche pathway candidat dans une population de femmes d'origine européenne.

Dans la première partie, nous n'avons pas observé d'associations entre les polymorphismes de gènes candidats et les CDT dans les deux populations. En revanche nous avons montré que la délétion des gènes *GSTT1* et *GSTM1* pouvait moduler l'association connue entre l'obésité et les CDT ainsi que l'association entre la consommation d'alcool et les CDT. Les polymorphismes identifiés par les études GWAS ont été répliqués dans la population européenne et dans une moindre mesure dans la population mélanésienne.

Dans le locus 9q22, la structure génétique comparable chez les européens et chez les mélanésiens, nous a permis d'identifier dans les deux populations un variant candidat à risque de CDT rs10759944 ayant des fonctions prédites intéressantes. Dans le locus 14q13, les déséquilibres de liaisons dans les deux populations sont très différentes. Dans la population européenne, nous avons identifié un haplotype associé au risque de CDT formé par les SNP rs944289, rs116909374 et rs999460. Dans la population mélanésienne, nous avons mis en évidence de nouveaux SNP associés au risque de CDT dont le SNP rs1755774 qui a montré le plus fort signal, et les SNP rs2787423 et rs67263215 pour lesquels nous avons observé des fonctions prédites intéressantes.

Dans la deuxième partie de cette thèse, nous nous sommes intéressés au rôle des gènes impliqués dans le « métabolisme des xénobiotiques » et dans la « biosynthèse et métabolisme

des hormones stéroïdiennes » dans les CDT. Nous n'avons pas rapporté d'association entre les SNP et gènes de ces pathways et les CDT. En revanche, nous avons observé des interactions avec les facteurs environnementaux ou liés au mode de vie. En effet, nous avons rapportés des interactions entre :

- L'IMC et les gènes de la famille des UDP-glucuronotransférases (UGT) responsable du métabolisme de phase 2
- La consommation de tabac et le gène *GSTA5* également impliqué dans le métabolisme de phase 2
- La consommation d'alcool et le gène *CYP24A1* impliqué dans le métabolisme de phase 1.

Nous avons également rapporté des interactions entre la consommation de tabac et le gène *SRD5A1* impliqué dans la synthèse des œstrogènes, ainsi qu'entre l'âge aux premières règles et les gènes *CYP11A1,* impliqué dans la synthèse des androgènes et *HSD17B8,* impliqué dans le métabolisme des œstrogènes.

Perspectives

A l'issue de cette thèse, de nouvelles questions de recherche et de mesures de santé publique peuvent être formulées.

La réplication des résultats de notre étude dans d'autres populations possédant des échantillons de plus grande taille permettrait de valider le rôle des facteurs de risque identifiés dans les CDT. Nous avons identifié de nouveaux variants associés au risque de CDT sur les loci 9q22 et 14q13 qui doivent être répliqués dans d'autres populations comme les populations africaines, polynésiennes ou asiatiques sur lesquels ces variants n'ont pas été étudiés. De même les interactions gène-environnement rapportées dans notre étude de manière générale, doivent être confirmées dans d'autres études.

Des études fonctionnelles visant à caractériser les régions contenant les variants identifiés dans les loci 9q22 et 14q13 peuvent être envisagées. L'utilisation des bases de données bio-informatiques disponibles nous a permis de prédire que ces polymorphismes sont situés :

- Soit sur des sites de fixation de facteurs de transcription ou de protéines régulatrices jouant un rôle important dans la transcription et la régulation des gènes et potentiellement impliqués dans la pathogénèse des CDT
- Soit dans des zones contenant des lincRNA exprimés dans les tissues thyroïdiens.

Il serait intéressant de conduire des études expérimentales pour voir si l'altération de ces sites de fixation entraine une modification de la fonction des FTs ou protéines régulatrices dans les tissus sains ou tumoraux de la thyroïde. Il serait également utile de caractériser, dans les tissus normaux et tumoraux de la thyroïde, les LincRNA non caractérisés observés dans le locus 14q13 et de voir si, ce lincRNA pourrait être impliqué dans la pathogénèse de la thyroïde.

L'application d'autres méthodes d'analyses par pathway permettant notamment de quantifier les effets propres au niveau des gènes et des pathways et de mesurer le sens des associations, permettrait de mieux comprendre le rôle des gènes de pathway candidats dans l'étiologie des CDT. De même, pour l'étude des interactions gène-environnement, des méthodes permettant de quantifier le sens des associations pourraient aider à une meilleure interprétation des résultats.

Les résultats des études génétiques laissent penser qu'il reste encore de nombreux facteurs génétiques à découvrir dans les CDT. Dans cette optique, un GWAS multi-ethnique est en cours de réalisation dans l'équipe et inclut au total 7 études cas-témoins (2050 cas/2410 témoins) avec des informations détaillées sur les facteurs environnementaux et liés au mode de vie. Cette étude porte sur 4 populations différentes : européennes, mélanésiennes, polynésiennes et cubaines avec des taux d'incidence très variés de CDT. La richesse des informations disponibles sur les expositions environnementales dans les différentes études permettrait d'étudier de façon approfondie les interactions gènes-environnement.

Ce GWAS multi-ethnique permettra d'avoir des données originales sur des populations peu étudiées d'un point de vue génétique. Les analyses sur ces différents groupes ethniques constituent des analyses de réplication indépendantes et pourrait faciliter la mise en évidence de variants causaux dont les signaux persisteraient d'une population à l'autre malgré les différences de structure de déséquilibre de liaison entre population. De nouveaux challenges sont posés par les GWAS multi-ethniques comme par exemple le fait de prendre en compte l'hétérogénéité génétique des populations étudiées dans une analyse globale. De nouvelles méthodes ont été récemment développées pour tenir compte de la proximité génétique des populations dans l'analyse (Shi & Lee 2016)

Cette étude menée sur des populations très différentes du point de vue génétique et avec des modes de vie et des expositions environnementales contrastées pourrait permettre d'identifier de nouveaux facteurs de risque génétiques dans les CDT et de mieux comprendre les différences d'incidence observées dans les différents groupes ethniques.

References

Adjadj, E., Schlumberger, M. & de Vathaire, F., 2009. Germ-line DNA polymorphisms and susceptibility to differentiated thyroid cancer. *The Lancet Oncology*, 10(2), pp.181–190. Available at: http://dx.doi.org/10.1016/S1470-2045(09)70020-8.

Akulevich, N.M. et al., 2009. Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma. *Endocrine-related cancer*, 16(2), pp.491–503. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19286843 [Accessed July 1, 2014].

Anon, Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on C. Available at: http://globocan.iarc.fr/Pages/references.aspx [Accessed October 30, 2014].

Aschebrook-Kilfoy, B. et al., 2012. Common genetic variants in metabolism and detoxification pathways and the risk of papillary thyroid cancer. *Endocrine-related cancer*, 19(3), pp.333–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3394851&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Austen, M., Lüscher, B. & Lüscher-Firzlaff, J.M., 1997. Characterization of the transcriptional regulator YY1. The bipartite transactivation domain is independent of interaction with the TATA box-binding protein, transcription factor IIB, TAFII55, or cAMP-responsive element-binding protein (CPB)-binding protei. *The Journal of biological chemistry*, 272(3), pp.1709–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8999850 [Accessed October 20, 2015].

Avbelj, M. et al., 2007. High prevalence of thyroid peroxidase gene mutations in patients with thyroid dyshormonogenesis. *European journal of endocrinology / European Federation of Endocrine Societies*, 156(5), pp.511–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17468186 [Accessed October 12, 2016].

Balhara, Y.P.S. & Deb, K.S., 2013. Impact of alcohol use on thyroid function. *Indian journal of endocrinology and metabolism*, 17(4), pp.580–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3743356&tool=pmcentrez&rendertype=abstract [Accessed April 10, 2016].

Bastos, H.N. et al., 2009. Association of polymorphisms in genes of the homologous recombination DNA repair pathway and thyroid cancer risk. *Thyroid : official journal of the American Thyroid Association*, 19(10), pp.1067–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19772428 [Accessed October 7, 2016].

Belin, R.M. et al., 2004. Smoke exposure is associated with a lower prevalence of serum thyroid autoantibodies and thyrotropin concentration elevation and a higher prevalence of mild thyrotropin concentration suppression in the third National Health and Nutrition Examination Surve. *The Journal of clinical endocrinology and metabolism*, 89(12), pp.6077–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15579761 [Accessed September 10, 2015].

Bignell, G.R. et al., 1997. Familial nontoxic multinodular thyroid goiter locus maps to chromosome 14q but does not account for familial nonmedullary thyroid cancer. *American journal of human genetics*, 61(5), pp.1123–30. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1716029&tool=pmcentrez&rendertype=abstract [Accessed August 5, 2015].

Binder-Foucard, F. et al., 2014. Cancer incidence and mortality in France over the 1980–2012 period: Solid tumors. *Revue d'Épidémiologie et de Santé Publique*, 62(2), pp.95–108.

Bonora, E. et al., 2013. The FOXE1 locus is a major genetic determinant for familial nonmedullary thyroid carcinoma. *International journal of cancer. Journal international du cancer*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24127282 [Accessed December 24, 2013].

Boyle, A.P. et al., 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*, 22(9), pp.1790–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3431494&tool=pmcentrez&rendertype=abstract [Accessed December 30, 2014].

Brenner, A. V et al., 2013. Common single nucleotide polymorphisms in genes related to immune function and risk of papillary thyroid cancer. *PloS one*, 8(3), p.e57243. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3592848&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Brindel, P. et al., 2009. Anthropometric factors in differentiated thyroid cancer in French Polynesia: a case-control study. *Cancer causes & control : CCC*, 20(5), pp.581–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19043789 [Accessed October 31, 2016].

Brindel, P. et al., 2008. Menstrual and reproductive factors in the risk of differentiated thyroid carcinoma in native women in French Polynesia: a population-based case-control study. *American journal of epidemiology*, 167(2), pp.219–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17965111 [Accessed October 31, 2016].

Bufalo, N.E. et al., 2006. Smoking and susceptibility to thyroid cancer: an inverse association with CYP1A1 allelic variants. *Endocrine-related cancer*, 13(4), pp.1185–93. Available at: http://erc.endocrinology-journals.org/content/13/4/1185.full [Accessed July 24, 2015].

Butler, M.W. et al., 2011. Glutathione S-transferase copy number variation alters lung gene expression. *European Respiratory Journal*, 38(1), pp.15–28.

Bůžková, P., Lumley, T. & Rice, K., 2011. Permutation and Parametric Bootstrap Tests for Gene-Gene and Gene-Environment Interactions. *Annals of Human Genetics*, 75(1), pp.36–45. Available at: http://doi.wiley.com/10.1111/j.1469-1809.2010.00572.x [Accessed December 21, 2016].

Caini, S. et al., 2015. Menstrual and reproductive history and use of exogenous sex hormones and risk of thyroid cancer among women: a meta-analysis of prospective studies. *Cancer Causes & Control*, 26(4), pp.511–518. Available at: http://link.springer.com/10.1007/s10552-015-0546-z [Accessed October 31, 2016].

Canbay, E. et al., 2003. Higher glutathione transferase GSTM1 0/0 genotype frequency in young thyroid carcinoma patients. *Current medical research and opinion*, 19(2), pp.102–106. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12740153 [Accessed July 1, 2014].

Cancemi, L. 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. *International Journal of Cancer*, 129(12), pp.2816–2824.

Canzian, F. et al., 1998. A gene predisposing to familial thyroid tumors with cell oxyphilia maps to chromosome 19p13.2. *American journal of human genetics*, 63(6), pp.1743–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1377646&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2015].

Cao, Y. et al., 2015. Reproductive Factors but Not Hormonal Factors Associated with Thyroid Cancer Risk: A Systematic Review and Meta-Analysis. *BioMed research international*, 2015, p.103515. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4538312&tool=pmcentrez&rendertype=abstract [Accessed September 8, 2015].

Cavaco, B.M. et al., 2008. Mapping a new familial thyroid epithelial neoplasia susceptibility locus to chromosome 8p23.1-p22 by high-density single-nucleotide polymorphism genome-wide linkage analysis. *The Journal of clinical endocrinology and metabolism*, 93(11), pp.4426–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18765515 [Accessed August 5, 2015].

Chen, X. et al., 2007. Learning position weight matrices from sequence and expression data. *Computational systems bioinformatics / Life Sciences Society. Computational Systems Bioinformatics Conference*, 6, pp.249–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17951829 [Accessed August 23, 2016].

Chiang, F.-Y. et al., 2008. Association between polymorphisms in DNA base excision repair genes XRCC1, APE1, and ADPRT and differentiated thyroid carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14(18), pp.5919–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18779313 [Accessed October 7, 2016].

Chiefari, E. et al., 2002. Increased expression of AP2 and Sp1 transcription factors in human thyroid tumors: a role in NIS expression regulation? *BMC cancer*, 2, p.35. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=139985&tool=pmcentrez&rendertype=abstract [Accessed May 16, 2015].

Cho, Y.A. & Kim, J., 2014. Thyroid cancer risk and smoking status: a meta-analysis. *Cancer causes & control : CCC*, 25(9), pp.1187–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24981099 [Accessed July 29, 2015].

Cipollini, M. et al., 2013. TPO genetic variants and risk of differentiated thyroid carcinoma in two European populations. *International Journal of Cancer*, 133(12), pp.2843–2851.

Cléro, E. et al., 2010. Pooled analysis of two case-control studies in New Caledonia and French Polynesia of body mass index and differentiated thyroid cancer: the importance of body surface area. *Thyroid : official journal of the American Thyroid Association*, 20(11), pp.1285–93. Available at: http://online.liebertpub.com.gate2.inist.fr/doi/abs/10.1089/thy.2009.0456 [Accessed September 9, 2014].

Clinckspoor, I. et al., 2013. Vitamin D in thyroid tumorigenesis and development. *Progress in Histochemistry and Cytochemistry*, 48(2), pp.65–98. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0079633613000120 [Accessed December 29, 2016].

Colonna, M. et al., 2015. Recent trends in incidence, geographical distribution, and survival of papillary thyroid cancer in France. *Cancer epidemiology*, 39(4), pp.511–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26003877 [Accessed October 25, 2016].

Cruzan, G. et al., 2013. Studies of styrene, styrene oxide and 4-hydroxystyrene toxicity in CYP2F2 knockout and CYP2F1 humanized mice support lack of human relevance for mouse lung tumors. *Regulatory Toxicology and Pharmacology*, 66(1), pp.24–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23454103 [Accessed January 21, 2017].

Czene, K., Lichtenstein, P. & Hemminki, K., 2002. Environmental and heritable causes of cancer

among 9.6 million individuals in the Swedish Family-Cancer Database. *International journal of cancer. Journal international du cancer*, 99(2), pp.260–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11979442 [Accessed October 30, 2014].

Derwahl, M. & Nicula, D., 2014. Estrogen and its role in thyroid cancer. *Endocrine-related cancer*, 21(5), pp.T273-83. Available at: http://erc.endocrinology-journals.org/content/21/5/T273.full [Accessed September 8, 2015].

Dou, R. et al., 2016. Vitamin D and colorectal cancer: molecular, epidemiological and clinical evidence. *British Journal of Nutrition*, 115(9), pp.1643–1660. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27245104 [Accessed December 29, 2016].

Fan, Z. et al., 2016. Association between the CYP11 family and six cancer types. *Oncology Letters*, 12(1), pp.35–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27347096 [Accessed January 10, 2017].

Fesinmeyer, M.D. et al., 2013. Genetic risk factors for BMI and obesity in an ethnically diverse population: results from the population architecture using genomics and epidemiology (PAGE) study. *Obesity (Silver Spring, Md.)*, 21(4), pp.835–46. Available at: http://doi.wiley.com/10.1002/oby.20268 [Accessed January 24, 2017].

Figlioli, G. et al., 2016. A comprehensive meta-analysis of case-control association studies to evaluate polymorphisms associated with the risk of differentiated thyroid carcinoma. *Cancer Epidemiology Biomarkers & Prevention*, 25(April), pp.1–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26843521 [Accessed October 12, 2016].

Figlioli, G. et al., 2015. Novel genetic variants in differentiated thyroid cancer and assessment of the cumulative risk. *Scientific reports*, 5, p.8922. Available at: http://www.nature.com/articles/srep08922 [Accessed December 2, 2016].

Figlioli, G. et al., 2014. Novel genome-wide association study-based candidate loci for differentiated thyroid cancer risk. *The Journal of clinical endocrinology and metabolism*, (September), p.jc20141734. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25029422 [Accessed September 8, 2014].

Franceschi, S. et al., 1999. A pooled analysis of case-control studies of thyroid cancer. IV. Benign thyroid diseases. *Cancer causes & control*, 10(6), pp.583–595. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10616827 [Accessed November 3, 2016].

Gaspar, J. et al., 2004. Combined effects of glutathione S-transferase polymorphisms and thyroid cancer risk. *Cancer genetics and cytogenetics*, 151(1), pp.60–7. Available at: http://www.sciencedirect.com/science/article/pii/S0165460803004321 [Accessed July 4, 2015].

Geisler, S.A., 2001. GSTM1, GSTT1, and the Risk of Squamous Cell Carcinoma of the Head and Neck: A Mini-HuGE Review. *American Journal of Epidemiology*, 154(2), pp.95–105. Available at: http://aje.oxfordjournals.org.gate2.inist.fr/content/154/2/95.long [Accessed August 25, 2015].

Gemignani, F. et al., 2007. Development of lung cancer before the age of 50: The role of xenobiotic metabolizing genes. *Carcinogenesis*, 28(6), pp.1287–1293. Available at: http://carcin.oxfordjournals.org/cgi/doi/10.1093/carcin/bgm021 [Accessed January 19, 2017].

Goldgar, D.E. et al., 1994. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *Journal of the National Cancer Institute*, 86(21), pp.1600–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7932824 [Accessed July 1, 2014].

Guan, Y. & Stephens, M., 2008. Practical issues in imputation-based association mapping. N. J. Schork, ed. *PLoS genetics*, 4(12), p.e1000279. Available at: http://dx.plos.org/10.1371/journal.pgen.1000279 [Accessed June 3, 2014].

Gudmundsson, J. et al., 2009. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nature genetics*, 41(4), pp.460–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3664837&tool=pmcentrez&rendertype=abstract [Accessed December 23, 2013].

Gudmundsson, J. et al., 2012. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nature genetics*, 44(3), pp.319–22. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3655412&tool=pmcentrez&rendertype=abstract [Accessed December 19, 2013].

Guignard, R. et al., 2007. Alcohol drinking, tobacco smoking, and anthropometric characteristics as risk factors for thyroid cancer: a countrywide case-control study in New Caledonia. *American journal of epidemiology*, 166(10), pp.1140–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2220030&tool=pmcentrez&rendertype=abstract [Accessed June 3, 2015].

Guilhen, A.C.T. et al., 2009. Role of the N-acetyltransferase 2 detoxification system in thyroid cancer susceptibility. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 15(1), pp.406–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19118072 [Accessed August 12, 2015].

Hayes, J.D. & Strange, R.C., 2000. Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology*, 61(3), pp.154–66. Available at: http://www.karger.com/Article/FullText/28396 [Accessed April 8, 2016].

He, C. et al., 2010. A large-scale candidate gene association study of age at menarche and age at natural menopause. *Human Genetics*, 128(5), pp.515–527. Available at: http://link.springer.com/10.1007/s00439-010-0878-4 [Accessed January 13, 2017].

He, H. et al., 2009. A Susceptibility Locus for Papillary Thyroid Carcinoma on Chromosome 8q24. *Cancer research*, 69(2), pp.625–631. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19147577 [Accessed February 6, 2014].

He, H., Li, W., Liyanarachchi, S., Jendrzejewski, J., et al., 2015. Genetic predisposition to papillary thyroid carcinoma: involvement of FOXE1, TSHR, and a novel lincRNA gene, PTCSC2. *The Journal of clinical endocrinology and metabolism*, 100(1), pp.E164-72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25303483 [Accessed October 21, 2014].

He, H., Li, W., Liyanarachchi, S., Srinivas, M., et al., 2015. Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 112(19), pp.6128–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25918370 [Accessed May 2, 2015].

Hemminki, K., Eng, C. & Chen, B., 2005. Familial risks for nonmedullary thyroid cancer. *The Journal of clinical endocrinology and metabolism*, 90(10), pp.5747–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16030170 [Accessed November 14, 2014].

Hemminki, K. & Vaittinen, P., 1999. Familial cancers in a nationwide family cancer database: age distribution and prevalence. *European Journal of Cancer*, 35(7), pp.1109–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10533456 [Accessed July 1, 2014].

Henderson, B.E. & Feigelson, H.S., 2000. Hormonal carcinogenesis. *Carcinogenesis*, 21(3), pp.427–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10688862 [Accessed December 29, 2016].

Hermann, D., Heinz, A. & Mann, K., 2002. Dysregulation of the hypothalamic-pituitary-thyroid axis in alcoholism. *Addiction*, 97(11), pp.1369–1381. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12410778 [Accessed March 9, 2016].

Hernández, A. et al., 2003. Glutathione S-transferase polymorphisms in thyroid cancer patients. *Cancer Letters*, 190(1), pp.37–44. Available at: http://www.sciencedirect.com/science/article/pii/S0304383502005803 [Accessed August 12, 2015].

Hirvonen, A., 1999. Polymorphisms of xenobiotic-metabolizing enzymes and susceptibility to cancer. *Environmental health perspectives*, 107 Suppl, pp.37–47. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1566364&tool=pmcentrez&rendertype=abstract [Accessed August 25, 2015].

Ho, T. et al., 2009. Association of XRCC1 polymorphisms and risk of differentiated thyroid carcinoma: a case-control analysis. *Thyroid : official journal of the American Thyroid Association*, 19(2), pp.129–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19191745 [Accessed June 17, 2015].

Ho, T. et al., 2006. Glutathione S-transferase polymorphisms and risk of differentiated thyroid carcinomas: a case-control analysis. *Archives of otolaryngology--head & neck surgery*, 132(7), pp.756–61. Available at: http://archotol.jamanetwork.com.gate2.inist.fr/article.aspx?articleid=484212 [Accessed August 12, 2015].

Hong, S.-H., Myung, S.-K. & Kim, H., 2016. Alcohol Intake and Risk of Thyroid Cancer: A Meta-analysis of Observational Studies. *Cancer Research and Treatment*. Available at: http://e-crt.org/journal/view.php?doi=10.4143/crt.2016.161 [Accessed November 2, 2016].

Horn-Ross, P.L. et al., 2011. Hormonal factors and the risk of papillary thyroid cancer in the California Teachers Study cohort. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 20(8), pp.1751–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3288117&tool=pmcentrez&rendertype=abstract [Accessed March 15, 2016].

Horn-Ross, P.L. et al., 2001. Iodine and thyroid cancer risk among women in a multiethnic population: the Bay Area Thyroid Cancer Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 10(9), pp.979–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11535551 [Accessed November 3, 2016].

Howie, B., Marchini, J. & Stephens, M., 2011. Genotype imputation with thousands of genomes. *G3 (Bethesda, Md.)*, 1(6), pp.457–70. Available at: http://www.g3journal.org/content/1/6/457.full [Accessed May 27, 2014].

Howie, B.N., Donnelly, P. & Marchini, J., 2009. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies N. J. Schork, ed. *PLoS Genetics*, 5(6), p.e1000529. Available at: http://dx.plos.org/10.1371/journal.pgen.1000529 [Accessed September 21, 2016].

Huang, R.S. et al., 2009. Population-specific GSTM1 copy number variation. *Human molecular genetics*, 18(2), pp.366–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2722188&tool=pmcentrez&rendertype=abstract [Accessed August 21, 2015].

Ionita-Laza, I. et al., 2013. Sequence kernel association tests for the combined effect of rare and common variants. *American journal of human genetics*, 92(6), pp.841–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23684009 [Accessed November 21, 2016].

Jendrzejewski, J. et al., 2012. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22), pp.8646–51. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3365219&tool=pmcentrez&rendertype=abstract [Accessed February 6, 2014].

Jones, A.M. et al., 2012. Thyroid cancer susceptibility polymorphisms: confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24. *Journal of medical genetics*, 49(3), pp.158–63. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3286794&tool=pmcentrez&rendertype=abstract [Accessed January 27, 2014].

Kabat, G.C. et al., 2012. Smoking and alcohol consumption in relation to risk of thyroid cancer in postmenopausal women. *Cancer epidemiology*, 36(4), pp.335–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22525339 [Accessed July 29, 2015].

Kallel, R. et al., 2010. Genetic investigation of FOXE1 polyalanine tract in thyroid diseases: New insight on the role of FOXE1 in thyroid carcinoma. *Cancer Biomarkers*, 8(1), pp.43–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21896990 [Accessed October 11, 2016].

Kilfoy, B.A. et al., 2009. International patterns and trends in thyroid cancer incidence, 1973-2002. *Cancer Causes and Control*, 20(5), pp.525–531. Available at: http://link.springer.com/10.1007/s10552-008-9260-4 [Accessed October 28, 2016].

Kitahara, C.M. et al., 2014. Childhood height and body mass index were associated with risk of adult thyroid cancer in a large cohort study. *Cancer research*, 74(1), pp.235–42. Available at: http://cancerres.aacrjournals.org/content/74/1/235.full [Accessed May 6, 2016].

Kitahara, C.M., Linet, M.S., et al., 2012. Cigarette smoking, alcohol intake, and thyroid cancer risk: a pooled analysis of five prospective studies in the United States. *Cancer causes & control : CCC*, 23(10), pp.1615–24. Available at: /pmc/articles/PMC3511822/?report=abstract [Accessed March 10, 2015].

Kitahara, C.M., Neta, G., et al., 2012. Common obesity-related genetic variants and papillary thyroid cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 21(12), pp.2268–71. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3518668&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Kitahara, C.M. et al., 2011. Obesity and thyroid cancer risk among U.S. men and women: a pooled analysis of five prospective studies. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 20(3), pp.464–72. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3079276&tool=pmcentrez&rendertype=abstract [Accessed May 6, 2016].

Kitahara, C.M. & Sosa, J.A., 2016. The changing incidence of thyroid cancer. *Nature reviews. Endocrinology*, 12(11), pp.646–653. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27418023 [Accessed October 25, 2016].

Klimkiewicz, A. et al., 2016. COMT and BDNF Gene Variants Help to Predict Alcohol Consumption in Alcohol-dependent Patients. *Journal of Addiction Medicine*, p.1. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27898499 [Accessed January 2, 2017].

Klug, J. & Beato, M., 1996. Binding of YY1 to a site overlapping a weak TATA box is essential for transcription from the uteroglobin promoter in endometrial cells. *Molecular and cellular biology*, 16(11), pp.6398–407. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=231641&tool=pmcentrez&rendertype=abstract [Accessed October 20, 2015].

Köhler, A. et al., 2013. Genome-wide association study on differentiated thyroid cancer. *The Journal of clinical endocrinology and metabolism*, 98(10), pp.E1674-81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23894154 [Accessed December 19, 2013].

Kolonel, L.N. et al., 1990. An epidemiologic study of thyroid cancer in Hawaii. *Cancer causes & control : CCC*, 1(3), pp.223–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2102295 [Accessed January 3, 2017].

Kreiger, N. & Parkes, R., 2000. Cigarette smoking and the risk of thyroid cancer. *European Journal of Cancer*, 36(15), pp.1969–1973. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11000579 [Accessed December 13, 2016].

Kristensen, V.N. & Børresen-Dale, A.L., 2000. Molecular epidemiology of breast cancer: genetic variation in steroid hormone metabolism. *Mutation Research/Reviews in Mutation Research*, 462(2–3), pp.323–333.

Kweon, S.-S. et al., 2014. Polymorphisms of methylenetetrahydrofolate reductase and glutathione S-transferase are not associated with the risk of papillary thyroid cancer in Korean population. *Molecular biology reports*, 41(6), pp.3793–9. Available at: http://link.springer.com/10.1007/s11033-014-3245-z [Accessed August 26, 2015].

Landa, I. et al., 2009. The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS genetics*, 5(9), p.e1000637. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2727793&tool=pmcentrez&rendertype=abstract [Accessed December 23, 2013].

Landa, I. & Robledo, M., 2011. Association studies in thyroid cancer susceptibility: are we on the right track? *Journal of molecular endocrinology*, 47(1).

Lao, X. et al., 2014. Glutathione S-transferase gene GSTM1, gene-gene interaction, and gastric cancer susceptibility: evidence from an updated meta-analysis. *Cancer cell international*, 14(1), p.127. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4255933&tool=pmcentrez&rendertype=abstract [Accessed July 23, 2015].

Leclerc, J. et al., 2011. Xenobiotic metabolism and disposition in human lung: Transcript profiling in non-tumoral and tumoral tissues. *Biochimie*, 93(6), pp.1012–1027.

Lemos, M.C. et al., 2008. Combined GSTM1 and GSTT1 null genotypes are associated with a lower risk of papillary thyroid cancer. *Journal of endocrinological investigation*, 31(6), pp.542–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18591888 [Accessed November 6, 2014].

Lemos, M.C. et al., 2007. Genetic polymorphism of CYP2D6 influences susceptibility to papillary thyroid cancer. *Clinical Endocrinology*, 67(2), pp.180–183. Available at: http://doi.wiley.com/10.1111/j.1365-2265.2007.02858.x [Accessed January 16, 2017].

Lesueur, F. et al., 2002. Specific haplotypes of the RET proto-oncogene are over-represented in patients with sporadic papillary thyroid carcinoma. *Journal of medical genetics*, 39(4), pp.260–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1735081&tool=pmcentrez&rendertype=abstract [Accessed December 23, 2013].

Leux, C. & Guénel, P., 2010. Risk factors of thyroid tumors: role of environmental and occupational exposures to chemical pollutants. *Revue d'épidémiologie et de santé publique*, 58(5), pp.359–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20980113 [Accessed September 9, 2014].

Li, J. et al., 2012. Glutathione S-transferase M1, T1, and P1 polymorphisms and thyroid cancer risk: a meta-analysis. *Cancer epidemiology*, 36(6), pp.e333-40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22765906 [Accessed November 6, 2014].

Li, M.-X. et al., 2012. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human genetics*, 131(5), pp.747–56. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3325408&tool=pmcentrez&rendertype=abstract [Accessed September 11, 2014].

Li, Y. et al., 2009. Genotype imputation. *Annual review of genomics and human genetics*, 10, pp.387–406. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2925172&tool=pmcentrez&rendertype=abstract [Accessed May 5, 2014].

Liyanarachchi, S. et al., 2013. Cumulative risk impact of five genetic variants associated with papillary thyroid carcinoma. *Thyroid : official journal of the American Thyroid Association*, 23(12), pp.1532–40. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3868253&tool=pmcentrez&rendertype=abstract [Accessed October 22, 2014].

Lönn, S. et al., 2007. Papillary thyroid cancer and polymorphic variants in TSHR- and RET-related genes: a nested case-control study within a cohort of U.S. radiologic technologists. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 16(1), pp.174–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17220349 [Accessed February 4, 2015].

Low, Y.L. et al., 2010. Multi-variant pathway association analysis reveals the importance of genetic determinants of estrogen metabolism in breast and endometrial cancer susceptibility K. Kerr, ed. *PLoS Genetics*, 6(7), pp.1–8. Available at: http://dx.plos.org/10.1371/journal.pgen.1001012 [Accessed January 9, 2017].

Loza, M.J. et al., 2007. Assembly of Inflammation-Related Genes for Pathway-Focused Genetic Analysis J. Gold, ed. *PLoS ONE*, 2(10), p.e1035. Available at: http://dx.plos.org/10.1371/journal.pone.0001035 [Accessed January 13, 2017].

Ma, J. et al., 2015. Obesity and risk of thyroid cancer: evidence from a meta-analysis of 21 observational studies. *Medical science monitor : international medical journal of experimental and clinical research*, 21, pp.283–91. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4315628&tool=pmcentrez&rendertype=abstract [Accessed July 29, 2015].

Maillard, S. et al., 2015. Common Variants at 9q22.33, 14q13.3, and ATM Loci, and Risk of Differentiated Thyroid Cancer in the French Polynesian Population. *PloS one*, 10(4), p.e0123700. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4388539&tool=pmcentrez&rendertype=abstract [Accessed April 10, 2015].

Malchoff, C.D. et al., 2000. Papillary thyroid carcinoma associated with papillary renal neoplasia: genetic linkage analysis of a distinct heritable tumor syndrome. *The Journal of clinical endocrinology and metabolism*, 85(5), pp.1758–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10843148 [Accessed August 5, 2015].

Mancikova, V. et al., 2015. Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations. *International journal of cancer. Journal international du cancer*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25855579 [Accessed June 11, 2015].

Männistö, P.T. & Kaakkola, S., 1999. Catechol-O-methyltransferase (COMT): Biochemistry, Molecular Biology, Pharmacology, and Clinical Efficacy of the New Selective COMT Inhibitors. *Pharmacological Reviews*, 51(4).

Manolio, T.A. et al., 2009. Finding the missing heritability of complex diseases. *Nature*, 461(7265). Available at: http://www.nature.com/doifinder/10.1038/nature08494 [Accessed January 7, 2017].

Mansoori, A.A. & Jain, S.K., 2015. Molecular Links between Alcohol and Tobacco Induced DNA Damage, Gene Polymorphisms and Patho-physiological Consequences: A Systematic Review of Hepatic Carcinogenesis. *Asian Pacific journal of cancer prevention : APJCP*, 16(12), pp.4803–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26163595 [Accessed September 1, 2015].

Marcello, M.A. et al., 2014. Obesity and thyroid cancer. *Endocrine-related cancer*, 21(5), pp.T255-71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24741026 [Accessed September 13, 2015].

Marchini, J. & Howie, B., 2010. Genotype imputation for genome-wide association studies. *Nature reviews. Genetics*, 11(7), pp.499–511. Available at: http://dx.doi.org/10.1038/nrg2796 [Accessed May 25, 2014].

McKay, J.D. et al., 2001. Localization of a susceptibility gene for familial nonmedullary thyroid carcinoma to chromosome 2q21. *American journal of human genetics*, 69(2), pp.440–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1235316&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2015].

McLellan, R.A. et al., 1997. Characterization of a human glutathione S-transferase mu cluster containing a duplicated GSTM1 gene that causes ultrarapid enzyme activity. *Molecular pharmacology*, 52(6), pp.958–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9415705 [Accessed September 7, 2015].

Meinhold, C.L. et al., 2009. Alcohol intake and risk of thyroid cancer in the NIH-AARP Diet and Health

Study. *British journal of cancer*, 101(9), pp.1630–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2778506&tool=pmcentrez&rendertype=abstract [Accessed July 29, 2015].

Meinhold, C.L. et al., 2010. Nonradiation risk factors for thyroid cancer in the US Radiologic Technologists Study. *American journal of epidemiology*, 171(2), pp.242–52. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3290908&tool=pmcentrez&rendertype=abstract [Accessed March 15, 2016].

Memon, A. et al., 2002. Epidemiology of reproductive and hormonal factors in thyroid cancer: evidence from a case-control study in the Middle East. *International journal of cancer*, 97(1), pp.82–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11774247 [Accessed January 12, 2017].

Memon, A., Varghese, A. & Suresh, A., 2002. Benign thyroid disease and dietary factors in thyroid cancer: a case–control study in Kuwait. *British Journal of Cancer*, 86(11), pp.1745–1750. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12087461 [Accessed November 3, 2016].

Mitrunen, K. & Hirvonen, A., 2003. Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. *Mutation research*, 544(1), pp.9–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12888106 [Accessed September 7, 2015].

Miyauchi, E. et al., 2016. Quantitative Atlas of Cytochrome P450, UDP-Glucuronosyltransferase, and Transporter Proteins in Jejunum of Morbidly Obese Subjects. *Molecular Pharmaceutics*, 13(8), pp.2631–2640. Available at: http://pubs.acs.org/doi/abs/10.1021/acs.molpharmaceut.6b00085 [Accessed December 30, 2016].

Mooney, M.A. & Wilmot, B., 2015. Gene set analysis: A step-by-step guide. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 168(7), pp.517–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26059482 [Accessed October 23, 2015].

Moore, M. a et al., 2010. Cancer epidemiology in the pacific islands - past, present and future. *Asian Pacific journal of cancer prevention : APJCP*, 11 Suppl 2, pp.99–106. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20553071.

Morari, E.C. et al., 2002. The null genotype of glutathione s-transferase M1 and T1 locus increases the risk for thyroid cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 11(11), pp.1485–8. Available at: http://cebp.aacrjournals.org/content/11/11/1485.full [Accessed August 12, 2015].

Morey, L. & Helin, K., 2010. Polycomb group protein-mediated repression of transcription. *Trends in biochemical sciences*, 35(6), pp.323–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20346678 [Accessed March 11, 2015].

Moriya, H. et al., 2014. Single-nucleotide polymorphisms and copy number variations of the FCGR2A and FCGR3A genes in healthy Japanese subjects. *Biomedical reports*, 2(2), pp.265–269. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3917743&tool=pmcentrez&rendertype=abstract [Accessed November 7, 2014].

Mota, P. et al., 2015. Genetic polymorphisms of phase I and phase II metabolic enzymes as

modulators of lung cancer susceptibility. *Journal of cancer research and clinical oncology*, 141(5), pp.851–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25388590 [Accessed September 10, 2015].

Nagano, J. et al., 2007. A case-control study in Hiroshima and Nagasaki examining non-radiation risk factors for thyroid cancer. *Journal of epidemiology*, 17(3), pp.76–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17545694 [Accessed December 13, 2016].

Negri, E. et al., 1999. A pooled analysis of case-control studies of thyroid cancer. II. Menstrual and reproductive factors. *Cancer causes & control : CCC*, 10(2), pp.143–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10231163 [Accessed May 6, 2016].

Neta, G. et al., 2011. Common genetic variants related to genomic integrity and risk of papillary thyroid cancer. *Access*, 32(8), pp.1231–1237. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3166197&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Neuhouser, M.L., Kristal, A.R. & Penson, D.F., 2004. Steroid hormones and hormone-related genetic and lifestyle characteristics as risk factors for benign prostatic hyperplasia: Review of epidemiologic literature. *Urology*, 64(2), pp.201–211. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15302461 [Accessed December 29, 2016].

Nørskov, M.S. et al., 2011. Copy number variation in glutathione-S-transferase T1 and M1 predicts incidence and 5-year survival from prostate and bladder cancer, and incidence of corpus uteri cancer in the general population. *The pharmacogenomics journal*, 11(4), pp.292–9. Available at: http://dx.doi.org/10.1038/tpj.2010.38 [Accessed March 3, 2016].

Nørskov, M.S. et al., 2009. High-throughput genotyping of copy number variation in glutathione S-transferases M1 and T1 using real-time PCR in 20,687 individuals. *Clinical biochemistry*, 42(3), pp.201–9. Available at: http://www.sciencedirect.com/science/article/pii/S0009912008005614 [Accessed March 4, 2016].

Omdahl, J.L., Morris, H.A. & May, B.K., 2002. H YDROXYLASE E NZYMES OF THE V ITAMIN D P ATHWAY : Expression, Function, and Regulation. *Annual Review of Nutrition*, 22(1), pp.139–166. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12055341 [Accessed December 29, 2016].

Onda, M. et al., 2004. Up-regulation of transcriptional factor E2F1 in papillary and anaplastic thyroid cancers. *Journal of human genetics*, 49(6), pp.312–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15118916 [Accessed May 8, 2015].

Park, S.-K. et al., 2003. Reproductive Factors, Glutathione S-Transferase M1 and T1 Genetic Polymorphism and Breast Cancer Risk. *Breast Cancer Research and Treatment*, 78(1), pp.89–96. Available at: http://link.springer.com.gate2.inist.fr/article/10.1023/A%25253A1022113713226 [Accessed April 26, 2016].

Pazaitou-Panayiotou, K., Polyzos, S.A. & Mantzoros, C.S., 2013. Obesity and thyroid cancer: epidemiologic associations and underlying mechanisms. *Obesity reviews : an official journal of the International Association for the Study of Obesity*, 14(12), pp.1006–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24034423 [Accessed March 30, 2016].

Penna-Martinez, M. et al., 2012. Impaired vitamin D activation and association with CYP24A1 haplotypes in differentiated thyroid carcinoma. *Thyroid : official journal of the American Thyroid Association*, 22(7), pp.709–16. Available at: http://online.liebertpub.com/doi/abs/10.1089/thy.2011.0330 [Accessed December 29, 2016].

Penning, T.M. et al., 2000. Human 3alpha-hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *The Biochemical journal*, 351(Pt 1), pp.67–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10998348 [Accessed January 10, 2017].

Penning, T.M. & Byrns, M.C., 2009. Steroid hormone transforming aldo-keto reductases and cancer. *Annals of the New York Academy of Sciences*, 1155, pp.33–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19250190 [Accessed January 10, 2017].

Pereda, C.M. et al., 2015. Common variants at the 9q22.33, 14q13.3 and ATM loci, and risk of differentiated thyroid cancer in the Cuban population. *BMC Genetics*, 16. Available at: http://www.biomedcentral.com/1471-2156/16/22.

Pilon, C. et al., 1999. Mutations in CYP11B1 gene converting 11??-hydroxylase into an aldosterone-producing enzyme are not present in aldosterone-producing adenomas. *Journal of Clinical Endocrinology and Metabolism*, 84(11), pp.4228–4231.

Prentice, R.L. et al., 1982. Radiation exposure and thyroid cancer incidence among Hiroshima and Nagasaki residents. *National Cancer Institute monograph*, 62, pp.207–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7167191 [Accessed October 30, 2016].

Rebaï, M. et al., 2009. Association of polymorphisms in estrogen and thyroid hormone receptors with thyroid cancer risk. *Journal of receptor and signal transduction research*, 29(2), pp.113–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19519176 [Accessed August 10, 2015].

Rebbeck, T.R., 1997. Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 6(9), pp.733–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9298582 [Accessed September 8, 2015].

Reis, A.A.S. et al., 2010. Involvement of CYP1A1, GST, 72TP53 polymorphisms in the pathogenesis of thyroid nodules. *Genetics and molecular research : GMR*, 9(4), pp.2222–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21086258 [Accessed August 12, 2015].

Ren, Y., Lence-Anta, J.J., et al., 2016. *Foxe1* Polymorphism Interacts with Dietary Iodine Intake in Differentiated Thyroid Cancer Risk in the Cuban Population. *Thyroid*, p.thy.2015.0594. Available at: http://online.liebertpub.com/doi/10.1089/thy.2015.0594 [Accessed November 2, 2016].

Ren, Y., Lence-Anta, J.J., et al., 2016. Foxe1 Polymorphism Interacts with Dietary Iodine Intake in Differentiated Thyroid Cancer Risk in the Cuban Population. *Thyroid : official journal of the American Thyroid Association*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27610545 [Accessed October 3, 2016].

Rinaldi, S. et al., 2012. Body size and risk of differentiated thyroid carcinomas: Findings from the EPIC study. *International Journal of Cancer*, 131(6), pp.E1004–E1014. Available at: http://doi.wiley.com/10.1002/ijc.27601 [Accessed October 31, 2016].

Ron, E. et al., 1995. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiation research*, 141(3), pp.259–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7871153 [Accessed October 7, 2016].

Rose-Zerilli, M.J. et al., 2009. Copy-number variation genotyping of GSTT1 and GSTM1 gene deletions by real-time PCR. *Clinical chemistry*, 55(9), pp.1680–5. Available at:

http://www.clinchem.org/content/55/9/1680.full [Accessed March 24, 2015].

Rossing, M.A. et al., 2000. Risk of papillary thyroid cancer in women in relation to smoking and alcohol consumption. *Epidemiology (Cambridge, Mass.)*, 11(1), pp.49–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10615843 [Accessed December 13, 2016].

Rotinen, M. et al., 2009. Estradiol induces type 8 17beta-hydroxysteroid dehydrogenase expression: crosstalk between estrogen receptor alpha and C/EBPbeta. *The Journal of endocrinology*, 200(1), pp.85–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18852215 [Accessed January 8, 2017].

Rudolph, A. et al., 2012. Copy number variations of GSTT1 and GSTM1, colorectal cancer risk and possible effect modification of cigarette smoking and menopausal hormone therapy. *International journal of cancer. Journal international du cancer*, 131(5), pp.E841-8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22234881 [Accessed November 7, 2014].

Saitou, M. & Ishida, T., 2015. Distributions of the GSTM1 and GSTT1 null genotypes worldwide are characterized by latitudinal clines. *Asian Pacific journal of cancer prevention : APJCP*, 16(1), pp.355–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25640380 [Accessed August 26, 2015].

Schlattl, A. et al., 2011. Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. *Genome research*, 21(12), pp.2004–13. Available at: http://genome.cshlp.org/cgi/doi/10.1101/gr.122614.111 [Accessed December 16, 2016].

Schonfeld, S.J. et al., 2012. Common genetic variants in sex hormone pathway genes and papillary thyroid cancer risk. *Thyroid : official journal of the American Thyroid Association*, 22(2), pp.151–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3271376&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Schonfeld, S.J. et al., 2010. Polymorphisms in oxidative stress and inflammation pathway genes, low-dose ionizing radiation, and the risk of breast cancer among US radiologic technologists. *Cancer causes & control : CCC*, 21(11), pp.1857–66. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3076104&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Secretan, B.L. et al., 2016. Spe ci a l R e p or t Body Fatness and Cancer — Viewpoint of the IARC Working Group.

Sen, A. et al., 2015. Baseline and lifetime alcohol consumption and risk of differentiated thyroid carcinoma in the EPIC study. *British journal of cancer*, 113(5), pp.840–7. Available at: http://www.nature.com.gate2.inist.fr/bjc/journal/vaop/ncurrent/full/bjc2015280a.html [Accessed September 1, 2015].

Shi, J. & Lee, S., 2016. A novel random effect model for GWAS meta-analysis and its application to trans-ethnic meta-analysis. , 72(3), pp.945–954.

Sigurdson, A.J. et al., 2003. Cancer incidence in the US radiologic technologists health study, 1983-1998. *Cancer*, 97(12), pp.3080–9. Available at: http://doi.wiley.com/10.1002/cncr.11444 [Accessed February 4, 2015].

Sigurdson, A.J. et al., 2009. Thyroid nodules, polymorphic variants in DNA repair and RET-related genes, and interaction with ionizing radiation exposure from nuclear tests in Kazakhstan.

Radiation research, 171(1), pp.77–88. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2875679&tool=pmcentrez&rendertype=abstract [Accessed July 1, 2014].

Simon, J.A. & Lange, C.A., 2008. Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 647(1), pp.21–29.

Singer, S. et al., 2012. Quality of Life in Patients with Thyroid Cancer Compared with the General Population. *Thyroid*, 22(2), pp.117–124. Available at: http://online.liebertpub.com/doi/abs/10.1089/thy.2011.0139 [Accessed October 26, 2016].

Siraj, A.K. et al., 2008. Polymorphisms of selected xenobiotic genes contribute to the development of papillary thyroid cancer susceptibility in Middle Eastern population. *BMC medical genetics*, 9(1), p.61. Available at: http://www.biomedcentral.com/1471-2350/9/61 [Accessed August 12, 2015].

Soldin, O.P. et al., 2009. Thyroid hormone levels associated with active and passive cigarette smoking. *Thyroid : official journal of the American Thyroid Association*, 19(8), pp.817–23. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3643222&tool=pmcentrez&rendertype=abstract [Accessed September 10, 2015].

Sprenger, R. et al., 2000. Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype-phenotype correlation. *Pharmacogenetics*, 10(6), pp.557–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10975610 [Accessed September 7, 2015].

Stankov, K. et al., 2006. GSTT1 and M1 polymorphisms in Hürthle thyroid cancer patients. *Cancer letters*, 240(1), pp.76–82. Available at: http://www.sciencedirect.com/science/article/pii/S0304383505007998 [Accessed August 13, 2015].

Stewart, S.H. et al., 2009. COMT Genotype Influences the Effect of Alcohol on Blood Pressure: Results From the COMBINE Study. *American Journal of Hypertension*, 22(1), pp.87–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19023276 [Accessed January 22, 2017].

Sturgis, E.M. et al., 2005. Radiation response genotype and risk of differentiated thyroid cancer: a case-control analysis. *The Laryngoscope*, 115(6), pp.938–945.

Suzuki, T. et al., 2008. Anthropometric factors at age 20 years and risk of thyroid cancer. *Cancer causes & control : CCC*, 19(10), pp.1233–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18618280 [Accessed October 31, 2016].

Takahashi, M. et al., 2010. The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Human molecular genetics*, 19(12), pp.2516–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20350937 [Accessed December 23, 2013].

Tian, C., Gregersen, P.K. & Seldin, M.F., 2008. Accounting for ancestry: Population substructure and genome-wide association studies. *Human Molecular Genetics*, 17(R2), pp.143–150. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18852203 [Accessed December 21, 2016].

Tournel, G. et al., 2007. CYP2F1 genetic polymorphism: Identification of interethnic variations. *Xenobiotica*, 37(12), pp.1433–1438. Available at: http://www.tandfonline.com/doi/full/10.1080/00498250701644403 [Accessed January 21, 2017].

Tournel, G. et al., 2007. Molecular analysis of the CYP2F1 gene: Identification of a frequent non-functional allelic variant. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 617(1–2), pp.79–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17327131 [Accessed January 22, 2017].

Truong, T. et al., 2010. Role of dietary iodine and cruciferous vegetables in thyroid cancer: a countrywide case-control study in New Caledonia. *Cancer causes & control : CCC*, 21(8), pp.1183–92. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3496161&tool=pmcentrez&rendertype=abstract [Accessed April 26, 2015].

Truong, T. et al., 2005. Role of goiter and of menstrual and reproductive factors in thyroid cancer: a population-based case-control study in New Caledonia (South Pacific), a very high incidence area. *American journal of epidemiology*, 161(11), pp.1056–65. Available at: http://aje.oxfordjournals.org.gate2.inist.fr/content/161/11/1056.long [Accessed September 9, 2014].

Truong, T. et al., 2007. Time trends and geographic variations for thyroid cancer in New Caledonia, a very high incidence area (1985-1999). *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)*, 16(1), pp.62–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17220706 [Accessed July 10, 2014].

Vaccarella, S. et al., 2015. The Impact of Diagnostic Changes on the Rise in Thyroid Cancer Incidence: A Population-Based Study in Selected High-Resource Countries. *Thyroid : official journal of the American Thyroid Association*, 25(10), pp.1127–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26133012 [Accessed October 25, 2016].

Valeix, P. et al., 2008. Effects of light to moderate alcohol consumption on thyroid volume and thyroid function. *Clinical endocrinology*, 68(6), pp.988–95. Available at: http://onlinelibrary.wiley.com.gate2.inist.fr/doi/10.1111/j.1365-2265.2007.03123.x/full [Accessed March 9, 2016].

Vigneri, R., Malandrino, P. & Vigneri, P., 2014. The changing epidemiology of thyroid cancer: why is incidence increasing? *Current opinion in oncology*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25310641 [Accessed October 24, 2014].

Di Vito, M. et al., 2011. Overexpression of estrogen receptor-α in human papillary thyroid carcinomas studied by laser- capture microdissection and molecular biology. *Cancer science*, 102(10), pp.1921–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21707866 [Accessed March 2, 2015].

Wang, K., Li, M. & Bucan, M., 2007. Pathway-based approaches for analysis of genomewide association studies. *American journal of human genetics*, 81(6), pp.1278–83. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0002929707637756 [Accessed December 27, 2016].

Wang, Y.-L. et al., 2013. Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population. *Journal of medical genetics*, 50(10), pp.689–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23847140 [Accessed July 15, 2014].

Ward, L.D. & Kellis, M., 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research*, 40(Database issue), pp.D930-4. Available at:

http://nar.oxfordjournals.org/content/early/2011/11/07/nar.gkr917.long [Accessed July 10, 2014].

Weber, F. & Eng, C., 2005. Editorial: Germline Variants within RET: Clinical Utility or Scientific Playtoy? *The Journal of Clinical Endocrinology & Metabolism*, 90(11), pp.6334–6336. Available at: http://press.endocrine.org/doi/10.1210/jc.2005-2030 [Accessed October 12, 2016].

Williams, D., 2002. Cancer after nuclear fallout: lessons from the Chernobyl accident. *Nature reviews. Cancer*, 2(7), pp.543–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12094241 [Accessed October 7, 2016].

Williams, J.A. & Phillips, D.H., 2000. Mammary Expression of Xenobiotic Metabolizing Enzymes and Their Potential Role in Breast Cancer. *Cancer Research*, 60(17).

Wiltshire, J.J. et al., 2016. Systematic Review of Trends in the Incidence Rates of Thyroid Cancer. *Thyroid*, p.thy.2016.0100. Available at: http://online.liebertpub.com/doi/10.1089/thy.2016.0100 [Accessed October 27, 2016].

Wu, L. & Zhu, J., 2015. Linear reduction in thyroid cancer risk by oral contraceptive use: a dose-response meta-analysis of prospective cohort studies. *Human reproduction (Oxford, England)*, 30(9), pp.2234–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26141711 [Accessed November 2, 2016].

Xhaard, C. et al., 2015. Anthropometric Risk Factors for Differentiated Thyroid Cancer in Young Men and Women From Eastern France: A Case-Control Study. *American journal of epidemiology*, 182(3), pp.202–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26133374 [Accessed October 31, 2016].

Xhaard, C. et al., 2014. Menstrual and reproductive factors in the risk of differentiated thyroid carcinoma in young women in France: a population-based case-control study. *American journal of epidemiology*, 180(10), pp.1007–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25269571 [Accessed February 24, 2015].

Xu, S. et al., 2013. Oestrogen action on thyroid progenitor cells: relevant for the pathogenesis of thyroid nodules? *The Journal of endocrinology*, 218(1), pp.125–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23645248 [Accessed August 10, 2015].

Yu, K. et al., 2009. Pathway analysis by adaptive combination of P-values. *Genetic epidemiology*, 33(8), pp.700–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2790032&tool=pmcentrez&rendertype=abstract [Accessed June 18, 2015].

Zablotska, L.B. et al., 2008. A cohort study of thyroid cancer and other thyroid diseases after the Chornobyl accident: dose-response analysis of thyroid follicular adenomas detected during first screening in Ukraine (1998-2000). *American journal of epidemiology*, 167(3), pp.305–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17989057 [Accessed October 30, 2016].

Zai, G. et al., 2017. A review of molecular genetic studies of neurocognitive deficits in schizophrenia. *Neuroscience & Biobehavioral Reviews*, 72, pp.50–67.

Zamora-Ros, R. et al., 2015. Reproductive and menstrual factors and risk of differentiated thyroid carcinoma: the EPIC study. *International journal of cancer*, 136(5), pp.1218–27. Available at: http://onlinelibrary.wiley.com.gate2.inist.fr/doi/10.1002/ijc.29067/full [Accessed September 8, 2015].

Zaykin, D.V. et al., 2002. Truncated product method for combiningP-values. *Genetic Epidemiology*, 22(2), pp.170–185. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11788962 [Accessed January 7, 2017].

Zhang, W. et al., 2014. Meta-analysis in the association between obesity and risk of thyroid cancer. *International journal of clinical and experimental medicine*, 7(12), pp.5268–74. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4307477&tool=pmcentrez&rendertype=abstract [Accessed September 10, 2015].

Zheng, T. et al., 2003. Glutathione S-transferase M1 and T1 genetic polymorphisms, alcohol consumption and breast cancer risk. *British journal of cancer*, 88(1), pp.58–62. Available at: http://www.nature.com.gate2.inist.fr/bjc/journal/v88/n1/full/6600708a.html [Accessed March 24, 2015].

Zhou, Q. et al., 2016. Long noncoding RNA PVT1 modulates thyroid cancer cell proliferation by recruiting EZH2 and regulating thyroid-stimulating hormone receptor (TSHR). *Tumor Biology*, 37(3), pp.3105–3113.

Zhu, J. et al., 2016. Parity and thyroid cancer risk: a meta-analysis of epidemiological studies. *Cancer medicine*, 5(4), pp.739–52. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4831293&tool=pmcentrez&rendertype=abstract [Accessed May 3, 2016].

Zimmermann, M.B. & Galetti, V., 2015. Iodine intake as a risk factor for thyroid cancer: a comprehensive review of animal and human studies. *Thyroid research*, 8, p.8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26146517 [Accessed November 2, 2016].

Annexes

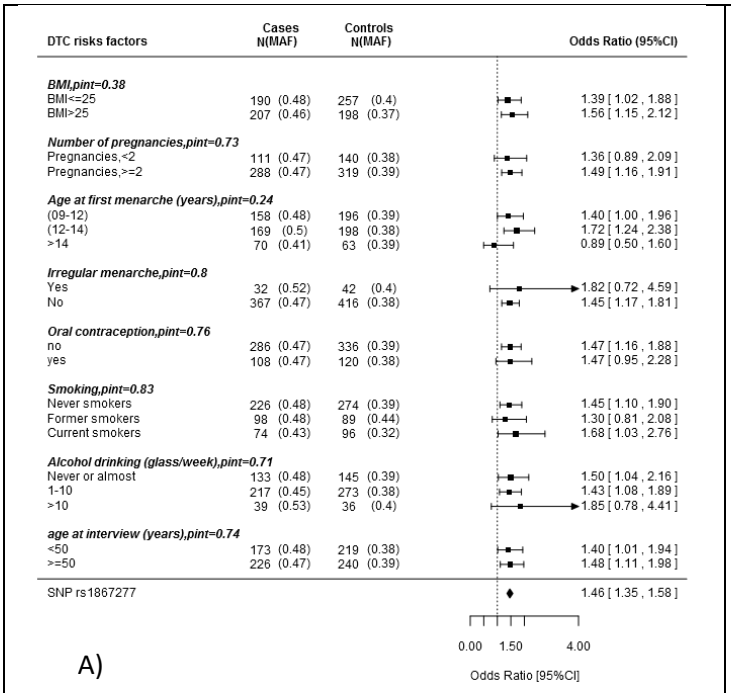
Annexe 1 : Méthode d'estimation du nombre de SNP indépendant pour la correction sur les tests multiples.

Pour l'estimation du nombre de tests indépendants, nous avons utilisé la méthode proposée par Li et al (M.-X. Li et al. 2012) qui implique l'estimation des Eigen values de la matrice de corrélation des p-value des test d'associations (M.-X. Li et al. 2012). Le nombre de test indépendant Me de M test d'association (correspondant au nombre de SNP tester) s'exprime de la manière suivante :

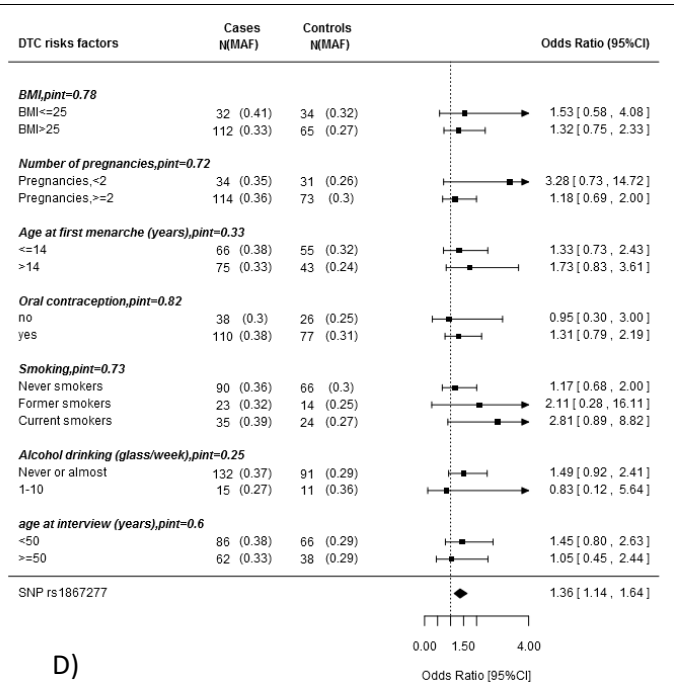
$$Me = M - \sum_{i=1}^{M} [\mathrm{I}(\lambda \mathrm{i} > 1)(\lambda \mathrm{i} - 1)]$$

λi correspond à l'Eigen value de la matrice de corrélation des p-values des M test d'associations associations. La somme $[\mathrm{I}(\lambda \mathrm{i} > 1)(\lambda \mathrm{i} - 1)]$ estime la redondance entre les tests et est la résultante de la dépendance entre les SNP.

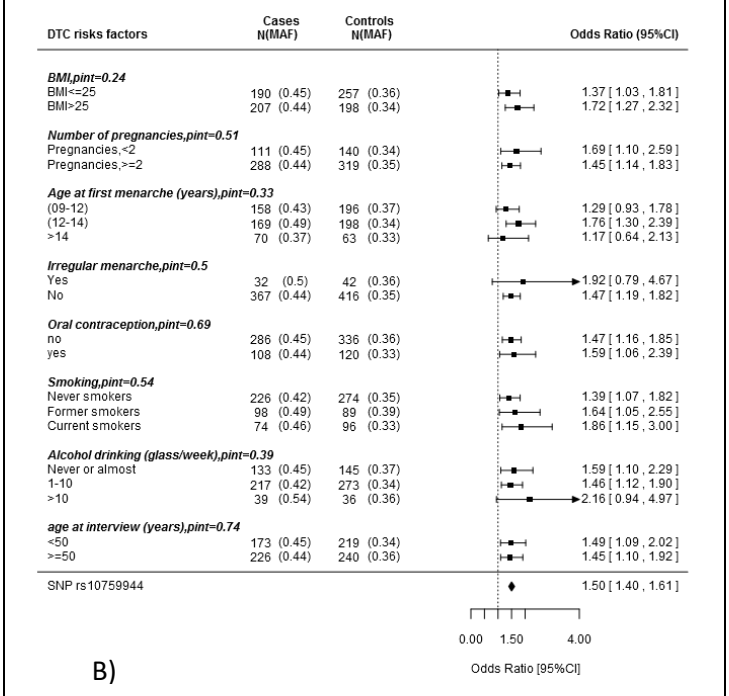
Annexe 2: Annexes des figures



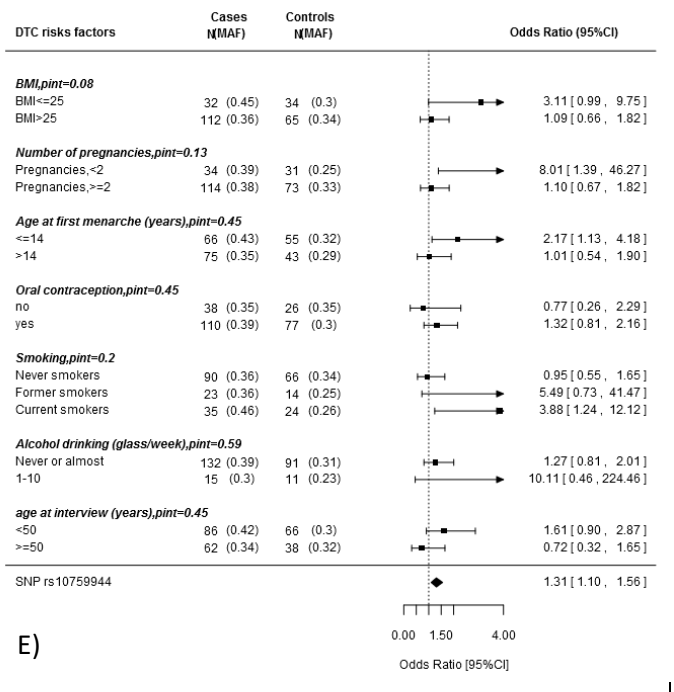
A)



D)



B)



E)

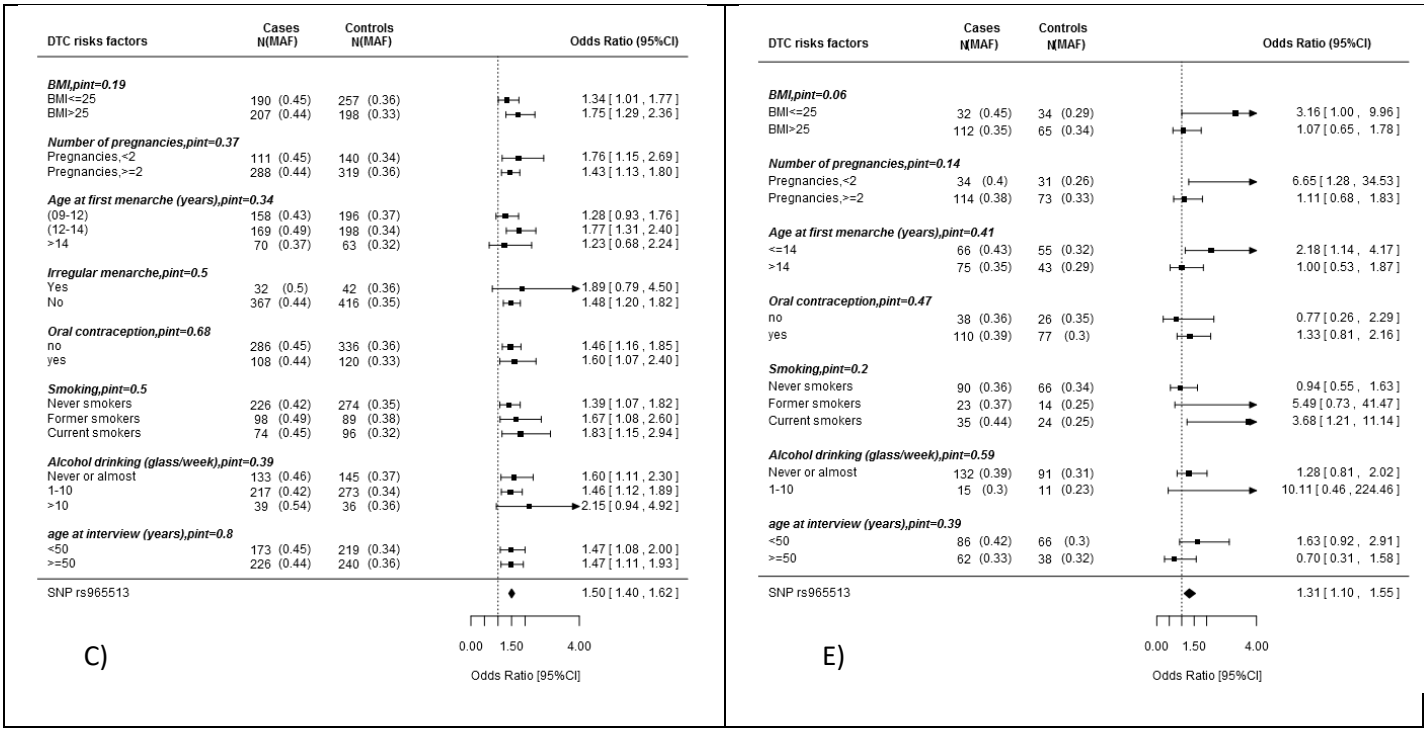
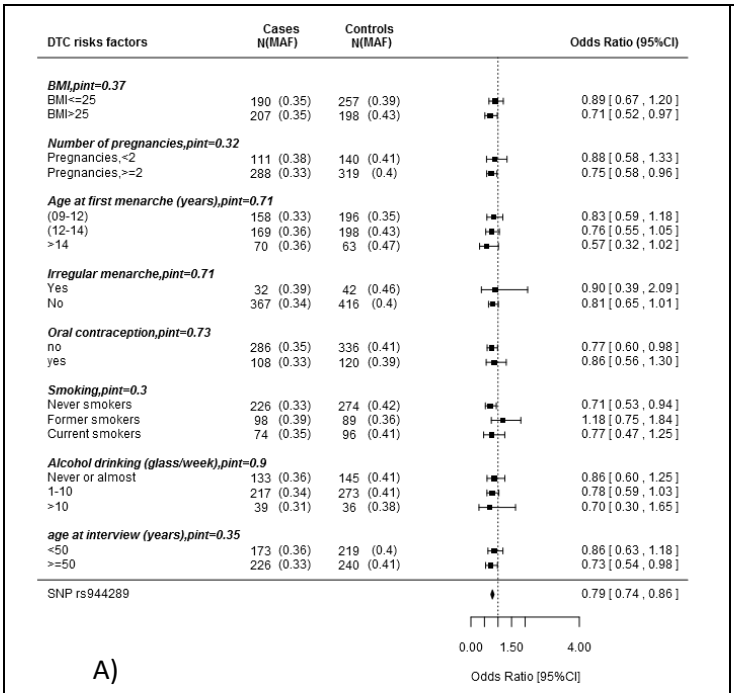
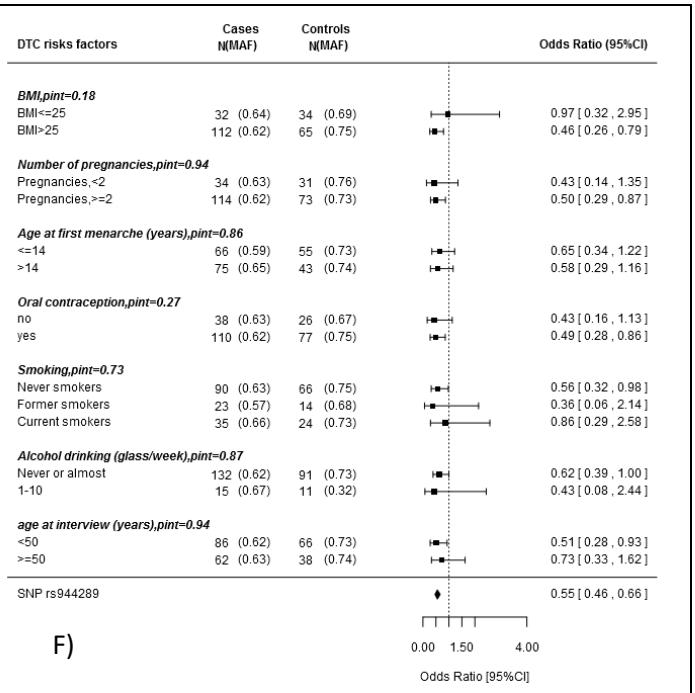


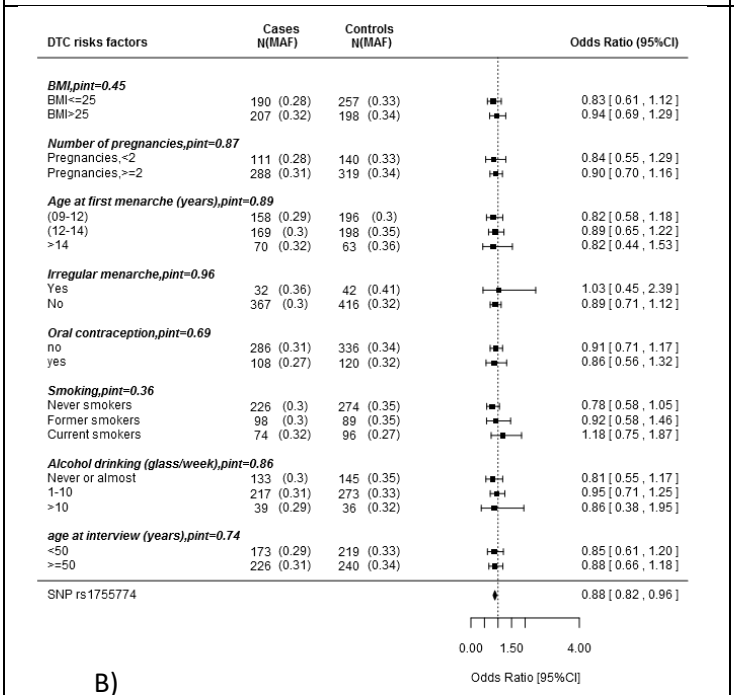
Figure 1 : les « forest plot » des associations entre les CDT et les SNP candidats des loci 9q22 (rs1867277, rs10759944 et rs965513) stratifiés sur l'IMC, la consommation d'alcool, le statut tabagique, l'âge aux premières règles, l'irrégularité des règles et le nombre de grossesses à terme chez femmes d'origine européenne (A, B, C) et mélanésienne (D,E,F)


DTC risks factors
Cases N(MAF)
Controls N(MAF)
Odds Ratio (95%CI)
BMI,pint=0.37
BMI<=25 190 (0.35) 257 (0.39) 0.89 [0.67 , 1.20]
BMI>25 207 (0.35) 198 (0.43) 0.71 [0.52 , 0.97]
Number of pregnancies,pint=0.32
Pregnancies,<2 111 (0.38) 140 (0.41) 0.88 [0.58 , 1.33]
Pregnancies,>=2 288 (0.33) 319 (0.4) 0.75 [0.58 , 0.96]
Age at first menarche (years),pint=0.71
(09-12) 158 (0.33) 196 (0.35) 0.83 [0.59 , 1.18]
(12-14) 169 (0.36) 198 (0.43) 0.76 [0.55 , 1.05]
>14 70 (0.36) 63 (0.47) 0.57 [0.32 , 1.02]
Irregular menarche,pint=0.71
Yes 32 (0.39) 42 (0.46) 0.90 [0.39 , 2.09]
No 367 (0.34) 416 (0.4) 0.81 [0.65 , 1.01]
Oral contraception,pint=0.73
no 286 (0.35) 336 (0.41) 0.77 [0.60 , 0.98]
yes 108 (0.33) 120 (0.39) 0.86 [0.56 , 1.30]
Smoking,pint=0.3
Never smokers 226 (0.33) 274 (0.42) 0.71 [0.53 , 0.94]
Former smokers 98 (0.39) 89 (0.36) 1.18 [0.75 , 1.84]
Current smokers 74 (0.35) 96 (0.41) 0.77 [0.47 , 1.25]
Alcohol drinking (glass/week),pint=0.9
Never or almost 133 (0.36) 145 (0.41) 0.86 [0.60 , 1.25]
1-10 217 (0.34) 273 (0.41) 0.78 [0.59 , 1.03]
>10 39 (0.31) 36 (0.38) 0.70 [0.30 , 1.65]
age at interview (years),pint=0.35
<50 173 (0.36) 219 (0.4) 0.86 [0.63 , 1.18]
>=50 226 (0.33) 240 (0.41) 0.73 [0.54 , 0.98]
SNP rs944289 0.79 [0.74 , 0.86]
0.00 1.50 4.00
Odds Ratio [95%CI]

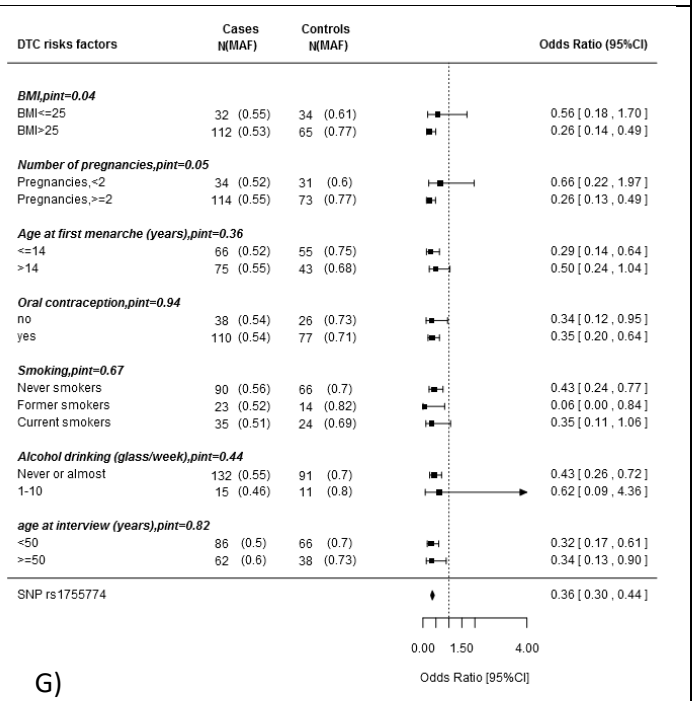
A)


DTC risks factors
Cases N(MAF)
Controls N(MAF)
Odds Ratio (95%CI)
BMI,pint=0.18
BMI<=25 32 (0.64) 34 (0.69) 0.97 [0.32 , 2.95]
BMI>25 112 (0.62) 65 (0.75) 0.46 [0.26 , 0.79]
Number of pregnancies,pint=0.94
Pregnancies,<2 34 (0.63) 31 (0.76) 0.43 [0.14 , 1.35]
Pregnancies,>=2 114 (0.62) 73 (0.73) 0.50 [0.29 , 0.87]
Age at first menarche (years),pint=0.86
<=14 66 (0.59) 55 (0.73) 0.65 [0.34 , 1.22]
>14 75 (0.65) 43 (0.74) 0.58 [0.29 , 1.16]
Oral contraception,pint=0.27
no 38 (0.63) 26 (0.67) 0.43 [0.16 , 1.13]
yes 110 (0.62) 77 (0.75) 0.49 [0.28 , 0.86]
Smoking,pint=0.73
Never smokers 90 (0.63) 66 (0.75) 0.56 [0.32 , 0.98]
Former smokers 23 (0.57) 14 (0.68) 0.36 [0.06 , 2.14]
Current smokers 35 (0.66) 24 (0.73) 0.86 [0.29 , 2.58]
Alcohol drinking (glass/week),pint=0.87
Never or almost 132 (0.62) 91 (0.73) 0.62 [0.39 , 1.00]
1-10 15 (0.67) 11 (0.32) 0.43 [0.08 , 2.44]
age at interview (years),pint=0.94
<50 86 (0.62) 66 (0.73) 0.51 [0.28 , 0.93]
>=50 62 (0.63) 38 (0.74) 0.73 [0.33 , 1.62]
SNP rs944289 0.55 [0.46 , 0.66]
0.00 1.50 4.00
Odds Ratio [95%CI]

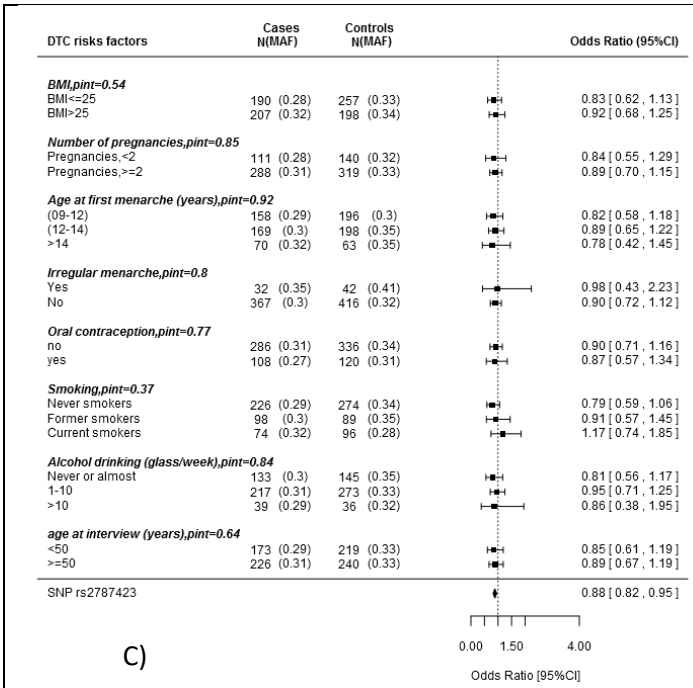
F)


DTC risks factors
Cases N(MAF)
Controls N(MAF)
Odds Ratio (95%CI)
BMI,pint=0.45
BMI<=25 190 (0.28) 257 (0.33) 0.83 [0.61 , 1.12]
BMI>25 207 (0.32) 198 (0.34) 0.94 [0.69 , 1.29]
Number of pregnancies,pint=0.87
Pregnancies,<2 111 (0.28) 140 (0.33) 0.84 [0.55 , 1.29]
Pregnancies,>=2 288 (0.31) 319 (0.34) 0.90 [0.70 , 1.16]
Age at first menarche (years),pint=0.89
(09-12) 158 (0.29) 196 (0.3) 0.82 [0.58 , 1.18]
(12-14) 169 (0.3) 198 (0.35) 0.89 [0.65 , 1.22]
>14 70 (0.32) 63 (0.36) 0.82 [0.44 , 1.53]
Irregular menarche,pint=0.96
Yes 32 (0.36) 42 (0.41) 1.03 [0.45 , 2.39]
No 367 (0.3) 416 (0.32) 0.89 [0.71 , 1.12]
Oral contraception,pint=0.69
no 286 (0.31) 336 (0.34) 0.91 [0.71 , 1.17]
yes 108 (0.27) 120 (0.32) 0.86 [0.56 , 1.32]
Smoking,pint=0.36
Never smokers 226 (0.3) 274 (0.35) 0.78 [0.58 , 1.05]
Former smokers 98 (0.3) 89 (0.35) 0.92 [0.58 , 1.46]
Current smokers 74 (0.32) 96 (0.27) 1.18 [0.75 , 1.87]
Alcohol drinking (glass/week),pint=0.86
Never or almost 133 (0.3) 145 (0.35) 0.81 [0.55 , 1.17]
1-10 217 (0.31) 273 (0.33) 0.95 [0.71 , 1.25]
>10 39 (0.29) 36 (0.32) 0.86 [0.38 , 1.95]
age at interview (years),pint=0.74
<50 173 (0.29) 219 (0.33) 0.85 [0.61 , 1.20]
>=50 226 (0.31) 240 (0.34) 0.88 [0.66 , 1.18]
SNP rs1755774 0.88 [0.82 , 0.96]
0.00 1.50 4.00
Odds Ratio [95%CI]

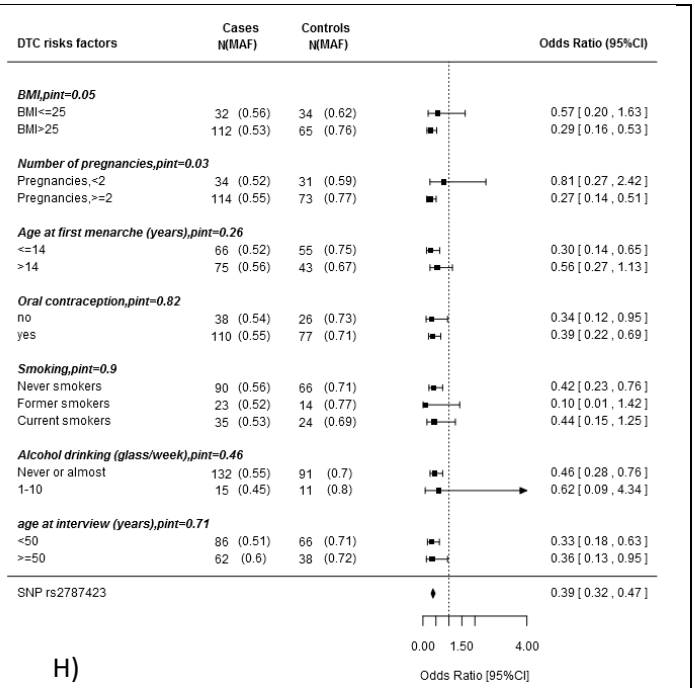
B)


DTC risks factors
Cases N(MAF)
Controls N(MAF)
Odds Ratio (95%CI)
BMI,pint=0.04
BMI<=25 32 (0.55) 34 (0.61) 0.56 [0.18 , 1.70]
BMI>25 112 (0.53) 65 (0.77) 0.26 [0.14 , 0.49]
Number of pregnancies,pint=0.05
Pregnancies,<2 34 (0.52) 31 (0.6) 0.66 [0.22 , 1.97]
Pregnancies,>=2 114 (0.55) 73 (0.77) 0.26 [0.13 , 0.49]
Age at first menarche (years),pint=0.36
<=14 66 (0.52) 55 (0.75) 0.29 [0.14 , 0.64]
>14 75 (0.55) 43 (0.68) 0.50 [0.24 , 1.04]
Oral contraception,pint=0.94
no 38 (0.54) 26 (0.73) 0.34 [0.12 , 0.95]
yes 110 (0.54) 77 (0.71) 0.35 [0.20 , 0.64]
Smoking,pint=0.67
Never smokers 90 (0.56) 66 (0.7) 0.43 [0.24 , 0.77]
Former smokers 23 (0.52) 14 (0.82) 0.06 [0.00 , 0.84]
Current smokers 35 (0.51) 24 (0.69) 0.35 [0.11 , 1.06]
Alcohol drinking (glass/week),pint=0.44
Never or almost 132 (0.55) 91 (0.7) 0.43 [0.26 , 0.72]
1-10 15 (0.46) 11 (0.8) 0.62 [0.09 , 4.36]
age at interview (years),pint=0.82
<50 86 (0.5) 66 (0.7) 0.32 [0.17 , 0.61]
>=50 62 (0.6) 38 (0.73) 0.34 [0.13 , 0.90]
SNP rs1755774 0.36 [0.30 , 0.44]
0.00 1.50 4.00
Odds Ratio [95%CI]

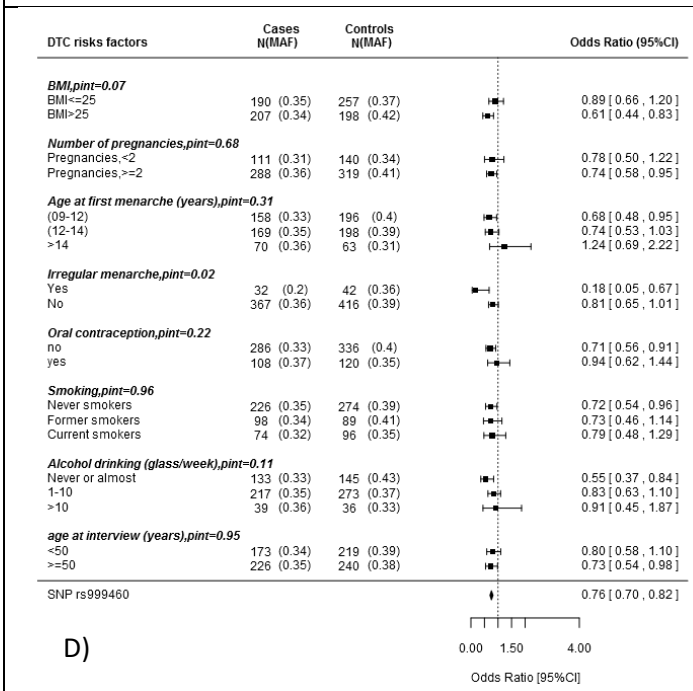
G)


DTC risks factors
Cases N(MAF)
Controls N(MAF)
Odds Ratio (95%CI)
BMI,pint=0.54
BMI<=25 190 (0.28) 257 (0.33) 0.83 [0.62 , 1.13]
BMI>25 207 (0.32) 198 (0.34) 0.92 [0.68 , 1.25]
Number of pregnancies,pint=0.85
Pregnancies,<2 111 (0.28) 140 (0.32) 0.84 [0.55 , 1.29]
Pregnancies,>=2 288 (0.31) 319 (0.33) 0.89 [0.70 , 1.15]
Age at first menarche (years),pint=0.92
(09-12) 158 (0.29) 196 (0.3) 0.82 [0.58 , 1.18]
(12-14) 169 (0.3) 198 (0.35) 0.89 [0.65 , 1.22]
>14 70 (0.32) 63 (0.35) 0.78 [0.42 , 1.45]
Irregular menarche,pint=0.8
Yes 32 (0.35) 42 (0.41) 0.98 [0.43 , 2.23]
No 367 (0.3) 416 (0.32) 0.90 [0.72 , 1.12]
Oral contraception,pint=0.77
no 286 (0.31) 336 (0.34) 0.90 [0.71 , 1.16]
yes 108 (0.27) 120 (0.31) 0.87 [0.57 , 1.34]
Smoking,pint=0.37
Never smokers 226 (0.29) 274 (0.34) 0.79 [0.59 , 1.06]
Former smokers 98 (0.3) 89 (0.35) 0.91 [0.57 , 1.45]
Current smokers 74 (0.32) 96 (0.28) 1.17 [0.74 , 1.85]
Alcohol drinking (glass/week),pint=0.84
Never or almost 133 (0.3) 145 (0.35) 0.81 [0.56 , 1.17]
1-10 217 (0.31) 273 (0.33) 0.95 [0.71 , 1.25]
>10 39 (0.29) 36 (0.32) 0.86 [0.38 , 1.95]
age at interview (years),pint=0.64
<50 173 (0.29) 219 (0.33) 0.85 [0.61 , 1.19]
>=50 226 (0.31) 240 (0.33) 0.89 [0.67 , 1.19]
SNP rs2787423 0.88 [0.82 , 0.95]
0.00 1.50 4.00
Odds Ratio [95%CI]

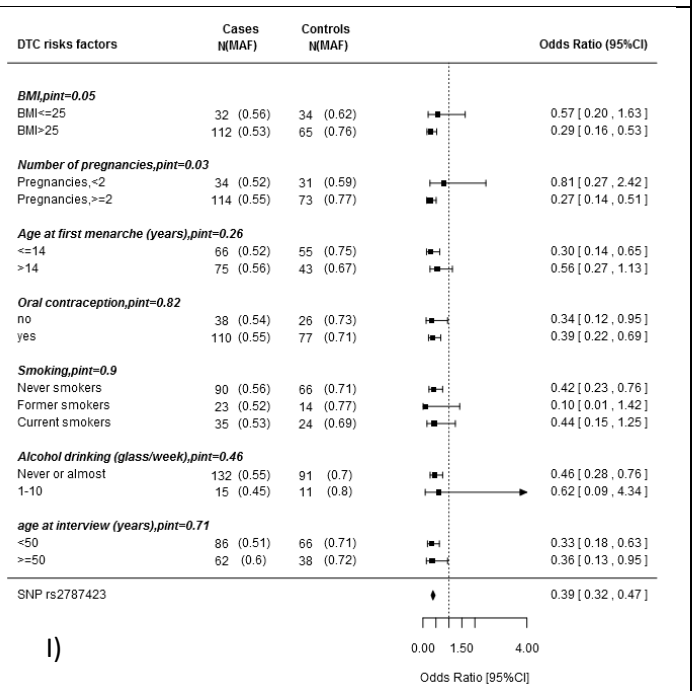
C)


DTC risks factors
Cases N(MAF)
Controls N(MAF)
Odds Ratio (95%CI)
BMI,pint=0.05
BMI<=25 32 (0.56) 34 (0.62) 0.57 [0.20 , 1.63]
BMI>25 112 (0.53) 65 (0.76) 0.29 [0.16 , 0.53]
Number of pregnancies,pint=0.03
Pregnancies,<2 34 (0.52) 31 (0.59) 0.81 [0.27 , 2.42]
Pregnancies,>=2 114 (0.55) 73 (0.77) 0.27 [0.14 , 0.51]
Age at first menarche (years),pint=0.26
<=14 66 (0.52) 55 (0.75) 0.30 [0.14 , 0.65]
>14 75 (0.56) 43 (0.67) 0.56 [0.27 , 1.13]
Oral contraception,pint=0.82
no 38 (0.54) 26 (0.73) 0.34 [0.12 , 0.95]
yes 110 (0.55) 77 (0.71) 0.39 [0.22 , 0.69]
Smoking,pint=0.9
Never smokers 90 (0.56) 66 (0.71) 0.42 [0.23 , 0.76]
Former smokers 23 (0.52) 14 (0.77) 0.10 [0.01 , 1.42]
Current smokers 35 (0.53) 24 (0.69) 0.44 [0.15 , 1.25]
Alcohol drinking (glass/week),pint=0.46
Never or almost 132 (0.55) 91 (0.7) 0.46 [0.28 , 0.76]
1-10 15 (0.45) 11 (0.8) 0.62 [0.09 , 4.34]
age at interview (years),pint=0.71
<50 86 (0.51) 66 (0.71) 0.33 [0.18 , 0.63]
>=50 62 (0.6) 38 (0.72) 0.36 [0.13 , 0.95]
SNP rs2787423 0.39 [0.32 , 0.47]
0.00 1.50 4.00
Odds Ratio [95%CI]

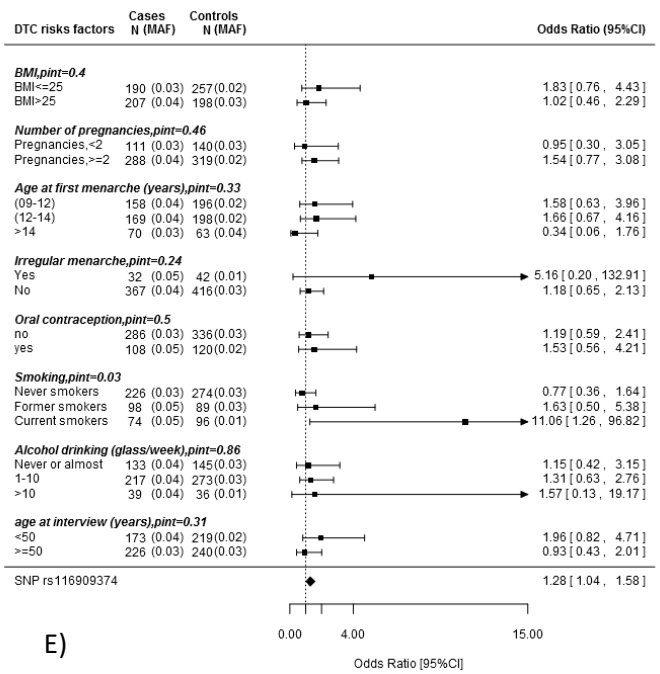
H)


DTC risks factors
Cases N(MAF)
Controls N(MAF)
Odds Ratio (95%CI)
BMI,pint=0.07
BMI<=25 190 (0.35) 257 (0.37) 0.89 [0.66 , 1.20]
BMI>25 207 (0.34) 198 (0.42) 0.61 [0.44 , 0.83]
Number of pregnancies,pint=0.68
Pregnancies,<2 111 (0.31) 140 (0.34) 0.78 [0.50 , 1.22]
Pregnancies,>=2 288 (0.36) 319 (0.41) 0.74 [0.58 , 0.95]
Age at first menarche (years),pint=0.31
(09-12) 158 (0.33) 196 (0.4) 0.68 [0.48 , 0.95]
(12-14) 169 (0.35) 198 (0.39) 0.74 [0.53 , 1.03]
>14 70 (0.36) 63 (0.31) 1.24 [0.69 , 2.22]
Irregular menarche,pint=0.02
Yes 32 (0.2) 42 (0.36) 0.18 [0.05 , 0.67]
No 367 (0.36) 416 (0.39) 0.81 [0.65 , 1.01]
Oral contraception,pint=0.22
no 286 (0.33) 336 (0.4) 0.71 [0.56 , 0.91]
yes 108 (0.37) 120 (0.35) 0.94 [0.62 , 1.44]
Smoking,pint=0.96
Never smokers 226 (0.35) 274 (0.39) 0.72 [0.54 , 0.96]
Former smokers 98 (0.34) 89 (0.41) 0.73 [0.46 , 1.14]
Current smokers 74 (0.32) 96 (0.35) 0.79 [0.48 , 1.29]
Alcohol drinking (glass/week),pint=0.11
Never or almost 133 (0.33) 145 (0.43) 0.55 [0.37 , 0.84]
1-10 217 (0.35) 273 (0.37) 0.83 [0.63 , 1.10]
>10 39 (0.36) 36 (0.33) 0.91 [0.45 , 1.87]
age at interview (years),pint=0.95
<50 173 (0.34) 219 (0.39) 0.80 [0.58 , 1.10]
>=50 226 (0.35) 240 (0.38) 0.73 [0.54 , 0.98]
SNP rs999460 0.76 [0.70 , 0.82]
0.00 1.50 4.00
Odds Ratio [95%CI]

D)


DTC risks factors
Cases N(MAF)
Controls N(MAF)
Odds Ratio (95%CI)
BMI,pint=0.05
BMI<=25 32 (0.56) 34 (0.62) 0.57 [0.20 , 1.63]
BMI>25 112 (0.53) 65 (0.76) 0.29 [0.16 , 0.53]
Number of pregnancies,pint=0.03
Pregnancies,<2 34 (0.52) 31 (0.59) 0.81 [0.27 , 2.42]
Pregnancies,>=2 114 (0.55) 73 (0.77) 0.27 [0.14 , 0.51]
Age at first menarche (years),pint=0.26
<=14 66 (0.52) 55 (0.75) 0.30 [0.14 , 0.65]
>14 75 (0.56) 43 (0.67) 0.56 [0.27 , 1.13]
Oral contraception,pint=0.82
no 38 (0.54) 26 (0.73) 0.34 [0.12 , 0.95]
yes 110 (0.55) 77 (0.71) 0.39 [0.22 , 0.69]
Smoking,pint=0.9
Never smokers 90 (0.56) 66 (0.71) 0.42 [0.23 , 0.76]
Former smokers 23 (0.52) 14 (0.77) 0.10 [0.01 , 1.42]
Current smokers 35 (0.53) 24 (0.69) 0.44 [0.15 , 1.25]
Alcohol drinking (glass/week),pint=0.46
Never or almost 132 (0.55) 91 (0.7) 0.46 [0.28 , 0.76]
1-10 15 (0.45) 11 (0.8) 0.62 [0.09 , 4.34]
age at interview (years),pint=0.71
<50 86 (0.51) 66 (0.71) 0.33 [0.18 , 0.63]
>=50 62 (0.6) 38 (0.72) 0.36 [0.13 , 0.95]
SNP rs2787423 0.39 [0.32 , 0.47]
0.00 1.50 4.00
Odds Ratio [95%CI]

I)



E)

Figure 2 : les « forest plot » des associations entre les CDT et les SNP candidats des loci 9q22 14q13 (rs944289, rs1755774, rs2787423, rs116909374 et rs999460) stratifiés sur l'IMC, la consommation d'alcool, le statut tabagique, l'âge aux premières règles, l'irrégularité des règles et le nombre de grossesses à terme chez les femmes d'origine européenne (A-E) et mélanésienne (F-I)

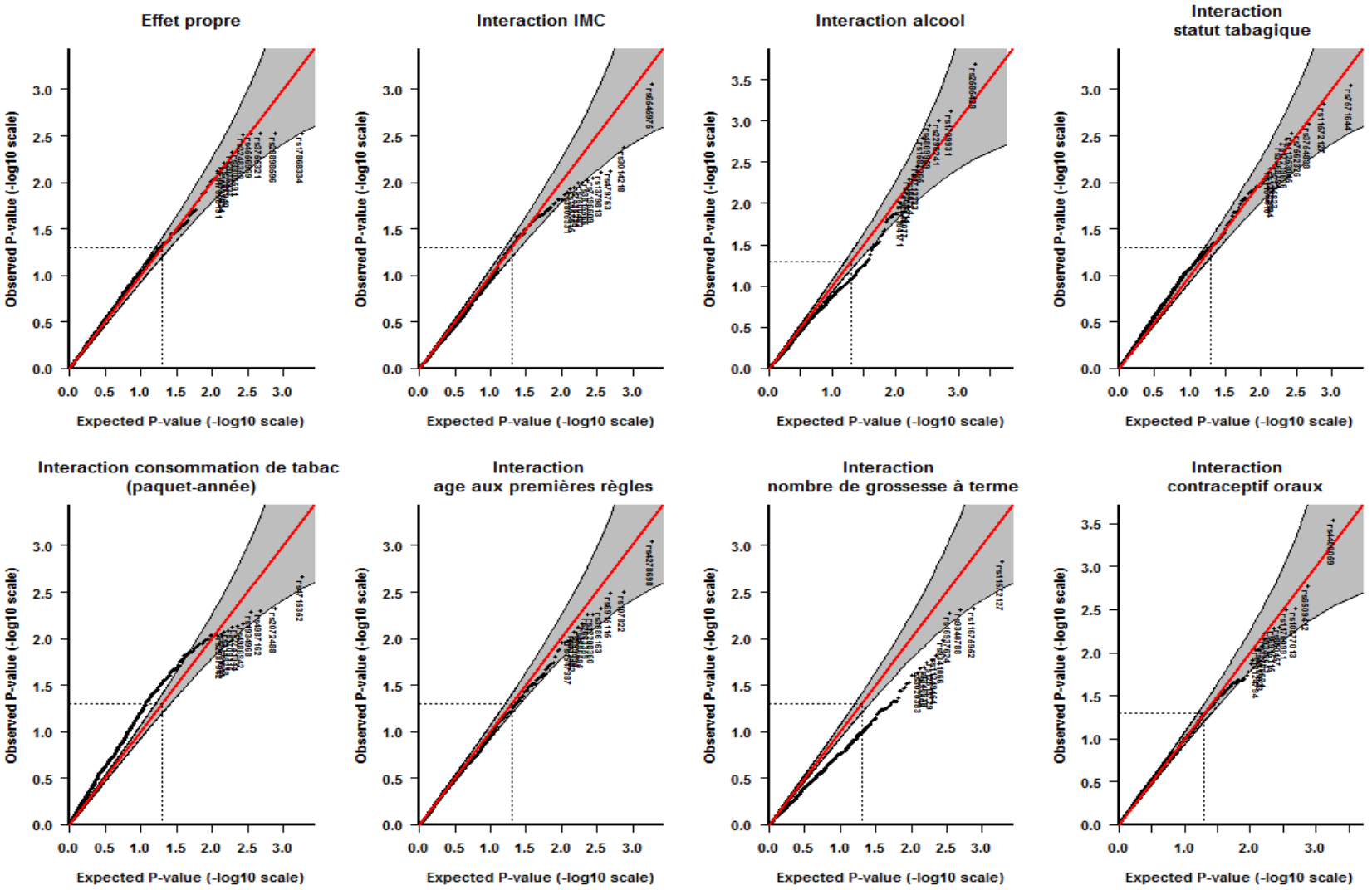


Figure 3 : QQ plot des résultats de p value d'interaction dans l'étude CATHY/Young-Thyr. L'axe des abscisses représente les p-values attendues sous l'hypothèse d'une distribution uniforme et l'axe des ordonnées représente les p-values. Les traits en pointillés représentent le seuil de significativité à 5%.

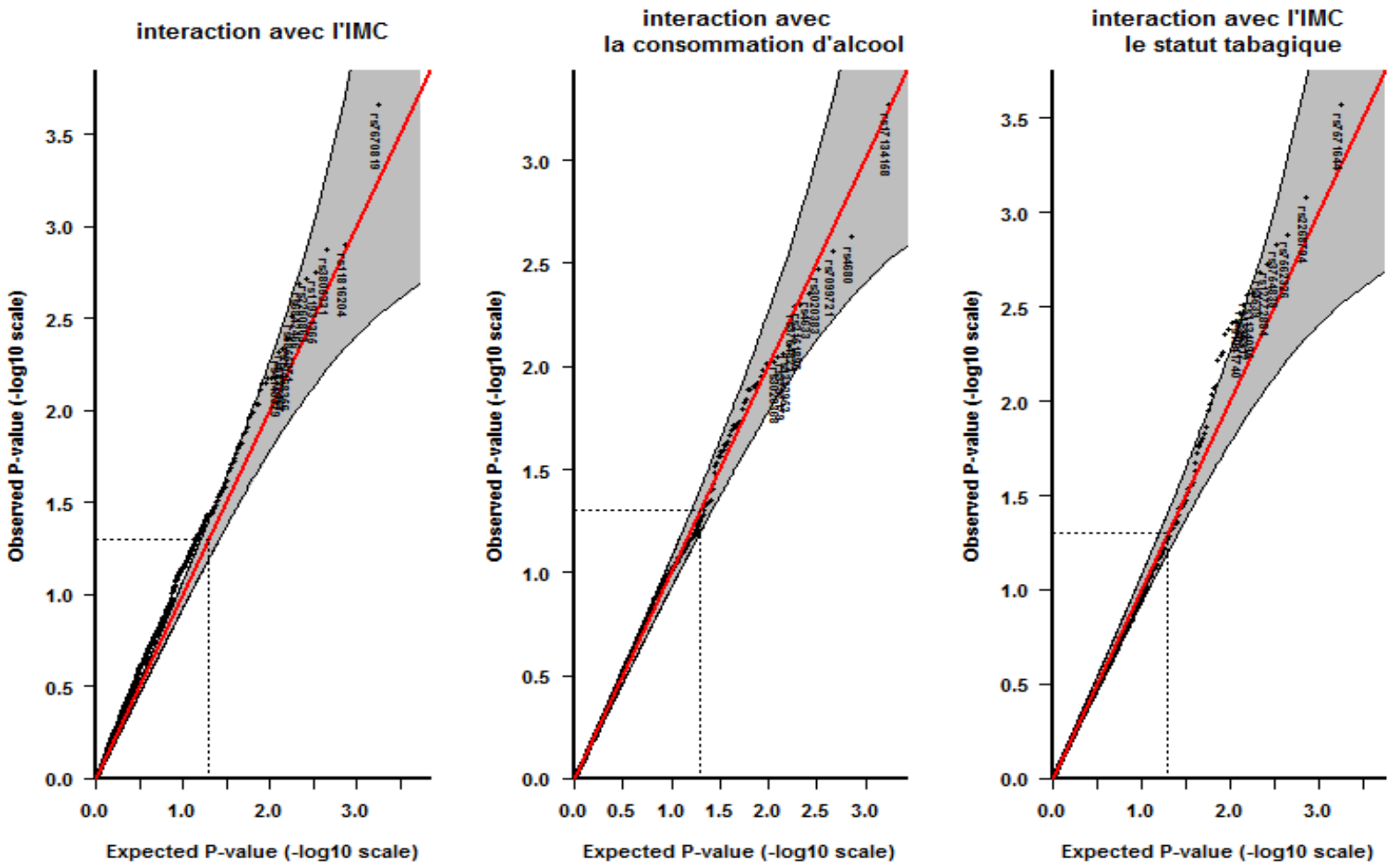
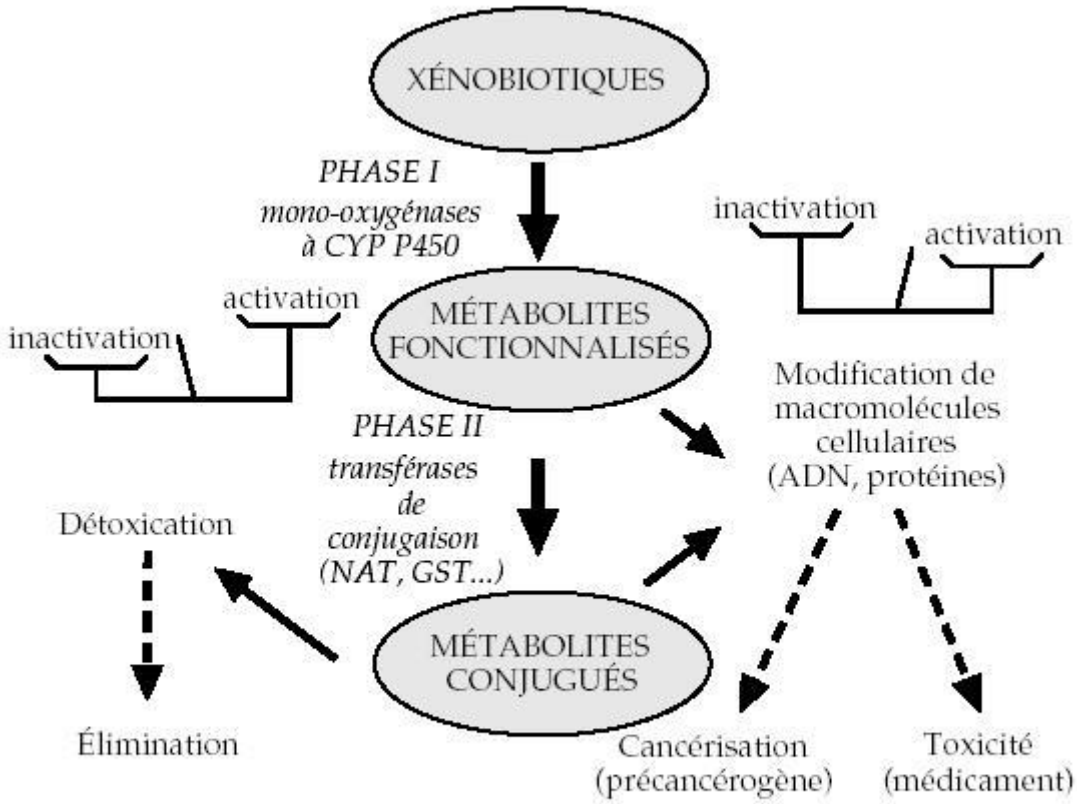


Figure 4 : QQ plot des pvalues d'interaction entre les SNp et l'IMC, la consommation d'alcool et de tabac dans le pool des données CATHY/Young-Thyr et USRT/UTMDACC



Biotransformation des xénobiotiques

CYP : mono-oxygénase à cytochrome P450 ; NAT : *N*-acétyltransférase ; GST : glutathion *S*-transférase

Figure 5 : figure résumant les phases du métabolisme des xénobiotiques. L'Expertise collective « Susceptibilités génétiques et expositions professionnelles », Inserm, 2001.

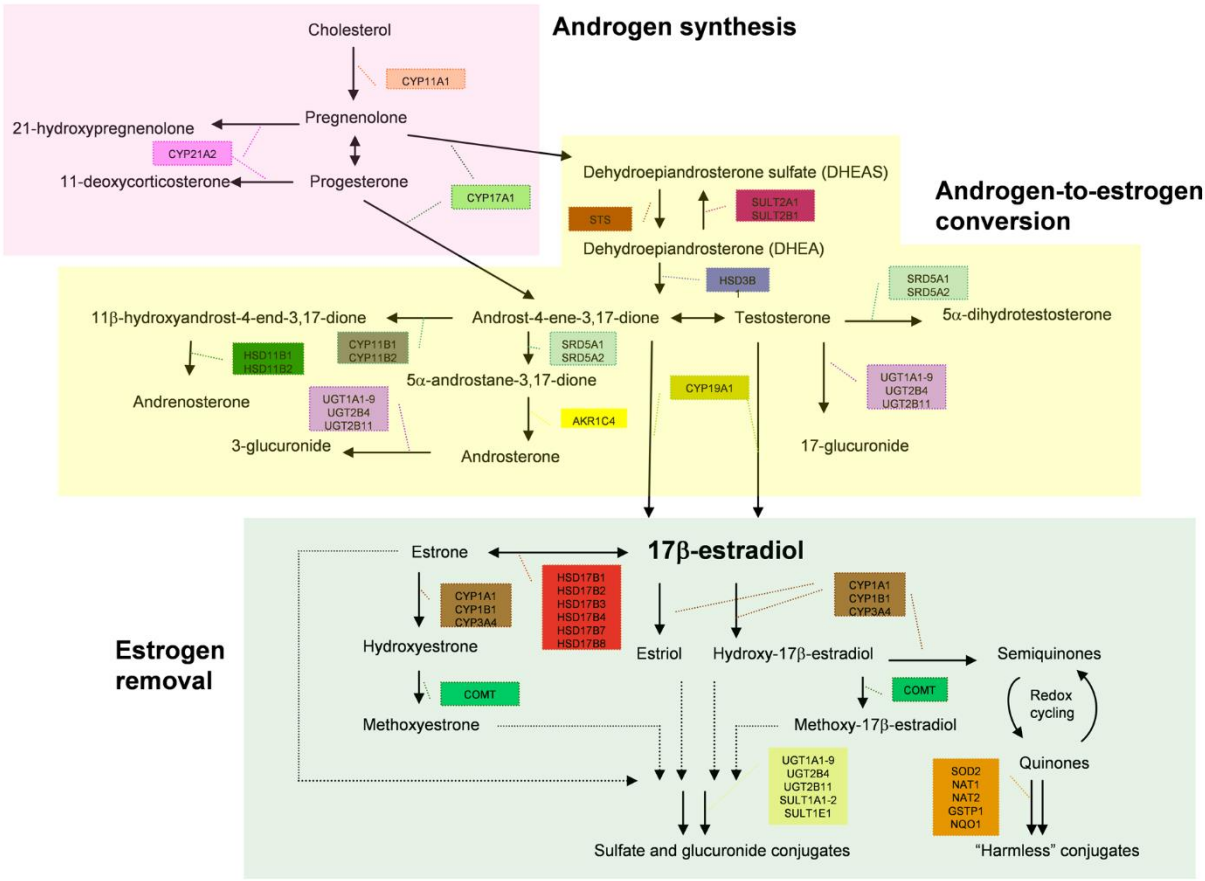


Figure 6 : Figurant résumant les étapes de la biosynthèse des hormones stéroïdiennes et les gènes impliqués dans chacune des phases. Figure tirée de l'aricle de Low YL et al, PLoS Genet, 2010.

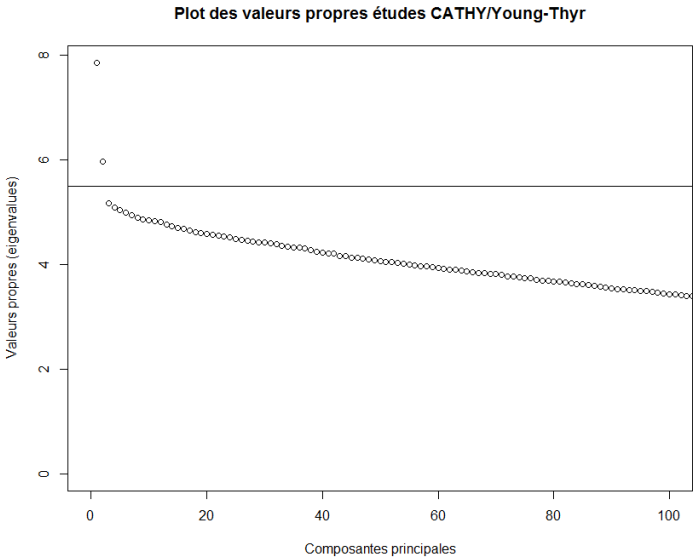


Figure 7 : Plot des valeurs propres des 100 premières composantes principales études CATHY/Youngthyr pour identification des axes à inclure dans le modèle finale. La ligne représente le décrochement (coude). Les axes dont les valeurs sont situées avant cette ligne explique la plus grande part de la variabilité.

Annexe 3: Résultats d'association entre les facteurs environnementaux et le risque de CDT dans les populations initiales des études CATHY et NC

Tableau 1 : Association entre l'IMC, la consommation et de tabac et le risque de CDT dans la population initiale de l'étude CATHY ; en population générale, chez les femmes et chez les hommes séparément.

	Population générale				Homme				Femmes			
Caractéristique démographique	**Cas 621(%)**	**Témoins 706(%)**	**OR**	**IC à 95%**	**Cas 138(%)**	**Témoins 201(%)**	**OR**	**IC à 95%**	**Cas 483(%)**	**Témoins 505(%)**	**OR**	**IC à 95%**
Origine ethnique												
Européenne	593 (95,5)	689 (97,6)	1,00	-	129 (93,5)	196 (97,5)	1,00	-	464 (96,1)	493 (97,6)	1,00	-
Magrébine	18 (2,9)	11 (1,6)	1,93	(0,84 -4,42)	5 (3,6)	4 (2,0)	2,50E+00	(0,52 -12,03)	13 (2,7)	7 (1,4)	1,4	(0,49 -3,98)
Africaine ou antillaise	5 (0,8)	2 (0,3)	3,65	(0,68 -19,59)	3 (2,2)	0 (0,0)	-	-	2 (0,4)	2 (0,4)	1,46	(0,19 -11,47)
Asiatique	2 (0,3)	0 (0,0)	-	-	-	-	-	-	2 (0,4)	0 (0,0)	5,60E+05	-
Autre	3 (0,5)	4 (0,6)	0,79	(0,16 -3,76)	1 (0,7)	1 (0,5)	2,17E+00	(0,12 -40,87)	2 (0,4)	3 (0,6)	6,30E-01	(0,10 -4,11)
IMC en classes, selon l'OMS												
Maigre	15 (2,4)	20 (2,8)	0,94	(0,84 -1,05)	1 (0,7)	1 (0,5)	1,83	(0,09 -35,93)	14 (2,9)	19 (3,8)	0,93	(0,43 -2)
Normal	281 (45,2)	348 (49,3)	1,00	-	51 (37,0)	77 (38,3)	-	-	230 (47,6)	271 (53,7)	1,00	-
Surpoids	200 (32,2)	219 (31,0)	1,13	(0,87 -1,47)	56 (40,6)	90 (44,8)	0,89	(0,52 -1,51)	144 (29,8)	129 (25,5)	1,23	(0,89 -1,71)
Obèse	122 (19,6)	114 (16,1)	1,27	(0,92 -1,75)	30 (21,7)	32 (15,9)	1,44	(0,73 -2,84)	92 (19,0)	82 (16,2)	1,25	(0,85 -1,84)
manquants	3 (0,5)	5 (0,7)	-	-	0 (0,0)	1 (0,5)	-	-	3 (0,6)	4 (0,8)	-	-
Nombre de verres d'alcool												
0 ou presque	183 (29,5)	161 (22,8)	1,00	-	21 (15,2)	19 (9,5)	1,00	-	162 (33,5)	142 (28,1)	1,00	-
1-10	343 (55,2)	426 (60,3)	0,76	(0,58 -1,01)	71 (51,4)	107 (53,2)	0,75	(0,33 -1,68)	272 (56,3)	319 (63,2)	0,73	(0,53 -0,99)
>10	89 (14,3)	113 (16,0)	0,89	(0,59 -1,32)	45 (32,6)	74 (36,8)	0,7	(0,31 -1,6)	44 (9,1)	39 (7,7)	0,96	(0,57 -1,63)
manquants	5 (0,8)	6 (0,8)	-	-	1 (0,7)	1 (0,5)	-	-	5 (1,0)	5 (1,0)	-	-
Statut tabagique												
Non-fumeur	333 (53,6)	357 (50,6)	1,00	-	49 (35,5)	60 (29,9)	1,00	-	284 (58,8)	297 (58,8)	1,00	-
Ancien Fumer	183 (29,5)	194 (27,5)	1,12	(0,85 -1,48)	65 (47,1)	88 (43,8)	1,02	(0,59 -1,79)	119 (24,6)	106 (21,0)	1,21	(0,86 -1,7)
Fumeur Actuel	104 (16,7)	155 (22,0)	0,76	(0,55 -1,05)	24 (17,4)	53 (26,4)	0,56	(0,29 -1,09)	80 (16,6)	102 (20,2)	0,87	(0,59 -1,29)
Nombre de paquets-an												
Non-fumeur	333 (53,6)	357 (50,6)	1,00	-	49 (35,5)	60 (29,9)	1,00	-	284 (58,8)	297 (58,8)	1,00	-
0-10	127 (20,5)	146 (20,7)	1,02	(0,75 -1,37)	21 (15,2)	45 (22,4)	0,65	(0,33 -1,28)	106 (21,9)	101 (20,0)	0,65	(0,33 -1,28)
10-20	77 (12,4)	90 (12,7)	1,06	(0,74 -1,53)	28 (20,3)	40 (19,9)	0,99	(0,51 -1,93)	49 (10,1)	50 (9,9)	0,99	(0,51 -1,93)
20-30	36 (5,8)	48 (6,8)	0,8	(0,49 -1,32)	11 (8,0)	15 (7,5)	1,01	(0,40 -2,59)	25 (5,2)	33 (6,5)	1,01	(0,40 -2,59)
>30	43 (6,9)	59 (8,4)	0,81	(0,51 -1,28)	26 (18,8)	39 (19,4)	0,78	(0,38 -1,58)	17 (3,5)	20 (4,0)	0,78	(0,38 -1,58)
manquants	5 (0,8)	6 (0,8)	-	-	3 (2,2)	2 (1,0)	-	-	2 (0,4)	4 (0,8)	-	-

Les ORs et IC sont ajustés sur l'âge en classe de 5, le sexe et la région de résidence

Tableau 2 : Association entre l'IMC, la consommation et de tabac et le risque de CDT dans la population initiale de l'étude de NC ; en population générale, chez les femmes et chez les hommes séparément

	Population générale				**Homme**				**Femme**			
caractéristique démographique	**Cas 332(%)**	**Témoins 412(%)**	**OR**	**IC à 95%**	**Cas 39(%)**	**Témoins 58(%)**	**OR**	**IC à 95%**	**Cas 293(%)**	**Témoins 353(%)**	**OR**	**IC à 95%**
Ethnie												
Européens	42 (12,7)	133 (32,3)	1,00	-	10 (25,6)	23 (39,7)	1,00	-	32 (10,9)	110 (31,1)	1,00	-
mélanésiens	244 (73,5)	189 (45,9)	2,47	(1,46 -4,17)	24 (61,5)	20 (34,5)	1,71	(0,27 -10,99)	220 (75,1)	169 (47,7)	2,45	(1,28 -4,67)
autres	46 (13,9)	90 (21,8)	1,29	(0,73 -2,29)	5 (12,8)	15 (25,9)	1,79	(0,22 -14,28)	41 (14,0)	75 (21,2)	1,33	(0,68 -2,59)
Indice de masse corporelle selon l'OMS												
Maigre	7 (2,1)	16 (3,9)	1,07	(0,39 -2,9)	0 (0,0)	1 (1,7)	-	-	7 (2,4)	15 (4,2)	1,15	(0,40 -3,37)
Normal	91 (27,4)	167 (40,5)	1,00	-	11 (28,2)	17 (29,3)	1,00	-	80 (27,3)	150 (42,4)	1,00	-
Surpoids	102 (30,7)	119 (28,9)	1,18	(0,78 -1,81)	15 (38,5)	23 (39,7)	0,79	(0,20 -3,09)	87 (29,7)	96 (27,1)	1,4	(0,85 -2,3)
Obèse	**72 (21,7)**	**59 (14,3)**	**2,17**	**(1,33 -3,53)**	**8 (20,5)**	**13 (22,4)**	**1,92**	**(0,32 -11,68)**	**64 (21,8)**	**46 (13,0)**	**2,28**	**(1,29 -4,04)**
obésité sévère	**43 (13,0)**	**31 (7,5)**	**2,09**	**(1,14 -3,84)**	**2 (5,1)**	**1 (1,7)**	**1,60**	**(0,04 -67,65)**	**41 (14,0)**	**30 (8,5)**	**2,24**	**(1,15 -4,39)**
DM	17 (5,1)	20 (4,9)	-	-	3 (7,7)	3 (5,2)	-	-	14 (4,8)	17 (4,8)	-	-
Nombre de verres par semaine												
jamais ou presque	277 (83,4)	301 (73,1)	1,00	-	17 (43,6)	26 (44,8)	1,00	-	260 (88,7)	275 (77,7)	1,00	-
1-10	37 (11,1)	84 (20,4)	0,58	(0,35 -0,96)	10 (25,6)	15 (25,9)	1,42	(0,30 -6,61)	27 (9,2)	69 (19,5)	0,48	(0,27 -0,87)
>10	16 (4,8)	25 (6,1)	1,15	(0,50 -2,62)	11 (28,2)	17 (29,3)	1,47	(0,30 -7,07)	5 (1,7)	8 (2,3)	1,07	(0,26 -4,41)
DM	2 (0,6)	2 (0,5)	-	-	1 (2,6)	0 (0,0)	-	-	1 (0,3)	2 (0,6)	-	-
Statut tabagique												
Non-fumeur	195 (58,7)	236 (57,3)	1,00	-	15 (38,5)	20 (34,5)	1,00	-	180 (61,4)	216 (61,0)	1,00	-
Ancien Fumeur	51 (15,4)	64 (15,5)	0,94	(0,57 -1,56)	9 (23,1)	17 (29,3)	4,18	(0,75 -23,41)	42 (14,3)	47 (13,3)	0,98	(0,54 -1,79)
Fumeur Actuel	85 (25,6)	112 (27,2)	0,89	(0,50 -1,57)	15 (38,5)	21 (36,2)	1,42	(0,25 -8,18)	70 (23,9)	91 (25,7)	1,16	(0,58 -2,33)
DM	1 (0,3)	0 (0,0)			-	-	-	-	1 (0,3)	0 (0,0)	-	-
Nombre de paquets an												
Non-fumeur	195 (58,7)	236 (57,3)	1,00	-	15 (38,5)	20 (34,5)	1,00	-	180 (61,4)	216 (61,0)	1,00	-
0-10	74 (22,3)	75 (18,2)	1,45	(0,92 -2,28)	7 (17,9)	8 (13,8)	0,73	(0,14 -3,91)	67 (22,9)	67 (18,9)	1,54	(0,92 -2,56)
10-20	12 (3,6)	32 (7,8)	0,37	(0,17 -0,83)	2 (5,1)	7 (12,1)	0,45	(0,03 -6,3)	10 (3,4)	25 (7,1)	0,45	(0,18 -1,15)
20-30	19 (5,7)	27 (6,6)	0,91	(0,44 -1,9)	1 (2,6)	5 (8,6)	0,03	(0,00 -1,49)	18 (6,1)	22 (6,2)	1,08	(0,46 -2,54)
>30	23 (6,9)	35 (8,5)	0,84	(0,42 -1,68)	12 (30,8)	15 (25,9)	0,22	(0,04 -1,34)	11 (3,8)	20 (5,6)	0,9	(0,34 -2,4)
DM	9 (2,7)	7 (1,7)	-	-	2 (5,1)	3 (5,2)	-	-	7 (2,4)	4 (1,1)	-	-

Les ORs et IC sont ajustés sur l'âge en classe de 5, le sexe et la province de résidence ; DM : données manquantes

Tableau 3 : Association entre les facteurs hormonaux et reproductifs et le risque de CDT chez les femmes dans la population initiale des études CATHY et NC.

	Etude CATHY				Etude NC			
Caractéristique démographique	**Cas 483(%)**	**Témoins 505(%)**	**OR**	**IC à 95%**	**Cas 293(%)**	**Témoins 354(%)**	**OR**	**IC à 95%**
Règles irrégulières								
Non	449 (93,0)	466 (92,3)	1,00	-	241 (82,3)	307 (86,7)	1,00	-
Oui	34 (7,0)	39 (7,7)	0,75	(0,45 -1,27)	44 (15,0)	39 (11,0)	1,71	(0,98 -2,99)
DM	0	0			8 (2,7)	8 (2,3)	-	-
Age aux premières règles								
09-12	189 (39,1)	213 (42,2)	1,00	-	42 (14,3)	76 (21,5)	1,00	-
13 ou (13-14)[1]	117 (24,2)	125 (24,8)	1,1	(0,78 -1,56)	113 (38,6)	151 (42,7)	1,07	(0,61 -1,89)
14	93 (19,3)	95 (18,8)	1,11	(0,76 -1,61)	120 (41,0)	-	-	-
>=15	82 (17,0)	69 (13,7)	1,55	(1,02 -2,34)	18 (6,1)	109 (30,8)	1,44	(0,80 -2,62)
DM	0	0			42 (14,3)	18 (5,1)	-	-
Nombre de grossesse à terme								
0	44 (9,1)	53 (10,5)	1,00	-	35 (11,9)	49 (13,8)	1,00	-
1	93 (19,3)	90 (17,8)	1,13	(0,63 -2,03)	36 (12,3)	55 (15,5)	0,89	(0,41 -1,9)
2	171 (35,4)	186 (36,8)	1,00	(0,56 -1,76)	31 (10,6)	77 (21,8)	0,63	(0,29 -1,36)
3	107 (22,2)	121 (24,0)	1,00	(0,55 -1,83)	42 (14,3)	47 (13,3)	1,18	(0,54 -2,59)
4-5	50 (10,4)	43 (8,5)	1,27	(0,63 -2,54)	50 (17,1)	61 (17,2)	1,05	(0,48 -2,28)
6-7 (ou≥6)[2]	18 (3,7)	12 (2,4)	2,26	(0,81 -6,33)	38 (13,0)	29 (8,2)	1,21	(0,50 -2,91)
≥ 8					61 (20,8)	36 (10,2)	2,01	(0,85 -4,75)
Manquant	0	0			0	0		
pilule								
NON	138 (28,6)	119 (23,6)	1,00	(0,97 -2,07)	195 (66,6)	214 (60,5)	1,00	-
OUI	340 (70,4)	384 (76,0)	0,71	(0,48 -1,03)	95 (32,4)	137 (38,7)	1,53	(0,94 -2,49)
DM	5 (1,0)	2 (0,4)	-	-	3 (1,0)	3 (0,8)	-	-
Statut ménopausique								
Pré ménopause	201 (41,6)	252 (49,9)	1,00	-	162 (55,3)	180 (50,8)	1,00	
Ménopause naturelle	261 (54,0)	236 (46,7)	0,67	(0,45 -1,00)	99 (33,8)	131 (37,0)	0,70	(0,40 -2,80)
Ménopause artificielle	21 (4,3)	17 (3,4)	1,10	(0,53 -2,28)	22 (7,5)	25 (7,1)	1,1	(0,50-2 ,70)
Ménopause indéterminé					6 (2,0)	11 (3,1)	0,60	(0,20-1,90)
Manquant	0	0			4 (1,4)	7 (2,0)		
THM								
NON	373 (77,2)	383 (75,8)	1,00	-				
OUI	104 (21,5)	118 (23,4)	0,66	(0,37 -0,98)				
DM	6 (1,2)	4 (0,8)						

Les ORs et IC ont été ajustés sur l'âge, la province de résidences, l'IMC, la consommation d'alcool et de tabac, [1](13-14) dans l'étude de NC ; [2](≥6) dans l'étude de CATHY, DM : données manquantes

Annexe 4: P-value d'interaction entre les variants candidats et les facteurs de risques environnementaux

						IMC		Nombre de grossesse		Age aux premières règles		Statut tabagique		Consommation d'alcool	
CHR	**SNP**	**Allele1**	**Allele2**	**Genes**	**A risque**	**Européens**	**Mélanésiens**	**Européens**	**Mélanésiens**	**Européens**	**Mélanésiens**	**Européens**	**Mélanésiens**	**Européens**	**Mélanésiens**
Variants gènes Candidat															
chr17	**rs1042522**	**C**	**G**	***TP53***	**C**	**4,58E-03**	**0,06**	0,43	0,15	0,72	0,22	0,47	0,41	0,93	0,28
chr11	rs10767664	A	T	*BDNF*	T	0,30	0,16	0,63	0,09	0,29	0,51	0,17	0,30	0,74	0,21
chr19	rs1799782	C	T	*XRCC1*	T	0,77	0,59	0,73	0,73	0,11	0,03	0,44	0,17	**0,03**	0,12
chr11	rs1800057	C	G	*ATM*	G	0,33	-	0,88	-	0,45	-	0,12	-	0,09	-
chr11	rs1801516	A	G	*ATM*	A	0,25	-	0,53	-	0,95	-	0,40	-	0,81	-
chr11	rs189037	A	G	*ATM*	G	0,51	0,15	0,15	0,67	0,94	0,61	0,88	0,46	0,46	0,57
chr2	rs2048722	C	T	*TPO*	C	0,87	0,76	0,21	0,06	0,82	0,47	0,58	0,87	0,88	**0,01**
chr8	rs2439302	C	G	*NRG1*	C	0,10	0,54	0,46	0,27	0,82	0,21	0,72	0,59	0,45	0,18
chr19	rs25487	A	G	*XRCC1*	A	0,48	0,65	0,69	0,76	0,36	0,16	0,60	0,28	0,62	0,52
chr5	rs2910164	C	G	*pre-miRNA-146a*	C	0,40	0,80	0,34	0,34	0,64	0,91	0,66	0,12	0,48	0,45
chr11	rs3092993	A	C	*ATM*	A	0,25	-	0,53	-	0,95	-	0,40	-	0,81	-
chr11	rs4752904	C	G	*PTPRJ*	G	0,18	0,15	0,07	0,32	0,95	0,07	0,44	0,50	0,50	0,78
chr11	rs664677	C	T	*ATM*	C	0,77	0,30	0,23	0,99	0,83	0,67	0,86	0,60	0,43	0,64
chr5	rs6885099	A	G	*PDE8B*	G	0,92	0,19	0,79	0,26	0,91	0,65	0,24	0,13	0,57	0,45
chr8	rs6983267	G	T	*gene desert*	G	0,51	0,63	0,93	0,39	0,85	0,82	0,99	0,06	0,80	0,67
chr2	**rs732609**	**A**	**C**	***TPO***	**C**	0,89	0,32	**0,33**	**4,79E-03**	0,58	0,18	0,63	0,61	0,74	0,82
chr2	rs966423	C	T	*DIRC3*	C	0,25	0,06	0,75	0,92	0,20	0,22	0,52	0,37	0,35	0,65
variants des cas familiaux															
chr8	rs1133076	A	G	*TG*	A	0,15	0,20	0,81	0,17	0,21	0,53	0,57	0,58	0,92	0,51
chr8	rs1445493	A	G	_	A	0,86	0,13	0,90	0,98	**9,94E-03**	0,34	0,53	0,34	0,78	**-**
chr8	rs17695552	C	T	*TG,SLA*	C	0,66	0,43	0,95	**0,03**	0,32	0,43	0,27	-	0,74	-
chr8	rs2252696	A	C	*TG,SLA*	C	0,87	0,63	0,70	0,43	0,66	0,04	0,67	0,95	0,66	0,95
chr8	rs2702967	C	T	*TG,SLA*	T	0,05	0,27	0,68	0,55	0,54	0,10	0,32	0,91	0,85	0,93
chr8	rs2702968	A	T	*TG,SLA*	T	0,05	0,27	0,68	0,55	0,54	0,10	0,32	0,91	0,85	0,93
chr8	rs2741205	A	G	*TG,SLA*	A	0,02	0,50	0,69	0,24	0,06	0,68	0,91	0,62	0,26	0,62
chr8	rs822318	C	T	_	C	0,36	0,35	0,74	0,19	0,39	0,19	0,49	0,56	0,46	0,26
chr8	rs9942754	C	T	*TG,SLA*	T	0,97	-	0,01	-	0,21	-	**3,68E-03**	**-**	0,59	-
Variants de l'obésité															
chr11	rs10838738	A	G	*MTCH2*	G	0,98	0,15	0,78	0,31	0,48	**2,19E-03**	0,43	0,26	0,12	0,91
chr19	rs11084753	A	G	*KCTD15*	A	0,15	0,61	0,95	0,23	0,83	0,18	0,40	0,73	0,69	0,98
chr18	rs12970134	A	G	*MC4R*	A	0,04	0,97	0,39	0,86	**6,25E-03**	0,48	0,93	0,41	0,65	0,70
chr4	rs13130484	C	T	_	T	0,63	0,93	0,04	0,10	0,24	0,19	0,14	0,27	0,97	0,94
chr18	rs17782313	C	T	*MC4R*	C	0,11	0,79	0,57	0,86	0,01	0,66	0,56	0,33	0,41	0,45
chr1	rs2815752	C	T	*NEGR1*	C	0,46	0,58	0,08	0,46	0,08	0,48	0,10	0,96	0,74	0,38
chr2	rs6548238	C	T	*TMEM18*	T	0,99	0,37	0,07	0,09	**7,40E-03**	0,37	0,22	0,09	0,93	0,37
chr16	rs9939609	A	T	*FTO*	A	0,43	**0,04**	0,33	0,37	0,35	0,16	0,42	0,68	0,54	0,40

Annexe 5: Association entre le risque de CDT et les acteurs de risque environnementaux stratifié sur les gènes GSTM1 et GSTT1

Tableau 1: Association entre le risque de CDT : le l'IMC, la consommation d'alcool et de tabac stratifié sur le nombre de copie de *GSTM1* et *GSTT1* dans la population européenne des études CATHY et NC,

	≥2 copies			1 copie			0 copie (nul)			
Variable	**Ca/Te**	**OR**	**(IC a 95%)**	**Ca/Te**	**OR**	**(IC a 95%)**	**Ca/Te**	**OR**	**(IC a 95%)**	***p***
GSTM1										
IMC (kg/m²)										0,58
≤25	24/32	1,00		81/125	1,00		120/154	1,00		
25-29,9	20/16	1,47	(0,53-4,09)	52/58	1,50	(0,92-2,44)	75/91	1,03	(0,68-1,55)	
≥30	9/8	1,04	(0,28-3,83)	35/44	1,07	(0,60- 1,88)	51/44	1,40	(0,86- 2,29)	
Consommation d'alcool [a]										0,22
Jamais	14/12			49/58			75/51			
Déjà pris	39/44	0,69	(0,24-2,01)	119/169	0,87	(0,54-1,40)	171/238	0,50	(0,32-0,76)	
Consommation d'alcool (verre/jour)[a]										0,18
Jamais	14/12	1,00		49/58	1,00		75/51	1,00		
<1	26/27	0,84	(0,27-2,61)	91/114	0,92	(0,56- 1,51)	119/174	0,47	(0,30-0,74)	
≥1	13/17	0,42	(0,11-1,62)	28/55	0,73	(0,38- 1,39)	52/64	0,59	(0,34-1,04)	
Statut tabagique [b]										0,43
Jamais	25/31	1,00		91/123	1,00		140/155	1,00		
Déjà fumé	28/25	1,37	(0,32-5,89)	77/104	0,65	(0,32-1,30)	106/134	0,85	(0,49-1,47)	
Consommation de tabac (paquets-année)[a]										0,22
Non-fumeur	25/31	1,00		91/123	1,00		140/155	1,00		
<20	19/19	1,24	(0,45-3,39)	62/67	1,51	(0,92- 2,47)	74/90	0,92	(0,61-1,39)	
≥20	9/6	1,74	(0,38-7,95)	15/37	0,67	(0,33-1,36)	32/44	0,84	(0,48-1,47)	
GSTT1										
IMC (kg/m²)										0,67
≤25	106/117	1,00		95/136	1,00		37/62	1,00		
25-29,9	64/68	0,98	(0,63-1,54)	60/70	1,29	(0,80-2,06)	35/38	1,86	(0,90-3,87)	
≥30	47/45	1,11	(0,67-1,84)	39/39	1,35	(0,77-2,36)	18/14	2,17	(0,86-5,47)	
Consommation d'alcool [a]										0,74
Jamais	62/54	1,00		52/52	1,00		25/20	1,00		
Déjà pris	155/176	0,78	(0,50-1,23)	142/193	0,70	(0,43-1,14)	65/94	0,61	(0,29-1,28)	
Consommation d'alcool (verre/jour)[a]										0,12
Jamais	62/54	1,00		52/52	1,00		25/20	1,00		
<1	100/120	0,73	(0,45-1,16)	100/140	0,66	(0,40- 1,09)	51/58	0,76	(0,35-1,64)	
≥1	55/56	0,97	(0,54-1,75)	42/53	0,84	(0,44-1,60)	14/36	0,33	(0,12-0,87)	
Statut tabagique [b]										0,66
Jamais	117/127	1,00		103/129	1,00		48/55	1,00		
Déjà fumé	100/103	1,17	(0,67-2,06)	91/116	0,58	(0,30- 1,12)	42/59	0,88	(0,34-2,23)	
Consommation de tabac (paquets-année)[a]										0,57
Non-fumeur	117/127	1,00		103/129	1,00		48/55	1,00		
<20	65/67	1,14	(0,73-1,80)	71/78	1,17	(0,74-1,86)	29/39	0,90	(0,45-1,81)	
≥20	35/36	1,15	(0,65-2,05)	20/38	0,57	(0,29-1,10)	13/20	1,00	(0,38-2,62)	

Tous les OR ont été ajustés sur l'âge, le sexe et le département de résidences,

[a] Le modèle multivariée inclut l'IMC, la consommation d'alcool en verre par jour et la consommation de tabac en paquets-année,

[b] The Le modèle multivariée inclut l'IMC, la consommation d'alcool en verre par jour et le statut tabagique,

Ca/Te: Cas/Témoins, OR: Odd ratios, IC à 95%: Intervalle de confiance à 95%, *p*: p-value du test d'interaction

Tableau 2 : Association entre le risque de CDT et la combinaison des génotypes GSTM1/GSTT1 et l'IMC

		Européens				Mélanésiens			
		Cas (%)	**Témoins (%)**	**OR**	**IC à 95%**	**Cas (%)**	**Témoins (%)**	**OR**	**IC à 95%**
IMC (kg/m²)	***GSTM1***								
<30	Non nul	175 (37,63)	226 (39,86)	1,00		16 (10,67)	12 (10,57)	1,00	
<30	Nul	195 (41,93)	245 (43,21)	1,04	(0,79-1,38)	72 (47,33)	65 (62,50)	0,87	(0,37-2,05)
≥30	Non nul	44 (9,46)	52 (9,17)	0,96	(0,60-1,52)	14 (9,33)	7 (6,73)	1,57	(0,46- 5,37)
≥30	Nul	51 (11,00)	44 (7,76)	1,51	(0,95-2,39)	48 (32,67)	21 (20,19)	2,19	(0,85- 5,66)
IMC (kg/m²)	***GSTT1***								
<30	Non nul	298 (63,67)	355 (64,66)	1,00		48 (38,10)	38 (41,11)	1,00	
<30	Nul	72 (15,38)	100 (18,21)	0,89	(0,62-1,26)	27 (20,63)	30 (33,33)	0,68	(0,33-1,37)
≥30	Non nul	80 (17,09)	80 (14,57)	1,10	(0,77-1,58)	28 (22,22)	16 (17,78)	1,83	(0,82-4,07)
≥30	Nul	18 (3,84)	14 (2,55)	1,36	(0,81-1,73)	23 (19,04)	7 (7,78)	3,03	(1,13-8,12)
IMC (kg/m²)	***GSTM1* and *GSTT1***								
<30	Non nul	315 (72,41)	377 (73,50)	1,00		54 (43,20)	40 (43,33)	1,00	
<30	Nul	30 (6,90)	47 (9,16)	0,84	(0,51-1,39)	20 (16,00)	28 (31,11)	0,47	(0,22-0,99)
≥30	Non nul	79 (18,16)	84 (16,37)	1,05	(0,73-1,49)	32 (25,60)	17 (18,90)	1,85	(0,86-3,98)
≥30	Nul	11 (2,53)	5 (1,00)	2,42	(0,81-7,25)	19 (15,20)	6 (6,67)	2,79	(0,97-7,99)

Tous les OR ont été ajustés sur l'âge, le sexe et le département de résidences, Le modèle multivariée inclut l'IMC, la consommation d'alcool en verre par jour et la consommation de tabac en paquets-année, OR: Odd ratios, IC à 95%: Intervalle de confiance à 95%

Tableau 3 : Association entre le risque de CDT et la combinaison des génotypes GSTM1/GSTT1 et la consommation d'alcool

		Européens				Mélanésiens			
		Cas(%)	**Témoins (%)**	**OR**	**95% CI**	**Cas(%)**	**Témoins (%)**	**OR**	**95% CI**
Consommation d'alcool	***GSTM1***								
Jamais	Non-nul	63 (13,55)	67 (11,82)	1,00		13 (8,67)	8 (7,62)	1,00	
Jamais	Nul	75 (16,13)	51 (8,99)	1,61	(0,97-2,66)	55 (36,67)	34 (32,38)	1,17	(0,42-3,29)
Déjà pris	Non-nul	156 (33,55)	211 (37,21)	0,81	(0,54-1,23)	17 (11,33)	11 (10,48)	0,86	(0,25-2,96)
Déjà pris	Nul	171 (36,77)	238 (41,98)	0,81	(0,54-1,22)	65 (43,33)	52 (49,52)	0,78	(0,28-2,21)
Consommation d'alcool	***GSTT1***								
Jamais	Non-nul	108 (23,08)	97 (17,67)	1,00		30 (23,81)	19 (20,88)	1,00	
Jamais	Nul	25 (5,34)	20 (3,64)	1,09	(0,56-2,12)	26 (20,63)	13 (14,29)	1,09	(0,43-2,75)
Déjà pris	Non-nul	270 (57,69)	338 (61,57)	0,73	(0,52-1,02)	46 (36,51)	35 (38,46)	0,70	(0,32- 1,54)
Déjà pris	Nul	65 (13,89)	94 (17,12)	0,64	(0,42-1,00)	24 (19,05)	24 (26,37)	0,54	(0,23-1,31)
Consommation d'alcool	***GSTM1/GSTT1***								
Jamais	Non-nul	119 (27,36)	102 (19,88)	1,00		33 (26,4)	20 (21,98)	1,00	
Jamais	Nul	12 (2,76)	6 (1,17)	1,75	(0,62-4,96)	23 (18,4)	12 (13,19)	1,04	(0,41-2,66)
Déjà pris	Non-nul	275 (63,22)	359 (69,98)	0,67	(0,49-0,93)	53 (42,4)	37 (40,66)	0,74	(0,34-1,60)
Déjà pris	Nul	29 (6,67)	46 (8,97)	0,58	(0,33-1,02)	16 (12,8)	22 (24,18)	0,36	(0,14-0,91)

Annexe 6 : ORs et IC à 95% des SNP significativement associés au risque de CDT seuil 5% dans le locus 9q22 chez les européens et les mélanésiens au

					Européens					Mélanésiens				
			Allèle		Fréquence de l'allèle à risque					Fréquence de l'allèle à risque				
Type de SNP	SNP	Position	Référence	Risque	Cas	Témoins	OR	IC (95%)	p-value	Cas	Témoins	OR	IC (95%)	p-value
imputé	rs1443438	100550028	C	T	0,43	0,33	1,52	(1,28-1,80)	7,12E-06	0,39	0,31	1,43	(0,99-2,06)	0,03
imputé	rs7030241	100550375	A	T	0,43	0,33	1,52	(1,28-1,80)	7,16E-06	0,39	0,31	1,43	(0,99-2,06)	0,03
imputé	rs7027030	100550455	C	A	0,43	0,33	1,52	(1,28-1,80)	7,01E-06	0,39	0,31	1,43	(0,99-2,06)	0,03
imputé	rs10739496	100552559	T	C	0,43	0,33	1,52	(1,28-1,80)	7,40E-06	0,38	0,30	1,43	(1,00-2,07)	0,03
imputé	rs10983761	100553957	C	A	0,43	0,33	1,52	(1,28-1,80)	7,34E-06	0,39	0,30	1,43	(1,00-2,06)	0,03
imputé	rs4743131	100554907	G	C	0,43	0,33	1,52	(1,28-1,80)	7,40E-06	0,38	0,30	1,43	(1,00-2,07)	0,03
genotypé	**rs965513[3]**	**100556109**	**G**	**A**	**0,43**	**0,33**	**1,52**	**(1,28-1,81)**	**5,87E-06**	**0,38**	**0,30**	**1,43**	**(1,00-2,07)**	**0,03**
genotypé	rs10759944	100556972	G	A	0,43	0,34	1,48	(1,25-1,76)	2,25E-05	0,39	0,31	1,42	(0,99-2,05)	0,03
imputé	rs13295081[2]	100559011	C	T	0,49	0,4	1,44	(1,22-1,70)	5,47E-06	-	-	-	-	-
imputé	rs10983796	100560368	T	C	0,31	0,38	0,73	(0,62-0,87)	3,58E-04	0,21	0,23	0,91	(0,60-1,37)	0,54
imputé	rs1867281	100561147	A	G	0,31	0,39	0,73	(0,62-0,87)	3,46E-04	0,21	0,23	0,91	(0,60-1,37)	0,54
imputé	rs10818090	100561486	T	C	0,43	0,33	1,52	(1,28-1,80)	7,56E-06	0,38	0,30	1,43	(1,00-2,07)	0,03
imputé	rs10983806	100562635	A	G	0,44	0,35	1,49	(1,26-1,76)	1,11E-05	0,39	0,31	1,43	(1,00-2,06)	0,02
imputé	rs1156850	100573164	A	G	0,34	0,29	1,23	(1,03-1,47)	0,03	0,06	0,04	1,28	(0,58-2,86)	0,50
imputé	rs7870540	100573925	C	T	0,34	0,29	1,23	(1,03-1,47)	0,03	0,04	0,04	1,27	(0,52-3,08)	0,62
genotypé	rs7870795	100574120	C	T	0,34	0,29	1,23	(1,03-1,47)	0,03	0,04	0,04	1,27	(0,52-3,08)	0,62
imputé	rs7859751	100575888	A	G	0,33	0,29	1,24	(1,04-1,48)	0,02	0,04	0,03	1,29	(0,51-3,26)	0,57
genotypé	rs1443433	100579219	A	G	0,19	0,15	1,29	(1,04-1,61)	0,03	0,37	0,30	1,36	(0,94-1,96)	0,06
genotypé	rs1443432	100583195	T	C	0,49	0,4	1,47	(1,24-1,73)	3,44E-05	0,38	0,30	1,43	(0,99-2,06)	0,03
imputé	rs4297160	100585506	A	G	0,51	0,44	1,36	(1,15-1,6)	2,81E-04	0,38	0,31	1,34	(0,93-1,93)	0,04
imputé	rs79332431	100591463	T	C	0,52	0,44	1,36	(1,15-1,61)	2,62E-04	0,39	0,32	1,35	(0,94-1,93)	0,04
imputé	rs7033765	100591705	T	A	0,52	0,44	1,36	(1,15-1,61)	2,63E-04	0,39	0,32	1,35	(0,94-1,93)	0,04
imputé	rs7043885	100596439	T	C	0,46	0,38	1,38	(1,17-1,63)	2,29E-04	0,36	0,29	1,38	(0,96-2,00)	0,04
imputé	rs4743134	100598568	A	C	0,46	0,38	1,38	(1,17-1,63)	2,27E-04	0,36	0,29	1,38	(0,96-2,00)	0,04
imputé	rs55960649	100598695	A	C	0,47	0,39	1,38	(1,17-1,63)	2,30E-04	0,36	0,29	1,38	(0,96-2,00)	0,04
imputé	rs4743135	100599493	T	C	0,46	0,38	1,38	(1,17-1,63)	2,26E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs7049054	100599969	G	T	0,46	0,38	1,38	(1,17-1,63)	2,26E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs7018683	100600241	C	T	0,46	0,38	1,38	(1,17-1,63)	2,26E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs7847126	100600635	C	A	0,46	0,38	1,38	(1,17-1,63)	2,25E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs7860598	100600702	A	G	0,46	0,38	1,38	(1,17-1,63)	2,25E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs10818124	100603008	T	C	0,46	0,38	1,38	(1,17-1,63)	2,24E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs7040492	100604213	G	A	0,46	0,38	1,38	(1,17-1,63)	2,23E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs7860651[2]	100606004	T	C	0,3	0,24	1,37	(1,14-1,66)	7,35E-04	-	-	-	-	-
imputé	rs6478437	100607497	G	A	0,46	0,39	1,38	(1,17-1,63)	2,30E-04	0,36	0,29	1,37	(0,94-1,98)	0,05
imputé	rs2401639	100607837	C	T	0,46	0,39	1,38	(1,17-1,63)	2,29E-04	0,36	0,29	1,37	(0,95-1,98)	0,05
imputé	rs12348691	100608682	A	G	0,46	0,39	1,38	(1,17-1,64)	2,23E-04	0,36	0,29	1,37	(0,95-1,98)	0,05
imputé	rs10759960	100608745	G	A	0,33	0,28	1,27	(1,06-1,52)	0,01	0,04	0,03	1,37	(0,54-3,46)	0,48

imputé	rs13288000	100608980	C	T	0,46	0,39	1,38	(1,17-1,64)	2,23E-04	0,36	0,29	1,37	(0,94-1,98)	0,05
imputé	rs7873389	100609230	T	C	0,46	0,39	1,38	(1,17-1,64)	2,23E-04	0,36	0,29	1,37	(0,95-1,98)	0,05
imputé	rs10818133	100610348	G	C	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs12341377	100610382	G	C	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs7851552	100610652	A	C	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs7851660	100610759	A	C	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs12348304	100611233	C	T	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs894673	100612270	T	A	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs3758251	100613700	C	G	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs10046805	100613956	A	G	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs3758249	100614140	C	T	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs3758248	100614601	A	G	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs907577	100615117	T	C	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs13302470	100615553	G	A	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
genotypé	**rs1867277[(3)]**	**100615914**	**G**	**A**	**0,46**	**0,38**	**1,38**	**(1,17-1,63)**	**2,59E-04**	**0,36**	**0,29**	**1,34**	**(0,93-1,94)**	**0,06**
imputé	rs1867278	100615949	A	C	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs1867280	100616066	C	G	0,46	0,38	1,39	(1,18-1,64)	1,86E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs3021526	100617021	T	C	0,46	0,38	1,39	(1,17-1,64)	1,90E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
genotypé	rs1443434	100617479	T	G	0,46	0,38	1,4	(1,18-1,65)	1,48E-04	0,36	0,30	1,28	(0,89-1,85)	0,10
imputé	rs1443435	100617583	C	T	0,46	0,38	1,39	(1,17-1,64)	1,90E-04	0,36	0,29	1,38	(0,95-1,99)	0,04
imputé	rs12347191	100619719	T	C	0,46	0,38	1,37	(1,16-1,62)	2,99E-04	0,36	0,29	1,38	(0,95-1,99)	0,04

Les ORs et IC ont été ajustés sur l'âge, le sexe et la région de résidence ; les SNP sont classés selon leurs positions sur le chromosome

Annexe 7: OR et IC à 95% des SNP significativement associés au risque de CDT seuil 5% dans le locus 14q13 chez les européens et les mélanésiens.

					Européens					Mélanésiens				
			Allèle		Fréquence de l'allèle à risque					Fréquence de l'allèle à risque				
Type of SNPs	SNPs	position	Référence	Risque	Cas	Témoins	OR	IC (95%)	p-value	Cas	Témoins	OR	IC (95%)	p-value
imputé	rs8013429	36601061	T	C	0,37	0,43	0,76	(0,64-0,90)	1,63E-03	0,63	0,74	0,59	(0,4-0,86)	1,25E-03
imputé	rs34407552	36602141	T	C	0,40	0,45	0,8	(0,68-0,95)	0,01	0,72	0,81	0,59	(0,39-0,89)	2,57E-03
imputé	rs12437001	36602209	C	T	0,24	0,29	0,76	(0,63-0,92)	4,36E-03	0,46	0,47	0,96	(0,68-1,35)	0,63
imputé	rs1958625	36602676	A	G	0,37	0,43	0,76	(0,64-0,90)	1,63E-03	0,63	0,74	0,59	(0,4-0,86)	1,25E-03
imputé	rs1952712	36603222	T	A	0,22	0,27	0,78	(0,64-0,94)	8,07E-03	-	-	-	-	-
imputé	rs4981323	36603629	A	G	0,26	0,32	0,75	(0,62-0,90)	1,70E-03	0,47	0,48	0,97	(0,69-1,37)	0,69
imputé	rs8005427	36603728	G	A	0,27	0,33	0,75	(0,63-0,90)	9,59E-04	0,46	0,47	0,97	(0,69-1,37)	0,70
imputé	rs8005584	36603729	G	C	0,27	0,33	0,75	(0,63-0,90)	9,59E-04	0,46	0,47	0,97	(0,69-1,37)	0,70
imputé	rs12890957	36605247	G	A	0,24	0,29	0,76	(0,63-0,92)	4,23E-03	0,46	0,47	0,98	(0,69-1,38)	0,72
imputé	rs12896537	36606076	G	C	0,24	0,29	0,77	(0,63-0,93)	4,75E-03	0,46	0,47	0,99	(0,7-1,39)	0,78
imputé	rs12897066	36606468	G	A	0,25	0,31	0,77	(0,64-0,92)	3,61E-03	0,46	0,47	0,98	(0,69-1,38)	0,72
imputé	rs12437348	36606550	A	G	0,24	0,29	0,76	(0,63-0,92)	4,13E-03	0,46	0,47	0,99	(0,70-1,39)	0,77
imputé	rs186170887	36608331	A	T	0,03	0,03	1,26	(0,77-2,05)	0,19	0,16	0,26	0,54	(0,35-0,82)	6,89E-04
imputé	rs10147735	36608953	C	T	0,24	0,30	0,76	(0,63-0,92)	5,21E-03	0,45	0,47	0,94	(0,67-1,33)	0,63
imputé	rs2415316	36609012	C	T	0,35	0,42	0,75	(0,63-0,89)	1,39E-03	0,64	0,76	0,55	(0,38-0,81)	4,53E-04
génotypé	rs2415317	36609678	A	G	0,34	0,41	0,74	(0,62-0,88)	9,57E-04	0,61	0,73	0,56	(0,39-0,82)	7,51E-04
imputé	rs2899846	36609943	C	A	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,73	0,56	(0,39-0,82)	7,29E-04
imputé	rs2899847	36610678	T	C	0,23	0,28	0,74	(0,61-0,90)	4,48E-03	0,43	0,45	0,92	(0,65-1,3)	0,51
imputé	rs4982337	36610778	T	C	0,34	0,41	0,73	(0,62-0,87)	8,25E-04	0,61	0,73	0,57	(0,39-0,83)	1,00E-03
imputé	rs4981324	36610851	C	G	0,34	0,41	0,73	(0,62-0,87)	8,25E-04	0,61	0,73	0,57	(0,39-0,83)	1,00E-03
imputé	rs10150608	36611427	T	C	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,73	0,56	(0,39-0,82)	7,24E-04
imputé	rs142424003	36614651	T	C	0,34	0,41	0,73	(0,62-0,87)	7,76E-04	0,61	0,74	0,57	(0,39-0,83)	8,67E-04
imputé	rs7494749	36615427	G	A	0,38	0,45	0,76	(0,64-0,90)	4,34E-04	0,64	0,76	0,57	(0,39-0,83)	7,94E-04
imputé	rs7494769	36615554	G	A	0,28	0,35	0,75	(0,63-0,90)	2,33E-03	0,61	0,74	0,56	(0,39-0,82)	7,48E-04
imputé	rs147733983	36616785	G	A	0,25	0,31	0,75	(0,62-0,90)	5,11E-03	0,61	0,73	0,56	(0,39-0,82)	6,13E-04
imputé	rs183373305	36616885	A	G	0,28	0,34	0,75	(0,63-0,90)	4,57E-03	0,61	0,74	0,56	(0,39-0,82)	6,40E-04
imputé	rs6571738	36617565	C	G	0,26	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,56	(0,38-0,81)	5,44E-04
imputé	rs2415319	36617793	C	T	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,73	0,56	(0,39-0,81)	6,43E-04
imputé	rs2415320	36617939	G	A	0,25	0,31	0,77	(0,64-0,92)	0,01	0,61	0,73	0,56	(0,39-0,81)	6,43E-04
imputé	rs1169122	36618102	T	A	0,27	0,33	0,75	(0,62-0,90)	3,93E-03	0,61	0,73	0,56	(0,39-0,81)	6,43E-04
imputé	rs1169123	36618493	C	G	0,27	0,33	0,75	(0,62-0,89)	3,53E-03	0,61	0,73	0,57	(0,39-0,82)	8,17E-04
imputé	rs1169124	36619288	T	G	0,34	0,41	0,73	(0,62-0,87)	7,72E-04	0,61	0,74	0,57	(0,39-0,83)	1,04E-03
imputé	rs1169125	36619766	A	G	0,34	0,41	0,73	(0,62-0,87)	7,67E-04	0,61	0,74	0,57	(0,39-0,83)	1,04E-03
imputé	rs1169127	36620680	C	A	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,57	(0,39-0,82)	8,31E-04
imputé	rs1169128	36620865	T	G	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,57	(0,39-0,82)	8,30E-04
imputé	rs1169129	36621825	A	G	0,34	0,41	0,73	(0,62-0,87)	7,55E-04	0,61	0,74	0,57	(0,39-0,83)	1,07E-03

imputé	rs1169130	36621907	C	A	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,57	(0,39-0,82)	8,36E-04
imputé	rs1169131	36622555	A	G	0,27	0,33	0,75	(0,63-0,90)	4,67E-03	0,61	0,74	0,56	(0,39-0,82)	8,39E-04
imputé	rs1169132	36622652	A	G	0,26	0,31	0,76	(0,63-0,91)	8,03E-03	0,61	0,74	0,56	(0,39-0,82)	8,40E-04
imputé	rs1169133	36622815	C	T	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,56	(0,39-0,82)	8,40E-04
génotypé	rs1169134	36624720	T	C	0,25	0,31	0,78	(0,64-0,93)	0,02	0,61	0,74	0,56	(0,39-0,82)	8,47E-04
imputé	rs1169135	36624902	C	T	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,56	(0,39-0,82)	8,50E-04
imputé	rs1169136	36625599	G	A	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,56	(0,39-0,82)	8,53E-04
imputé	rs1169137	36625754	G	A	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,56	(0,39-0,82)	8,52E-04
imputé	rs1169138	36625884	C	T	0,26	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,56	(0,39-0,82)	8,43E-04
imputé	rs1169139	36627136	T	C	0,34	0,41	0,74	(0,62-0,87)	8,49E-04	0,61	0,74	0,57	(0,39-0,83)	1,11E-03
imputé	rs146562657	36628232	G	C	0,40	0,46	0,8	(0,68-0,95)	0,01	0,76	0,86	0,53	(0,34-0,84)	3,20E-03
imputé	rs1169140	36628366	G	C	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,56	(0,39-0,81)	6,79E-04
imputé	rs1169142	36628814	C	T	0,26	0,31	0,77	(0,64-0,93)	0,01	0,62	0,74	0,58	(0,39-0,84)	1,25E-03
imputé	rs1169143	36629212	G	A	0,26	0,31	0,78	(0,65-0,94)	0,02	0,61	0,74	0,57	(0,39-0,82)	9,03E-04
imputé	rs1169144	36629769	A	T	0,25	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,57	(0,39-0,82)	8,75E-04
imputé	rs1169145	36630318	G	A	0,27	0,33	0,75	(0,63-0,90)	4,71E-03	0,61	0,74	0,57	(0,39-0,82)	8,78E-04
imputé	rs1305576	36630734	T	G	0,26	0,31	0,77	(0,64-0,93)	0,01	0,61	0,74	0,57	(0,39-0,82)	8,80E-04
génotypé	rs1177590	36632555	T	C	0,23	0,29	0,74	(0,62-0,90)	4,55E-03	0,43	0,44	0,94	(0,66-1,33)	0,59
génotypé	rs1169146	36632769	A	G	0,34	0,41	0,73	(0,61-0,87)	5,93E-04	0,62	0,74	0,58	(0,4-0,84)	1,40E-03
imputé	rs1169147	36632903	A	C	0,26	0,31	0,76	(0,63-0,92)	9,74E-03	0,62	0,74	0,57	(0,39-0,83)	1,03E-03
imputé	rs1169148	36633415	T	C	0,23	0,29	0,74	(0,61-0,90)	4,20E-03	0,43	0,45	0,94	(0,67-1,33)	0,60
imputé	rs1169149	36634225	T	G	0,23	0,29	0,74	(0,61-0,89)	3,39E-03	0,43	0,45	0,95	(0,67-1,34)	0,60
imputé	rs1182929	36635152	T	G	0,34	0,41	0,73	(0,61-0,87)	6,15E-04	0,62	0,74	0,58	(0,4-0,84)	1,28E-03
imputé	rs934075	36638222	A	G	0,25	0,31	0,72	(0,60-0,87)	1,18E-03	0,43	0,45	0,94	(0,67-1,33)	0,58
imputé	rs2774166	36638655	G	A	0,28	0,34	0,74	(0,62-0,89)	3,23E-03	0,62	0,74	0,57	(0,39-0,83)	1,13E-03
imputé	rs1820604	36639206	C	G	0,28	0,34	0,75	(0,62-0,89)	3,75E-03	0,61	0,74	0,57	(0,39-0,83)	8,90E-04
imputé	rs1183904	36640069	T	C	0,25	0,31	0,72	(0,60-0,87)	1,19E-03	0,43	0,45	0,94	(0,66-1,33)	0,57
imputé	rs1169150	36640590	G	A	0,26	0,31	0,77	(0,64-0,92)	0,01	0,61	0,74	0,57	(0,39-0,83)	9,63E-04
génotypé	rs1169151	36640601	A	G	0,34	0,42	0,73	(0,62-0,87)	7,54E-04	0,62	0,74	0,58	(0,4-0,84)	1,47E-03
imputé	rs1834854	36640635	T	C	0,23	0,29	0,74	(0,61-0,89)	3,73E-03	0,43	0,45	0,93	(0,66-1,32)	0,54
imputé	rs1834855	36640638	G	A	0,26	0,31	0,77	(0,64-0,92)	0,01	0,61	0,74	0,57	(0,39-0,83)	9,63E-04
imputé	rs11623884	36644537	G	A	0,03	0,02	1,07	(0,63-1,80)	0,60	0,18	0,29	0,55	(0,37-0,83)	3,77E-03
imputé	rs34992253	36647163	C	T	0,34	0,41	0,73	(0,62-0,87)	6,41E-04	0,62	0,74	0,57	(0,39-0,83)	1,13E-03
génotypé	**rs944289**	**36649246**	**T**	**C**	**0,34**	**0,41**	**0,73**	**(0,61-0,86)**	**4,78E-04**	**0,62**	**0,74**	**0,57**	**(0,39-0,83)**	**1,13E-03**
imputé	rs1619784	36649589	C	T	0,23	0,29	0,74	(0,61-0,89)	3,32E-03	0,43	0,45	0,93	(0,66-1,31)	0,53
imputé	rs1766115	36650371	G	A	0,27	0,32	0,78	(0,65-0,94)	9,33E-03	0,58	0,53	1,22	(0,87-1,73)	0,29
imputé	rs1766116	36650375	G	T	0,27	0,32	0,78	(0,65-0,94)	8,32E-03	0,58	0,53	1,22	(0,87-1,73)	0,29
imputé	rs2787417	36651803	C	T	0,22	0,28	0,73	(0,60-0,89)	3,95E-03	0,46	0,49	0,89	(0,63-1,26)	0,34
imputé	rs2993417	36651854	C	T	0,37	0,41	0,83	(0,7-0,99)	0,08	0,78	0,88	0,49	(0,3-0,78)	1,98E-03
imputé	rs2993418	36651921	C	T	0,26	0,31	0,78	(0,65-0,94)	0,02	0,54	0,55	0,97	(0,69-1,37)	0,73
imputé	rs2988704	36651934	A	G	0,25	0,31	0,78	(0,65-0,94)	0,02	0,54	0,55	0,96	(0,68-1,35)	0,68
imputé	rs2774159	36652248	A	G	0,37	0,41	0,83	(0,70-0,99)	0,08	0,77	0,88	0,48	(0,3-0,78)	1,63E-03
imputé	rs2787418	36652309	A	G	0,32	0,34	0,91	(0,76-1,09)	0,58	0,76	0,87	0,48	(0,3-0,77)	1,00E-03

imputé	rs75570808	36652443	C	T	0,03	0,03	1,13	(0,71-1,82)	0,37	0,18	0,29	0,53	(0,35-0,8)	2,12E-03
imputé	rs78024166	36653813	G	A	0,03	0,03	1,12	(0,70-1,81)	0,41	0,17	0,28	0,53	(0,35-0,8)	2,24E-03
imputé	rs2787419	36654120	T	C	0,26	0,27	0,93	(0,77-1,12)	0,84	0,76	0,87	0,48	(0,3-0,76)	7,33E-04
imputé	rs138797476	36654831	C	G	0,40	0,44	0,84	(0,71-0,99)	0,08	0,78	0,88	0,49	(0,3-0,79)	2,41E-03
imputé	rs150371780	36654894	A	G	0,28	0,29	0,92	(0,77-1,11)	0,76	0,77	0,87	0,48	(0,3-0,77)	1,06E-03
imputé	rs142697020	36655790	A	T	0,03	0,03	1,12	(0,70-1,80)	0,41	0,17	0,28	0,53	(0,35-0,8)	2,25E-03
imputé	rs144059594	36655833	T	C	0,28	0,30	0,92	(0,77-1,11)	0,74	0,76	0,87	0,49	(0,31-0,78)	1,32E-03
imputé	rs139280637	36656583	A	C	0,23	0,27	0,84	(0,69-1,01)	0,16	0,63	0,77	0,50	(0,34-0,73)	6,79E-05
imputé	rs149119426	36656744	T	C	0,27	0,31	0,84	(0,70-1,01)	0,12	0,65	0,78	0,51	(0,35-0,76)	1,19E-04
imputé	rs185812681	36657151	A	G	0,03	0,03	1,12	(0,69-1,80)	0,42	0,17	0,28	0,53	(0,35-0,8)	2,25E-03
imputé	rs11621627	36660440	A	G	0,03	0,03	1,13	(0,70-1,82)	0,40	0,17	0,28	0,53	(0,35-0,8)	2,24E-03
imputé	rs7157340	36662185	G	A	0,03	0,03	1,13	(0,70-1,81)	0,40	0,17	0,28	0,53	(0,35-0,8)	2,22E-03
imputé	rs75538239	36663632	C	T	0,03	0,03	1,12	(0,70-1,80)	0,41	0,17	0,28	0,53	(0,35-0,79)	2,13E-03
imputé	rs77972831	36663788	A	G	0,03	0,03	1,12	(0,70-1,80)	0,41	0,17	0,28	0,53	(0,35-0,79)	2,12E-03
imputé	rs1958611	36664777	G	A	0,04	0,03	1,22	(0,79-1,88)	0,21	0,17	0,28	0,52	(0,35-0,79)	2,00E-03
imputé	rs7154109	36665403	C	T	0,04	0,03	1,21	(0,79-1,87)	0,22	0,17	0,28	0,52	(0,35-0,79)	2,01E-03
imputé	rs1766123	36666488	G	C	0,32	0,34	0,91	(0,77-1,09)	0,59	0,76	0,87	0,47	(0,29-0,75)	7,95E-04
imputé	rs1766124	36666499	T	C	0,37	0,41	0,84	(0,71-0,99)	0,09	0,77	0,88	0,48	(0,3-0,77)	1,70E-03
imputé	rs1755770	36666760	T	C	0,37	0,41	0,84	(0,71-0,99)	0,09	0,77	0,88	0,48	(0,3-0,77)	1,70E-03
imputé	rs1766126	36667286	T	C	0,37	0,41	0,84	(0,71-0,99)	0,09	0,77	0,88	0,48	(0,3-0,77)	1,73E-03
imputé	rs11621565	36667501	A	G	0,04	0,03	1,21	(0,79-1,87)	0,22	0,17	0,28	0,53	(0,35-0,8)	2,11E-03
imputé	rs17553241	36669321	T	C	0,04	0,03	1,2	(0,77-1,86)	0,26	0,17	0,28	0,53	(0,35-0,8)	2,73E-03
imputé	rs77963526	36669955	G	A	0,04	0,03	1,21	(0,78-1,86)	0,23	0,17	0,28	0,53	(0,35-0,8)	2,24E-03
imputé	rs946068	36670083	A	G	0,32	0,34	0,91	(0,76-1,09)	0,57	0,76	0,87	0,47	(0,29-0,75)	8,03E-04
imputé	rs1114365	36671982	C	A	0,32	0,34	0,93	(0,78-1,11)	0,8	0,76	0,87	0,47	(0,29-0,75)	5,87E-04
imputé	rs1114852	36672026	T	G	0,22	0,24	0,89	(0,73-1,08)	0,55	0,61	0,76	0,49	(0,33-0,72)	3,85E-05
imputé	rs28396553	36673392	T	C	0,39	0,43	0,86	(0,73-1,02)	0,17	0,77	0,87	0,48	(0,3-0,77)	1,37E-03
imputé	rs28758606	36673410	C	G	0,39	0,43	0,86	(0,73-1,02)	0,17	0,77	0,87	0,48	(0,3-0,77)	1,29E-03
imputé	rs1958612	36673875	C	T	0,29	0,32	0,83	(0,69-0,99)	0,12	0,61	0,74	0,56	(0,39-0,82)	7,92E-04
imputé	rs10139973	36675034	G	A	0,34	0,36	0,9	(0,75-1,07)	0,44	0,74	0,84	0,54	(0,35-0,84)	3,47E-03
génotypé	rs1952707	36675857	T	C	0,29	0,33	0,83	(0,70-1,00)	0,12	0,62	0,74	0,57	(0,39-0,83)	1,15E-03
imputé	rs10143891	36676667	T	A	0,34	0,37	0,9	(0,76-1,07)	0,48	0,74	0,84	0,55	(0,35-0,85)	3,91E-03
imputé	rs10147188	36677739	T	C	0,29	0,33	0,84	(0,7-1,00)	0,15	0,61	0,74	0,56	(0,38-0,81)	7,42E-04
génotypé	rs7140782	36681985	C	T	0,10	0,11	0,85	(0,65-1,11)	0,23	0,18	0,28	0,57	(0,38-0,86)	9,42E-03
imputé	rs10129798	36682357	C	T	0,10	0,11	0,85	(0,65-1,11)	0,23	0,18	0,28	0,57	(0,38-0,86)	9,41E-03
génotypé	rs7148603	36683779	G	A	0,32	0,36	0,85	(0,71-1,01)	0,14	0,65	0,79	0,49	(0,33-0,73)	9,56E-05
imputé	rs1766135	36686181	C	T	0,28	0,34	0,78	(0,65-0,94)	8,90E-03	0,53	0,69	0,51	(0,36-0,73)	3,37E-05
imputé	rs10132289	36686604	T	C	0,42	0,45	0,87	(0,74-1,03)	0,09	0,56	0,71	0,52	(0,36-0,74)	2,90E-05
imputé	rs11845130	36686606	T	A	0,28	0,34	0,78	(0,65-0,94)	8,96E-03	0,53	0,69	0,52	(0,36-0,74)	4,90E-05
imputé	rs2787423	36687054	G	A	0,29	0,34	0,78	(0,65-0,94)	8,78E-03	0,54	0,69	0,51	(0,35-0,73)	3,69E-05
imputé	rs67263215	36687199	G	T	0,29	0,34	0,78	(0,65-0,94)	8,89E-03	0,53	0,69	0,50	(0,35-0,72)	3,02E-05
imputé	rs17553775	36687724	C	G	0,29	0,34	0,78	(0,65-0,94)	8,47E-03	0,54	0,70	0,51	(0,35-0,73)	3,73E-05
imputé	rs1766136	36687974	A	C	0,29	0,34	0,78	(0,65-0,93)	7,62E-03	0,54	0,70	0,51	(0,35-0,72)	2,21E-05

imputé	rs1766137	36688171	G	T	0,29	0,34	0,78	(0,65-0,94)	8,49E-03	0,53	0,69	0,51	(0,35-0,73)	2,50E-05
imputé	rs1755774	36688516	A	G	0,29	0,34	0,78	(0,65-0,93)	7,29E-03	0,53	0,70	0,50	(0,35-0,71)	1,62E-05
imputé	rs1766138	36688560	G	C	0,29	0,34	0,78	(0,65-0,94)	8,57E-03	0,53	0,69	0,51	(0,35-0,73)	2,72E-05
imputé	rs1766139	36689922	T	C	0,29	0,34	0,78	(0,65-0,93)	8,06E-03	0,54	0,70	0,51	(0,36-0,73)	4,39E-05
imputé	rs1755775	36691187	T	A	0,26	0,32	0,76	(0,63-0,91)	4,02E-03	0,52	0,64	0,60	(0,42-0,85)	1,62E-03
imputé	rs1766140	36691817	T	G	0,29	0,34	0,78	(0,65-0,93)	7,21E-03	0,54	0,70	0,49	(0,34-0,71)	1,78E-05
imputé	rs2774164	36691832	A	G	0,23	0,28	0,74	(0,61-0,90)	1,62E-03	0,39	0,46	0,74	(0,52-1,05)	0,02
imputé	rs1755776	36693285	G	A	0,29	0,34	0,79	(0,66-0,94)	8,77E-03	0,54	0,69	0,51	(0,36-0,73)	4,91E-05
imputé	rs1755777	36693370	G	A	0,29	0,34	0,78	(0,66-0,94)	8,42E-03	0,53	0,69	0,51	(0,35-0,73)	4,50E-05
imputé	rs1755778	36693991	A	G	0,29	0,34	0,78	(0,65-0,93)	6,89E-03	0,53	0,69	0,51	(0,35-0,73)	4,23E-05
imputé	rs1766141	36694219	G	A	0,29	0,34	0,78	(0,65-0,93)	7,42E-03	0,53	0,69	0,51	(0,35-0,73)	4,40E-05
imputé	rs1755779	36694638	C	A	0,26	0,32	0,76	(0,63-0,91)	3,81E-03	0,43	0,58	0,56	(0,39-0,79)	4,24E-04
imputé	rs66606098	36694823	T	C	0,49	0,46	1,13	(0,95-1,33)	0,11	0,38	0,24	1,9	(1,3-2,78)	1,21E-04
génotypé	rs1766142	36696327	C	A	0,40	0,44	0,84	(0,71-0,99)	0,04	0,56	0,72	0,49	(0,34-0,71)	5,03E-05
imputé	rs1766143	36696342	A	G	0,27	0,32	0,76	(0,63-0,91)	5,11E-03	0,54	0,66	0,60	(0,42-0,85)	3,33E-03
imputé	rs1766144	36697247	G	A	0,29	0,34	0,79	(0,66-0,94)	0,01	0,55	0,71	0,50	(0,35-0,72)	1,05E-04
imputé	rs1755781	36698516	A	G	0,27	0,32	0,76	(0,63-0,91)	5,27E-03	0,54	0,66	0,60	(0,42-0,85)	3,72E-03
imputé	rs1755782	36698522	C	T	0,29	0,34	0,79	(0,66-0,94)	0,01	0,55	0,71	0,50	(0,35-0,72)	1,15E-04
imputé	rs1755783	36699362	T	A	0,24	0,29	0,75	(0,62-0,91)	2,81E-03	0,40	0,48	0,73	(0,52-1,04)	0,04
génotypé	rs1766145	36699641	A	T	0,26	0,32	0,76	(0,63-0,91)	4,96E-03	0,53	0,66	0,60	(0,42-0,85)	3,75E-03
imputé	rs1755784	36700375	A	G	0,27	0,32	0,76	(0,63-0,91)	4,81E-03	0,53	0,66	0,60	(0,42-0,85)	3,72E-03
imputé	rs1766146	36701241	G	A	0,29	0,34	0,78	(0,65-0,93)	8,48E-03	0,55	0,71	0,50	(0,35-0,73)	1,25E-04
imputé	rs1755785	36701802	C	T	0,29	0,34	0,78	(0,65-0,94)	9,29E-03	0,55	0,71	0,50	(0,35-0,73)	1,21E-04
imputé	rs1755786	36703291	G	T	0,29	0,34	0,78	(0,65-0,93)	7,40E-03	0,56	0,71	0,51	(0,35-0,73)	1,48E-04
imputé	rs1632250	36704436	T	C	0,27	0,32	0,76	(0,63-0,91)	4,25E-03	0,53	0,66	0,60	(0,42-0,85)	3,77E-03
imputé	rs1755787	36705085	A	T	0,27	0,32	0,76	(0,63-0,91)	4,55E-03	0,53	0,66	0,60	(0,42-0,85)	3,79E-03
imputé	rs1755788	36705394	C	T	0,29	0,35	0,78	(0,65-0,93)	8,33E-03	0,56	0,71	0,51	(0,35-0,73)	1,40E-04
imputé	rs1755789	36705445	T	C	0,27	0,32	0,76	(0,63-0,91)	4,90E-03	0,53	0,66	0,60	(0,42-0,85)	3,77E-03
imputé	rs1618855	36705696	C	A	0,27	0,32	0,76	(0,63-0,91)	4,23E-03	0,53	0,66	0,60	(0,42-0,85)	3,84E-03
imputé	rs1755791	36707251	C	G	0,27	0,32	0,76	(0,64-0,92)	5,48E-03	0,54	0,66	0,60	(0,42-0,85)	3,57E-03
imputé	rs1766147	36708179	C	T	0,26	0,31	0,79	(0,66-0,95)	0,02	0,53	0,65	0,60	(0,42-0,85)	3,92E-03
imputé	rs1766148	36708225	A	G	0,42	0,45	0,91	(0,77-1,08)	0,25	0,57	0,72	0,51	(0,35-0,74)	1,14E-04
imputé	rs1766149	36708406	G	C	0,29	0,33	0,81	(0,67-0,97)	0,03	0,55	0,71	0,51	(0,36-0,74)	1,60E-04
imputé	rs1895804	36709801	A	T	0,26	0,31	0,78	(0,65-0,94)	0,01	0,53	0,65	0,60	(0,42-0,85)	3,89E-03
imputé	rs1895803	36709809	G	A	0,26	0,31	0,78	(0,65-0,94)	0,01	0,53	0,65	0,60	(0,42-0,85)	3,89E-03
imputé	rs1895802	36709826	G	A	0,26	0,31	0,78	(0,65-0,94)	0,01	0,53	0,65	0,60	(0,42-0,85)	3,89E-03
imputé	rs2787424	36710378	T	C	0,26	0,31	0,78	(0,65-0,94)	0,01	0,53	0,66	0,58	(0,41-0,83)	2,17E-03
imputé	rs2787425	36710394	G	T	0,26	0,31	0,79	(0,66-0,95)	0,02	0,53	0,66	0,58	(0,41-0,83)	2,17E-03
imputé	rs1863348	36711554	G	A	0,24	0,29	0,75	(0,62-0,91)	3,05E-03	0,41	0,49	0,73	(0,51-1,03)	0,03
imputé	rs1863347	36711569	A	G	0,24	0,29	0,75	(0,62-0,91)	2,90E-03	0,41	0,49	0,72	(0,51-1,02)	0,03
imputé	rs1863346	36711706	C	T	0,24	0,29	0,75	(0,62-0,91)	3,11E-03	0,41	0,48	0,73	(0,52-1,04)	0,04
imputé	rs2553570	36712126	C	A	0,24	0,29	0,75	(0,62-0,91)	3,02E-03	0,41	0,49	0,73	(0,51-1,03)	0,03
imputé	rs2761340	36712129	T	G	0,42	0,44	0,91	(0,77-1,08)	0,26	0,58	0,74	0,50	(0,35-0,73)	7,95E-05

imputé	rs2764575	36712476	C	T	0,24	0,29	0,75	(0,62-0,91)	3,07E-03	0,41	0,48	0,73	(0,52-1,04)	0,04
imputé	rs2254613	36713154	T	G	0,42	0,49	0,78	(0,66-0,92)	2,09E-03	0,42	0,50	0,73	(0,52-1,04)	0,04
imputé	rs111347631	36713487	G	A	0,15	0,12	1,3	(1,02-1,65)	0,03	0,16	0,21	0,71	(0,46-1,11)	0,27
imputé	rs13379166	36714025	G	A	0,16	0,12	1,3	(1,02-1,65)	0,03	0,16	0,21	0,72	(0,46-1,12)	0,29
imputé	rs2761339	36714788	G	A	0,40	0,45	0,82	(0,69-0,97)	0,02	0,42	0,49	0,74	(0,52-1,05)	0,04
imputé	rs77069096	36714792	A	G	0,16	0,13	1,28	(1,01-1,63)	0,03	0,17	0,22	0,72	(0,46-1,11)	0,26
imputé	rs113803960	36715058	T	A	0,16	0,13	1,28	(1,01-1,63)	0,03	0,17	0,22	0,72	(0,46-1,11)	0,26
imputé	rs142043845	36715526	C	A	0,40	0,45	0,82	(0,69-0,97)	0,02	0,42	0,49	0,74	(0,52-1,05)	0,04
imputé	rs7400734	36715956	G	T	0,44	0,43	1,07	(0,91-1,27)	0,36	0,43	0,30	1,71	(1,19-2,46)	3,46E-03
imputé	rs2787426	36716321	C	T	0,44	0,42	1,08	(0,91-1,27)	0,36	0,42	0,29	1,74	(1,21-2,51)	2,18E-03
imputé	rs114605448	36716686	A	C	0,15	0,12	1,3	(1,02-1,65)	0,02	0,17	0,21	0,76	(0,49-1,18)	0,38
imputé	rs141380976	36716712	T	A	0,16	0,13	1,28	(1,01-1,63)	0,03	0,16	0,21	0,75	(0,48-1,16)	0,34
imputé	rs2263571	36716828	C	T	0,44	0,42	1,08	(0,91-1,27)	0,36	0,42	0,29	1,74	(1,21-2,51)	2,19E-03
imputé	rs2415337	36716853	G	A	0,44	0,42	1,08	(0,91-1,27)	0,36	0,42	0,29	1,74	(1,21-2,51)	2,19E-03
imputé	rs7401714	36718441	G	A	0,44	0,42	1,08	(0,91-1,27)	0,35	0,42	0,29	1,74	(1,21-2,51)	2,20E-03
imputé	rs10149813	36718728	A	G	0,42	0,40	1,07	(0,90-1,26)	0,43	0,40	0,25	2,01	(1,38-2,92)	8,48E-05
imputé	rs3909610	36719782	T	C	0,42	0,40	1,07	(0,90-1,26)	0,42	0,40	0,25	1,98	(1,36-2,89)	1,13E-04
imputé	rs3909611	36720044	A	G	0,44	0,42	1,08	(0,91-1,27)	0,35	0,42	0,30	1,68	(1,17-2,41)	3,59E-03
imputé	rs12590774	36720462	C	T	0,44	0,42	1,08	(0,91-1,27)	0,35	0,42	0,30	1,68	(1,17-2,41)	3,60E-03
imputé	rs2244319	36720659	C	T	0,40	0,45	0,82	(0,69-0,97)	0,02	0,42	0,49	0,74	(0,52-1,05)	0,04
imputé	rs2244414	36720733	C	T	0,44	0,42	1,08	(0,91-1,27)	0,35	0,42	0,30	1,68	(1,17-2,41)	3,61E-03
imputé	rs28602170	36721236	G	A	0,16	0,12	1,31	(1,03-1,66)	0,03	0,15	0,19	0,75	(0,48-1,18)	0,38
imputé	rs8005096	36721373	A	C	0,45	0,43	1,08	(0,91-1,27)	0,35	0,43	0,31	1,65	(1,15-2,36)	5,60E-03
imputé	rs28397092	36721389	G	A	0,16	0,12	1,31	(1,03-1,66)	0,03	0,15	0,19	0,75	(0,48-1,18)	0,38
génotypé	rs2553576	36722498	T	C	0,44	0,43	1,06	(0,90-1,26)	0,42	0,43	0,30	1,7	(1,19-2,45)	3,78E-03
génotypé	rs17764769	36723256	G	T	0,16	0,12	1,31	(1,03-1,67)	0,02	0,16	0,21	0,72	(0,46-1,13)	0,30
génotypé	rs2553577	36723777	A	T	0,40	0,45	0,81	(0,69-0,96)	0,01	0,42	0,49	0,74	(0,52-1,05)	0,04
imputé	rs2553578	36724607	T	C	0,40	0,45	0,81	(0,69-0,96)	0,01	0,42	0,49	0,74	(0,52-1,05)	0,04
imputé	rs11621553	36725221	A	G	0,15	0,12	1,31	(1,03-1,66)	0,03	0,16	0,21	0,72	(0,46-1,12)	0,28
imputé	rs2553574	36726409	C	T	0,40	0,45	0,81	(0,69-0,96)	0,01	0,42	0,49	0,74	(0,52-1,05)	0,04
imputé	rs2764572	36727061	G	A	0,45	0,43	1,08	(0,92-1,28)	0,32	0,43	0,30	1,7	(1,19-2,45)	3,71E-03
imputé	rs2787428	36727898	G	A	0,45	0,43	1,08	(0,92-1,28)	0,32	0,43	0,32	1,61	(1,12-2,3)	8,32E-03
imputé	rs78098267	36728338	T	A	0,15	0,12	1,3	(1,02-1,65)	0,03	0,16	0,19	0,8	(0,51-1,25)	0,50
imputé	rs2761336	36728531	A	G	0,44	0,43	1,07	(0,91-1,27)	0,38	0,43	0,32	1,59	(1,11-2,27)	9,93E-03
imputé	rs2553573	36728874	T	C	0,42	0,40	1,07	(0,90-1,26)	0,43	0,40	0,26	1,92	(1,32-2,8)	1,99E-04
imputé	rs2761337	36728890	G	A	0,44	0,43	1,07	(0,91-1,27)	0,37	0,42	0,31	1,65	(1,15-2,36)	5,81E-03
imputé	rs11626666	36729145	G	A	0,15	0,12	1,3	(1,02-1,65)	0,03	0,15	0,19	0,8	(0,5-1,26)	0,54
imputé	rs2553572	36729213	A	G	0,42	0,40	1,07	(0,90-1,27)	0,41	0,40	0,26	1,92	(1,32-2,8)	1,96E-04
imputé	rs2761338	36730556	T	C	0,42	0,48	0,79	(0,67-0,93)	3,59E-03	0,43	0,50	0,75	(0,53-1,06)	0,05
imputé	rs2787429	36730903	T	C	0,42	0,48	0,79	(0,67-0,93)	3,66E-03	0,43	0,51	0,75	(0,53-1,06)	0,05
imputé	rs2787430	36730916	T	C	0,24	0,20	1,26	(1,03-1,53)	0,04	0,07	0,04	1,7	(0,77-3,71)	0,09
génotypé	rs2553571	36732149	G	A	0,25	0,21	1,24	(1,02-1,51)	0,06	0,05	0,02	2,52	(0,92-6,95)	0,04
génotypé	**rs116909374**	**36738362**	**C**	**T**	**0,05**	**0,03**	**1,97**	**(1,26-3,10)**	**3,25E-03**	**0,01**	**0**	**1,44**	**(0,13-15,96)**	**0,76**

imputé	rs2415329	36740806	T	A	0,33	0,28	1,22	(1,02-1,46)	0,02	-	-	-	-	-
génotypé	rs3762903	36791410	A	G	0,33	0,28	1,23	(1,03-1,47)	0,02	0,10	0,04	2,54	(1,18-5,47)	0,01
imputé	rs8016464	36841338	A	G	0,33	0,36	0,85	(0,72-1,01)	0,04	0,05	0,04	1,1	(0,48-2,53)	0,81
génotypé	rs12147399	36875728	C	A	0,28	0,23	1,26	(1,05-1,53)	0,01	0,38	0,39	0,96	(0,67-1,36)	0,91
imputé	rs4901957	36884085	C	T	0,32	0,28	1,18	(0,99-1,41)	0,04	0,45	0,45	0,98	(0,69-1,38)	0,97
imputé	rs3844017	36887280	T	C	0,32	0,28	1,18	(0,99-1,41)	0,04	0,45	0,45	0,98	(0,69-1,38)	0,97
génotypé	rs3850279	36889182	G	A	0,32	0,28	1,18	(0,98-1,41)	0,04	0,45	0,45	0,98	(0,69-1,38)	0,97
génotypé	rs11156914	36889689	C	T	0,27	0,27	1	(0,83-1,21)	0,85	0,28	0,36	0,68	(0,47-0,99)	0,04
imputé	rs17104442	36890683	A	G	0,26	0,30	0,81	(0,68-0,98)	0,01	0,17	0,22	0,71	(0,46-1,09)	0,13
imputé	rs1951163	36891638	G	A	0,26	0,22	1,22	(1,00-1,47)	0,02	0,23	0,15	1,7	(1,08-2,67)	0,03
imputé	rs7140510	36891663	T	C	0,26	0,31	0,81	(0,67-0,97)	8,66E-03	0,17	0,22	0,73	(0,48-1,12)	0,18
imputé	rs67033133	36892219	A	G	0,26	0,30	0,8	(0,67-0,97)	8,27E-03	0,17	0,22	0,73	(0,48-1,13)	0,18
imputé	rs17104448	36892493	A	G	0,26	0,31	0,81	(0,67-0,97)	8,68E-03	0,17	0,22	0,73	(0,48-1,13)	0,19
imputé	rs17765274	36893215	G	T	0,24	0,28	0,81	(0,67-0,98)	9,36E-03	0,04	0,04	1,05	(0,43-2,54)	0,81
imputé	rs56374528	36896846	T	C	0,26	0,31	0,81	(0,67-0,97)	8,61E-03	0,17	0,22	0,74	(0,48-1,13)	0,19
imputé	rs59266981	36896854	G	T	0,26	0,31	0,81	(0,67-0,97)	8,61E-03	0,17	0,22	0,74	(0,48-1,13)	0,19
imputé	rs3909616	36896963	C	G	0,43	0,37	1,27	(1,07-1,50)	1,98E-03	0,47	0,47	1,01	(0,71-1,42)	0,85
imputé	rs67169817	36897503	C	T	0,28	0,33	0,79	(0,66-0,94)	2,94E-03	0,18	0,23	0,74	(0,48-1,13)	0,19
imputé	rs12433390	36897744	A	C	0,26	0,30	0,81	(0,67-0,97)	8,56E-03	0,17	0,22	0,74	(0,48-1,13)	0,19
imputé	rs12433472	36897747	G	A	0,26	0,30	0,81	(0,67-0,97)	8,56E-03	0,17	0,22	0,74	(0,48-1,13)	0,19
imputé	rs12891306	36897836	A	G	0,43	0,37	1,27	(1,07-1,50)	1,98E-03	0,47	0,47	1,01	(0,71-1,42)	0,85
imputé	rs56237712	36899888	C	T	0,26	0,30	0,81	(0,67-0,97)	8,54E-03	0,17	0,22	0,74	(0,48-1,13)	0,19
imputé	rs11156915	36900583	C	T	0,26	0,31	0,8	(0,67-0,97)	7,95E-03	0,17	0,22	0,73	(0,48-1,13)	0,18
imputé	rs11156916	36900603	G	A	0,26	0,30	0,81	(0,67-0,97)	8,51E-03	0,17	0,22	0,73	(0,48-1,13)	0,18
imputé	rs11620693	36900642	G	C	0,24	0,28	0,81	(0,67-0,98)	9,30E-03	0,04	0,04	1,05	(0,43-2,54)	0,81
imputé	rs111671641	36901629	A	G	0,28	0,33	0,78	(0,65-0,94)	2,63E-03	0,18	0,22	0,75	(0,49-1,14)	0,20
imputé	rs10132817	36901937	G	A	0,43	0,37	1,27	(1,07-1,50)	1,95E-03	0,47	0,47	1	(0,71-1,42)	0,85
imputé	rs11156917	36903143	T	C	0,26	0,30	0,8	(0,66-0,96)	5,55E-03	0,17	0,22	0,73	(0,48-1,13)	0,18
génotypé	rs12894630	36903354	C	G	0,44	0,38	1,28	(1,08-1,51)	1,46E-03	0,47	0,48	0,96	(0,68-1,35)	0,62
imputé	rs17175660	36904006	G	A	0,26	0,30	0,79	(0,66-0,96)	4,87E-03	0,17	0,22	0,73	(0,48-1,13)	0,18
imputé	rs112872564	36904182	A	G	0,26	0,30	0,8	(0,66-0,96)	5,84E-03	0,17	0,22	0,74	(0,48-1,13)	0,19
imputé	rs112864555	36904224	T	A	0,26	0,30	0,8	(0,66-0,96)	5,84E-03	0,17	0,22	0,74	(0,48-1,13)	0,19
imputé	rs3909604	36905021	A	G	0,26	0,31	0,8	(0,67-0,97)	8,33E-03	0,18	0,22	0,74	(0,48-1,13)	0,19
imputé	rs6571748	36905646	T	C	0,26	0,31	0,8	(0,67-0,97)	8,47E-03	0,18	0,22	0,74	(0,48-1,14)	0,20
imputé	rs7149799	36905960	G	T	0,24	0,28	0,82	(0,68-0,99)	0,01	0,17	0,22	0,73	(0,47-1,12)	0,17
imputé	rs10483464	36906144	C	A	0,28	0,33	0,78	(0,65-0,94)	2,87E-03	0,18	0,23	0,74	(0,49-1,14)	0,20
imputé	rs11628001	36906292	C	A	0,24	0,28	0,81	(0,67-0,98)	9,60E-03	0,04	0,04	1,07	(0,45-2,58)	0,74
imputé	rs112551702	36909568	G	A	0,28	0,33	0,78	(0,65-0,94)	2,81E-03	0,18	0,22	0,74	(0,48-1,14)	0,21
imputé	rs10444702	36910545	G	A	0,45	0,40	1,24	(1,05-1,46)	3,90E-03	0,61	0,66	0,8	(0,56-1,15)	0,16
imputé	rs11626643	36910704	G	A	0,24	0,29	0,81	(0,67-0,97)	7,06E-03	0,04	0,04	1,07	(0,45-2,55)	0,73
génotypé	rs3850281	36912537	T	C	0,26	0,31	0,8	(0,66-0,96)	4,67E-03	0,05	0,05	0,98	(0,44-2,17)	0,91
imputé	rs11851850	36912658	G	C	0,42	0,37	1,26	(1,06-1,49)	2,95E-03	0,48	0,48	1,01	(0,71-1,42)	0,85
imputé	rs17104494	36913825	T	C	0,26	0,31	0,8	(0,66-0,96)	4,43E-03	0,05	0,04	1,12	(0,49-2,54)	0,65

imputé	rs72677904	36915171	G	A	0,25	0,29	0,81	(0,67-0,98)	7,67E-03	0,04	0,04	1,08	(0,44-2,62)	0,73
imputé	rs72677905	36915177	G	A	0,23	0,26	0,83	(0,69-1,01)	0,02	0,04	0,04	1,08	(0,44-2,62)	0,73
imputé	rs17104500	36915469	C	T	0,25	0,29	0,81	(0,67-0,98)	7,94E-03	0,04	0,04	1,07	(0,45-2,54)	0,74
imputé	rs3850282	36916334	C	T	0,22	0,26	0,83	(0,69-1,01)	0,02	0,04	0,04	1,04	(0,44-2,47)	0,81
imputé	rs75569493	36916520	A	C	0,25	0,29	0,81	(0,67-0,98)	7,71E-03	0,05	0,04	1,05	(0,46-2,41)	0,78
imputé	rs3850284	36916521	C	T	0,24	0,28	0,82	(0,68-0,99)	0,01	0,04	0,04	1,02	(0,43-2,44)	0,81
imputé	rs8004126	36916860	G	A	0,25	0,29	0,81	(0,67-0,98)	7,73E-03	0,05	0,04	1,07	(0,46-2,47)	0,73
imputé	rs8008525	36917230	A	G	0,26	0,31	0,8	(0,66-0,96)	4,44E-03	0,05	0,04	1,13	(0,49-2,59)	0,64
imputé	rs8009476	36917462	G	A	0,24	0,29	0,81	(0,67-0,98)	7,96E-03	0,04	0,04	1,05	(0,44-2,47)	0,79
imputé	rs17104524	36921330	T	C	0,23	0,26	0,82	(0,68-1,00)	0,02	-	-	-	-	-
imputé	rs10047808	36925048	G	A	0,45	0,40	1,24	(1,04-1,46)	3,77E-03	0,61	0,66	0,84	(0,59-1,2)	0,23
imputé	rs12882950	36931357	C	T	0,45	0,40	1,24	(1,05-1,46)	3,19E-03	0,61	0,65	0,84	(0,59-1,2)	0,24
imputé	rs1547048	36939259	G	A	0,30	0,36	0,79	(0,66-0,94)	4,22E-03	0,32	0,29	1,19	(0,82-1,74)	0,25
imputé	rs1956956	36939663	G	A	0,30	0,35	0,8	(0,67-0,95)	5,90E-03	0,32	0,29	1,19	(0,82-1,73)	0,25
imputé	rs8015629	36941068	T	A	0,45	0,39	1,25	(1,06-1,48)	2,80E-03	0,49	0,48	1,04	(0,74-1,47)	0,96
génotypé	rs4902397	36941142	G	C	0,21	0,25	0,81	(0,66-0,98)	0,02	0,05	0,05	1,12	(0,51-2,43)	0,66
imputé	rs12436124	36941269	G	A	0,24	0,28	0,81	(0,67-0,97)	6,22E-03	-	-	-	-	-
génotypé	rs7156370	36945922	A	G	0,28	0,26	1,06	(0,88-1,28)	0,43	0,39	0,28	1,68	(1,16-2,43)	9,55E-03
imputé	rs3825634	36946617	G	T	0,28	0,26	1,12	(0,93-1,35)	0,17	0,39	0,29	1,54	(1,07-2,23)	9,47E-03
imputé	rs927286	36947756	G	C	0,30	0,27	1,11	(0,92-1,33)	0,19	0,41	0,32	1,47	(1,03-2,12)	0,02
imputé	rs1951157	36951814	T	A	0,28	0,26	1,1	(0,91-1,32)	0,24	0,39	0,31	1,46	(1,02-2,1)	0,02
imputé	rs4902467	36959027	A	G	0,27	0,24	1,12	(0,93-1,36)	0,18	0,37	0,28	1,5	(1,03-2,17)	0,02
imputé	rs7144500	36959783	A	C	0,28	0,26	1,12	(0,93-1,36)	0,18	0,39	0,29	1,55	(1,07-2,24)	0,01
imputé	rs3825635	36960618	A	T	0,32	0,30	1,12	(0,94-1,34)	0,18	0,48	0,38	1,45	(1,02-2,05)	0,04
génotypé	rs11845120	36961489	G	A	0,33	0,30	1,13	(0,94-1,35)	0,17	0,47	0,38	1,45	(1,02-2,05)	0,04
génotypé	rs17765618	36969462	C	A	0,29	0,24	1,24	(1,03-1,50)	0,01	-	-	-	-	-
génotypé	rs999460	36984511	C	T	0,34	0,39	0,8	(0,68-0,96)	5,18E-03	0,21	0,19	1,14	(0,74-1,76)	0,59

Les ORs et IC ont été ajustés sur l'âge, le sexe et la région de résidence ; les SNP sont classés selon leurs positions sur le chromosome

Annexe 8 : P value d'association entre les gènes et le risque de CDT dans l'étude CATHY/Young-thyr et interaction avec les facteurs de risque d'intérêt. Les gènes sont classé selon la plus petite p-value observée pour l'effet propre des gènes

		effet propre		Alcool		statut tabagique		tabac (pq-année)		IMC		pilule		Age première règles		nb grossesse à terme	
Gene	**N.SNP**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**	**pvalue**	**FDR**
CYP2F1	7	2.37E-02	0.92	8.74E-01	1.00	2.02E-01	0.99	2.80E-03	0.09	7.28E-01	1.00	2.94E-01	1.00	8.24E-01	1.00	3.50E-01	1.00
UGT1A8	69	4.77E-02	0.92	1.00E+00	1.00	4.15E-01	1.00	2.16E-02	0.22	8.08E-01	1.00	9.95E-01	1.00	9.00E-01	1.00	1.00E+00	1.00
MGST1	10	5.13E-02	0.92	5.62E-02	0.92	8.19E-01	1.00	2.12E-01	0.71	6.00E-03	0.38	1.00E+00	1.00	2.96E-02	0.64	4.61E-01	1.00
GSTA2	6	5.38E-02	0.92	9.99E-01	1.00	5.83E-01	1.00	4.51E-01	0.88	7.99E-01	1.00	6.26E-02	0.58	9.82E-01	1.00	9.86E-01	1.00
CYP2E1	9	5.96E-02	0.92	2.16E-01	1.00	7.22E-01	1.00	2.34E-01	0.72	9.39E-01	1.00	8.04E-01	1.00	8.14E-01	1.00	7.68E-01	1.00
MGST2	26	6.00E-02	0.92	5.46E-02	0.92	5.28E-01	1.00	1.62E-01	0.59	9.96E-01	1.00	3.38E-01	1.00	8.42E-01	1.00	8.58E-01	1.00
CYP24A1	27	7.74E-02	0.92	1.00E-04	0.01	4.42E-01	1.00	2.97E-01	0.76	9.99E-01	1.00	3.94E-01	1.00	6.07E-02	0.78	1.00E+00	1.00
CYP11A1	8	8.36E-02	0.92	1.42E-01	1.00	4.96E-01	1.00	4.46E-01	0.88	1.00E+00	1.00	1.00E+00	1.00	4.00E-04	0.05	6.87E-02	1.00
ADH6	8	8.47E-02	0.92	9.91E-01	1.00	1.73E-02	0.28	2.43E-02	0.22	8.03E-01	1.00	3.32E-01	1.00	1.71E-01	1.00	1.00E+00	1.00
AKR1C2	16	1.04E-01	0.92	5.46E-01	1.00	6.98E-01	1.00	6.84E-01	0.99	3.19E-01	1.00	3.08E-01	1.00	7.78E-01	1.00	9.79E-01	1.00
GSTZ1	10	1.17E-01	0.92	7.15E-01	1.00	3.39E-01	1.00	8.47E-01	0.99	9.08E-01	1.00	3.53E-01	1.00	9.58E-01	1.00	4.34E-01	1.00
ADH1A	3	1.18E-01	0.92	9.62E-01	1.00	5.13E-01	1.00	8.65E-02	0.44	6.97E-01	1.00	6.24E-01	1.00	3.76E-01	1.00	9.59E-01	1.00
HSD17B6	4	1.26E-01	0.92	8.86E-01	1.00	8.01E-01	1.00	9.74E-01	0.99	6.77E-01	1.00	8.14E-01	1.00	8.60E-01	1.00	7.70E-01	1.00
ADH5	7	1.27E-01	0.92	9.90E-01	1.00	1.78E-02	0.28	8.56E-02	0.44	9.75E-01	1.00	5.58E-02	0.58	5.76E-01	1.00	9.99E-01	1.00
HSD17B12	20	1.28E-01	0.92	5.64E-01	1.00	8.56E-01	1.00	5.64E-01	0.99	6.24E-01	1.00	4.68E-01	1.00	8.07E-01	1.00	1.94E-01	1.00
NAT2	11	1.50E-01	0.92	2.23E-02	0.92	3.97E-01	1.00	3.56E-01	0.82	4.40E-02	0.68	7.80E-01	1.00	9.00E-01	1.00	9.99E-01	1.00
APOA4	3	1.54E-01	0.92	7.11E-01	1.00	9.13E-01	1.00	7.10E-03	0.13	1.35E-01	0.87	9.29E-01	1.00	1.38E-01	1.00	6.01E-01	1.00
APOC3	3	1.56E-01	0.92	9.72E-01	1.00	6.04E-01	1.00	3.50E-03	0.09	9.41E-01	1.00	9.17E-01	1.00	1.23E-01	1.00	4.54E-01	1.00
ADH4	7	1.64E-01	0.92	9.59E-01	1.00	1.90E-03	0.12	6.30E-03	0.13	9.66E-01	1.00	1.25E-01	0.70	5.37E-01	1.00	9.77E-01	1.00
APOA2	8	1.71E-01	0.92	9.66E-01	1.00	2.90E-01	1.00	2.35E-01	0.72	9.00E-01	1.00	7.35E-01	1.00	6.28E-01	1.00	8.49E-01	1.00
AKR1B10	4	1.82E-01	0.92	7.21E-01	1.00	6.00E-02	0.59	1.05E-01	0.46	3.00E-01	1.00	8.82E-01	1.00	4.61E-01	1.00	9.30E-01	1.00
CYP1B1	8	1.85E-01	0.92	8.79E-01	1.00	1.14E-01	0.85	8.02E-02	0.44	4.28E-01	1.00	8.52E-01	1.00	8.80E-01	1.00	9.97E-01	1.00
CYP2C9	18	1.91E-01	0.92	9.52E-01	1.00	1.27E-01	0.86	9.80E-03	0.14	9.93E-01	1.00	6.39E-02	0.58	8.03E-01	1.00	5.88E-01	1.00
ALDH3B1	7	1.96E-01	0.92	3.33E-02	0.92	1.56E-01	0.90	3.65E-01	0.82	1.13E-01	0.87	5.59E-01	1.00	9.10E-01	1.00	9.23E-01	1.00
HSD17B3	16	2.00E-01	0.92	1.78E-01	1.00	7.27E-01	1.00	4.54E-01	0.88	1.64E-01	0.87	1.81E-01	0.86	9.09E-01	1.00	9.93E-01	1.00
GSTA1	2	2.04E-01	0.92	5.03E-01	1.00	6.15E-01	1.00	9.93E-02	0.46	4.78E-01	1.00	9.56E-02	0.68	6.98E-01	1.00	8.24E-01	1.00
GSTP1	5	2.04E-01	0.92	7.76E-01	1.00	9.00E-01	1.00	8.21E-01	0.99	4.57E-01	1.00	3.85E-01	1.00	8.92E-01	1.00	9.19E-01	1.00
CYP2D6	4	2.34E-01	0.92	9.73E-01	1.00	4.79E-01	1.00	2.84E-01	0.76	8.03E-01	1.00	7.22E-02	0.60	3.23E-01	1.00	8.71E-01	1.00
ESR2	15	2.36E-01	0.92	8.91E-01	1.00	5.17E-01	1.00	8.67E-01	0.99	9.99E-01	1.00	9.90E-01	1.00	7.24E-01	1.00	8.62E-01	1.00

HSD17B8	2	2.45E-01	0.92	7.44E-01	1.00	8.59E-01	1.00	8.14E-01	0.99	7.21E-02	0.71	7.15E-01	1.00	8.00E-04	0.05	8.33E-01	1.00
LHCGR	44	2.47E-01	0.92	7.80E-01	1.00	8.22E-01	1.00	9.43E-01	0.99	9.76E-01	1.00	9.93E-01	1.00	9.90E-01	1.00	3.18E-01	1.00
AKR1C3	8	2.57E-01	0.92	4.84E-01	1.00	1.62E-01	0.90	1.67E-01	0.59	4.13E-01	1.00	9.57E-01	1.00	3.99E-01	1.00	9.66E-01	1.00
HSD11B1	5	2.62E-01	0.92	5.31E-01	1.00	9.04E-01	1.00	7.20E-01	0.99	8.39E-01	1.00	8.03E-01	1.00	6.24E-01	1.00	4.94E-01	1.00
STS	6	2.72E-01	0.92	9.09E-01	1.00	5.13E-01	1.00	3.41E-01	0.82	3.86E-01	1.00	9.19E-01	1.00	5.37E-02	0.76	3.97E-01	1.00
POMC	7	2.75E-01	0.92	5.74E-01	1.00	7.78E-01	1.00	9.78E-01	0.99	1.51E-02	0.52	9.19E-01	1.00	1.94E-01	1.00	7.45E-01	1.00
ALDH1A3	6	2.77E-01	0.92	5.04E-01	1.00	9.43E-01	1.00	6.53E-01	0.99	9.95E-01	1.00	8.27E-01	1.00	9.37E-01	1.00	6.06E-01	1.00
GSTM5	3	2.83E-01	0.92	6.04E-01	1.00	6.95E-01	1.00	8.89E-01	0.99	5.31E-01	1.00	9.93E-01	1.00	8.65E-01	1.00	9.62E-01	1.00
LHB	1	2.85E-01	0.92	3.35E-01	1.00	5.90E-01	1.00	6.80E-02	0.44	7.40E-01	1.00	2.90E-03	0.15	8.67E-01	1.00	8.45E-01	1.00
SULT2A1	7	2.86E-01	0.92	2.14E-01	1.00	5.11E-01	1.00	1.23E-01	0.49	9.90E-01	1.00	6.78E-01	1.00	4.63E-02	0.74	9.49E-01	1.00
HSD17B4	13	2.91E-01	0.92	4.90E-01	1.00	7.66E-01	1.00	7.60E-01	0.99	9.33E-01	1.00	5.63E-01	1.00	7.00E-01	1.00	4.31E-01	1.00
GSTM3	5	2.97E-01	0.92	9.89E-01	1.00	8.70E-01	1.00	4.28E-01	0.88	4.06E-01	1.00	5.73E-01	1.00	8.83E-01	1.00	6.52E-01	1.00
ALDH3B2	3	3.03E-01	0.92	3.87E-01	1.00	3.24E-02	0.46	7.64E-02	0.44	2.43E-02	0.52	8.36E-01	1.00	2.84E-01	1.00	6.09E-01	1.00
NR3C1	18	3.18E-01	0.94	9.59E-01	1.00	8.20E-01	1.00	5.96E-01	0.99	3.34E-01	1.00	1.59E-01	0.78	9.22E-01	1.00	9.95E-01	1.00
ADH1B	11	3.38E-01	0.94	8.81E-01	1.00	1.19E-01	0.85	4.01E-01	0.84	9.26E-01	1.00	6.16E-01	1.00	4.97E-01	1.00	9.99E-01	1.00
AKR1B1	11	3.49E-01	0.94	8.98E-01	1.00	9.66E-01	1.00	9.76E-01	0.99	8.85E-01	1.00	9.95E-01	1.00	1.00E+00	1.00	9.49E-01	1.00
CYP2C19	14	3.50E-01	0.94	3.49E-01	1.00	7.33E-01	1.00	5.82E-01	0.99	9.14E-01	1.00	1.32E-02	0.34	1.91E-01	1.00	9.97E-01	1.00
SRD5A2	7	3.51E-01	0.94	9.98E-01	1.00	2.00E-04	0.03	3.10E-03	0.09	9.40E-01	1.00	5.72E-02	0.58	3.34E-01	1.00	9.72E-01	1.00
CYP19A1	44	3.62E-01	0.94	9.20E-01	1.00	5.72E-02	0.59	3.01E-01	0.76	4.84E-01	1.00	5.84E-01	1.00	4.12E-01	1.00	1.00E+00	1.00
GSTM2	3	3.71E-01	0.94	5.48E-01	1.00	7.42E-01	1.00	7.93E-01	0.99	6.39E-01	1.00	9.05E-01	1.00	4.23E-01	1.00	9.68E-01	1.00
CYP2A6	4	3.80E-01	0.94	4.70E-01	1.00	6.67E-01	1.00	9.44E-01	0.99	5.87E-01	1.00	1.20E-01	0.70	1.71E-01	1.00	8.09E-01	1.00
CYP2R1	11	3.80E-01	0.94	5.80E-01	1.00	8.07E-01	1.00	9.00E-01	0.99	8.28E-01	1.00	9.98E-01	1.00	5.85E-01	1.00	3.10E-01	1.00
DHDH	6	3.81E-01	0.94	5.41E-01	1.00	3.83E-01	1.00	1.15E-02	0.15	2.10E-01	0.97	3.22E-01	1.00	7.82E-01	1.00	5.78E-01	1.00
MGST3	17	4.01E-01	0.94	6.68E-01	1.00	6.76E-01	1.00	8.13E-01	0.99	7.75E-01	1.00	9.11E-01	1.00	3.26E-01	1.00	7.25E-01	1.00
CYP2S1	12	4.03E-01	0.94	7.29E-01	1.00	9.94E-01	1.00	8.88E-01	0.99	9.99E-01	1.00	5.11E-01	1.00	7.36E-01	1.00	9.78E-01	1.00
CYP2C8	17	4.14E-01	0.94	9.99E-01	1.00	5.30E-02	0.59	6.40E-01	0.99	7.18E-01	1.00	7.57E-01	1.00	9.58E-01	1.00	5.47E-02	1.00
GSTM4	6	4.36E-01	0.94	8.12E-01	1.00	8.25E-01	1.00	6.21E-01	0.99	6.26E-01	1.00	8.73E-01	1.00	1.91E-01	1.00	9.85E-01	1.00
CYP3A43	6	4.37E-01	0.94	9.77E-01	1.00	4.64E-02	0.59	9.99E-02	0.46	8.49E-01	1.00	2.28E-01	0.94	9.15E-01	1.00	2.79E-01	1.00
CYP46A1	12	4.37E-01	0.94	7.16E-01	1.00	1.03E-01	0.82	1.82E-02	0.21	1.28E-01	0.87	8.05E-01	1.00	2.99E-02	0.64	7.46E-01	1.00
SULT4A1	12	4.37E-01	0.94	6.45E-01	1.00	2.12E-01	0.99	8.39E-01	0.99	7.78E-01	1.00	8.17E-01	1.00	8.94E-01	1.00	9.83E-01	1.00
UGT2B7	5	4.46E-01	0.94	7.13E-01	1.00	3.48E-01	1.00	6.18E-01	0.99	3.36E-02	0.61	7.45E-02	0.60	8.98E-01	1.00	2.24E-01	1.00
FSHR	47	4.51E-01	0.94	9.99E-01	1.00	3.13E-01	1.00	9.10E-03	0.14	2.65E-01	0.99	8.83E-01	1.00	1.00E+00	1.00	8.06E-01	1.00
CYP3A4	3	4.54E-01	0.94	9.39E-01	1.00	7.59E-01	1.00	6.26E-01	0.99	1.72E-02	0.52	1.26E-01	0.70	3.24E-01	1.00	7.46E-01	1.00
GSTA4	10	4.63E-01	0.94	9.78E-01	1.00	6.47E-01	1.00	1.11E-01	0.46	2.37E-01	0.99	9.99E-01	1.00	9.96E-01	1.00	9.56E-01	1.00

CYP1A2	4	4.84E-01	0.94	9.35E-01	1.00	6.71E-01	1.00	8.01E-01	0.99	2.70E-01	0.99	5.26E-01	1.00	5.71E-01	1.00	1.89E-01	1.00
UGT2A1	30	4.86E-01	0.94	9.98E-01	1.00	9.34E-01	1.00	8.23E-01	0.99	1.29E-01	0.87	9.35E-01	1.00	7.40E-01	1.00	3.61E-01	1.00
MC2R	6	5.09E-01	0.94	3.78E-01	1.00	1.54E-01	0.90	7.48E-01	0.99	4.14E-01	1.00	6.72E-01	1.00	9.07E-02	0.97	9.94E-01	1.00
APOB	8	5.10E-01	0.94	5.74E-02	0.92	5.52E-01	1.00	2.36E-01	0.72	2.64E-01	0.99	3.10E-01	1.00	7.52E-01	1.00	2.18E-01	1.00
SULT1B1	9	5.11E-01	0.94	3.98E-01	1.00	7.20E-01	1.00	9.06E-01	0.99	9.98E-01	1.00	1.49E-01	0.78	8.52E-01	1.00	9.70E-01	1.00
NAT1	10	5.25E-01	0.94	8.02E-01	1.00	9.60E-01	1.00	9.14E-01	0.99	8.25E-01	1.00	9.66E-01	1.00	6.66E-01	1.00	9.26E-01	1.00
AKR1C4	13	5.26E-01	0.94	1.29E-01	1.00	1.30E-02	0.28	5.76E-02	0.43	8.18E-01	1.00	6.98E-01	1.00	4.21E-01	1.00	9.39E-01	1.00
GSTK1	1	5.32E-01	0.94	1.05E-02	0.67	8.23E-01	1.00	7.33E-01	0.99	8.18E-01	1.00	5.51E-01	1.00	2.80E-01	1.00	9.13E-01	1.00
HSD17B13	10	5.35E-01	0.94	9.36E-01	1.00	5.47E-01	1.00	9.69E-01	0.99	9.03E-01	1.00	6.06E-01	1.00	7.78E-01	1.00	1.00E+00	1.00
CYP3A7	3	5.36E-01	0.94	8.40E-01	1.00	9.58E-01	1.00	8.62E-01	0.99	3.80E-03	0.38	2.21E-01	0.94	2.08E-01	1.00	6.83E-01	1.00
CYP2B6	16	5.45E-01	0.94	9.10E-01	1.00	2.79E-01	1.00	4.33E-01	0.88	9.87E-01	1.00	6.95E-01	1.00	5.44E-01	1.00	4.52E-01	1.00
SULT1E1	10	5.71E-01	0.95	1.63E-01	1.00	7.32E-01	1.00	9.58E-01	0.99	7.61E-01	1.00	2.20E-01	0.94	8.33E-01	1.00	6.26E-01	1.00
AR	4	5.73E-01	0.95	6.05E-01	1.00	9.69E-01	1.00	2.16E-01	0.71	1.51E-01	0.87	6.68E-01	1.00	5.91E-01	1.00	8.74E-01	1.00
CYP27B1	2	5.77E-01	0.95	4.49E-01	1.00	2.43E-01	1.00	1.02E-01	0.46	8.24E-01	1.00	1.07E-02	0.34	9.07E-01	1.00	4.45E-01	1.00
SRD5A1	11	5.83E-01	0.95	2.29E-01	1.00	9.69E-01	1.00	9.29E-01	0.99	9.85E-01	1.00	8.76E-01	1.00	7.27E-01	1.00	3.41E-01	1.00
CYP1A1	4	5.84E-01	0.95	9.70E-01	1.00	8.02E-01	1.00	6.88E-01	0.99	1.60E-01	0.87	8.44E-02	0.64	1.38E-01	1.00	5.08E-01	1.00
IL4	3	6.05E-01	0.96	8.69E-01	1.00	9.74E-01	1.00	7.99E-01	0.99	6.87E-02	0.71	4.02E-01	1.00	2.25E-01	1.00	2.05E-01	1.00
ASMT	13	6.33E-01	0.96	1.18E-01	1.00	2.98E-01	1.00	1.27E-01	0.49	8.39E-01	1.00	4.30E-01	1.00	7.55E-01	1.00	7.59E-01	1.00
CYP39A1	25	6.36E-01	0.96	9.72E-01	1.00	9.18E-01	1.00	1.56E-01	0.59	7.89E-01	1.00	6.79E-01	1.00	2.97E-02	0.64	8.77E-01	1.00
CYP7B1	20	6.55E-01	0.96	8.74E-01	1.00	7.62E-01	1.00	7.40E-02	0.44	1.00E+00	1.00	5.36E-01	1.00	9.93E-01	1.00	9.99E-01	1.00
AKR1D1	8	6.60E-01	0.96	9.44E-01	1.00	4.87E-01	1.00	6.97E-02	0.44	7.33E-01	1.00	3.50E-03	0.15	6.91E-02	0.80	8.75E-01	1.00
AKR1A1	3	6.61E-01	0.96	8.36E-01	1.00	4.54E-01	1.00	6.50E-01	0.99	2.10E-02	0.52	8.88E-01	1.00	1.48E-01	1.00	5.92E-01	1.00
GNRHR	9	6.61E-01	0.96	8.26E-01	1.00	2.01E-01	0.99	2.41E-02	0.22	5.41E-01	1.00	2.70E-01	1.00	7.57E-01	1.00	6.14E-01	1.00
AHR	10	6.71E-01	0.96	7.56E-01	1.00	9.60E-01	1.00	7.63E-01	0.99	9.20E-01	1.00	9.99E-01	1.00	9.87E-01	1.00	8.84E-01	1.00
SULT2B1	18	6.71E-01	0.96	9.95E-01	1.00	6.91E-01	1.00	3.46E-01	0.82	9.36E-01	1.00	9.34E-01	1.00	1.38E-01	1.00	9.21E-01	1.00
CYP7A1	7	6.76E-01	0.96	1.34E-01	1.00	4.60E-01	1.00	3.03E-01	0.76	1.03E-01	0.87	6.51E-01	1.00	4.81E-01	1.00	5.79E-01	1.00
ADH1C	9	7.14E-01	0.96	7.43E-01	1.00	8.31E-01	1.00	8.25E-01	0.99	8.10E-01	1.00	1.15E-01	0.70	4.88E-01	1.00	9.20E-01	1.00
CYP17A1	7	7.20E-01	0.96	9.70E-01	1.00	3.72E-01	1.00	3.80E-01	0.84	9.05E-01	1.00	9.27E-01	1.00	9.90E-01	1.00	9.96E-01	1.00
SRD5A3	6	7.21E-01	0.96	6.01E-01	1.00	7.20E-01	1.00	6.81E-01	0.99	7.92E-01	1.00	3.29E-01	1.00	9.97E-01	1.00	6.77E-01	1.00
SULT1A1	1	7.41E-01	0.96	7.35E-01	1.00	7.60E-02	0.65	7.41E-02	0.44	1.85E-01	0.94	8.44E-01	1.00	5.87E-01	1.00	6.51E-01	1.00
GSTA5	2	7.44E-01	0.96	9.13E-02	1.00	5.20E-03	0.17	5.00E-04	0.06	3.29E-01	1.00	5.91E-01	1.00	6.23E-01	1.00	4.51E-01	1.00
COMT	13	7.52E-01	0.96	3.78E-02	0.92	5.16E-01	1.00	7.56E-01	0.99	3.50E-01	1.00	8.21E-01	1.00	9.56E-01	1.00	9.38E-01	1.00
NQO1	7	7.53E-01	0.96	9.68E-01	1.00	9.99E-01	1.00	9.20E-01	0.99	8.33E-01	1.00	9.66E-01	1.00	6.96E-01	1.00	7.28E-02	1.00
SHBG	5	7.59E-01	0.96	9.69E-01	1.00	2.22E-01	0.99	3.58E-01	0.82	2.13E-01	0.97	8.72E-01	1.00	5.12E-01	1.00	4.13E-01	1.00

EPHX1	17	7.61E-01	0.96	9.08E-01	1.00	7.68E-01	1.00	8.31E-01	0.99	3.21E-01	1.00	8.01E-01	1.00	9.41E-01	1.00	9.24E-01	1.00
HSD17B1	1	7.64E-01	0.96	8.49E-01	1.00	9.88E-01	1.00	6.91E-01	0.99	6.05E-01	1.00	4.08E-01	1.00	7.42E-01	1.00	9.93E-01	1.00
AKR1C1	8	7.81E-01	0.96	6.42E-01	1.00	4.48E-01	1.00	2.98E-01	0.76	2.82E-01	1.00	1.59E-01	0.78	9.96E-01	1.00	9.51E-01	1.00
ALDH3A1	1	7.84E-01	0.96	4.21E-01	1.00	7.00E-02	0.64	1.08E-01	0.46	2.48E-01	0.99	4.15E-01	1.00	8.83E-01	1.00	2.89E-01	1.00
CYP11B2	5	7.87E-01	0.96	5.16E-01	1.00	1.96E-01	0.99	5.24E-01	0.99	7.32E-01	1.00	9.25E-01	1.00	7.83E-01	1.00	6.89E-01	1.00
CYP27A1	5	7.88E-01	0.96	5.29E-01	1.00	9.79E-01	1.00	9.46E-01	0.99	1.49E-01	0.87	9.35E-01	1.00	9.99E-01	1.00	7.01E-01	1.00
HSD17B7	7	7.88E-01	0.96	8.35E-01	1.00	9.80E-01	1.00	9.04E-01	0.99	2.54E-01	0.99	6.77E-01	1.00	9.48E-01	1.00	3.74E-01	1.00
GSTO1	2	8.02E-01	0.96	9.19E-01	1.00	7.28E-01	1.00	8.08E-01	0.99	5.23E-01	1.00	4.75E-01	1.00	3.14E-01	1.00	9.83E-01	1.00
CGA	5	8.05E-01	0.96	5.27E-01	1.00	4.40E-03	0.17	2.10E-03	0.09	9.57E-01	1.00	8.11E-01	1.00	9.98E-01	1.00	9.97E-01	1.00
ESR1	54	8.06E-01	0.96	9.46E-01	1.00	9.93E-01	1.00	8.43E-01	0.99	1.16E-01	0.87	9.83E-01	1.00	8.90E-01	1.00	1.72E-01	1.00
GSTT2	1	8.23E-01	0.96	5.76E-01	1.00	6.26E-01	1.00	5.64E-01	0.99	4.60E-01	1.00	4.50E-02	0.58	2.87E-02	0.64	5.88E-01	1.00
UGT2B4	8	8.26E-01	0.96	9.89E-01	1.00	2.25E-01	0.99	3.95E-01	0.84	9.53E-01	1.00	4.33E-02	0.58	6.81E-01	1.00	9.93E-01	1.00
UGT2B11	3	8.35E-01	0.96	9.57E-01	1.00	2.90E-01	1.00	2.95E-01	0.76	5.65E-02	0.71	1.40E-03	0.15	6.64E-01	1.00	8.37E-01	1.00
GSTA3	4	8.54E-01	0.96	7.85E-01	1.00	8.57E-01	1.00	7.50E-01	0.99	3.62E-01	1.00	6.93E-01	1.00	1.85E-01	1.00	8.26E-01	1.00
HSD17B2	14	8.55E-01	0.96	1.00E+00	1.00	8.76E-01	1.00	8.60E-01	0.99	2.22E-01	0.98	8.94E-01	1.00	1.00E+00	1.00	7.07E-01	1.00
CYP2C18	12	8.59E-01	0.96	8.77E-01	1.00	7.59E-01	1.00	9.66E-01	0.99	7.71E-01	1.00	2.18E-01	0.94	7.03E-01	1.00	8.12E-01	1.00
CYP11B1	6	8.73E-01	0.96	5.80E-01	1.00	5.01E-01	1.00	5.42E-01	0.99	9.55E-01	1.00	1.00E+00	1.00	7.26E-01	1.00	9.28E-01	1.00
GSTO2	5	8.73E-01	0.96	4.11E-01	1.00	7.64E-01	1.00	7.95E-01	0.99	8.87E-01	1.00	5.30E-01	1.00	2.56E-01	1.00	1.00E+00	1.00
SREBF2	15	8.93E-01	0.96	7.19E-01	1.00	9.46E-01	1.00	6.41E-01	0.99	9.94E-01	1.00	6.89E-01	1.00	9.95E-01	1.00	8.40E-01	1.00
ADH7	10	8.94E-01	0.96	9.89E-01	1.00	6.73E-01	1.00	2.84E-01	0.76	1.91E-01	0.94	2.61E-01	1.00	8.67E-01	1.00	9.82E-01	1.00
SOD2	6	8.95E-01	0.96	9.42E-01	1.00	1.53E-02	0.28	2.62E-01	0.76	9.64E-01	1.00	4.07E-02	0.58	7.11E-01	1.00	5.26E-01	1.00
PPARD	21	9.01E-01	0.96	7.39E-01	1.00	9.25E-01	1.00	1.00E+00	1.00	1.00E+00	1.00	1.07E-01	0.70	4.05E-02	0.74	9.99E-01	1.00
HSD3B2	4	9.02E-01	0.96	8.92E-01	1.00	4.77E-01	1.00	2.85E-01	0.76	3.60E-01	1.00	3.88E-01	1.00	3.39E-01	1.00	4.67E-01	1.00
CYP3A5	2	9.15E-01	0.96	8.36E-01	1.00	2.49E-01	1.00	2.04E-01	0.71	6.80E-02	0.71	2.16E-02	0.46	5.49E-01	1.00	8.47E-01	1.00
UGT2B10	2	9.21E-01	0.96	3.87E-01	1.00	9.31E-01	1.00	9.66E-01	0.99	9.66E-01	1.00	4.96E-01	1.00	5.32E-01	1.00	8.87E-01	1.00
SULT1A2	1	9.28E-01	0.96	1.21E-01	1.00	4.86E-01	1.00	5.11E-02	0.41	4.93E-01	1.00	3.01E-02	0.55	9.15E-01	1.00	1.86E-01	1.00
GNRH1	4	9.29E-01	0.96	3.27E-01	1.00	6.31E-01	1.00	8.76E-01	0.99	9.34E-01	1.00	9.76E-01	1.00	9.49E-01	1.00	5.05E-01	1.00
PGR	14	9.46E-01	0.96	8.19E-01	1.00	5.98E-01	1.00	5.82E-01	0.99	1.49E-01	0.87	5.36E-01	1.00	7.99E-01	1.00	9.47E-01	1.00
LDLR	8	9.48E-01	0.96	2.56E-01	1.00	6.54E-01	1.00	3.22E-01	0.79	8.10E-01	1.00	4.52E-01	1.00	8.23E-01	1.00	3.36E-01	1.00
FSHB	2	9.63E-01	0.97	9.85E-01	1.00	1.52E-01	0.90	4.16E-02	0.35	2.99E-01	1.00	6.77E-01	1.00	8.36E-01	1.00	7.11E-01	1.00
UGT2A3	3	9.99E-01	1.00	6.69E-01	1.00	2.75E-01	1.00	3.93E-01	0.84	4.76E-02	0.68	2.66E-01	1.00	9.57E-01	1.00	9.26E-01	1.00

Annexe 9 : Résultats des p-value des effets propres et des interactions dans l'analyse poolée des études CATHY/Young-Thyr et USRT/UTMDACC

		effet propre		IMC		Alcool		tabac	
Gene	**N.SNP**	**Pvalue**	**FRD**	**Pvalue**	**FRD**	**Pvalue**	**FRD**	**Pvalue**	**FRD**
CYP19A1	44	1.27E-02	0.47	9.12E-01	0.99	1.00E+00	1.00	9.82E-01	1.00
NR3C1	15	1.38E-02	0.47	6.05E-01	0.95	9.47E-01	1.00	7.25E-01	1.00
UGT2B7	5	1.58E-02	0.47	8.30E-03	0.10	1.11E-01	0.87	9.34E-01	1.00
ADH5	7	1.61E-02	0.47	8.08E-01	0.95	1.21E-01	0.87	7.62E-01	1.00
MGST2	22	4.38E-02	0.70	7.69E-01	0.95	2.39E-01	0.87	7.10E-01	1.00
AKR1C3	8	5.82E-02	0.70	4.74E-01	0.93	2.31E-02	0.38	7.43E-02	0.80
ADH4	7	6.03E-02	0.70	9.92E-01	1.00	6.01E-01	0.96	6.00E-01	1.00
HSD17B3	16	6.79E-02	0.70	5.24E-01	0.94	4.81E-01	0.91	3.89E-01	1.00
ALDH3B1	7	7.55E-02	0.70	6.11E-01	0.95	1.52E-02	0.36	3.17E-01	1.00
NQO1	5	8.42E-02	0.70	8.18E-01	0.95	8.83E-01	1.00	9.34E-01	1.00
UGT1A5	35	8.57E-02	0.70	5.40E-03	0.07	3.45E-01	0.88	9.97E-01	1.00
ADH6	8	9.26E-02	0.70	9.69E-01	1.00	2.29E-01	0.87	1.22E-01	1.00
ADH1C	9	1.08E-01	0.70	8.43E-01	0.95	6.10E-01	0.96	9.85E-01	1.00
CYP11A1	8	1.09E-01	0.70	9.63E-01	1.00	2.08E-01	0.87	5.26E-01	1.00
CYP2E1	7	1.15E-01	0.70	9.72E-01	1.00	2.03E-02	0.38	2.98E-01	1.00
UGT1A6	48	1.17E-01	0.70	**3.00E-04**	**0.01**	5.24E-01	0.96	1.00E+00	1.00
GSTO2	4	1.25E-01	0.70	7.16E-01	0.95	8.85E-01	1.00	6.73E-01	1.00
UGT1A7	55	1.30E-01	0.70	**9.00E-04**	**0.02**	6.27E-01	0.96	9.99E-01	1.00
CYP2C9	17	1.31E-01	0.70	8.36E-01	0.95	9.38E-01	1.00	3.68E-02	0.70
CYP2A6	2	1.40E-01	0.70	6.58E-01	0.95	8.76E-01	1.00	5.29E-02	0.70
MGST1	10	1.42E-01	0.70	7.70E-02	0.35	7.40E-03	0.22	2.59E-02	0.61
ESR2	15	1.43E-01	0.70	9.01E-01	0.98	2.17E-01	0.87	6.97E-01	1.00
ADH1A	3	1.44E-01	0.70	8.06E-01	0.95	1.30E-01	0.87	4.21E-01	1.00
IL4	3	1.54E-01	0.70	1.21E-01	0.46	3.55E-01	0.88	8.63E-01	1.00
UGT1A9	60	1.54E-01	0.70	**1.00E-03**	**0.02**	6.64E-01	0.99	9.97E-01	1.00
UGT1A4	30	1.57E-01	0.70	4.60E-03	0.07	2.65E-01	0.87	9.94E-01	1.00
UGT1A10	65	1.65E-01	0.70	**9.00E-04**	**0.02**	6.72E-01	0.99	9.96E-01	1.00
UGT1A8	68	1.73E-01	0.70	**1.00E-03**	**0.02**	4.74E-01	0.91	9.98E-01	1.00
CYP46A1	11	1.74E-01	0.70	7.51E-01	0.95	9.54E-01	1.00	5.17E-02	0.70
AKR1C4	13	1.78E-01	0.70	7.31E-01	0.95	6.50E-03	0.22	6.56E-01	1.00
CYP3A4	1	1.93E-01	0.74	7.22E-01	0.95	1.92E-01	0.87	5.77E-01	1.00
UGT2B4	8	2.02E-01	0.74	4.97E-01	0.94	3.36E-01	0.88	8.33E-01	1.00
NAT2	11	2.08E-01	0.74	7.71E-02	0.35	3.67E-01	0.88	5.64E-01	1.00
CYP27A1	4	2.17E-01	0.75	4.44E-02	0.31	3.66E-01	0.88	9.55E-01	1.00
GSTO1	2	2.29E-01	0.76	5.32E-01	0.94	8.98E-01	1.00	3.98E-01	1.00
AKR1C2	16	2.31E-01	0.76	1.15E-02	0.12	**5.00E-04**	**0.03**	6.21E-01	1.00
EPHX1	15	2.39E-01	0.76	6.05E-02	0.35	5.27E-01	0.96	8.57E-01	1.00
GSTA4	10	2.55E-01	0.78	7.21E-01	0.95	3.28E-01	0.88	1.94E-01	1.00
CYP2S1	10	2.62E-01	0.78	9.95E-01	1.00	3.05E-01	0.88	8.53E-01	1.00
FSHB	1	2.64E-01	0.78	2.51E-01	0.70	7.26E-01	1.00	7.11E-01	1.00
SRD5A2	7	2.76E-01	0.78	6.98E-01	0.95	9.22E-01	1.00	**1.00E-04**	**3.93E-03**
SULT4A1	11	2.78E-01	0.78	5.12E-01	0.94	9.96E-01	1.00	9.75E-01	1.00
CYP2F1	5	2.84E-01	0.78	3.81E-01	0.86	4.45E-01	0.91	2.23E-01	1.00
CYP24A1	27	3.00E-01	0.80	6.86E-01	0.95	5.92E-01	0.96	9.89E-01	1.00
SREBF2	2	3.25E-01	0.82	7.27E-01	0.95	6.01E-01	0.96	7.40E-01	1.00
CYP1B1	8	3.27E-01	0.82	5.19E-01	0.94	7.69E-01	1.00	7.80E-01	1.00
CYP3A5	2	3.30E-01	0.82	1.56E-01	0.55	6.95E-01	0.99	5.52E-01	1.00
NAT1	9	3.36E-01	0.82	8.21E-01	0.95	9.71E-01	1.00	8.91E-01	1.00
UGT2B11	3	3.45E-01	0.82	5.05E-02	0.31	6.45E-01	0.98	5.64E-01	1.00
PGR	13	3.49E-01	0.82	8.49E-02	0.37	5.93E-01	0.96	6.34E-01	1.00
SULT1E1	10	3.63E-01	0.82	4.35E-01	0.88	2.68E-01	0.87	7.88E-01	1.00
DHDH	6	3.65E-01	0.82	1.93E-01	0.64	6.08E-01	0.96	6.31E-02	0.74
GSTM3	4	3.67E-01	0.82	6.31E-02	0.35	9.99E-01	1.00	8.51E-01	1.00
ALDH3A1	1	3.81E-01	0.83	3.93E-01	0.86	5.66E-01	0.96	7.85E-01	1.00
SOD2	5	4.13E-01	0.88	6.39E-01	0.95	4.14E-01	0.89	7.25E-01	1.00
LHCGR	44	4.28E-01	0.88	5.50E-01	0.94	9.26E-01	1.00	1.00E+00	1.00
SULT1A1	1	4.28E-01	0.88	2.54E-01	0.70	2.66E-01	0.87	1.95E-01	1.00
AKR1A1	2	4.52E-01	0.88	3.73E-01	0.86	2.78E-01	0.87	3.54E-01	1.00
MGST3	17	4.56E-01	0.88	5.60E-01	0.94	6.22E-01	0.96	4.55E-01	1.00
HSD17B2	11	4.89E-01	0.88	4.32E-01	0.88	4.68E-01	0.91	9.85E-01	1.00
GSTA5	2	4.90E-01	0.88	7.50E-02	0.35	2.58E-02	0.38	1.18E-01	1.00

ALDH1A3	3	5.13E-01	0.88	1.16E-01	0.46	2.69E-01	0.87	4.56E-01	1.00
COMT	12	5.21E-01	0.88	6.78E-01	0.95	**3.00E-04**	**0.03**	5.37E-01	1.00
LHB	1	5.22E-01	0.88	8.39E-01	0.95	1.22E-01	0.87	6.07E-01	1.00
CYP11B1	6	5.25E-01	0.88	8.33E-01	0.95	1.15E-01	0.87	**1.00E-04**	**3.93E-03**
CYP7A1	7	5.36E-01	0.88	2.20E-01	0.67	3.59E-01	0.88	4.45E-01	1.00
GSTA2	6	5.46E-01	0.88	7.24E-01	0.95	4.29E-01	0.90	8.97E-01	1.00
CYP2C8	16	5.49E-01	0.88	5.21E-01	0.94	4.49E-01	0.91	3.00E-01	1.00
GSTA1	2	5.51E-01	0.88	1.54E-01	0.55	6.80E-01	0.99	9.28E-01	1.00
SULT1B1	9	5.64E-01	0.88	4.22E-01	0.88	5.56E-01	0.96	1.68E-01	1.00
CYP39A1	24	5.78E-01	0.88	3.08E-01	0.76	9.95E-01	1.00	9.06E-01	1.00
POMC	7	5.80E-01	0.88	6.73E-01	0.95	5.18E-01	0.96	6.45E-01	1.00
CYP1A1	2	5.87E-01	0.88	5.64E-01	0.94	8.09E-01	1.00	8.45E-01	1.00
CYP3A43	4	5.94E-01	0.88	2.71E-01	0.71	8.66E-01	1.00	8.05E-01	1.00
APOA2	8	6.03E-01	0.88	9.32E-02	0.39	2.86E-01	0.87	1.64E-01	1.00
UGT2A2	25	6.08E-01	0.88	1.00E-04	0.01	9.16E-01	1.00	9.94E-01	1.00
ADH1B	11	6.17E-01	0.88	8.69E-01	0.96	2.73E-01	0.87	3.12E-01	1.00
AKR1D1	8	6.22E-01	0.88	9.58E-02	0.39	1.89E-01	0.87	3.48E-01	1.00
UGT1A1	22	6.22E-01	0.88	2.59E-02	0.24	2.06E-01	0.87	9.84E-01	1.00
APOB	6	6.23E-01	0.88	1.58E-01	0.55	3.61E-01	0.88	8.54E-01	1.00
UGT1A3	26	6.34E-01	0.88	4.21E-02	0.31	2.81E-01	0.87	9.88E-01	1.00
CYP11B2	5	6.35E-01	0.88	6.29E-01	0.95	7.52E-02	0.87	**1.00E-04**	**3.93E-03**
LDLR	8	6.38E-01	0.88	2.24E-01	0.67	3.69E-01	0.88	9.92E-01	1.00
GNRH1	2	6.38E-01	0.88	2.28E-01	0.67	4.75E-01	0.91	6.89E-01	1.00
UGT2A1	30	6.41E-01	0.88	2.00E-04	0.01	9.54E-01	1.00	9.82E-01	1.00
CYP2C19	13	6.48E-01	0.88	6.64E-01	0.95	9.13E-01	1.00	1.52E-01	1.00
CYP7B1	19	6.50E-01	0.88	1.00E+00	1.00	9.99E-01	1.00	9.95E-01	1.00
CYP2C18	10	6.63E-01	0.89	5.56E-01	0.94	7.45E-01	1.00	1.49E-01	1.00
HSD11B1	5	7.01E-01	0.93	4.47E-01	0.89	6.88E-01	0.99	6.84E-01	1.00
CYP2D6	2	7.31E-01	0.95	3.30E-02	0.26	9.76E-01	1.00	4.83E-01	1.00
SULT1A2	1	7.33E-01	0.95	6.10E-01	0.95	1.23E-01	0.87	1.84E-01	1.00
CYP2B6	16	7.43E-01	0.95	9.91E-01	1.00	9.58E-01	1.00	2.13E-02	0.61
GSTT2	1	7.56E-01	0.95	3.08E-01	0.76	7.40E-01	1.00	5.43E-01	1.00
AHR	9	7.60E-01	0.95	1.96E-01	0.64	1.86E-01	0.87	8.85E-01	1.00
HSD17B12	2	7.70E-01	0.96	2.68E-01	0.71	5.82E-01	0.96	8.46E-01	1.00
GSTM5	3	7.81E-01	0.96	7.02E-02	0.35	8.99E-01	1.00	9.27E-01	1.00
HSD17B1	1	8.16E-01	0.98	7.78E-01	0.95	1.73E-01	0.87	7.81E-01	1.00
HSD17B4	11	8.32E-01	0.98	3.44E-01	0.81	4.10E-01	0.89	6.83E-01	1.00
SHBG	5	8.38E-01	0.98	7.22E-01	0.95	2.70E-01	0.87	2.52E-01	1.00
ALDH3B2	3	8.44E-01	0.98	2.13E-01	0.67	2.27E-01	0.87	4.00E-01	1.00
ESR1	50	8.48E-01	0.98	4.77E-02	0.31	1.77E-01	0.87	3.16E-01	1.00
CYP17A1	7	8.48E-01	0.98	7.24E-02	0.35	4.14E-01	0.89	8.95E-01	1.00
SRD5A1	11	8.62E-01	0.98	6.73E-01	0.95	1.57E-01	0.87	6.18E-01	1.00
CYP1A2	3	8.69E-01	0.98	3.88E-01	0.86	7.52E-01	1.00	9.88E-01	1.00
ADH7	10	8.72E-01	0.98	2.81E-02	0.24	9.61E-01	1.00	9.36E-01	1.00
GSTA3	4	8.81E-01	0.98	2.48E-01	0.70	3.74E-01	0.88	9.32E-01	1.00
SULT2A1	7	9.10E-01	1.00	9.95E-01	1.00	3.85E-01	0.89	9.37E-01	1.00
SULT2B1	17	9.18E-01	1.00	6.45E-01	0.95	9.50E-01	1.00	5.33E-02	0.70
GSTZ1	6	9.35E-01	1.00	8.30E-01	0.95	3.75E-02	0.49	5.70E-01	1.00
GNRHR	9	9.41E-01	1.00	7.92E-01	0.95	5.69E-01	0.96	9.41E-01	1.00
AKR1C1	8	9.56E-01	1.00	4.17E-01	0.88	1.30E-01	0.87	6.63E-01	1.00
GSTM4	6	9.60E-01	1.00	3.43E-01	0.81	8.06E-01	1.00	8.83E-01	1.00
PPARD	14	9.61E-01	1.00	1.00E+00	1.00	7.26E-01	1.00	7.21E-01	1.00
CYP2R1	10	9.75E-01	1.00	8.57E-01	0.95	7.54E-01	1.00	9.94E-01	1.00
CYP27B1	2	9.75E-01	1.00	3.06E-01	0.76	2.22E-01	0.87	4.31E-01	1.00
GSTP1	4	9.87E-01	1.00	1.75E-02	0.17	9.32E-01	1.00	9.95E-01	1.00
GSTM2	3	9.92E-01	1.00	7.92E-01	0.95	7.92E-01	1.00	9.81E-01	1.00
SRD5A3	2	9.98E-01	1.00	5.58E-01	0.94	4.00E-01	0.89	3.89E-01	1.00

Annexes 10 : comparaison des fréquences alléliques entre témoins des populations Européennes des études CATHY et de NC (puce à 96 SNP), cut off =0.008, test effectué avec la méthode proposé par Jon Tyr pour les données de concorsium diff= (p1-p2-0.01)^2/((p1+p2)(2-p1-p2)) avec p1 = MAF étude CATHY et p1= MAF étude NC

Chr	snp	pos_pb	A1	A2	Genes	A_min	CATHY	NC	diff
11	rs1800057	1.08E+08	C	G	ATM	G	0.0304	0.0671	1.18E-02
19	rs11084753	3.43E+07	A	G	KCTD15	A	0.3204	0.4136	1.15E-02
14	rs8018390	3.67E+07	C	T	PTCS3-NKX2-1	T	0.0562	0.1037	1.12E-02
14	rs4902397	3.69E+07	C	G	PTCS3-NKX2-1	C	0.2505	0.1603	9.85E-03
11	rs4752904	4.82E+07	C	G	PTPRJ	G	0.4329	0.5183	9.12E-03
14	rs3850281	3.69E+07	C	T	PTCS3-NKX2-1	C	0.319	0.2256	8.78E-03
16	rs3764340	7.85E+07	C	G	WWOX	G	0.0404	0.0732	8.55E-03
16	rs7498665	2.89E+07	A	G	SH2B1	G	0.3706	0.4512	8.48E-03
14	rs3916179	3.69E+07	C	T	PTCS3-NKX2-1	C	0.4134	0.4939	8.26E-03
14	rs2553576	3.67E+07	C	T	PTCS3-NKX2-1	C	0.4164	0.4939	7.72E-03
8	rs2741205	1.34E+08	A	G	TG,SLA	A	0.2849	0.2012	7.38E-03
14	rs7156370	3.69E+07	A	G	PTCS3-NKX2-1	G	0.2555	0.3232	7.34E-03
19	rs1799782	4.41E+07	C	T	XRCC1	T	0.0671	0.1049	7.27E-03
9	rs1443432	1.01E+08	A	G	FOXE1	G	0.4076	0.3171	7.01E-03
14	rs17765618	3.70E+07	A	C	PTCS3-NKX2-1	A	0.2362	0.2988	6.72E-03
14	rs999460	3.70E+07	A	G	PTCS3-NKX2-1	A	0.4004	0.3171	5.84E-03
14	rs3850279	3.69E+07	A	G	PTCS3-NKX2-1	A	0.2757	0.3354	5.72E-03
14	rs116909374	3.67E+07	C	T	_	T	0.0294	0.001	5.65E-03
14	rs2553577	3.67E+07	A	T	PTCS3-NKX2-1	T	0.4605	0.378	5.40E-03
2	rs966423	2.18E+08	C	T	DIRC3	C	0.4145	0.4756	5.12E-03
2	rs2048722	1.50E+06	C	T	TPO	C	0.4391	0.3598	5.00E-03
14	rs11845120	3.70E+07	A	G	PTCS3-NKX2-1	A	0.2924	0.3476	4.88E-03
2	rs732609	1.50E+06	A	C	TPO	C	0.3998	0.3232	4.80E-03
8	rs2702968	1.34E+08	A	T	TG,SLA	T	0.4853	0.5427	4.55E-03
8	rs2702967	1.34E+08	C	T	TG,SLA	T	0.4853	0.5427	4.55E-03
8	rs1133076	1.34E+08	A	G	TG	A	0.4743	0.5305	4.38E-03
14	rs7158323	3.69E+07	A	C	PTCS3-NKX2-1	A	0.17	0.1159	3.97E-03
5	rs2910164	1.60E+08	C	G	pre-miRNA-146a	C	0.2224	0.2622	3.38E-03
14	rs7159699	3.69E+07	A	T	PTCS3-NKX2-1	T	0.08	0.0427	3.24E-03
16	rs9939609	5.38E+07	A	T	FTO	A	0.43	0.4756	3.12E-03
14	rs3909623	3.69E+07	C	T	PTCS3-NKX2-1	C	0.4274	0.4695	2.74E-03
14	rs12147399	3.69E+07	A	C	PTCS3-NKX2-1	A	0.2298	0.2622	2.42E-03
14	rs1766120	3.67E+07	C	T	PTCS3-NKX2-1	C	0.2127	0.2439	2.41E-03
14	rs7148603	3.67E+07	A	G	PTCS3-NKX2-1	A	0.3539	0.3902	2.29E-03
14	rs3762903	3.68E+07	C	T	PTCS3-NKX2-1	C	0.2781	0.311	2.21E-03
14	rs1766118	3.67E+07	A	C	PTCS3-NKX2-1	A	0.2046	0.2308	1.92E-03
9	rs10759944	1.01E+08	A	G	FOXE1	A	0.3426	0.2927	1.84E-03
14	rs861539	1.04E+08	C	T	XRCC3,KLC1	T	0.3796	0.3293	1.77E-03
9	rs1867277	1.01E+08	A	G	FOXE1	A	0.3895	0.3395	1.73E-03
14	rs12894630	3.69E+07	C	G	PTCS3-NKX2-1	C	0.3696	0.4	1.72E-03
14	rs4903957	8.14E+07	A	G	TSHR	A	0.3594	0.311	1.65E-03
9	rs965513	1.01E+08	A	G	FOXE1	A	0.3398	0.2927	1.59E-03
14	rs4301936	3.67E+07	C	T	PTCS3-NKX2-1	T	0.1878	0.2073	1.37E-03
14	rs7141157	3.69E+07	A	G	PTCS3-NKX2-1	G	0.3527	0.378	1.34E-03
9	rs7870795	1.01E+08	C	T	FOXE1	T	0.2987	0.2561	1.33E-03
14	rs1755768	3.67E+07	A	G	PTCS3-NKX2-1	A	0.2702	0.2927	1.31E-03
9	rs1443434	1.01E+08	G	T	FOXE1	G	0.3861	0.3415	1.29E-03
18	rs17782313	5.79E+07	C	T	MC4R	C	0.2209	0.2407	1.25E-03
14	rs2553571	3.67E+07	C	T	PTCS3-NKX2-1	T	0.2142	0.1768	1.19E-03
14	rs17764769	3.67E+07	G	T	PTCS3-NKX2-1	T	0.1213	0.1341	1.17E-03
19	rs25487	4.41E+07	A	G	XRCC1	A	0.3552	0.378	1.16E-03
8	rs9942754	1.34E+08	C	T	TG,SLA	T	0.0101	0.0061	1.12E-03
14	rs12879339	3.67E+07	A	C	PTCS3-NKX2-1	A	0.102	0.0732	1.11E-03
14	rs17461837	3.66E+07	C	T	PTCS3-NKX2-1	T	0.0781	0.0854	9.97E-04

8	rs822318	1.64E+07	C	T	_	C	0.4724	0.4939	9.93E-04
8	rs6983267	1.28E+08	G	T	gene desert	G	0.4871	0.5061	8.41E-04
14	rs2787431	3.67E+07	G	T	PTCS3-NKX2-1	T	0.0588	0.0366	8.19E-04
14	rs17462372	3.67E+07	C	T	PTCS3-NKX2-1	C	0.0414	0.0427	7.92E-04
14	rs11156914	3.69E+07	C	T	PTCS3-NKX2-1	T	0.2656	0.2805	7.81E-04
14	rs1952707	3.67E+07	C	T	PTCS3-NKX2-1	C	0.3253	0.3415	7.72E-04
8	rs17695552	1.34E+08	C	T	TG,SLA	C	0.2246	0.2378	7.57E-04
8	rs1445493	8.98E+06	A	G	_	A	0.3646	0.3293	7.06E-04
14	rs7140782	3.67E+07	C	T	PTCS3-NKX2-1	T	0.1112	0.1159	5.37E-04
4	rs13130484	4.52E+07	C	T	_	T	0.4347	0.4024	5.11E-04
9	rs1443433	1.01E+08	C	T	FOXE1	C	0.1535	0.128	4.97E-04
9	rs4743136	1.01E+08	C	G	FOXE1	G	0.147	0.1524	4.66E-04
14	rs1177590	3.66E+07	C	T	PTCS3-NKX2-1	C	0.2885	0.2593	4.63E-04
14	rs3861479	3.68E+07	A	G	PTCS3-NKX2-1	G	0.0552	0.0549	4.52E-04
14	rs12587845	3.69E+07	A	G	PTCS3-NKX2-1	G	0.2482	0.2561	4.25E-04
14	rs1766142	3.67E+07	A	C	PTCS3-NKX2-1	A	0.4449	0.4146	4.20E-04
14	rs2415328	3.69E+07	C	G	PTCS3-NKX2-1	C	0.2311	0.2378	3.88E-04
14	rs17175332	3.68E+07	A	G	PTCS3-NKX2-1	G	0.4733	0.4451	3.33E-04
11	rs3092993	1.08E+08	A	C	ATM	A	0.1262	0.128	3.14E-04
18	rs12970134	5.79E+07	A	G	MC4R	A	0.2542	0.2593	2.99E-04
14	rs4898569	3.68E+07	A	C	PTCS3-NKX2-1	A	0.057	0.0549	2.95E-04
11	rs1801516	1.08E+08	A	G	ATM	A	0.1268	0.128	2.82E-04
5	rs6885099	7.65E+07	A	G	PDE8B	G	0.4337	0.439	2.38E-04
9	rs10217225	1.01E+08	C	T	FOXE1	T	0.093	0.0915	2.16E-04
2	rs6548238	6.35E+05	C	T	TMEM18	T	0.1728	0.1524	1.99E-04
11	rs189037	1.08E+08	A	G	ATM	G	0.4156	0.3951	1.14E-04
11	rs10838738	4.77E+07	A	G	MTCH2	G	0.3603	0.3598	9.79E-05
1	rs2815752	7.28E+07	C	T	NEGR1	C	0.3293	0.311	7.91E-05
14	rs1169134	3.66E+07	C	T	PTCS3-NKX2-1	C	0.3072	0.2901	6.02E-05
8	rs2252696	1.34E+08	A	C	TG,SLA	C	0.4844	0.4817	5.34E-05
9	rs7871887	1.01E+08	C	T	FOXE1	C	0.1553	0.1402	5.16E-05
8	rs2439302	3.24E+07	C	G	NRG1	C	0.4605	0.4451	2.94E-05
14	rs1766145	3.67E+07	A	T	PTCS3-NKX2-1	T	0.3226	0.3171	2.33E-05
17	rs1042522	7.58E+06	C	G	TP53	C	0.2592	0.2531	2.00E-05
14	rs1169146	3.66E+07	A	G	PTCS3-NKX2-1	G	0.4144	0.4085	1.74E-05
14	rs944289	3.66E+07	C	T	NKX2-1 (TFF1)	C	0.4145	0.4085	1.65E-05
14	rs1169151	3.66E+07	A	G	PTCS3-NKX2-1	G	0.4138	0.4074	1.34E-05
14	rs2415317	3.66E+07	A	G	PTCS3-NKX2-1	G	0.4108	0.4024	2.65E-06
14	rs17104528	3.69E+07	C	T	PTCS3-NKX2-1	C	0.0717	0.061	1.98E-06
9	rs10818094	1.01E+08	A	G	FOXE1	A	0.2164	0.2073	1.21E-06
11	rs664677	1.08E+08	C	T	ATM	C	0.4053	0.3963	1.04E-06
11	rs10767664	2.77E+07	A	T	BDNF	T	0.2297	0.2195	5.74E-08

Annexes 11 : comparaison fréquences nombre de copies des gènes GSTM1 et GSTT1 chez les témoins d'origine européennes des études CATHY et NC

	GSTM1				**GSTT1**			
	CATHY		**NC**		**CATHY**		**NC**	
CNV	**count**	**freq**	**count**	**freq**	**count**	**freq**	**count**	**freq**
0	251	0.50	46	0.56	104	0.20	12	0.16
1	202	0.40	26	0.32	224	0.42	29	0.39
2/3	50	0.10	10	0.12	202	0.38	34	0.45
total	**573**		**82**	**1**	**530**	**1**	**75**	**1**

GSTM1 : Khi 2 = 2.17, df=2 ; p=0.33; non significatif ; ***GSTM1*** : Khi 2 = 1.53, df=2 ; p=0.46; non significatif

Annexe article

IJC International Journal of Cancer

Fine-Mapping of two differentiated thyroid carcinoma susceptibility loci at 9q22.33 and 14q13.3 detects novel candidate functional SNPs in Europeans from metropolitan France and Melanesians from New Caledonia

Journal:	*International Journal of Cancer*
Manuscript ID	IJC-15-2347.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	05-Feb-2016
Complete List of Authors:	Catherine, Tcheandjieu; Université Paris-Saclay, Université Paris-sud, UVSQ, Inserm, CESP (center for research in epidemiology and population health) Lesueur, Fabienne; Curie Institut,Mines Paris Tech, Inserm U 900 Sanchez, Marie; Université Paris-Saclay, Université Paris-sud, UVSQ, Inserm, CESP (center for research in epidemiology and population health) Baron_dubourdieu, Dominique; laboratory of Anatomy and cytopathology Guizard, Anne-Valerie; centre Francois Baclesse, registre general des Tumeurs du calvados et Enquete parmantier Cancers; Inserm U1086, Cancers and prevention Mulot, Claire; Université Paris Descartes, INSERM UMR 5775 EPIGENETEC Laurent-Puig, Pierre; Université Paris Descartes, INSERM UMR 5775 EPIGENETEC Schartz, Claire; Institut de Cancerologie Jean Godinot, Centre de lutte contre le Cancer Truong, Therese; Université Paris-Saclay, Université Paris-sud, UVSQ, Inserm, CESP (center for research in epidemiology and population health) Guenel, Pascal; Université Paris-Saclay, Université Paris-sud, UVSQ, Inserm, CESP (center for research in epidemiology and population health)
Key Words:	thyroid cancer, case-control study, fine-mapping study, cancer genetics

SCHOLARONE™
Manuscripts

Fine-Mapping of two differentiated thyroid carcinoma susceptibility loci at 9q22.33 and 14q13.3 detects novel candidate functional SNPs in Europeans from metropolitan France and Melanesians from New Caledonia

Tcheandjieu C[1], Lesueur F[2], Sanchez M[1], Baron-Dubourdieu D[3], Guizard AV[4,5], Mulot C[6], Puig PL[6], Schvartz C[7], Truong T[1]* and Guenel P[1]*

***as co-last authors**

Corresponding author:
Therese Truong, PhD
Epidemiology of Cancers: Genes and Environment
16 avenue Paul Vaillant Couturier - Villejuif, France
Tel +33 1 45 59 50 31
Email: therese.truong@inserm.fr

Key Word: Thyroid cancer, Cancer genetics, case-control study, Fine-mapping study

Abbreviations
TC: Thyroid Cancer
DTC: Differentiated Thyroid Cancer
PTC: Papillary Thyroid Cancer
FTC: Follicular Thyroid Cancer
PWM: Position Weight Matrix
eQTL: Expression Quantitative Traits
GWAS: Genome Wise Association Study
BMI: Body Mass Index
SNP: Single Nucleotide Polymorphism

Article category: Cancer genetics and epigenetics

Novelty and Impact: We performed a multiethnic fine-mapping study at two differentiated thyroid cancer (DTC) susceptibility regions, using data from two population-based case-control studies conducted in Metropolitan France and New Caledonia. This is the first time that genetics risk factors of thyroid cancer were investigated in Melanesians, a population with particularly high incidence rates of DTC. The results suggested novel variants to be prioritized for further functional studies.

For Peer Review

Author affiliations

1: Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, 94807, France

2: Inserm U900, Curie Institute, Mines ParisTech, Paris, 75005, France

3: Laboratory of Anatomy and Cytopathology, Noumea, New Caledonia, France

4: Registre Général des tumeurs du Calvados & Enquête Permanente Cancer, Centre François Baclesse, Caen, France

5: Inserm U1086, Cancers and Prevention, Caen, France

6: Université Paris Descartes, Inserm UMR 5775 EPIGENETEC, Paris, France

7: Centre de Lutte Contre le Cancer Jean GODINOT, Reims, France

For Peer Review

Abstract

Incidence of differentiated thyroid carcinoma varies considerably between countries and ethnic groups, with particularly high incidence rates in Melanesians of New Caledonia. DTC has a familial relative risk higher than other cancers, highlighting the contribution of inherited factors to the disease. Recently, GWAS identified several DTC susceptibility loci. The most robust associations were reported at loci 9q22 (rs965513 and rs1867277) and 14q13 (rs944289 and rs116909734). In this study, we performed a fine-mapping study of the two gene regions among Europeans and Melanesians from Metropolitan France and New Caledonia.

We examined 81 SNPs at 9q22 and 561 SNPs at 14q13 in Europeans (625 cases/776 controls) and in Melanesians (244 cases/189 controls). The association with the 4 SNPs previously identified in GWAS was replicated in Europeans while only rs944289 was replicated in Melanesians. Among Europeans, we found that the 2 SNPs previously reported at 9q22 were not independently associated to DTC and that rs965513 was the predominant signal; at 14q13, we showed that the haplotype rs944289[C]-rs116909374[C]-rs999460[T] was significantly associated with DTC risk and that the association with rs116909374 differed by smoking status (p-interaction=0.03). Among Melanesians, a new independent signal was observed at 14q13 for rs1755774 which is strongly correlated to rs2787423; this latter is potentially a functional variant. Significant interactions with parity ($p<0.05$) and BMI were observed for rs1755774 and rs2787423.

This study contributed to a better characterization of the DTC loci 9q22 and 14q13 in Europeans and in Melanesians and has identified novel variants to be prioritized for further functional studies.

For Peer Review

Introduction

Thyroid cancer (TC) is the most common malignancy of the human endocrine system; its incidence is characterized by considerable ethnic and geographic variation[1,2]. One of the highest incidence rates was observed in Melanesian women of New Caledonia (71.4/100,000 person-years)[1–3], a French territory in South Pacific. In metropolitan France, the incidence rate was 12.6/100,000 in women and 4.1/100,000 in men[2] in 2012.

Papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC) are the most frequent subtypes of differentiated thyroid carcinoma (DTC), representing about 90% of all thyroid cancers. Exposure to ionizing radiations during childhood is a well-established risk factor for DTC. Overweight, deficiencies in iodine intake, parity, and late age at menarche are suspected DTC risk factors[4–6]. In addition excess risk for DTC has consistently been observed among first-degree relatives of DTC patients[7] and it has been shown that TC is one of the cancers with the highest familial risk[7–9]. Linkage studies in multiple-case DTC families have reported several susceptibility loci but no high-penetrance gene could be identified[10–12]. Studies on candidate gene or pathways have reported associations of DTC with common susceptibility variants but only a few have been replicated in independent investigations[13–15].

Recently, genome-wide association studies (GWAS) on DTC have identified several other susceptibility loci. Rs965513 at locus 9q22 and rs944289 at locus 14q13 were first identified in a GWAS on DTC conducted in Iceland in 192 DTC cases and 37,196 controls[16]. Rs965513 was also associated with PTC in a GWAS conducted in a population exposed to radiation in the Chernobyl area[17] and this association was independently confirmed in a population of Belarusian children exposed to fallout from the power plant accident[18]. Another GWAS on TSH levels highlighted 22 single nucleotide polymorphisms (SNPs), among which rs116909374 on 14q13 was also associated with DTC risk in addition to rs965513 and rs944289[19]. These findings were replicated in a GWAS conducted in 690 DTC cases and 479 controls from an Italian population with high DTC incidence and in another GWAS conducted in 398 cases and 502 controls from Spain[20,21]. These two GWAS identified additional susceptibility SNPs at other loci[20,21].

For Peer Review

So far, the most robust associations from GWAS were reported for the SNPs rs965513 (9q22, near *FOXE1*) and rs944289 (14q13, between *PTCSC3* and *NKX2-1*). Both *FOXE1* and *NKX2-1* encode thyroid specific transcription factors (TTFs) involved in thyroid development, differentiation and regulation of thyroid function making them strong candidate susceptibility genes that may harbor likely pathogenic sequence variants.

Follow-up studies of the 9q22 locus identified a functional SNP rs1867277 located in the 5'UTR of *FOXE1*, that affects the transcriptional regulation of *FOXE1* through the recruitment of USF1/USF2 (upstream stimulating factors)[22–24]. Rs1867277 was suggested as the *FOXE1* causal variant at 9q22 for TC risk by functional studies[22]. Further studies also consistently associated DTC risk with the expansion of a trinucleotide repeat in the unique exon of *FOXE1* encoding poly-alanine tract which could result in a gain of function mutation[18,23,25]. Recently, He and al. showed that rs965513 may also have a functional role in DTC as this SNP is associated with decreased expression of *PTCSC2* [a new long intergenic non-coding RNA (lincRNA) gene] expressed in PTC tissues[26]. Rs965513 is located in the unspliced transcript of *PTCSC2* and in an intron of the spliced *PTCSC2,* and is 60 kb upstream of rs1867277; these two variants are not correlated in Europeans (r^2=0.39 and D'=0.73 in the European sample of the 1000 genomes (EUR)).

Follow-up study of the 14q13 locus described a lincRNA named *PTCSC3* located 64 kb upstream of the GWAS SNP rs944289[27]. The authors demonstrated that the expression of *PTCSC3* was strongly down regulated in PTC and that, the risk allele T of rs944289 was associated with low expression of *PTCSC3* in PTC. The second GWAS SNP at 14q13, rs116909374, is located in a gene desert region, 89 kb downstream of rs944289; these 2 SNPs are not correlated (r^2=0.006, D'=0.4 in the EUR population) suggesting the presence of two distinct signals at this locus.

Following the mapping of the two DTC loci, association studies on familial DTC, sporadic DTC and radio-induced DTC have consistently replicated the association between DTC and 9q22 polymorphisms but were less consistent for 14q13 polymorphisms[18,25,28,29]. However, most of

For Peer Review

the variants identified in these loci are located outside of gene regions. Also it remains unclear whether the polymorphisms identified within each of these two loci are independent and whether they are causally related to DTC.

In this study, using data from studies conducted in Europeans and Melanesians, we aimed to conduct a fine-mapping analysis at loci 9q22 and 14q13 in these two populations and to functionally annotate the most associated variants in order to propose novel candidate risk variants of DTC. Another objective was to examine whether differences in frequency of risk alleles between Europeans and Melanesians might account for the difference in DTC incidence rates between the two groups. We also explored the interaction between susceptibility alleles and non-genetic risk factors for DTC such as body mass index (BMI), parity, age at menarche, smoking, and alcohol drinking.

Material and Methods

Study Populations

We used data from two case-control studies on thyroid cancer conducted in metropolitan France (CATHY study) and in New Caledonia, a French territory in South Pacific. All participants signed written consent.

The CATHY study is a population-based case-control study conducted in Marne, Ardennes and Calvados, three French administrative areas ("départements") covered by a cancer registry. Eligible cases were all patients diagnosed with DTC between 2002 and 2007 residing in these areas. Controls were frequency-matched to cases by sex and 5-year age groups and were randomly selected among individuals aged 25 years and over, residing in the same areas at the time of cases' diagnosis. From the 621 cases and 706 controls initially included in the study, saliva (Oragene DNA®) or cells from buccal swabs were available for 583 cases and 643 controls; among those, 556 cases and 630 controls declared to be of European ancestry.

The New Caledonia (NC) study is a country-wide population-based case-control study[6]. The case group included patients with DTC diagnosed between 1993 and 1999, who had been

For Peer Review

living for at least 5 years in NC at the time of diagnosis. The cases were identified from the two pathology laboratories of NC and active searches in medical records of the main hospitals. Age and sex frequency-matched controls were randomly selected from recently updated electoral rolls. A total of 332 cases and 412 controls were included. Of these, 244 cases and 189 controls declared themselves as Melanesians and 42 cases and 133 controls declared themselves as Europeans. DNA was obtained from saliva samples available for 284 Melanesians (164 cases and 120 controls) and 110 Europeans (28 cases and 82 controls).

In both studies, information on ethnicity, personal and familial history of thyroid disease, reproductive factors, exogenous hormone use, weight, height, dietary habits, alcohol intake, tobacco smoking, residential and occupational histories was collected during in-person interviews.

Markers selection and genotyping

For the fine-mapping analysis, we defined at locus 9q22, an interval that encompassed rs965513 and the *FOXE1* gene (62 kb, positions 100,556,108-100,671,478 [NCBI build 37/hg19 assembly]) and at locus 14q13, an interval that encompassed *PTCSC3* and *NKX2-1* genes (375 kb; positions 36,601,061-36,999,637 [NCIB build 37/hg19 assembly]). We used tagger software[30] (pairwise approach with $r^2 \geq 0.8$) to capture the genetic variation within the selected regions. We selected tagging SNPs with a minimum minor allele frequency (MAF) of 0.05 in the Caucasian population (CEU) as reported in the HapMap Project (Data Release 21/Phase II).

At locus 9q22, 11 tagging SNPs (including the GWAS SNP rs965513 and the functional SNP rs1867277) were selected. At locus 14q13, 47 tagging SNPs (including the GWAS SNPS rs944589 and rs116909374) were selected. A total of 58 SNPs were genotyped using the Fluidigm Dynamic array chip.

Quality control

Oligonucleotide primers and probes were validated on 90 CEPH DNA samples. All genotyped SNPs had a call rate>95%. From The CATHY study, 64 cases/65 controls with DNA prepared from oral brush sample were not genotyped due to their lower DNA concentration

For Peer Review

(<5µl). We excluded 19 cases/11 controls because of genotyping completion rate less than 90%. Among those, 9 cases/9 controls were from NC study. The deviation of the genotype proportions from Hardy-Weinberg equilibrium (HWE) was assessed in both European controls and Melanesian controls using Chi-squared test with one degree of freedom. We applied a Bonferroni correction for multiple testing using a p-value threshold of $8x10^{-4}$; no SNP deviated from HWE. Because the genotype frequencies of controls in Europeans from the New-Caledonia study and in Europeans from the CATHY study were similar, we combined these two populations in the analyses. In total, 1134 Europeans (508 cases/626 controls) and 268 Melanesians (156 cases/112 controls) were considered for analysis (table 1).

Regional imputation

Ungenotyped SNPs with MAF>0.02 in the selected regions were imputed using IMPUTE2.2 software[31] and the cosmopolitan reference panel from the 1000 genomes dataset (September 2013 release). The use of this panel as reference haplotype is particularly recommended for admixed population such as the Oceanian population[32]. We imputed separately genotypes of Europeans and genotypes of Melanesians for each locus. We evaluated the accuracy of the imputation by masking a random subset of the available genotypes and checking whether these genotypes could be recovered by imputation. The concordance between imputed and true genotypes was 95% in the two populations. For the analyses, we excluded the imputed SNPs that deviated from HWE in controls and the imputed SNPs with imputation performance (INFO-score<0.70).

Among Europeans, respectively 56 imputed SNPs at loci 9q22 and 478 imputed SNPs 14q13 were considered for analysis; in Melanesians 69 imputed SNPs at locus 9q22 and 464 imputed SNPs at locus 14q13 were considered for analysis.

Statistical analyses

Single marker association tests were performed using unconditional logistic regression assuming a log-additive model. Imputed SNPs were analyzed as expected genotype count (gene dosage) with SNPtest V2.4.1 software. We conducted conditional analyses to test for residual

For Peer Review

association after accounting for a key SNP. We applied Bonferroni correction based on the effective number of independent tests estimated by the method of Li and al.[33]. We considered that SNPs with p-value<$4.4x10^{-4}$ (based on our estimation of 115 independent tests) were significantly associated with DTC. Haplotype analysis was carried out using the R package Haplot.Stat. All the analyses were conducted separately in Europeans and Melanesians and were adjusted on age, sex and area of residence. Europeans of NC were grouped into one single area of residence (table 1).

We investigated the interaction between SNPs showing the strongest association with DTC risk (including the GWAS SNPs) at each locus and suspected lifestyle risk factors of DTC such as BMI, alcohol intake and cigarette smoking; we also investigated interaction with parity, age at menarche and irregular periods in women. We used the likelihood ratio test comparing models with and without the interaction term. These models were adjusted for age, sex, area of residence and all the suspected lifestyle risk factors.

We used the R package ggplot2 and locuszoom1.3 software[34] for plotting our result of single markers association and conditional regression at each locus.

Functional annotation

To assess the possible functional role of the most associated SNPs we used the following freely available tools that investigate noncoding genome variants: RegulomeDB (http://regulome.stanford.edu) and Haploreg V2 (www.broadinstitute.org/mammals/haploreg). These tools predict whether a given SNP may alter the regulatory protein binding sites, chromatin structure, histone modifications and putative transcription factors (TFs) binding sites. The prediction is based on information from Chip-Seq data or position weight matrix analysis (PWMs). In addition, we used is-rSNP (http://bioinformatics.research.nicta.com.au/software/is-rsnp) to determine the statistical significance of the nucleotide variation for altering binding of TFs using a PWM score. To examine the association between associated SNPs and expression level of a nearby gene, we used the freely available expression quantitative trait loci (eQTL) data of SNPexp tools (http://app3.titan.uio.no/biotools). With this tool, we calculated an adjusted p-value (with

For Peer Review

Bonferroni correction) applying additive model, the correlation between genotype of each SNP and expression levels for probes representing the gene of interested using the 210 HapMap phase II dataset. We also used UCSC (http://genome-euro.ucsc.edu) and Ensembl (http://www.ensembl.org) genomes databases to identify possible uncharacterized lincRNA, predicted transcripts and ESTs (Express Sequence Tags) at each locus.

Results

The characteristics of the study sample are presented in table 1. Men represented 25% of the cases in Europeans and 5% in Melanesians.

Locus 9q22

In Europeans and in Melanesians, rs965513 and rs1867277, previously associated with DTC risk in GWAS or functional studies, were in separate LD blocks. The LD block containing rs1867277 encompasses *FOXE1* and part of a gene desert region; this LD block was slightly larger in Melanesians than in Europeans (figure S1).

In Europeans, we analyzed 69 genotyped or imputed SNPs. The associations with alleles rs965513[A] (OR=1.52, $p=6x10^{-6}$) and rs1867277[A] (OR=1.38, $p=3x10^{-4}$) were replicated (table 2; figure 1A). The most significant association was observed for the imputed SNP rs13295081[T] (OR=1.44, $p=5x10^{-6}$). This marker was strongly correlated with rs965513 ($r^2=1$) that displayed the second most significant association (figure 2A). From the 50 other SNPs that reached statistical significance ($p<4.4x10^{-4}$) (figure 1A and table S1), 12 SNPs were strongly to moderately correlated with rs13295081 ($0.6<r^2\leq1.0$) (figure 2A) and 34 SNPs were strongly correlated with rs1867277 ($r^2>0.8$).

When conditioning the regression test on rs13295081 or rs965513 (figure 2B), no SNP (including rs1867277) remained significantly associated with DTC. When conditioning on rs1867277, the p-values for rs13295081 and rs965513 remained suggestive of an association ($p=6x10^{-3}$) (figure 2C); indicating predominant signal from the LD block containing rs965513 and a non-independence of the signals from the LD blocks containing rs965513 and rs1867277.

For Peer Review

Haplotype analysis did not reveal any association beyond the single SNP analysis (data not shown).

In Melanesians, we analyzed 75 genotyped or imputed SNPs. No SNP was statistically significantly associated with DTC (figure 1B), but suggestive associations were observed for rs965513[A] (OR=1.43 ; p=0.03) and rs1867277[A] (OR=1.34 ; p=0.06), with effects of the same order of magnitude as in Europeans (table 2).

Locus 14q13

The size of the LD blocks at 14q13 was slightly smaller in Melanesians than in Europeans. Moreover, the allele frequencies of several SNPs differed considerably between the two ethnic groups (figure S2; table S3).

In Europeans, 522 genotyped or imputed SNPs were considered. Association between DTC and GWAS SNPs rs944289[C] (OR=0.74, $p=4.7.10^{-4}$) and rs116909374[T] (OR=1.97; $p=3.3x10^{-3}$) were at the limit of statistical significance after correction for multiple testing (table 2), whereas the association with the imputed SNP rs7494749[A] (OR=0.73, $p=4.3x10^{-4}$) was statistically significant (table 2, figure 1C). This marker is located in the *PTCSC3* third intron and is strongly correlated with rs944289 ($r^2=0.98$; figure 2D), which displayed the second strongest association with DTC in Europeans.

Conditional analysis on rs7494749 or rs944289 attenuated the association p-values of DTC with other SNPs, except for the GWAS SNP rs116909374 and SNPs near *NKX2-1* for which, the p-values remained suggestive of an association ($p<4x10^{-3}$; figure 2E). When conditioning the analysis on the GWAS SNPs rs944289 and rs116909374, rs999460 remained the most associated SNP with a p-value at the limit of significance (figure 2F). These results suggest that several SNPs in the locus are independently associated with DTC risk. We performed an haplotype analysis based on rs944289, rs116909374 and rs999460 and we observed a significantly decreased risk of DTC in individuals carrying the haplotype rs944289[C]-rs116909374[C]-rs999460[T] (OR=0.51, $p=3.11x10^{-6}$) when compared to individuals with the most frequent haplotype rs944289[T]-rs116909374[C]-rs999460[C] (table 3).

For Peer Review

In Melanesians, 511 imputed and genotyped SNPs were analyzed. The per-allele OR reported for the GWAS SNP rs944289[C] (OR=0.57, $p=1.1x10^{-3}$) was at the limit of statistical significance whereas the other GWAS SNP rs116909374 was not associated with DTC (OR=1.44, p=0.76; table 2). The OR for rs944289 in Melanesians was close to that reported in Europeans, but a much higher frequency of the rs944289[T] allele was observed. In this population, the imputed SNP rs1755774[G] was the most strongly associated SNP with DTC (OR=0.50, $p=1.6x10^{-5}$; table 2, figure 1D). Rs1755774 was weakly correlated with rs944289 ($r^2=0.27$) and rs116909374 ($r^2=0.14$) (figure 3A) in Melanesians. Of note, rs1755774 was also associated with DTC in Europeans but the association was at the limit of significance (OR=0.76, $p=7.3x10^{-3}$; figure 1C). The additional 37 SNPs that reached statistical significance in Melanesians ($p<4.4x10^{-4}$; table s3) were all located in the intergenic region between *PTCSC3* and *MBIP* and most of them were strongly correlated with rs1755774 ($r^2>0.8$; figure 3A). No SNP remained statistically significant after conditioning the association tests on rs1755774 (figure 3B) suggesting one independent signal only at 14q13 in Melanesians.

In Europeans, 88% of cases were PTC while 12% were FTC; in Melanesians, the numbers were 84% and 16%, respectively (table 1). When restricting the analyses to PTC, the ORs in both ethnic groups were similar to those observed for PTC and FTC combined (tables S1 and S2).

Difference in allelic frequency between Europeans and Melanesians

We compared frequencies of risk alleles of the most significant SNPs at 9q22 and 14q13 between Melanesians and Europeans to see whether they may account for the differences in DTC incidence rates between ethnic groups. We found that the risk alleles were more frequent in Europeans than in Melanesians, and hence could not explain the high DTC incidence rates among Melanesians relative to Europeans.

Effect modification by non-genetic risk factors of DTC

Supplementary figures 4 and 5 show forest plots of the associations between DTC risk and candidate SNPs at loci 9q22 and 14q13, stratified on categories of BMI, alcohol consumption, smoking status, age at first menarche, irregular menstrual cycles and number of

For Peer Review

pregnancies among European and Melanesian women separately. Analysis of interaction in European men was not possible due to limited sample size in each strata.

At locus 9q22, there was no evidence of interaction between the previously highlighted SNPs and the non-genetic risk factors (figure S4).

At locus 14q13, we observed a statistically significant interaction between the GWAS SNP rs116909374 and smoking status among European women (p-interaction=0.03). The association between rs116909374[T] allele and DTC risk was stronger in current smoker [OR=11.06 (95%CI: 1.26–96.82)] than in former and never smokers [OR=1.63 (95% CI: 0.50–5.38); OR=0.77 (95%CI: 0.36–1.64), respectively] (figure S5). Interaction between smoking status and rs116909374 could not be examined in Melanesian women due to a very low risk allele frequency (MAF≤0.01).

In Melanesians, we found that the association between DTC risk and rs1755774 differed by parity (p-interaction=0.04) and BMI (p-interaction=0.05). Indeed, the association between DTC and rs1755774 was weaker in women who had 2 or more pregnancies (OR=0.26 (95% CI: 0.13–0.49)) than and in women who had less than 2 pregnancies (OR=0.66 (95% CI: 0.22–1.97)) (figure S5). For the same SNP, women with BMI>25 kg/m² had an OR of 0.26 (95% CI: 0.14–0.49) while women with BMI≤25 kg/m² had an OR of 0.56 (95% CI: 0.18–1.70). No such effect was apparent in European women.

Discussion

At 9q22, the previously described associations between DTC risk and SNPs rs965513 and rs1867277[16,19,20,22,29], were replicated in our European population (OR=1.52 for rs965513 ($p=6x10^{-6}$) and OR=1.38 for of rs1867277 ($p=3x10^{-4}$)). In Melanesians, effects of same magnitude were observed for rs965513 and rs1867277 although the p-values did not reach the level of statistical significance, possibly due to small sample size (OR=1.43, p=0.03 and OR=1.34, p=0.06 respectively for rs965513 and rs1867277).

For Peer Review

Using conditional analyses we showed that the associations of DTC with rs965513 and rs1867277 were not independent. This finding is consistent with the candidate gene study by Jones et *al.*[35] that reported reduced signals of the two SNPs after reciprocal adjustment. Because the predominant association signal at locus 9q22 lied in the LD block containing rs965513, we annotated this block in silico looking for the presence of a potential functional SNP.

Among SNPs predicted to alter transcription factors (TFs) or regulatory protein binding sites, we identified rs10759944 (r^2=0.99 with rs965513) as a possible functional SNP as it is located in the binding site of the regulatory proteins ERALPHA (ESR1) according to chip-Seq data and its risk allele A was suggested by PWM (position weight matrix) analysis to disrupt the binding site of some TFs such as EWSR1-FLI1, E2F1 and AP-2α (TFAP2A) (data not shown). Indeed, ESR1 is a transcription factor activated by estrogen binding. Steroid hormones and their receptors are involved in the regulation of gene expression and affect cellular proliferation and differentiation in target tissues. Overexpression of ESR1 receptors (ERα) was reported in PTC[36] or thyroid cells[37]. E2F1 is a TF controlling cell cycle and acting as tumor suppressor protein[38]. AP-2α (TFAP2A) is a protein that interacts with cellular enhancer elements to regulate transcription of selected genes. These proteins are overexpressed in PTC[36,38,39] and seem to play an important role in PTC pathogenesis.

The biological role of rs10759944 in thyroid tissue was investigated in a recent study that performed in vivo experiments on variants prioritized from a larger region of locus 9q22 than in our study[40]. This study reported that at least three regulatory elements functioning as an enhancers lies in a LD block containing rs965513, and that these enhancers harbor at least four functional SNPs: rs7864322, rs12352658, rs7847449, and rs10759944[40]. They also showed that the three enhancer elements interact with the shared promoter region of *FOXE1* and *PTCSC2* in a human papillary thyroid carcinoma cell line and this interaction was also retrieved in unaffected thyroid tissue for the enhancer containing rs1079944. Among these 4 SNPs, we analyzed only rs10759944; rs7864322 was not imputed due to its location over 1000 kb upstream of our fine-mapping boundary; imputation of rs12352658 and rs7847449 did not pass the quality controls and were excluded from the analysis (info-score<0.7).

For Peer Review

We reported that rs10759944 was associated with DTC risk in Europeans (OR=1.48, $p=2x10^{-5}$) and in Melanesians (OR=1.42, p=0.03). This association was also reported by the GWAS conducted in Iceland, the GWAS on radiation-related PTC in Belarussian patients, and a study on familial DTC cases[16,17,25].

At locus 14q13, the associations previously reported for the GWAS SNPs rs944289 (OR=0.74, $p=5x10^{-4}$) and rs116909374 (OR=1.97, $p=3x10^{-3}$) were replicated in Europeans. In Melanesians, rs944289 was replicated (OR=0.57, $p=1x10^{-3}$), but not rs116909374 (OR=1.44, p=0.76), which is rare in this population (MAF=0.004). Moreover, in Europeans we found one haplotype rs944289[C]-rs116909374[C]-rs999460[T], that was significantly associated with DTC risk (OR=0.51, $p=3x10^{-6}$). This result suggests a real combined effect of rs944289, rs116909374 and rs999460 on DTC risk or a yet to be characterized causal SNP tagged by this haplotype.

Rs944289[T] is positioned in the binding site of the TFs CEBP/α and CEBP/β and had been associated with variation in expression levels of *PTCSC3* in PTC tissues[27]. To further investigate whether the risk alleles rs116909374 and rs999460 could also be biologically relevant to DTC pathogenesis, we used range of annotation tools.

Interestingly, chip-Seq data revealed that rs116909374 was located in the binding site of the regulatory protein ESR1. Rs116909374 is also located in an enhancer region in H1 cell line and in a promoter of an uncharacterized lincRNA named TCONS_0022711, which is expressed in thyroid tissues.

Rs999460 is positioned in the poised promoter in H1 cell line and in a repressed polycomb of numerous cell lines. Chip-Seq data revealed that rs999460 falls also in the binding site of EZH2, and analysis using Position Weight Matrix (PWMs) indicates that its risk allele may alter the binding of the TFs ZNF143 and NKX2-1 (data not shown). EZH2 is the functional enzymatic component of the polycomb repressive complex 2 (PRC2) responsible for healthy embryonic development through the epigenetic maintenance of genes regulating development and differentiation[41]. EZH2 is also involved in up regulation of NKX2-1[41]; however, using SNPexp tools, we did not find significant association between the expression level of *NXK2-1* gene and

For Peer Review

rs999460. Therefore, further *in vivo* investigation regarding the potential role of rs116909374 and rs999460 in DTC pathogenesis is warranted.

Following the identification of the 14q13 locus in the first GWAS on DTC[16], few studies replicated the association with rs944289 in other populations. Bonora et *al*[25] investigated the candidate SNP and four other SNPs tagging the 14q33 region in families highly predisposed to DTC and showed that in addition to rs944289, rs999460 was significantly associated with the disease[25]. Also, consistently with the study of Gudmundsson and *al.*[19], we showed that the GWAS SNPs rs944289 and rs116909374 are independently associated with DTC in Europeans.

In Melanesians, at 14q13, the pattern is different. The strongest association was reported for rs1755774 (OR=0.50, $p=1.6x10^{-5}$), which is not correlated with rs944289 or rs116909374. No haplotype outperforming the single-SNP analysis was shown to be associated with DTC, and conditional analysis suggested that rs1755774 is the only signal of association at this locus in Melanesians.

14 SNPs located in the LD block containing rs1755774 and significantly associated to DTC in Melanesians showed putative functional prediction. Among those, rs2787423 and rs67263215 had predictive function that can be related to thyroid cancer. These SNPs were located in the binding site of the regulatory proteins STAT3 and FOS (figure S3). Rs2787423[G] also seems to alter the binding site of p300, CEBP/α and CEBP/β transcriptions factors (TFs), which are all expressed in thyroid and suppressed in tumor tissue[27] while rs67263215[T] alter the binding site of BDP1, TATA box and YY1 which form together, a multiprotein complexes involved in the initiation of the transcription of some gene[42,43]. Moreover rs2787423 and rs67263215 also falls in the enhancer region in human mammary epithelial cells (HMEC) upstream of an uncharacterized lincRNA or uncertain coding transcript named TCONS_0002269, which is abundantly expressed in thyroid tissues (figure S3). However, functional studies are necessary to confirm the importance of this variant and to better characterize the region.

So far, genetics studies on DTC have been conducted in different populations such as Chinese, Japanese, Cuban or Polynesians[28,29,44,45] but to our knowledge, this is the first study

For Peer Review

that investigated the role of common SNPs among Melanesians of New-Caledonia, a population with one of the highest incidence rates of DTC in the world[3].The incidence of DTC in Melanesian women was 10-fold higher than in European women[3]. No clear explanation is currently available for this observed high incidence rates. Over diagnosis of occult thyroid micro-carcinomas due to intensive medical screening practices of thyroid nodules cannot be a unique explanation, since the incidence of large size carcinomas >10 mm remains elevated in Melanesians. In contrast to French Polynesia, no nuclear experiment was conducted in New Caledonia, making radiation exposure an unlikely explanation for the elevated incidence. In addition, striking difference in incidence rates between the ethnic groups in this territory does not support this hypothesis. We have shown previously that high parity and obesity may partly account for the elevated incidence of DTC in Melanesian women of New Caledonia[46,47]. In the present study, we showed that the difference in allelic distribution of susceptibility SNPs at loci 9q22 and 14q13 may not contribute to the difference in DTC incidence between Europeans and Melanesians.

In addition to be the first study on the GWAS loci 9q22 and 14q13 in Melanesians, this is also to our knowledge, the first fine mapping study of 14q13 locus in Europeans. Other strength of our study is the availability of data on lifestyle-related risk factors, which allowed studying gene-environment interactions. Therefore we examined whether risk factors for DTC, including BMI, parity, tobacco smoking and alcohol consumption, modified the associations between SNPs and DTC. Gene-environment interactions in DTC have been investigated in only two previous studies where the authors investigated interaction between parity, BSA (body surface area), BMI, iodine intake or thyroid radiation dose and the GWAS SNPs at loci 9q22 and 14q13[47,48]. Consistently with these studies, we did not observe interactions between the SNPs and parity or BMI in Europeans women. However, we reported that the association of rs116909374 with DTC in European women was stronger in current smokers (OR=11.06, $p=2.9x10^{-2}$) than in never or former smokers (p-interaction=0.03) (figure S5-E). This interaction may be explained by the independent effects of cigarette smoking and rs116909374[T] on decreased TSH levels[19,48], which in turn are associated with increased risk of DTC[49]. Gudmunsson et al[19] hypothesized that low TSH levels could lead to less differentiation of the

thyroid epithelium, causing a higher predisposition to malignant cell transformation. In Melanesians, we found that parity modified the association of DTC with rs2787423 or rs1755774 at 14q13. However, these results should be interpreted with caution given the small number of subjects in each stratum and need to be confirmed in other studies.

Our study has some limitation: first, the lack of power to detect associations in Melanesians; second, the fine-mapping study was based on imputed SNPs and we used *in silico* analysis to predict a potential biological function. This implied the need for replication of our results in other studies and the need for functional studies to confirm the biological prediction. The ethnicity was self-declared by study subjects possibly leading to misclassification particularly among Europeans. Also, DTC constitutes a heterogeneous set of diseases with various molecular signatures in the tumor genotype that have been associated with tumor aggressiveness[50]. It is possible that molecular signature of the tumor also differs according to the etiology of DTC, but this could not be examined in our study as no information on tumor molecular profile was available.

In summary, the association of DTC with SNPs at loci 9q22 and 14q13 was confirmed in Europeans. We showed that the 2 GWAS SNPs at locus 9q22 were not independently associated to DTC and that rs965513 is the predominant signal in this region. We also showed that rs10759944, strongly correlated to rs965513 and recently reported as a functional SNP, is associated with PTC in Europeans and Melanesians. At locus 14q13, we found a haplotype associated with DTC risk in Europeans and identified among Melanesians of New Caledonia, new susceptibility SNPs that are also associated to DTC risk in Europeans. We also performed a comprehensive bio-informatics assessment of the associated SNPs at the two loci and proposed some novel candidate SNPs that may be related to DTC pathogenesis although further functional studies are needed to confirm their role in DTC. Further studies on other suspected DTC susceptibility loci should also be conducted to explain the elevated incidence rates in Melanesians.

Funding

The NC study was supported by grants from the "Fondation de France" (FDF), the "Association pour la Recherche contre le Cancer" (ARC) and the "Agence Française de Sécurité Sanitaire de l'Environnement et du Travail" (AFSSET). The CATHY study was funded by AFSSET, Ligue Nationale Contre le Cancer, Institut National du Cancer, EDF and FDF.

Conflict of interested statement: None declared

References

1. Moore M a, Baumann F, Foliaki S, Goodman MT, Haddock R, Maraka R, Koroivueta J, Roder D, Vinit T, Whippy HJD, Sobue T. Cancer epidemiology in the pacific islands - past, present and future. *Asian Pac J Cancer Prev [Internet]* 2010;11 Suppl 2:99–106. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20553071

2. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on C [Internet]. [cited 2014 Oct 30];Available from: http://globocan.iarc.fr/Pages/references.aspx

3. Truong T, Rougier Y, Dubourdieu D, Guihenneuc-Jouyaux C, Orsi L, Hémon D, Guénel P. Time trends and geographic variations for thyroid cancer in New Caledonia, a very high incidence area (1985-1999). *Eur J Cancer Prev [Internet]* 2007 [cited 2014 Jul 10];16:62–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17220706

4. Truong T, Orsi L, Dubourdieu D, Rougier Y, Hémon D, Guénel P. Role of goiter and of menstrual and reproductive factors in thyroid cancer: a population-based case-control study in New Caledonia (South Pacific), a very high incidence area. *Am J Epidemiol [Internet]* 2005 [cited 2014 Sep 9];161:1056–65. Available from: http://aje.oxfordjournals.org.gate2.inist.fr/content/161/11/1056.long

5. Leux C, Guénel P. Risk factors of thyroid tumors: role of environmental and occupational exposures to chemical pollutants. *Rev Epidemiol Sante Publique [Internet]* 2010 [cited 2014 Sep 9];58:359–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20980113

6. Cléro E, Leux C, Brindel P, Truong T, Anger A, Teinturier C, Diallo I, Doyon F, Guénel P, de

For Peer Review

Vathaire F. Pooled analysis of two case-control studies in New Caledonia and French Polynesia of body mass index and differentiated thyroid cancer: the importance of body surface area. *Thyroid [Internet]* 2010 [cited 2014 Sep 9];20:1285–93. Available from: http://online.liebertpub.com.gate2.inist.fr/doi/abs/10.1089/thy.2009.0456

7. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst [Internet]* 1994 [cited 2014 Jul 1];86:1600–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7932824

8. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer [Internet]* 2002 [cited 2014 Oct 30];99:260–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11979442

9. Hemminki K, Eng C, Chen B. Familial risks for nonmedullary thyroid cancer. *J Clin Endocrinol Metab [Internet]* 2005 [cited 2014 Nov 14];90:5747–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16030170

10. Robledo M. Association studies in thyroid cancer susceptibility : are we on the right track ? *J Mol Endocrinol* 2011;2.

11. Lesueur F, Stark M, Tocco T, Ayadi H, Delisle MJ, Goldgar DE, Schlumberger M, Romeo G, Canzian F. Genetic heterogeneity in familial nonmedullary thyroid carcinoma: exclusion of linkage to RET, MNG1, and TCO in 56 families. NMTC Consortium. *J Clin Endocrinol Metab [Internet]* 1999 [cited 2015 Mar 13];84:2157–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10372725

12. Malchoff CD, Sarfarazi M, Tendler B, Forouhar F, Whalen G, Joshi V, Arnold A, Malchoff DM. Papillary thyroid carcinoma associated with papillary renal neoplasia: genetic linkage analysis of a distinct heritable tumor syndrome. *J Clin Endocrinol Metab [Internet]* 2000 [cited 2015 Aug 5];85:1758–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10843148

13. Granja F, Morari J, Morari EC, Correa LAC, Assumpção LVM, Ward LS. Proline homozygosity in codon 72 of p53 is a factor of susceptibility for thyroid cancer. *Cancer Lett [Internet]* 2004 [cited 2014 Jul 1];210:151–7. Available from: http://www.sciencedirect.com/science/article/pii/S030438350400076X

14. Akulevich NM, Saenko VA, Rogounovitch TI, Drozd VM, Lushnikov EF, Ivanov VK, Mitsutake N, Kominami R, Yamashita S. Polymorphisms of DNA damage response genes

For Peer Review

in radiation-related and sporadic papillary thyroid carcinoma. *Endocr Relat Cancer [Internet]* 2009 [cited 2014 Jul 1];16:491–503. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19286843

15. Sigurdson AJ, Land CE, Bhatti P, Pineda M, Brenner A, Carr Z, Gusev BI, Zhumadilov Z, Simon SL, Bouville A, Rutter JL, Ron E, et al. Thyroid nodules, polymorphic variants in DNA repair and RET-related genes, and interaction with ionizing radiation exposure from nuclear tests in Kazakhstan. *Radiat Res [Internet]* 2009 [cited 2014 Jul 1];171:77–88. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2875679&tool=pmcentrez&rendertype=abstract

16. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M, Magnusdottir DN, Matthiasdottir S, et al. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet [Internet]* 2009 [cited 2013 Dec 23];41:460–4. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3664837&tool=pmcentrez&rendertype=abstract

17. Takahashi M, Saenko VA, Rogounovitch TI, Kawaguchi T, Drozd VM, Takigawa-Imamura H, Akulevich NM, Ratanajaraya C, Mitsutake N, Takamura N, Danilova LI, Lushchik ML, et al. The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Hum Mol Genet [Internet]* 2010 [cited 2013 Dec 23];19:2516–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20350937

18. Damiola F, Byrnes G, Moissonnier M, Pertesi M, Deltour I, Fillon A, Le Calvez-Kelm F, Tenet V, McKay-Chopin S, McKay JD, Malakhova I, Masyakin V, et al. Contribution of ATM and FOXE1 (TTF2) to risk of papillary thyroid carcinoma in Belarusian children exposed to radiation. *Int J Cancer [Internet]* 2014 [cited 2014 Feb 13];134:1659–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24105688

19. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Masson G, He H, Jonasdottir A, Sigurdsson A, Stacey SN, Johannsdottir H, Helgadottir HT, Li W, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat Genet [Internet]* 2012 [cited 2013 Dec 19];44:319–22. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3655412&tool=pmcentrez&rendertype=abstract

20. Köhler A, Chen B, Gemignani F, Elisei R, Romei C, Figlioli G, Cipollini M, Cristaudo A, Bambi F, Hoffmann P, Herms S, Kalemba MMM, et al. Genome-wide association study on

For Peer Review

differentiated thyroid cancer. *J Clin Endocrinol Metab [Internet]* 2013 [cited 2013 Dec 19];98:E1674–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23894154

21. Mancikova V, Cruz R, Inglada-Pérez L, Fernández-Rozadilla C, Landa I, Cameselle-Teijeiro J, Celeiro C, Pastor S, Velázquez A, Marcos R, Andía V, Álvarez-Escolá C, et al. Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations. *Int J Cancer [Internet]* 2015 [cited 2015 Jun 11];Available from: http://www.ncbi.nlm.nih.gov/pubmed/25855579

22. Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Pérez L, Schiavi F, Leskelä S, Pita G, Milne R, Maravall J, Ramos I, Andía V, Rodríguez-Poyo P, et al. The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS Genet [Internet]* 2009 [cited 2013 Dec 23];5:e1000637. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2727793&tool=pmcentrez&rendertype=abstract

23. Bullock M, Duncan EL, O'Neill C, Tacon L, Sywak M, Sidhu S, Delbridge L, Learoyd D, Robinson BG, Ludgate M, Clifton-Bligh RJ. Association of FOXE1 polyalanine repeat region with papillary thyroid cancer. *J Clin Endocrinol Metab [Internet]* 2012 [cited 2014 Jun 4];97:E1814–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22736773

24. Fan Y, Ding Z, Yang Z, Deng X, Kang J, Wu B, Zheng Q. Expression and clinical significance of FOXE1 in papillary thyroid carcinoma. *Mol Med Rep [Internet]* 2013 [cited 2013 Dec 24];8:123–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23715628

25. Bonora E, Rizzato C, Diquigiovanni C, Oudot-Mellakh T, Campa D, Vargiolu M, Guedj M, McKay JD, Romeo G, Canzian F, Lesueur F. The FOXE1 locus is a major genetic determinant for familial nonmedullary thyroid carcinoma. *Int J Cancer [Internet]* 2013 [cited 2013 Dec 24];Available from: http://www.ncbi.nlm.nih.gov/pubmed/24127282

26. He H, Li W, Liyanarachchi S, Jendrzejewski J, Srinivas M, Davuluri R V, Nagy R, de la Chapelle A. Genetic predisposition to papillary thyroid carcinoma: involvement of FOXE1, TSHR and a novel lincRNA gene, PTCSC2. *J Clin Endocrinol Metab [Internet]* 2014 [cited 2014 Oct 21];jc20142147. Available from: http://press.endocrine.org.gate2.inist.fr/doi/abs/10.1210/jc.2014-2147?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dpubmed

27. Jendrzejewski J, He H, Radomska HS, Li W, Tomsic J, Liyanarachchi S, Davuluri R V, Nagy R, de la Chapelle A. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type.

For Peer Review

Proc Natl Acad Sci U S A [Internet] 2012 [cited 2014 Feb 6];109:8646–51. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3365219&tool=pmcentrez&rendertype=abstract

28. Zheng J, Li C, Wang C, Ai Z. Common genetic variant on 14q13.3 contributes to thyroid cancer susceptibility: evidence based on 12 studies. *Mol Genet Genomics [Internet]* 2015 [cited 2015 Feb 11];Available from: http://www.ncbi.nlm.nih.gov/pubmed/25552255

29. Kang J, Deng X-Z, Fan Y-B, Wu B. Relationships of FOXE1 and ATM genetic polymorphisms with papillary thyroid carcinoma risk: a meta-analysis. *Tumour Biol [Internet]* 2014 [cited 2014 May 22];1–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24756757

30. de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet [Internet]* 2005 [cited 2014 Dec 29];37:1217–23. Available from: http://www.nature.com.gate2.inist.fr/ng/journal/v37/n11/full/ng1669.html

31. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda) [Internet]* 2011 [cited 2014 May 27];1:457–70. Available from: http://www.g3journal.org/content/1/6/457.full

32. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet [Internet]* 2009 [cited 2014 May 5];10:387–406. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2925172&tool=pmcentrez&rendertype=abstract

33. Li M-X, Yeung JMY, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet [Internet]* 2012 [cited 2014 Sep 11];131:747–56. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3325408&tool=pmcentrez&rendertype=abstract

34. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics [Internet]* 2010 [cited 2014 Jul 10];26:2336–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2935401&tool=pmcentrez&rendertype=abstract

35. Jones AM, Howarth KM, Martin L, Gorman M, Mihai R, Moss L, Auton A, Lemon C, Mehanna H, Mohan H, Clarke SEM, Wadsley J, et al. Thyroid cancer susceptibility

For Peer Review

polymorphisms: confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24. *J Med Genet [Internet]* 2012 [cited 2014 Jan 27];49:158–63. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3286794&tool=pmcentrez&rendertype=abstract

36. Di Vito M, De Santis E, Perrone GA, Mari E, Giordano MC, De Antoni E, Coppola L, Fadda G, Tafani M, Carpi A, Russo MA. Overexpression of estrogen receptor-α in human papillary thyroid carcinomas studied by laser- capture microdissection and molecular biology. *Cancer Sci [Internet]* 2011 [cited 2015 Mar 2];102:1921–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21707866

37. Xu S, Chen G, Peng W, Renko K, Derwahl M. Oestrogen action on thyroid progenitor cells: relevant for the pathogenesis of thyroid nodules? *J Endocrinol [Internet]* 2013 [cited 2015 Aug 10];218:125–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23645248

38. Onda M, Nagai H, Yoshida A, Miyamoto S, Asaka S-I, Akaishi J, Takatsu K, Nagahama M, Ito K, Shimizu K, Emi M. Up-regulation of transcriptional factor E2F1 in papillary and anaplastic thyroid cancers. *J Hum Genet [Internet]* 2004 [cited 2015 May 8];49:312–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15118916

39. Chiefari E, Brunetti A, Arturi F, Bidart J-M, Russo D, Schlumberger M, Filetti S. Increased expression of AP2 and Sp1 transcription factors in human thyroid tumors: a role in NIS expression regulation? *BMC Cancer [Internet]* 2002 [cited 2015 May 16];2:35. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=139985&tool=pmcentrez&rendertype=abstract

40. He H, Li W, Liyanarachchi S, Srinivas M, Wang Y, Akagi K, Wang Y, Wu D, Wang Q, Jin V, Symer DE, Shen R, et al. Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. *Proc Natl Acad Sci U S A [Internet]* 2015 [cited 2015 May 2];112:6128–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25918370

41. Morey L, Helin K. Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci [Internet]* 2010 [cited 2015 Mar 11];35:323–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20346678

42. Klug J, Beato M. Binding of YY1 to a site overlapping a weak TATA box is essential for transcription from the uteroglobin promoter in endometrial cells. *Mol Cell Biol [Internet]* 1996 [cited 2015 Oct 20];16:6398–407. Available from:

For Peer Review

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=231641&tool=pmcentrez&rendertype=abstract

43. Austen M, Lüscher B, Lüscher-Firzlaff JM. Characterization of the transcriptional regulator YY1. The bipartite transactivation domain is independent of interaction with the TATA box-binding protein, transcription factor IIB, TAFII55, or cAMP-responsive element-binding protein (CPB)-binding protei. *J Biol Chem [Internet]* 1997 [cited 2015 Oct 20];272:1709–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8999850

44. Pereda CM, Lesueur F, Pertesi M, Robinot N, Lence-Anta JJ, Turcios S, Velasco M, Chappe M, Infante I, Bustillo M, García A, Clero E, et al. Common variants at the 9q22.33, 14q13.3 and ATM loci, and risk of differentiated thyroid cancer in the Cuban population. *BMC Genet [Internet]* 2015;16. Available from: http://www.biomedcentral.com/1471-2156/16/22

45. Maillard S, Damiola F, Clero E, Pertesi M, Robinot N, Rachédi F, Boissin J-L, Sebbag J, Shan L, Bost-Bezeaud F, Petitdidier P, Doyon F, et al. Common Variants at 9q22.33, 14q13.3, and ATM Loci, and Risk of Differentiated Thyroid Cancer in the French Polynesian Population. *PLoS One [Internet]* 2015 [cited 2015 Apr 10];10:e0123700. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4388539&tool=pmcentrez&rendertype=abstract

46. Guignard R, Truong T, Rougier Y, Baron-Dubourdieu D, Guénel P. Alcohol drinking, tobacco smoking, and anthropometric characteristics as risk factors for thyroid cancer: a countrywide case-control study in New Caledonia. *Am J Epidemiol [Internet]* 2007 [cited 2015 Jun 3];166:1140–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2220030&tool=pmcentrez&rendertype=abstract

47. Truong T, Baron-Dubourdieu D, Rougier Y, Guénel P. Role of dietary iodine and cruciferous vegetables in thyroid cancer: a countrywide case-control study in New Caledonia. *Cancer Causes Control [Internet]* 2010 [cited 2015 Apr 26];21:1183–92. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3496161&tool=pmcentrez&rendertype=abstract

48. Wiersinga WM. Smoking and thyroid. Clin Endocrinol (Oxf) [Internet] 2013 [cited 2016 Jan 25];79:145–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23581474

49. Rinaldi S, Plummer M, Biessy C, Tsilidis KK, Østergaard JN, Overvad K, Tjønneland A, Halkjaer J, Boutron-Ruault M-C, Clavel-Chapelon F, Dossus L, Kaaks R, et al. Thyroid-

For Peer Review

stimulating hormone, thyroglobulin, and thyroid hormones and risk of differentiated thyroid carcinoma: the EPIC study. J Natl Cancer Inst [Internet] 2014 [cited 2016 Jan 25];106:dju097. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24824312

50. Yip L, Nikiforova MN, Yoo JY, McCoy KL, Stang MT, Armstrong MJ, Nicholson KJ, Ohori NP, Coyne C, Hodak SP, Ferris RL, LeBeau SO, et al. Tumor genotype determines phenotype and disease-related outcomes in thyroid cancer: a study of 1510 patients. Ann Surg [Internet] 2015 [cited 2016 Jan 25];262:519–25; discussion 524–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26258321

For Peer Review

Table 1: Characteristics of the study sample stratified by ethnicity

Characteristics	Europeans		Melanesians	
	Cases	Controls	Cases	Controls
Mean age (years)	50.9	51.4	45.5	45.7
Sex (%)				
Females	399(75.2)	459(64.4)	148(94.9)	104(92.8)
Males	109(24.8)	167 (35.6)	8(5.1)	8(7.1)
Study population				
CATHY	481(94.7)	544(87.0)	-	-
NC	29(5.3)	82(13.0)	156(100)	112(100)
Area of residence in CATHY				
Ardennes	111(21.8)	99(15.8)	-	-
Calvados	152(29.9)	189(30.2)	-	-
Marne	218(42.9)	256(40.9)	-	-
Province of residence in NC[1]				
North	3(0.11)	6(0.07)	32(20.5)	26(23.2)
South	23(0.85)	75(0.91)	54(34.6)	47(41.9)
Island	1(0.04)	1(0.04)	70(44.8)	39(34.8)
Type of cancer (%)				
papillary	447(88.0)	-	131(84.0)	-
vesicular	61(12.0)	-	25(16.0)	-
Total (N)	508	626	156	112

(1)NC: New Caledonia

Table 2. Single markers association results for DTC, for a selection of SNPs (GWAS SNP and possible functional SNP that are highly correlated to a top SNP) for loci 9q22.33 and 14q13.3 in Europeans and Melanesians.

					Europeans				Melanesians			
			Allele		MAF[4]				MAF[4]			
Locus	SNPs	Position	Ref[3]	risk	Case	Control	OR (95% IC)	p-values	Case	Control	OR(95% IC)	p-value
9q22.33	rs13295081[1]	10055901	C	T	0.49	0.40	1.44 (1.22 - 1.70)	5.46×10^{-6}	-	-	-	-
	rs965513[2]	10055610	G	A	0.43	0.33	1.52 (1.28 - 1.81)	5.87×10^{-6}	0.38	0.30	1.43 (1.00 - 2.07)	0.03
	rs10759944	10055697	G	A	0.43	0.34	1.48 (1.25 - 1.76)	2.25×10^{-5}	0.39	0.31	1.42 (0.95 – 2.05)	0.03
	rs1867277[2]	10061591	G	A	0.46	0.38	1.38 (1.17 - 1.63)	2.59×10^{-4}	0.36	0.29	1.34 (0.93 - 1.94)	0.06
14q13.3	rs1755774	36688516	A	G	0.29	0.34	0.76 (0.63- 0.92)	7.29×10^{-3}	0.53	0.70	0.50 (0.35 - 0.71)	1.62×10^{-5}
	rs2787423	36687054	G	A	0.29	0.34	0.78 (0.65 - 0.94)	8.78×10^{-3}	0.54	0.69	0.51 (0.35 - 0.73)	3.69×10^{-5}
	rs7494749	36615427	G	A	0.38	0.45	0.73 (0.62 - 0.87)	4.34×10^{-4}	0.64	0.76	0.57 (0.39 - 0.83)	7.94×10^{-4}
	rs944289[2]	36649246	T	C	0.34	0.41	0.74 (0.61 - 0.89)	4.78×10^{-4}	0.62	0.74	0.57 (0.39 - 0.83)	1.13×10^{-3}
	rs116909374[2]	36738362	C	T	0.05	0.03	1.97 (1.26 - 3.10)	3.25×10^{-3}	0.01	0.004	1.44 (0.13 - 15.96)	0.76
	rs999460	36984511	C	T	0.34	0.39	0.8 (0.68 - 0.96)	5.18×10^{-3}	0.21	0.19	1.14 (0.74 - 1.76)	0.59

[1]rs13295081 did not pass quality control in Melanesians.[2]GWAS SNPs, [3]ref, reference allele.[4]MAF, minor allele frequency and frequency of risk allele. The odds ratios are adjusted for age, sex and area of residence.

Table 3: OR estimations for common haplotypes based on rs944289, rs116909374, rs999460 at locus 14q13 in Europeans.

				frequencies					
Haplotype	rs944289	rs116909374	rs999460	cases	controls	pools	OR	IC	p-value
5	T	C	C	0.39	0.36	0.38	ref	-	-
1	C	C	C	0.23	0.23	0.23	0.93	(0.74-1.15)	0.48
6	T	C	T	0.22	0.21	0.21	0.93	(0.74-1.16)	0.52
2	**C**	**C**	**T**	**0.1**	**0.17**	**0.13**	**0.51**	**(0.38-0.68)**	**3.11×10^{-6}**
8	**T**	**T**	**C**	**0.03**	**0.01**	**0.02**	**2.47**	**(1.25-4.86)**	**8.9×10^{-3}**
7	C	T	T	0.02	0.006	0.01	3.16	(1.27-7.86)	0.01

The odds ratios are adjusted for age, sex and area of residence.

For Peer Review

Legends to figure

Fig. 1. Manhattan plot of the DTC risk loci 9q22 and 14q13 and heatmap of correlation between SNPs at each locus for Europeans (1A) and for Melanesians (1B).

Genotyped and imputed SNPs are plotted according to their chromosomal positions on the x axis and their overall p value (-log10 p values) on the y axis. The black rhombuses indicate observed SNPs and the red circle, the imputed SNP. The red line indicates the threshold of significance, the red arrow indicates the GWAS SNPs and, the blue arrow indicates the most robust associated SNPs in our study sample. On the heat map, the blue Arrow indicates the gene and its position on the chromosome. A.) Manhattan plot of locus and heatmap of correlation between SNPs at locus 9q22 in Europeans. B.) Manhattan plot of locus and heatmap of correlation between SNPs at locus 9q22 in Melanesians. C.) Manhattan plot of locus and heatmap of correlation between SNPs at locus 14q13 in Europeans. D.) Manhattan plot of locus and heatmap of correlation between SNPs at locus 14q13 in Melanesians.

Fig. 2. Regional plot of results of single-marker association and, results of conditional regression, at loci 9q22 and 14q13, in Europeans.

Figure 2A and 2D: plot of –log10 (p-value) of a single markers association results at loci 9q22 (figure 2A) and 14q13 (figure 2D) in Europeans. Figure 2B and 2C indicate the results of –log10 (p-value) of SNPs after conditioning the regression test on the most robust signal rs13295081 at locus 9q22 (2B) or on the previously describe functional SNP rs1867277 (2C) in Europeans. Figure 2E and 2F represented the results of –log10 (p-value) of SNPs after conditioning the regression test on the most robust signal rs7494749 at locus 14q13 (2E) or on the GWAS SNP rs116909374 (2F) in Europeans.

For Peer Review

On Y axis, are the –log10 (p-value); on X axis, the position of SNPs and gene on the chromosome. The color of each SNP spot reflects its r^2 with the most associated SNP (in purple). The correlation between SNP is ranged from high (in red) to low (in blue); the arrows show the GWAS SNP in the region. The square indicated observed SNPs and the circle, the imputed SNPs.

Fig. 3. Regional plot of results of single-marker association and, results of conditional regression, at locus 14q13, in Melanesians.

Figure 3A: plot of –log10 (p-value) of a single markers association results at loci 14q13 in Melanesians. Figure 3B: plot of –log10 (p-value) of SNPs after conditioning the regression test on the most robust signal rs1755774 at locus 14q13 in Melanesians. On Y axis, the –log10 (p-value); on X axis, the position of SNPs and gene on the chromosome. The color of each SNP spot reflects its r^2 with the most associated SNP in purple. The correlation is ranged from high (in red) to low (in blue); the arrows show the GWAS SNP in the region. The squares represent observed SNPs and the circles, the imputed SNPs.

For Peer Review

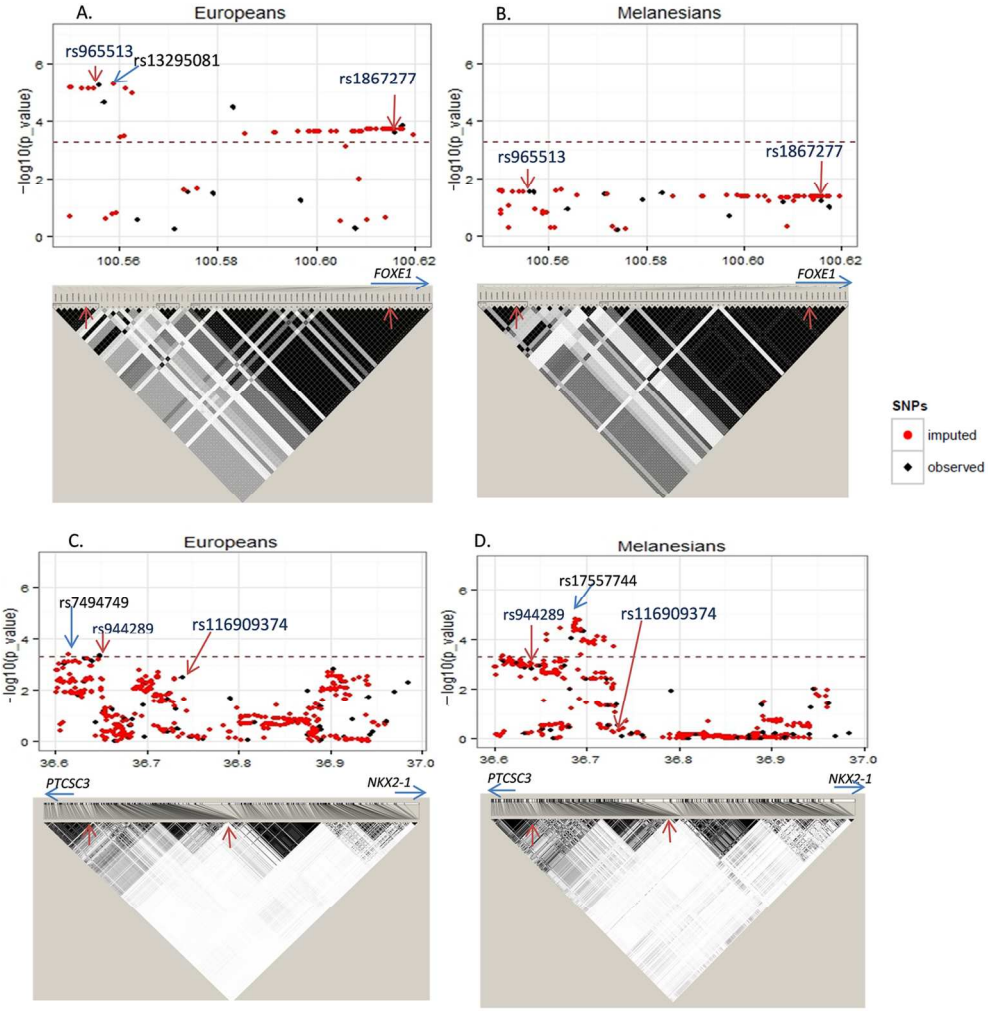


Fig. 1. Manhattan plot of the DTC risk loci 9q22 and 14q13 and heatmap of correlation between SNPs at each locus for Europeans (1A) and for Melanesians (1B).
Genotyped and imputed SNPs are plotted according to their chromosomal positions on the x axis and their overall p value (-log10 p values) on the y axis. The black rhombuses indicate observed SNPs and the red circle, the imputed SNP. The red line indicates the threshold of significance, the red arrow indicates the GWAS SNPs and, the blue arrow indicates the most robust associated SNPs in our study sample. On the heat map, the blue Arrow indicates the gene and its position on the chromosome. A.) Manhattan plot of locus and heatmap of correlation between SNPs at locus 9q22 in Europeans. B.) Manhattan plot of locus and heatmap of correlation between SNPs at locus 9q22 in Melanesians. C.) Manhattan plot of locus and heatmap of correlation between SNPs at locus 14q13 in Europeans. D.) Manhattan plot of locus and heatmap of correlation between SNPs at locus 14q13 in Melanesians.

131x135mm (300 x 300 DPI)

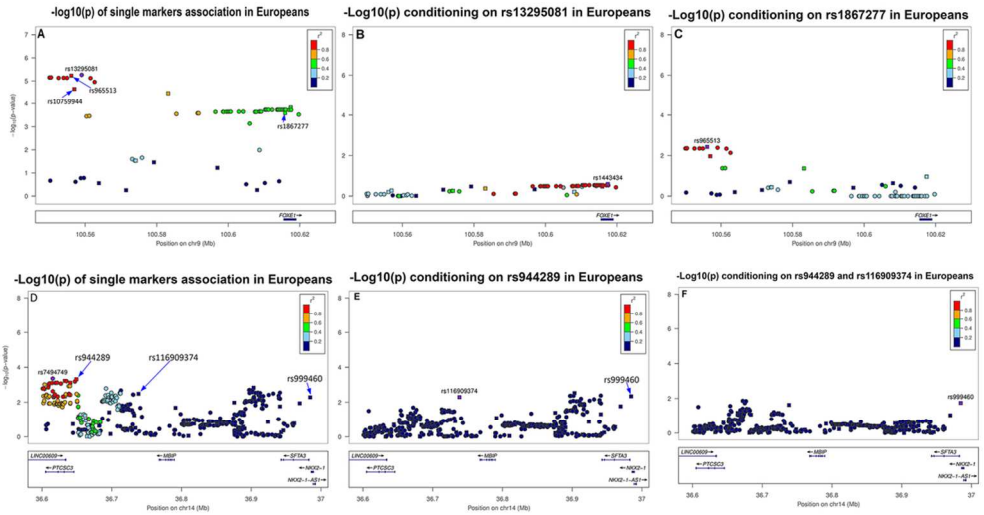


Fig. 2. Regional plot of results of single-marker association and, results of conditional regression, at loci 9q22 and 14q13, in Europeans.

Figure 2A and 2D: plot of –log10 (p-value) of a single markers association results at loci 9q22 (figure 2A) and 14q13 (figure 2D) in Europeans. Figure 2B and 2C indicate the results of –log10 (p-value) of SNPs after conditioning the regression test on the most robust signal rs13295081 at locus 9q22 (2B) or on the previously describe functional SNP rs1867277 (2C) in Europeans. Figure 2E and 2F represented the results of –log10 (p-value) of SNPs after conditioning the regression test on the most robust signal rs7494749 at locus 14q13 (2E) or on the GWAS SNP rs116909374 (2F) in Europeans.

On Y axis, are the –log10 (p-value); on X axis, the position of SNPs and gene on the chromosome. The color of each SNP spot reflects its r^2 with the most associated SNP (in purple). The correlation between SNP is ranged from high (in red) to low (in blue); the arrows show the GWAS SNP in the region. The square indicated observed SNPs and the circle, the imputed SNPs.

108x58mm (300 x 300 DPI)

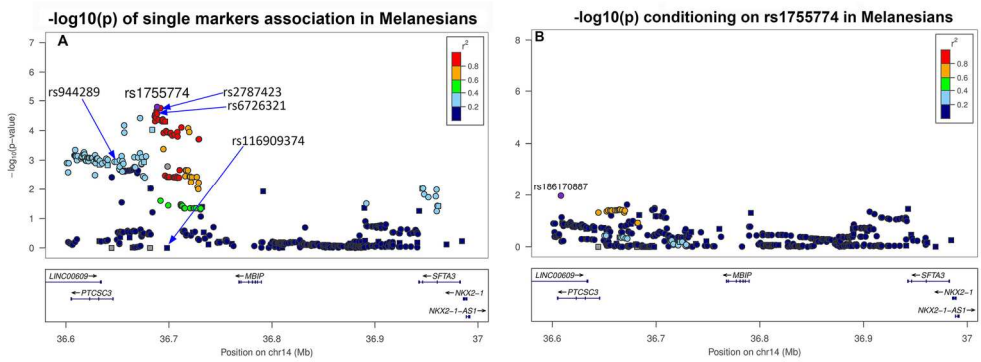


Fig. 3. Regional plot of results of single-marker association and, results of conditional regression, at locus 14q13, in Melanesians.

Figure 3A: plot of –log10 (p-value) of a single markers association results at loci 14q13 in Melanesians. Figure 3B: plot of –log10 (p-value) of SNPs after conditioning the regression test on the most robust signal rs1755774 at locus 14q13 in Melanesians. On Y axis, the –log10 (p-value); on X axis, the position of SNPs and gene on the chromosome. The color of each SNP spot reflects its r^2 with the most associated SNP in purple. The correlation is ranged from high (in red) to low (in blue); the arrows show the GWAS SNP in the region. The squares represent observed SNPs and the circles, the imputed SNPs.

169x67mm (300 x 300 DPI)

er Review

Role of *GSTM1* and *GSTT1* genotypes in differentiated thyroid cancer and interaction with lifestyle, hormonal and reproductive characteristics: a study in Europeans and Melanesians from Metropolitan France and New Caledonia

Catherine Tcheandjieu[1], Emilie Cordina-Duverger[1], Marie Sanchez[1], Claire Mulot[2], Dominique Baron-Dubourdieu[3], Anne-Valérie Guizard[4,5], Claire Schvartz[6], Pierre Laurent-Puig[2], Pascal Guénel[1]* and Thérèse Truong[1]*

*as co-last authors

Author affiliations

1: CESP - Cancer and Environment team, INSERM U1018, Université Paris-Sud, Université Paris-Saclay, Villejuif, France

2: Center for Biological Resources EPIGENETEC, Université Paris Descartes, Inserm UMR 5775, Paris, France

3: Laboratory of Anatomy and Cytopathology, Noumea, New Caledonia, France

4: Registre Général des tumeurs du Calvados, Centre François Baclesse, Caen, France

5: U1086 Inserm-UCNB, Cancers and Prevention, Caen, France

6: Centre de Lutte Contre le Cancer Jean GODINOT, Reims, France

Corresponding author:

Pascal Guenel, MD, PhD

Epidemiology of Cancers: Genes and Environment

16 avenue Paul Vaillant Couturier - Villejuif, France

Tel +33 1 45 59 50 31

Email: pascal.guenel@inserm.fr

Key Word: Thyroid cancer, Cancer genetics, case-control study, *GSTM1* and *GSTT1*

Acknowledgements

The NC study was supported by grants from the “Fondation de France” (FDF), the “Association pour la Recherche contre le Cancer” (ARC) and the “Agence Française de Sécurité Sanitaire de l’Environnement et du Travail” (AFSSET). The CATHY study was funded by AFSSET, Ligue Nationale Contre le Cancer, Institut National du Cancer, EDF and FDF.

Abstract:

GSTM1 and *GSTT1* genes encode for phase 2 enzymes involved in detoxification pathways. They exhibit copy number variation (CNV) and a dose-effect relationship between gene copy number and enzymatic activity, with gene deletion eliminating the gene function. *GSTM1* and *GSTT1* genotype has been associated with risk of several cancers, but their role in differentiated thyroid carcinoma (DTC) is not known. We investigated how *GSTM1* and *GSTT1* genotypes modify DTC risk using data from two case-control studies in France and New Caledonia, including 505 cases and 622 controls of European descent, and 156 cases and 114 controls of Melanesian descent. Analyses were conducted separately in each ethnic group. Gene copy number were not associated with DTC risk, but modified the association of DTC with several risk factors for the disease. In particular, the positive association of DTC with body mass index (BMI) and the inverse association of DTC with alcohol drinking were strengthened in GST-null individuals, among both Europeans and Melanesians. These findings suggest that *GSTM1* and *GSTT1* genotypes interact with risk factors for DTC. Different frequencies of genotype distribution between populations may explain part of the variations in DTC incidence rates.

INTRODUCTION

Thyroid cancer is the most common malignancy of the human endocrine system. Approximately 90% of thyroid cancers are represented by differentiated thyroid carcinomas (DTC), which are subdivided into papillary or follicular carcinomas. The incidence of thyroid cancer is characterized by considerable ethnic and geographic variations (Moore et al. 2010; Anon n.d.). One of the highest incidence rates has been observed in Melanesian women of New Caledonia (71.4 per 100 000 person-years) (Moore et al. 2010; Anon n.d.; Truong et al. 2007). In metropolitan France, the incidence rate was 12.6/100 000 in women and 4.1/100 000 in men in 2012 (Anon n.d.). Although part of the variations in incidence are explained by differences in medical screening practices enhancing detection of small dormant micro carcinomas (Vaccarella et al. 2015; Kitahara & Sosa 2016; Colonna et al. 2015), the causes of the disease have not been fully elucidated.

Exposure to ionizing radiations during childhood is a well-established risk factor for DTC. Overweight, iodine intake deficiency, reduced consumption of alcohol and tobacco, high parity and late age at menarche have also been associated with DTC risk (Truong et al. 2005; Leux & Guénel 2010; Cléro et al. 2010; Kabat et al. 2012; Kitahara, Linet, et al. 2012). It was shown that thyroid cancer is one of the cancers with the highest familial risk (Hemminki & Vaittinen 1999), suggesting a role for genetic risk factors. However, only few variants have been identified so far (Landa & Robledo 2011) and to our knowledge no convincing gene-environment interactions have been reported for DTC.

Genes that encode for enzymes involved in the metabolism of exogenous or endogenous substances are suspected determinants of thyroid or other cancers (Hirvonen 1999). Genes of the Glutathione S-Transferase (GST) family, such as *GSTM1* and *GSTT1* that encode for phase II enzymes are involved in the detoxification and neutralization of environmental carcinogens, therapeutic drugs, or products of oxidative stress. The enzyme activity depends in a dose-dependent manner on the number of copies of *GSTM1* and *GSTT1* (copy number variations, CNVs) in an individual's genome (McLellan et al. 1997; Sprenger et al. 2000) with complete gene deletion resulting in a loss of function. Thus, it is believed that gene deletion may confer a higher vulnerability to carcinogen exposure (Nørskov et al. 2011).

In a recent meta-analysis of 12 epidemiological studies on the association between *GSTM1* or *GSTT1*-null genotypes and DTC, no overall association was detected (J. Li et al. 2012). However,

there was a statistically significant heterogeneity between studies, possibly resulting from differences in the prevalence of exposure to risk factors and frequencies of *GSTM1*- or *GSTT1*-null genotypes (Saitou & Ishida 2015; Geisler 2001). Besides genotype data on GSTs, assessing exposure to DTC risk factors and investigating gene-environment interactions is thus critical to understand the role of GST in the etiology of DTC. In addition, accounting for CNVs of GSTs may provide further insight by investigating dose-response relationships.

In the present study, we analyzed the association between CNVs of *GSTM1* and *GSTT1* and DTC incidence in Europeans and in Melanesians from New Caledonia, two populations with distinct genetic background and different patterns of exposure to DTC risk factors. We also investigated in these two populations separately, the modification effects of *GSTM1* and *GSTT1* genotypes on the association between DTC and previously suggested risk factors such as body mass index (BMI), alcohol consumption and cigarette smoking, as well as in women age at menarche, parity, oral contraception and menopausal status.

MATERIAL AND METHODS

Study population

We used data from two case-control studies on thyroid cancer conducted in metropolitan France (CATHY study) and in New Caledonia (NC study), a French territory in South Pacific. All participants signed informed consent.

The CATHY study is a population-based case-control study conducted in Marne, Ardennes and Calvados, three French administrative areas ("départements") covered by a cancer registry. All patients aged 25 and over residing in these areas and diagnosed with DTC between 2002 and 2007 were eligible for inclusion in the group of cases. Controls were matched to the cases by sex, age (5-year age groups) and study area. Controls were recruited using a telephone number selected at random. For recruiting the controls, predefined numbers of subjects by socioeconomic status (SES), reflecting the distribution by SES in the study areas, were applied to prevent possible selection bias that may arise from differential participation rates across SES categories. From 621 cases and 706 controls recruited for the study, DNA samples from saliva (Oragen®) were obtained for 482 cases and 565 controls of self-declared European ancestry.

The New Caledonia (NC) study is a country-wide population-based case-control study (Truong et al. 2007; Truong et al. 2005; Guignard et al. 2007). The case group included patients with DTC diagnosed between 1993 and 1999, who had been living for at least 5 years in NC at the time of diagnosis. The cases were identified from the two pathology laboratories of NC completed by active searches in medical records of the main hospitals. Age and sex matched controls were randomly selected from recently updated electoral rolls. A total of 332 cases and 412 controls were included; among those, 42 cases and 133 controls self-declared Europeans and 206 cases and 156 controls self-declared Melanesians. DNA was obtained from saliva samples (Oragen®) available for 284 Melanesians (164 cases and 120 controls) and 108 Europeans (27 cases and 81 controls).

In the two studies, information on ethnicity, personal and familial history of thyroid disease, menstrual and reproductive factors, exogenous hormone use, weight, height, dietary habits, alcohol intake, tobacco smoking, residential and occupational histories was collected during in-person interviews by trained interviewers.

The year of reference was defined as the year of diagnosis for the cases and their matched controls. BMI was defined as weight (kg) at reference year divided by the height squared (m^2) (4,18). Alcohol drinking was assessed as the lifetime average number of glasses per week, elicited from the questionnaire for each type of beverage separately (beer, wine, aperitif, and liqueur). Because the ethanol content is approximately the same for an ordinary glass of any alcoholic beverage, the total number of drinks per week was used as an indicator of total alcohol intake. Subjects who had quitted smoking for at least one year at reference year were classified as former smokers. Pack-years of cigarette smoking were calculated from the total number of years of tobacco smoking and from the number of cigarettes smoked per day for each smoking period. In these calculations we assumed that one pipe or one cigarillo were equivalent to 2 cigarettes.

In women, the number of full term pregnancies was calculated as the sum of the live born children and stillbirths. Women were considered as users of oral contraceptives if they had ever taken pills for at least 6 months. Women were considered postmenopausal if they reported no menstruation for at least one year or used menopausal hormone therapy (MHT) (natural menopause), or if they had bilateral ovariectomy (artificial menopause).

Genotyping

DNA was extracted from salivary DNA samples; *GSTM1* and *GSTT1* CNVs were determined using TaqMan gene copy number detection designed by Applied Biosystems; the genotyping was processed by Integragen (Evry, France). The real-time PCR were run on an Applied Biosystems 7900HT Fast system with gene-specific primers for *GSTM1* probes (Hs02575461_cn) and *GSTT1* probes (Hs00010004_cn) along with primer for the RNase P gene as the reference gene; each sample were run in triplicate using 50 ng of genomic DNA. The gene specific primers were first validated in 90 CEPH individuals and genotyping was carried out blindly as to the case-control status. The CopyCaller software V1 (Applied Biosystems) was used to quantify the number of copies in each sample.

The genotype was missing for either *GSTM1* in 65 subjects or *GSTT1* in 75 subjects. In addition, genotypes for both *GSTM1* and *GSTT1* could not be determined in 28 subjects (5 cases/23 controls) of the CATHY study and 12 subjects (8 cases/6 controls) of the NC study and were excluded from the analysis. As the frequencies of *GSTM1* and GSTT1 CNVs were similar in Europeans from the CATHY and the NC studies, they were combined in the analysis.

In total, 1127 Europeans (504 cases and 620 controls) and 270 Melanesians (156 cases and 114 controls) were included in the analyses.

Statistical Analysis

Odds ratios for *GSTM1* and *GSTT1* genotypes were calculated, using unconditional logistic regression, comparing categories defined by the gene copy number coded as a categorical variable (0, 1 or ≥2 copies), and adjusting for age (5-year intervals), sex and area of residence. Europeans and Melanesians were analyzed separately. We also compared non-carriers (null genotype) to carriers of at least one copy of the gene (non-null genotype). Analysis was also conducted using the total copy number of both *GSTM1* and *GSTT1* genes.

We tested the interaction between *GSTM1* and *GSTT1* genotypes and tobacco smoking, alcohol intake, BMI, hormonal and reproductive factors using the likelihood ratio test comparing models with and without the interaction term. Tobacco smoking, alcohol intake and BMI were analyzed in men and women combined. Hormonal and reproductive factors were analyzed in women only. To account for the potential confounding effect of hormonal and reproductive factors, we also conducted the analysis for tobacco smoking, alcohol drinking and BMI in women separately. The results were similar to those obtained for men and women combined, and are not shown.

RESULTS

Selected characteristics of the two study populations are shown in Table 1. Among Europeans, the odds ratios for DTC increased slightly in individuals with BMI ≥ 25 kg/m^2. In women, age at menarche ≥ 13 years and menopause were associated with increased odds ratios at the limit of statistical significance. The odds ratio was below unity for alcohol drinkers and users of oral contraceptives as compared to non-users, and it was close to one for parous compared to non-parous women, whatever the number of children. Among Melanesians, the odds ratio clearly increased with BMI, but no statistically significant association was detected for other risk factors. Associations of DTC with lifestyle habits or hormonal and reproductive risk factors are described in more details somewhere else for the CATHY (Cordina-Duverger et al., submitted) and the NC (Guignard et al. 2007; Truong et al. 2005) studies.

The proportion of *GSTM1-null* genotypes among controls was 51.2% in Europeans and 82.5% in Melanesians. The corresponding figures for the *GSTT1-null* genotype were 20.5% and 39.4%, respectively. The proportion of controls with the null genotype for both *GSTM1* and *GSTT1* was 10.1% in Europeans and 36.4% in Melanesians. Accordingly, there were less Melanesians with 2 or more copies of the genes (Table 2). Table 2 also shows the odds ratios for *GSTM1* and *GSTT1* genotypes, comparing DTC risk for the null versus non-null genotype, or by number of gene copies. No statistically significant association was observed either by considering each gene separately or in combination.

Odds ratios for BMI, alcohol drinking and tobacco smoking are shown separately for the null and the non-null genotypes in Table 3. In both ethnic groups, obese subjects (BMI ≥ 30 kg/m^2) with *GSTM1*- or *GSTT1*-null genotypes had higher odds ratios than obese subjects with non-null genotypes. In obese subjects with deletion of both genes, the odds ratios were 4.09 in Europeans and 13.28 in Melanesians, while the corresponding odds ratio in obese subjects with at least one copy of either gene were 1.20 and 1.26, respectively (Table 3). However, none of the p-values for interaction between BMI and genotypes was statistically significant. Among Europeans, the odd ratios for obesity increased with decreasing number of copies of *GSTT1*. This trend was not seen for *GSTM1* (supplementary table 1).

Odds ratios for alcohol drinking were below unity in both ethnic groups. This decrease was particularly noticeable in subjects with *GSTM1* or *GSTT1*-null genotypes (Table 3). Among Europeans, the p-values for interaction between genotype and number of glasses per day

were 0.048 for *GSTM1* and 0.03 for *GSTT1*. Among Melanesians, the p-values for interaction between genotype and ever alcohol drinking were not statistically significant. In subjects with null genotypes for both *GSTM1* and *GSTT1,* the odds ratio for ever drinking alcohol were 0.28 (95%CI: 0.08, 1.02) and 0.24 (95%CI: 0.07, 0.81) in Europeans and Melanesians, respectively (p for interaction not significant) (Table 3). For tobacco smoking, no difference between *GSTM1* and/or *GSTT1* emerged (Table 3).

The odds ratios for hormonal and reproductive risk factors of DTC among women, after stratification on null and non-null GSTs genotypes are shown in Table 4. For *GSTM1*, no statistically significant interaction was detected in either ethnic group with age at menarche, parity, oral contraceptive, or menopausal status. For *GSTT1* among Europeans, the odds ratios for age at menarche >13 years were 0.57 (95% CI) and 1.79 (95% CI) in women with *GSTT1* null- and non-null genotype respectively (p for interaction 0.02); similarly, the odds ratios for parous vs. nulliparous women were 5.71 [95%CI 1.29, 25.2] and 0.86 (NS) for *GSTT1*-null and non-null genotypes, respectively (p for interaction 0.02). These findings for *GSTT1* were not seen in Melanesian women.

DISCUSSION

This study conducted in Europeans and Melanesians provides new clues for understanding the etiology of thyroid cancer. In these two populations, there was convincing evidence that BMI and alcohol consumption were more strongly associated with DTC in individuals with *GSTM1* and *GSTT1*-null than non-null genotypes, suggesting that gene deletion may enhance the effect of these risk factors in thyroid carcinogenesis. Statistically significant interactions between age at menarche or parity and *GSTT1* genotype was observed in European but not in Melanesian women, possibly because of small numbers.

In a meta-analysis of 12 studies conducted in several countries (Brazil, Italy, Spain, Portugal, Poland, Turkey, Saudi Arabia and China), no overall association was observed between DTC risk and GST genotypes (J. Li et al. 2012). However significant heterogeneity between study findings was reported. Assuming gene-environment interaction between DTC risk factors and GST genotypes, this heterogeneity suggests that the distribution of exposure to risk factors and of GST genotypes in the populations studied were dissimilar (J. Li et al. 2012). The present study in two populations with different genetic background and distinct patterns of exposure to environmental factors was conducted to shed further light on gene-environment interaction with GST genotypes.

CNV distribution in Europeans and Melanesians

The distribution of CNVs of *GSTM1* and *GSTT1* among Europeans in our study were similar to previous reports (Huang et al. 2009; Nørskov et al. 2009). To our knowledge, this is the first report of CNV distribution in Melanesians. The frequency of null genotypes was notably higher in Melanesians than in Europeans, a finding consistent with data in other Pacific islanders (Rebbeck 1997).

BMI

Overweight has consistently been associated with risk of DTC (Kitahara et al. 2014; Kitahara et al. 2011). In the present study, obesity (BMI>30 kg/m^2) was more strongly associated with DTC in *GST-null* than non-null individuals, particularly when both genes were deleted. Although the interaction between BMI and GST genotypes was not statistically significant, this finding was consistent in both ethnic groups suggesting a true modification effect of the BMI-DTC association by the genes. To our knowledge, the only study conducted

in Korea that investigated the joint effect of BMI and GST in DTC found no indication of an interaction (Kweon et al. 2014), but the low prevalence of obesity in this population may explain the null results. Studies with sufficient sample size particularly in populations with high prevalence of obesity would be helpful to confirm our finding.

Obesity may increase the risk of DTC through complex mechanisms (Marcello et al. 2014). The chronic systemic inflammation induced by obesity may favor cancer development through formation of reactive oxygen species (ROS) (Pazaitou-Panayiotou et al. 2013). Because *GSTM1* and *GSTT1* enzymes contribute to reducing oxidative stress by conjugating and eliminating ROS products (Hayes & Strange 2000), obese subjects with deletion of *GSTM1* and *GSTT1* could be at particularly high risk of DTC.

GST-null genotypes and obesity were more frequently observed in Melanesians than in Europeans. Interaction between obesity and *GSTM1*/*GSTT1*-null genotype may thus explain part of the elevated incidence of thyroid cancer among Melanesians (Huang et al. 2009). We estimated that 9.9% of the difference in incidence rates between Melanesians and Europeans was explained by the greater the proportion of obese subjects (BMI ≥ 30 kg/m2) with deletion of both GSTM1 and GSTT1 in Melanesians (6.6%) than in Europeans (1.0%) (Supplementary table 2).

Interaction with alcohol drinking and tobacco smoking

The inverse association of alcohol drinking with DTC observed in our data is in agreement with previous studies (Kabat et al. 2012; Kitahara, Linet, et al. 2012; Meinhold et al. 2009; Sen et al. 2015; Cho & Kim 2014) and a recent meta-analysis (Hong et al. 2016). We found that this inverse association was greater in *GSTM1*/*GSTT1*-null than non-null subjects, particularly in those with a deletion of both genes. Interaction between alcohol intake and *GSTM1* or *GSTT1* gene was reported previously in cancer of the breast, lung and stomach (Zheng et al. 2003; Lao et al. 2014; Mota et al. 2015). To our knowledge, this is the first report such an interaction in thyroid cancer.

The underlying mechanisms explaining the protective effect of alcohol drinking in thyroid carcinogenesis are still unclear and potentially complex. Some studies suggested that free radical generated from alcohol metabolism might have either a direct toxic effect on the thyroid tissues or disturb the hypothalamus–pituitary–thyroid axis (Sen et al. 2015; Valeix et

al. 2008; Hermann et al. 2002). However, it remains unclear how this can lead to a decrease risk of DTC (Mansoori & Jain 2015) and do not explain a synergistic effect of alcohol and GST-null genotypes.

Unlike alcohol, *GSTM1* or *GSTT1* were not found to modify the association between tobacco smoking and DTC. Interaction between GST genes and smoking was also investigated in a previous study but no evidence of interaction was reported (22).

Interaction with hormonal and reproductive factors

Given the higher incidence of DTC in women than in men, female sex hormones have been suspected to play a major role in thyroid carcinogenesis. We found that late age at menarche was associated with increased DTC risk. This association was reported by several studies (Negri et al. 1999; Horn-Ross et al. 2011; Truong et al. 2005), but no conclusive evidence was found in a recent meta-analysis (Cao et al. 2015). We also reported that ever users of oral contraceptives were at decreased risk, a finding supported by recent studies (Zamora-Ros et al. 2015; Xhaard et al. 2014). Conversely, parity was not associated with DTC in our data. There was also some indication that menopause increases DTC risk. These findings are reported in details somewhere else in European (Cordina-Duverger et al.) and Melanesian women (Truong et al. 2005)

Our study is the first to investigate the interaction between GSTs genotype and reproductive factors in thyroid cancer. No interaction between *GSTM1* and age at menarche, use of oral contraceptives, parity or menopausal status was observed in our data for either ethnic groups. A statistically significant interaction of *GSTT1* with age at menarche and with parity was observed in European, but not Melanesian women.

The associations of DTC with reproductive factors have been inconsistent in epidemiological studies, and the mechanisms leading to thyroid cancer are still unclear. It was hypothesized that elevated estrogen levels during pregnancy may influence the proliferation of malignant thyroid cells (Xu et al. 2013; Derwahl & Nicula 2014; Di Vito et al. 2011). Estrogen metabolism results in generation of quinones and others cytotoxic molecules which are detoxified by GSTs genes (Mitrunen & Hirvonen 2003). Genetic variation of these genes could modify the ability to detoxify products of estrogen metabolism. Interaction between parity and GSTs

polymorphisms has also been reported in one previous study on breast cancer (Park et al. 2003).

Strengths and limitations

Our study was based on a relatively high number of subjects from 2 population-based case-control studies. With the exception of one study in Korea that included 1372 cases and 1669 controls (Kweon et al. 2014), previous studies of DTC risk in relation to *GSTM1* or *GSTT1* genotype had limited sample size with less than 300 cases. Unlike previous studies contrasting DTC risk in null vs. non-null genotypes, copy number variations of *GSTM1* and *GSTT1* were also examined in our study to investigate a potential dose-effect relationship between enzymatic activity and cancer risk. However, such a dose-effect association was not observed between copy number of the genes and DTC risk. In addition, we also examined gene-environment interactions between *GSTM1* and *GSTT1* genotypes and environmental or hormonal risk factors in DTC. This study used data collected from two case-control studies in Europeans and Melanesians with a population-based design and with an exhaustive identification of thyroid cancer cases. The analyses of risk factors in these two ethnic groups with diverse prevalence of exposure to risk factors and frequency of GST null genotypes could help to explain part of the difference in DTC incidence between populations.

The main limitation of our study is the limited statistical power in some strata. A potential recall bias, inherent to case-control studies, may have occurred when assessing exposure to several risk factors, but this is unlikely for most risk factors investigated such as parity, age at menarche, or menopausal status. Differential misclassification was also unlikely since cases and controls were interviewed in the same conditions by trained interviewers.

In conclusion, our results suggest that GSTM1 and GSTT1 may modify the associations between DTC risk and obesity and alcohol consumption, and possibly hormonal factors. Disparities in frequencies of *GSTM1* and *GSTT1* deletion and in prevalence of obesity between populations across the world, may explain part of the differences in thyroid cancer incidence.

The authors declare that they have no conflict of interest

References

Adjadj, E., Schlumberger, M. & de Vathaire, F., 2009. Germ-line DNA polymorphisms and susceptibility to differentiated thyroid cancer. *The Lancet Oncology*, 10(2), pp.181–190. Available at: http://dx.doi.org/10.1016/S1470-2045(09)70020-8.

Akulevich, N.M. et al., 2009. Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma. *Endocrine-related cancer*, 16(2), pp.491–503. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19286843 [Accessed July 1, 2014].

Anon, Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on C. Available at: http://globocan.iarc.fr/Pages/references.aspx [Accessed October 30, 2014].

Aschebrook-Kilfoy, B. et al., 2012. Common genetic variants in metabolism and detoxification pathways and the risk of papillary thyroid cancer. *Endocrine-related cancer*, 19(3), pp.333–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3394851&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Austen, M., Lüscher, B. & Lüscher-Firzlaff, J.M., 1997. Characterization of the transcriptional regulator YY1. The bipartite transactivation domain is independent of interaction with the TATA box-binding protein, transcription factor IIB, TAFII55, or cAMP-responsive element-binding protein (CPB)-binding protei. *The Journal of biological chemistry*, 272(3), pp.1709–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8999850 [Accessed October 20, 2015].

Avbelj, M. et al., 2007. High prevalence of thyroid peroxidase gene mutations in patients with thyroid dyshormonogenesis. *European journal of endocrinology / European Federation of Endocrine Societies*, 156(5), pp.511–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17468186 [Accessed October 12, 2016].

Balhara, Y.P.S. & Deb, K.S., 2013. Impact of alcohol use on thyroid function. *Indian journal of endocrinology and metabolism*, 17(4), pp.580–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3743356&tool=pmcentrez&rendertype=abstract [Accessed April 10, 2016].

Bastos, H.N. et al., 2009. Association of polymorphisms in genes of the homologous recombination DNA repair pathway and thyroid cancer risk. *Thyroid : official journal of the American Thyroid Association*, 19(10), pp.1067–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19772428 [Accessed October 7, 2016].

Belin, R.M. et al., 2004. Smoke exposure is associated with a lower prevalence of serum thyroid autoantibodies and thyrotropin concentration elevation and a higher prevalence of mild thyrotropin concentration suppression in the third National Health and Nutrition Examination Surve. *The Journal of clinical endocrinology and metabolism*, 89(12), pp.6077–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15579761 [Accessed September 10, 2015].

Bignell, G.R. et al., 1997. Familial nontoxic multinodular thyroid goiter locus maps to

chromosome 14q but does not account for familial nonmedullary thyroid cancer. *American journal of human genetics*, 61(5), pp.1123–30. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1716029&tool=pmcentrez&rendertype=abstract [Accessed August 5, 2015].

Binder-Foucard, F. et al., 2014. Cancer incidence and mortality in France over the 1980–2012 period: Solid tumors. *Revue d'Épidémiologie et de Santé Publique*, 62(2), pp.95–108.

Bonora, E. et al., 2013. The FOXE1 locus is a major genetic determinant for familial nonmedullary thyroid carcinoma. *International journal of cancer. Journal international du cancer*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24127282 [Accessed December 24, 2013].

Boyle, A.P. et al., 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*, 22(9), pp.1790–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3431494&tool=pmcentrez&rendertype=abstract [Accessed December 30, 2014].

Brenner, A. V et al., 2013. Common single nucleotide polymorphisms in genes related to immune function and risk of papillary thyroid cancer. *PloS one*, 8(3), p.e57243. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3592848&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Brindel, P. et al., 2009. Anthropometric factors in differentiated thyroid cancer in French Polynesia: a case-control study. *Cancer causes & control : CCC*, 20(5), pp.581–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19043789 [Accessed October 31, 2016].

Brindel, P. et al., 2008. Menstrual and reproductive factors in the risk of differentiated thyroid carcinoma in native women in French Polynesia: a population-based case-control study. *American journal of epidemiology*, 167(2), pp.219–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17965111 [Accessed October 31, 2016].

Bufalo, N.E. et al., 2006. Smoking and susceptibility to thyroid cancer: an inverse association with CYP1A1 allelic variants. *Endocrine-related cancer*, 13(4), pp.1185–93. Available at: http://erc.endocrinology-journals.org/content/13/4/1185.full [Accessed July 24, 2015].

Butler, M.W. et al., 2011. Glutathione S-transferase copy number variation alters lung gene expression. *European Respiratory Journal*, 38(1), pp.15–28.

Bůžková, P., Lumley, T. & Rice, K., 2011. Permutation and Parametric Bootstrap Tests for Gene-Gene and Gene-Environment Interactions. *Annals of Human Genetics*, 75(1), pp.36–45. Available at: http://doi.wiley.com/10.1111/j.1469-1809.2010.00572.x [Accessed December 21, 2016].

Caini, S. et al., 2015. Menstrual and reproductive history and use of exogenous sex hormones and risk of thyroid cancer among women: a meta-analysis of prospective studies. *Cancer Causes & Control*, 26(4), pp.511–518. Available at: http://link.springer.com/10.1007/s10552-015-0546-z [Accessed October 31, 2016].

Canbay, E. et al., 2003. Higher glutathione transferase GSTM1 0/0 genotype frequency in young thyroid carcinoma patients. *Current medical research and opinion*, 19(2), pp.102–106. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12740153 [Accessed July 1,

2014].

Cancemi, L. 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. *International Journal of Cancer*, 129(12), pp.2816–2824.

Canzian, F. et al., 1998. A gene predisposing to familial thyroid tumors with cell oxyphilia maps to chromosome 19p13.2. *American journal of human genetics*, 63(6), pp.1743–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1377646&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2015].

Cao, Y. et al., 2015. Reproductive Factors but Not Hormonal Factors Associated with Thyroid Cancer Risk: A Systematic Review and Meta-Analysis. *BioMed research international*, 2015, p.103515. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4538312&tool=pmcentrez&rendertype=abstract [Accessed September 8, 2015].

Cavaco, B.M. et al., 2008. Mapping a new familial thyroid epithelial neoplasia susceptibility locus to chromosome 8p23.1-p22 by high-density single-nucleotide polymorphism genome-wide linkage analysis. *The Journal of clinical endocrinology and metabolism*, 93(11), pp.4426–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18765515 [Accessed August 5, 2015].

Chen, X. et al., 2007. Learning position weight matrices from sequence and expression data. *Computational systems bioinformatics / Life Sciences Society. Computational Systems Bioinformatics Conference*, 6, pp.249–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17951829 [Accessed August 23, 2016].

Chiang, F.-Y. et al., 2008. Association between polymorphisms in DNA base excision repair genes XRCC1, APE1, and ADPRT and differentiated thyroid carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14(18), pp.5919–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18779313 [Accessed October 7, 2016].

Chiefari, E. et al., 2002. Increased expression of AP2 and Sp1 transcription factors in human thyroid tumors: a role in NIS expression regulation? *BMC cancer*, 2, p.35. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=139985&tool=pmcentrez&rendertype=abstract [Accessed May 16, 2015].

Cho, Y.A. & Kim, J., 2014. Thyroid cancer risk and smoking status: a meta-analysis. *Cancer causes & control : CCC*, 25(9), pp.1187–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24981099 [Accessed July 29, 2015].

Cipollini, M. et al., 2013. TPO genetic variants and risk of differentiated thyroid carcinoma in two European populations. *International Journal of Cancer*, 133(12), pp.2843–2851.

Cléro, E. et al., 2010. Pooled analysis of two case-control studies in New Caledonia and French Polynesia of body mass index and differentiated thyroid cancer: the importance of body surface area. *Thyroid : official journal of the American Thyroid Association*, 20(11), pp.1285–93. Available at: http://online.liebertpub.com.gate2.inist.fr/doi/abs/10.1089/thy.2009.0456 [Accessed September 9, 2014].

Clinckspoor, I. et al., 2013. Vitamin D in thyroid tumorigenesis and development. *Progress in Histochemistry and Cytochemistry*, 48(2), pp.65–98. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0079633613000120 [Accessed December 29, 2016].

Colonna, M. et al., 2015. Recent trends in incidence, geographical distribution, and survival of papillary thyroid cancer in France. *Cancer epidemiology*, 39(4), pp.511–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26003877 [Accessed October 25, 2016].

Cruzan, G. et al., 2013. Studies of styrene, styrene oxide and 4-hydroxystyrene toxicity in CYP2F2 knockout and CYP2F1 humanized mice support lack of human relevance for mouse lung tumors. *Regulatory Toxicology and Pharmacology*, 66(1), pp.24–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23454103 [Accessed January 21, 2017].

Czene, K., Lichtenstein, P. & Hemminki, K., 2002. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *International journal of cancer. Journal international du cancer*, 99(2), pp.260–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11979442 [Accessed October 30, 2014].

Derwahl, M. & Nicula, D., 2014. Estrogen and its role in thyroid cancer. *Endocrine-related cancer*, 21(5), pp.T273-83. Available at: http://erc.endocrinology-journals.org/content/21/5/T273.full [Accessed September 8, 2015].

Dou, R. et al., 2016. Vitamin D and colorectal cancer: molecular, epidemiological and clinical evidence. *British Journal of Nutrition*, 115(9), pp.1643–1660. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27245104 [Accessed December 29, 2016].

Fan, Z. et al., 2016. Association between the CYP11 family and six cancer types. *Oncology Letters*, 12(1), pp.35–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27347096 [Accessed January 10, 2017].

Fesinmeyer, M.D. et al., 2013. Genetic risk factors for BMI and obesity in an ethnically diverse population: results from the population architecture using genomics and epidemiology (PAGE) study. *Obesity (Silver Spring, Md.)*, 21(4), pp.835–46. Available at: http://doi.wiley.com/10.1002/oby.20268 [Accessed January 24, 2017].

Figlioli, G. et al., 2016. A comprehensive meta-analysis of case-control association studies to evaluate polymorphisms associated with the risk of differentiated thyroid carcinoma. *Cancer Epidemiology Biomarkers & Prevention*, 25(April), pp.1–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26843521 [Accessed October 12, 2016].

Figlioli, G. et al., 2015. Novel genetic variants in differentiated thyroid cancer and assessment of the cumulative risk. *Scientific reports*, 5, p.8922. Available at: http://www.nature.com/articles/srep08922 [Accessed December 2, 2016].

Figlioli, G. et al., 2014. Novel genome-wide association study-based candidate loci for differentiated thyroid cancer risk. *The Journal of clinical endocrinology and metabolism*, (September), p.jc20141734. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25029422 [Accessed September 8, 2014].

Franceschi, S. et al., 1999. A pooled analysis of case-control studies of thyroid cancer. IV. Benign thyroid diseases. *Cancer causes & control*, 10(6), pp.583–595. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/10616827 [Accessed November 3, 2016].

Gaspar, J. et al., 2004. Combined effects of glutathione S-transferase polymorphisms and thyroid cancer risk. *Cancer genetics and cytogenetics*, 151(1), pp.60–7. Available at: http://www.sciencedirect.com/science/article/pii/S0165460803004321 [Accessed July 4, 2015].

Geisler, S.A., 2001. GSTM1, GSTT1, and the Risk of Squamous Cell Carcinoma of the Head and Neck: A Mini-HuGE Review. *American Journal of Epidemiology*, 154(2), pp.95–105. Available at: http://aje.oxfordjournals.org.gate2.inist.fr/content/154/2/95.long [Accessed August 25, 2015].

Gemignani, F. et al., 2007. Development of lung cancer before the age of 50: The role of xenobiotic metabolizing genes. *Carcinogenesis*, 28(6), pp.1287–1293. Available at: http://carcin.oxfordjournals.org/cgi/doi/10.1093/carcin/bgm021 [Accessed January 19, 2017].

Goldgar, D.E. et al., 1994. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *Journal of the National Cancer Institute*, 86(21), pp.1600–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7932824 [Accessed July 1, 2014].

Guan, Y. & Stephens, M., 2008. Practical issues in imputation-based association mapping. N. J. Schork, ed. *PLoS genetics*, 4(12), p.e1000279. Available at: http://dx.plos.org/10.1371/journal.pgen.1000279 [Accessed June 3, 2014].

Gudmundsson, J. et al., 2009. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nature genetics*, 41(4), pp.460–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3664837&tool=pmcentrez&rendertype=abstract [Accessed December 23, 2013].

Gudmundsson, J. et al., 2012. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nature genetics*, 44(3), pp.319–22. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3655412&tool=pmcentrez&rendertype=abstract [Accessed December 19, 2013].

Guignard, R. et al., 2007. Alcohol drinking, tobacco smoking, and anthropometric characteristics as risk factors for thyroid cancer: a countrywide case-control study in New Caledonia. *American journal of epidemiology*, 166(10), pp.1140–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2220030&tool=pmcentrez&rendertype=abstract [Accessed June 3, 2015].

Guilhen, A.C.T. et al., 2009. Role of the N-acetyltransferase 2 detoxification system in thyroid cancer susceptibility. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 15(1), pp.406–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19118072 [Accessed August 12, 2015].

Hayes, J.D. & Strange, R.C., 2000. Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology*, 61(3), pp.154–66. Available at: http://www.karger.com/Article/FullText/28396 [Accessed April 8, 2016].

He, C. et al., 2010. A large-scale candidate gene association study of age at menarche and age at natural menopause. *Human Genetics*, 128(5), pp.515–527. Available at: http://link.springer.com/10.1007/s00439-010-0878-4 [Accessed January 13, 2017].

He, H. et al., 2009. A Susceptibility Locus for Papillary Thyroid Carcinoma on Chromosome 8q24. *Cancer research*, 69(2), pp.625–631. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19147577 [Accessed February 6, 2014].

He, H., Li, W., Liyanarachchi, S., Jendrzejewski, J., et al., 2015. Genetic predisposition to papillary thyroid carcinoma: involvement of FOXE1, TSHR, and a novel lincRNA gene, PTCSC2. *The Journal of clinical endocrinology and metabolism*, 100(1), pp.E164-72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25303483 [Accessed October 21, 2014].

He, H., Li, W., Liyanarachchi, S., Srinivas, M., et al., 2015. Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 112(19), pp.6128–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25918370 [Accessed May 2, 2015].

Hemminki, K., Eng, C. & Chen, B., 2005. Familial risks for nonmedullary thyroid cancer. *The Journal of clinical endocrinology and metabolism*, 90(10), pp.5747–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16030170 [Accessed November 14, 2014].

Hemminki, K. & Vaittinen, P., 1999. Familial cancers in a nationwide family cancer database: age distribution and prevalence. *European Journal of Cancer*, 35(7), pp.1109–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10533456 [Accessed July 1, 2014].

Henderson, B.E. & Feigelson, H.S., 2000. Hormonal carcinogenesis. *Carcinogenesis*, 21(3), pp.427–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10688862 [Accessed December 29, 2016].

Hermann, D., Heinz, A. & Mann, K., 2002. Dysregulation of the hypothalamic-pituitary-thyroid axis in alcoholism. *Addiction*, 97(11), pp.1369–1381. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12410778 [Accessed March 9, 2016].

Hernández, A. et al., 2003. Glutathione S-transferase polymorphisms in thyroid cancer patients. *Cancer Letters*, 190(1), pp.37–44. Available at: http://www.sciencedirect.com/science/article/pii/S0304383502005803 [Accessed August 12, 2015].

Hirvonen, A., 1999. Polymorphisms of xenobiotic-metabolizing enzymes and susceptibility to cancer. *Environmental health perspectives*, 107 Suppl, pp.37–47. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1566364&tool=pmcentrez&rendertype=abstract [Accessed August 25, 2015].

Ho, T. et al., 2009. Association of XRCC1 polymorphisms and risk of differentiated thyroid carcinoma: a case-control analysis. *Thyroid : official journal of the American Thyroid Association*, 19(2), pp.129–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19191745 [Accessed June 17, 2015].

Ho, T. et al., 2006. Glutathione S-transferase polymorphisms and risk of differentiated thyroid carcinomas: a case-control analysis. *Archives of otolaryngology--head & neck surgery*, 132(7), pp.756–61. Available at: http://archotol.jamanetwork.com.gate2.inist.fr/article.aspx?articleid=484212 [Accessed August 12, 2015].

Hong, S.-H., Myung, S.-K. & Kim, H., 2016. Alcohol Intake and Risk of Thyroid Cancer: A Meta-

analysis of Observational Studies. *Cancer Research and Treatment*. Available at: http://e-crt.org/journal/view.php?doi=10.4143/crt.2016.161 [Accessed November 2, 2016].

Horn-Ross, P.L. et al., 2011. Hormonal factors and the risk of papillary thyroid cancer in the California Teachers Study cohort. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 20(8), pp.1751–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3288117&tool=pmcentrez&rendertype=abstract [Accessed March 15, 2016].

Horn-Ross, P.L. et al., 2001. Iodine and thyroid cancer risk among women in a multiethnic population: the Bay Area Thyroid Cancer Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 10(9), pp.979–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11535551 [Accessed November 3, 2016].

Howie, B., Marchini, J. & Stephens, M., 2011. Genotype imputation with thousands of genomes. *G3 (Bethesda, Md.)*, 1(6), pp.457–70. Available at: http://www.g3journal.org/content/1/6/457.full [Accessed May 27, 2014].

Howie, B.N., Donnelly, P. & Marchini, J., 2009. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies N. J. Schork, ed. *PLoS Genetics*, 5(6), p.e1000529. Available at: http://dx.plos.org/10.1371/journal.pgen.1000529 [Accessed September 21, 2016].

Huang, R.S. et al., 2009. Population-specific GSTM1 copy number variation. *Human molecular genetics*, 18(2), pp.366–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2722188&tool=pmcentrez&rendertype=abstract [Accessed August 21, 2015].

Ionita-Laza, I. et al., 2013. Sequence kernel association tests for the combined effect of rare and common variants. *American journal of human genetics*, 92(6), pp.841–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23684009 [Accessed November 21, 2016].

Jendrzejewski, J. et al., 2012. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22), pp.8646–51. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3365219&tool=pmcentrez&rendertype=abstract [Accessed February 6, 2014].

Jones, A.M. et al., 2012. Thyroid cancer susceptibility polymorphisms: confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24. *Journal of medical genetics*, 49(3), pp.158–63. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3286794&tool=pmcentrez&rendertype=abstract [Accessed January 27, 2014].

Kabat, G.C. et al., 2012. Smoking and alcohol consumption in relation to risk of thyroid cancer in postmenopausal women. *Cancer epidemiology*, 36(4), pp.335–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22525339 [Accessed July 29, 2015].

Kallel, R. et al., 2010. Genetic investigation of FOXE1 polyalanine tract in thyroid diseases:

New insight on the role of FOXE1 in thyroid carcinoma. *Cancer Biomarkers*, 8(1), pp.43–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21896990 [Accessed October 11, 2016].

Kilfoy, B.A. et al., 2009. International patterns and trends in thyroid cancer incidence, 1973-2002. *Cancer Causes and Control*, 20(5), pp.525–531. Available at: http://link.springer.com/10.1007/s10552-008-9260-4 [Accessed October 28, 2016].

Kitahara, C.M. et al., 2014. Childhood height and body mass index were associated with risk of adult thyroid cancer in a large cohort study. *Cancer research*, 74(1), pp.235–42. Available at: http://cancerres.aacrjournals.org/content/74/1/235.full [Accessed May 6, 2016].

Kitahara, C.M., Linet, M.S., et al., 2012. Cigarette smoking, alcohol intake, and thyroid cancer risk: a pooled analysis of five prospective studies in the United States. *Cancer causes & control : CCC*, 23(10), pp.1615–24. Available at: /pmc/articles/PMC3511822/?report=abstract [Accessed March 10, 2015].

Kitahara, C.M., Neta, G., et al., 2012. Common obesity-related genetic variants and papillary thyroid cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 21(12), pp.2268–71. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3518668&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Kitahara, C.M. et al., 2011. Obesity and thyroid cancer risk among U.S. men and women: a pooled analysis of five prospective studies. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 20(3), pp.464–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3079276&tool=pmcentrez&rendertype=abstract [Accessed May 6, 2016].

Kitahara, C.M. & Sosa, J.A., 2016. The changing incidence of thyroid cancer. *Nature reviews. Endocrinology*, 12(11), pp.646–653. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27418023 [Accessed October 25, 2016].

Klimkiewicz, A. et al., 2016. COMT and BDNF Gene Variants Help to Predict Alcohol Consumption in Alcohol-dependent Patients. *Journal of Addiction Medicine*, p.1. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27898499 [Accessed January 2, 2017].

Klug, J. & Beato, M., 1996. Binding of YY1 to a site overlapping a weak TATA box is essential for transcription from the uteroglobin promoter in endometrial cells. *Molecular and cellular biology*, 16(11), pp.6398–407. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=231641&tool=pmcentrez&rendertype=abstract [Accessed October 20, 2015].

Köhler, A. et al., 2013. Genome-wide association study on differentiated thyroid cancer. *The Journal of clinical endocrinology and metabolism*, 98(10), pp.E1674-81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23894154 [Accessed December 19, 2013].

Kolonel, L.N. et al., 1990. An epidemiologic study of thyroid cancer in Hawaii. *Cancer causes & control : CCC*, 1(3), pp.223–34. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/2102295 [Accessed January 3, 2017].

Kreiger, N. & Parkes, R., 2000. Cigarette smoking and the risk of thyroid cancer. *European Journal of Cancer*, 36(15), pp.1969–1973. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11000579 [Accessed December 13, 2016].

Kristensen, V.N. & Børresen-Dale, A.L., 2000. Molecular epidemiology of breast cancer: genetic variation in steroid hormone metabolism. *Mutation Research/Reviews in Mutation Research*, 462(2–3), pp.323–333.

Kweon, S.-S. et al., 2014. Polymorphisms of methylenetetrahydrofolate reductase and glutathione S-transferase are not associated with the risk of papillary thyroid cancer in Korean population. *Molecular biology reports*, 41(6), pp.3793–9. Available at: http://link.springer.com/10.1007/s11033-014-3245-z [Accessed August 26, 2015].

Landa, I. et al., 2009. The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS genetics*, 5(9), p.e1000637. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2727793&tool=pmcentrez&rendertype=abstract [Accessed December 23, 2013].

Landa, I. & Robledo, M., 2011. Association studies in thyroid cancer susceptibility: are we on the right track? *Journal of molecular endocrinology*, 47(1).

Lao, X. et al., 2014. Glutathione S-transferase gene GSTM1, gene-gene interaction, and gastric cancer susceptibility: evidence from an updated meta-analysis. *Cancer cell international*, 14(1), p.127. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4255933&tool=pmcentrez&rendertype=abstract [Accessed July 23, 2015].

Leclerc, J. et al., 2011. Xenobiotic metabolism and disposition in human lung: Transcript profiling in non-tumoral and tumoral tissues. *Biochimie*, 93(6), pp.1012–1027.

Lemos, M.C. et al., 2008. Combined GSTM1 and GSTT1 null genotypes are associated with a lower risk of papillary thyroid cancer. *Journal of endocrinological investigation*, 31(6), pp.542–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18591888 [Accessed November 6, 2014].

Lemos, M.C. et al., 2007. Genetic polymorphism of CYP2D6 influences susceptibility to papillary thyroid cancer. *Clinical Endocrinology*, 67(2), pp.180–183. Available at: http://doi.wiley.com/10.1111/j.1365-2265.2007.02858.x [Accessed January 16, 2017].

Lesueur, F. et al., 2002. Specific haplotypes of the RET proto-oncogene are over-represented in patients with sporadic papillary thyroid carcinoma. *Journal of medical genetics*, 39(4), pp.260–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1735081&tool=pmcentrez&rendertype=abstract [Accessed December 23, 2013].

Leux, C. & Guénel, P., 2010. Risk factors of thyroid tumors: role of environmental and occupational exposures to chemical pollutants. *Revue d'épidémiologie et de santé publique*, 58(5), pp.359–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20980113 [Accessed September 9, 2014].

Li, J. et al., 2012. Glutathione S-transferase M1, T1, and P1 polymorphisms and thyroid

cancer risk: a meta-analysis. *Cancer epidemiology*, 36(6), pp.e333-40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22765906 [Accessed November 6, 2014].

Li, M.-X. et al., 2012. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human genetics*, 131(5), pp.747–56. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3325408&tool=pmcentrez&rendertype=abstract [Accessed September 11, 2014].

Li, Y. et al., 2009. Genotype imputation. *Annual review of genomics and human genetics*, 10, pp.387–406. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2925172&tool=pmcentrez&rendertype=abstract [Accessed May 5, 2014].

Liyanarachchi, S. et al., 2013. Cumulative risk impact of five genetic variants associated with papillary thyroid carcinoma. *Thyroid : official journal of the American Thyroid Association*, 23(12), pp.1532–40. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3868253&tool=pmcentrez&rendertype=abstract [Accessed October 22, 2014].

Lönn, S. et al., 2007. Papillary thyroid cancer and polymorphic variants in TSHR- and RET-related genes: a nested case-control study within a cohort of U.S. radiologic technologists. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 16(1), pp.174–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17220349 [Accessed February 4, 2015].

Low, Y.L. et al., 2010. Multi-variant pathway association analysis reveals the importance of genetic determinants of estrogen metabolism in breast and endometrial cancer susceptibility K. Kerr, ed. *PLoS Genetics*, 6(7), pp.1–8. Available at: http://dx.plos.org/10.1371/journal.pgen.1001012 [Accessed January 9, 2017].

Loza, M.J. et al., 2007. Assembly of Inflammation-Related Genes for Pathway-Focused Genetic Analysis J. Gold, ed. *PLoS ONE*, 2(10), p.e1035. Available at: http://dx.plos.org/10.1371/journal.pone.0001035 [Accessed January 13, 2017].

Ma, J. et al., 2015. Obesity and risk of thyroid cancer: evidence from a meta-analysis of 21 observational studies. *Medical science monitor : international medical journal of experimental and clinical research*, 21, pp.283–91. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4315628&tool=pmcentrez&rendertype=abstract [Accessed July 29, 2015].

Maillard, S. et al., 2015. Common Variants at 9q22.33, 14q13.3, and ATM Loci, and Risk of Differentiated Thyroid Cancer in the French Polynesian Population. *PloS one*, 10(4), p.e0123700. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4388539&tool=pmcentrez&rendertype=abstract [Accessed April 10, 2015].

Malchoff, C.D. et al., 2000. Papillary thyroid carcinoma associated with papillary renal neoplasia: genetic linkage analysis of a distinct heritable tumor syndrome. *The Journal of clinical endocrinology and metabolism*, 85(5), pp.1758–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10843148 [Accessed August 5, 2015].

Mancikova, V. et al., 2015. Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations. *International journal of cancer. Journal international du cancer*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25855579 [Accessed June 11, 2015].

Männistö, P.T. & Kaakkola, S., 1999. Catechol-O-methyltransferase (COMT): Biochemistry, Molecular Biology, Pharmacology, and Clinical Efficacy of the New Selective COMT Inhibitors. *Pharmacological Reviews*, 51(4).

Manolio, T.A. et al., 2009. Finding the missing heritability of complex diseases. *Nature*, 461(7265). Available at: http://www.nature.com/doifinder/10.1038/nature08494 [Accessed January 7, 2017].

Mansoori, A.A. & Jain, S.K., 2015. Molecular Links between Alcohol and Tobacco Induced DNA Damage, Gene Polymorphisms and Patho-physiological Consequences: A Systematic Review of Hepatic Carcinogenesis. *Asian Pacific journal of cancer prevention : APJCP*, 16(12), pp.4803–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26163595 [Accessed September 1, 2015].

Marcello, M.A. et al., 2014. Obesity and thyroid cancer. *Endocrine-related cancer*, 21(5), pp.T255-71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24741026 [Accessed September 13, 2015].

Marchini, J. & Howie, B., 2010. Genotype imputation for genome-wide association studies. *Nature reviews. Genetics*, 11(7), pp.499–511. Available at: http://dx.doi.org/10.1038/nrg2796 [Accessed May 25, 2014].

McKay, J.D. et al., 2001. Localization of a susceptibility gene for familial nonmedullary thyroid carcinoma to chromosome 2q21. *American journal of human genetics*, 69(2), pp.440–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1235316&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2015].

McLellan, R.A. et al., 1997. Characterization of a human glutathione S-transferase mu cluster containing a duplicated GSTM1 gene that causes ultrarapid enzyme activity. *Molecular pharmacology*, 52(6), pp.958–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9415705 [Accessed September 7, 2015].

Meinhold, C.L. et al., 2009. Alcohol intake and risk of thyroid cancer in the NIH-AARP Diet and Health Study. *British journal of cancer*, 101(9), pp.1630–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2778506&tool=pmcentrez&rendertype=abstract [Accessed July 29, 2015].

Meinhold, C.L. et al., 2010. Nonradiation risk factors for thyroid cancer in the US Radiologic Technologists Study. *American journal of epidemiology*, 171(2), pp.242–52. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3290908&tool=pmcentrez&rendertype=abstract [Accessed March 15, 2016].

Memon, A. et al., 2002. Epidemiology of reproductive and hormonal factors in thyroid cancer: evidence from a case-control study in the Middle East. *International journal of cancer*, 97(1), pp.82–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11774247 [Accessed January 12, 2017].

Memon, A., Varghese, A. & Suresh, A., 2002. Benign thyroid disease and dietary factors in

thyroid cancer: a case–control study in Kuwait. *British Journal of Cancer*, 86(11), pp.1745–1750. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12087461 [Accessed November 3, 2016].

Mitrunen, K. & Hirvonen, A., 2003. Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. *Mutation research*, 544(1), pp.9–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12888106 [Accessed September 7, 2015].

Miyauchi, E. et al., 2016. Quantitative Atlas of Cytochrome P450, UDP-Glucuronosyltransferase, and Transporter Proteins in Jejunum of Morbidly Obese Subjects. *Molecular Pharmaceutics*, 13(8), pp.2631–2640. Available at: http://pubs.acs.org/doi/abs/10.1021/acs.molpharmaceut.6b00085 [Accessed December 30, 2016].

Mooney, M.A. & Wilmot, B., 2015. Gene set analysis: A step-by-step guide. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 168(7), pp.517–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26059482 [Accessed October 23, 2015].

Moore, M. a et al., 2010. Cancer epidemiology in the pacific islands - past, present and future. *Asian Pacific journal of cancer prevention : APJCP*, 11 Suppl 2, pp.99–106. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20553071.

Morari, E.C. et al., 2002. The null genotype of glutathione s-transferase M1 and T1 locus increases the risk for thyroid cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 11(11), pp.1485–8. Available at: http://cebp.aacrjournals.org/content/11/11/1485.full [Accessed August 12, 2015].

Morey, L. & Helin, K., 2010. Polycomb group protein-mediated repression of transcription. *Trends in biochemical sciences*, 35(6), pp.323–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20346678 [Accessed March 11, 2015].

Moriya, H. et al., 2014. Single-nucleotide polymorphisms and copy number variations of the FCGR2A and FCGR3A genes in healthy Japanese subjects. *Biomedical reports*, 2(2), pp.265–269. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3917743&tool=pmcentrez&rendertype=abstract [Accessed November 7, 2014].

Mota, P. et al., 2015. Genetic polymorphisms of phase I and phase II metabolic enzymes as modulators of lung cancer susceptibility. *Journal of cancer research and clinical oncology*, 141(5), pp.851–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25388590 [Accessed September 10, 2015].

Nagano, J. et al., 2007. A case-control study in Hiroshima and Nagasaki examining non-radiation risk factors for thyroid cancer. *Journal of epidemiology*, 17(3), pp.76–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17545694 [Accessed December 13, 2016].

Negri, E. et al., 1999. A pooled analysis of case-control studies of thyroid cancer. II. Menstrual and reproductive factors. *Cancer causes & control : CCC*, 10(2), pp.143–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10231163 [Accessed May 6, 2016].

Neta, G. et al., 2011. Common genetic variants related to genomic integrity and risk of papillary thyroid cancer. *Access*, 32(8), pp.1231–1237. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3166197&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Neuhouser, M.L., Kristal, A.R. & Penson, D.F., 2004. Steroid hormones and hormone-related genetic and lifestyle characteristics as risk factors for benign prostatic hyperplasia: Review of epidemiologic literature. *Urology*, 64(2), pp.201–211. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15302461 [Accessed December 29, 2016].

Nørskov, M.S. et al., 2011. Copy number variation in glutathione-S-transferase T1 and M1 predicts incidence and 5-year survival from prostate and bladder cancer, and incidence of corpus uteri cancer in the general population. *The pharmacogenomics journal*, 11(4), pp.292–9. Available at: http://dx.doi.org/10.1038/tpj.2010.38 [Accessed March 3, 2016].

Nørskov, M.S. et al., 2009. High-throughput genotyping of copy number variation in glutathione S-transferases M1 and T1 using real-time PCR in 20,687 individuals. *Clinical biochemistry*, 42(3), pp.201–9. Available at: http://www.sciencedirect.com/science/article/pii/S0009912008005614 [Accessed March 4, 2016].

Omdahl, J.L., Morris, H.A. & May, B.K., 2002. H YDROXYLASE E NZYMES OF THE V ITAMIN D P ATHWAY : Expression, Function, and Regulation. *Annual Review of Nutrition*, 22(1), pp.139–166. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12055341 [Accessed December 29, 2016].

Onda, M. et al., 2004. Up-regulation of transcriptional factor E2F1 in papillary and anaplastic thyroid cancers. *Journal of human genetics*, 49(6), pp.312–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15118916 [Accessed May 8, 2015].

Park, S.-K. et al., 2003. Reproductive Factors, Glutathione S-Transferase M1 and T1 Genetic Polymorphism and Breast Cancer Risk. *Breast Cancer Research and Treatment*, 78(1), pp.89–96. Available at: http://link.springer.com.gate2.inist.fr/article/10.1023/A%25253A1022113713226 [Accessed April 26, 2016].

Pazaitou-Panayiotou, K., Polyzos, S.A. & Mantzoros, C.S., 2013. Obesity and thyroid cancer: epidemiologic associations and underlying mechanisms. *Obesity reviews : an official journal of the International Association for the Study of Obesity*, 14(12), pp.1006–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24034423 [Accessed March 30, 2016].

Penna-Martinez, M. et al., 2012. Impaired vitamin D activation and association with CYP24A1 haplotypes in differentiated thyroid carcinoma. *Thyroid : official journal of the American Thyroid Association*, 22(7), pp.709–16. Available at: http://online.liebertpub.com/doi/abs/10.1089/thy.2011.0330 [Accessed December 29, 2016].

Penning, T.M. et al., 2000. Human 3alpha-hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *The Biochemical journal*, 351(Pt 1), pp.67–77. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/10998348 [Accessed January 10, 2017].

Penning, T.M. & Byrns, M.C., 2009. Steroid hormone transforming aldo-keto reductases and cancer. *Annals of the New York Academy of Sciences*, 1155, pp.33–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19250190 [Accessed January 10, 2017].

Pereda, C.M. et al., 2015. Common variants at the 9q22.33, 14q13.3 and ATM loci, and risk of differentiated thyroid cancer in the Cuban population. *BMC Genetics*, 16. Available at: http://www.biomedcentral.com/1471-2156/16/22.

Pilon, C. et al., 1999. Mutations in CYP11B1 gene converting 11??-hydroxylase into an aldosterone-producing enzyme are not present in aldosterone-producing adenomas. *Journal of Clinical Endocrinology and Metabolism*, 84(11), pp.4228–4231.

Prentice, R.L. et al., 1982. Radiation exposure and thyroid cancer incidence among Hiroshima and Nagasaki residents. *National Cancer Institute monograph*, 62, pp.207–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7167191 [Accessed October 30, 2016].

Rebaï, M. et al., 2009. Association of polymorphisms in estrogen and thyroid hormone receptors with thyroid cancer risk. *Journal of receptor and signal transduction research*, 29(2), pp.113–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19519176 [Accessed August 10, 2015].

Rebbeck, T.R., 1997. Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 6(9), pp.733–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9298582 [Accessed September 8, 2015].

Reis, A.A.S. et al., 2010. Involvement of CYP1A1, GST, 72TP53 polymorphisms in the pathogenesis of thyroid nodules. *Genetics and molecular research : GMR*, 9(4), pp.2222–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21086258 [Accessed August 12, 2015].

Ren, Y., Lence-Anta, J.J., et al., 2016. *Foxe1* Polymorphism Interacts with Dietary Iodine Intake in Differentiated Thyroid Cancer Risk in the Cuban Population. *Thyroid*, p.thy.2015.0594. Available at: http://online.liebertpub.com/doi/10.1089/thy.2015.0594 [Accessed November 2, 2016].

Ren, Y., Lence-Anta, J.J., et al., 2016. Foxe1 Polymorphism Interacts with Dietary Iodine Intake in Differentiated Thyroid Cancer Risk in the Cuban Population. *Thyroid : official journal of the American Thyroid Association*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27610545 [Accessed October 3, 2016].

Rinaldi, S. et al., 2012. Body size and risk of differentiated thyroid carcinomas: Findings from the EPIC study. *International Journal of Cancer*, 131(6), pp.E1004–E1014. Available at: http://doi.wiley.com/10.1002/ijc.27601 [Accessed October 31, 2016].

Ron, E. et al., 1995. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiation research*, 141(3), pp.259–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7871153 [Accessed October 7, 2016].

Rose-Zerilli, M.J. et al., 2009. Copy-number variation genotyping of GSTT1 and GSTM1 gene deletions by real-time PCR. *Clinical chemistry*, 55(9), pp.1680–5. Available at:

http://www.clinchem.org/content/55/9/1680.full [Accessed March 24, 2015].

Rossing, M.A. et al., 2000. Risk of papillary thyroid cancer in women in relation to smoking and alcohol consumption. *Epidemiology (Cambridge, Mass.)*, 11(1), pp.49–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10615843 [Accessed December 13, 2016].

Rotinen, M. et al., 2009. Estradiol induces type 8 17beta-hydroxysteroid dehydrogenase expression: crosstalk between estrogen receptor alpha and C/EBPbeta. *The Journal of endocrinology*, 200(1), pp.85–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18852215 [Accessed January 8, 2017].

Rudolph, A. et al., 2012. Copy number variations of GSTT1 and GSTM1, colorectal cancer risk and possible effect modification of cigarette smoking and menopausal hormone therapy. *International journal of cancer. Journal international du cancer*, 131(5), pp.E841-8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22234881 [Accessed November 7, 2014].

Saitou, M. & Ishida, T., 2015. Distributions of the GSTM1 and GSTT1 null genotypes worldwide are characterized by latitudinal clines. *Asian Pacific journal of cancer prevention : APJCP*, 16(1), pp.355–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25640380 [Accessed August 26, 2015].

Schlattl, A. et al., 2011. Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. *Genome research*, 21(12), pp.2004–13. Available at: http://genome.cshlp.org/cgi/doi/10.1101/gr.122614.111 [Accessed December 16, 2016].

Schonfeld, S.J. et al., 2012. Common genetic variants in sex hormone pathway genes and papillary thyroid cancer risk. *Thyroid : official journal of the American Thyroid Association*, 22(2), pp.151–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3271376&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Schonfeld, S.J. et al., 2010. Polymorphisms in oxidative stress and inflammation pathway genes, low-dose ionizing radiation, and the risk of breast cancer among US radiologic technologists. *Cancer causes & control : CCC*, 21(11), pp.1857–66. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3076104&tool=pmcentrez&rendertype=abstract [Accessed February 4, 2015].

Secretan, B.L. et al., 2016. Spe ci a l R e p or t Body Fatness and Cancer — Viewpoint of the IARC Working Group.

Sen, A. et al., 2015. Baseline and lifetime alcohol consumption and risk of differentiated thyroid carcinoma in the EPIC study. *British journal of cancer*, 113(5), pp.840–7. Available at: http://www.nature.com.gate2.inist.fr/bjc/journal/vaop/ncurrent/full/bjc2015280a.html [Accessed September 1, 2015].

Shi, J. & Lee, S., 2016. A novel random effect model for GWAS meta-analysis and its application to trans-ethnic meta-analysis. , 72(3), pp.945–954.

Sigurdson, A.J. et al., 2003. Cancer incidence in the US radiologic technologists health study, 1983-1998. *Cancer*, 97(12), pp.3080–9. Available at:

http://doi.wiley.com/10.1002/cncr.11444 [Accessed February 4, 2015].

Sigurdson, A.J. et al., 2009. Thyroid nodules, polymorphic variants in DNA repair and RET-related genes, and interaction with ionizing radiation exposure from nuclear tests in Kazakhstan. *Radiation research*, 171(1), pp.77–88. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2875679&tool=pmcentrez&rendertype=abstract [Accessed July 1, 2014].

Simon, J.A. & Lange, C.A., 2008. Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 647(1), pp.21–29.

Singer, S. et al., 2012. Quality of Life in Patients with Thyroid Cancer Compared with the General Population. *Thyroid*, 22(2), pp.117–124. Available at: http://online.liebertpub.com/doi/abs/10.1089/thy.2011.0139 [Accessed October 26, 2016].

Siraj, A.K. et al., 2008. Polymorphisms of selected xenobiotic genes contribute to the development of papillary thyroid cancer susceptibility in Middle Eastern population. *BMC medical genetics*, 9(1), p.61. Available at: http://www.biomedcentral.com/1471-2350/9/61 [Accessed August 12, 2015].

Soldin, O.P. et al., 2009. Thyroid hormone levels associated with active and passive cigarette smoking. *Thyroid : official journal of the American Thyroid Association*, 19(8), pp.817–23. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3643222&tool=pmcentrez&rendertype=abstract [Accessed September 10, 2015].

Sprenger, R. et al., 2000. Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype-phenotype correlation. *Pharmacogenetics*, 10(6), pp.557–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10975610 [Accessed September 7, 2015].

Stankov, K. et al., 2006. GSTT1 and M1 polymorphisms in Hürthle thyroid cancer patients. *Cancer letters*, 240(1), pp.76–82. Available at: http://www.sciencedirect.com/science/article/pii/S0304383505007998 [Accessed August 13, 2015].

Stewart, S.H. et al., 2009. COMT Genotype Influences the Effect of Alcohol on Blood Pressure: Results From the COMBINE Study. *American Journal of Hypertension*, 22(1), pp.87–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19023276 [Accessed January 22, 2017].

Sturgis, E.M. et al., 2005. Radiation response genotype and risk of differentiated thyroid cancer: a case-control analysis. *The Laryngoscope*, 115(6), pp.938–945.

Suzuki, T. et al., 2008. Anthropometric factors at age 20 years and risk of thyroid cancer. *Cancer causes & control : CCC*, 19(10), pp.1233–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18618280 [Accessed October 31, 2016].

Takahashi, M. et al., 2010. The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Human molecular genetics*, 19(12), pp.2516–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20350937 [Accessed December 23, 2013].

Tian, C., Gregersen, P.K. & Seldin, M.F., 2008. Accounting for ancestry: Population substructure and genome-wide association studies. *Human Molecular Genetics*, 17(R2), pp.143–150. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18852203 [Accessed December 21, 2016].

Tournel, G. et al., 2007. CYP2F1 genetic polymorphism: Identification of interethnic variations. *Xenobiotica*, 37(12), pp.1433–1438. Available at: http://www.tandfonline.com/doi/full/10.1080/00498250701644403 [Accessed January 21, 2017].

Tournel, G. et al., 2007. Molecular analysis of the CYP2F1 gene: Identification of a frequent non-functional allelic variant. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 617(1–2), pp.79–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17327131 [Accessed January 22, 2017].

Truong, T. et al., 2010. Role of dietary iodine and cruciferous vegetables in thyroid cancer: a countrywide case-control study in New Caledonia. *Cancer causes & control : CCC*, 21(8), pp.1183–92. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3496161&tool=pmcentrez&rendertype=abstract [Accessed April 26, 2015].

Truong, T. et al., 2005. Role of goiter and of menstrual and reproductive factors in thyroid cancer: a population-based case-control study in New Caledonia (South Pacific), a very high incidence area. *American journal of epidemiology*, 161(11), pp.1056–65. Available at: http://aje.oxfordjournals.org.gate2.inist.fr/content/161/11/1056.long [Accessed September 9, 2014].

Truong, T. et al., 2007. Time trends and geographic variations for thyroid cancer in New Caledonia, a very high incidence area (1985-1999). *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)*, 16(1), pp.62–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17220706 [Accessed July 10, 2014].

Vaccarella, S. et al., 2015. The Impact of Diagnostic Changes on the Rise in Thyroid Cancer Incidence: A Population-Based Study in Selected High-Resource Countries. *Thyroid : official journal of the American Thyroid Association*, 25(10), pp.1127–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26133012 [Accessed October 25, 2016].

Valeix, P. et al., 2008. Effects of light to moderate alcohol consumption on thyroid volume and thyroid function. *Clinical endocrinology*, 68(6), pp.988–95. Available at: http://onlinelibrary.wiley.com.gate2.inist.fr/doi/10.1111/j.1365-2265.2007.03123.x/full [Accessed March 9, 2016].

Vigneri, R., Malandrino, P. & Vigneri, P., 2014. The changing epidemiology of thyroid cancer: why is incidence increasing? *Current opinion in oncology*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25310641 [Accessed October 24, 2014].

Di Vito, M. et al., 2011. Overexpression of estrogen receptor-α in human papillary thyroid carcinomas studied by laser- capture microdissection and molecular biology. *Cancer science*, 102(10), pp.1921–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21707866 [Accessed March 2, 2015].

Wang, K., Li, M. & Bucan, M., 2007. Pathway-based approaches for analysis of genomewide

association studies. *American journal of human genetics*, 81(6), pp.1278–83. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0002929707637756 [Accessed December 27, 2016].

Wang, Y.-L. et al., 2013. Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population. *Journal of medical genetics*, 50(10), pp.689–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23847140 [Accessed July 15, 2014].

Ward, L.D. & Kellis, M., 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research*, 40(Database issue), pp.D930-4. Available at: http://nar.oxfordjournals.org/content/early/2011/11/07/nar.gkr917.long [Accessed July 10, 2014].

Weber, F. & Eng, C., 2005. Editorial: Germline Variants within RET: Clinical Utility or Scientific Playtoy? *The Journal of Clinical Endocrinology & Metabolism*, 90(11), pp.6334–6336. Available at: http://press.endocrine.org/doi/10.1210/jc.2005-2030 [Accessed October 12, 2016].

Williams, D., 2002. Cancer after nuclear fallout: lessons from the Chernobyl accident. *Nature reviews. Cancer*, 2(7), pp.543–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12094241 [Accessed October 7, 2016].

Williams, J.A. & Phillips, D.H., 2000. Mammary Expression of Xenobiotic Metabolizing Enzymes and Their Potential Role in Breast Cancer. *Cancer Research*, 60(17).

Wiltshire, J.J. et al., 2016. Systematic Review of Trends in the Incidence Rates of Thyroid Cancer. *Thyroid*, p.thy.2016.0100. Available at: http://online.liebertpub.com/doi/10.1089/thy.2016.0100 [Accessed October 27, 2016].

Wu, L. & Zhu, J., 2015. Linear reduction in thyroid cancer risk by oral contraceptive use: a dose-response meta-analysis of prospective cohort studies. *Human reproduction (Oxford, England)*, 30(9), pp.2234–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26141711 [Accessed November 2, 2016].

Xhaard, C. et al., 2015. Anthropometric Risk Factors for Differentiated Thyroid Cancer in Young Men and Women From Eastern France: A Case-Control Study. *American journal of epidemiology*, 182(3), pp.202–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26133374 [Accessed October 31, 2016].

Xhaard, C. et al., 2014. Menstrual and reproductive factors in the risk of differentiated thyroid carcinoma in young women in France: a population-based case-control study. *American journal of epidemiology*, 180(10), pp.1007–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25269571 [Accessed February 24, 2015].

Xu, S. et al., 2013. Oestrogen action on thyroid progenitor cells: relevant for the pathogenesis of thyroid nodules? *The Journal of endocrinology*, 218(1), pp.125–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23645248 [Accessed August 10, 2015].

Yu, K. et al., 2009. Pathway analysis by adaptive combination of P-values. *Genetic epidemiology*, 33(8), pp.700–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2790032&tool=pmcentrez

&rendertype=abstract [Accessed June 18, 2015].

Zablotska, L.B. et al., 2008. A cohort study of thyroid cancer and other thyroid diseases after the Chornobyl accident: dose-response analysis of thyroid follicular adenomas detected during first screening in Ukraine (1998-2000). *American journal of epidemiology*, 167(3), pp.305–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17989057 [Accessed October 30, 2016].

Zai, G. et al., 2017. A review of molecular genetic studies of neurocognitive deficits in schizophrenia. *Neuroscience & Biobehavioral Reviews*, 72, pp.50–67.

Zamora-Ros, R. et al., 2015. Reproductive and menstrual factors and risk of differentiated thyroid carcinoma: the EPIC study. *International journal of cancer*, 136(5), pp.1218–27. Available at: http://onlinelibrary.wiley.com.gate2.inist.fr/doi/10.1002/ijc.29067/full [Accessed September 8, 2015].

Zaykin, D.V. et al., 2002. Truncated product method for combiningP-values. *Genetic Epidemiology*, 22(2), pp.170–185. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11788962 [Accessed January 7, 2017].

Zhang, W. et al., 2014. Meta-analysis in the association between obesity and risk of thyroid cancer. *International journal of clinical and experimental medicine*, 7(12), pp.5268–74. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4307477&tool=pmcentrez&rendertype=abstract [Accessed September 10, 2015].

Zheng, T. et al., 2003. Glutathione S-transferase M1 and T1 genetic polymorphisms, alcohol consumption and breast cancer risk. *British journal of cancer*, 88(1), pp.58–62. Available at: http://www.nature.com.gate2.inist.fr/bjc/journal/v88/n1/full/6600708a.html [Accessed March 24, 2015].

Zhou, Q. et al., 2016. Long noncoding RNA PVT1 modulates thyroid cancer cell proliferation by recruiting EZH2 and regulating thyroid-stimulating hormone receptor (TSHR). *Tumor Biology*, 37(3), pp.3105–3113.

Zhu, J. et al., 2016. Parity and thyroid cancer risk: a meta-analysis of epidemiological studies. *Cancer medicine*, 5(4), pp.739–52. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4831293&tool=pmcentrez&rendertype=abstract [Accessed May 3, 2016].

Zimmermann, M.B. & Galetti, V., 2015. Iodine intake as a risk factor for thyroid cancer: a comprehensive review of animal and human studies. *Thyroid research*, 8, p.8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26146517 [Accessed November 2, 2016].

Table articles GSTs

Table 1: Selected Characteristics of the Study Sample and Association Between a Selection of Lifestyle, Hormonal and Reproductive Factors and DTC Risk in Europeans and Melanesians.

	Europeans				Melanesians			
variable	**Cases 504(%)**	**Controls 620(%)**	**OR**	**(95% CI)**	**Cases 156(%)**	**Controls 114(%)**	**OR**	**(95% CI)**
Mean age (years)	51.3	51			45.6	46		
Sex								
Women	402 (79.76)	463 (74.68)			148 (94.87)	105 (92.11)		
Men	102 (20.24)	157 (25.32)			8 (5.13)	9 (7.89)		
Study population								
CATHY	477 (94.64)	539 (86.94)						
New Caledonia	27 (5.36)	81 (13.06)			156 (100)	114 (100)		
Area of residence in CATHY study								
Calvados	151 (29.96)	192 (30.97)						
Marne	217 (43.06)	250 (40.32)						
Ardennes	109 (21.63)	97 (15.65)						
Area of residence in NC study								
North	3 (11.11)	6 (7.41)			35 (22.44)	29 (25.44)		
South	23 (85.19)	74 (91.36)			54 (34.62)	46 (40.35)		
Island	1 (3.70)	1 (1.23)			67 (42.95)	39 (34.21)		
Histological type of cancer								
Papillary	444 (88.10)				132 (84.62)			
Follicular	60 (11.90)				24 (15.38)			
Both genders								
BMI (kg/m^2) [a]								
<25	237 (47.02)	323 (52.1)	1.00		38 (24.36)	38 (33.33)	1.00	
25-29.9	162 (32.14)	188 (30.32)	1.37	1.01, 1.87	52 (33.33)	41 (35.96)	1.28	0.65, 2.50
≥30	104 (20.63)	105 (16.94)	1.33	0.93, 1.91	63 (40.38)	29 (25.44)	2.70	1.34, 5.44
Missing	1 (0.20)	4 (0.65)			3 (1.92)	6 (5.26)		
Alcohol drinking (glass/day) [a]								
Never	141 (27.98)	130 (20.97)	1.00		69 (44.23)	45 (39.47)	1.00	
<1	249 (49.4)	333 (53.71)	0.65	0.47, 0.89	66 (42.31)	53 (46.49)	0.68	0.37, 1.26
≥1	111 (22.02)	152 (24.52)	0.64	0.42, 0.96	19 (12.18)	14 (12.28)	1.08	0.42, 2.75
Missing	3 (0.60)	5 (0.81)			2 (1.28)	2 (1.75)		
Smoking status [b]								
Never	269 (53.37)	322 (51.94)	1.00		95 (60.9)	70 (61.4)	1.00	
Former	147 (29.17)	159 (25.65)	1.09	0.81, 1.47	25 (16.03)	15 (13.16)	0.86	0.38, 1.95
Current	88 (17.46)	139 (22.42)	0.80	0.57, 1.12	36 (23.08)	29 (25.44)	1.07	0.55, 2.06
Missing	0 (0.00)	0 (0.00)			0 (0.00)	0 (0.00)		
Cigarette smoking (pack-years) [a]								
Never	269 (53.37)	322 (51.94)	1.00		95 (60.9)	70 (61.4)	1.00	
<20	165 (32.74)	191 (30.81)	1.14	0.84, 1.55	42 (26.92)	26 (22.81)	1.15	0.59, 2.25
≥20	68 (13.49)	99 (15.97)	0.82	0.54, 1.26	19 (12.18)	17 (14.91)	0.82	0.36, 1.85
Missing	2 (0.40)	8 (1.29)			0 (0.00)	1 (0.88)		
Women								
Age at menarche (years) [c]								
≤13	329 (81.84)	398 (85.96)	1.00		63 (42.57)	56 (53.33)	1.00	
>13	71 (17.66)	63 (13.61)	1.47	0.99, 2.17	78 (52.7)	43 (40.95)	1.31	0.72, 2.41
Missing	2 (0.50)	2 (0.43)			7 (4.73)	6 (5.71)		
Oral contraception use [c]								
Ever	288 (71.64)	338 (73)	0.72	0.49, 1.06	38 (25.68)	27(25.71)	1.27	0.62, 2.61
Never	109 (27.11)	122 (26.35)	1.00		110 (74.32)	77 (73.33)	1.00	
Missing	5 (1.24)	3 (0.65)			0 (0.00)	1 (0.95)		
Number of full term pregnancy [c]								
0	36 (8.96)	53 (11.45)	1.00		18 (12.16)	13 (12.38)	1.00	
1-2	225 (55.97)	251 (54.21)	1.20	0.72, 1.99	28 (18.92)	29 (27.62)	0.52	0.17, 1.59
≥3	141 (35.07)	159 (34.34)	1.23	0.72, 2.12	95 (64.19)	60 (57.14)	0.67	0.23, 1.91
Missing	0 (0.00)	0 (0.00)			7 (4.73)	3 (2.86)		
Menopausal status [c]								
Pre-menopause	140(41.90)	170(50.40)	1.00		84 (63.10)	58 (64.40)	1.00	
Post-menopause	194(58.10)	167(49.60)	1.30	0.94, 1.80	49 (36.80)	167 (35.60)	1.04	0.34, 3.63

ORs are estimated from multivariate analysis based on a total of 498 cases /603 controls in Europeans and 105 cases /144 controls in Melanesians. In analyses restricted to women, 391 cases /448 controls in Europeans and 98 cases /142 controls in Melanesians were included. All ORs are adjusted for age, sex and area of residence.
[a] The multivariate regression model includes BMI, alcohol consumption in glass per day and cigarette smoking in pack-years.
[b] The multivariate regression model include BMI, alcohol consumption in glass per day and smoking status.
[c] The multivariate regression model include BMI, alcohol consumption in glass per day, cigarette smoking in pack-years, age at menarche, number of full term pregnancy and oral contraception.
Ca/Co: Cases/Controls, OR: Odd ratios, 95% IC: 95% confident Interval, *p*: p-value of interaction test

Table 2: Distribution of *GSTM1* and *GSTT1* Genotypes and Association With DTC Risk in Europeans and Melanesians.

Variables	Europeans				Melanesians			
	Cases 504(%)	Controls 620(%)	OR	(95% CI)	Cases 156(%)	Controls 114(%)	OR	(95% CI)
***GSTM1* number of copies**								
0	248 (52.65)	298 (51.20)	0.92	0.60, 1.41	124 (80)	94 (82.46)	0.80	0.21, 3.08
1	172 (36.52)	231 (39.69)	0.80	0.52, 1.25	25 (16.13)	16 (14.04)	0.82	0.19, 3.54
≥2	51 (10.83)	53 (9.11)	1.00		6 (3.87)	4 (3.51)	1.00	
Missing	33	38			1	0		
***GSTM1* genotype**								
Null	248 (52.65)	298 (51.2)	1.09	0.85, 1.40	124 (80)	94 (82.46)	0.94	0.49, 1.82
Non-null	223 (47.35)	284 (48.8)	1.00		31 (20)	20 (17.54)	1.00	
Missing	33	38			1	0		
***GSTT1* number of copies**								
0	90 (18.99)	116 (20.53)	0.81	0.57, 1.15	53 (40.46)	39 (39.39)	0.80	0.36, 1.75
1	196 (41.35)	254 (44.96)	0.80	0.61, 1.06	48 (36.64)	44 (44.44)	0.61	0.28, 1.33
≥2	188 (39.66)	195 (34.51)	1.00		30 (22.9)	16 (16.16)	1.00	
Missing	30	55			25	15		
***GSTT1* genotype**								
Null	90 (18.99)	116 (20.53)	0.92	0.67, 1.26	53 (40.46)	39 (39.39)	1.12	0.64, 1.97
Non-null	384 (81.01)	449 (79.47)	1.00		78 (59.54)	60 (60.61)	1.00	
Missing	30	55			25	15		
Combination of *GSTM1* and *GSTT1*								
***GSTM1/GSTT1* number of copies**								
0 copy for *GSTM1* and *GSTT1*	41 (9.30)	53 (10.06)	0.97	0.60, 1.55	42 (32.31)	36 (36.36)	0.97	0.38, 2.50
At least 1 copy for *GSTM1* or *GSTT1*	229 (51.93)	274 (51.99)	1.01	0.76, 1.31	72 (55.38)	52 (52.53)	1.18	0.48, 2.91
At least 1 copy for *GSTM1* and *GSTT1*	171 (38.78)	200 (37.95)	1.00		16 (12.31)	11 (11.11)	1.00	
Missing	63	93			26	15		
***GSTM1/GSTT1* genotype**								
Null/null	41 (9.30)	53 (10.06)	0.97	0.60, 1.55	42 (32.31)	36 (36.36)	0.97	0.38, 2.50
Null/non null	189 (42.86)	222 (42.13)	1.02	0.77, 1.36	62 (47.69)	49 (49.49)	1.06	0.43, 2.63
Non-null/null	40 (9.07)	52 (9.87)	0.90	0.56, 1.44	10 (7.69)	3 (3.03)	3.44	0.69, 17.19
Non-null/non-null	171 (38.78)	200 (37.95)	1.00		16 (12.31)	11 (11.11)	1.00	
Missing	63	93			26	15		

All ORs are adjusted for age, sex and area of residence. Ca/Co: Cases/Controls. OR: Odd ratios, 95% IC: 95% confident Interval

Table 3: Association Between DTC Risk and BMI, Alcohol Consumption and Tobacco Smoking Stratified by Genotypes of *GSTM1* and *GSTT1* in Europeans and Melanesians.

	Europeans							Melanesians						
	Non-null			Null				Non-null			Null			
Variable	Ca/Co	OR	95%IC	Ca/C	O	95% IC	*p*	Ca/C	OR	95% IC	Ca/C	OR	95% IC	*p*
GSTM1														
BMI (kg/m²) [a]							0.							0.
<25	105/157	1.0		120/	1.			10/6	1.00	1.00	27/3	1.0		
25-29.9	72/74	1.4	0.97,	75/9	1.	0.68,		6/6	0.26	0.03,	45/3	1.6	0.79,	
≥30	44/52	1.1	0.69,	51/4	1.	0.86,		14/7	1.39	0.25,	48/2	3.2	1.47,	
Alcohol drinking [a]							0.							0.
Never	63/70			75/5				13/8	1.00	1.00	55/3	1.0	1.00	
Ever	158/213	0.8	0.55,	171/	0.	0.32,		17/1	0.85	0.17,	65/5	0.7	0.37,	
Alcohol drinking							0.							
Never	63/70	1.0		75/5	1.									
<1	117/141	0.8	0.58,	119/	0.	0.3, 0.74								
≥1	41/72	0.6	0.38,	52/6	0.	0.34,								
Smoking status [b]							0.							0.
Never	116/154	1.0		140/	1.			73/5	1.00	1.00	17/1	1.0		
Ever	105/129	0.7	0.42,	106/	0.	0.49,		47/3	1.19	0.64,	13/7	0.7	0.14,	
Cigarette smoking (pack-							0.							
Never	116/154	1.0		140/	1.									
<20	81/86	1.3	0.89,	74/9	0.	0.61,								
≥20	24/43	0.8	0.45,	32/4	0.	0.48,								
GSTT1														
BMI (kg/m²)[a]							0.							0.
<25	201/253	1.0		37/6	1.			19/1	1.00	1.00	14/1	1.0		
25-29.9	124/138	1.1	0.83,	35/3	1.	0.9, 3 .87		29/2	0.64	0.24,	13/1	2.1	0.65,	
≥30	86/84	1.2	0.87,	18/1	2.	0.86,		28/1	1.30	0.48,	23/7	7.6	2.1,	
Alcohol drinking [a]														
Never	114/106	1.0		25/2	1.			30/1			26/1	1.0	1.00	
Ever	297/369	0.7	0.57,	65/9	0.	0.29,		46/3	0.81	0.36,	24/2	0.3	0.13,	
Alcohol drinking							0.							
Never	114/106	1.0		25/2	1.									
<1	200/260	0.7	0.53,	51/5	0.	0.35,								
≥1	97/109	0.9	0.63,	14/3	0.	0.12,								
Smoking status [b]							0.							0.
Never	220/256	1.0		48/5	1.	1.00		45/3	1.00	1.00	29/2	1.0		
Ever	191/219	0.8	0.59,	42/5	0.	0.34,		31/1	1.30	0.58,	21/1	1.0	0.39,	
Cigarette smoking (pack-							0.							
Never	220/256	1.0		48/5	1.									
<20	136/145	1.1	0.84,	29/3	0.	0.45,								
≥20	55/74	0.8	0.57,	13/2	1.	0.38,								
GSTM1 an *GSTT1*[c]														
BMI (kg/m²)														
≤25	190/252	1.0		15/3	1.			23/1	1.00	1.00	10/1	1.0		
25-29.9	125/125	1.3	0.97,	15/1	1.	0.54,		31/2	0.61	0.24,	10/1	2.4	0.64,	
≥30	79/84	1.2	0.82,	11/5	4.	0.98,		32/1	1.26	0.49,	19/6	13.	2.79,	0.
Alcohol drinking [a]														
Never	119/102	1.0		12/6	1.			33/2	1.00	1.00	23/1	1.0		
Ever	275/359	0.6	0.5,	29/4	0.	0.08,		53/3	0.80	0.37,	16/2	0.2	0.07,	
Alcohol drinking							0.							
Never	119/102	1.0		12/6	1.									
<1	198/254	0.6	0.48,	22/3	0.	0.09,								
≥1	77/105	0.7	0.48,	7/15	0.	0.03,								
Smoking status [b]							0.							0.
Never	215/253	1.0		24/2	1.			48/3	1.00	1.00	25/1	1.0		
Ever	179/208	0.7	0.49,	17/2	0.	0.21, 4.1		38/2	1.45	0.68,	14/1	1.1	0.35,	

Cigarette smoking									
Never	215/253	1.0		24/2	1.				
<20	137/139	1.2	0.91,	10/1	0.	0.18, 1.8			
≥20	42/69	0.7	0.49,	7/8	1.	0.29,			

All ORs are adjusted for age, sex and area of residence.

[a] The multivariate regression model includes BMI, alcohol consumption in glass per day and cigarette smoking in pack-years.

[b] The multivariate regression model includes BMI, alcohol consumption in glass per day and smoking status.

[c] Null genotype refers to individuals with null genotypes for both genes, while non-null genotype refers to subjects with at least one copy of *GSTM1* or *GSTT1*

Ca/Co: Cases/Controls, OR: Odd ratios, 95% IC: 95% confident Interval, *p*: p-value of interaction test

Table 4: Association Between DTC Risk and Hormonal and Reproductive Factors stratified by Genotypes of *GSTM1* and *GSTT1* in European and Melanesian Women.

	Europeans							Melanesians						
	Non-null			Null				Non-null			Null			
Variable	**Ca/Co**	**OR**	**95%IC**	**Ca/Co**	**OR**	**95%IC**	***P***	**Ca/Co**	**OR**	**95%IC**	**Ca/Co**	**OR**	**95%IC**	***P***
GSTM1														
Age at menarche [a] (years)							0.81							0.71
≤13	148/186	1.00		168/196	1.00			11/12	1.00		50/41	1.00		
>13	35/33	1.53	0.88, 2.65	32/28	1.41	0.79, 2.52		16/6	3.16	0.55, 18.08	59/34	1.39	0.74, 2.63	
Oral contraception[a]							0.22							0.86
Never	50/47	1.00		55/69	1.00			21/14	1.00		79/54	1.00		
Ever	133/172	0.55	0.31, 0.97	145/155	1.01	0.59, 1.74		6/4	0.84	0.10, 7.42	30/21	1.16	0.54, 2.52	
Parity[a]							0.36							0.90
Nulliparous	16/28	1.00		18/22	1.00			2/1	1.00		14/9	1.00		
Parous	167/191	1.51	0.75, 3.05	182/202	0.87	0.42, 1.83		25/17	0.92	0.03, 26.45	95/66	0.79	0.29, 2.19	
Menopausal status[a]							0.84							0.79
Pre-menopause	72/81	1.00		65/84	1.00			13/9	1.00		17/11	1.00		
Post-menopause	99/83	1.12	0.23, 5.35	91/81	3.43	1.36, 8.66		92/64	0.75	0.25, 2.23	10/6	7E, 10	0, inf	
GSTT1														
Age at menarche[a] (year)							0.02							0.73
≤13	262/311	1.00		59/63	1.00			29/26	1.00		25/19	1.00		
>13	59/38	1.79	1.14, 2.82	9/20	0.57	0.21, 1.54		39/21	1.54	0.67, 3.54	24/15	1.47	0.50, 4.36	
Oral contraception[a]							0.42							0.87
Never	88/84	1.00		20/30	1.00			49/33	1.00		36/25	1.00		
Ever	233/265	0.73	0.48, 1.13	48/53	0.8	0.28, 2.33		19/14	1.28	0.47, 3.52	13/9	0.79	0.21, 2.99	
Parity[a]							0.02							0.67
Nulliparous	33/33	1.00		3/13	1.00			5/2	1.00		8/6			
Parous	288/316	0.86	0.49, 1.51	65/70	5.71	1.29, 25.2		63/45	0.44	0.06, 3.04	41/28	0.84	0.20, 3.63	
Menopausal status[a]							0.89							0.50
Pre-menopause	23/29	1.00		109/130	1.00			32/21	1.00		44/30	1.00		
Post-menopause	35/32	1.19	0.25, 5.62	156/125	1.86	0.95, 3.64		14/13	1.24	0.18, 8.84	24/15	0.55	0.09, 3.24	
***GSTM1* and *GSTT1*[b]**														
Age at menarche[a](years)							0.14							0.58
≤13	268/311	1.00		27/34	1.00			31/27	1.00		23/18	1.00		
>13	61/44	1.66	1.07, 2.58	4/10	0.47	0.10, 2.23		46/23	1.65	0.75, 3.63	16/13	1.28	0.39, 4.27	
Oral contraception[a]							0.92							0.92
Ever	240/266	0.77	0.50, 1.18	20/27	0.56	0.12, 2.72		22/14	1.18	0.45, 3.08	10/9	0.91	0.20, 4.1	
Never	89/89	1.00		11/17	1.00			55/36	1.00		29/22			
Parity[a]														0.65
Nulliparous	32/37	1.00		0/7				5/2	1.00		7/6	1.00		
Parous	297/318	1.01	0.59, 1.74	31/37				72/48	0.52	0.08, 3.4	32/25	0.71	0.13, 3.97	

Menopausal status[a]							0.40							0.49
Pre-menopause	11/12	1.00		118/141	1.00			23/21	1.00		52/30	1.00		
Post-menopause	15/19	0.43	0.03, 5.44	172/135	1.90	1.00, 3.63		13/10	3.3	0.34, 31.81	25/18	0.43	0.07, 2.46	

. All ORs are adjusted for age, sex, area of residence

[a] The multivariate regression model include BMI, alcohol consumption in glass per day, cigarette smoking in pack-years, age at menarche, number of full term pregnancy and oral contraception.

[b] Null genotype refers to individuals with null genotypes for both genes, while non-null genotype refers to subjects with at least one copy of *GSTM1* or *GSTT1*

Ca/Co: Cases/Controls, OR: Odd ratios, 95% IC: 95% confident Interval, *p*: p-value of interaction test