



Étude des facteurs de risque génétiques des cancers différenciés de la thyroïde dans une population multi-ethnique : cartographie fine des régions 2q35 et 8p12 et caractérisation des régions d'homozygotie associées à ces cancers

Julie Guibon

► **To cite this version:**

Julie Guibon. Étude des facteurs de risque génétiques des cancers différenciés de la thyroïde dans une population multi-ethnique : cartographie fine des régions 2q35 et 8p12 et caractérisation des régions d'homozygotie associées à ces cancers. Cancer. Université Paris-Saclay, 2021. Français. NNT : 2021UPASR027 . tel-03741829

HAL Id: tel-03741829

<https://theses.hal.science/tel-03741829>

Submitted on 2 Aug 2022

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.

Étude des facteurs de risque génétiques des
cancers différenciés de la thyroïde dans une
population multi-ethnique : cartographie fine
des régions 2q35 et 8p12 et caractérisation des
régions d'homozygotie associées à ces cancers

*Genetic risk factors of differentiated thyroid cancer in a multi-ethnic
population: fine-mapping of loci 2q35 and 8p12, and characterization of
runs of homozygosity associated with this cancer*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 570, santé publique (EDSP)
Spécialité de doctorat : santé publique - épidémiologie
Graduate School : Santé Publique. Référent : Faculté de Médecine

Thèse préparée dans les unités de recherche **U900, Institut Curie (Université PSL, Mines Paris Tech, Inserm)** et **CESP (université Paris-Saclay, UVSQ, Inserm)**, sous la direction de **Fabienne LESUEUR**, Chargée de recherche, INSERM, université PSL, et le co-encadrement de **Thérèse TRUONG**, Chargée de recherche, INSERM, université Paris-Saclay

Thèse soutenue à Paris-Saclay, le 7 décembre 2021, par

Julie GUIBON

Composition du Jury

Philippe BROËT Professeur, université Paris-Saclay	Président
Nabila BOUATIA-NAJI Directrice de Recherche, Inserm, Université de Paris	Rapporteur & Examinatrice
Federico CANZIAN Directeur de Recherche, Centre allemand de recherche sur le cancer (DKFZ), Heidelberg	Rapporteur & Examineur
Stefano LANDI Professeur, Université de Pise	Examineur
Fabienne LESUEUR Chargée de Recherche, Inserm, université PSL	Directrice de thèse
Thérèse TRUONG Chargée de Recherche, Inserm, université Paris-Saclay	Co-encadrante de thèse

UNITES D'ACCUEIL

- Centre de recherche en épidémiologie et santé des populations (CESP), INSERM U1018, équipe « Exposome, hérédité, cancer et santé », Hôpital Paul Brousse, 16 avenue Paul-Vaillant Couturier, Villejuif, France.
- INSERM, U900, équipe « Epidémiologie génétique des cancers », Institut Curie, Paris, France.

FINANCEMENTS

Cette thèse a été financée par une bourse doctorale ARDoc allouée par la Région Ile-de-France et par une allocation doctorale de la Ligue contre le Cancer (4e année).

Remerciements

Durant ces quatre années de thèse j'ai eu la chance d'être entourée par deux équipes de recherche qui m'ont permise de m'épanouir tant sur le plan professionnel que personnel et d'y rencontrer de belles personnes qui, pour certaines, occupent désormais une place à part entière dans ma vie.

Tout d'abord je tiens à remercier très chaleureusement mes directrices de thèse Fabienne et Thérèse de m'avoir permise de réaliser cette thèse. Merci pour votre confiance, votre complémentarité dans l'encadrement scientifique apporté et votre rigueur qui pousse à donner le meilleur de soi. Merci pour la compréhension, la patience ainsi que la bienveillance dont vous avez su faire part dans les périodes professionnelles ou personnelles difficiles que j'ai traversées. Je vous remercie pour cet enrichissement tant humain que scientifique que vous m'avez apporté durant ces quatre années.

Je remercie tous les membres du jury d'avoir accepté d'y prendre part pour évaluer mon travail de thèse et d'être venus de plus ou moins loin pour en discuter le jour de ma soutenance malgré ces temps très particuliers de crise sanitaire. Merci à mes deux rapporteurs Nabila Bouati-Naji et Federico Canzian pour le temps qu'ils ont dédié à la relecture de mon manuscrit et pour leurs commentaires très enrichissants. Merci à Philippe Broët de me faire l'honneur de présider le jury et d'occuper la place d'examineur. Et merci à Stefano Landi d'avoir accepté d'apporter son expertise en tant qu'examineur.

Je remercie la Région Ile-de-France ainsi que la Ligue contre le Cancer pour avoir financé ma thèse ainsi que mes formations et déplacements. Je tiens à remercier la directrice de l'Ecole Doctorale de Santé Publique (EDSP), Florence Ménégaux, ainsi que Fabienne et Lucie pour leur disponibilité et leur aide dans la résolution de tous les petits tracas du doctorat. Un merci tout singulier à Sabine Barbut et Carine Benie qui ont su gérer une situation administrative très délicate et ont pu m'enlever une très grosse épine du pied en cette fin de thèse.

Merci aux membres du consortium EPITHYR pour m'avoir permise d'utiliser les données et particulièrement à Monia vers qui j'ai pu me tourner pour toutes mes questions « polynésiennes ».

Je tiens à remercier Anne-Louise et Hervé pour leur expertise ainsi que leurs précieux conseils au cours de ma thèse.

Merci aux membres, anciens et actuels, de l'équipe Exposome et Hérité du CESP du 3^e étage de Paul Brousse pour la bonne humeur et l'ambiance apportées ainsi que toutes ces discussions échangées autour d'un repas à la cantine ou d'un café. Un merci tout particulier à Pierre-Emmanuel, alias le chat noir des technologies, co-bureau parfait bien qu'un peu râleur parfois (mais je ne peux pas t'en vouloir après tout, c'était souvent des plaintes informatiques je l'avoue...). Merci à mes autres amies du CESP : Clémentine, Elise, Emilie, Soumaya pour votre soutien et vos conseils, pour nos discussions et nos fous rires partagés.

Je remercie l'équipe quasi-100% féminine du Boul'mich : Dorothée, Eve, Gaëlle, Juana, Juliette, Marie-Gabrielle, Noura, Sarah, Sèverine, Sophia et Yue et les deux garçons, qui à tour de rôle ont permis de maintenir l'unique présence masculine au sein de cette équipe, Nicolas et Om. Un merci tout particulier à Nadine de m'avoir accueillie dans cette équipe et pour ces shots de self-confiance qui boostent. Un merci particulier à Dorothée qui est devenue plus qu'une collègue cette dernière année. Merci à toi pour ces discussions de cœur autour d'un verre au Gay Lussac (avec des olives bien sûr !), ces soirées appart-Café Chéri-crêpes, et toutes celles à venir. ☺

Bien sûr, un merci tout particulier à une personne qui occupe une place à part entière dans ma vie depuis 11 longues années maintenant. Merci à toi Cloé, meilleure amie, coloc, compagne de VDM. Merci d'être présente depuis tout ce temps pour partager mes bonheurs de vie mais aussi de me soutenir lors des coups durs... Je pense notamment à cette période difficile commune vécue il y a deux ans, qui nous a permis aussi de mieux nous retrouver. Je te remercie pour ta bienveillance, ta générosité inépuisable, ta gentillesse, tes conseils, ton écoute si

précieuse mais aussi pour tes phrases discontinues (« minus...cule ! »), ton accent et tes mots si singuliers du Sud de la France (je mets mon véto sur l'utilisation de «patin-couffin »), nos planques de détectives (pas si privés...),... =D pour tous ces souvenirs inoubliables. Merci tout simplement pour l'Amitié que tu me portes !

Merci à la bande de copains Baptiste, Fanny, Robin, pour ces moments de décompression.

Je remercie toute la bande de l'ASFI. Merci à vous tous mes copains-athlètes de participer à cette belle atmosphère qui règne au sein de ce club. A toutes nos séances de souffrance autour de la piste qui se finissent bien souvent par des moments de franches rigolades. Un merci tout particulier à Antoine, Laure, Laurie, Léa et Samuel. Merci à toi Alain, qui occupe bien plus qu'une place de coach. Je te remercie pour ta bienveillance inégalable, tes conseils et ton soutien apporté tant sportivement qu'humainement. Et bien sûr, cette parenthèse ASFI ne serait rien sans elle ; merci à toi ma rayonnante Céline.

Je remercie ma famille, Papa, Maman, James ainsi que mon grand-frère David et ma petite sœur Alice. Merci d'avoir cru en moi toutes ces années et pour votre soutien qui m'a permis d'aller au bout de cette thèse. Merci pour l'équilibre que vous me permettez d'avoir au quotidien.

Et merci aussi à tout le reste de ma famille, oncles et tantes, cousins et cousines, même si je sais que pour vous mon doctorat est tellement abstrait que vous m'imaginez voguer entre l'école et la bibliothèque universitaire, le nez dans les bouquins à passer mes derniers partiels... « Alors alors, quelle note t'as eu à ta soutenance ? » 😊 .

Table des matières

Valorisations scientifiques	14
Liste des tableaux	17
Liste des figures	19
Liste des tableaux supplémentaires	22
Liste des figures supplémentaires	23
Liste des abréviations	25
Introduction générale.....	27
I. La glande thyroïdienne.....	27
II. Les cancers de la thyroïde	28
III. Epidémiologie descriptive des cancers de la thyroïde.....	29
IV. Les facteurs de risque non génétiques des cancers de la thyroïde	33
1. Exposition aux rayonnements ionisants	33
2. Les antécédents personnels de pathologie bénigne de la thyroïde	34
3. Les facteurs anthropométriques.....	34
4. Les facteurs hormonaux et reproductifs	35
i. Les facteurs menstruels	35
ii. Les facteurs reproductifs	35
iii. Les traitements hormonaux	36
5. La consommation d'alcool et de tabac	36
6. L'alimentation	37
7. Autres facteurs de risque	38
V. Les facteurs de risque génétiques des cancers de la thyroïde	38

1. Antécédents familiaux des CDT	38
i. Syndromes héréditaires	40
ii. Loci de prédisposition aux CDT familiaux	41
2. Les études cas-témoins en population générale	46
i. Approche « gène candidat »	46
3. Les études pangénomiques de type « GWAS »	49
4. L'identification de variants récessifs	50
Objectifs de la thèse	51
Etudes	53
I. Le consortium EPITHYR	53
1. L'étude CATHY	53
2. L'étude Nouvelle-Calédonie (NC)	54
3. L'étude YoungThyr	55
4. L'étude CUBA	56
5. L'Étude de Polynésie française	57
6. L'étude de Tchernobyl	58
7. La cohorte E3N	58
II. La cohorte EPIC	60
Données de génotypage	63
I. La puce OncoArray	63
II. Contrôle qualité des données de génotypage	64
III. Imputation des SNPs non génotypés	64
IV. Définition des groupes ethniques	68
Chapitre 1	73

Etude d'association génome entier dans la population multi-ethnique du consortium EPITHYR	73
I. Introduction	73
II. Données	75
III. Stratégie d'analyse	77
IV. Méthodes	78
1. Test d'association avec un modèle logistique mixte	78
2. QQ plot	80
3. Méta-analyse	80
4. Test d'hétérogénéité	81
5. Analyses stratifiées	81
V. Résultats	81
1. Analyses descriptives	81
2. Réplication des loci de prédisposition connus	82
3. Identification de nouvelles régions de prédisposition au CDT	86
4. Analyse de réplication des locus mis en évidence dans la méta-analyse de GWAS	90
VI. Discussion	90
Chapitre 2	93
Cartographie fine des régions 2q35 et 8p12 dans les populations européennes et océaniques d'EPITHYR et recherche de variants causaux	93
I. Introduction	93
II. Matériel et méthodes	98
1. La population d'étude : le consortium EPITHYR	98
i. Définition des groupes de population	98
ii. Description de la population	101

2.	Les données génétiques	102
i.	Imputation des génotypes manquants dans les deux régions.....	103
3.	Stratégie d'analyse	105
i.	Analyse conditionnelle.....	105
ii.	Identification des variants causaux	106
	Analyse de colocalisation avec PancanQTL	106
iii.	Annotation fonctionnelle des variants.....	107
III.	Résultats	107
1.	Locus 2q35	107
i.	Test d'association univarié et analyses conditionnelles.....	107
ii.	Analyses de colocalisation	112
iii.	Analyses stratifiées pour rs16857609	114
2.	Locus 8p12	115
i.	Test d'association univarié et analyses conditionnelles.....	115
ii.	Analyses de colocalisation	119
iii.	Analyses stratifiées pour rs2439304 et rs7844425.....	121
IV.	Discussion	123
	Chapitre 3	127
	Etude des régions homozygotes du génome et recherche de facteurs récessifs.....	127
I.	Introduction	127
II.	Matériel et méthodes	129
1.	Données	129
i.	Définition des groupes de population dans EPITHYR	131
ii.	Estimation du taux d'apparementement entre individus.....	135
iii.	Description des populations	136

2.	Stratégie d'analyse	140
i.	Identification des ROH dans chaque population.....	140
ii.	Méthode d'identification des ROH : algorithme et paramètres	141
iii.	Identification des ROH partagés entre individus	142
iv.	Estimation du coefficient de consanguinité des individus	143
v.	Analyses statistiques	144
III.	Résultats	144
1.	Hétérozygotie et longueur minimale des ROH dans les différents groupes de population étudiés	144
2.	Caractéristiques globales des ROH dans les différentes populations	145
3.	Association entre la longueur ou le nombre de ROH et le risque de CDT	147
4.	Consanguinité et risque de CDT	150
5.	Nombre de ROH partagés et risque de CDT	152
6.	Analyses de sensibilité	155
i.	Comparaison Nord/Sud des caractéristiques des ROH identifiés chez les Européens	155
ii.	Influence du type de prélèvement à partir duquel est extrait l'ADN sur la caractérisation des ROH	157
IV.	Discussion	158
	Conclusions et perspectives	163
	Références bibliographiques	167
	Annexes	199

Valorisations scientifiques

Articles issus du travail de thèse :

Julie Guibon, Pierre-Emmanuel Sugier, Mojgan Karimi, Om Kulkarni, Anne Boland-Auge, Delphine Bacq-Daian, Céline Besse, Elisabeth Adjadj, Frédérique Rachédi, Carole Rubino, Constance Xhaard, Pierre Laurent-Puig, Claire Mulot, Anne-Valérie Guizard, Rosa Maria Ortiz, Yan Ren, Claire Schvartz, Evgenia Ostroumova, Jean-François Deleuze, Marie-Christine Boutron-Ruault, Ausrele Kesminiene, Florent De Vathaire, Pascal Guénel, Fabienne Lesueur*, Thérèse Truong*. **Fine-mapping of two differentiated thyroid carcinoma susceptibility loci at 2q35 and 8p12 detects novel candidate functional SNPs in Europeans, Melanesians and Polynesians.** *Oncotarget*, 2021 Mar 2;12(5):493-506

Julie Guibon, Anne-Louise Leutenegger, Mojgan Karimi, Pierre-Emmanuel Sugier, Sabina Rinaldi, Laure Dossus, Anne Boland-Auge, Delphine Bacq-Daian, Céline Besse, Elisabeth Adjadj, Frédérique Rachédi, Carole Rubino, Constance Xhaard, Pierre Laurent-Puig, Claire Mulot, Anne-Valérie Guizard, Rosa Maria Ortiz, Yan Ren, Claire Schvartz, Evgenia Ostroumova, Jean-François Deleuze, Marie-Christine Boutron-Ruault, Ausrele Kesminiene, Florent De Vathaire, Pascal Guénel, Fabienne Lesueur*, Thérèse Truong*. **Runs of homozygosity and differentiated thyroid cancer risk.** *En préparation*

Thérèse Truong, Fabienne Lesueur, Pierre-Emmanuel Sugier, Julie Guibon, Constance Xhaard, Mojgan Karimi, Om Kulkarni, Elise A Lucotte, Delphine Bacq-Daian, Anne Boland-Auge, Claire Mulot, Pierre Laurent-Puig, Claire Schvartz, Anne-Valérie Guizard, Yan Ren, Elisabeth Adjadj, Frédérique Rachédi, Françoise Borson-Chazot, Rosa Maria Ortiz, Juan J Lence-Anta, Celia María Pereda, Daniel F Comiskey Jr, Huiling He, Sandya Liyanarachchi, Albert de la Chapelle, Rossella Elisei, Federica Gemignani, Hauke Thomsen, Asta Forsti, Anthony F Herzig, Anne-Louise Leutenegger, Carole Rubino, Evgenia Ostroumova, Ausrele Kesminiene, Marie-Christine Boutron-Ruault, Jean-François Deleuze, Pascal Guénel, Florent de Vathaire. **Multiethnic genome-wide association study of differentiated thyroid cancer in the EPITHYR consortium.** *International Journal of Cancer*, 2021 Jun 15;148(12):2935-2946.

Om Kulkarni, Pierre-Emmanuel Sugier, Julie Guibon, Anne Boland-Auge, Christine Lonjou, Delphine Bacq-Daian, Robert Olasso, Carole Rubino, Vincent Souchard, Frédérique Rachedi, Juan Jesus Lence, Rosa Maria Ortiz7, Constance Xhaard, Pierre Laurent-Puig, Claire Mulot, Anne-Valérie Guizard, Claire Schvartz, Marie-Christine Boutron-Ruault, Evgenia Ostroumova, Ausrele Kesminiene, Jean-François Deleuze, Pascal Guénel, Florent De Vathaire, Thérèse Truong*, Fabienne Lesueur*. **Pathway-based and gene-network analyses of GWAS data identify functionally relevant gene modules for differentiated thyroid carcinomas.** Scientific Report. 2021 Apr 26;11(1):8932

Monia Zidane, Thérèse Truong, Fabienne Lesueur, Constance Xhaard, Emilie Cordina-Duverger, Jean-François Deleuze, Anne Boland, Catherine Ory, Sylvie Chevillard, Hélène Blanché, Vincent Souchard, Yan Ren, Mohammed Zakarya Zemmache, Sandra Canale, Françoise Borson-Chazot, Claire Schvartz, Eugénia Mariné Barjoan, Anne-Valérie Guizard, Pierre Laurent-Puig, Claire Mulot, Julie Guibon, Mojgan Karimi, Elizabeth Adjadj, Carole Rubino, Pascal Guenel, Jean-Baptiste Cazier, Florent de Vathaire. **Variants on DNA repair gene and diagnosis irradiation implication in differentiated thyroid cancer risk: A pooled-analysis of two case-control studies.** Cancer Epidemiology Biomarkers & Prevention. 2021 Jun;30(6):1208-1217

Communications affichées:

American Society of Human Genetic meeting. Houston, Texas, October 14th – 19th 2019. Julie Guibon, Pierre-Emmanuel Sugier, Mojgan Karimi, Om Kulkarni, Anne Boland-Auge, Delphine Bacq-Daian, Céline Besse, Elisabeth Adjadj, Frédérique Rachédi, Carole Rubino, Constance Xhaard1, Pierre Laurent-Puig, Claire Mulot, Anne-Valérie Guizard, Rosa Maria Ortiz, Yan Ren, Claire Schvartz, Evgenia Ostroumova, Jean-François Deleuze, Marie-Christine Boutron-Ruault, Ausrele Kesminiene, Florent De Vathaire, Pascal Guénel, Fabienne Lesueur*, Thérèse Truong*. **Fine-mapping of two differentiated thyroid carcinoma susceptibility loci at 2q35 and 8p12 detects novel candidate functional SNPs in Europeans, Melanesians and Polynesians.**

10èmes Assises de Génétiques Humaines et Médicales, Tours, 21-24 janvier 2020. Julie Guibon, Pierre-Emmanuel Sugier, Mojgan Karimi, Om Kulkarni, Anne Boland-Auge, Delphine Bacq-Daian, Céline Besse, Elisabeth Adjadj, Frédérique Rachédi, Carole Rubino, Constance Xhaard¹, Pierre Laurent-Puig, Claire Mulot, Anne-Valérie Guizard, Rosa Maria Ortiz, Yan Ren, Claire Schvartz, Evgenia Ostroumova, Jean-François Deleuze, Marie-Christine Boutron-Ruault, Ausrele Kesminiene, Florent De Vathaire, Pascal Guénel, Fabienne Lesueur*, Thérèse Truong*. **Cartographie fine des loci 2q35 et 8p12 associés à la prédisposition au cancer différencié de la thyroïde chez les Européens et Océaniens.**

Prix Josué Feingold de la Société Française de Génétique Humaine pour la meilleure communication affichée en génétique épidémiologique et bioinformatique

Liste des tableaux

Tableau 1 : Récapitulatif des différents loci de prédisposition mis en évidence dans les études familiales sur le CDT.	45
Tableau 2 : Principaux polymorphismes associés au risque de CDT dans des études "gène ou SNP candidat"	47
Tableau 3 : Descriptif des populations européennes et océaniques d'EPITHYR selon l'étude, l'âge, le sexe et le type histologique des carcinomes.	82
Tableau 4 : Association entre le CDT et les SNPs rapportés dans les précédentes GWAS aux loci 2q35, 8p12, 9q22.33 et 14q13.3 dans les populations européennes et océaniques d'EPITHYR.....	85
Tableau 5 : Résultats de l'analyse de réplication pour les SNPs identifiés	86
Tableau 6 : description des populations européennes, polynésiennes et mélanésiennes d'EPITHYR selon l'étude, l'âge, le sexe, le type histologique et la taille des carcinomes. ..	101
Tableau 7 : Résultats de l'analyse de colocalisation des SNPs ayant une probabilité postérieure (PP) > 0,10 aux loci 2q35 et 8p12 chez les Européens, les Polynésiens et les Mélanésiens.	119
Tableau 8 : Nombre de cas et témoins par étude selon le degré de métissage.	135
Tableau 9 : Description de la population européenne EPITHYR selon l'étude, l'âge, le sexe et l'histologie de la tumeur (après exclusion des apparentés).	137
Tableau 10 : Description des populations asiatiques, mélanésiennes et polynésiennes selon l'étude, l'âge, le sexe et l'histologie de la tumeur (après exclusion des apparentés).	138
Tableau 11 : Description des populations européennes d'EPIC selon l'étude, l'âge, le sexe et l'histologie de la tumeur (après exclusion des apparentés).	139
Tableau 12 : Caractéristiques des différentes populations et longueurs minimales des ROH (en nombre de SNPs) calculées pour les Européens, les Polynésiens, les Mélanésiens et les Asiatiques d'EPITHYR et d'EPIC.	145
Tableau 13 : Comparaison des caractéristiques globales des ROH chez les cas et les témoins européens d'EPITHYR et d'EPIC.....	146
Tableau 14 : Caractéristiques globales des ROH chez les cas et les témoins chez les Polynésiens, les Mélanésiens et les Asiatiques d'EPITHYR.	147

Tableau 15 : Association entre le nombre de ROH ou la longueur totale ROH et le risque de CDT chez les Européens.	148
Tableau 16 : Association entre le nombre de ROH ou la longueur totale des ROH et le risque de CDT chez les Polynésiens, les Mélanésiens et les Asiatiques d’EPITHYR.	149
Tableau 17 : Comparaison des coefficients de consanguinité (Froh) entre les cas et les témoins des différents groupes de population.....	151
Tableau 18 : Nombre de ROH partagés identifiés dans chaque population.	153
Tableau 19 : ROH partagés et associés au risque de CDT dans les différentes populations d’EPITHYR et d’EPIC.....	154
Tableau 20 : Description des individus des groupes EPITHYR+EPIC NORD et EPITHYR+EPIC SUD selon les études, l’âge le sexe et le type histologique.....	156

Liste des figures

Figure 1: La glande thyroïdienne (psychomedia.ca)	27
Figure 2 : Incidence standardisée des cancers de la thyroïde en 2020.	32
Figure 3 : Différents types de variations génétiques existantes sur le génome. Source : Frazer et al, 2009.	46
Figure 4 : Manhattan plot des p-valeurs d'association (exprimées en $-\log_{10}(p)$) des 7 millions de SNPs analysés dans la méta-analyse européenne (Gudmundsson et al. 2017).	49
Figure 5 : Etudes participantes au consortium EPITHYR.	60
Figure 6 : Etudes participantes au consortium EPIC.	61
Figure 7 : Principe de l'imputation. (Illustration de (Marchini and Howie 2010)).	66
Figure 8: Structure de la population d'EPITHYR.	69
Figure 9 : ACP de la population d'EPIC.	70
Figure 10 : Etapes du contrôle qualité des données génotypage d'EPITHYR et sélection des SNPs et des individus retenus pour les analyses.	76
Figure 11 : Stratégie d'analyse du GWAS.	78
Figure 12 : Miami plot des résultats de l'étude GWAS chez les Européens et chez les Océaniens.	84
Figure 13 : Analyses stratifiées des SNPs : A. rs334729 (1p31.3) ; B. rs16950982 (16q23) ; C. rs7726159 (5p15) ; D. rs10415826 (19p12).	87
Figure 14 : Stratégie de la cartographie fine utilisée dans notre étude.	97
Figure 15 : ACP des 649 individus Océaniens d'EPITHYR et des 940 individus du panel HGDP.	99
Figure 16 : Régions génomiques d'intérêt en 2q35 et 8p12 définies pour l'analyse de fine-mapping.	102
Figure 17 : Schéma des étapes d'imputation des SNPs manquants au locus 2q35 pour les trois populations d'EPITHYR.	104
Figure 18 : Schéma des étapes d'imputation au locus 8p12 pour les trois populations d'EPITHYR.	105

Figure 19 : Locuszoom de l'association entre les SNPs de la région 2q35 et le CDT chez (a) les Européens, (c) les Mélanésiens et (d) les Polynésiens. (b) Association conditionnellement au SNP rs57481445 chez les Européens.	109
Figure 20 : Locuszoom de l'association entre les SNPs des régions 2q35 et 8p12 et le CDT et résultats de la régression conditionnelle dans l'analyse poolée des Européens, Mélanésiens et Polynésiens. (a) Association au locus 2q35, (b) Association conditionnellement au SNP rs3821098 au locus 2q35, (c) Association au locus 8p12, (d) Association conditionnellement au SNP rs28406305.....	110
Figure 21 : Cartes de corrélation génétiques (r^2) entre les SNPs du locus 2q35 chez (a) les Européens, (b) les Mélanésiens et (c) les Polynésiens.	111
Figure 22 : Plots représentant les (a) $-\log_{10}$ (p-valeurs) d'association entre les SNPs au locus 2q35 et l'expression de différents gènes dans les cellules tumorales de la thyroïde (CPT), et résultats de probabilités postérieures obtenues pour chaque SNP dans l'analyse de colocalisation chez (b) les Européens, (c) les Mélanésiens et (d) les Polynésiens.	113
Figure 23 : Forest plots des analyses par sous-groupe pour le SNP rs16857609 (allèle à risque : T) chez les Européens, et par groupe ethnique.	115
Figure 24 : Locuszoom de l'association entre les SNPs de la région 8p12 et le CDT chez (a) les Européens, (c) les Mélanésiens et (d) les Polynésiens et de (b) l'association conditionnellement au SNP rs28406305 chez les Européens.	117
Figure 25 : Cartes de corrélation génétiques (r^2) entre les SNPs du locus 8p12 chez (a) les Européens, (b) les Mélanésiens et (c) les Polynésiens.	118
Figure 26 : Plots représentant les p-valeurs d'association des variants au locus 8p12.....	120
Figure 27 : Forest plots des analyses par sous-groupes pour A) le SNP rs2439304 (allèle à risque : A) chez les Européens, les Mélanésiens et les Polynésiens), B) le SNP rs7844425 (allèle à risque : A) chez les Européens, les Mélanésiens et les Polynésiens).	122
Figure 28 : Etapes du contrôle qualité des données génotypage d'EPIC et sélection des SNPs et des individus retenus pour les analyses.	130
Figure 29 : Analyse en composantes principales des individus classés dans les groupes "Asiatiques" et « Autres » d'EPITHYR par Fastpop réalisée en prenant en compte les populations européennes, asiatiques et océaniques du panel de référence HGDP.....	132

Figure 30 : Histogramme représentant le mélange de population des individus d'EPITHYR précédemment classés "Asiatique" ou "Autre" en utilisant comme populations de référence les individus européens, asiatiques et océaniens d'HGDP.	134
Figure 31 : Principe d'identification de ROH par le logiciel PLINK.....	142
Figure 32 : Illustration d'un ROH partagé entre individus.	143

Liste des tableaux supplémentaires

Tableau supplémentaire 1 : Polymorphismes associés au risque de CDT dans les études GWAS et dans la méta-analyse européenne.....	199
Tableau supplémentaire 2 : Résultats des analyses de réplication des tops SNPs chez les Européens aux loci ayant au moins 2 SNPs associés au CDT (p -valeur $< 10^{-05}$) dans l'analyse poolée.	202
Tableau supplémentaire 3 : Résultats d'association des précédents SNP GWAS identifiés aux locus 2q35 et 8p12 chez les Européens, les Mélanésiens et les Polynésiens.	204
Tableau supplémentaire 4 : Caractéristiques globales des ROH identifiés dans les groupes Nord et SUD d'EPITHYR et d'EPIC poolés.	205
Tableau supplémentaire 5 : Association entre le nombre de ROH ou la longueur totale ROH et le risque de CDT dans les groupes NORD et SUD des données européennes EPIC et EPITHYR poolées.....	206
Tableau supplémentaire 6 : Nombre de ROH partagés au sein des groupes NORD et SUD des données européennes d'EPIC et d'EPITHYR poolées.	207
Tableau supplémentaire 7 : Comparaison des coefficients de consanguinité (Froh) chez les cas et témoins des groupes NORD et SUD des données européennes d'EPIC et EPITHYR poolées.	208
Tableau supplémentaire 8 : Comparaison et caractéristiques des ROHs partagés chez les individus des études Polynésie Française selon le type de prélèvement d'ADN et selon différents seuils d'exclusion de SNPS avec nombre de cas et témoins similaires dans les deux groupes.	209

Liste des figures supplémentaires

Figure supplémentaire 1 : Diagrammes quantile-quantile des p-valeurs de GWAS chez (A) les Européens, (B) les Océaniens, (C) Analyse combinant les deux populations.....	210
Figure supplémentaire 2 : Manhattan plot des résultats de l'analyse GWAS menée chez les Européens de CPT.....	212
Figure supplémentaire 3 : Locuszoom des résultats d'association GWAS chez les Européens.	213
Figure supplémentaire 4 : Diagramme du contrôle qualité réalisé sur les données EPITHYR après ajout des individus de Polynésie Française 2.....	214
Figure supplémentaire 5 : Manhattan plot des p-valeurs d'associations obtenues avec un modèle logistique mixte chez les Européens d'EPITHYR (modèle récessif).....	216
Figure supplémentaire 6 : Analyse en composantes principales des individus d'EPITHYR et d'EPIC	217

Liste des abréviations

ADN : Acide Désoxyribonucléique
CDT : Cancer Différencié de la Thyroïde
CIRC : Centre International de Recherche sur le Cancer
CMT : Cancer Médullaire de la Thyroïde
CNMT : Cancer Non Médullaire de la Thyroïde
CNRGH : Centre National de Recherche en Génomique Humaine
CNV : Copy Number Variation
CP : Composante Principale
CPT : Cancer Papillaire de la Thyroïde
CVT : Cancer Vésiculaire de la Thyroïde
DL : Déséquilibre de Liaison
eQTL : Expression Quantitative Trait Locus
FAM : Fréquence de l'Allèle Mineur
GO : Gene Ontology
GWAS : Genome-Wide Association Study
HR : Hazard Ratio
HWE : Hardy-Weinberg Equilibrium
IC95% : Intervalle de Confiance à 95%
IMC : Indice de Masse Corporelle
KEGG : Kyoto Encyclopedia of Genes and Genomes
MMC : Modèle de Markov Caché
NC : Nouvelle-Calédonie
OR : Odds Ratio
PP : Probabilité *a posteriori* de causalité
PQL : Quasi Vraisemblance Périalisée
RR : Risque Relatif
SNP : Single Nucleotide Polymorphism

T3 : Triiodothyronine

T4 : Tétraiodothyronine

TCGA : The Cancer Genome Atlas Program

THM : Traitement Hormonal de la Ménopause

TRH : Thyréolibérine

TSH : Thyréostimuline

Introduction générale

I. 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 1). Elle a la forme d'un papillon et est composée de deux lobes (droit et gauche) reliés par un fin morceau de tissu appelé isthme. La 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 glande thyroïdienne est composée de deux types de cellules : les cellules folliculaires, encore appelées cellules vésiculaires ou thyrocytes, et les cellules parafolliculaires, encore appelées paravésiculaires ou cellules C. Les cellules folliculaires incorporent l'iode et synthétisent les hormones thyroïdiennes : la triiodothyronine (T3) et la tétraiodothyronine ou thyroxine (T4), sous le contrôle des hormones de l'axe hypothalamo-hypophysaire (thyroïdostimuline (TSH) et thyroïdolibérine (TRH)). Les cellules parafolliculaires régulent le métabolisme du calcium, par l'intermédiaire de la sécrétion de calcitonine.

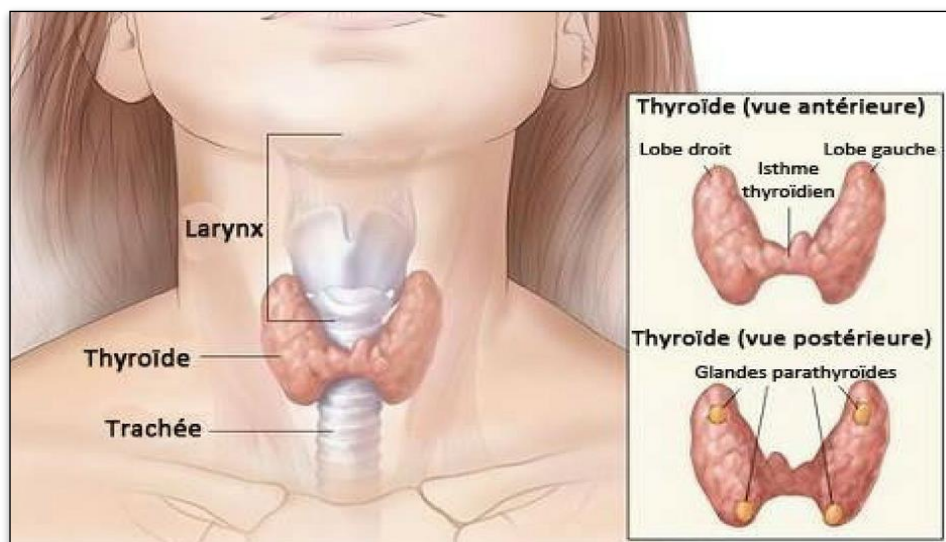


Figure 1: La glande thyroïdienne (psychomedia.ca)

II. Les cancers de la thyroïde

Les cancers de la thyroïde se manifestent principalement par un nodule, découvert le plus souvent à la palpation du cou ou lors d'examens d'imagerie. La plupart des nodules thyroïdiens sont indolores et asymptomatiques et correspondent à des adénomes hyperplasiques bénins. Cependant, environ 5% des nodules thyroïdiens sont des lésions malignes (cancers) qui peuvent avoir pour origine les cellules folliculaires ou les cellules C. Plus de 90% des cancers de la thyroïde se développent à partir des cellules folliculaires. Dans 5 à 10% des cas, le cancer se développe à partir des cellules C.

Lorsque le cancer de la thyroïde se développe à partir des cellules C, on parle alors de **cancers médullaires de la thyroïde (CMT)**. Bien que les CMT se présentent sous forme sporadique dans la majorité des cas, une prédisposition génétique existe pour ces cancers. Environ 25 % des CMT surviennent dans un contexte de syndrome autosomique dominant de néoplasie endocrinienne multiple de type 2 (NEM2) causée par des mutations germinales du proto-oncogène *RET* localisé sur le chromosome 10 dans la région 10q11.2 (Weber and Eng 2005). La NEM2 est une affection rare (prévalence estimée à 0,2/1000), héréditaire, de transmission autosomique dominante dont la pénétrance est voisine de 100 %. Il existe trois variants phénotypiques de la maladie :

- la NEM de type 2A (syndrome de Sipple) est le variant le plus fréquent (60 % des NEM2) ; elle associe un CMT présent dans 100 % des cas, à un phéochromocytome dans 20 à 50 % des cas et à une hyperparathyroïdie primaire dans 5 à 20 % des cas.
- la NEM de type 2B (syndrome de Gorlin), représente 5% des NEM2 ; elle associe au CMT constant un phéochromocytome dans 50 % des cas, une dysmorphie marfanoïde et une ganglioneuromatose cutanéomuqueuse et digestive.
- les formes familiales isolées de CMT pour lesquelles les autres composantes de la maladie sont absentes initialement et tout au long de l'évolution et qui représentent environ 35% des NEM2.

Les **cancers dérivés des cellules folliculaires** sont classés en « cancers différenciés » et « cancers non différenciés ». Les **cancers différenciés de la thyroïde (CDT)** incluent les cancers papillaires de la thyroïde ou CPT (80-85%), les cancers vésiculaires (ou folliculaires) de la thyroïde ou CVT (15-20%) et plus rarement les carcinomes à cellules de Hürthle et se caractérisent généralement par un excellent pronostic, avec une survie globale relative à 5 ans des CDT est supérieure à 90%. Les **cancers non différenciés** (10%), ou anaplasiques, sont au contraire agressifs et caractérisés par un mauvais pronostic. Malgré leur bon pronostic, les CDT sont responsables d'une altération de la qualité de vie des patients (Singer, Lincke et al. 2012). En effet, suite à l'ablation de la thyroïde une hormonothérapie est mise en place après la chirurgie et doit être prise à vie. Le traitement hormonal consiste à prendre des hormones thyroïdiennes de synthèse sous forme d'un médicament, la lévothyroxine, qui remplace les hormones naturelles produites par la thyroïde avant qu'elle ne soit retirée lors de la chirurgie.

III. Épidémiologie descriptive des cancers de la thyroïde

L'incidence des cancers thyroïdiens était d'environ 6,6 pour 100 000 personnes année (PA) en 2020 dans le monde, ce qui représente près de 586 000 nouveaux cas diagnostiqués, soit environ 3% de l'ensemble des cancers. Il est le cinquième cancer le plus fréquent chez les femmes après les cancers du sein, du colon, du poumon et du col de l'utérus (GLOBOCAN 2020).

L'incidence des cancers thyroïdiens est en effet environ trois à cinq fois plus élevée chez les femmes que les hommes (Kilfoy et al. 2009). Pour les deux sexes, la plus forte incidence des cancers thyroïdiens est observée à l'âge adulte 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).

Les taux d'incidence des cancers thyroïdiens sont également caractérisés par de **fortes variations ethniques et géographiques** (Figure 2). Les taux d'incidence les plus élevés en 2020 sont observés en Corée du Sud (44,7 pour 100 000 PA chez les femmes et 9,4 pour 100 000 PA chez les hommes) et sont expliqués en grande partie par le dépistage systématique mis

en place en 1999 dans ce pays (Vaccarella et al. 2015). Les taux d'incidence observés en Nouvelle-Calédonie (24,7 pour 100 000 PA chez les femmes et 7,3 pour 100 000 PA chez les hommes) et en Polynésie Française (26,3 pour 100 000 PA chez les femmes et 6,2 pour 100 000 PA chez les hommes) figurent toujours parmi les plus élevés (Figure 2). En Nouvelle-Calédonie, les plus forts taux sont observés dans de la population d'origine mélanésienne. Une étude menée sur des cas de CPT de Nouvelle-Calédonie 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 mélanésien sur la période 1985-1999 (Truong et al. 2007). Dans cette étude, l'incidence des CPT est quatre fois plus élevée chez les femmes d'origine mélanésienne que les femmes d'origine européenne (Truong et al. 2007).

Ces disparités géographiques et ethniques pourraient refléter des inégalités d'accès aux soins, des différences de pratiques diagnostiques ou encore des facteurs de risque environnementaux ou liés au mode de vie ainsi qu'une prédisposition génétique de certaines populations.

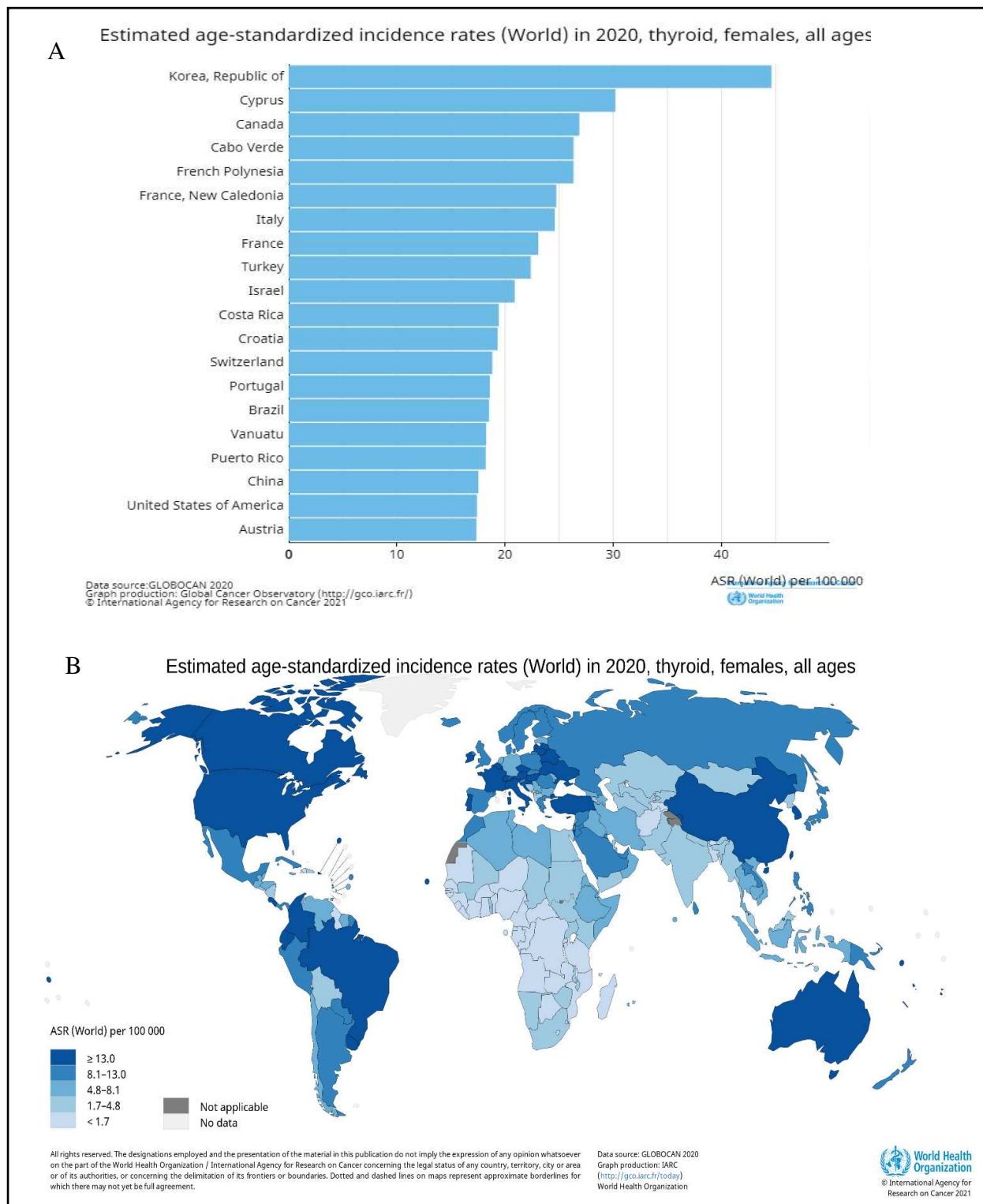
L'incidence des cancers thyroïdiens est en constante augmentation ces dernières décennies dans la plupart des pays du monde (La Vecchia et al. 2014). Ainsi, de 1990 à 2013, l'incidence globale des cancers thyroïdiens a augmenté de 20%, avec une plus forte augmentation observée dans les pays à faible revenu (33%) par rapport aux pays à fort revenu (19%) (Kim, Gosnell, and Roman 2019). En France, l'incidence est passée de 13,8 pour 100 000 PA chez les femmes et de 2,7 pour 100 000 PA chez les hommes en 2012 à respectivement 23,1 pour 100 000 PA et 6,1 pour 100 000 PA en 2020 (GLOBOCAN 2020).

Cette augmentation de l'incidence concerne surtout les CPT et est en partie expliquée par l'évolution des pratiques médicales (augmentation de la proportion de thyroïdectomie totale, utilisation plus fréquente de l'échographie, développement de la biopsie par aspiration à l'aiguille fine, évolution des pratiques anatomopathologiques permettant la réalisation de sections de plus en plus minces) amenant à un diagnostic plus fréquent de microcarcinomes latents (Colonna et al. 2020; Kitahara and Sosa 2016; Vaccarella et al. 2015; Vigneri and Malandrino 2015).

Cette dernière décennie, on observe une stabilisation, voire un léger déclin de l'incidence globale des cancers thyroïdiens dans certains pays comme les Etats-Unis, le Canada ou la France, où les recommandations de prise en charge des patients ont évolué et où il est préconisé de ne plus explorer les nodules thyroïdiens de taille inférieure à 1 cm (Colonna et al. 2020; Ellison and T. Bushnik 2020; Lim et al. 2017). Cette baisse d'incidence concerne principalement les CPT de petite taille mais on continue à observer en parallèle une augmentation des cancers thyroïdiens de stade avancé ou de grande taille (>4 cm) dans ces mêmes pays (Colonna et al. 2002; Kitahara and Sosa 2020; Wiltshire et al. 2016). Aussi, si le surdiagnostic était la raison exclusive de l'augmentation de l'incidence, on devrait s'attendre à une diminution du taux de mortalité liée aux cancers thyroïdiens due à une prise en charge plus précoce des patients, or il est observé globalement une stabilisation des taux de mortalité, voire parfois une légère augmentation dans certains pays. Ainsi, une récente étude menée sur les données de registres de cancers aux Etats-Unis (SEER-9) montre que l'incidence globale des cancers de la thyroïde a augmenté de 3,6 % par an entre 1974 et 2013 dans ce pays (Lim et al. 2017) et qu'en parallèle la mortalité liée aux cancers thyroïdiens a augmenté de 1,1 % par an, passant de 0,40 pour 100 000 PA en 1994-1997 à 0,46 pour 100 000 PA en 2010-2013. Une autre étude américaine menée sur les mêmes données de registres de cancer SEER-9 met en évidence à la fois un effet période et un effet cohorte dans l'augmentation de l'incidence des cancers de la thyroïde (Enewold et al. 2009) (Kitahara and Sosa 2016).

L'augmentation de l'incidence des CPT de grande taille, la mise en évidence d'un effet cohorte et l'évolution des taux de mortalité sont autant d'arguments suggérant que le surdiagnostic n'est pas la seule explication à l'augmentation de l'incidence des cancers thyroïdiens et qu'un rôle de facteurs de risque environnementaux ou individuels peut être suspecté. Dans les paragraphes suivants, nous faisons un état des lieux des facteurs de risque connus ou suspectés des cancers thyroïdiens.

Figure 2 : Incidence standardisée des cancers de la thyroïde en 2020.



A. Représentation des 20 pays ayant les plus fortes incidences de cancers de la thyroïde ; B. Cartographie de l'incidence des cancers de la thyroïde dans le monde (GLOBOCAN 2020).

IV. Les facteurs de risque non génétiques des cancers de la thyroïde

1. Exposition aux rayonnements ionisants

A ce jour, l'exposition aux rayonnements ionisants durant l'enfance constitue le seul facteur de risque environnemental clairement établi pour les CDT (RAPPORT UNSCEAR 2001). Il existe de multiples sources d'exposition aux radiations ionisantes telles que les rayonnements ionisants à visée médicale (diagnostique et thérapeutique) et les retombées radioactives des accidents nucléaires de Tchernobyl (Ukraine) en 1986, de Fukushima (Japon) en 2011, des bombardements atomiques d'Hiroshima et Nagasaki (Japon) en 1944.

Les niveaux d'expositions et les effets des radiations ionisantes varient selon leur nature et de la dose absorbée. En France, la dose maximale d'exposition artificielle est limitée à 1 mSv. Dans le cadre des pratiques médicales, les doses émises sont être relativement faibles avec par exemple une dose de 0,113 mSv reçue (en moyenne) pour une mammographie, tandis que la dose moyenne reçue par les habitants des régions à proximité de la centrale de Tchernobyl lors de l'accident était d'environ 300 mSv.

En Ukraine, il a été démontré que des doses de rayonnement élevées (950 - 17472 mGray) reçues à la glande thyroïdienne sont associées à des CPT plus agressifs sur le plan biologique et à une probabilité plus élevée de développer des cancers multifocaux et des pathologies nodulaires multiples chez des personnes ayant moins de 18 ans au moment de la catastrophe nucléaire (Zablotska et al. 2015). Aussi, une augmentation de l'incidence des cancers thyroïdiens de la population de Fukushima a été rapportée durant les années suivant l'accident nucléaire japonais (Shibata et al. 2021). De même, une très forte incidence de CDT a été observée chez les survivants aux bombes atomiques d'Hiroshima et Nagasaki et tout particulièrement chez les personnes ayant moins de 30 ans au moment de l'irradiation (Prentice et al. 1982).

2. Les antécédents personnels de pathologie bénigne de la thyroïde

Plusieurs études ont mis en évidence une forte association entre les antécédents personnels de goitre ou de nodules thyroïdiens bénins et le risque de cancer de la thyroïde. Des antécédents de goitre sont associés à un risque relatif de développer un cancer de la thyroïde proche de 5 (Franceschi et al. 1999; Horn-Ross et al. 2001; Memon et al. 2002; Truong et al. 2005) et un OR de 29,9 a été rapporté dans une méta-analyse chez les sujets ayant un antécédent de nodules thyroïdiens (Franceschi et al 1999). Malgré ces fortes associations, le lien entre ces deux pathologies et le cancer de la thyroïde est peu clair car les goitres ou nodules thyroïdiens pourraient constituer une pathologie précancéreuse plutôt qu'un facteur de risque à proprement parler du cancer thyroïdien, dont ils partageraient les mêmes facteurs étiologiques.

Une revue de la littérature récente (Tran et al. 2020) met également en évidence un risque de cancer de la thyroïde associé à un antécédent d'hypothyroïdie (RR = 3,31) ou d'hyperthyroïdie (RR = 3,31) mais ces associations sont basées sur des études de petites tailles qui ne tiennent pas toujours compte des facteurs de confusion importants tels que la prise de traitements médicaux.

3. Les facteurs anthropométriques

Une grande taille et un indice de masse corporelle (IMC) élevé ont été associés à un risque de cancer de la thyroïde dans de nombreuses études. Ainsi, une récente analyse poolée de 22 études prospectives incluant plus de 800 000 hommes et 1,2 M de femmes (Kitahara et al. 2016) rapporte une association positive entre la taille et le risque de CDT avec un HR de 1,07 (IC95% : 1,04 – 1,10) pour une augmentation de la taille de 5 cm et aussi une relation positive entre l'IMC et le risque de CDT avec un HR de 1,06 (IC95% : 1,02 – 1,10) pour une augmentation de 5kg/m².

En 2016, un rapport du Centre International de Recherche sur le Cancer (CIRC) sur l'obésité et le risque de cancers (Lauby-Secretan et al. 2016) conclut sur le fait que l'absence d'excès de masse adipeuse diminue le risque de cancer de la thyroïde chez l'Homme, avec un OR

comparant la catégorie la plus haute d'IMC par rapport à un IMC normal (18,5 – 24,9) de 1,1 (IC95% : 1,00 - 1,10).

4. Les facteurs hormonaux et reproductifs

Les taux d'incidence plus élevées chez la femme ainsi que le pic d'incidence observée chez ces dernières pendant la période reproductive ont conduit à suspecter un rôle des facteurs menstruels et reproductifs dans l'étiologie des CDT.

i. Les facteurs menstruels

Les facteurs menstruels le plus souvent étudiés dans la littérature sont l'âge aux premières règles, le statut ménopausique et, le cas échéant, l'âge à la ménopause, et les antécédents de chirurgie gynécologique. 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. Bien que les résultats rapportés dans la littérature sur ces indicateurs restent assez hétérogènes, certaines associations avec les cancers thyroïdiens ont été mises en évidence. Une méta-analyse de l'ensemble des études cas-témoins et des études de cohorte existantes sur l'association entre les facteurs reproductifs et le risque de cancer de la thyroïde au cours des deux dernières décennies a montré une association positive entre **un âge aux premières règles** supérieur à 14 ans et le risque de CDT (OR = 1,43 (IC95% : 1,10 - 1,77) (Mannathazhathu et al. 2019). Dans une autre méta-analyse récente portant sur 24 études incluant des études cas-témoins et des études de cohortes, **un âge tardif à la ménopause** est associé à une augmentation de risque de CDT (Cao et al. 2015).

ii. Les facteurs reproductifs

Les facteurs reproductifs les plus étudiés concernent le nombre de grossesses à terme, l'âge à la première grossesse, le délai depuis la dernière grossesse, les issues de grossesse et

l'allaitement. Les résultats sur ces facteurs ne sont pas toujours convergents. Les études dans des populations avec une parité élevée mettent en évidence une association positive avec le risque de CDT (Brindel et al. 2008; Memon et al. 2002; Truong et al. 2005). Ce résultat est retrouvé dans la méta-analyse de Mannathazhathu et coll. qui met en évidence une association positive significative chez les femmes ayant au moins deux enfants par rapport aux femmes nullipares (OR = 1,10, (IC95% : 0,94 - 1,27) (Mannathazhathu et al. 2019).

iii. Les traitements hormonaux

Le rôle des contraceptifs oraux dans les cancers de la thyroïde a été beaucoup étudié mais là encore, les résultats ne sont pas toujours convergents. Dans une analyse groupant 13 études cas-témoins le risque de cancer de la thyroïde est légèrement augmenté chez les femmes qui ont pris une contraception orale au cours de leur vie par rapport à celles qui n'en ont jamais pris (OR=1,2, IC95% : 1,0-1,4) (Vecchia et al. 1999). A l'inverse, une étude française plus récente a mis en évidence une association négative (OR = 0,55, IC95% 0,31–0,96) (Cordina-Duverger et al. 2017) alors que d'autres études ne montrent pas d'association (Brindel et al. 2009; Horn-Ross et al. 2011; Truong et al. 2005; Xhaard et al. 2015; Zamora-Ros et al. 2015).

De la même façon, les résultats concernant la prise d'un traitement hormonal de la ménopause (THM) était associée à une faible diminution du risque de cancer de la thyroïde dans l'analyse poolée de Vecchia et coll (Vecchia et al. 1999) alors que deux méta-analyses récentes ont conclu à l'absence d'association entre les THM et le risque de CDT (Caini et al. 2015; Cao et al. 2015).

5. La consommation d'alcool et de tabac

Les consommations de tabac et d'alcool sont des facteurs de risque avérés pour de nombreuses localisations cancéreuses. Les résultats des études épidémiologiques semblent cependant montrer une relation inverse entre ces consommations et le risque de cancer de la thyroïde.

Deux grandes méta-analyses, l'une basée sur 33 études (Wang et al 2016) et l'autre sur 27 études (Hong et al 2016) ont montré une association inverse entre la **consommation d'alcool** et le risque de CDT avec respectivement un RR de 0,75 (IC95% 0,69–0,83) et un RR de 0,80 (IC95% : 0,71 - 0,90) pour les plus fortes consommations par rapport aux plus faibles consommations.

Plusieurs études ont rapporté une association inverse entre la **consommation de tabac** et le risque de CDT (Belin et al. 2004; Guignard et al. 2007; Kabat et al. 2012; Meinhold et al. 2010; Soldin et al. 2009) ainsi qu'une relation dose-effet inverse avec l'intensité du tabagisme (Kitahara et al. 2012; Kreiger and Parkes 2000; Nagano et al. 2007), la durée (Kreiger and Parkes 2000; Rossing et al. 2000) et le nombre de paquets années consommés (Kitahara et al. 2012; Rossing et al. 2000). D'autres études n'ont pas rapporté d'association entre le tabac et le risque de CDT (Iribarren et al. 2001; Silvera, Miller, and Rohan 2005). Une méta-analyse publiée en 2014 à partir de 31 études observationnelles incluant des études cas-témoins et de cohorte, a montré une association inverse entre les fumeurs actuels et le risque de CDT comparé aux non-fumeurs (Cho and Kim 2014). Plus récemment, une étude de cohorte coréenne a mis en évidence, chez les hommes fumeurs actuels, une relation dose-effet inverse avec le nombre de paquets-années consommés (Cho et al. 2018).

6. L'alimentation

La plupart des études alimentaires sur les cancers de la thyroïde se sont focalisées sur le rôle des aliments riches en iode (Nettore, Colao, and Macchia 2018). Il a été suggéré qu'un excès d'iode alimentaire pouvait être associé à une augmentation de risque de CPT (Kolonel et al. 1990) alors qu'un déficit pouvait plutôt être associé au risque de CVT (Zimmermann and Galetti 2015). Une revue de la littérature souligne que les études expérimentales confirment l'association positive entre le déficit en iode et le risque de CVT, mais ne confirment pas la relation avec l'excès d'iode alimentaire (Zimmermann and Galetti 2015).

7. Autres facteurs de risque

D'autres facteurs de risque moins étudiés ont été évoqués dans la littérature tels que les perturbateurs endocriniens ou encore l'exposition aux métaux lourds dans les régions volcaniques. Cette dernière hypothèse est intéressante car elle pourrait expliquer l'incidence plus élevée observée dans les îles du Pacifique, à Hawaï, en Islande ou encore en Sicile, qui sont des îles volcaniques. Ainsi, dans les régions autour des volcans, la pollution par certains métaux lourds spécifiques (tels que l'arsenic, le cadmium, le chrome, le cobalt, le mercure, le tungstène et le zinc, entre autres) a été documentée comme une conséquence de l'émission de gaz, de cendres et de lave (Vigneri et al. 2017). Cette pollution entraînerait une contamination du sol, de l'eau et de l'atmosphère qui pourrait entrer dans la chaîne alimentaire. Ainsi, des taux de métaux lourds anormalement élevés ont été observés dans l'urine et les cheveux des personnes vivant dans des régions volcaniques (Vigneri et al. 2017). Il a été mis en évidence qu'à des concentrations plus élevées, les métaux lourds peuvent être toxiques pour les cellules vivantes et agir à la fois comme perturbateurs endocriniens et comme agents cancérigènes (Tchounwou et al. 2012). Le lien entre un environnement volcanique et le cancer de la thyroïde reste toutefois encore hypothétique.

V. Les facteurs de risque génétiques des cancers de la thyroïde

1. Antécédents familiaux des CDT

L'association entre la survenue d'un cancer de la thyroïde et l'existence d'antécédents familiaux de cancer de la thyroïde a fait l'objet d'études cas-témoins en population (Brindel et al. 2009; E Ron et al. 1987; Galanti et al. 1997; Horn-Ross et al. 2001; Leux et al. 2012; Mack et al. 2002), d'études cas-témoins hospitalières (Xu et al. 2012) et d'études de cohortes historiques à partir de bases de données issues de registres nationaux de cancers dans l'Utah, en Islande et en Suède (Czene, Lichtenstein, and Hemminki 2002; Goldgar et al. 1994; Hemminki et al. 2021; Hemminki, Eng, and Chen 2005; Hemminki and Dong 2000; Hemminki and Eng 2004; Hemminki and Vaitinen 1999; Hrafnkelsson et al. 2001). Les estimations du

risque relatif de cancer de la thyroïde chez les sujets ayant un apparenté de premier degré atteint de cancer de la thyroïde varient entre trois et huit. Le cancer de la thyroïde est ainsi un cancer pour lequel le risque familial est un des plus élevé. A titre d'exemple, le risque familial pour le cancer du sein est de l'ordre de deux, et ce, malgré une composante génétique importante et notamment l'implication de gènes majeurs de prédisposition comme *BRCA1* et *BRCA2* de pénétrance élevée expliquant environ 20% des formes familiales (Nelson et al. 2014). Dans l'étude suédoise estimant la part des cas attribuable aux facteurs génétiques et environnementaux pour différentes localisations de cancers à partir du registre national des familles et du registre national des cancers, la composante héréditaire est estimée à 53% pour les cancers de la thyroïde et est la plus élevée de tous les cancers étudiés (Czene et al. 2002). Ces résultats suggèrent un rôle important des facteurs de risque génétiques pour ce cancer.

Comme évoqué précédemment, le proto-oncogène *RET* est en cause dans la prédisposition au CMT mais la composante génétique des CDT, qui représentent 90% des cancers de la thyroïde, est loin d'être aujourd'hui élucidée. En effet, aucun gène de prédisposition majeur permettant de proposer un diagnostic moléculaire aux personnes présentant une prédisposition familiale n'a été mis en évidence et comme nous le verrons plus loin, tous les facteurs de risque génétiques identifiés sont associés à des risques relatifs relativement faibles.

Les formes familiales de CDT sont définies par la présence d'un CPT ou d'un CVT chez au moins deux apparentés au premier degré et ce en l'absence d'autres causes de prédisposition héréditaires ou environnementales. On estime que seulement 3 à 9% des CDT sont d'**origine familiale**. La grande majorité des CDT diagnostiqués sont ainsi qualifiés de CDT "sporadiques". Dans les premières études sur le CDT familial, la maladie était décrite comme étant un cancer multifocal dont les rechutes sont plus fréquentes et se manifestant généralement à un âge plus précoce que les formes sporadiques (Grossman et al. 1995). A partir de ces observations, il est apparu comme étant particulièrement important d'identifier les individus à haut risque de CDT afin de pouvoir établir un meilleur diagnostic, proposer un meilleur suivi aux patients et éviter ainsi des complications de la maladie telles qu'une extension extra-thyroïdienne du cancer ou encore des métastases ganglionnaires.

i. Syndromes héréditaires

Les études familiales ont suggéré un mode de transmission autosomique dominant de la maladie (Malchoff and Malchoff 2017; Sippel, Caron, and Clark 2007). Les formes familiales de CDT peuvent être séparées en deux groupes distincts : les formes syndromiques et les formes non-syndromiques. Dans les formes syndromiques, une prépondérance de tumeurs non thyroïdiennes est observée dans les familles. Ces syndromes à hérédité mendélienne ou “monogéniques” incluent le syndrome de Cowden (Liaw et al. 1997), la polypose adénomateuse familiale (PAF) ou syndrome de Gardner (Cetta et al. 2000), le complexe de Carney (Stratakis et al. 1997) et les syndromes de Werner et de DICER1 (Yu et al. 1996; Frio et al. 2011; Yang and Ngeow 2016).

Le **syndrome de Cowden** est un trouble multisystémique caractérisé par des hamartomes bénins de la peau, de la muqueuse buccale et de la langue, de la thyroïde, du sein, de l'utérus et du côlon (Pilarski 2009; Pilarski et al. 2013). Environ 75 % des personnes atteintes du syndrome de Cowden développeront une maladie bénigne de la thyroïde, le plus souvent des nodules multiples, tels que des adénomes folliculaires, ou un goitre adénomateux. Ces personnes ont également un risque accru de plusieurs types de cancer, notamment de CDT. Des mutations germinales du gène *PTEN* sont responsables de la prédisposition à ce syndrome (Liaw et al. 1997; Ngeow et al. 2011).

La **polypose adénomateuse familiale** se caractérise par le développement de polypes adénomateux dans le côlon et le rectum, avec un risque extrêmement élevé de développer un cancer du côlon au cours de la vie. Les individus atteints de ce syndrome présentent un risque plus élevé de développer une tumeur extra-colique, dont le CDT (Cetta et al. 2000, 2011). Il s'agit d'une maladie autosomique dominante causée par des mutations germinales du gène *APC*. Le **complexe de Carney** se caractérise par des anomalies pigmentaires de la peau, des myxomes, des schwannomes et des tumeurs endocrines et environ 75 % des personnes atteintes de ce syndrome présentent des adénomes folliculaires multiples ou d'autres nodules thyroïdiens. Cette maladie autosomique dominante est causée par des mutations germinales du gène *PRKARIA* (Stratakis et al. 1997).

Les caractéristiques cliniques du **syndrome de Werner** comprennent des changements cutanés semblables à la sclérodermie, des cataractes bilatérales, le grisonnement et la perte des cheveux, le diabète sucré, l'ostéoporose, l'athérosclérose et un risque accru de plusieurs tumeurs malignes, notamment de mélanome, de méningiome, de sarcome des tissus mous, de leucémie, d'ostéosarcome et de CDT (Epstein et al. 1966; Muftuoglu et al. 2008), le plus souvent de CVT (Lauper et al. 2013). Ce trouble autosomique récessif rare est causé par des mutations germinales du gène *WRN* (Muftuoglu et al. 2008).

Cependant ces syndromes n'expliquent qu'environ 5% des CDT familiaux et les gènes de prédisposition à la majorité des CDT restent à être identifiés.

ii. Loci de prédisposition aux CDT familiaux

L'étude de familles présentant des cas multiples de CDT suggère l'existence de gènes de prédisposition à pénétrance variable ou un modèle polygénique impliquant probablement de multiples allèles à faible pénétrance (Moses, Weng, and Kebebew 2011), ainsi qu'un degré élevé d'hétérogénéité génétique. Ainsi, à la fin des années 90, des approches de liaison génétique (ou *linkage*) reposant sur le fait que deux gènes ou deux marqueurs proches sur un même chromosome auront tendance à être transmis ensemble d'un individu à sa descendance, ont été réalisées dans des familles avec de multiples cas de CDT afin d'identifier les gènes de prédisposition à ce type cancer. Ces premières études menées sur l'ensemble du génome ont été possibles grâce à l'utilisation des premières cartes génétiques établies à partir de marqueurs polymorphes multi-alléliques appelés **microsatellites** (séquences d'ADN répétées en tandem d'un motif de deux à 10 pb correspondant à un locus unique dans le génome, parfaitement défini par les séquences uniques qui encadrent la répétition) permettant de baliser le génome. Cependant seule une faible proportion du génome est couverte par ces marqueurs. Une des méthodes les plus employées pour rechercher des régions du génome liées à une maladie ou à un trait phénotypique est la méthode du LOD score.

*Le **LOD score** (logarithm of odds, logarithme des probabilités) est une valeur numérique qui quantifie le degré de liaison génétique à partir des probabilités de transmission des allèles de deux gènes ou deux marqueurs génétiques observés dans la descendance. Une liaison entre le marqueur étudié et le gène morbide recherché est déclarée significative lorsque le LOD score, ou le cumul des scores obtenus pour différentes familles, est supérieur ou égal à 3. Une liaison peut être définitivement exclue, pour une valeur de recombinaison donnée entre le marqueur et le gène recherché, lorsque le LOD score est strictement inférieur à -2.*

Plusieurs loci ont été ainsi identifiés dans des familles présentant deux cas de CDT ou plus. Le premier locus, **MNG1**, a ainsi identifié en **14q32** grâce à l'étude d'une famille canadienne présentant deux cas de CPT et 18 individus ayant développé un goitre multinodulaire (LOD score maximal à 4,88) (Bignell et al. 1997). A ce locus, le séquençage direct du gène **DICER1** chez 53 membres issus de cinq familles (y compris la famille d'origine où la liaison a été identifiée pour la première fois en 1997 (Bignell et al. 1997)) a permis l'identification de "mutations" germinales, c'est à dire de "variants causaux" ou "pathogènes" (Frio et al. 2011). Les personnes porteuses de ces mutations de **DICER1** présentent une plus grande prédisposition au CDT (Khan et al. 2017), aux tumeurs des cellules de Sertoli-Leydig de l'ovaire (Frio et al. 2011) et aux blastomes pleuropulmonaires (Hill et al. 2009).

Par la suite, le locus **TCO** (pour *thyroid tumors with cell oxyphilia*) situé au locus **19p13.2**, a été identifié grâce à l'étude d'une famille française présentant une forme rare de tumeur thyroïdienne à cellules oxyphiles, avec un LOD score de 3,01 (Canzian et al. 1998). Cependant, l'étude d'autres familles à cas multiples de CDT n'a pas confirmé **TCO** comme locus de prédisposition majeur au CDT (Lesueur et al. 1999; McKay et al. 1999; Tsilchorozidou et al. 2006). Cependant, le rôle de **TCO** nécessite d'être étudié dans un plus grand nombre de familles prédisposées au CDT, présentant ou non le phénotype d'oxyphilie cellulaire (Bevan et al. 2001). En 2018, le séquençage de l'exome entier a permis de mettre en évidence un variant rare du gène **MYO1F** situé au locus **TCO** probablement impliqué dans la prédisposition au CDT.

Le potentiel tumorigène de ce variant a été mis en évidence dans des études fonctionnelles *in vitro* réalisées sur des lignées de cellules thyroïdiennes de rat hautement différenciées et *in vivo* chez le poisson zèbre avec une surexpression de la mutation induisant la croissance et la prolifération des cellules tumorales (Diquigiovanni et al. 2018).

Par la suite dans une deuxième étude reposant sur l'analyse de l'exome constitutionnel de cas de CDT familiaux a identifié des variants rares probablement délétères et situés dans les gènes pertinents d'un point de vue fonctionnel. Le séquençage ciblé de ces gènes chez 77 cas de CPT issus d'autres familles présentant une prédisposition à la maladie a identifié deux variants rares potentiellement causaux dans le gène **MAP2K5** (c.961G>A et c.1100T>C), tous deux situés dans la région codant le domaine kinase de MAP2K5. Les études fonctionnelles ont ensuite suggéré que la voie de signalisation alternative MAP2K5-ERK5 est l'une des voies impliquées dans l'étiologie du CDT familial (Ye et al. 2019).

Par analyse de liaison génétique, le locus **fPTC/PRN1** situé au niveau de la **région 1q21** a été identifié grâce à l'étude d'une famille américaine présentant cinq cas de CPT et pour dans laquelle un LOD score de 3,58 est obtenu pour ce locus (Malchoff et al. 2000) et le **locus NMTC1 a été identifié en 2q21** dans une famille Tasmanienne présentant 8 cas de CPT (McKay et al. 2001).

D'autres études de liaison plus récentes réalisées avec des puces à ADN permettant un criblage plus dense du génome en utilisant des variants nucléotidiques bialléliques ou **SNPs** (*single nucleotide polymorphisms*) communs ont permis d'identifier d'autres loci de prédisposition au CPT familial dont le **locus 8p23-p22 (FTEN)** lié au risque de CDT dans une famille portugaise (Cavaco et al. 2008) et le locus **8q24** dans une famille américaine (He et al. 2009). A ce locus, He et coll. ont réalisé une étude de cartographie fine (*fine-mapping*) de cette région à partir de 10 familles à cas multiple de CPT et ont proposé l'**ARN non codant AK023948**, ayant une expression plus faible au sein des tissus de carcinomes papillaires thyroïdiens qu'au sein des tissus thyroïdiens sains représente, comme candidat (He et al. 2009).

Le locus **12p14** a été mis en évidence grâce à l'étude d'un échantillon de 38 familles européennes dont 32 familles présentaient au moins trois membres atteints de CPT. Dans cette

étude, le séquençage direct de Sanger de la région d'intérêt chez les patients de 21 de ces familles a conduit à l'identification de quatre variants faux-sens dans le gène **SRGAP1**, co-ségrégeant avec la maladie (p.Gln149His, p.Ala275Thr, p.Arg617Cys, et p.His875Arg) (He et al. 2013). *SRGAP1* est un gène qui code une enzyme interagissant avec le complexe CDC42 jouant un rôle dans la régulation de diverses activités cellulaires dont la tumorigénèse (Etienne-Manneville 2004). Des études fonctionnelles *in vitro* menées sur des lignées cellulaires non cancéreuses de patients atteints de CPT et porteurs de l'une des quatre mutations ont montré ensuite que les variants faux-sens p.Gln149His et p.Ala275Thr conduisent à une activité fortement réduite de CDC42-GAP et entraînent le développement des cellules cancéreuses (He et al. 2013).

Une étude de séquençage de l'exome de patient atteints de CPT familial suivie d'une analyse des insertions et délétions dans différents gènes chez les membres affectés de 44 autres familles a mis en évidence un variant situé dans le gène **NOP53** gène (Orois et al. 2019). Un variant générant le changement faux-sens p.Asp31His, dans le gène *NOP53* et l'analyse immunohistochimique d'échantillons tumoraux de patients et porteurs du variant au niveau constitutionnel a montré une augmentation de l'expression de NOP53 dans le tissu tumoral par rapport au tissu thyroïdien sain adjacent (Orois et al. 2019). Le rôle de *NOP53* dans la morphologie nucléaire, la stabilité chromosomique et l'intégrité mitotique soutient l'hypothèse que ce gène pourrait être impliqué dans la carcinogénèse des cellules folliculaires de la thyroïde (Lee et al. 2018).

Enfin, signalons une dernière étude s'intéressant aux rôles des micro-ARN (miARN) dans l'étiologie du CDT. Les *miRNA* sont des ARN non codants de 20 à 22 nucléotides très conservés et impliqués dans la régulation l'expression des gènes au niveau post-transcriptionnel. Xiong et coll. ont montré que les **miR-886-3p** et **miR-20a** sont respectivement trois et quatre fois plus exprimés dans les tumeurs des patients présentant une forme familiale de CPT par rapport aux patients atteints de CPT sporadique (Xiong et al. 2011).

Ainsi, les études familiales et en particulier les études de liaison suivies du reséquençage des régions liées au CPT ou l'analyse d'exome de patients atteints de CPT familial ont permis

l'identification de 18 gènes candidats répliqués ou non dans d'autres séries de familles. Le **tableau 1** récapitule les loci de prédisposition au CDT décrits à ce jour.

Tableau 1 : Récapitulatif des différents loci de prédisposition mis en évidence dans les études familiales sur le CDT.

Position	Nom du Locus	N cas, familles	Gène candidat	Réplication	Référence	Technique d'analyse
1q21	<i>fPTC1/PRN1</i>	1 famille, 5 cas CPT	-		(Malchoff et al. 2000)	Etudes de liaison / séquençage de l'exome (WES)
2q21	<i>NMTC1</i>	1 famille, 8 cas CPT	-		(McKay et al. 2001)	
12p14	-	38 familles, 108 cas	<i>SRGAP1</i>		(He et al. 2013)	
14q32	<i>MNG1</i>	1 famille, 18 cas 5 familles, 37 cas	<i>DICER1</i>		(Bignell et al. 1997; Frio et al. 2011)	
	-	-	miR-886-3p et miR-20a		(Xiong et al. 2011)	
16p13.3	-	1 famille, 2 cas	<i>SRRM2</i>		(Tomsic et al. 2015)	
19p13.2	<i>TCO</i>	1 famille, 9 cas 1 famille, 3 cas	<i>MYO1F</i>	Oui	(Canzian et al. 1998; Diquigiovanni et al. 2018)	
8q24	-	25 familles, 72 cas	AK023948		(He et al. 2009)	
8p23.1-p22	<i>FTEN</i>	1 famille, 17 cas	-		(Cavaco et al. 2008)	
15q23	-	34 familles, 77 cas	<i>MAP2K5</i>		(Ye et al. 2019)	
19q13.33	-	44 familles, 95 cas	<i>NOP53</i>		(Orois et al. 2019)	Etudes d'association génome entier (GWAS)
9q22.33	-	378 cas	<i>FOXE1</i>		(Gudmundsson et al. 2009)	
9q22.33	-	65 cas	<i>PTCS2</i>		(He et al. 2015)	
9q22.33	-	-	<i>MYH9</i>		(Wang et al. 2017)	
14q13.3	-	378 cas 284 cas	<i>NKX2-1</i>	Oui	(Gudmundsson et al. 2009; Ngan et al. 2009)	
2q35		561 cas 1 003 cas 1 085 cas	<i>DIRC3</i>	Oui	(Gudmundsson et al. 2012, 2017; Son et al. 2017)	Séquençage du génome (WGS)
22q12.1	-	5 familles, 23 cas	<i>CHEK2</i>		(Srivastava et al. 2019)	
14q12	-	24 familles, 34 cas	<i>TINF2</i>		(He et al. 2020)	
7q31.33	-	1 famille, 8 cas	<i>POT1</i>		(Srivastava et al. 2020)	

2. Les études cas-témoins en population générale

i. Approche « gène candidat »

Avant l'avènement des études GWAS, des approches “**gènes candidats**” se focalisant sur des variations génétiques (Figure 3) localisées dans des gènes impliqués dans une fonction biologique pertinente pour le CDT ont été menées. Les résultats de ces études sont hétérogènes et peu ont été par la suite répliqués dans des études indépendantes.

Les **SNPs (pour Single Nucleotide Polymorphism)** constituent la forme la plus abondante de variations génétiques dans le génome humain. Il s'agit d'une variation d'une seule paire de base du génome entre individus. Les SNPs représentent plus de 90% des différences génétiques entre deux individus (figure 3).

Figure 3 : Différents types de variations génétiques existantes sur le génome. Source : Frazer et al, 2009.

Single nucleotide variant	ATTGGCCTTAAC C CCGATTATCAGGAT ATTGGCCTTAAC T CCGATTATCAGGAT	Structural variants
Insertion-deletion variant	ATTGGCCTTAAC CCGAT CCGATTATCAGGAT ATTGGCCTTAAC --- CCGATTATCAGGAT	
Block substitution	ATTGGCCTTAAC CCCC GATTATCAGGAT ATTGGCCTTAAC AGTC GATTATCAGGAT	
Inversion variant	ATTGGCCTTA ACCCCCG ATTATCAGGAT ATTGGCCTT CGGGGGT TATTATCAGGAT	
Copy number variant	ATT GGCCTTAGGCCTTA ACCCCCGATTATCAGGAT ATT GGCCTTA -----ACCTCCGATTATCAGGAT	

Les principaux polymorphismes, pour la plupart des SNPs, associés au risque de CDT dans au moins deux études sont rapportés dans le tableau 2. La majorité de ces polymorphismes sont situés dans la séquence codante de gènes candidats et ont pour conséquence un changement

d'acide aminé dans la séquence protéique. D'autres SNPs ont été étudiés car ils altèrent une région régulatrice ou UTR ou par exemple un site de fixation de facteurs de transcription pouvant être à l'origine d'une modification de l'expression des gènes.

Tableau 2 : Principaux polymorphismes associés au risque de CDT dans des études "gène ou SNP candidat"

Position	SNP	Allèle à risque	Freq	Gène candidat	Fonction du gène	Référence
14q32.33	rs861539	A	0,36	<i>XRCC3</i>	Réparation de l'ADN et de la régulation du cycle cellulaire	(Bastos et al. 2009; Sturgis et al. 2005)
19q13.31	rs25487 rs1799782	C A	0,73 0,06	<i>XRCC1</i>		(Akulevich et al. 2009; Chiang et al. 2008; Ho et al. 2009)
11q22.3	rs1801516	A	0,13	<i>ATM</i>		(Akulevich et al. 2009)
17p13.1	rs1042522	C	0,71	<i>TP53</i>	Stabilité du génome ; gène suppresseur de tumeur	(Akulevich et al. 2009)
1p13.3	Délétion	-	-	<i>GSTM1</i>	Détoxification. La délétion des gènes correspondants entraîne une perte fonction	(Hernández et al. 2008; Landa and M. Robledo 2011; Tcheandjieu et al. 2020)
22q11.33	Délétion	-	-	<i>GSTT1</i>		(Hernández et al. 2008; Landa and M. Robledo 2011; Tcheandjieu et al. 2020)
8p22	rs1801280	C	0,53	<i>NAT2</i>	Catalyse la N-acétylation et l'O-acétylation des acides aminés aromatiques et hétérocycliques cancérigènes	(Hernández et al. 2008; Landa and M. Robledo 2011)
2p25.3	rs732609 rs2048722	C G	0,40 0,44	<i>TPO</i>	Catalyse de biosynthèse des hormones thyroïdiennes	(Cipollini et al. 2013; Landa and M. Robledo 2011)

XRCC1*, *XRCC3* et *ATM sont des gènes de la réparation de l'ADN et de la régulation du cycle cellulaire ; ils jouent notamment un rôle important dans la réparation des cassures double brin de l'ADN causées par l'exposition aux radiations ionisantes. ***TP53***, le « gardien du génome »,

est un gène essentiel au maintien de la stabilité du génome. C'est un gène suppresseur de tumeur qui joue un rôle essentiel lors de l'apparition d'anomalies en régulant la prolifération ou la mort cellulaire. *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 spécifiques à la thyroïde comme *TPO* et *TG* (Landa and M. Robledo 2011). TPO est une peroxydase thyroïdienne ayant pour rôle de catalyser la première étape de 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).

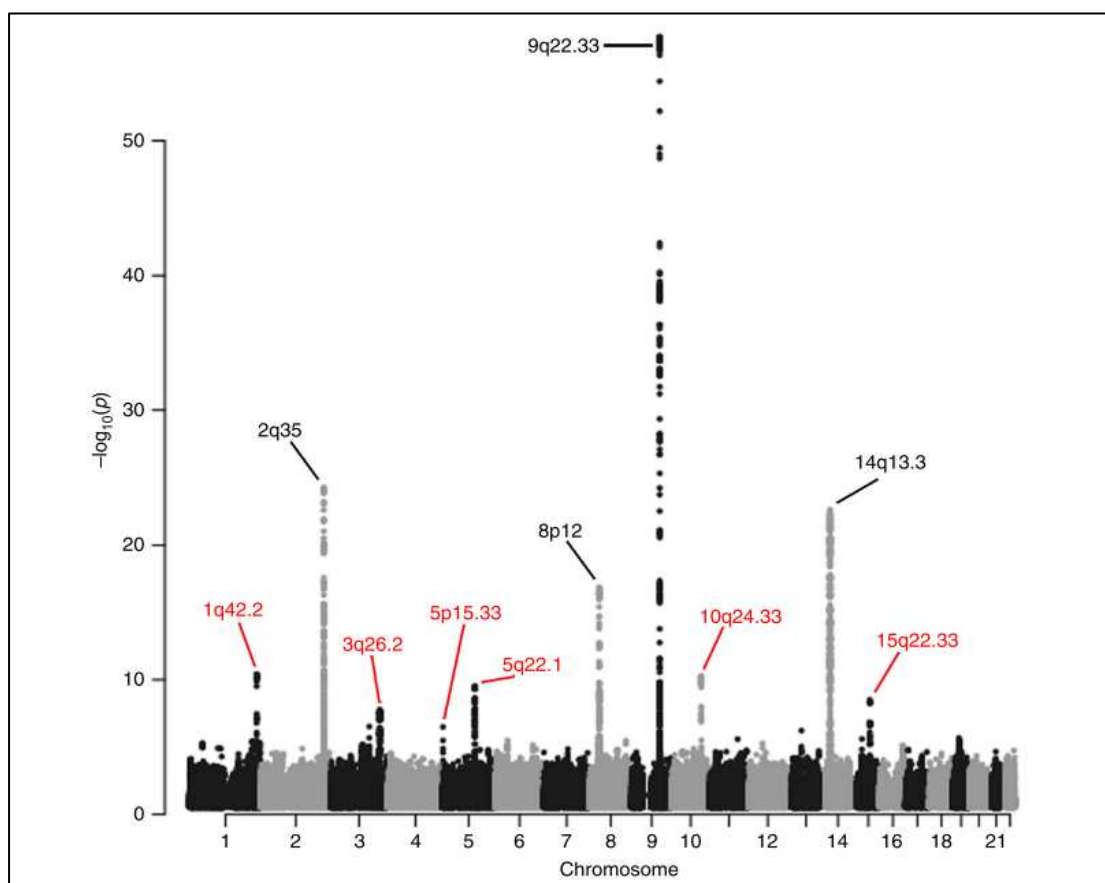
GSTM1, *GSTT1* et *NAT2* sont des gènes impliqués dans le métabolisme des xénobiotiques, codants pour des enzymes de phase II, dites de conjugaison. Les enzymes GST (glucosyltransférases) jouent un rôle important dans la détoxification, la délétion des gènes correspondants entraînent une perte fonction et ont été associés à un risque plus élevé de cancer (Hernandez 2008). Les résultats d'association entre les polymorphismes des gènes *GSTM1* et *GSTT1* et le risque de CDT sont hétérogènes. Les génotypes nuls (délétions du gène entier) de *GSTM1* et *GSTT1* ont été associés à un risque de CPT (Lemos et al. 2008). Ces deux associations ont également été mises en évidence dans une étude portant sur 223 patients saoudiens atteints de CPT (Siraj et al. 2008), mais un effet protecteur a été associé à *GSTT1*. Une étude menée chez des populations européennes et mélanésiennes a montré que les gènes *GSTM1* et *GSTT1* modulaient l'association entre le risque de CDT et l'obésité (Tcheandjieu et al. 2020).

Les deux isozymes de la N-acétyltransférase, le NAT1 et le NAT2, sont polymorphes et catalysent à la fois la N-acétylation (généralement la désactivation) et l'O-acétylation (généralement l'activation) des acides aminés aromatiques et hétérocycliques cancérigènes. Des études épidémiologiques suggèrent que les polymorphismes d'acétylation de *NAT1* et de *NAT2* modifient le risque de développer des cancers de la vessie, colorectal, du sein, de la tête et du cou, du poumon et de la prostate. Des polymorphismes dans des régions codantes de *NAT2* ont permis de classer des populations en fonction de leurs profils d'acétylation (lent ou rapide). Un phénotype acétyleur rapide et certains polymorphismes de *NAT2* ont été associés au risque de cancer de la thyroïde (Guilhen et al. 2009; Hernández et al. 2008).

3. Les études pangénomiques de type « GWAS »

Lorsque j'ai débuté cette thèse, six GWAS avaient été menées sur le CDT. Au total, 39 SNPs situés dans 20 loci distincts du génome avaient été identifiés dans des GWAS réalisées dans différentes populations principalement d'origine européenne (Tableau supplémentaire 1, figure 4) et seules les quatre régions 2q35, 8p12, 9q22.33 et 14q13.3 ont été répliquées dans au moins une étude indépendante. La méta-analyse des études GWAS européennes publiée en 2017 (Gudmundsson et al. 2017) a permis de mettre en évidence 5 nouveaux loci de prédisposition au CDT en 1q42.2, 3q26.2, 5p15.33, 10q24.33 et 15q22.33 dans la phase de découverte mais n'ont pas été répliqués faute d'échantillon indépendant disponible.

Figure 4 : Manhattan plot des p-valeurs d'association (exprimées en $-\log_{10}(p)$) des 7 millions de SNPs analysés dans la méta-analyse européenne (Gudmundsson et al. 2017).



Sont indiqués tous les loci significativement associés au risque de CDT dans les populations européennes. Source : (Gudmundsson et al. 2017).

Une étude menée sur des cohortes a démontré que cinq SNPs (rs965513, rs944289, rs966423, rs2439302 et rs116909374) identifiés dans les quatre loci principaux de prédisposition au CDT (2q35, 8p12, 9q22.33, 14q13.3) n'expliqueraient que 11% du risque familial de CDT dans une cohorte de l'Ohio et 6% dans une cohorte polonaise (Liyanarachchi et al. 2013.).

4. L'identification de variants récessifs

Les études familiales permettent de mettre en évidence des SNPs à transmission autosomique récessive en comparant les segments génomiques identiques (segments IBD ou identiques par descendance) entre individus sains et malades d'une même famille. En population générale, une stratégie qui est de plus en plus fréquemment utilisée ces dernières années consiste à détecter les régions d'homozygotie (ou runs of homozygosity, ROH), les ROH les plus longs correspondant le plus souvent aux segments homozygotes par descendance (ou *Homozygous By Descent*, HBD). Ainsi, la comparaison des caractéristiques d'homozygotie (fréquence, longueur,...) chez les cas et les témoins pour une maladie donnée permet d'évaluer le rôle de variants récessifs dans le risque de cette maladie. Par exemple, cette stratégie a permis d'identifier des segments homozygotes associés à des risques plus élevés de développer différents types de cancers tel que le cancer colorectal, le cancer du poumon ou encore le cancer dans des populations européennes (Bacolod et al. 2015; Orloff et al. 2012; Thomsen et al. 2015).

Objectifs de la thèse

En résumé, les études épidémiologiques ont montré que l'exposition aux radiations ionisantes particulièrement durant l'enfance et l'obésité sont les seuls facteurs de risque environnementaux ou liés au mode de vie bien établis pour les CDT. Concernant les facteurs génétiques, quatre loci de prédisposition au CDT contenant des gènes d'intérêt, en 2q35 (DIRC3), 8p12 (NRG1), 9q22.33 (FOXE1) et 14q13 (NKX2-1), ont été identifiés dans les GWAS et répliqués dans plusieurs études. D'autres loci ont été suggérés dans les études familiales ou de population mais n'ont pas été à ce jour confirmés dans d'autres études indépendantes.

L'objectif général de cette thèse est d'étudier les facteurs de risque génétiques des CDT.

Dans le premier chapitre de cette thèse, je présenterai les résultats du GWAS réalisé dans la population multi-ethnique du consortium EPITHYR. Dans le deuxième chapitre, je présenterai les résultats de la cartographie fine ou « fine-mapping » des régions 2q35 et 8p12 répliquées dans le GWAS EPITHYR qui a permis d'identifier de potentiels variants causaux dans les populations Européennes et Polynésiennes et Mélanésiennes d'EPITHYR. Dans le troisième chapitre, je décrirai l'étude des régions d'homozygotie du génome que j'ai réalisée dans les populations d'EPITHYR afin d'identifier de nouveaux facteurs génétiques impliqués dans la prédisposition au CDT. Cette dernière analyse est motivée par l'hypothèse d'un mode de transmission récessif des CDT qui n'a pas été exploré dans les études GWAS.

Etudes

Mes travaux ont été effectués en utilisant les données des consortia EPITHYR et EPIC décrits ci-dessous.

I. Le consortium EPITHYR

Le consortium EPITHYR a pour objectif d'améliorer les connaissances sur l'étiologie des cancers de la thyroïde en explorant le rôle des facteurs génétiques en combinaison avec des facteurs environnementaux ou liés au mode de vie. Ce consortium initié en 2015 et coordonné par Thérèse Truong, réunit les données de sept études cas-témoins sur les CDT menées dans des populations très contrastées originaires de France métropolitaine (études CATHY, YoungThyr, E3N), de Nouvelle-Calédonie (étude NC), de Polynésie Française (étude Polynésie Française), de Cuba (étude CUBA) et de Biélorussie (étude Tchernobyl) (Figure 5). Les données épidémiologiques et les prélèvements biologiques (sang, salive) sont disponibles pour plus de 2 800 cas et 4 200 témoins, faisant d'EPITHYR une ressource unique pour étudier les facteurs de risque génétiques et non génétiques impliqués dans l'étiologie du CDT.

1. L'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 (Cordina-Duverger et al. 2017).

Sélection des cas. Les cas éligibles correspondent à l'ensemble des patients avec un CDT diagnostiqués entre le 1er janvier 2002 et le 31 décembre 2007, résidant dans l'un des trois départements concernés 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-Ardenne. Sur 853 cas éligibles, **621** (73%) ont accepté de participer.

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 à l'année de référence et résidant dans les départements du Calvados, de la Marne et des Ardennes. Le tirage au sort a été 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 aux cas sur le sexe et l'âge quinquennal. Le tirage au sort des témoins a été stratifié sur la catégorie socioprofessionnelle de manière à ce que la répartition des catégories socioprofessionnelles des témoins soit représentative de celle de leur département de résidence.

Sur 943 cas éligibles, 706 (75%) cas ont accepté de participer à l'étude et un prélèvement salivaire est disponible pour 516 d'entre eux. Parmi les 706 témoins de l'étude, **569** ont accepté de fournir un prélèvement salivaire.

2. L'étude Nouvelle-Calédonie (NC)

L'étude NC est une étude cas-témoins en population générale conduite en Nouvelle-Calédonie entre 1997 et 1999 (Guignard et al. 2007).

Sélection des cas. Tous les patients atteints d'un CPT ou d'un CVT diagnostiqué entre le 1er janvier 1993 et le 31 décembre 1999 et ayant résidé en Nouvelle-Calédonie pendant au moins 5 ans au moment du diagnostic du cancer étaient éligibles pour l'étude. Les cas ont été identifiés à partir des dossiers médicaux et des comptes rendus d'anatomopathologie conservés par les deux laboratoires d'histopathologie du territoire. Tous les cas diagnostiqués au cours de la période de 5 ans précédant le début de la collecte des données en 1998 et tous les cas nouvellement diagnostiqués jusqu'à la fin de la période d'étude en 1999 ont été inclus dans l'étude. Le bon pronostic des tumeurs thyroïdiennes différenciées et la stabilité résidentielle ont permis le recrutement des sujets plusieurs années après le diagnostic initial.

Sélection des témoins. Les témoins ont été sélectionnés à partir de listes électorales récemment mises à jour, qui comprenaient le nom, l'adresse et la date de naissance des résidents de

Nouvelle-Calédonie âgés de 18 ans ou plus. Les témoins ont été appariés par fréquence aux cas par sexe et par groupe d'âge de 5 ans. Les témoins se sont vus attribuer une année de référence égale à l'année de diagnostic du groupe de cas pour lequel ils ont été sélectionnés. Les témoins ont été exclus s'ils ne résidaient pas en Nouvelle-Calédonie depuis au moins 5 ans à la date de référence ou s'ils avaient eu un cancer de la thyroïde avant cette date.

Au total, 332 cas et 412 témoins sont inclus dans cette étude et **224 cas et 261 témoins** ont fourni un prélèvement salivaire.

3. L'étude YoungThyr

L'étude YoungThyr est une étude cas-témoins menée dans l'Est de la France entre 2005 et 2010 dans les régions ayant potentiellement été exposées aux retombées nucléaires après l'accident de Tchernobyl en 1986 (Xhaard et al. 2015).

Sélection des cas. Les cas éligibles sont les patients ayant un CDT diagnostiqué entre le 1er janvier 2002 et le 31 décembre 2006, nés après le 1er janvier 1971 et résidant dans une région de l'est de la France (Alsace, Champagne-Ardenne, Corse, Franche-Comté, Lorraine, Rhône-Alpes ou Provence-Alpes-Côte d'Azur). Les cas incidents ont été identifiés d'après trois sources principales, selon la région : 1) le Registre général des incidences du cancer en Champagne-Ardenne, Alsace et Rhône-Alpes ; 2) le Registre national des cancers de l'enfant (qui contient des informations sur les enfants de moins de 15 ans) dans toutes les régions ; et 3) les hôpitaux privés et publics de Lorraine, Franche-Comté, Corse et Provence-Alpes-Côte d'Azur. Tous les cas de CDT ont été confirmés histologiquement. Sur 1049 cas éligibles, **805 cas (79%)** ont été inclus dans l'étude et ont fourni un prélèvement salivaire.

Sélection des témoins. Les témoins ont été sélectionnés en population générale et appariés individuellement aux cas sur le sexe, la date de naissance (plus ou moins 1 an) et la région de résidence à l'année de référence. Les témoins ont été sélectionnés à partir de l'annuaire des abonnés au téléphone fixe. Cette sélection a été réalisée à l'aide d'un générateur de nombres aléatoires et d'un fichier de France Télécom contenant tous les numéros de téléphone fixes (environ 100 millions) en France. Étant donné qu'une grande partie de la population de la

tranche d'âge ciblée (10-35 ans au moment de l'interview) ne possède qu'un téléphone portable et qu'un certain nombre de personnes ne figurent pas dans l'annuaire, les témoins incluent en plus des personnes sélectionnées dans l'annuaire les proches des témoins déjà sélectionnés. Au moment du premier contact téléphonique, les témoins ont dû préciser qu'ils n'avaient pas eu de cancer de la thyroïde, mais ils n'ont pas subi d'examen médical. Au total, **876 témoins** ont été inclus, certains cas étant appariés à 2 témoins. Tous les témoins ont fourni un prélèvement salivaire.

4. L'étude CUBA

L'étude CUBA est une étude cas-témoins en population générale conduite à La Havane entre 2011 et 2012.

Sélection des cas. Les patients âgés de 17 à 60 ans, vivant à La Havane et dans la municipalité voisine de Jaruco (à 30 km de La Havane) et traités pour un CPT entre 2000 et 2011 à l'Institut National d'Oncologie et de Radiobiologie (INOR) et à l'Institut d'endocrinologie, étaient éligibles pour l'étude. Les cas potentiels ont été sélectionnés dans les bases de données du Registre national du cancer (C73 ICD-10) et leurs informations ont été croisées avec le registre des pathologies de l'INOR. Vingt-cinq à 30 patients répondant à ces critères ont été identifiés pour chaque année de la période d'étude. Afin d'atteindre la taille d'échantillon estimée, la sélection a été étendue aux cas incidents. Sur les 240 cas de CPT éligibles, **203 cas** (85%) ont été inclus. Un prélèvement salivaire a été effectué pour l'ensemble des cas (Juan J. Lence-Anta et al. 2014).

Sélection des témoins. Deux cent vingt-neuf témoins de la population générale vivants dans la ville de La Havane et ses environs, la municipalité de Jaruco, ont été sélectionnés en utilisant les dossiers de consultation des unités de soins primaires (médecins de famille). Ils ont été appariés par fréquence avec les cas sur l'âge au moment du diagnostic des cas (± 5 ans) et le sexe. Parmi ces 229 témoins, **212** ont accepté de participer à l'étude (Juan J. Lence-Anta et al. 2014) et ont fourni un prélèvement salivaire.

5. L'Étude de Polynésie française

Sélection des cas. Tous les patients diagnostiqués avec un cancer de la thyroïde avant l'âge de 56 ans, nés et résidant en Polynésie française en 2004 étaient éligibles pour l'étude. Les cas prévalents ont été identifiés à partir du registre du cancer de la Polynésie française, des fichiers de l'assurance médicale et des quatre endocrinologues de Tahiti. Les caractéristiques pathologiques des tumeurs ont été obtenues auprès des deux laboratoires d'histopathologie et à partir des dossiers médicaux des endocrinologues (Brindel et al. 2008).

Sélection des témoins. Pour chaque cas éligible, deux témoins potentiels, appariés sur la date de naissance (± 6 mois) et le sexe, ont été sélectionnés au hasard dans le registre des naissances de la Polynésie française, qui recense tous les habitants qui y sont nés. Chacun de ces témoins a été apparié à un cas du même sexe et ayant la même année de naissance. Les données épidémiologiques ont été recueillies et portaient sur l'origine ethnique, le niveau d'éducation, les habitudes tabagiques, l'activité physique récréative au cours de la vie, les facteurs gynécologiques et reproductifs pour les femmes, l'exposition aux radiographies médicales, le poids à différents âges et la taille lors de l'entretien.

Deux échantillons de la population de Polynésie française ont été utilisés pour les analyses présentées dans cette thèse. Le premier échantillon comprend **225 cas et 368 témoins** (82 cas ont été appariés avec un seul témoin et 143 cas ont été appariés avec deux témoins) (Brindel et al. 2009), dont 144 cas et 231 témoins ont fourni soit un prélèvement salivaire, soit un prélèvement de cellules buccales (Brindel et al. 2008). Le deuxième échantillon (« Polynésie Française 2 »), dont le recueil des données et le génotypage ont été finalisés en 2020, comprend **166 cas et 225 témoins**.

Les analyses des chapitres 1 (analyses GWAS EPITHYR) et 2 (Fine-mapping des régions 2q35 et 8p12) se sont appuyées sur les données du premier échantillon et les analyses du chapitre 3 (régions d'homozygotie du génome) ont été réalisées sur l'ensemble des données disponibles. La sélection des cas et témoins est identique à celle décrite précédemment ainsi que le recueil des données épidémiologiques effectué. L'ensemble des cas et témoins ont fourni

un prélèvement salivaire et ont été génotypés avec la même puce que les autres 7 autres études d'EPITHYR.

6. L'étude de Tchernobyl

L'étude Tchernobyl est une étude de cohorte en population générale menée dans les zones les plus contaminées du Bélarus pour évaluer le risque de cancer de la thyroïde après une exposition à l'iode radioactif dans l'enfance (Cardis et al. 2005; Stezhko et al. 2004).

L'étude cas-témoin sur laquelle j'ai travaillée dans le cadre de cette thèse a été menée sur un sous-groupe de sujets pour lesquels un prélèvement sanguin était disponible.

Sélection des cas. La cohorte totale inclut 114 537 personnes nées entre le 26 avril 1968 et le 26 avril 1986 et dont la dose de radioactivité reçue à la thyroïde a été mesurée en 1986, peu après l'accident. Ces personnes résidaient alors dans l'une des zones hautement contaminées du Bélarus ou de l'Ukraine.

Les cas ont été diagnostiqués dans les 6 à 12 ans qui ont suivi l'accident avec un CPT vérifié histologiquement et confirmé par le panel international de pathologistes. Deux tiers des cas ont développé un CPT avant l'âge de 15 ans et les autres avant l'âge de 25 ans. Pour plus de 60 % des cas, la latence (temps entre l'exposition et le diagnostic) était inférieure à 10 ans.

Sélection des témoins. Les témoins ont été appariés aux cas en fonction de l'âge (dans l'année pour ceux qui avaient 18 mois ou plus au moment de l'accident ; dans les 6 mois pour ceux qui avaient entre 12 et 18 mois et dans le mois pour ceux qui avaient moins de 12 mois), du sexe et du type d'établissement.

Les sujets inclus dans l'étude génétique comprenaient **83 cas de CPT et 324 témoins** appariés qui ont consenti à donner un échantillon de sang.

7. La cohorte E3N

L'étude E3N est une cohorte prospective qui a inclus 98 995 femmes en 1990. Ces femmes éligibles devaient résider en France, être nées entre 1925 et 1950 et avoir souscrit à la MGEN,

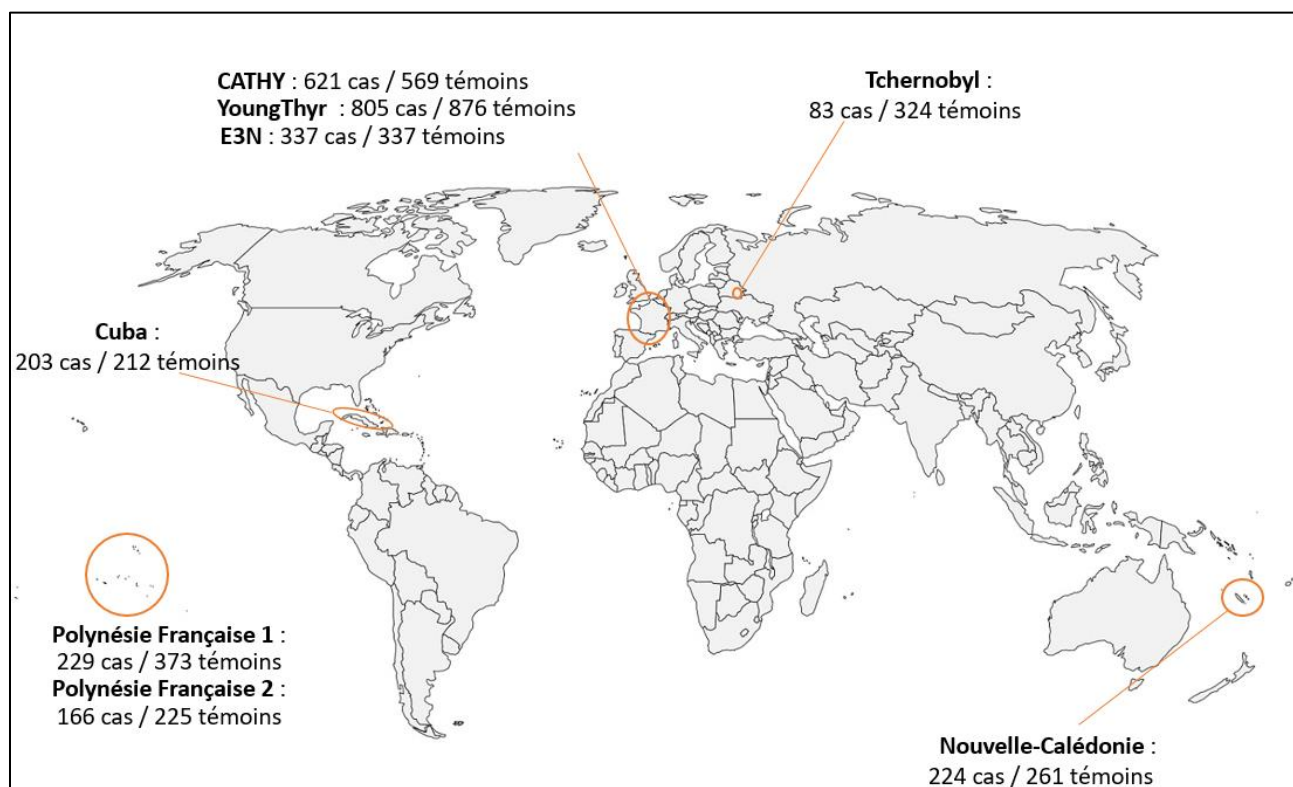
un régime national d'assurance maladie pour les enseignants et les collègues de travail (Clavel-Chapelon et al. 2010). Les participantes, qui ont donné leur consentement éclairé écrit pour un suivi médical externe par l'intermédiaire de l'assureur maladie, ont rempli des questionnaires auto-administrés bisannuels, demandant des informations sur les caractéristiques du mode de vie et les conditions de santé, les antécédents médicaux et chirurgicaux, la prise de médicaments, le passé gynécologique, le statut tabagique, les caractéristiques staturo-pondérales, l'activité physique, les antécédents familiaux de cancer, le niveau socio-économique ... Les participants ont été invités à signaler tout diagnostic de cancer sur chaque questionnaire Q1 à Q8, et les diagnostics de cancer ont ensuite été vérifiés et détaillés en demandant des informations supplémentaires aux médecins des sujets. Le compte-rendu d'anatomopathologie a été recueilli pour tous les cas incidents identifiés, sauf deux. Pour ces deux cas, l'histologie du CDT a été confirmée par le clinicien en charge du cas correspondant. Ainsi tous les cas de la cohorte ayant développé un CDT primitif (vésiculaire ou papillaire) ont été inclus dans le GWAS, excluant ainsi les 19 cas de cancers de la thyroïde anaplasique ou médullaire identifiés dans la cohorte.

Les témoins ont été tirés au sort parmi les sujets n'ayant pas eu de cancer et ont été appariés individuellement aux cas sur l'âge. Au total **337 cas et 337 témoins ont été inclus** dans l'étude cas-témoin d'EPITHYR, dont 279 cas et 289 témoins avaient un prélèvement salivaire ou sanguin exploitable.

Les protocoles d'études et questionnaires pour les études NC et CATHY (PI : Pascal Guénel), YoungThyr, Polynésie Française et CUBA (PI : Florent de Vathaire) sont très proches. Les données épidémiologiques disponibles concernent les facteurs hormonaux et reproductifs, les facteurs anthropométriques, les habitudes alimentaires ou liés au mode de vie, les antécédents personnels et familiaux de pathologie de la thyroïde, l'historique résidentiel, l'historique professionnel ... Ces facteurs ont été largement explorés et analysés précédemment dans les études de Nouvelle-Calédonie (Guignard et al. 2007; Leux et al. 2012; Paquet et al. 2020; Truong et al. 2005, 2010), Polynésie Française (Brindel et al. 2009; Cléro et al. 2010; Drozdovitch et al. 2019; 0; C. Xhaard et al. 2014; Xhaard et al. 2015; Zidane et al. 2019),

CATHY (Cordina-Duverger et al. 2017), YoungThyr (Constance Xhaard et al. 2014; Xhaard et al. 2015), Cuba (Juan J Lence-Anta et al. 2014; Turcios et al. 2015) et ne seront pas discutés dans ce manuscrit.

Figure 5 : Etudes participantes au consortium EPITHYR.



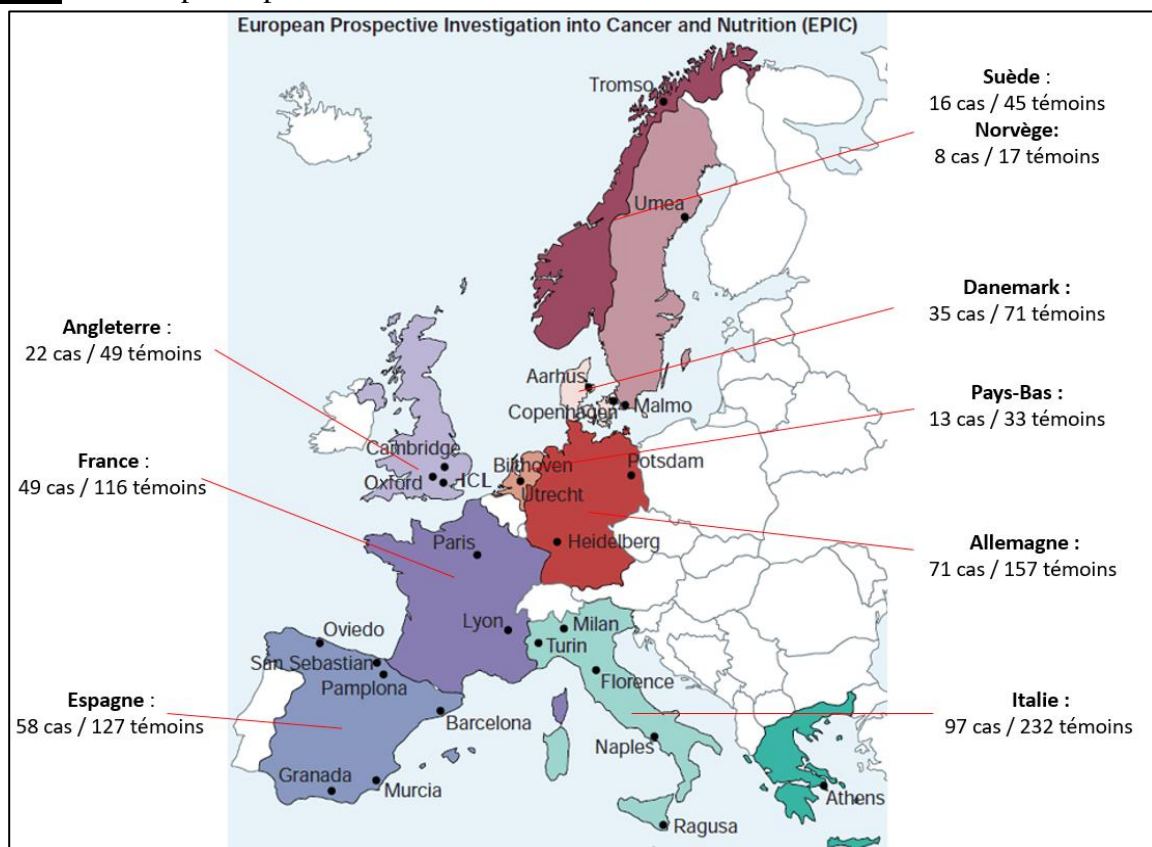
Les effectifs indiqués correspondent aux nombres de cas et témoins pour lesquels les données génétiques sont disponibles.

II. La cohorte EPIC

La cohorte EPIC est une étude prospective multicentrique, conçue pour étudier les associations entre les facteurs de risque nutritionnels, liés au style de vie, métaboliques et génétiques et le risque de cancer (<https://epic.iarc.fr/>, Kaaks, Slimani, and Riboli 1997; Riboli 1992). Elle a été initiée en 1992 dans 10 pays européens (Danemark, France, Allemagne, Grèce, Italie, Pays-Bas, Norvège, Espagne, Suède et Royaume-Uni) et inclut environ 370 000 femmes et 150 000 hommes, parmi lesquels environ 246 000 femmes et 140 000 hommes ont également

fourni un échantillon de sang (Riboli et al. 2002). Les questionnaires comprenaient des données sur le régime alimentaire, les antécédents reproductifs, l'utilisation d'hormones exogènes, le tabagisme et la consommation d'alcool, le niveau d'éducation, les antécédents professionnels, les variables anthropométriques, l'activité physique et les antécédents de certaines maladies. La population d'EPIC sur laquelle j'ai travaillé correspond à celle de l'étude cas-témoins nichée sur les CDT sur laquelle a été réalisée une étude sur les biomarqueurs de l'inflammation (Dossus et al. 2018). Les cas incidents de cancers de la thyroïde ont été identifiés à l'aide des registres régionaux du cancer dans la plupart des centres, sauf en France, en Allemagne, en Grèce et à Naples (Italie), où le suivi était basé sur à la fois les données de l'assurance maladie, les registres de cancer et le suivi actif des participants à l'étude. Chaque cas de CDT identifié a été apparié à 2 témoins chez les femmes et 3 témoins chez les hommes. Un projet de génotypage de cette collection coordonné par Thérèse Truong a été initié en 2019 (Figure 6).

Figure 6 : Etudes participantes au consortium EPIC.



Les effectifs indiqués en figure 6 correspondent au nombre de sujets inclus dans l'étude cas-témoins nichée sur les CDT pour laquelle des données de génotypes sont disponibles.

Données de génotypage

I. La puce OncoArray

Le génotypage des sujets d'EPITHYR et d'EPIC a été réalisé à l'aide de la puce à ADN OncoArray d'Illumina (Amos et al. 2017) au Centre National de Recherche en Génomique Humaine (CNRGH) entre 2016 et 2020, en plusieurs phases. Une première phase de génotypage a été réalisée sur les participants aux 7 études d'EPITHYR en 2016, dont le premier échantillon de la population de Polynésie française, l'échantillon Polynésie 2 a été génotypé en 2018 et les participants de l'étude cas-témoins nichée d'EPIC ont été génotypés en 2020.

La puce OncoArray a été conçue par le consortium GAME-ON, regroupant des consortia internationaux sur les cancers du poumon, du sein, de la prostate, du colon et ovaire avec l'hypothèse qu'il existe des mécanismes de carcinogénèse communs à ces cinq cancers qui sont les plus fréquents (Amos et al. 2017). La puce inclut ainsi environ 500 000 SNPs, dont 250 000 répartis le long du génome capturant le maximum de variation génétique. La moitié des SNPs de la puce a été sélectionnée pour leur capacité à marquer un groupe de SNPs ou même une région entière en déséquilibre de liaison (DL) avec eux. Ces 250 000 SNPs « marqueurs » permettent de prédire (ou “imputer”) le génotype des SNPs manquants sur le même haplotype. Les SNPs supplémentaires sont des SNPs d'intérêt pour les cinq cancers ciblés (régions de fine mapping, SNPs associés à des facteurs de risque de ces cancers, etc).

Pour l'étude des cancers thyroïdiens, notre équipe en collaboration avec le CNRGH a sélectionné des SNPs localisés dans des gènes impliqués dans des voies métaboliques spécifiques à la signalisation cellulaire, à la synthèse des hormones thyroïdiennes d'après la base de données KEGG (Kanehisa and Goto 2000) ainsi que dans des gènes candidats ou dans des régions identifiées dans les précédents GWAS sur le CDT (loci 2q35, 8p12, 9q22.33 et 14q13.3). Au total, 14 407 variants supplémentaires ayant une fréquence allélique mineure supérieure à 2% dans la population européenne du projet 1000Genomes (Clarke et al. 2012) ont

été sélectionnés pour être ajouté sur la puce OncoArray mais seules les sondes spécifiques à 13 759 d'entre eux ont pu être conçues et ajoutées sur la puce.

II. Contrôle qualité des données de génotypage

Les différentes étapes du contrôle qualité des données de génotypage que j'ai effectuées sur les données sont détaillées dans le diagramme Figure 10. Ce contrôle qualité a été appliqué aux données génotypées d'EPITHYR et d'EPIC. Nous avons appliqué les mêmes seuils d'exclusion de SNPs et d'individus que ceux utilisés par le consortium OncoArray pour l'étude d'autres cancers. Dans une première étape, j'ai exclu les variants dupliqués, ceux ayant un taux de données manquantes >5%, les SNPs monomorphes et ceux ayant une mauvaise qualité de génotypage (listes de SNPs fournies par le consortium OncoArray sur la base des clusters plots obtenus sur les données de consortia). J'ai ensuite exclu les individus ayant plus de 5% de données manquantes, les individus ayant un sexe génétique discordant par rapport au sexe déclaré et les doublons intra ou inter-études (individus inclus dans deux études distinctes ou pour lesquels 2 échantillons d'ADN ont été génotypés).

La dernière étape a consisté à exclure les variants monomorphes et ayant une fréquence allélique mineure (FAM) inférieure à 1% et ne respectant pas l'équilibre d'Hardy Weinberg. J'ai également exclu les individus ayant des taux d'hétérozygotie anormaux par rapport aux taux médians des populations européenne ou asiatique de référence, selon l'analyse.

Aucun individu n'a été exclu sur la base de l'apparentement pour les analyses GWAS (chapitre 1) et de fine-mapping (chapitre 2) mais les individus apparentés au premier degré ont été exclus dans chacune des populations étudiées pour les analyses ROH (chapitre 3). Les populations analysées étant sensiblement différentes selon le type d'analyse, le détail des effectifs du contrôle qualité sera précisé en début de chaque chapitre.

III. Imputation des SNPs non génotypés

Des millions de SNPs répartis sur le génome entier sont aujourd'hui répertoriés dans la base de données publique dbSNP (Smigielski et al. 2000), mais la puce OncoArray n'en cible

qu'environ 500 000 SNPs. Toutefois, à partir des SNPs génotypés, il est possible de prédire la valeur des SNPs non génotypés grâce au DL entre les SNPs d'une même région. L'imputation des génotypes des SNPs non présents sur la puce est réalisé à l'aide de base de données décrivant les variants de génomes entiers de référence de plusieurs groupes de populations (européens, asiatiques, africains, amérindiens etc), comme ceux du panel de référence 1000Genomes (Howie, Marchini, and Stephens 2011). Imputer des SNPs manquants permet donc d'augmenter la couverture génétique dans l'échantillon étudié et de caractériser plus finement des régions chromosomiques d'intérêt. Cela permet également de combiner les données de génotypage provenant de différentes études et générées avec des puces différentes afin de pouvoir réaliser des méta-analyses.

Principe de l'imputation

Les données génotypiques que l'on obtient grâce à une puce à ADN comme la puce OncoArray ne permettent pas de savoir quels sont les allèles présents sur un même chromosome. Cependant, pour pouvoir prédire la valeur des SNPs non génotypés, nous devons connaître la combinaison allélique des SNPs de chaque sujet sur un même chromosome (ou haplotype). Une étape de phasage est donc nécessaire pour reconstruire cette information à partir des SNPs génotypés dans notre population, du DL et des taux de recombinaison à partir d'haplotypes connus de populations de référence. Cette étape peut se faire avant l'imputation ce qui permet d'optimiser le temps de calcul lorsqu'on impute un grand jeu de données. Cette méthode d'imputation avec pré-phasage a été appliquée au jeu de données total d'EPITHYR. Aussi, le phasage et l'imputation peuvent être réalisées simultanément ce qui permet d'augmenter la qualité et la précision de l'imputation (Herzig et al. 2018). J'ai ainsi imputé les données en une seule étape sur les régions sur lesquelles j'ai réalisé l'étude de cartographie fine. Le phasage peut être réalisé par différents logiciels tels que SHAPEIT (O'Connell et al. 2014) ou beagle (Browning and Browning 2007) qui se basent sur des modèles de Markov cachés (MMC). Le principe de l'imputation est représenté sur la Figure 7.

Figure 7 : Principe de l'imputation. (Illustration de (Marchini and Howie 2010)).

a) L'échantillon d'étude représente la base de données génétiques d'intérêt avec des génotypes manquants, b) Groupe d'haplotypes de la population de référence, c) L'haplotype de la population d'étude est estimé à partir de des haplotypes des différents panels de référence, d) L'haplotype de référence est utilisé pour imputer les génotypes manquants dans l'échantillon d'étude.

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

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

Une mesure de l'évaluation de l'incertitude de l'imputation est donnée par les différents logiciels. Nous avons utilisé le logiciel IMPUTE2 pour l'imputation, qui donne une valeur INFO correspondant 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. Un info score =1 signifie que le variant a été imputé avec une précision parfaite tandis qu'un info score égale ou proche de 0 indique une qualité d'imputation très faible.

Panel de référence

Nous avons utilisé comme panel de référence la Phase 3 du projet 1000Genomes (Howie et al. 2011), qui répertorie la majorité des variants génétiques ayant une fréquence supérieure à 1 % à partir du séquençage du génome de sujets de différentes populations. Le projet 1000Genomes a été rendu possible grâce au développement des technologies de séquençage dont le coût a diminué de façon drastique. La base de données de référence 1000Genomes Phase3 est composée des haplotypes de 2 504 sujets de différentes origines ethniques, avec 503 sujets d'origine européenne, 347 sujets d'origine américaine, 489 sujets d'origine d'Asie du sud, 504 sujets d'origine d'Asie de l'est et 661 sujets d'origine africaine. Ce panel capture un large éventail de la diversité génétique humaine.

Pour les données génétiques de la population européenne, nous avons imputé les SNPs du génome non génotypés en utilisant le logiciel IMPUTE2 avec comme panel de référence, les génotypes de l'ensemble des populations du projet 1000Genomes (Howie et al. 2011). La même stratégie a été appliquée pour l'imputation des données génétiques des populations asiatiques, mais afin d'améliorer la précision de l'imputation, nous avons ajouté comme panel de référence supplémentaire les 26 individus de la population océanienne du panel de référence « Simon's diversity project » (<https://www.simonsfoundation.org/life-sciences/simons-genome-diversity-project/>). Les 26 individus Océaniens d'HGDP sont originaires de l'île de Bougainville en Papouasie-Nouvelle-Guinée, située en Mélanésie.

Contrôle qualité post-imputation par SNP

Plus les SNPs imputés ont un déséquilibre de liaison élevé avec les SNPs génotypés, plus leur **qualité d'imputation (info score)** est élevée. Pour les analyses présentées dans ce mémoire, étant dans une approche exploratoire, j'ai conservé tous les SNPs imputés ayant un info score supérieur ou égal à 0,3. Cependant lors de l'interprétation des résultats des analyses GWAS et

de *fine-mapping*, j'ai systématiquement vérifié que les variants significativement associés au CDT avaient une bonne qualité d'imputation (avec un info score supérieur à 0,8).

IV. Définition des groupes ethniques

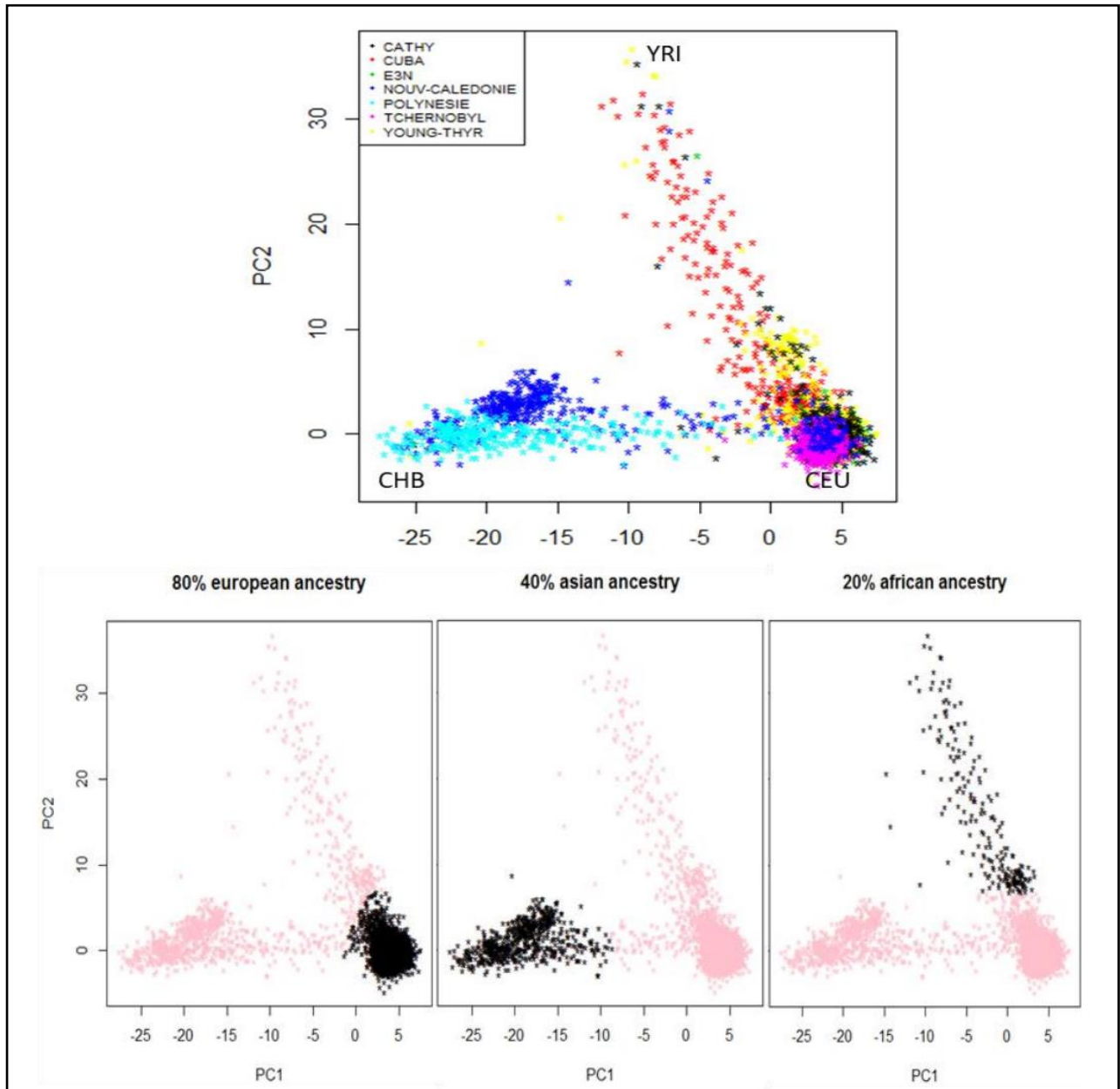
EPITHYR

La population d'EPITHYR est composée de populations d'origines très diverses. Pour inférer les coefficients de métissage ou coefficients d'ascendance, j'ai réalisé une **analyse en composantes principales (ACP)**. Cette méthode statistique multivariée consiste à transformer des variables liées entre elles (dites « corrélées ») en nouvelles variables décorrélatées les unes des autres (ou appelées composantes principales) à partir d'une matrice de données. Les composantes principales (CP) sont calculées de manière à ce que la première composante principale rende compte de la plus grande variation possible des données en une seule composante, suivie de la deuxième composante et ainsi de suite.

Pour définir les différents groupes ethniques, nous avons utilisé la même méthodologie que celle employée par le consortium Oncoarray Game-ON (Amos et al. 2017). L'ACP est d'abord réalisée sur l'ensemble des individus retenus pour l'analyse après les différentes étapes de contrôle qualité décrites précédemment. Le programme Fastpop (Li et al. 2016) calcule un pourcentage d'ascendance (ou « ancestry ») pour chaque individu de notre échantillon en utilisant les scores ACP d'environ 47 000 individus du consortium GAME-ON et ceux des individus européens (CEU), asiatiques (JPT + CHB) et africains (YRI) du panel de référence HapMap préalablement calculés sur un panel de 2318 SNPs et intégrés dans le programme (Li et al. 2016). Ces scores reflètent la distance « physique » de chaque individu par rapport aux centroïdes de chaque population des échantillons **HapMap** d'origine européenne (CEU), africaine (YRI) et asiatique (CHB). La figure 8 montre la répartition des individus d'EPITHYR en fonction de leur appartenance aux différents groupes ethniques identifiés par Fastpop. Sont

ainsi définis comme Européens tous les individus ayant un **score européen > 80%**, comme Asiatiques tous les individus ayant un **score asiatique > 40%** et comme Africains ceux ayant un **score africain > 20%**. Les seuils sont moins stricts pour les populations asiatiques et africaines pour des raisons d'effectifs de population (Figure 8).

Figure 8: Structure de la population d'EPITHYR.



Proportion d'origine ethnique attribuée utilisant le premier et le second score CP. Les axes X et Y indiquent le premier et le second score CP générés par FastPop. Les centroïdes de chaque

population ont été calculés pour les échantillons HapMap d'origine européenne (CEU), africaine (YRI) et asiatique (CHB).

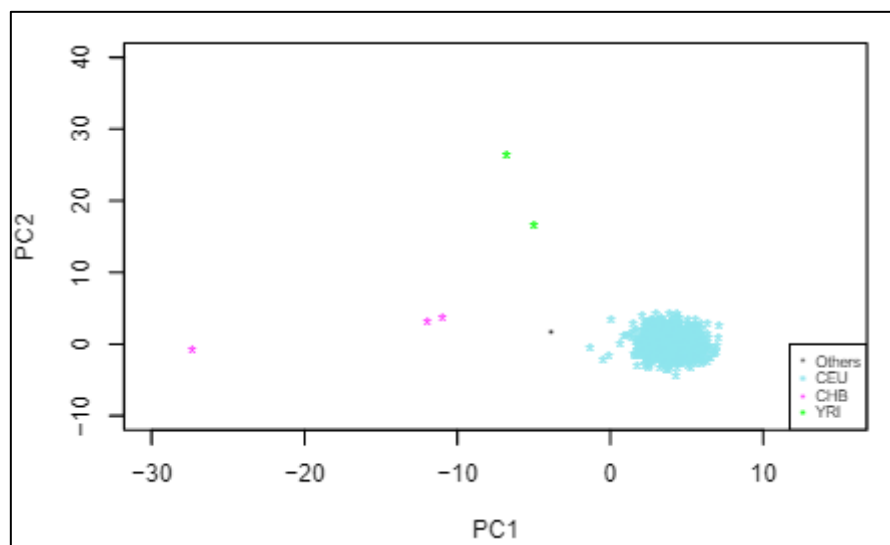
Etant donné le peu d'individus d'ascendance africaine et l'hétérogénéité de cette population dans les études d'EPITHYR, ce dernier groupe de population n'a pas été pris en compte dans les analyses effectuées dans le cadre de cette thèse.

Les effectifs des populations seront détaillés dans chacun des chapitres car la population d'étude d'EPITHYR a évolué durant la période de ma thèse.

EPIC

Une ACP avec Fastpop a été réalisée en appliquant les mêmes critères de définition des ethnies que pour EPITHYR (Figure 9).

Figure 9 : ACP de la population d'EPIC.



Proportion d'origine ethnique attribuée utilisant le premier et le second score CP. Les axes X et Y indiquent le premier et le second score CP générés par FastPop. Les points bleus correspondent aux individus considérés comme Européens, les trois points roses sont les individus considérés comme Asiatiques et en vert les individus considérés comme Africains.

Cette population est beaucoup plus homogène, et seuls six individus considérés comme non européens ont été exclus des analyses.

Les données disponibles pour les analyses présentées dans les différents chapitres de ce mémoire ont évolué au cours des quatre années de thèse. Ainsi pour l'étude polynésienne, de nouveaux individus ont été recrutés et génotypés en 2019 et ont pu être exploitées pour la recherche de variants récessifs (chapitre 3). Les données de génotypage d'EPIC n'ont été disponibles qu'en 2019 et ont également été exploitées dans le chapitre 3. De plus, les groupes ethniques ont été définis selon différents critères selon les analyses effectuées. Pour plus de clarté, j'ai préféré décrire les populations analysées au début de chaque chapitre de la thèse.

Chapitre 1

Etude d'association génome entier dans la population multi-ethnique du consortium EPITHYR

I. Introduction

Depuis 2009, six GWAS et une méta-analyse ont été réalisés sur le CDT (tableau supplémentaire 1). La première GWAS (Gudmundsson et al. 2009) a été menée en 2009 dans une population islandaise incluant 378 cas de CDT et 37 196 témoins. Cette étude a mis en évidence deux loci majeurs sur les chromosomes 9 et 14. Les signaux les plus forts concernent les SNPs rs965513 (OR = 1,75, IC95% : 1,59–2,05, p -valeur = $1,7 \times 10^{-27}$) situé sur le locus 9q22.33 en 5' de *FOXEL* et rs944289 (OR = 1,37, IC95% : 1,24–1,52), p -valeur = $1,0 \times 10^{-9}$) situé en 5' de *NKX2-1* sur le locus 14q13.3. Comme *FOXEL*, *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 seconde étude GWAS a été menée en 2010 dans 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). Cette étude a permis d'identifier une association entre le risque de CDT et quatre SNPs de la région 9q22.33 dont rs965513 (OR = 1,65, IC95% : 1,43–1,91), p -valeur = $5,9 \times 10^{-9}$). Cette étude ne réplique pas le SNP rs944289 au locus 14q13.3 précédemment rapporté pour les CPT sporadiques (Takahashi et al. 2010).

En 2012, Gudmundsson et coll. ont conduit une nouvelle étude GWAS sur les taux de TSH (thyroéostimuline) circulant dans le sang chez 27 758 individus islandais et identifient 22 loci significativement associés au CDT (Gudmundsson et al. 2012). Dans la même étude, les 22 SNPs identifiés ont ensuite été testés dans un échantillon indépendant comprenant 561 cas de CDT et 40 013 témoins islandais ainsi que 595 cas de CDT et 2 604 témoins non islandais mais toujours d'ascendance européenne. En plus des loci 9q22.33 et 14q13.3 connus, d'autres SNPs ont été identifiés : rs966423 dans *DIRC3* au locus 2q35 (OR = 1,26, IC95% : 1,11–1,43, p -valeur = $3,8 \times 10^{-4}$) ; rs2439302 dans *NRG1* au locus 8p12 (OR = 1,41, IC95% : 1,23–1,62, p -valeur = $1,3 \times 10^{-6}$) et rs116909374 près de *NKX2-1* au locus 14q13.3 (OR = 2,03, IC95% : 1,54–2,67), p -valeur = $5,4 \times 10^{-7}$).

En 2013, une quatrième étude GWAS (Köhler et al. 2013) a été menée dans une population italienne. La phase exploratoire a inclus 690 cas de CDT et 497 témoins, et la phase de réplication des SNPs significatifs a inclus 2 958 cas and 3 727 témoins. En plus de la réplication de l'association de la région 9q22.33, cette étude a identifié de nouveaux allèles à risque de CDT près des gènes *DICR3* (2q35), *IMMP2L* (7q21), *RARRES1* (3q25) et *SNAPC4/CARD9* (9q34).

La cinquième étude GWAS, réalisée sur une population espagnole a inclus 398 cas de CDT et 502 témoins dans la phase de découverte (Mancikova et al. 2015), et 1 422 cas et 1 908 témoins provenant de trois études cas-témoins espagnoles et d'une étude cas-témoins italienne dans la phase de réplication. De plus, cette étude a identifié de nouveaux SNPs en 9q22.33, 6q14 et 10q26 (tableau supplémentaire 1) en plus de la réplication des variants rs965513 et rs1867277 du locus 9q22.33 et rs944289 du locus 14q13.3 précédemment décrits (Mancikova et al. 2015). En 2017, la première étude GWAS menée dans une population non-européenne a inclus 470 cas de CDT et 8 279 témoins coréens (Son et al. 2017). En plus de la réplication du SNP rs2439302 au locus 8p12, une nouvelle association a été observée pour le SNP rs6996585 (OR= 1,39, IC95% : 0,29-0,23, p -valeur = $1,08 \times 10^{-10}$) situé dans un intron du gène *NRG1*. D'autres variants précédemment identifiés dans les GWAS européennes ont été répliqués, dont rs12990503, rs2439302, rs72753537 confirmant le rôle des loci 2q35, 9q22.33, 14q13.3. Par ailleurs ce dernier GWAS a permis d'identifier sept nouveaux loci contenant les gènes *VAV3*

(1p13.3), *PCNXL2* (1q42.2), *INSR* (19p13.2), *MRSB3* (12q14.3), *FHIT* (3p14.2), *SEPT11* (4q21.1) et *SLC24A6* (12q24.13).

En 2017, une méta-analyse des études européennes précédemment décrites portant sur 3 001 cas et 287 550 témoins confirme les associations entre le risque de CDT et les loci 2q35, 8p12, 9q22.33 et 14q13.3 et fait état de cinq nouveaux loci de susceptibilité en 1q42.2, 3q26.2, 5q22.1, 10q24.33 et 15q22.33 qui n'ont pas encore été répliqués dans un échantillon indépendant (Figure 4, tableau supplémentaire 1) (Gudmundsson et al. 2017).

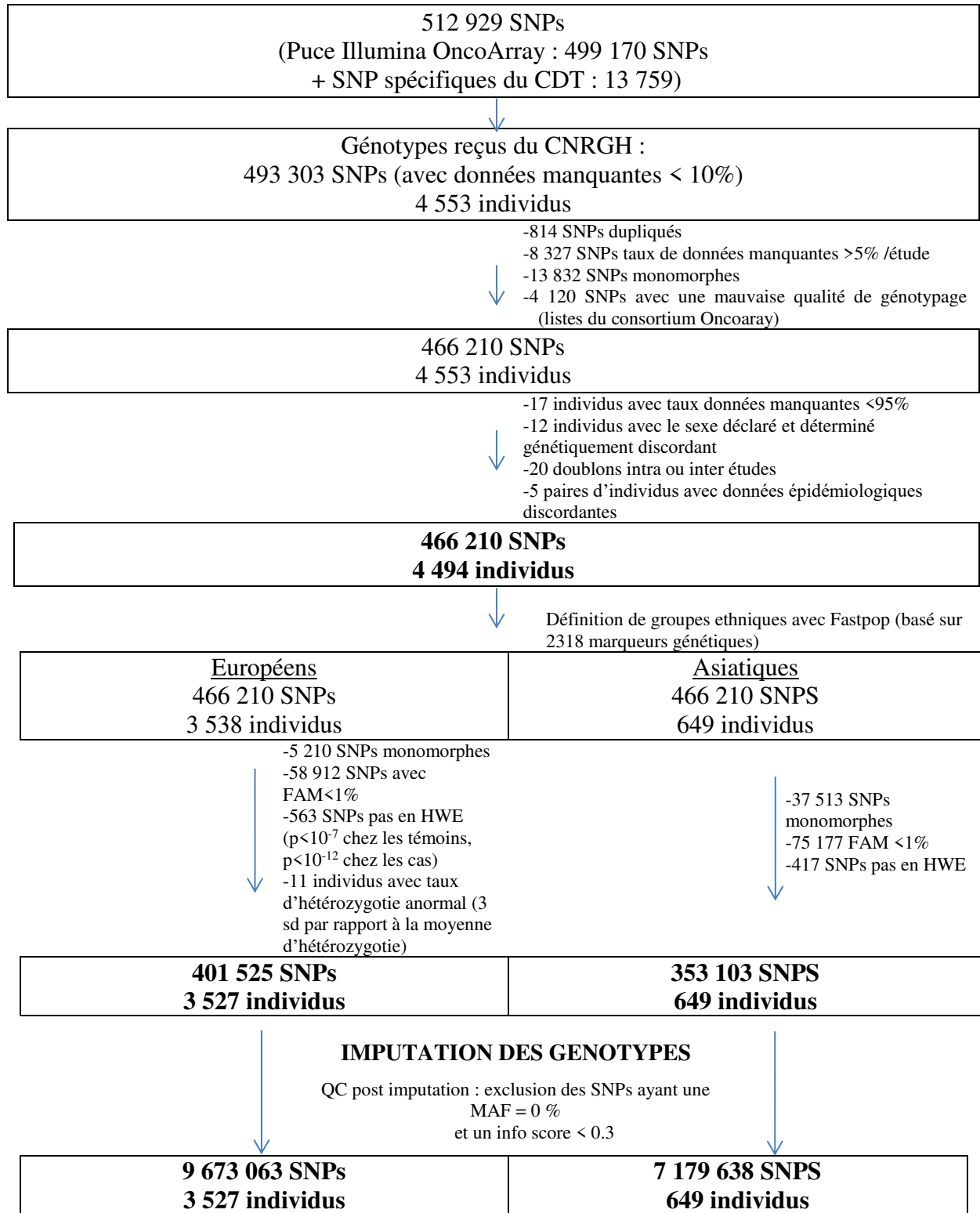
Cependant, l'ensemble des loci de susceptibilité identifiés n'explique qu'une partie de la prédisposition génétique au CDT (Liyanarachchi et al. 2013).

Afin d'évaluer la contribution des différents loci de prédisposition au CDT décrits précédemment et d'identifier de nouveaux loci, nous avons conduit une GWAS dans les populations européennes et océaniques d'EPITHYR. Cette étude a fait l'objet d'un article publié au sein l'International Journal of Cancer et est annexé à cette thèse.

II. Données

La figure 10 détaille le contrôle qualité de la population d'EPITHYR. L'analyse GWAS a été menée sur 1554 cas et 1973 témoins européens et 301 cas et 348 témoins océaniques. Au total, respectivement 9 673 063 SNPs et 7 179 638 SNPs génotypés ou imputés ont été analysés chez Européens et les Océaniques.

Figure 10 : Etapes du contrôle qualité des données génotypage d'EPITHYR et sélection des SNPs et des individus retenus pour les analyses.

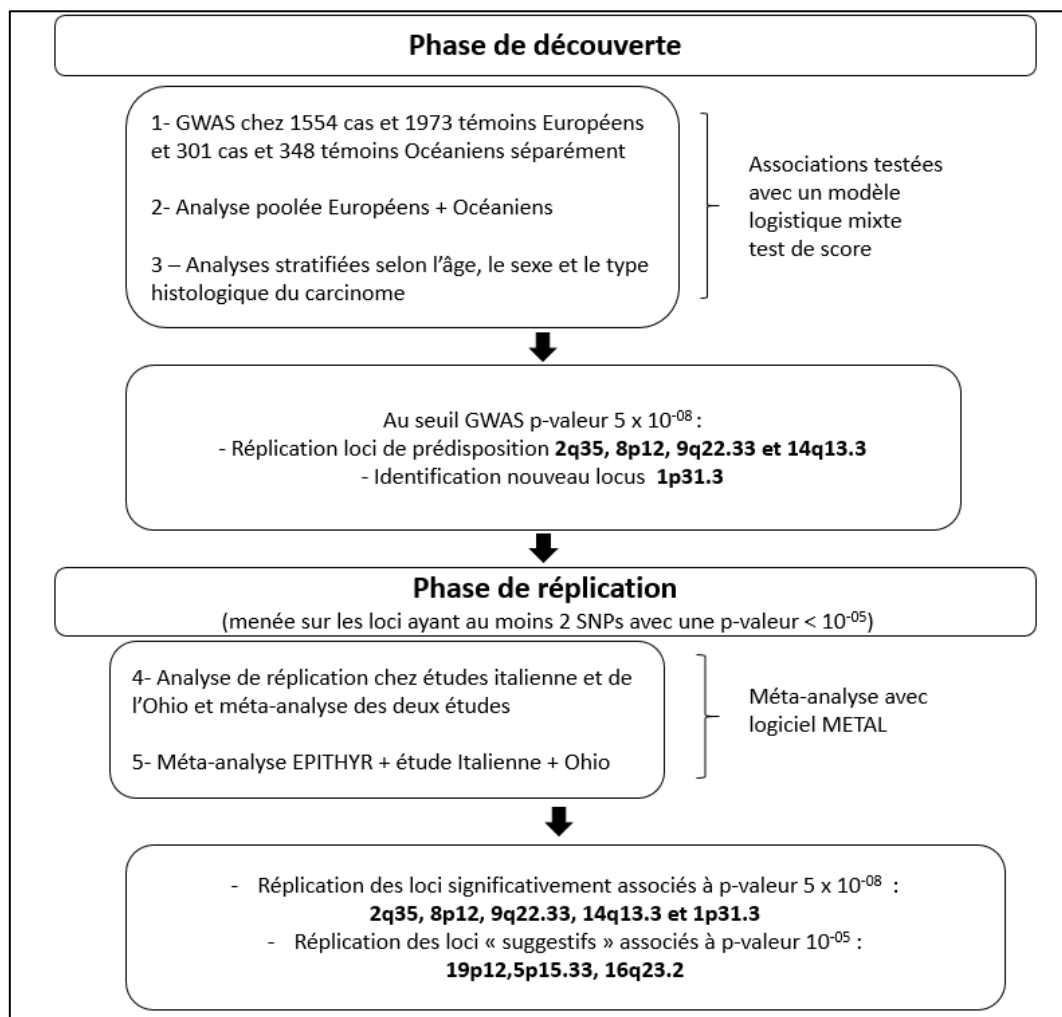


Pour la phase de réplique, nous avons utilisé les données du GWAS italien publié en 2013 (Köhler et al. 2013) et de l'étude américaine qui était déjà incluse dans la méta-analyse de GWAS (Gudmundsson et al. 2017). L'échantillon italien dont nous disposons comprend 649 cas de CDT recrutés consécutivement au département d'endocrinologie de l'hôpital universitaire de Pise. Les témoins (n=431) sont des sujets sains recrutés parmi les donneurs de sang et les personnes travaillant au sein de l'hôpital. L'étude américaine est une étude cas-témoins incluant des sujets se déclarant d'origine européenne, dont 1 580 cas de CDT confirmés histologiquement et 1 628 témoins de la région centrale de l'Ohio.

III. Stratégie d'analyse

Les différentes étapes effectuées pour l'étude GWAS sont détaillées dans la figure 11.

Figure 11 : Stratégie d'analyse du GWAS.



IV. Méthodes

1. Test d'association avec un modèle logistique mixte

Dans les études cas-témoins, il est usuel d'estimer les associations entre chacun des SNPs et le phénotype à l'aide d'un modèle logistique qui fait l'hypothèse que les données sont indépendantes et identiquement distribuées. Dans notre cas, afin de maximiser le nombre de sujets inclus dans les analyses, nous avons choisi de ne pas exclure les sujets apparentés. A noter qu'il y avait respectivement 16 individus européens et 36 individus océaniens apparentés au 1^{er} ou 2nd degré. Nous avons également choisi d'analyser les Mélanésiens et les Polynésiens

ensemble malgré l'hétérogénéité observée entre ces deux groupes (Figure 15), les effectifs étant trop faibles pour les analyser séparément.

Afin de prendre en compte à la fois l'apparement des individus et la stratification de la population induisant une structure de corrélation entre les individus, nous avons utilisé un modèle logistique mixte, avec un effet aléatoire qui dépend de la matrice de corrélation génétique des individus.

Ainsi, si l'on considère Y , un phénotype binaire mesuré sur n individus, X la matrice des covariables sur lesquelles on souhaite ajuster le modèle, G_i le génotype du SNP testé pour l'individu i , le modèle peut s'écrire :

$$\text{logit } P(Y_i = 1) = X_i\beta + G_i\gamma + \mu_i$$

avec :

- β et γ correspondants à des effets fixes. β est le vecteur des effets associés aux covariables et γ l'effet du SNP que l'on cherche à estimer,
- X_i le vecteur des covariables pour l'individu i ,
- μ_i l'effet aléatoire associé à l'individu i ; le vecteur $\mu = (\mu_1 \dots \mu_n)$ suit une loi $N(0, \tau K)$, avec K la matrice de corrélation génétique des individus.

Le temps de calcul de la vraisemblance de ce modèle étant trop long pour permettre une analyse sur le génome entier, nous avons dans un premier temps appliqué la méthode proposée par Chen et coll. (Chen et al. 2016) pour estimer rapidement les p -valeurs d'association de chaque SNP du GWAS avec un test du score (GMMAT) mais cette méthode ne permet pas d'estimer les effets associés. Dans un second temps, nous avons estimé les effets des SNPs des régions d'intérêt en utilisant une deuxième méthode proposée par Chen et coll. (Chen et al. 2016) basée sur la quasi-vraisemblance pénalisée (PQL) mais dont le temps de calcul est plus long que la méthode GMMAT. Les associations sont alors testées par un test de Wald qui est équivalent au test du score.

Nous avons utilisé le package R de GASTON (<https://cran.r-project.org/web/packages/gaston/index.html>) pour réaliser les modèles logistiques mixtes. Nous avons d'abord testé les associations séparément chez les Européens et les Océaniens puis

nous avons réalisé une analyse poolée en groupant les deux populations. Le seuil traditionnel de Bonferroni pour les GWAS (p -valeur $< 5 \times 10^{-8}$) a été utilisé pour évaluer la significativité des associations.

Pour l'étude des régions d'intérêt où une association est suggérée (signal d'association p -valeur $< 10^{-05}$), les associations ont été réestimées en utilisant la PQL pour obtenir une estimation des effets des SNPs (Figure 11). Tous les tests d'association ont été ajustés sur l'âge, le sexe et l'étude.

2. QQ plot

Les diagrammes quantile-quantile (QQ-plot) des p -valeurs d'association permettent de comparer la distribution des p -valeurs observées pour chaque SNP à la distribution des p -valeurs attendues sous l'hypothèse nulle H_0 d'absence d'association (i.e sous H_0 les p -valeurs suivent une loi uniforme sur $[0,1]$). Une trop grande déviation par rapport à la première bissectrice indique un biais de stratification de la population. Le QQ-plot s'accompagne d'un facteur d'inflation génomique correspondant au ratio entre la médiane de la distribution des p -valeurs observée et celle de la distribution attendue sous H_0 . Ce facteur permet de quantifier le taux de résultats faux positifs. Une valeur proche de 1 indique que le modèle est correctement ajusté.

3. Méta-analyse

Dans la méta-analyse des études italiennes et de l'Ohio ainsi que celles d'EPITHYR, seuls les SNPs situés dans les loci associés au risque de CDT à un seuil de p -valeur $< 1 \times 10^{-05}$ ont été analysés. Nous avons utilisé la méthode à effets fixes en pondérant l'effet observée dans chaque étude par l'inverse de la variance telle que mise en œuvre dans le logiciel METAL (Willer, Li, and Abecasis 2010).

4. Test d'hétérogénéité

L'hétérogénéité des associations entre les études ainsi qu'entre les groupes de population a été testée avec un test de Cochran. Il s'agit d'un test de Chi² à k-1 degrés de liberté où k est le nombre d'études.

5. Analyses stratifiées

Afin de mettre en évidence des associations spécifiques à certains groupes de la population, nous avons effectué des analyses stratifiées selon le sexe, la tranche d'âge (< 50, ≥ 50 ans), l'histologie (carcinome papillaire/vésiculaire) et la taille du carcinome pour les CPT (<10 mm / ≥10 mm).

V. Résultats

1. Analyses descriptives

Les sujets inclus dans la phase de découverte du GWAS sont décrits dans le tableau 3. Plus de 80 % des populations européenne et océanienne d'EPITHYR sont des femmes. La majorité des participants sont âgés de moins de 50 ans dans les deux groupes ethniques. La majorité des cas ont développé un CPT (91% des Européens et 84% des Océaniens). Chez les Européens, 64% des cas de CPT ont une taille supérieure à 10 mm. Chez les Océaniens, 54% des cas de CPT ont été diagnostiqués avec un microcarcinome (Tableau 3).

Tableau 3 : Descriptif des populations européennes et océaniques d'EPITHYR selon l'étude, l'âge, le sexe et le type histologique des carcinomes.

		Européens (N=3527)				Océaniens (N=649)			
		Cas		Témoins		Cas		Témoins	
		N=1554	%	N=1973	%	N=301	%	N=348	%
Etudes	Cathy	451	28,9	534	27,1	1	0,3	0	
	YoungThyr	637	41,0	661	33,5	3	1,0	0	
	E3N	277	17,8	287	14,5	0		0	
	Tchernobyl	66	4,2	304	15,4	0		0	
	Cuba	102	6,6	113	5,7	0		0	
	Nouvelle-Calédonie	21	1,3	70	3,5	183	60,8	161	46,3
	Polynésie Française	0	0,0	4	0,2	114	37,9	181	52,0
Age	<25	272	17,5	513	26,0	0		0	
	[25-50[	763	49,1	887	45,2	202	67,1	255	73,3
	≥50	519	33,4	573	29,0	99	32,9	93	26,7
Sexe	Femme	1279	82,3	1519	77,0	278	92,4	318	91,4
	Homme	275	17,7	454	23,0	23	7,6	30	8,6
Type histologique	Vésiculaire	137	8,8			46	16,4		
	Papillaire	1417	91,2			253	83,9		
	< 10 mm	506	35,7			136	53,6		
	≥ 10 mm	904	63,8			107	42,1		

2. Réplication des loci de prédisposition connus

La figure 12 montre les résultats du GWAS chez les Européens et chez les Océaniens. Il n'y avait aucune indication d'inflation génomique dans les deux GWAS ($\lambda = 1,02$ chez les Européens et $\lambda = 1,04$ chez les Océaniens), ce qui suggère l'absence de confusion due à une structure de population hétérogène (Figure supplémentaire 1).

Aucune association significative n'a été observée chez les Océaniens. Chez les **Européens**, parmi les cinq régions significativement associées aux CDT au seuil GWAS p -

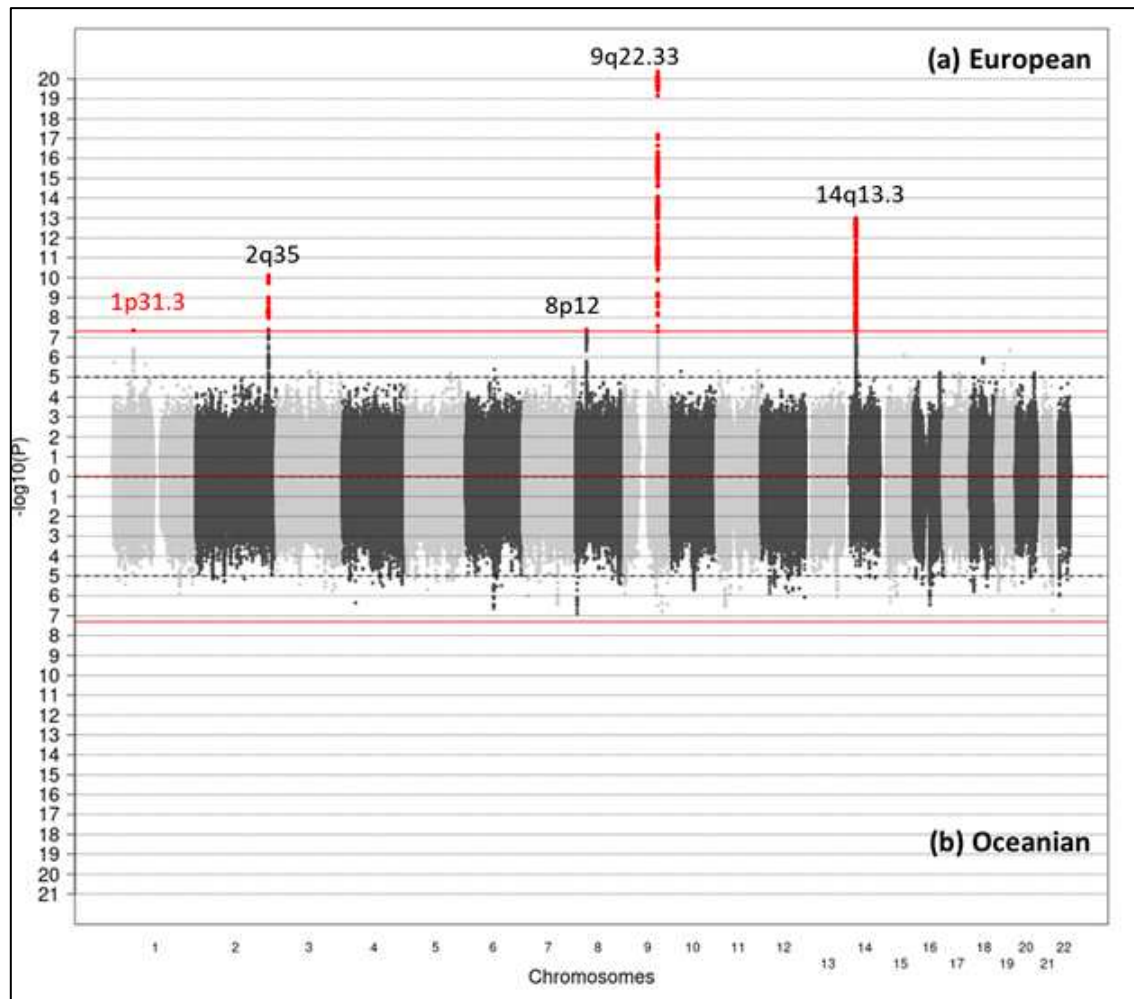
valeur $\leq 5 \times 10^{-8}$ (Figure 12), quatre loci (**2q35**, **8p12**, **9q22.33** et **14q13.3**) avaient déjà été rapportées dans les précédentes études GWAS.

Les SNPs mis en évidence dans les précédents GWAS sont répliqués avec des associations allant dans le même sens chez les Européens et les Océaniens d'EPITHYR (Tableau 4). Ces SNPs sont fortement corrélés aux SNPs les plus significatifs à ces loci chez les Européens d'EPITHYR (Tableau 4), mis à part le SNP rs142450470 au locus 8p12. Ce SNP est le SNP le plus fortement associé au CDT dans EPITHYR chez les Européens mais n'est que modérément corrélé au SNP rs2439302 rapporté dans les autres GWAS (Gudmundsson et al. 2009, 2012) ($r^2 = 0,45$) (Tableau 4).

L'analyse conditionnelle du SNP rs2439302 ajusté sur rs142450470 et inversement, du SNP rs142450470 ajusté sur rs2439302 diminue fortement leur association avec le CDT ($OR_{rs2439302} = 1,32$, p -valeur = $1,3 \times 10^{-4}$ et $OR_{rs142450470} = 1,01$, p -valeur = 0,97) suggérant qu'il existe un signal d'association unique au locus 8p12 (r^2 1000Genomes entre ces deux SNPs = 0,67).

L'analyse par type histologique chez les Européens montre que l'association avec le SNP rs1588635 en 9q22.33 est plus forte pour les CPT, en particulier pour les carcinomes de taille >10 mm, et l'association avec rs142450470 en 8p12 est spécifique aux CPT. L'analyse GWAS stratifiée sur la taille du carcinome montre de façon intéressante que les loci connus restent significatifs uniquement pour les carcinomes de taille >10 mm (Figure supplémentaire 2).

Figure 12 : Miami plot des résultats de l'étude GWAS chez les Européens et chez les Océaniens.



(a) Européens : 9 673 063 SNPs provenant de 1 554 cas et 1 973 témoins, (b) Océaniens : 7 179 638 SNPs provenant de 301 cas et 348 témoins. L'axe des abscisses représente l'emplacement chromosomique et l'axe des ordonnées représente les $-\log_{10}(p\text{-valeurs})$ du modèle logistique mixte (test de score). Les lignes horizontales en pointillés indiquent le seuil de suggestivité ($p\text{-valeur} = 10^{-5}$), tandis que les lignes rouges indiquent un seuil de significativité (à $p\text{-valeur} = 5 \times 10^{-8}$). Les loci déjà connus sont indiqués en noir et le nouveau locus identifié est indiqué en rouge.

Tableau 4 : Association entre le CDT et les SNPs rapportés dans les précédentes GWAS aux loci 2q35, 8p12, 9q22.33 et 14q13.3 dans les populations européennes et océaniques d'EPITHYR.

locus	SNP	position (hg19)	AA	AR	Européens					Océaniques					Corrélation avec le SNP de la région ayant la plus petite <i>p</i> -valeur chez les Européens d'EPITHYR	
					FAR	OR	IC95%		<i>p</i>	FAR	OR	IC95%		<i>p</i>	SNP le plus associé	r ²
2q35	rs966423	218,310,340	C	T	0,53	0,79	0,72	0,88	7,0x10 ⁻⁶	0,05	0,65	0,42	1,01	0,06	rs3821098	0,48
	rs11693806*	218,292,158	C	G	0,66	0,70	0,62	0,78	1,2x10 ⁻¹⁰	0,10	0,57	0,40	0,80	1,4x10 ⁻³		0,98
8p12	rs2439302	32,432,369	C	G	0,50	1,19	1,08	1,32	4,8x10 ⁻⁴	0,37	1,16	0,92	1,46	0,20	rs142450470	0,45
	rs2466076*	32,432,796	T	G	0,51	1,21	1,09	1,33	2,1x10 ⁻⁴	0,35	1,07	0,84	1,36	0,58		0,40
9q22.33	rs1588635*	100,537,802	A	C	0,55	0,61	0,55	0,68	2,1x10 ⁻²¹	0,75	0,85	0,65	1,12	0,24	rs1588635	-
	rs965513	100,556,109	A	G	0,55	0,62	0,56	0,68	1,1x10 ⁻²⁰	0,69	0,77	0,61	0,99	0,04		0,99
14q13.3	rs368187*	36,532,576	C	G	0,63	1,47	1,32	1,63	3,8x10 ⁻¹³	0,31	1,51	1,16	1,96	2,3x10 ⁻³	rs368187	-
	rs944289	36,649,246	C	T	0,65	1,36	1,22	1,51	8,5x10 ⁻⁹	0,35	1,35	1,06	1,72	0,02		0,75
	rs116909374*	36,738,361	T	C	0,94	0,43	0,33	0,55	1,6x10 ⁻¹⁰	0,99	0,24	0,05	1,17	0,08	rs116909374	-

AR : Allèle à risque ; AA : Allèle alternatif ; FAR : fréquence de l'allèle à risque chez les témoins ; OR : Odds ratio ; 95%CI : 95% intervalle de confiance, *p* : *p*-valeur, * SNP ayant la plus petite *p*-valeur dans la méta-analyse de Gudmundsson et coll, 2017.

3. Identification de nouvelles régions de prédisposition au CDT

Tableau 5 : Résultats de l'analyse de réplication pour les SNPs identifiés

locus	SNP			Européens (1554 cas/1973 témoins)					Océaniens (301 cas/348 témoins)				
		AA	AR	RAF	OR	95% CI		<i>p</i>	RAF	OR	95% CI		<i>p</i>
1p31.3	rs334729	G	C	0.05	1.89	1.52	2.44	8.7x10 ⁻⁸	0.14	0.98	0.69	1.39	0.92
5p15.33	rs7726159	C	A	0.39	1.26	1.13	1.40	2.0x10 ⁻⁵	0.70	1.42	1.11	1.82	5.6x10 ⁻³
16q23.2	rs16950982	A	G	0.37	1.26	1.13	1.40	2.4x10 ⁻⁵	0.58	1.15	0.92	1.45	0.22
19p12	rs10415826	A	T	0.20	1.35	1.19	1.53	5.0x10 ⁻⁶	0.20	1.33	0.99	1.79	0.06

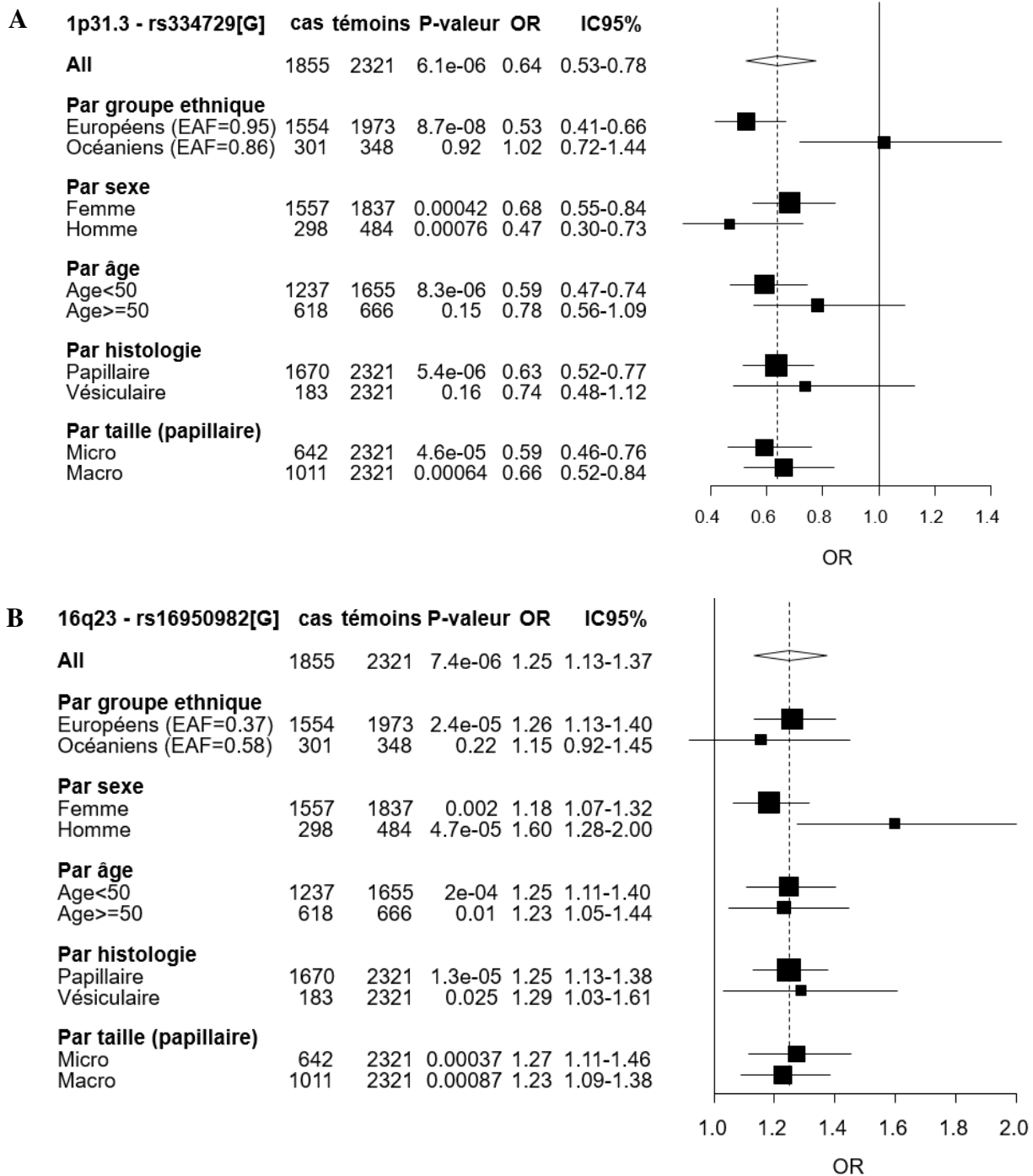
AR : Allèle à risque ; AA : Allèle alternatif ; RAF : fréquence de l'allèle à risque chez les témoins ; OR : Odds ratio ; 95%CI : intervalle de confiance à 95%, *p* : *p*-valeur

Un nouveau locus significativement associé au risque de CDT chez les Européens d'EPITHYR a été mis en évidence au locus 1p31.3. Le SNP le plus associé, rs334729 (OR = 1,89, IC95% : 1,52 - 2,44, *p*-valeur = 8,7 x 10⁻⁸) (Tableau 5), se situe dans l'intron 3 du gène *NFIA* (Figure 13 A). L'analyse de ce variant dans l'étude italienne (Köhler et al. 2013) et l'étude de l'Ohio (Gudmundsson et al. 2017) nous a permis de confirmer l'association (*p*-valeur = 0,03) chez les Européens malgré une hétérogénéité significative observée entre les études (*p*-hétérogénéité = 8,6 × 10⁻³) (Tableau supplémentaire 2).

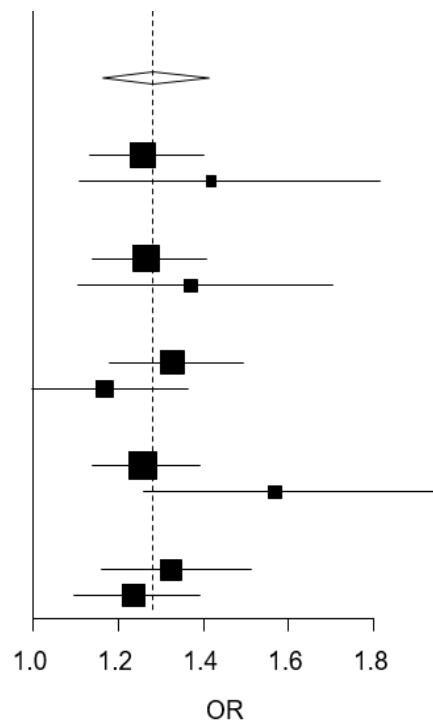
Toutes les régions chromosomiques présentant au moins deux SNPs associés au CDT avec des *p*-valeurs <10⁻⁵ dans l'analyse d'association principale ou dans une des analyses stratifiées menées chez les Européens (tableau supplémentaire 2), ou dans l'analyse poolée des deux populations (tableau supplémentaire 2) ont été sélectionnées pour une analyse de réplication dans une population européenne indépendante incluant les données des études cas-témoins de l'Ohio et de l'Italie.

Au total, les SNPs les plus associés de 20 loci d'intérêt ont été analysés dans l'échantillon de réplication, et seuls les SNPs situés aux loci **5p15.33**, **16q23.2** et **19p12** ont été répliqués (Tableau 5 et Tableau supplémentaire 2).

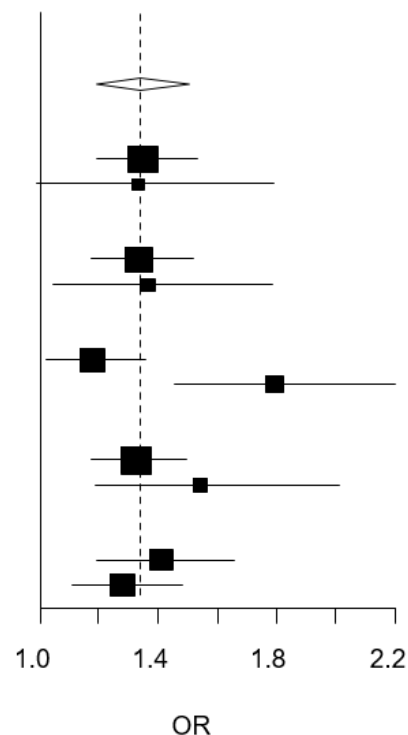
Figure 13 : Analyses stratifiées des SNPs : A.rs334729 (1p31.3) ; B. rs16950982 (16q23) ; C. rs7726159 (5p15) ; D. rs10415826 (19p12).



C	5p15 - rs7726159[A]	cas	témoins	P-valeur	OR	IC95%
	All	1855	2321	4.2e-07	1.28	1.17-1.41
	Par groupe ethnique					
	Européens (EAF= 0.39)	1554	1973	2e-05	1.26	1.13-1.40
	Océaniens (EAF=0.70)	301	348	0.0056	1.42	1.11-1.82
	Par sexe					
	Femme	1557	1837	1.3e-05	1.27	1.14-1.41
	Homme	298	484	0.0043	1.37	1.10-1.70
	Par âge					
	Age<50	1237	1655	2.3e-06	1.33	1.18-1.49
	Age>=50	618	666	0.054	1.17	1.00-1.36
	Par histologie					
	Papillaire	1670	2321	6.1e-06	1.26	1.14-1.39
	Vésiculaire	183	2321	6.4e-05	1.57	1.26-1.95
	Par taille (papillaire)					
	Micro	642	2321	2.9e-05	1.32	1.16-1.51
	Macro	1011	2321	0.00055	1.23	1.10-1.39



D	19q12 - rs10415826[T]	cas	témoins	P-valeur	OR	IC95%
	All	1855	2321	9.8e-07	1.34	1.19-1.51
	Par groupe ethnique					
	Européens (EAF=0.20)	1554	1973	5e-06	1.35	1.19-1.53
	Océaniens (EAF=0.20)	301	348	0.06	1.33	0.99-1.79
	Par sexe					
	Femme	1557	1837	1.2e-05	1.34	1.17-1.52
	Homme	298	484	0.02	1.36	1.04-1.78
	Par âge					
	Age<50	1237	1655	0.02	1.18	1.02-1.36
	Age>=50	618	666	3.8e-08	1.79	1.45-2.20
	Par histologie					
	Papillaire	1670	2321	5.6e-06	1.32	1.17-1.50
	Vésiculaire	183	2321	0.001	1.54	1.18-2.01
	Par taille (papillaire)					
	Micro	642	2321	4.2e-05	1.41	1.19-1.66
	Macro	1011	2321	7e-04	1.28	1.11-1.48



Les carrés représentent les odds ratios (ORs) ; la taille du carré représente l'inverse de la variance du logarithme de l'OR ; les lignes horizontales représentent les intervalles de confiance à 95 % ; les diamants représentent les estimations de l'analyse groupée, La ligne verticale continue représente un OR de 1 ; la ligne verticale pointillée représente l'OR global, Ca : nombre de cas, témoins : nombre de témoins, EAF : fréquence de l'allèle à effet, OR : Odds ratio, IC : intervalle de confiance.

Seuls deux variants, rs16950982 et rs17767383, au locus **16q23.2** en fort DL chez les Européens ($r^2 = 0,88$), mais plus faiblement corrélé chez les Océaniens ($r^2 = 0,37$) ont atteint le seuil de significativité GWAS de p -valeur = 5×10^{-8} dans la méta-analyse des études EPITHYR, Ohio et Italie.

Nous avons observé que l'association du SNP rs16950982 [G] allait dans le même sens chez les Océaniens et les Européens (Tableau 5) et que l'association est plus forte chez les hommes (OR = 1,60, p -valeur = $4,7 \times 10^{-5}$) que chez les femmes (OR = 1,18, p -valeur = 2×10^{-3}) (p -hétérogénéité = 0,02) (Figure 13B).

En **5p15.33**, l'association avec le SNP rs7726159 est répliquée dans l'échantillon de réplication (p -valeur = 0,02, Tableau supplémentaire 2), mais dans la méta-analyse combinant les phases de découverte et de réplication, l'association reste non significative au seuil de p -valeur = 5×10^{-8} (p -valeur = $4,9 \times 10^{-6}$) (Tableau 5). L'association avec ce SNP chez les Océaniens est dans la même direction que pour les Européens, avec une fréquence plus élevée de l'allèle à risque rs7726159[A] (Tableau 5, Figure 13 C).

En **19p12**, l'association avec le rs10415826 est significative dans l'analyse de réplication (p -valeur = $7,9 \times 10^{-3}$) (Tableau supplémentaire 2) et nous obtenons dans la méta-analyse combinant les données d'EPITHYR avec les données de réplication, un OR = 1,22, p -valeur = $8,7 \times 10^{-7}$. L'association chez les Océaniens n'est pas significative mais va dans la même direction (Tableau 5, Figure 13 D).

4. Analyse de réplication des locus mis en évidence dans la méta-analyse de GWAS (Gudmundsson et al 2017)

La méta-analyse de GWAS publiée en 2017 par Gudmundsson, a mis en évidence 5 nouveaux loci de prédisposition (1q42.2, 3q.26.2, 5q22.1, 10q24.33) mais n'incluait pas de phase de réplication dans un échantillon indépendant. A notre connaissance, aucune étude n'a depuis confirmé ces associations. Nous avons donc conduit une étude de réplication en utilisant les données d'EPITHYR.

L'association avec tous les SNPs testés vont dans la même direction dans la méta-analyse de GWAS et chez les européens d'EPITHYR, mais seules les *p*-valeurs d'associations avec les SNPs des loci 1q42.2, 5q22.1 et 10q24 était inférieures à 0.001 chez les Européens d'EPITHYR. Chez les Océaniens, les associations vont toutes dans le sens inverse et aucune n'est significative.

VI. Discussion

Ce GWAS réalisé dans la population d'EPITHYR nous a permis de confirmer les associations connues avec les loci de susceptibilité au CDT en **2q35**, **8p12**, **9q22.33** et **14q13.3** dans la population européenne et de mettre en évidence de nouveaux loci d'intérêt en **1p31.3** (rs334729, rs334725), **5p15.33** (rs7726159), **16q23.2** (rs16950982) et **19p12** (rs10415826). A l'inverse aucune association significative n'a pu être mise en évidence dans la population océanienne, pour laquelle nous ne disposions pas d'échantillon de réplication.

Dans notre étude, l'allèle C du SNP rs334729 au locus **1p31.3** est associé à une augmentation du risque de CDT et est en DL avec l'allèle [C] du SNP rs334725 qui a été associé à des niveaux réduits de TSH dans une étude GWAS menée dans une population islandaise ($\beta = -0,13$, *p*-valeur = $1,0 \times 10^{-20}$) (Gudmundsson et al, 2012). Dans l'étude de Gudmundsson et coll., l'allèle [C] du SNP rs334725 est également associé au risque de CPT (OR = 1,30, *p*-valeur = $6,6 \times 10^{-3}$). Ces deux SNPs sont situés dans le gène *NF1A*, qui code une protéine de la famille des facteurs de transcription NF1 (facteur nucléaire 1). Des études expérimentales ont d'ailleurs

montré que les isoformes de NF1 interagissent avec le facteur de transcription thyroïdien 1 (TTF-1) dans la régulation des fonctions cellulaires (Bachurski et al, 2003) soulignant l'hypothèse d'un rôle de ces variants dans la prolifération de cellules tumorales thyroïdiennes via une dérégulation l'expression de la voie métabolique NAF1 – NF1.

Au locus **16q23.2**, le variant rs16950982 [G] est associé à une augmentation du risque de CDT (OR = 1,35). Ce variant est corrélé avec l'allèle rs3813582[C] ($r^2 = 0,88$) associé à une diminution du niveau de TSH circulante ($\beta = -0,082$, p -valeur = $8,45 \times 10^{-18}$) avec une association qui est spécifiquement observée chez les hommes (Porcu et al. 2013), comme dans notre étude. Rs16950982 [G] est également en DL avec l'allèle rs17767419[T] ($r^2 = 0,88$) qui a été associé à une augmentation du volume de la thyroïde ($\beta = 0,068$, p -valeur = $9,4 \times 10^{-15}$) (Teumer et al, 2011). Ces deux SNPs de 16q23.2 sont situés dans une région intergénique proche du proto-oncogène *MAF* codant pour un facteur de transcription (Figure supplémentaire 3 C). Selon les données du GTEx, référençant les données d'expression des gènes dans de multiples tissus (<https://www.gtexportal.org/home/>), rs16950982[G] et rs17767383[A] sont significativement associés à une plus faible expression de *MAFTRR* dans de nombreux tissus, y compris le tissu thyroïdien (respectivement $\beta = -0,55$, p -valeur = $2,8 \times 10^{-36}$ et $\beta = -0,60$, p -valeur = $2,7 \times 10^{-43}$), ce qui suggère que *MAFTRR* joue ici un rôle de suppresseur de tumeurs (Cheng et al, 2017).

En **5p15.33**, nous avons mis en évidence le SNP rs7726159, qui est situé dans l'intron 3 de *TERT* codant la transcriptase inverse de la télomérase jouant un rôle clé dans le maintien des télomères. Le SNP le plus associé de la première méta-analyse publiée (Gudmundsson et al, 2017), rs10069690[T] (OR = 1,20, p -valeur = $3,2 \times 10^{-7}$), est modérément corrélé avec le SNP rs7726159 ($r^2 = 0,46$) que nous avons mis en évidence chez les Européens. Aussi, une étude menée dans une population chinoise (Ge et al, 2016) et analysant les haplotypes de la région *TERT-CLPTM1* à l'aide de 15 marqueurs a mis en évidence un risque plus élevé de CPT pour les porteurs de l'allèle rs2736100[G] qui est fortement corrélé avec rs7726159[A] chez les Asiatiques ($r^2 = 0,92$ dans la population 1000Genomes CHB) mais en plus faible DL dans les populations européennes ($r^2 = 0,43$ dans la population 1000Genomes CEU). De plus, il a été

démontré que l'allèle à risque [A] du SNP rs7726159 augmente la transcription de *TERT* et cet allèle a été précédemment associé à une longueur accrue des télomères dans les populations européennes (Pooley et al, 2013) et asiatiques (Li et al, 2019). Il est également associé au risque de multiples cancers, cancers du sein et de l'ovaire inclus (Pooley et al, 2013; Rode, Nordestgaard, and Bojesen 2016). Les mêmes observations ont été faites dans une étude danoise portant sur 95 568 personnes où les auteurs concluent que les déterminants génétiques des télomères longs sont associés à un risque accru de cancer. Ainsi, l'ensemble de ces données suggère que les télomères longs favorisent la survie des cellules précancéreuses et donc la survenue du cancer. Les résultats de notre étude vont aussi dans ce sens.

En **19p12**, rs10415826 est situé dans l'intron 1 de *ZNF257* (Figure supplémentaire 3 D). Aucune association n'a été mise en évidence dans la littérature pour ce variant et la fonction du facteur de transcription *ZNF257* appartenant à la famille des protéines possédant des motifs de liaison à l'ADN de type « doigts de zinc », a été peu étudiée.

Finalement, il est important de souligner que cette étude est la première étude GWAS réalisée dans une population océanienne, incluant des individus recrutés en Polynésie française et en Nouvelle-Calédonie, deux territoires où l'incidence du cancer de la thyroïde est parmi les plus élevées. À l'exception du locus 1p31.3, toutes les associations que nous avons mises en évidence chez les Européens vont dans le même sens dans la population océanienne. Des études pangénomiques supplémentaires dans les populations océaniques sont nécessaires afin d'identifier des variants de prédisposition spécifiques à ces populations où l'incidence élevée reste encore inexpliquée. Les faibles effectifs pour cette population n'ont pas permis d'avoir la puissance statistique suffisante pour confirmer les variants identifiés chez les Européens. Il est peu probable que les facteurs génétiques expliquent à eux seuls l'incidence très élevée de CDT observés dans les populations polynésiennes et mélanésiennes. Les données du consortium EPITHYR permettront d'étudier l'effet combiné entre les facteurs génétiques et d'autres facteurs de risque liés au mode de vie ou à l'environnement pour le CDT et de réaliser des études d'interactions "Gène-environnement".

Chapitre 2

Cartographie fine des régions 2q35 et 8p12 dans les populations européennes et océaniques d'EPITHYR et recherche de variants causaux

I. Introduction

Les polymorphismes associés au risque de maladie dans les GWAS sont le plus souvent dans des régions non codantes du génome, avec une fonction biologique non connue. Du fait des blocs de DL, le variant ayant la plus petite p -valeur dans un locus d'intérêt n'est pas forcément celui qui est le plus susceptible d'avoir un impact fonctionnel délétère. L'objectif des analyses de cartographie fine (ou fine-mapping) est de prioriser les variants d'un loci mis en évidence dans les GWAS, en fonction de leur fonction biologique connue.

Les études GWAS sur les CDT, dont le GWAS d'EPITHYR, ont permis de mettre en évidence des associations robustes avec plusieurs SNPs aux loci 2q35, 8p12, 9q22.33 et 14q13.3 dans les populations européennes et asiatiques (Gudmundsson et al. 2009, 2012; Köhler et al. 2013; Mancikova et al. 2015; Son et al. 2017; Takahashi et al. 2010; Truong et al. 2021).

Au locus 14q13, une étude fonctionnelle a mis en évidence le rôle de *PTSC53* dans la prédisposition au CDT via une sous-expression du gène cible *S100A4* (Jendrzewski et al. 2015). Au locus 9q22.33, un nouveau gène à long ARN non codant intergénique, *PTCSC2*, a

été identifié pour lequel les transcrits sont sous-exprimés dans les tissus tumoraux de CPT (He et al. 2015) suggérant que *PTCSC2* pourrait avoir un rôle régulateur dans des mécanismes biologiques thyroïdiens.

Dans une étude d'EPITHYR, la cartographie fine des régions 9q22.33 et 14q13.3 réalisée dans un échantillon européen (625 cas et 776 témoins) et mélanésien (244 cas et 189 témoins) de France métropolitaine et de Nouvelle-Calédonie (Tcheandjieu et al. 2016) a permis de montrer que l'haplotype rs944289[C]-rs116909374[C]-rs999460[T] est associé de manière significative au risque de CDT chez les Européens. De plus dans cette étude l'association entre CDT et rs116909374 diffère selon le statut tabagique des individus chez les Européens. Enfin, la même étude montre, un nouveau signal indépendant de ceux déjà connus, au niveau du SNP rs1755774 au locus 14q13.3 chez les Mélanésiens.

A ce jour, aucune étude de cartographie fine n'a portée sur les loci 2q35 et 8p12 qui sont les deux autres loci les plus significativement associés au risque de CDT.

Locus 2q35

Le SNP **rs966423**, situé dans le gène *DIRC3*, est le premier SNP au locus 2q35 à avoir été significativement associé au risque de CDT (OR = 1,26 ; IC95% : 1,11 - 1,43, p -valeur = $3,8 \times 10^{-4}$). Ce locus a été également mis en évidence dans une étude GWAS de CDT dans la population italienne (Köhler et al. 2013) avec l'identification du SNP **rs6759952** comme SNP le plus associé ("majeur" ou "top SNP") (OR = 1,21, IC95% : 1,18-1,43 ; p -valeur = $6,4 \times 10^{-10}$). Ce SNP est en fort DL avec le SNP rs966423 ($r^2 = 0,74$). Une méta-analyse de cinq études GWAS (Gudmundsson et al. 2017) menée sur des populations européennes a également identifié le SNP **rs11693806** situé au sein de *DIRC3*, avec un OR = 1,43 (IC95% : 1,33 - 1,54, p -valeur = $1,5 \times 10^{-24}$). Rs11693806 est situé dans la même région et est en DL avec les variants rs6759952 et rs966423 ($r^2 = 0,47$).

Aussi, le locus 2q35 est associé au risque de CDT dans la GWAS menée dans la population Coréenne. Le SNP majeur identifié dans ce GWAS est **rs12990503** (OR = 1,34, p -valeur = 3,55

$\times 10^{-09}$) et est en DL avec les trois variants identifiés dans les études européennes citées plus haut dans les populations asiatiques de 1000Genomes ($r^2 > 0.61$).

Ainsi, plusieurs SNPs de ce locus, plus ou moins corrélés entre eux, ont été mis en évidence dans différentes populations et dans différents GWAS sur les CDT.

Locus 8p12

Le locus 8p12 a été associé au CDT pour la première fois dans la GWAS islandaise dont l'objectif premier était d'identifier des facteurs génétiques associés au taux de TSH sanguin (Gudmundsson et al. 2012). Dans cette étude, les auteurs rapportent une association entre CDT et **rs2439302** (OR = 1,41 ; IC95% : 1,23 - 1,62 ; p -valeur = $1,3 \times 10^{-6}$). Ce variant, situé dans le gène *NRG1* a été répliqué dans la population coréenne avec un OR = 1,37 (p -valeur = $1,42 \times 10^{-09}$) (Son et al. 2017). Cette même GWAS coréenne a mis en évidence un autre variant **rs12542743** en 8p12 (OR = 1,36, p -valeur = $4,61 \times 10^{-10}$). Selon les données de 1000Genomes, ces deux variants sont en faible DL dans les populations asiatique ($r^2 = 0,27$) et européenne ($r^2 = 0,57$). Dans la méta-analyse des études GWAS européennes (Gudmundsson et al. 2017) un OR de 1,32 (IC 95% : 1,23 - 1,41) ; p -valeur = $1,50 \times 10^{-17}$) est obtenu pour **rs2466076**. Ce SNP est en fort DL avec rs2439302 ($r^2 = 0,94$).

Des hypothèses fonctionnelles ont été avancées pour expliquer le rôle de cette région dans l'étiologie du CDT. Dans l'étude sur les facteurs génétique associés au niveau de TSH circulant, l'allèle à risque rs2439302[G] a été associé à une expression plus élevée de *NRG1* (Gudmundsson et al. 2012) ; cet allèle a été également corrélé à une forte expression de différents isoformes de ce gène (He et al. 2018). Ces résultats suggèrent un rôle possible de ce gène dans le développement du CDT.

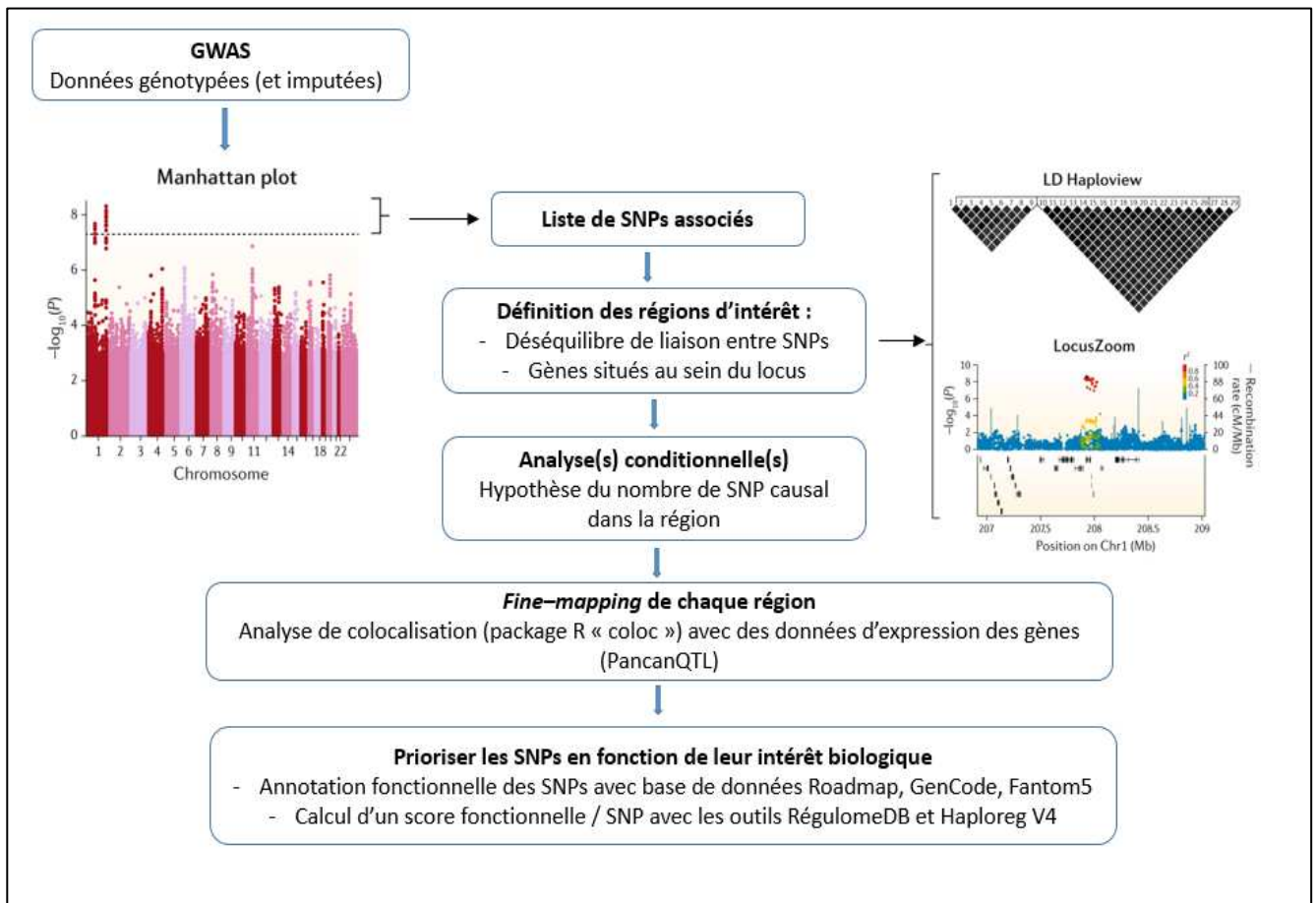
Enfin, il a été démontré que les allèles à risque rs966423[C] au locus 2q35 et rs2439302[G] au locus 8p12 sont associés à un niveau d'expression plus faible de TSH dans le sang (Gudmundsson et al. 2012).

Stratégie du fine-mapping

Le but du fine-mapping est d'identifier le (ou les) variant(s) génétique(s) responsable(s) d'une maladie dans une région de susceptibilité génétique où les variants sont en forts DL, en supposant qu'il existe au moins un variant causal. Après qu'une étude GWAS ait identifié au moins un SNP fortement associé au risque de la maladie (avec une p -valeur significative $< 5 \times 10^{-8}$), l'étude de fine-mapping se déroule généralement en suivant différentes étapes.

La stratégie générale consiste à explorer les régions dans lesquelles se situent les SNPs significatifs du GWAS. On s'intéresse à la structure de DL qui existe au sein de la région, à la position des marqueurs associés à la maladie étudiée et aux gènes présents sur le locus. L'objectif est d'enrichir la région en annotations fonctionnelles permettant de prioriser les variants significatifs selon leur fonction biologique connue ou probable et dont le rôle pourra être confirmé dans de futures études fonctionnelles. J'ai détaillé la stratégie que j'ai spécifiquement appliquée ainsi que les outils utilisés dans mon étude de fine-mapping en figure 14.

Figure 14 : Stratégie de la cartographie fine utilisée dans notre étude.



Diverses raisons peuvent expliquer les effets différents des SNPs d'une population à une autre. Le variant mis en évidence peut être en DL avec le variant causal dans une population donnée mais pas dans une autre population avec une structure de DL différente. L'effet des gènes peut aussi être modulé par des facteurs environnementaux ou liés au mode de vie spécifiques dont la prévalence peut être différente d'une population à l'autre. L'objectif de notre étude est d'identifier et de prioriser des variants les plus susceptibles d'être délétères au sein des régions 2q35 et 8p12 en exploitant les différences de DL entre les populations d'EPITHYR et en utilisant des données d'annotation fonctionnelle disponibles.

II. Matériel et méthodes

1. La population d'étude : le consortium EPITHYR

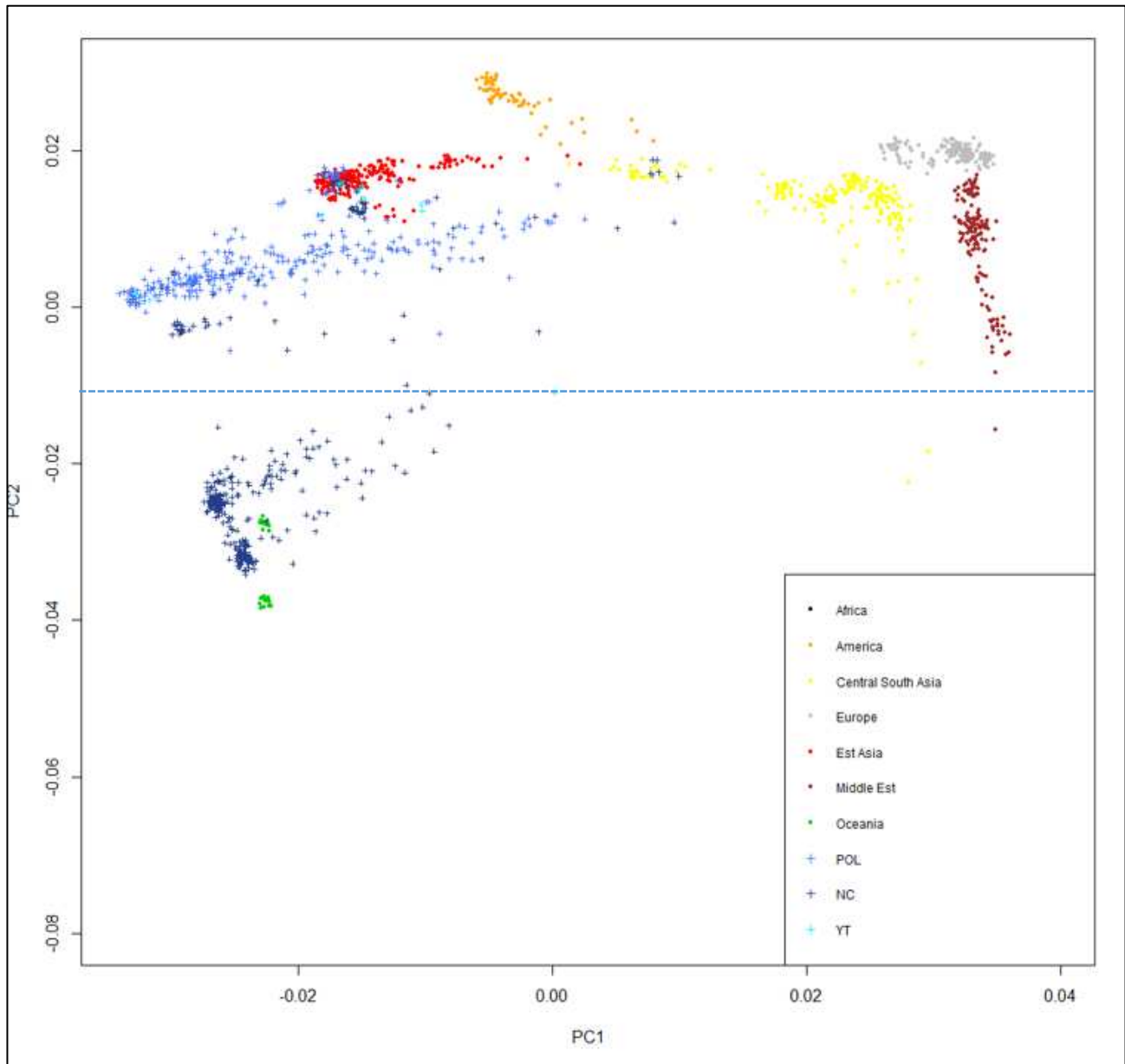
Les analyses ont été effectuées sur le même échantillon d'EPITHYR que l'étude GWAS présentée dans le chapitre précédent, c'est à dire sur 1554 cas et 1973 témoins européens et 301 cas et 348 témoins océaniens.

i. Définition des groupes de population

Les ACP réalisées lors du contrôle qualité ont montré une certaine hétérogénéité au sein de la population océanienne d'EPITHYR. Nous avons donc d'abord réalisé une nouvelle ACP des 649 Océaniens avec le panel de référence Human Genome Diversity Panel (HGDP) (Cavalli-Sforza 2005) qui est composé de 940 individus non apparentés aux origines ethniques diverses et ayant été génotypés avec la puce Illumina GSAMD-24v1-0_20011747_A. Nous avons ainsi récupéré les génotypes des 940 individus pour 642 810 variants. Le panel HGDP940 est composé de 101 individus originaires d'Afrique, 64 individus d'Amérique centrale ou latine, 200 individus du Sud de l'Asie ou d'Asie centrale, 228 individus d'Asie de l'Est, 157 individus d'Europe, 162 individus du Moyen-Orient et 27 individus d'origine Océanienne (Bougainville et Nouvelle-Guinée).

Une ACP a été réalisée sur 75 588 SNPs ayant une FAM >5% et en faible DL ($r^2 < 0,8$) (Figure 15).

Figure 15 : ACP des 649 individus Océaniens d'EPITHYR et des 940 individus du panel HGDP.



Les croix bleues représentent les individus EPITHYR des études Nouvelle-Calédonie (NC), Polynésie-française (POL) et YoungThyr (YT). Les points de différentes couleurs correspondent aux différentes populations du panel de référence HGDP. La ligne en pointillés bleue correspond au seuil de CP2 choisi afin de séparer les Océaniens d'EPITHYR en deux sous-groupes de populations.

On observe une stratification de la population Océanienne d'EPITHYR qui forme deux groupes distincts d'individus. Un groupe d'individus est plus proche de la population d'Asie de l'Est (en rouge) est un second groupe est proche de la population mélanésienne de Bougainville (en vert). Nous avons choisi de diviser les 649 Océaniens en deux sous-populations selon la valeur de CP2 égale à -0.01 sur l'axe des ordonnées (représentée par une ligne pointillée bleue sur la figure 15).

Les individus océaniens ayant une valeur de $CP2 > 0,01$ sont dorénavant définis comme **Polynésiens** (146 cas / 242 témoins) et correspondent majoritairement à des individus de l'étude Polynésie française et de quelques individus de l'étude Nouvelle-Calédonie. L'ensemble des individus océaniens ayant une valeur de $CP2 < 0,01$ sont définis comme **Mélanésiens**. Ce groupe est composé uniquement d'individus de l'étude de Nouvelle-Calédonie (155 cas / 106 témoins).

Nous avons effectué différentes étapes de contrôle qualité sur les individus et les SNPs dans les populations mélanésiennes et polynésiennes, en considérant pour chacun des deux groupes, le taux d'hétérozygotie moyen, en excluant les SNPs ne respectant pas l'équilibre d'Hardy-Weinberg et en excluant les SNPs monomorphes. Aucun individu n'a été exclu suite à ces étapes.

ii. Description de la population

La description des populations d'étude selon l'étude, le sexe, l'âge et le type histologique des carcinomes est détaillée dans le tableau 6.

Tableau 6 : description des populations européennes, polynésiennes et mélanésiennes d'EPITHYR selon l'étude, l'âge, le sexe, le type histologique et la taille des carcinomes.

	Européens (N=3 527)				Polynésiens (N=388)				Mélanésiens (N=261)			
	Cas		Témoins		Cas		Témoins		Cas		Témoins	
	N=1554	%	N=1973	%	N=146	%	N=242	%	N=155	%	N=106	%
Etude												
CATHY	451	28,9	534	27,1	1	0,7						
Cuba	102	6,6	113	5,7								
Tchernobyl	66	4,2	304	15,4								
E3N	277	17,8	287	14,5								
Nouvelle Calédonie	21	1,3	70	3,5	28	19,2	55	22,7	155	100	106	100
Polynésie Française			4	0,2	114	78,1	181	74,8				
YoungThyr	637	41,0	661	33,5	3	2,0	6	2,5				
Age (années)												
[0-25[	269	17,3	513	26,0	15	10,3	23	7,5	7	4,5	3	2,8
[25-50[	765	49,3	891	45,2	98	67,1	163	65,8	82	52,9	66	62,3
≥50	520	33,4	569	28,8	33	22,6	56	26,7	66	42,6	37	34,9
Sexe												
Homme	275	17,7	454	23,0	14	9,6	22	8,6	9	5,8	8	7,5
Femme	1279	82,3	1519	77,0	132	92,4	220	91,4	146	94,2	98	92,5
Type histologique												
Vésiculaire	137	8,8			25	17,1			23	14,8		
Papillaire	1417	91,2			121	82,9			132	85,2		
< 10 mm	506	35,7			67	45,9			69	44,5		
≥ 10 mm	904	63,8			45	30,8			62	40		

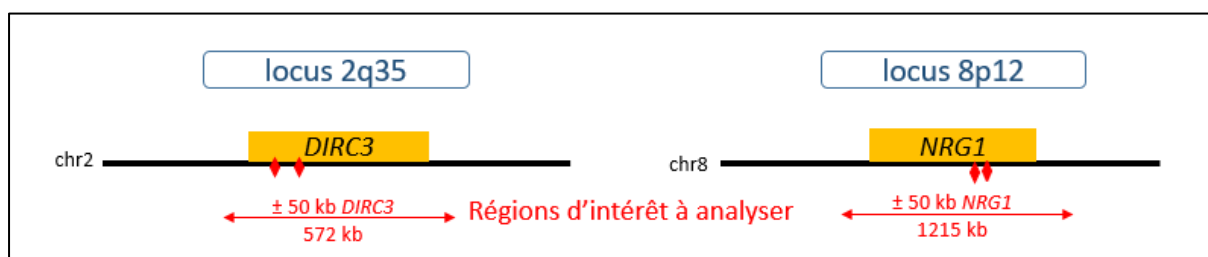
2. Les données génétiques

La puce OncoArray utilisée pour EPITHYR a été enrichie dans des régions potentiellement liées à des fonctions thyroïdiennes ou au CDT. En particulier, les régions 2q35 et 8p12 ont été enrichies en ajoutant les SNPs avec une FAM > 0,02 et corrélés avec un $r^2 > 0,2$ aux SNPs identifiés dans les précédentes GWAS sur le CDT. Ainsi, 226 SNPs et 492 SNPs ont été respectivement ajoutés à la puce dans les régions 2q35 et 8p12 (Figure 16).

Les variants de 2q35 (rs12129938, rs966423, rs6759952, rs12990503) mis en évidence dans les GWAS étant localisés dans le gène *DIRC3* et ceux de 8p12 (rs2439302, rs6996585, rs12542743, rs2439302) étant localisés dans le gène *NRG1*, nous avons défini les régions à analyser à plus ou moins 50 kb en amont et en aval des bornes de ces gènes.

- Pour le locus 2q35, une région de 572 kb a été considérée (positions 218 098 746 - 218 671 316 sur l'assemblage NCBI build 37 ;
- Pour le locus 8p12, une région de 1 215 kb a été considérée (positions 31 456 820 - 32 672 558 sur l'assemblage NCBI build 37.

Figure 16 : Régions génomiques d'intérêt en 2q35 et 8p12 définies pour l'analyse de fine-mapping.



Au locus **2q35**, 452 SNPs génotypés chez les Européens et 441 SNPs génotypés chez les Océaniens (Mélanésiens et Polynésiens) sont retenus pour les analyses après la phase de contrôle qualité des données (Figure 17). Au locus **8p12**, 552 SNPs génotypés chez les

Européens et 497 SNPs génotypés chez les Océaniens (Polynésiens et Mélanésiens) sont retenus pour les analyses après la phase de contrôle qualité (Figure 18).

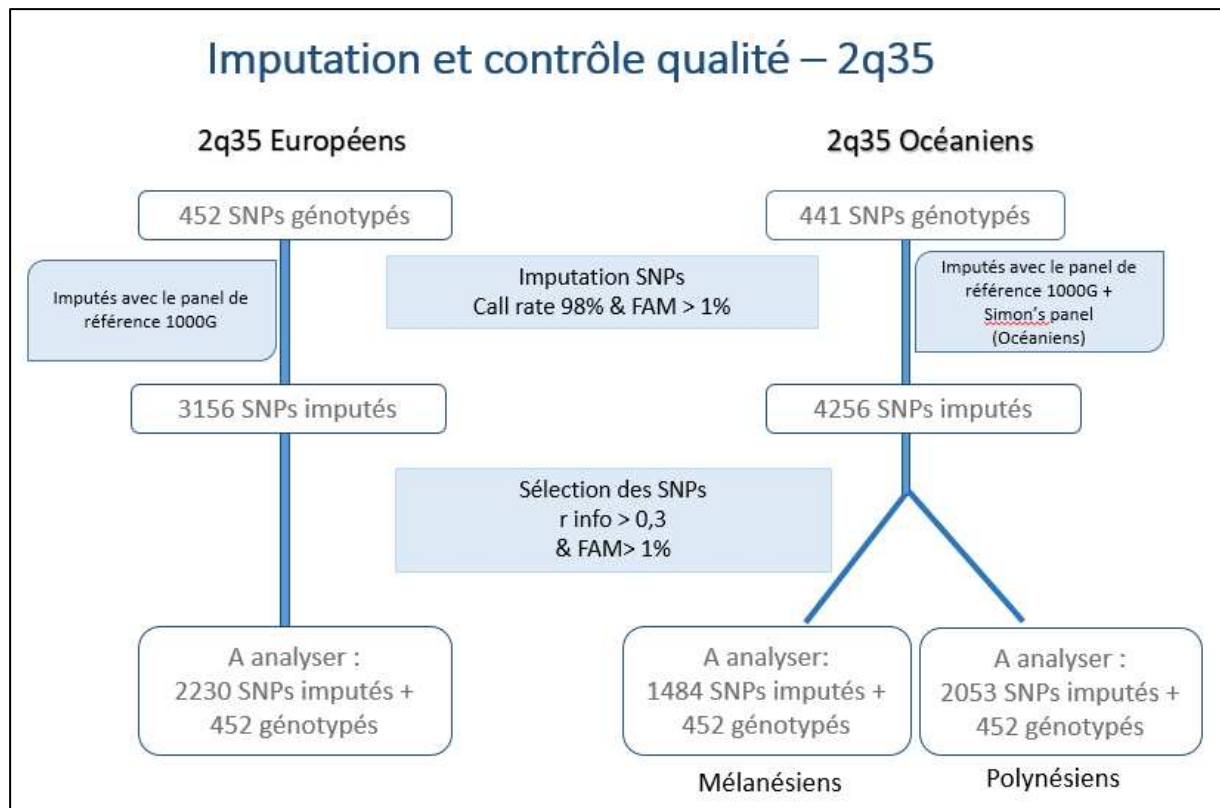
i. Imputation des génotypes manquants dans les deux régions

Les données génotypées ont été utilisées pour imputer les SNPs non génotypés en utilisant IMPUTEv2 et les données de 1000Genomes (phase 3) comme panel de référence. Le phasage des données et l'imputation ont été réalisés en une seule étape afin d'améliorer la qualité de l'imputation. Pour l'imputation chez les individus océaniens, nous avons utilisé les 26 individus océaniens du Simons Genome Diversity Panel (SGDP, phase 3) en plus du panel de référence du projet 1000Genomes. Ces individus provenaient majoritairement de l'île de Bougainville en Papouasie-Nouvelle-Guinée, située en Mélanésie.

L'étape d'imputation a été décrite plus en détail précédemment dans la partie « Données de génotypage » (page 57).

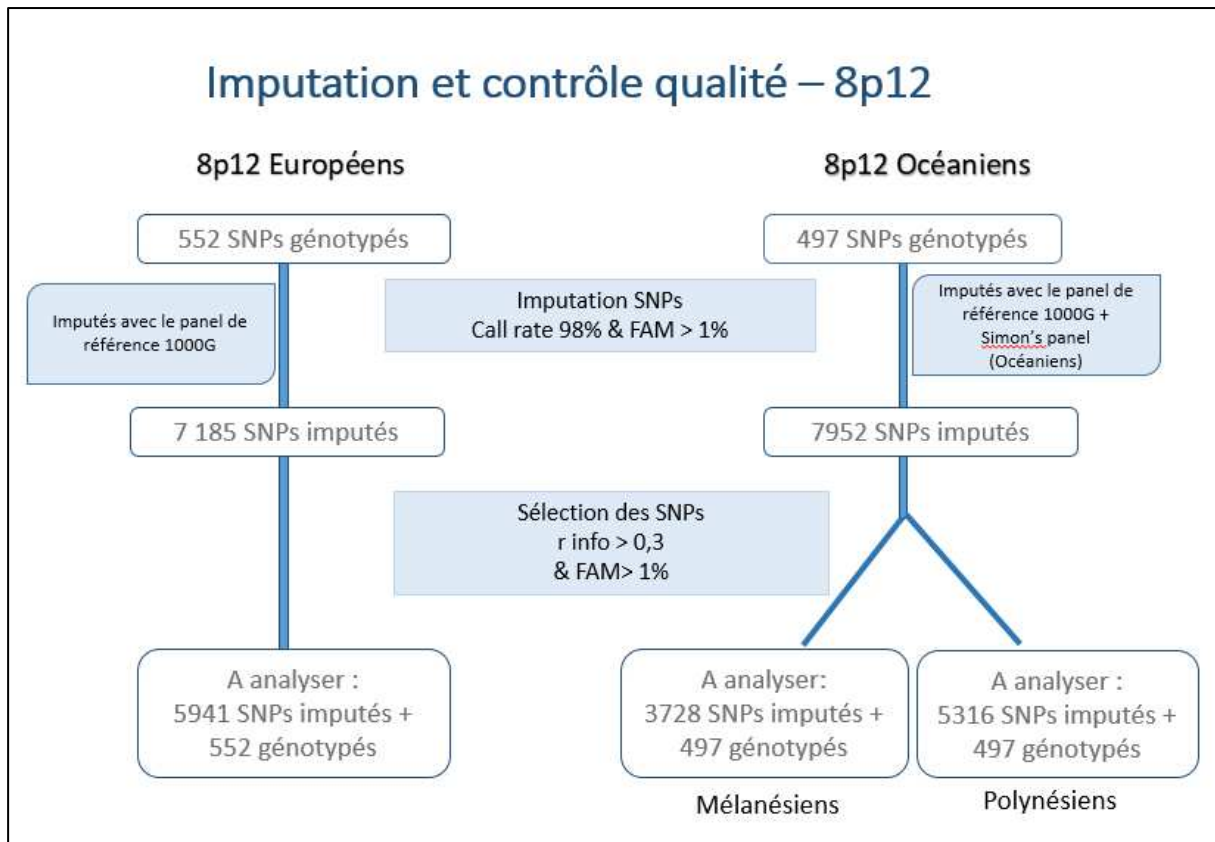
Après imputation, tous les SNPs monomorphes ont été exclus et seuls les SNPs imputés ayant un score de qualité d'imputation (info score) supérieur à 0,3 ont été conservés dans les analyses. Finalement, l'analyse du locus 2q35 a été basée sur 452 SNPs génotypés et 2 230 SNPs imputés chez les Européens, 441 SNPs génotypés et 2 053 SNPs imputés chez les Polynésiens, et 441 SNPs génotypés et 1 484 SNPs imputés chez les Mélanésiens (Figure 17).

Figure 17 : Schéma des étapes d'imputation des SNPs manquants au locus 2q35 pour les trois populations d'EPITHYR.



L'analyse du locus 8p12 a été basée sur 552 SNPs génotypés et 5 941 SNPs imputés chez les Européens, 497 SNPs génotypés et 5 316 SNPs imputés chez les Polynésiens, et 497 SNPs génotypés et 3 728 SNPs imputés chez les Mélanésiens (Figure 18).

Figure 18 : Schéma des étapes d'imputation au locus 8p12 pour les trois populations d'EPITHYR.



3. Stratégie d'analyse

i. Analyse conditionnelle

L'association entre chacun des SNPs (imputés et génotypés) aux loci 2q25 et 8p12 et le risque de DTC a été testée à l'aide d'un modèle logistique mixte ajusté sur l'âge, le sexe et l'étude avec un effet aléatoire qui dépend de la matrice de corrélation génétique des individus (obtenue avec le package R Gaston) pour chaque population.

Dans un second temps, nous avons réalisé des **analyses conditionnelles** afin de déterminer le nombre de signaux indépendants dans chacune des régions. Ainsi, pour chaque locus, un second test d'association a été fait en ajustant sur le top SNP. Si aucun signal d'association ne persiste après avoir conditionné les analyses sur ce top SNP, alors nous faisons l'hypothèse qu'il n'y a

pas d'autres signaux d'associations dans la région. Si après ajustement un signal persiste, l'analyse conditionnelle est répétée autant de fois qu'un SNP demeure significativement associé au seuil GWAS ($p\text{-valeur} = 5 \times 10^{-8}$). Le nombre de signaux indépendants dans la région est alors déterminé par le nombre d'étapes nécessaires pour que plus aucun signal ne soit significatif.

ii. Identification des variants causaux

Analyse de colocalisation avec PancanQTL

Les données sur les loci de caractère quantitatif d'expression (eQTL) permettent d'identifier des variants influençant le niveau d'expression d'un ou plusieurs gènes proches sur le même chromosome (cis) ou distants ou sur un chromosome différent (trans). Les données eQTL peuvent ainsi aider à prioriser les variants mis en évidence dans les GWAS pour des analyses fonctionnelles et à identifier les gènes qui pourraient jouer un rôle dans la maladie étudiée. En effet, bien que certains SNPs identifiés dans les GWAS soient situés dans des introns de gènes, cela ne garantit pas que le gène responsable de la prédisposition soit celui contenant ces SNPs. Ces SNPs peuvent en effet altérer les régions régulatrices et donc l'expression de gènes voisins.

La base de données PancanQTL ([http : //bioinfo.life.hust.edu.cn/PancanQTL/](http://bioinfo.life.hust.edu.cn/PancanQTL/)), répertorie les eQTL en cis et en trans identifiés par GWAS dans différents tissus tumoraux. Plus spécifiquement, cette base contient 5 606 570 paires "Gène-eQTL en cis" et 231 210 paires "Gène-eQTL en trans" identifiés dans 33 types de tissus tumoraux provenant du Cancer Genome Atlas (TCGA).

J'ai utilisé dans le cadre de mes analyses les données sur 10 766 gènes et 659 323 cis-eQTL obtenues à partir de 503 CPT.

L'analyse de colocalisation génétique de deux phénotypes consiste à évaluer si ces phénotypes partagent un ou plusieurs variants génétiques causaux et à les identifier. Pour ce faire, j'ai utilisé le package coloc de R (<https://cran.r-project.org/web/packages/coloc/index.html>) qui

permet de calculer pour chaque variant une **probabilité postérieure de causalité (PP)**. J'ai utilisé les résultats du test d'association entre chaque SNP et le risque de CDT dans EPITHYR et les résultats d'association entre chaque SNP et l'expression génétique dans les cellules tumorales de la thyroïde. Les hypothèses testées sont :

- H0 : Aucun SNP n'est causal au sein de la région
- H1 : Il existe un variant causal commun aux deux jeux de données d'association.

Ainsi, les PP obtenues correspondent à la probabilité de l'hypothèse H1.

iii. Annotation fonctionnelle des variants

Pour annoter les SNPs identifiés dans l'analyse de colocalisation et identifier les éléments régulateurs altérés par ces SNPs, j'ai utilisé les bases de données fonctionnelles de Roadmap (données génomiques de modifications d'histones, méthylation de l'ADN, DNase) (<http://www.roadmapepigenomics.org/>), Gencode (éléments fonctionnels situés dans des gènes) (<https://www.encodegenes.org/human/>), Fantom5 (éléments régulateurs transcriptomiques) ainsi que les outils d'annotation en ligne Haploreg V4 (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>) et RegulomeDB (<https://regulomedb.org/regulome-search/>).

III. Résultats

1. Locus 2q35

i. Test d'association univarié et analyses conditionnelles

Quarante-deux des 2 682 SNPs analysés au locus 2q35, sont associés au CDT chez les Européens au seuil p -valeur $< 5 \times 10^{-8}$. Le top SNP est rs57481445 (OR = 1,42 : p -valeur = $1,9 \times 10^{-10}$) (Figure 19 a). L'analyse conditionnelle sur ce SNP suggère qu'il n'existe qu'un seul signal d'association dans cette région (Figure 19 b).

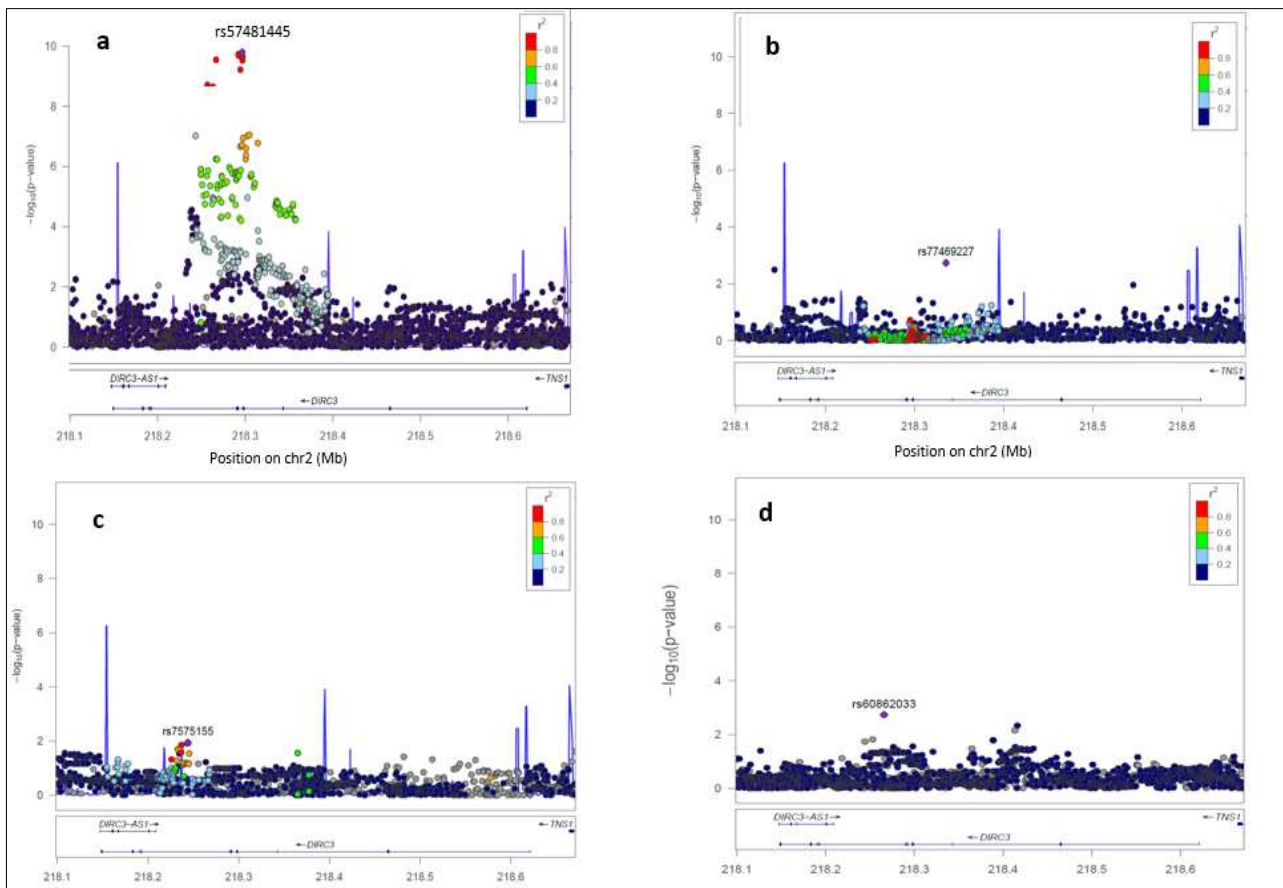
Chez les Mélanésien (Figure 19 c) et les Polynésien (Figure 19 d), aucun SNP n'est associé significativement au risque de CDT à ce locus. Les associations les plus significatives sont

retrouvées pour les SNPs rs7575155 chez les Mélanésien (OR = 2,42 : p -valeur = $1,1 \times 10^{-2}$) et rs60862033 (OR = 6,3 : p -valeur = $1,8 \times 10^{-3}$) chez les Polynésien (Figures 19.c et 19.d).

L'analyse poolée des trois groupes ethniques met en évidence le SNP rs3821098 (OR = 1,44, p -valeur = $1,6 \times 10^{-12}$) qui est fortement corrélé avec le SNP rs57481445 ($r^2 > 0,87$ dans les trois populations) (Figure 20 a). Suite à l'analyse conditionnelle sur ce SNP aucun signal supplémentaire significativement associé au risque de CDT n'est détecté (Figure 20 b).

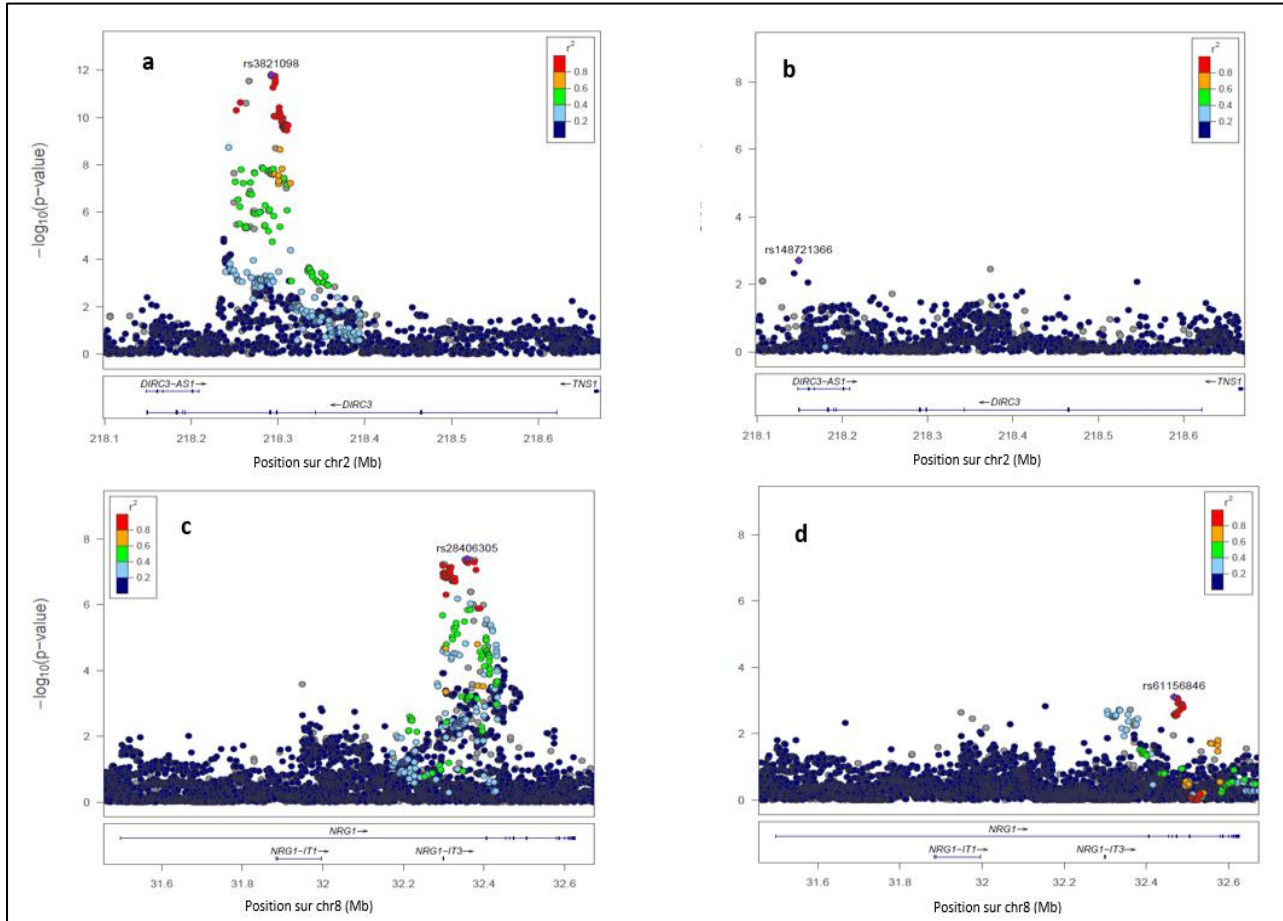
Les structures de DL au locus 2q35 dans les trois populations diffèrent avec des blocs de DL plus longs chez les Mélanésien et Polynésien comparés aux Européen. Toutefois, les SNPs rs57481445, rs16857609 et rs3821098 identifiés par l'analyse de colocalisation se situent dans le même bloc de DL dans les trois groupes ethniques (Figure 21).

Figure 19 : Locuszoom de l'association entre les SNPs de la région 2q35 et le CDT chez (a) les Européens, (c) les Mélanésiens et (d) les Polynésiens. (b) Association conditionnellement au SNP rs57481445 chez les Européens.



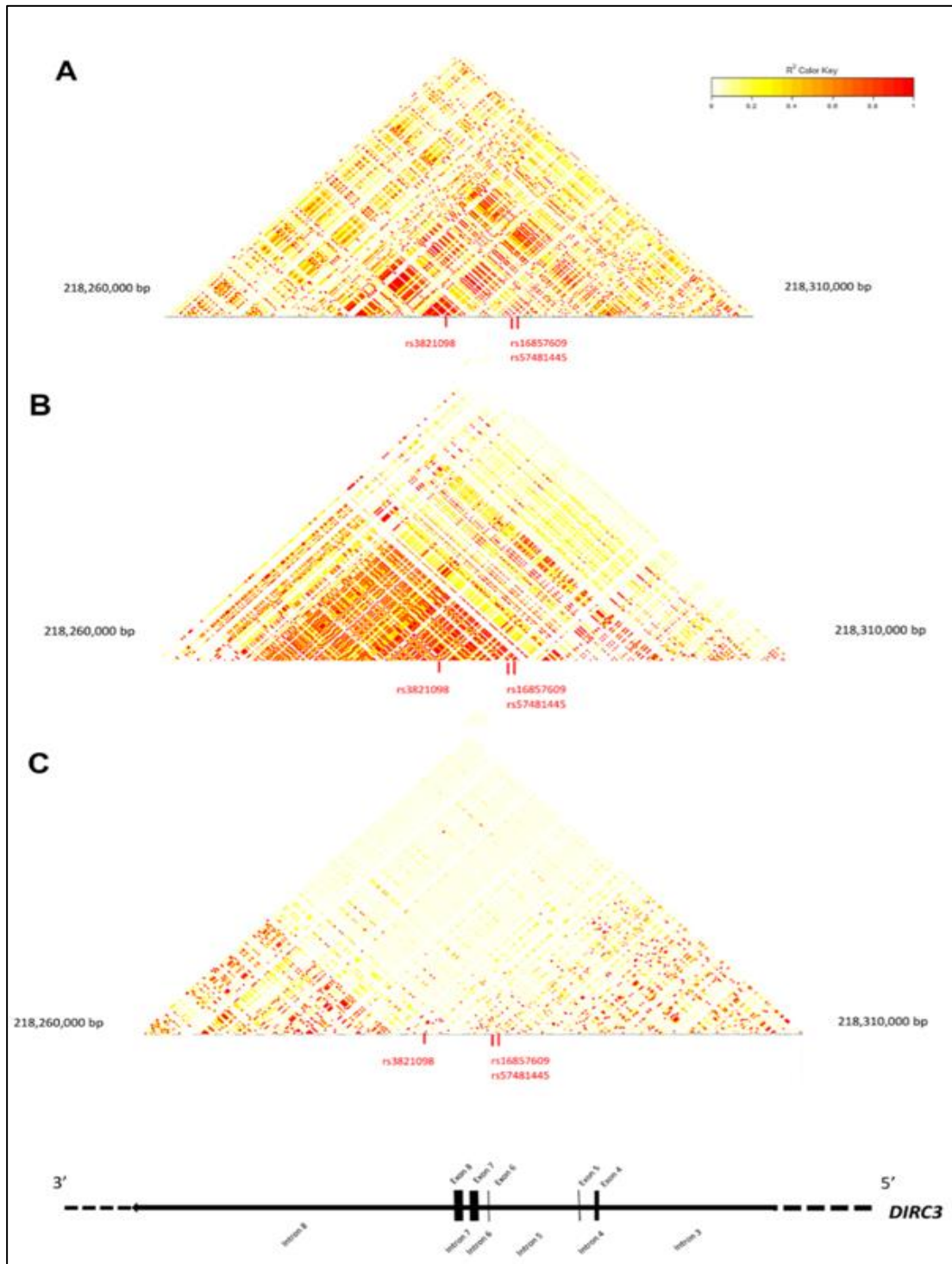
Axe Y : $-\log_{10}(p\text{-valeurs})$; Axe X : position des SNPs et des gènes sur le chromosome. La couleur de chaque SNP reflète son r^2 . Le SNP le plus associé est indiqué en violet. Les points gris correspondent aux SNPs pour lesquels le r^2 est manquant dans 1000Genomes.

Figure 20 : Locuszoom de l'association entre les SNPs des régions 2q35 et 8p12 et le CDT et résultats de la régression conditionnelle dans l'analyse poolée des Européens, Mélanésiens et Polynésiens. (a) Association au locus 2q35, (b) Association conditionnellement au SNP rs3821098 au locus 2q35, (c) Association au locus 8p12, (d) Association conditionnellement au SNP rs28406305.



Axe Y : $-\log_{10}(\text{p-valeurs})$; Axe X: position des SNPs et des gènes sur le chromosome. La couleur de chaque SNP reflète son r^2 avec le SNP le plus associé indiqué en violet, calculé dans la population européenne (CEU) de 1000Genomes. Les points gris correspondent aux SNPs pour lesquels le r^2 est manquant dans 1000Genomes.

Figure 21 : Cartes de corrélation génétiques (r^2) entre les SNPs du locus 2q35 chez (a) les Européens, (b) les Mélanésiens et (c) les Polynésiens.



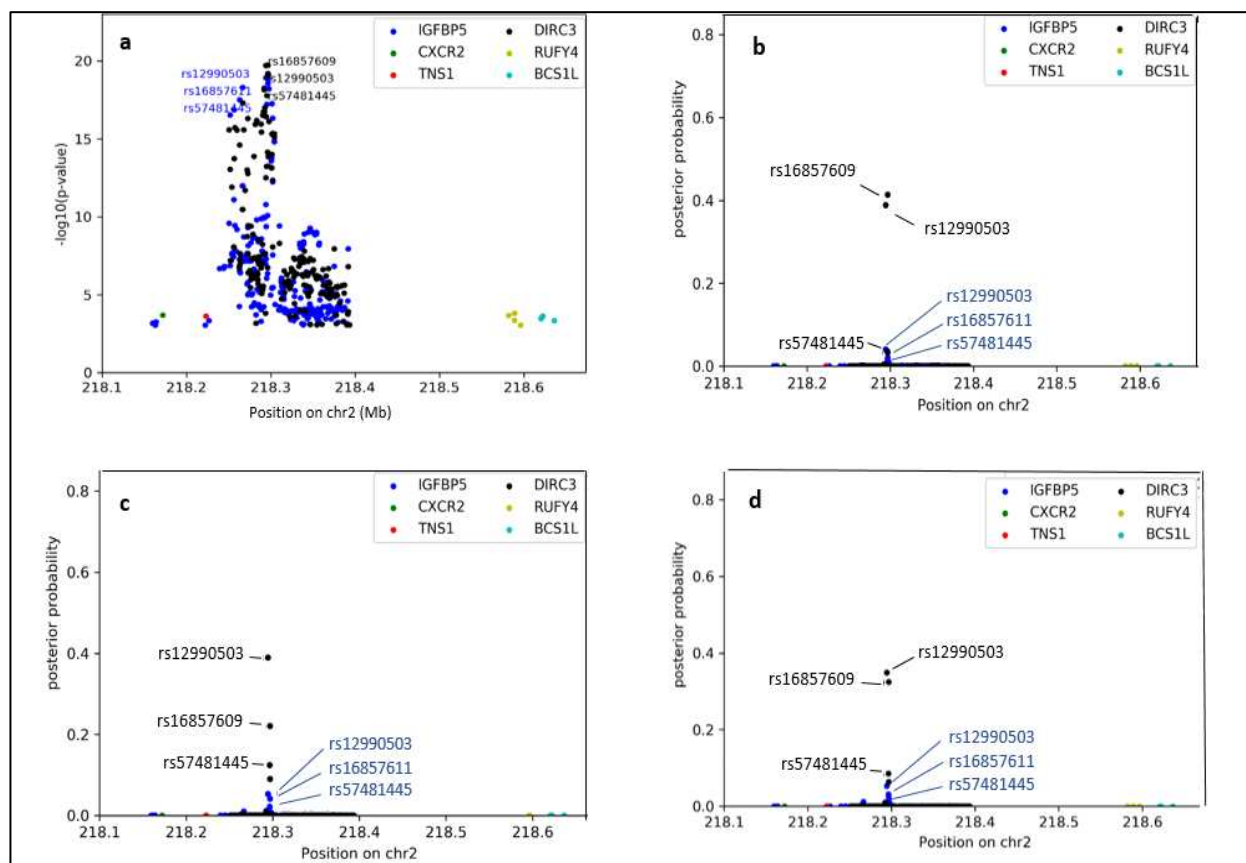
Région 2q35 (chr2 :218 260 000 - 218 310 000) indiquant le sens 5' et 3' de DIRC3. Les exons et introns sont numérotés en fonction du transcrit NR_026597.2. Un bloc de DL est ici défini comme un ensemble de SNPs ayant un $r^2 > 0.6$ (couleur orange à rouge sur les heatmaps).

ii. Analyses de colocalisation

Chez les Européens, l'analyse de colocalisation a permis de mettre en évidence deux variants introniques de *DIRC3* ayant les valeurs de PP les plus élevées : **rs16857609** ($PP_{\text{Européens}} = 0,41$) et **rs12990503** ($PP_{\text{Européens}} = 0,39$) (Tableau 7 et Figure 22 b). Ces deux SNPs sont fortement corrélés à rs57481445 ($r^2 > 0.9$) chez les Européens et sont fortement corrélés entre eux dans les trois populations ($r^2 > 0,88$).

Chez les Mélanésien et les Polynésien, les PP les plus élevées sont observées pour ces deux mêmes SNPs : pour **rs16857609**, $PP_{\text{Mélanésien}} = 0,22$ et $PP_{\text{Polynésien}} = 0,32$ et pour **rs12990503**, $PP_{\text{Mélanésien}} = 0,39$ et une $PP_{\text{Polynésien}} = 0,35$ (Figures 22 c et d). De plus, d'après les données de TCGA, les deux SNPs sont fortement associés à l'expression de *DIRC3* ($p\text{-valeur} = 1,9 \times 10^{-20}$ et $p\text{-valeur} = 1,9 \times 10^{-20}$) *IGFBP5* (*insulin growth factor binding protein 5*) ($p\text{-valeur} = 2,8 \times 10^{-19}$ et $p\text{-valeur} = 1,2 \times 10^{-19}$) dans le tissu tumoral de patients ayant développé un CPT (Figure 22 a et Tableau 7).

Figure 22 : Plots représentant les (a) $-\log_{10}$ (p-valeurs) d'association entre les SNPs au locus 2q35 et l'expression de différents gènes dans les cellules tumorales de la thyroïde (CPT), et résultats de probabilités postérieures obtenues pour chaque SNP dans l'analyse de colocalisation chez (b) les Européens, (c) les Mélanésien et (d) les Polynésien.



Annotation fonctionnelle des variants

Rs16857609 et rs57481445 sont distants de moins de 240 bases et sont dans le même bloc de DL dans les 3 groupes ethniques (Figure 21). Ils sont situés dans des sites de modification des histones, notamment des sites d'acétylation et de méthylation sur la sous-unité protéique de l'histone H3 et altèrent des régions régulatrices consensus pouvant être impliquées dans la transcription de gènes (enhancers) exprimés dans les adipocytes. En outre, d'après la base de données Roadmap, le SNP rs16857609 est situé dans des sites cibles hypersensibles à la DNase dans trois types cellulaires différents (ovaire, cœur fœtal et cellules pluripotentes) et

rs12990503 est situé dans une région de la chromatine au niveau d'un enhancer faible du gène *DIRC3* exprimé dans les adipocytes.

Ainsi, des trois variants (rs16857609, rs12990503 et rs57481445) mis en évidence dans les analyses de colocalisation, rs16857609 paraît être le plus intéressant d'après les annotations fonctionnelles que nous avons utilisées.

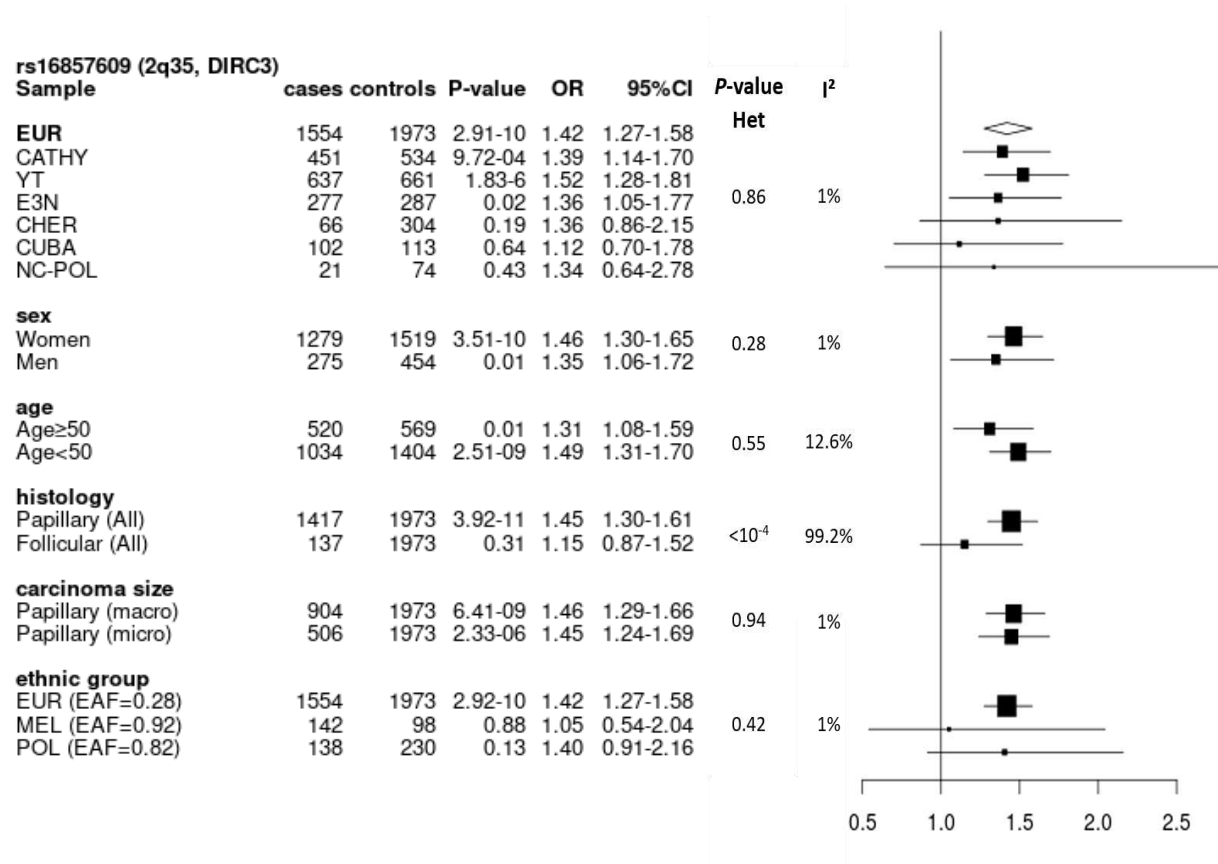
iii. Analyses stratifiées pour rs16857609

Afin d'étudier plus spécifiquement l'effet de rs16857609, j'ai réalisé des analyses stratifiées par sexe, classe d'âge, type histologique, et pour les CPT, selon la taille de carcinomes (< ou $\geq$ à 10 mm). J'ai également comparé l'association entre rs16857609 et CDT entre les différents groupes ethniques (Figure 23).

Cette analyse montre que ce SNP est plus spécifiquement associé au risque de développer un CPT plutôt qu'un CVT (Test d'hétérogénéité : $I^2 = 99,2\%$, p -valeur $< 10^{-4}$).

Aucune hétérogénéité n'est détectée entre ethnies. Bien que l'association soit non significative chez les Polynésiens et les Mélanésiens, les OR associés à ce SNP vont dans le même sens dans les trois populations. L'allèle à risque [T] est plus fréquent chez les Océaniens (fréquence de l'allèle = 0,92 et 0,82 chez les Mélanésiens et les Polynésiens, respectivement) que chez les Européens (fréquence de l'allèle = 0,28) (Figure 23).

Figure 23 : Forest plots des analyses par sous-groupe pour le SNP rs16857609 (allèle à risque : T) chez les Européens, et par groupe ethnique.



EAF, fréquence de l'allèle à risque ; EUR, Européens ; MEL, Mélanésiens ; POL, Polynésiens ; Het, hétérogénéité. OR, Odds ratio ; I², pourcentage d'hétérogénéité.

2. Locus 8p12

i. Test d'association univarié et analyses conditionnelles

Chez les Européens, le SNP le plus associé au CDT au locus 8p12 est rs28406305 (OR = 1,32, p -valeur = $5,9 \times 10^{-8}$) (Figure 24 a) et 37 autres SNPs sont associés au seuil p -valeur $< 10^{-7}$. L'analyse conditionnelle sur ce SNP ne met en évidence aucun autre signal supplémentaire (Figure 24 b) suggérant l'existence d'un seul variant causal dans cette région. Chez les Européens rs28406305 est modérément corrélé au SNP rs6996585 mis en évidence dans l'étude

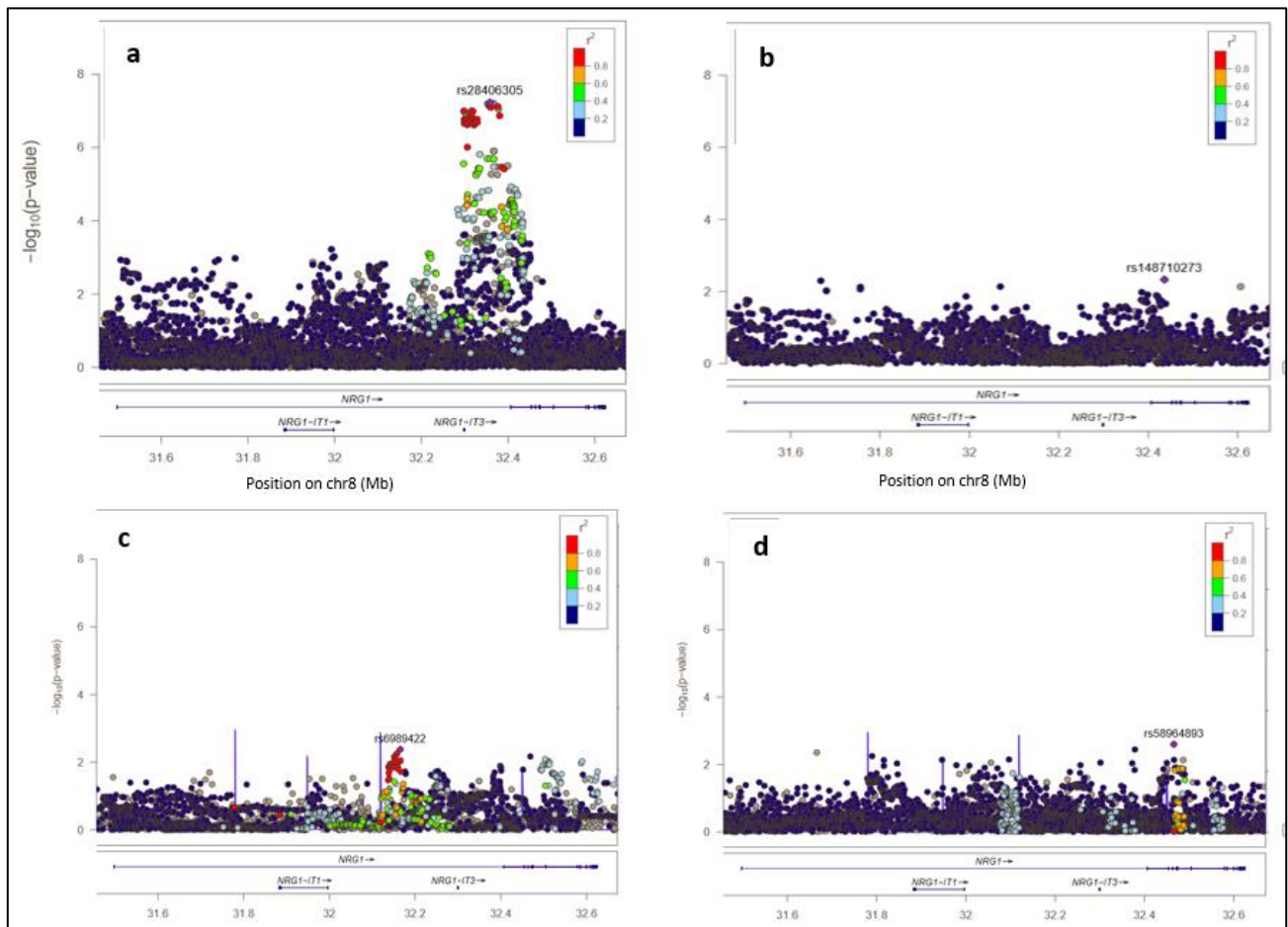
GWAS menée dans la population coréenne ($r^2 = 0,35$) (Son et al. 2017). L'analyse conditionnelle sur ce SNP ne détecte aucun signal supplémentaire dans cette région (Figure 24 b).

Chez les Mélanésien (Figure 24 c) et les Polynésien (Figure 24 d) aucun SNP n'est significativement associé au risque de CDT. Si une association est suggérée pour le SNP rs6989422 chez les Mélanésien (OR = 2,06, p -valeur = $4,1 \times 10^{-3}$) et pour le SNP rs58964893 chez les Polynésien (OR = 1,71, p -valeur = $2,5 \times 10^{-3}$), aucun de ces SNPs n'est corrélé avec rs28406305 associé au CDT chez les Européen ($r^2 < 0,01$).

L'analyse poolée des trois groupes ethniques met également en avant le SNP rs28406305 (OR = 1,30, p -valeur = $4,1 \times 10^{-8}$) (Figure 20 c).

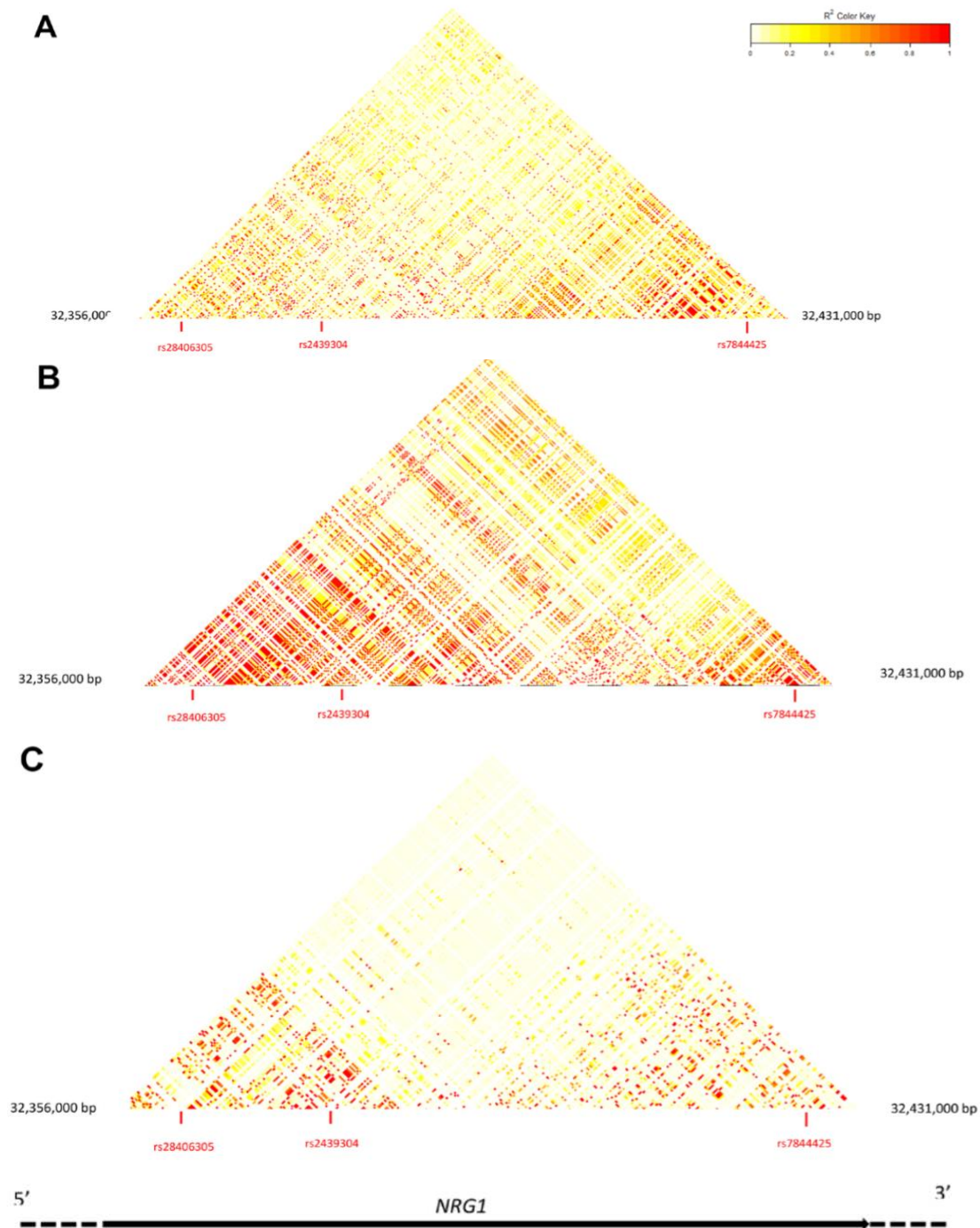
Les structures de DL au locus 8p12 dans les trois populations diffèrent avec des blocs de DL plus longs chez les Mélanésien comparés aux Européen et aux Polynésien (Figure 25).

Figure 24 : Locuszoom de l'association entre les SNPs de la région 8p12 et le CDT chez (a) les Européens, (c) les Mélanésiens et (d) les Polynésiens et de (b) l'association conditionnellement au SNP rs28406305 chez les Européens.



Axe Y : $-\log_{10}(p\text{-valeurs})$; Axe X : position des SNPs et des gènes sur le chromosome. La couleur de chaque SNP reflète son r^2 avec le SNP le plus associé indiqué en violet, calculé dans la population européenne (CEU) de 1000Genomes. Les points gris correspondent aux SNPs pour lesquels le r^2 est manquant dans 1000Genomes.

Figure 25 : Cartes de corrélation génétiques (r^2) entre les SNPs du locus 8p12 chez (a) les Européens, (b) les Mélanésiens et (c) les Polynésiens.



Région 8p12 (chr8 : 32 356 000 - 32 431 000) indiquant la région d'intérêt avec *NRG1*. Un bloc de DL est ici défini comme un ensemble de SNPs ayant un $r^2 > 0.6$ (couleur orange à rouge sur les heatmaps).

ii. Analyses de colocalisation

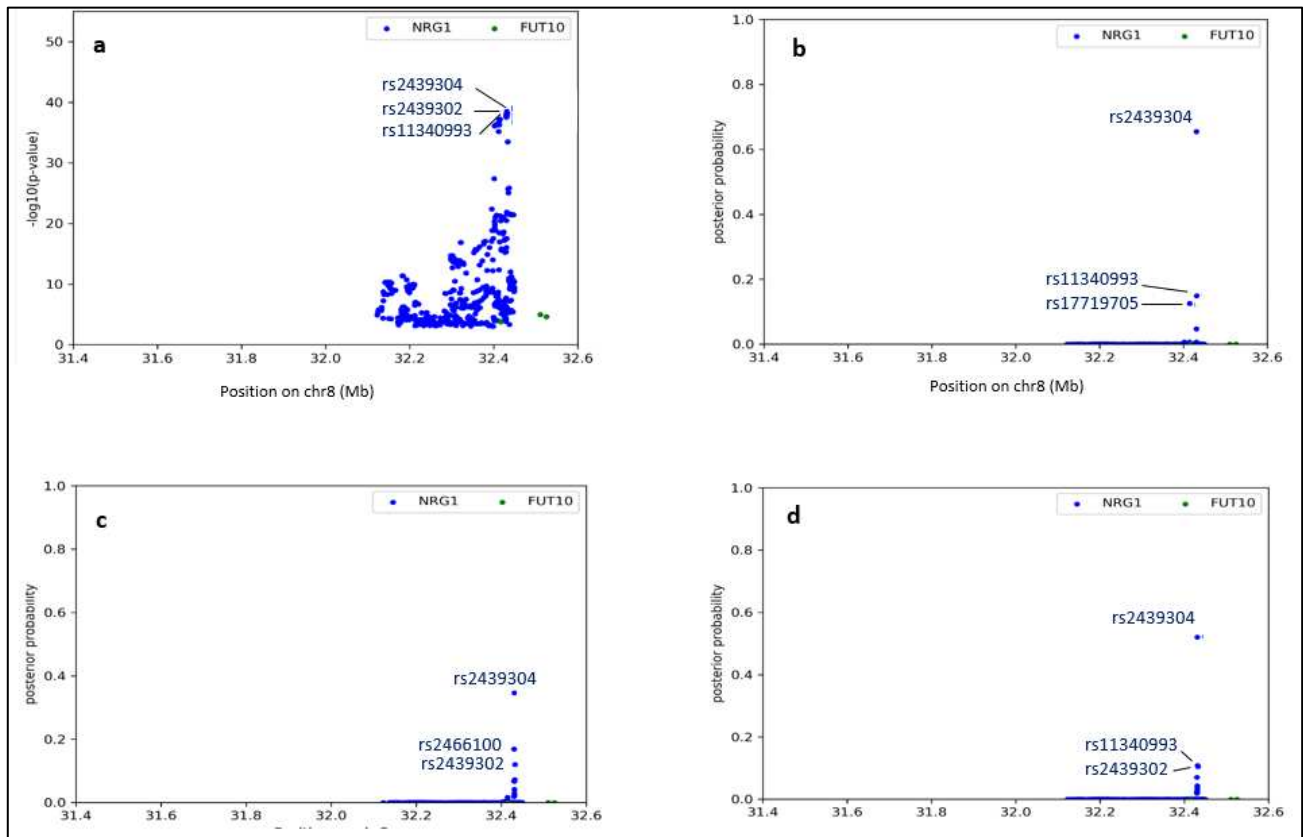
L'analyse de colocalisation estime les valeurs de PP les plus élevées pour le SNP **rs2439304** dans les trois groupes ethniques ($PP_{\text{Européens}} = 0,65$, $PP_{\text{Mélanésien}} = 0,52$ et $PP_{\text{Polynésien}} = 0,35$) (Tableau 7 et Figure 26). De plus **rs2439304** est un eQTL fortement associé à l'expression de *NRG1* dans les cellules tumorales thyroïdiennes avec p -valeur = $3,1 \times 10^{-39}$ dans PancanQTL (Tableau 7 et Figure 26). Cependant, notons que ce SNP est peu corrélé avec rs28406305 qui est le SNP le plus fortement associé au CDT chez les Européens d'EPITHYR ($r^2 = 0,5$) et dont la PP est très faible ($PP_{\text{Européens}} = 2,1 \times 10^{-4}$). Dans les trois groupes ethniques, rs2439304 est fortement corrélé avec le SNP rs2439302 identifié dans les précédents GWAS sur les CDT ($r^2 = 1$) (Gudmundsson et al. 2012), mais n'est pas significatif dans EPITHYR ($OR_{\text{Européens}} = 1,20$, p -valeur = $2,6 \times 10^{-4}$; $OR_{\text{Polynésien}} = 1,10$, p -valeur = 0,55 ; $OR_{\text{Mélanésien}} = 1,05$, p -valeur = 0,79).

Tableau 7 : Résultats de l'analyse de colocalisation des SNPs ayant une probabilité postérieure (PP) > 0,10 aux loci 2q35 et 8p12 chez les Européens, les Polynésiens et les Mélanésiens.

Locus	SNP	AR ^a	EUROPEENS				POLYNESEIENS				MELANESIENS				eQTL	
			Freq ^c	OR ^d	p^e	PP ^f	Freq	OR	p	PP	Freq	OR	p	PP	p -valeur	Gènes associés
2q35	rs16857609	T	0,28	1,42	$3,7 \times 10^{-10}$	0,41	0,82	1,42	0,12	0,32	0,92	1,05	0,88	0,22	$1,9 \times 10^{-20}$	<i>DIRC3</i>
	rs12990503	C	0,28	1,41	$3,4 \times 10^{-10}$	0,39	0,17	1,48	0,09	0,35	0,08	0,90	0,75	0,39	$2,8 \times 10^{-19}$	<i>IGFBP5</i>
8p12	rs2439304 ^b	A	0,50	1,16	$1,7 \times 10^{-03}$	0,65	0,30	1,11	0,51	0,52	0,41	1,03	0,84	0,35	$3,1 \times 10^{-39}$	<i>NRG1</i>

a AR : Allèle à risque ; b info score de l'imputation (valeur entre 0 et 1) pour ce SNP est de 1 pour les trois groupes ethniques ; c Fréquence de l'allèle de risque ; d OR : odds ratio d'association avec le risque de CDT ; e p -valeurs de l'association avec le CDT ; f Probabilité *a posteriori* de causalité calculée avec l'analyse de colocalisation (obtenue avec les p -valeurs eQTL de *DIRC3* pour 2q35 et les p -valeurs eQTL de *NRG1* pour 8p12 de PancanQTL)

Figure 26 : Plots représentant les p -valeurs d'association des variants au locus 8p12.



Plots représentant les (a) $-\log_{10}(p\text{-value})$ d'association entre les SNPs au locus 8p12 et l'expression de différents gènes dans les cellules tumorales de la thyroïde (CPT), et résultats de probabilités postérieures obtenues pour chaque SNP dans l'analyse de colocalisation chez (b) les Européens, (c) les Mélanésien et (d) les Polynésien.

Annotation fonctionnelle de rs2439304 et rs28406305

Rs2439304 est situé dans la région promotrice de *NRG1* et altère de nombreux sites promoteurs de modifications d'histones et de marqueurs d'états de la chromatine dans plusieurs tissus notamment dans les adipocytes. En particulier, rs2439304 est situé dans un site de liaison de l'histone H3 activant la méthylation de l'ADN dans ces cellules.

Le top SNP rs28406305 identifié chez les Européens est situé dans deux sites consensus de méthylation sur la sous-unité protéique des histones H3 exprimés dans les cellules adipocytaires

et dans les cellules hépatiques. Aussi, je me suis intéressée à l'intérêt biologique des SNPs du locus 8p12 associés au CDT et fortement corrélés à rs28406305 et le SNP rs7844425 ($r^2 = 0,96$, OR = 1,2, p -valeur = $7,6 \times 10^{-8}$) apparaît comme le SNP ayant un fort effet délétère. En effet, d'après les données de RegulomeDB et Haploreg V4, ce SNP est situé dans un site de méthylation et deux sites d'acétylation sur la sous-unité protéique des histones H3 au sein de multiple tissus. Surtout, d'après les données PanCanQTL, rs7844425 est fortement associé à l'expression de *NRG1* dans les tissus thyroïdiens avec une p -valeur = $2,3 \times 10^{-17}$. Malgré son intérêt fonctionnel important, ce SNP n'a pas été mis en évidence dans l'analyse de colocalisation avec des faibles PP associées dans les trois populations ($PP_{\text{Européens}} = 8,8 \times 10^{-28}$, $PP_{\text{Mélanésien}} = 3,4 \times 10^{-27}$ et $PP_{\text{Polynésien}} = 3,4 \times 10^{-27}$).

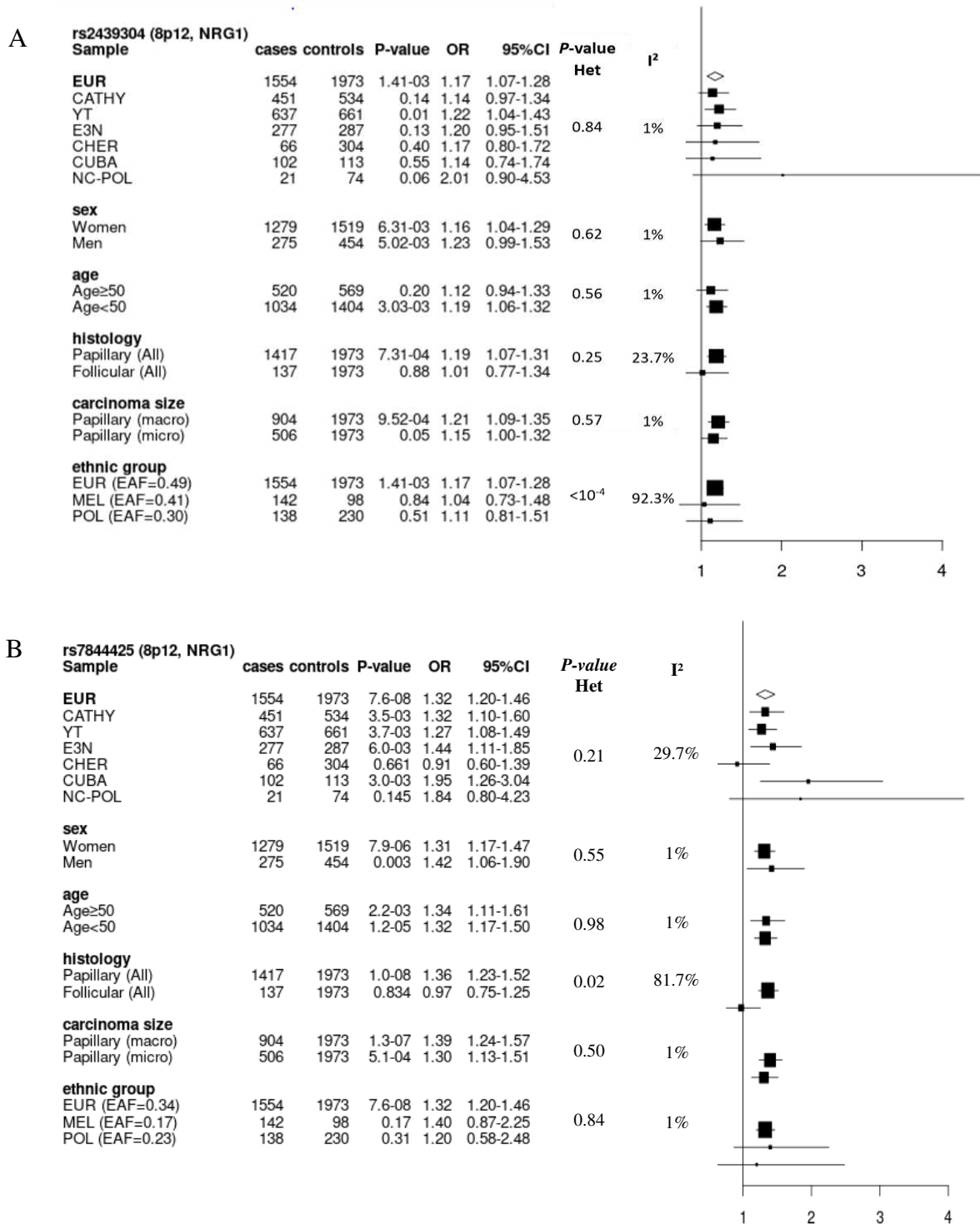
iii. Analyses stratifiées pour rs2439304 et rs7844425

Les analyses par sous-groupe pour rs2439304 ne montrent pas d'hétérogénéité entre études, sexes, classes d'âge, types histologiques ou taille du CPT chez les Européens d'EPITHYR.

En revanche, nous observons une hétérogénéité significative ($I^2 = 92,3\%$, p -valeur $< 10^{-4}$) dans l'analyse stratifiée par population (Figure 27 A) qui n'est pas due à une mauvaise qualité d'imputation du SNP dans l'une des populations (info score $> 0,99$ dans les 3 populations).

Nous avons également effectué des analyses par sous-groupe pour rs7844425, qui était le SNP le plus intéressant d'un point de vue fonctionnel associé au CDT, d'après les résultats de Regulome DB et Haploreg V4. Chez les Européens, nous avons observé une hétérogénéité significative par type histologique de CPT ($I^2 > 81\%$, p -valeur = 0,02), avec une association positive observée uniquement pour le CPT (Figure 27 B).

Figure 27 : Forest plots des analyses par sous-groupes pour A) le SNP rs2439304 (allèle à risque : A) chez les Européens, les Mélanésiens et les Polynésiens), B) le SNP rs7844425 5(allèle à risque : A) chez les Européens, les Mélanésiens et les Polynésiens).



EAF, fréquence de l'allèle à risque ; EUR, Européens ; MEL, Mélanésiens ; POL, Polynésiens ; Het, hétérogénéité. Papillary (macro), carcinome papillaire de la thyroïde $\geq 10\text{mm}$; Papillary (micro), carcinome papillaire de la thyroïde $< 10\text{mm}$.

IV. Discussion

Notre étude a permis de prioriser des variants de susceptibilité au CDT dans les régions 2q35 et 8p12 en utilisant les données d'une population multiethnique incluant des Européens, des Mélanésiens et des Polynésiens.

Au locus 2q35, nous avons identifié le SNP rs16857609 comme un SNP potentiellement causal dans les trois groupes de population. Ce SNP est associé à l'expression des gènes *DIRC3* et *IGFBP5* dans les cellules tumorales de la thyroïde, selon les données de PancanQTL. *DIRC3* est un gène non codant et *IGFBP5* code la protéine 5 de liaison au facteur de croissance analogue à l'insuline qui fait partie de la famille des « facteurs de croissance insulino-mimétiques » (IGF). Les IGF jouent un rôle essentiel dans l'homéostasie des tissus, en régulant la prolifération, la différenciation et la migration des cellules au cours du développement et également à l'âge adulte. Les données d'annotation fonctionnelle nous ont permis de montrer que rs16857609 est situé dans un site de régulation de la sous-protéine de l'histone H3 dans plusieurs tissus. Nous pouvons émettre l'hypothèse de l'implication de ce SNP dans la régulation de l'expression de gènes via l'altération de la décondensation de la chromatine dans les adipocytes bloquant ainsi la liaison entre l'enhancer et le promoteur du gène. La liaison entre ces deux séquences régulatrices de l'ADN est essentielle pour initier la transcription et activer l'expression du gène. Ces résultats sont intéressants car l'obésité a été systématiquement associée à un risque accru de CDT dans les études épidémiologiques (Lauby-Secretan et al. 2016). Nos observations suggèrent que le variant rs16857609 joue un rôle dans l'étiologie du CDT par le biais de voies complexes associées à l'obésité.

Dans EPITHYR, la fréquence de l'allèle à risque [T] de ce variant est beaucoup plus élevée chez les Océaniens que chez les Européens et pourrait donc contribuer en partie à la plus forte incidence de CDT observée chez les Océaniens si l'association est confirmée.

De manière intéressante, le variant rs16857609 a également été associé à un risque accru de cancer du sein chez les Européens (Michailidou et al. 2013). *IGFBP5* étant un régulateur physiologique de la survie des cellules épithéliales dans la glande mammaire (Marshman 2003), il a été proposé qu'une dérégulation de l'expression d'*IGFBP5* induite par des variants situés en amont de *DIRC3* (Wyszynski et al. 2018) pourrait conduire au développement d'une tumeur mammaire. L'ensemble de ces résultats suggère qu'une dérégulation de l'expression d'*IGFBP5* pourrait également conduire à la prolifération des cellules thyroïdiennes et finalement au développement du CDT.

L'analyse de fine-mapping au locus 8p12 nous a permis d'identifier le SNP rs7844425 comme étant le variant le plus susceptible d'avoir un impact fonctionnel délétère. En effet, nous avons observé que ce SNP est situé dans la région de transcription en amont des principales isoformes du gène *NRG1* exprimées dans le tissu thyroïdien et les lignées cellulaires thyroïdiennes (NM_004495, NM_013958 et NM_001160008). Ce gène code une glycoprotéine membranaire qui sert de médiateur à la signalisation cellulaire et qui joue un rôle essentiel dans la croissance et le développement de nombreux organes.

À notre connaissance, aucune étude GWAS n'a mis en évidence d'association entre le variant rs7844425 ou d'autres SNPs qui seraient situés dans le même bloc de DL. Dans notre population d'étude, ce SNP est fortement corrélé avec le SNP rs17716295 chez les Européens ($r^2 = 0,93$). Rs17716295 a été associé à la réponse au traitement antipsychotique chez les patients schizophrènes (Jajodia et al. 2016) et de manière intéressante, il a été observé que les perturbations de l'équilibre des hormones thyroïdiennes contribuent à la dérégulation de l'état mental, avec une proportion plus élevée d'hypothyroïdie chez les patients atteints de schizophrénie (Sharif et al. 2018). Un lien est ainsi possible entre rs7844425, ou d'autres variants à ce locus, et la dérégulation des hormones thyroïdiennes.

L'analyse de colocalisation a également mis en évidence le SNP rs2439404 au locus 8p12 qui est modérément corrélé avec le SNP rs2439304 ($r^2 = 0,51$ chez les Européens). Bien que rs2439304 ne soit pas significativement associé au risque de CDT dans nos populations (OR = 1,2, p -valeur = 0,001 chez les Européens), il présente la probabilité *a posteriori* de

causalité la plus élevée des SNPs testés dans les trois populations. Rs2439304 est fortement corrélé à rs2439302 mis en évidence dans une autre GWAS sur le CDT ($r^2 = 1$) (He et al. 2018). Ces deux SNPs sont situés dans un bloc de DL d'une longueur de 32kb dans l'intron 1 de *NRG1*. He et coll. ont analysé les haplotypes de ce bloc de DL et leurs travaux suggèrent que la régulation transcriptionnelle de *NRG1* pourrait être associée à la combinaison de plusieurs allèles à risque et conférer une susceptibilité au CDT. Aussi, en utilisant une approche CHIP-seq et le test de la luciférase sur des cellules thyroïdiennes primaires en culture, les auteurs ont montré que les allèles à risque rs2439304[A] et rs2439302[G] augmentent l'expression des trois principales isoformes de *NRG1*. Ces observations sont semblables à celles retrouvées dans l'étude GWAS coréenne sur le CDT qui suggère que l'expression élevée de *NRG1* dans le tissu thyroïdien pourrait influencer le développement ou la progression de ce cancer (Son et al. 2017). Cependant, ces résultats sont contradictoires avec ceux de Gudmundsson et coll. qui ont montré une expression plus faible de *NRG1* chez les porteurs de l'allèle à risque rs2439302[G] (Gudmundsson et al. 2009).

L'aspect trans-ethnique est un atout de l'étude EPITHYR. En effet, les méta-analyses trans-ethniques qui combinent les résultats d'un même trait dans des populations génétiquement diverses peuvent aider à la cartographie fine en capitalisant sur les différences ethniques dans les patrons de DL (Li and Keating 2014). Notre étude de fine-mapping a pour avantage d'étudier les populations polynésiennes et mélanésiennes qui ont les taux d'incidence de CDT parmi les plus élevées dans le monde.

Dans notre population d'étude, les blocs de DL des loci 2q35 et 8p12 diffèrent avec des blocs de DL plus longs chez les Mélanésien et Polynésien comparés aux Européens mais les variants mis en évidence sont dans les mêmes blocs de DL dans les trois groupes ethniques. Pour exploiter au mieux l'aspect multi-ethnique d'EPITHYR, j'ai exploré des méthodes de méta-analyses trans-ethniques telles que Metasoft (Han and Eskin 2011), mettant en œuvre une forme modifiée de l'approche traditionnelle à effets aléatoires et qui est une méthode populaire parmi les études trans-ethniques (Larson et al. 2015; Wang et al. 2013). Cependant les résultats ne nous ont pas permis de mettre en évidence de SNPs causaux dans les trois populations en raison des effectifs limités. J'ai également exploré l'utilisation d'une méthode bayésienne, mise

en œuvre dans le logiciel PAINTOR (Kichaev et al. 2017). Cette méthode permet d'associer deux sources de données pour estimer les probabilités de causalité spécifiques aux variants : la pertinence fonctionnelle génomique (avec les annotations fonctionnelles) et l'association génotype-phénotype. Cette méthode permet d'analyser simultanément plusieurs loci et/ou populations en prenant en compte la structure de DL spécifique à chaque groupe. PAINTOR propose d'utiliser les DL calculés sur la population de 1000Genomes. Si l'application de PAINTOR à la population européenne a pu être faite avec les données européennes (EUR) de 1000Genomes, l'utilisation de PAINTOR a posé des problèmes pour l'analyse des populations polynésiennes et mélanésiennes pour lesquelles j'ai essayé de forcer les DL calculés dans la population d'EPITHYR. Ce logiciel était relativement nouveau à l'époque où j'ai débuté les analyses et certaines des fonctionnalités proposées étaient encore en phase de développement. Malgré les échanges avec Dr. Kichaev, auteur de PAINTOR, je n'ai pas pu résoudre les problèmes techniques rencontrés pour l'analyse des données d'EPITHYR pour finaliser ces analyses.

Une autre limite de nos analyses est l'utilisation de PancanQTL qui est une base d'expression de gènes exprimés dans les tissus tumoraux et non dans les tissus sains pour lesquels les eQTL pourraient être différents. Aussi, PancanQTL fournit des eQTL des individus d'origine européenne. L'extrapolation aux populations océaniques pourrait être inexacte. C'est d'ailleurs le manque de populations de référence pour les populations polynésiennes et mélanésiennes qui ne m'a pas permis d'exploiter efficacement l'approche multi-ethnique dans mes analyses.

Nous avons confirmé les SNPs de prédisposition au CDT précédemment identifiés chez les Européens et les Asiatiques aux loci 2q35 et 8p12 et l'étude de cartographie fine a permis d'identifier de nouveaux SNPs potentiellement causaux dans chaque locus (rs16857609 à 2q35, rs7844425 et rs2439304 à 8p12) à privilégier dans de futures études fonctionnelles.

Chapitre 3

Etude des régions homozygotes du génome et recherche de facteurs récessifs

I. Introduction

Le risque de CDT plus élevé chez les sujets ayant un membre de la fratrie atteint que chez les sujets ayant un parent atteint (Hemminki, Sundquist, and Lorenzo Bermejo 2008) peut évoquer un mode de transmission récessif. Dans ce cas, le risque de CDT augmenterait avec le niveau de consanguinité de la population. Un rôle des variants récessifs peut ainsi expliquer également les incidences plus élevées de CDT observées dans les populations insulaires dans lesquelles les taux de consanguinité peuvent être plus élevés (Bener et al, 2009).

Les études GWAS manquant souvent de puissance pour détecter les variants récessifs (Lettre, Lange, and Hirschhorn 2007), une autre stratégie consiste à analyser les segments d'ADN homozygotes par descendance (ou *Homozygous By Descent*, HBD) associés au risque de CDT. Lorsque les données de pedigrees ne sont pas disponibles, les segments HBD ne sont pas directement observables mais peuvent être identifiés en recherchant des régions d'homozygotie (ou *Runs Of Homozygosity*, ROH) dépassant une longueur donnée (Todd Lencz et al, 2007; McQuillan et al, 2008). Il a été montré que les longs ROH sont des régions du génome enrichies en variants délétères (Szpiech et al. 2019). Ainsi l'étude de leur rôle dans les maladies

complexes suscite de plus en plus d'intérêt ces dernières années (Ceballos et al 2018). Par ailleurs, la proportion de ROH dans le génome peut être utilisée pour estimer le coefficient de consanguinité d'un individu (McQuillan et al 2008). Ainsi, en contrastant la présence, le nombre, la taille des ROH et les coefficients de consanguinité chez les cas et les témoins pour une maladie donnée, il est possible d'évaluer le rôle de variants récessifs dans le risque de cette maladie.

Les **Runs of Homozygosity (ROH)** ou segments homozygotes sont définis comme étant des segments du génome sur lesquels une série de marqueurs polymorphes (SNPs, microsatellites) consécutifs présentent chez un individu un génotype homozygote (allèles identiques sur chaque brin du chromosome). Un individu est considéré comme **consanguin** lorsqu'il est le fruit d'une union de deux individus apparentés. Chez les individus consanguins, les allèles homozygotes et identiques par descendance (IBD, *Identity By Descent*) (i.e. issus d'un ancêtre commun), sont dits **homozygotes par descendance** (ou *Homozygous By Descent*, **HBD**) ou autozygotes. Les loci HBD d'un individu consanguin ne sont pas distribués aléatoirement sur son génome, mais sont regroupés par blocs ou **segments HBD**, dont la taille et le nombre fluctuent en fonction des processus de recombinaison qui sont survenus depuis les ancêtres communs. Ainsi, chaque individu consanguin a une proportion HBD du génome qui lui est propre et qui reflète le processus de recombinaison depuis les ancêtres communs de ses parents. Dans les études GWAS, ces segments HBD peuvent être analysés indirectement en recherchant des régions d'homozygotie (ou *runs of homozygosity*, ROH) dépassant une longueur donnée.

Des ROH ont été associés à des risques plus élevés de développer différents types de cancers tel que le cancer colorectal, le cancer du poumon ou encore le cancer du sein dans des populations européennes (Bacolod et al. 2015; Orloff et al. 2012; Thomsen et al. 2015). A ce

jour, seule une étude menée sur une population italienne a étudié le lien entre ROH et risque de CDT (Thomsen et al. 2016). Dans cette étude, les auteurs ont mis en évidence un niveau de consanguinité plus élevé ainsi qu'un nombre plus élevé de ROH chez les cas que chez les témoins. Cependant ces résultats n'ont pas été répliqués dans un échantillon européen indépendant et aucune autre étude n'a étudié l'association entre ROH et le risque de CDT dans les populations insulaires océaniques. Une association entre la consanguinité et le CPT a également été mise en évidence dans une étude cas-témoin réalisée en Jordanie sur 200 patients atteints de CPT et 515 témoins qui montre une association entre **la consanguinité** parentale (cousins germains, cousins au second degré) et le risque de CPT (Zayed et al. 2021).

Notre objectif était d'évaluer le rôle des variants récessifs dans le risque de CDT, en analysant l'association entre les caractéristiques des ROH et le risque de CDT dans les populations d'EPITHYR et d'EPIC. Nous avons également caractérisé les ROH partagés par plusieurs sujets dans notre population d'étude pour évaluer leur association avec le risque de CDT.

II. Matériel et méthodes

1. Données

EPITHYR

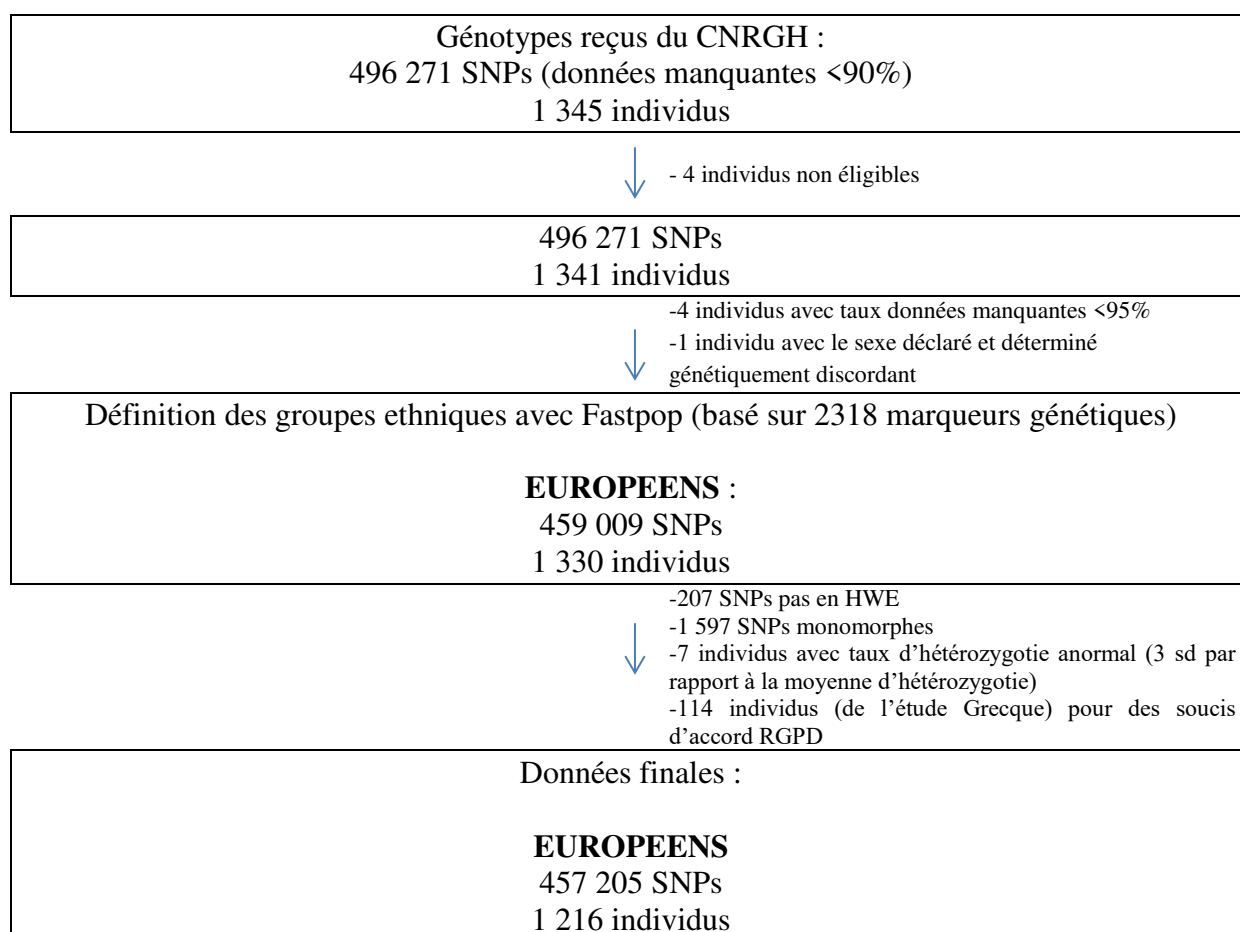
Pour ces analyses j'ai mis à jour les données du GWAS d'EPITHYR avec les données d'une deuxième étude menée en Polynésie française pour laquelle la même puce de génotypage a été utilisée. Les données d'EPITHYR ont été mises à jour avec des exclusions de sujets (histologie manquante, doublon, etc) et en incluant cette dernière étude. Les mêmes étapes de contrôle qualité ont alors été effectuées sur les données génétiques (cf Données de génotypage, parties II et IV).

EPIC

Dans cette partie nous utiliserons également les données européennes d'EPIC décrites précédemment (page 54). Les différentes étapes du contrôle qualité des données de génotypage d'EPIC sont détaillées dans la Figure 28.

Des individus E3N ont été inclus dans EPITHYR et dans EPIC. L'absence de doublons inclus dans les deux consortia a été vérifiée lors de la sélection des individus à génotyper en amont des analyses ROH et a également été contrôlé lors du QC sur les données génotypées.

Figure 28 : Etapes du contrôle qualité des données génotypage d'EPIC et sélection des SNPs et des individus retenus pour les analyses.



i. Définition des groupes de population dans EPITHYR

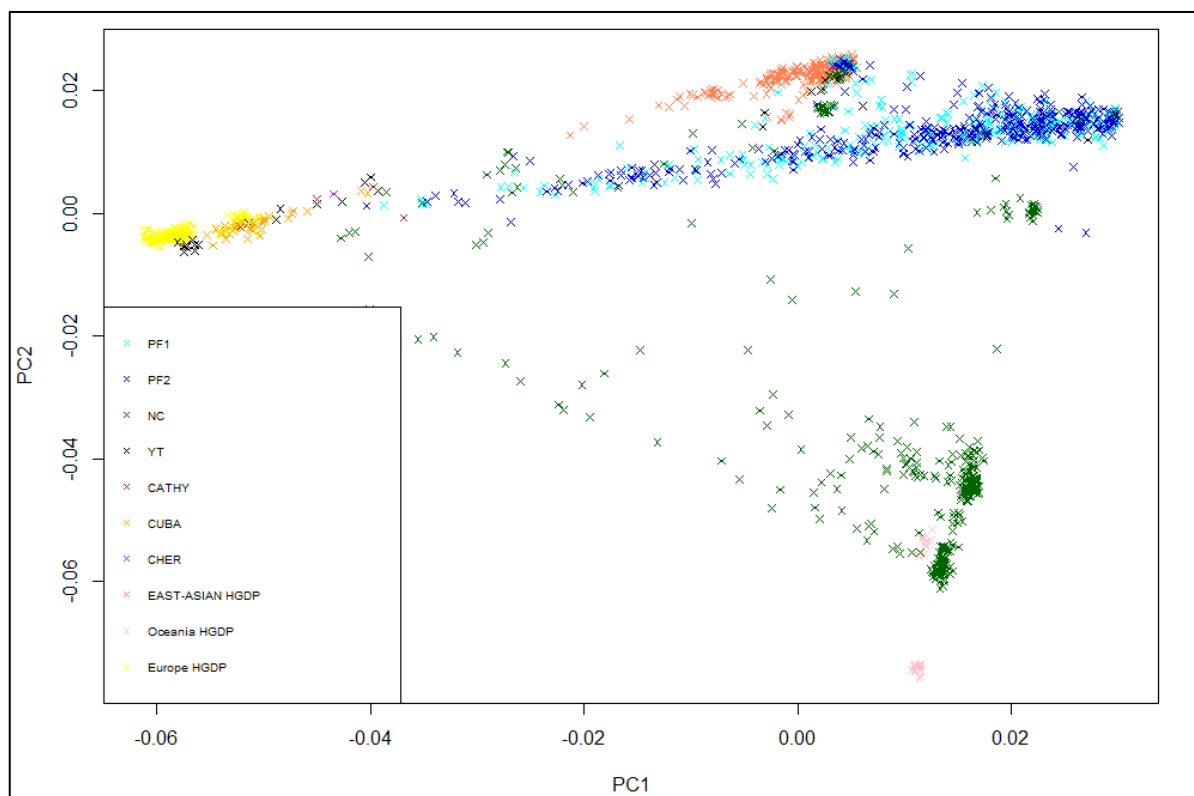
En plus des étapes de contrôle qualité appliquées pour l'analyse GWAS, j'ai dû réaliser pour l'étude des ROH des étapes supplémentaires pour définir les différents groupes de populations et estimer le taux d'apparentement des individus inclus dans les différentes études cas-témoins. Ainsi, après définition des ethnies avec Fastpop, 3 525 individus sont identifiés comme étant Européens (1 557 cas et 1 968 témoins) et 1 001 individus sont identifiés comme étant Asiatiques (Figure supplémentaire 4).

Cependant, une hétérogénéité dans la structure de population est observée pour le groupe « asiatique » d'EPITHYR. Lors des analyses GWAS et de fine-mapping nous avons pu prendre en compte cette hétérogénéité en utilisant un modèle linéaire mixte afin de tester l'association. Pour la recherche des ROH et suite à l'ajout de nouveaux individus de Polynésie française, il est important de définir des sous-groupes d'individus plus homogènes génétiquement. En effet, lors de la première définition des ethnies avec Fastpop, différents groupes ethniques avaient été créés : Européens (CEU), Asiatiques (CHB), Africains (YRI) et « Autres ». Le groupe « Autres » correspond aux individus n'ayant aucune ethnie attribuée selon les seuils définis (cf partie « Données de génotypage », Figure 8). Nous avons cherché à définir de nouveaux groupes de populations permettant d'estimer le métissage des individus à partir de données génétiques. Pour ce faire, nous avons utilisé le logiciel ADMIXTURE (Alexander and Lange 2011) avec les données génétiques de populations de référence du panel HGDP.

Dans un premier temps, les individus classés dans les groupes « Asiatiques » et « Autres » d'EPITHYR (503 cas et 612 témoins) par FastPop ont été regroupés avec les 157 Européens, 228 Asiatiques de l'Est et 26 Océaniens (Nouvelle-Guinée et Bougainville) du panel de référence Human Genome Diversity Panel (HGDP) (Cavalli-Sforza 2005).

J'ai également réalisé une ACP sur ce jeu de données afin de mieux visualiser l'hétérogénéité de cette population d'EPITHYR avant la définition des nouvelles populations avec ADMIXTURE (Figure 29).

Figure 29 : Analyse en composantes principales des individus classés dans les groupes «Asiatiques» et « Autres » d’EPITHYR par Fastpop réalisée en prenant en compte les populations européennes, asiatiques et océaniques du panel de référence HGDP.

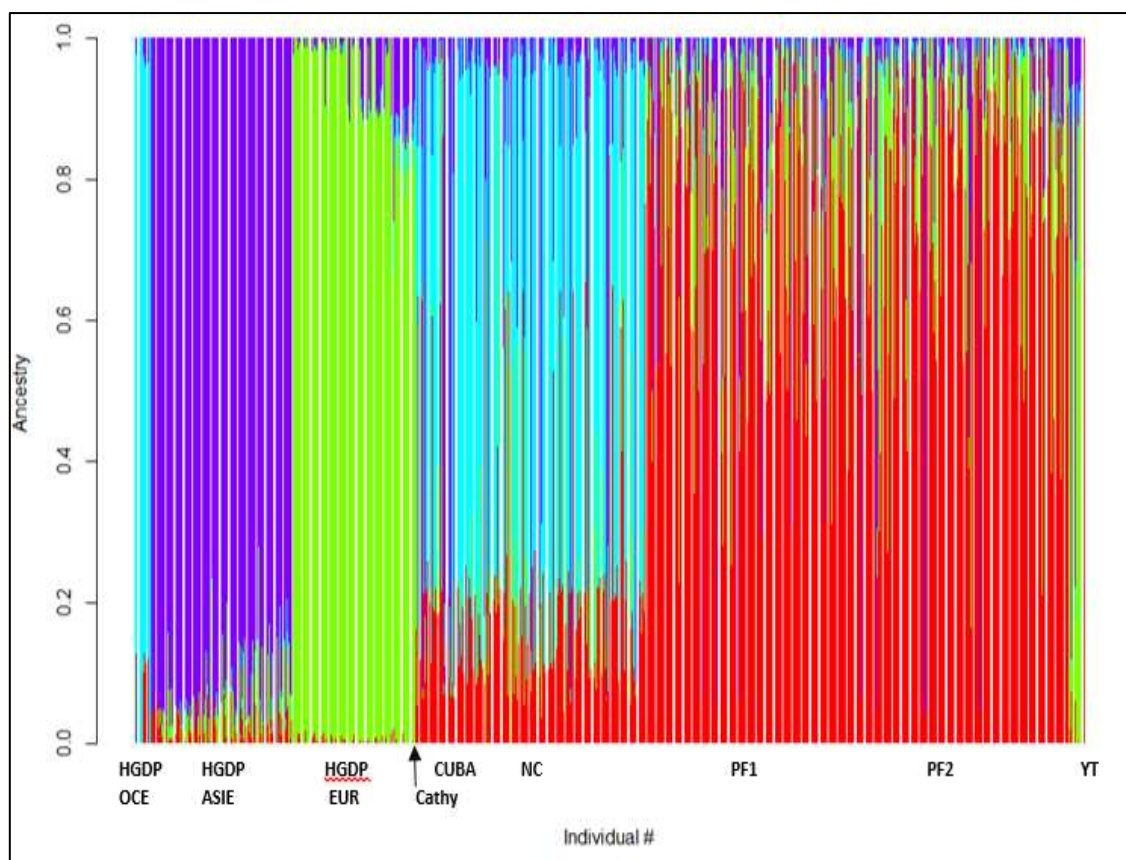


Chaque croix correspond à un individu. Les croix bleues foncées correspondent aux individus d’EPITHYR de l’étude Polynésie Française, les croix de couleur cyan aux individus de l’étude Polynésie Française 2, les croix vertes aux individus de l’étude Nouvelle-Calédonie, les croix noires aux individus de l’étude YoungThyr. Les autres couleurs correspondent aux populations du panel de référence HGDP.

L’analyse ADMIXTURE a été réalisée sur 1 527 individus (503 cas et 1 024 témoins) et 73 360 variants indépendants (exclusion des SNPs avec $r^2 > 0,8$ et une FAM < 5%) en faisant l’hypothèse que les individus d’EPITHYR analysés sont issus d’un mélange de quatre populations : européenne, asiatique, océanienne et polynésienne. Ainsi, le pourcentage de mélange de population (aussi appelé « admixture » en génétique des populations) pour chaque individu obtenu avec ADMIXTURE sont visualisables sur la Figure 30. Nous observons que

les individus des études Polynésie Française et Polynésie Française 2 sont majoritairement considérés comme Polynésiens (en rouge) et que les individus de l'étude Nouvelle-Calédonie sont eux considérés comme Océaniens (en bleu). Sur la base de ces résultats, j'ai classé les individus d'EPITHYR en différents sous-groupes de populations (Tableau 8) selon différents seuils que j'ai choisis afin d'exclure le minimum d'individus en attribuant un groupe ethnique unique à chaque individu.

Figure 30 : Histogramme représentant le mélange de population des individus d'EPITHYR précédemment classés "Asiatique" ou "Autre" en utilisant comme populations de référence les individus européens, asiatiques et océaniens d'HGDP.



En abscisse les individus des différentes populations, en ordonnées le pourcentage de métissage. Le bleu correspond aux Océaniens ; le violet aux Asiatiques, le vert aux Européens et le rouge aux Polynésiens.

Tableau 8 : Nombre de cas et témoins par étude selon le degré de métissage.

Population	N total		Nouvelle-Calédonie		Polynésie Française		Polynésie Française 2		YoungThyr		Cathy		Cuba		Tchernobyl	
	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
Asiatique ^a	28	84	13	26	6	29	5	25	3	4	1	-	-	-	-	-
Mélanésienne ^b	155	109	155	109	-	-	-	-	-	-	-	-	-	-	-	-
Polynésienne ^c	248	317	12	20	82	129	154	166	-	2	-	-	-	-	-	-
Européenne ^d	40	40	5	6	1	3	2	7	10	7	2	2	20	14	-	1
Autres ^e	32	62	6	17	12	17	14	28	-	-	-	-	-	-	-	-
Total	503	612	191	178	101	178	175	226	13	13	3	2	20	14	-	1

^a: Sont définis comme Asiatiques tous les individus étant au moins à 45% Asiatiques.

^b: Sont définis comme Mélanésiens, tous les individus étant au moins 45% Mélanésiens.

^c: Sont définis comme Polynésiens tous les individus étant au moins 55% Polynésiens.

^d: Sont définis comme Européens tous les individus étant au moins 60% Européens.

^e: Sont définis comme « autres » les individus ne répondant à aucun des critères a, b, c ou d.

A la suite de cette étape de définition des groupes de populations, la population d'EPITHYR est composée de 112 Asiatiques, 264 Mélanésiens et 565 Polynésiens ainsi que des 3 525 Européens définis précédemment.

ii. Estimation du taux d'apparentement entre individus

Pour les analyses ROH, il est important d'exclure les individus apparentés au premier degré (parents-enfants, frères-sœurs) du jeu de données afin de n'identifier que les segments homozygotes partagés reflétant le taux de consanguinité et non ceux expliqués par un lien de parenté connu entre les individus. Le taux d'apparentement entre individus a été calculé avec le logiciel PLINK (v1,06). La valeur « PI_HAT » (= Proportion du génome IBD, c'est à dire $P(\text{IBD} = 2) + 0,5 * P(\text{IBD} = 1)$) attribuée à chaque paire d'individus correspond à leur taux d'apparentement. Une valeur $\text{PI-HAT} \geq 0,5$ indique que deux individus sont des apparentés de premier degré. J'ai appliqué les critères d'exclusion suivants afin de choisir l'individu de chaque paire à exclure (par ordre d'importance) :

- 1 - Exclusion de l'individu s'il est impliqué dans plusieurs apparentements
- 2 - Si la paire comprend un cas et un témoin, le témoin est systématiquement exclu
- 3 - Si la paire comprend deux cas ou deux témoins, l'individu ayant le taux de données génétiques manquantes le plus élevé est exclu.

Ainsi, j'ai exclu 3 cas et 6 témoins chez les Européens d'EPITHYR, 7 cas et 7 témoins chez les Polynésiens, 4 cas et 1 témoin chez les Mélanésiens et 2 cas et 1 témoin Asiatiques. Aucun individu d'EPIC n'a été exclu en tenant compte de ce critère.

iii. Description des populations

EPITHYR

Les analyses ROH ont été menées sur 3 510 individus Européens (1 552 cas et 1 958 témoins), 551 Polynésiens (241 cas et 310 témoins), 259 individus Mélanésiens (151 cas et 108 témoins) et 109 Asiatiques (26 cas et 83 témoins) d'EPITHYR. Le descriptif de ces populations est détaillé Tableaux 9 et 10.

EPIC

Aucun individu et aucun SNP n'a été exclu pour les analyses ROH qui ont donc inclus 1 216 individus (369 cas et 847 témoins) (Tableau 11).

Tableau 9 : Description de la population européenne EPITHYR selon l'étude, l'âge, le sexe et l'histologie de la tumeur (après exclusion des apparentés).

		Européens (N=3 510)			
		Cas		Témoins	
		N=1 552	%	N=1 958	%
Etude					
	CATHY	459	28,9	534	27,3
	Cuba	102	6,6	104	5,3
	Tchernobyl	66	4,2	304	15,5
	E3N	276	17,8	287	14,6
	Nouvelle Calédonie	22	1,4	69	3,6
	Polynésie Française			4	0,2
	Polynésie Française 2				
	YoungThyr	627	41,1	656	33,5
Age (années)					
	[0-25[	272	17,5	508	25,9
	[25-50[	763	49,2	878	44,8
	≥50	516	33,3	573	29,3
Sexe					
	Homme	275	17,7	449	22,9
	Femme	1277	82,3	1509	77,1
Type histologique					
	Papillaire	1398	90,1		
	Vésiculaire	154	9,9		
Taille du carcinome (pour les CPT)					
	< 10 mm	503	36,0		
	≥ 10 mm	888	63,5		
	NA	7	0,5		

Tableau 10 : Description des populations asiatiques, mélanésiennes et polynésiennes selon l'étude, l'âge, le sexe et l'histologie de la tumeur (après exclusion des apparentés).

	Polynésiens (N=551)				Mélanésiens (N=259)				Asiatiques (N=109)			
	Cas		Témoins		Cas		Témoins		Cas		Témoins	
	N=241	%	N=310	%	N=151	%	N=108	%	N=26	%	N=83	%
Etude												
CATHY											1	1,2
Cuba												
Tchernobyl												
E3N												
Nouvelle-Calédonie	11	4,6	20	6,5	151	100	108	100	11	42,3	26	31,3
Polynésie Française	97	40,2	133	42,9					6	23,1	28	33,7
Polynésie Française 2	133	55,2	135	43,6					5	19,2	25	30,1
YoungThyr			2						3	11,5	4	4,8
Age (années)												
[0-25[	14	5,8	28	9,0	7	4,6	3	2,8	3	11,5	4	4,8
[25-50[	146	60,6	172	55,5	78	51,6	66	61,1	15	57,7	55	66,3
≥50	81	33,6	84	27,1	66	43,7	39	36,1	8	30,8	24	28,9
Sexe												
Homme	33	13,7	44	14,2	8	5,3	8	7,4	5	19,2	13	15,7
Femme	208	86,3	266	85,8	143	94,7	100	92,6	21	80,8	70	84,3
Type histologique												
Papillaire	199	82,6			129	85,4			25	96,2		
Vésiculaire	40	16,6			22	14,6			1	3,8		
NA	2	0,8										
Taille du carcinome (pour les CPT)												
< 10 mm	98	49,5			68	52,7			8	32,0		
≥ 10 mm	32	16,2			61	47,3			10	40,0		
NA	69	34,3							7	28,0		

Tableau 11 : Description des populations européennes d'EPIC selon l'étude, l'âge, le sexe et l'histologie de la tumeur (après exclusion des apparentés).

		EPIC (N=1 216)			
		Cas		Témoins	
		N=369	%	N=847	%
Etude					
	Danemark	35	8,5	71	8,9
	France	49	22,7	116	8,9
	Allemagne	71	17,1	157	19,6
	Italie	97	23,4	232	28,9
	Pays-Bas	13	9,2	33	4,1
	Norvège	8	1,9	17	2,1
	Espagne	58	14,0	127	15,8
	Suède	16	3,9	45	5,6
	Angleterre	22	5,3	49	6,1
Age (années)					
	[0-25[	0	0	0	0
	[25-50[	45	12,2	29	3,4
	≥50	324	87,8	818	96,6
Sexe					
	Homme	53	12,8	177	22,1
	Femme	316	87,2	670	77,9
Type histologique					
	Papillaire	243	65,6		
	Vésiculaire	65	17,6		
	NA	61	16,8		
Taille du carcinome (chez les CPT)					
	< 10 mm	115	47,3		
	≥ 10 mm	31	12,7		
	NA	97	39,9		

2. Stratégie d'analyse

i. Identification des ROH dans chaque population

De nombreux ROH sont identifiés dans le génome du fait des blocs de DL. Afin de repérer les régions HBD, nous limitons l'étude aux ROH les plus longs. Pour chaque individu, nous avons repéré les régions homozygotes ayant un nombre minimal de SNPs consécutifs homozygotes. Il n'y a pas de méthode de référence pour définir ce seuil minimal dans les analyses ROH aussi nous avons choisi de le définir selon deux méthodes. Dans un premier temps, j'ai considéré les ROH d'une longueur d'au moins 1000kb contenant des SNPs consécutifs avec un génotype homozygote contenus dans ce segment.

Dans un second temps, j'ai appliqué la méthode décrite par Lencz et al (T, Lencz et al, 2007) qui consiste à définir une longueur minimale de ROH en calculant **le nombre minimum de SNPs consécutifs** avec des génotypes homozygotes. Dans ce cas, la « longueur minimale des ROHs » y ou nombre de SNPs minimum définissant un ROH dans la population étudiée doit vérifier cette équation :

$$[(1 - \text{Hétérozygotie moyenne})^y \times \text{Nb SNPs} \times \text{Nb individus}] \leq \alpha$$

Avec :

y = nombre minimum de SNPs pour définir un ROH

Nb SNP = nombre de SNP génotypés par individu

Nb individus = nombre d'individus génotypés

α = % de faux positifs ROH

Cette méthode suppose que les SNPs sont indépendants. J'ai ainsi exclu les variants ayant une FAM supérieure à 0,01 et un r^2 par paire de SNPs supérieur à 0,80. Cette méthode permet de limiter l'erreur de type I avec l'identification de ROH dus au DL qui sont parmi les plus courts. J'ai ainsi utilisé un seuil $\alpha = 5\%$.

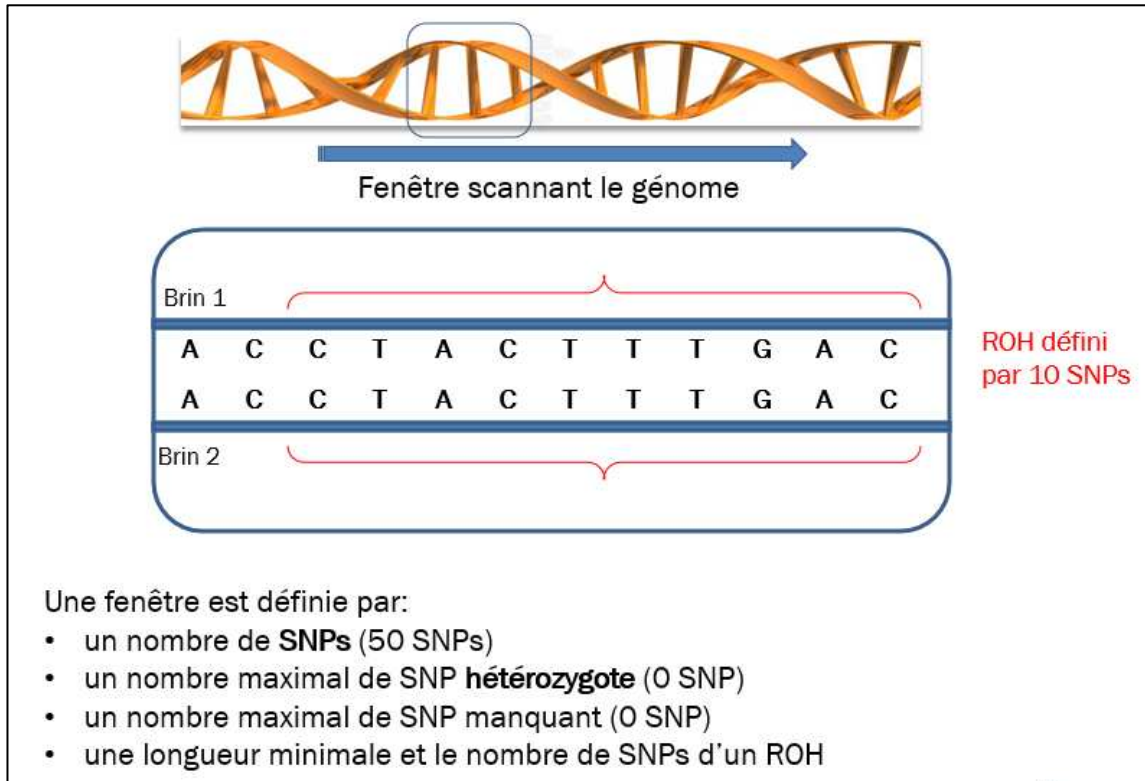
Dans ce mémoire, seuls les ROH identifiés avec la seconde définition seront présentés dans les résultats.

Les ROH identifiés selon les critères de notre première définition (i.e. segments > 1000 kb) étaient présents en plus faible nombre dans les différentes populations et étaient pour beaucoup inclus dans les ROH identifiés selon notre deuxième définition. Les associations entre ces ROH d'une longueur 1000 kb et le risque de CDT ont été testé et aucune association significative n'a été mise en évidence.

ii. Méthode d'identification des ROH : algorithme et paramètres

Les segments ROH ont été identifiés avec le logiciel PLINK (v1.06) à partir des données de génotypage du GWAS. Les algorithmes utilisés par PLINK, scannent chaque chromosome en déplaçant une fenêtre de taille fixe le long du génome afin d'identifier les segments de SNPs homozygotes consécutifs (Figure 31). La **longueur de la fenêtre, exprimée en nombre de SNPs**, est définie par l'utilisateur. Ici la fenêtre choisie est de 50 SNPs, valeur implémentée par défaut dans le logiciel. Un nombre variable de **SNPs hétérozygotes** ou **manquants** par fenêtre peut être spécifié afin de tolérer les erreurs et les échecs de génotypage ainsi que les événements mutationnels récents rares. Dans l'analyse, je n'ai toléré aucun SNP manquant ou hétérozygote afin de limiter le nombre de faux positifs. Enfin, un ROH est identifié si le nombre de SNPs consécutifs dans un segment homozygote dépasse un seuil prédéfini en termes de nombre de SNPs et/ou de longueur en kilobase. Les seuils définis pour chaque population sont détaillés dans le tableau 12 dans la partie résultats.

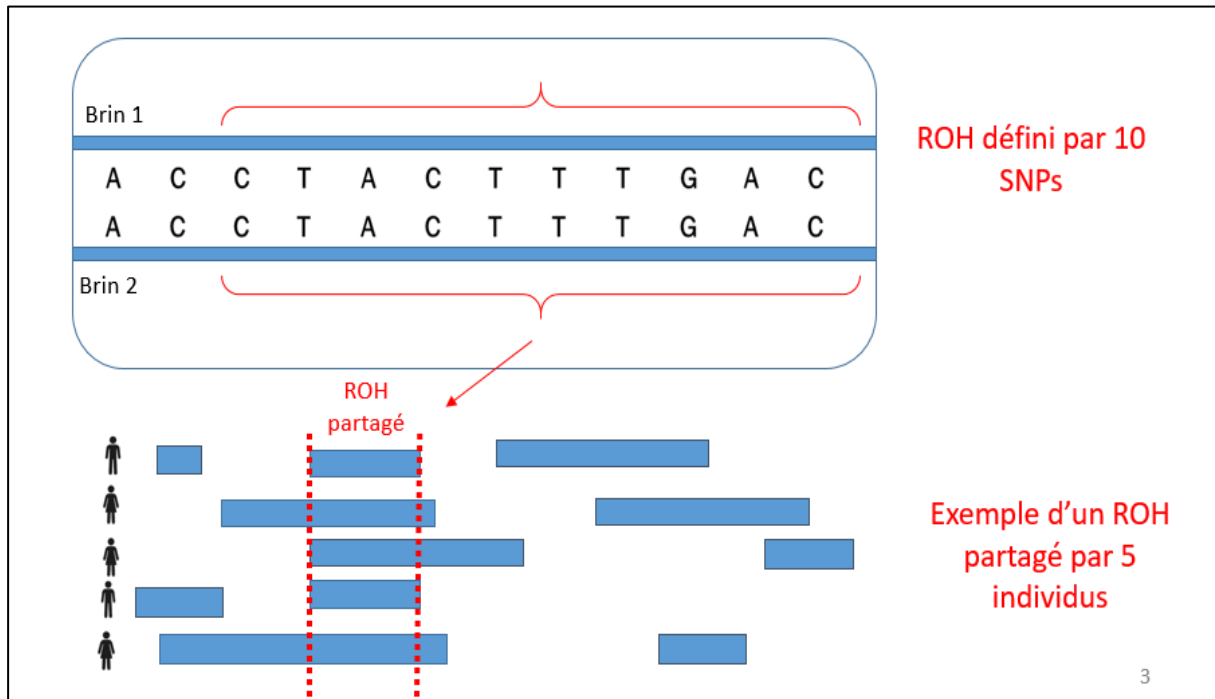
Figure 31 : Principe d'identification de ROH par le logiciel PLINK.



iii. Identification des ROH partagés entre individus

Pour nos analyses, un ROH est défini comme « partagé » s'il est présent chez au moins 5 individus (cas et/ou témoins). Un ROH est considéré comme identique chez deux individus s'il débute et se termine à la même position sur le génome (position « start » et position « end ») (Figure 32).

Figure 32 : Illustration d'un ROH partagé entre individus.



iv. Estimation du coefficient de consanguinité des individus

J'ai utilisé la méthode décrite par McQuillan et coll (McQuillan et al, 2008) pour calculer le taux de consanguinité de chaque individu.

Une mesure génomique de l'autozygotie individuelle (Froh) a été dérivée, définie comme la proportion du génome autosomique dans les séries d'homozygotie au-dessus d'un seuil de longueur spécifié :

$$Froh = \sum L_{roh} / L_{auto}$$

avec :

- L_{roh} : la longueur totale cumulée de tous les ROH d'un individu
- L_{auto} : la longueur du génome autosomique couverte par les SNPs (en excluant les centromères).

v. Analyses statistiques

Toutes les analyses statistiques ont été réalisées sous R. La longueur et la fréquence des ROH ont été comparées entre cas et témoins par un test du Chi2. L'association entre la longueur ou la fréquence des ROH et le risque de DTC dans chaque population a été testée avec un modèle logistique ajusté sur l'âge, le sexe, l'étude ou le pays ainsi que les 10 premières composantes principales. L'association entre chaque ROH partagé et le risque de CDT a été testée avec un test exact de Fisher. Aucune correction sur les tests multiples n'a été effectuée.

III. Résultats

1. Hétérozygotie et longueur minimale des ROH dans les différents groupes de population étudiés

Comme décrit précédemment, seuls les SNPs faiblement corrélés sont pris en compte pour estimer la longueur minimale des ROH dans un groupe de population donnée. Les nombres de SNPs utilisés pour ce calcul dans chaque population sont indiqués dans le tableau 12.

Les longueurs minimales pour identifier un ROH dans les différentes populations étaient respectivement de 76, 73, 65 et 55 SNPs (Tableau 12).

Tableau 12 : Caractéristiques des différentes populations et longueurs minimales des ROH (en nombre de SNPs) calculées pour les Européens, les Polynésiens, les Mélanésiens et les Asiatiques d'EPITHYR et d'EPIC.

Populations***	N SNPs**	N individus	Hétérozygotie moyenne* (%)	Longueur minimale ROH ^a
Européens d'EPITHYR	320 345	3510	27,0	76
Européens d'EPIC	320 079	1216	27,4	72 ^b
Européens d'EPITHYR + EPIC	316 203	4726	27,7	75 ^b
Polynésiens	239 360	551	26,0	73
Mélanésiens	227 758	259	27,8	65
Asiatiques	255 436	109	30,7	55

*** D'après les groupes de populations définis avec Admixture.

** Nombre de SNPs faiblement corrélés du GWAS pris en compte dans l'analyse.

* L'hétérozygotie moyenne est calculée pour chaque population pour tous les SNPs (aucun SNP n'a été exclu selon la FAM ou le DL).

^a Les longueurs minimales de SNP pour définir un ROH.

^b Pour les analyses ces seuils ne seront pas utilisés. Pour les Européens d'EPIC et d'EPITHYR ainsi que dans l'analyse poolée EPITHYR+EPIC, nous avons considéré la longueur minimale de ROH de 76 SNPs définie dans EPITHYR afin de pouvoir analyser tous les Européens selon le même critère.

2. Caractéristiques globales des ROH dans les différentes populations

Nous avons identifié en moyenne six ROH avec une longueur totale moyenne de 11 503 kb par individu chez les témoins Européens d'EPITHYR. Le nombre moyen de segments ROH par individu est significativement plus élevé chez les témoins que chez les cas chez les Européens d'EPITHYR (p -valeur = $1,2 \times 10^{-4}$) et de l'analyse poolée (p -valeur = $5,7 \times 10^{-5}$) (Tableau 13). Bien que l'association ne soit pas significative, nous avons observé la tendance inverse chez les Européens d'EPIC, avec un nombre moyen de ROH plus élevé chez les cas que chez les témoins (p -valeur = 0,53).

Tableau 13 : Comparaison des caractéristiques globales des ROH chez les cas et les témoins européens d'EPITHYR et d'EPIC.

	EPITHYR (1552 cas/ 1958 témoins) 320 345 SNPs			EPIC^b (369 cas/ 847 témoins) 320 079 SNPs			EPITHYR + EPIC^c (1921 cas/2805 témoins) 316 203 SNPs		
	Cas	Témoins	<i>P^d</i>	Cas	Témoins	<i>p</i>	Cas	Témoins	<i>p</i>
-Nombre moyen de ROH / individu	5,5	6,0	1,2x10⁻⁴	6,8	6,6	0,53	5,7	6,0	5,7x10⁻⁵
-Longueur moyenne de ROH (kb) / individu	1 785	1 784	0,97	1 910	1 999	0,17	1 817	1 861	0,19
-Longueur moyenne cumulée totale de ROH / individu (kb)	11 088	11 503	0,41	14 471	15 124	0,57	11 415	12 209	0,06

Chez les Polynésiens, les Mélanésiens et les Asiatiques, le nombre de ROH par individu est plus élevé que chez les Européens, avec respectivement 65,7, 24,9 et 8,5 ROH/individu dans ces populations (Tableau 14), ce qui traduit une plus grande homozygotie par rapport aux Européens.

De plus, nous avons observé chez les Polynésiens une longueur moyenne des ROH significativement plus élevée chez les témoins que chez les cas (*p*-valeur = 0,01).

Tableau 14 : Caractéristiques globales des ROH chez les cas et les témoins chez les Polynésiens, les Mélanésiens et les Asiatiques d'EPITHYR.

	Polynésiens^a (241 cas / 310 témoins), 239 360 SNPs			Mélanésiens^b (151 cas / 108 témoins), 227 758 SNPs			Asiatiques^c (26 cas /83 témoins), 255 436 SNPs		
	Cas	Témoins	<i>P</i>^d	Cas	Témoins	<i>p</i>	Cas	Témoins	<i>p</i>
-Nombre moyen de ROH / individu	71,0	65,7	0,01	28,7	24,9	0,01	8,9	8,5	0,76
-Longueur moyenne de ROH (kb) / individu	2 440	2 535	0,01	2 356	2 312	0,75	2 472	3 713	0,27
-Longueur moyenne cumulée totale de ROH / individu (kb)	175 252	167 568	0,22	69 659	60 355	0,10	24 288	26 936	0,77

^d Test de Student

3. Association entre la longueur ou le nombre de ROH et le risque de CDT

Dans chaque population, nous avons testé l'association entre les ROH et la prédisposition au CDT en créant différentes classes d'individus selon le nombre de ROH et la longueur totale des ROH en Mb, en utilisant des quartiles définis chez les témoins (Tableaux 15 et 16).

Globalement aucune association significative entre le risque de CDT et le nombre ou la longueur de ROH par individu n'a été identifiée dans les groupes de population européens, mélanésiens, polynésiens et asiatiques. Chez les Européens les associations ne vont pas dans le même sens dans la population d'EPIC et la population d'EPITHYR. Chez les Asiatiques, on observe une association négative entre CDT et longueur totale cumulée de ROH supérieure à 12 Mb alors que chez les Mélanésiens, la tendance observée est inverse (Tableau 16).

Tableau 15 : Association entre le nombre de ROH ou la longueur totale ROH et le risque de CDT chez les Européens.

	EPITHYR (N=3510, 1552 cas/ 1958 témoins) 320 345 SNPs					EPIC (369 cas/ 847 témoins) 321 112 SNPs					EPITHYR EUR + EPIC (N=4726, 1921 cas/ 2805 témoins) 316 203 SNPs				
Nombre de ROH / individu ^a	Cas	Témoins	OR ^b	IC 95%	P	Cas	Témoins	OR ^b	IC 95%	P	Cas	Témoins	OR ^b	IC 95%	P
< 4	369	414	1,00	Ref	Ref	63	158	1,00	Ref	Ref	503	646	1,00	Ref	Ref
4 - 6	517	556	1,08	0,89-1,31	0,41	98	234	1,03	0,69-1,53	0,88	609	793	1,06	0,90-1,25	0,47
6 - 8	350	481	0,92	0,75-1,13	0,45	85	202	1,05	0,70-1,60	0,79	406	655	0,93	0,78-1,11	0,42
≥8	316	507	0,93	0,75-1,15	0,51	123	253	1,05	0,81-1,88	0,32	403	711	0,98	0,82-1,18	0,88
Longueur totale de ROH /individu (Mb)															
< 5,8	435	493	1,00	Ref	Ref	65	178	1,00	Ref	Ref	566	718	1,00	Ref	Ref
5,8 - 8,9	457	483	1,12	0,93-1,35	0,22	87	194	1,30	0,87-1,95	0,20	539	700	1,05	0,89-1,23	0,55
8,9 - 13	341	484	0,90	0,74-1,09	0,29	89	203	1,33	0,89-2,00	0,16	409	654	0,90	0,76-1,07	0,25
≥ 13	319	498	0,94	0,77-1,15	0,57	128	272	1,34	0,90-2,00	0,14	407	733	0,93	0,77-1,10	0,40

^b Régression logistique ajustée sur l'âge, le sexe, l'étude et les 10 premières CPs.

Tableau 16 : Association entre le nombre de ROH ou la longueur totale des ROH et le risque de CDT chez les Polynésiens, les Mélanésiens et les Asiatiques d'EPITHYR.

	Polynésiens (N=551, 241 cas/ 310 témoins) 239 360 SNPs						Mélanésiens (N=259, 151 cas/ 108 témoins) 227 758 SNPs						Asiatiques (N=109, 26 cas/ 83 témoins) 255 436 SNPs					
	Seuils ^a	Cas	Témoins	OR ^b	IC 95%	P	Seuils ^a	Cas	Témoins	OR ^c	IC 95%	P	Seuils ^a	Cas	Témoins	OR ^c	IC 95%	P
Nombre de ROH / individu ^a	< 65	92	156	1,00	Ref	Ref	<23	50	50	1,00	Ref	Ref	<9	12	40	1,00	Ref	Ref
	≥ 65	149	154	1,38	0,75-2,54	0,29	≥23	101	58	1,05	0,48-2,24	0,90	≥9	14	43	0,49	0,15-1,53	0,22
Longueur totale de ROH /individu (Mb)	< 162	100	156	1,00	Ref	Ref	<52	48	54	1,00	Ref	Ref	<12	13	42	1,00	Ref	Ref
	≥ 162	141	151	0,91	0,52-1,59	0,75	≥52	103	54	1,59	0,77-3,27	0,20	≥12	13	41	0,48	0,14-1,55	0,23

^a Les seuils définis sont spécifiques à chaque population. Ces quartiles ont été définis dans le but d'obtenir des classes de témoins aux effectifs similaires. ^b Modèle logistique mixte ajusté selon l'âge le sexe et le type de prélèvement (sanguin, salivaire ou brosse buccale) et les 10 premières CPs. ^c Modèle logistique mixte ajusté selon l'âge le sexe et les 10 premières CPs.

4. Consanguinité et risque de CDT

Les coefficients de consanguinité (Froh) ont été calculés en utilisant la méthode décrite par McQuillan et coll. (McQuillan et al, 2008). Les valeurs moyennes de ces coefficients pour les cas et les témoins de chaque population sont rapportées dans le tableau 17.

Les taux de consanguinité observés sont plus élevés chez les Mélanésiens, les Polynésiens et les Asiatiques que chez les Européens (Tableau 17). Chez les Européens d'EPITHYR et d'EPIC, les coefficients de consanguinité sont similaires entre les cas et les témoins. Chez les Mélanésiens et les Polynésiens les cas ont un taux de consanguinité plus élevé que chez les témoins contrairement aux Asiatiques pour lesquels on observe une tendance inverse (Tableau 17).

Tableau 17 : Comparaison des coefficients de consanguinité (Froh) entre les cas et les témoins des différents groupes de population.

	Coefficient de consanguinité (Froh)		
	Cas	Témoins	p^a
Européens EPITHYR	0,0098	0,0098	0,75
Européens EPIC	0,011	0,011	0,69
Européens EPITHYR+EPIC	0,0095	0,0099	0,12
Polynésiens	0,083	0,080	0,25
Mélanésiens	0,041	0,037	0,10
Asiatiques	0,013	0,016	0,05

^a p -valeur calculée avec le test de Student.

5. Nombre de ROH partagés et risque de CDT

Chez les 3 510 individus Européens d'EPITHYR analysés, 135 ROH de longueur supérieure à 500 kb et avec au moins 76 SNPs homozygotes consécutifs sont partagés par au moins cinq individus (cas ou témoins). Si l'on s'intéresse aux ROH d'une longueur supérieure à 1000 kb et sans fixer un nombre minimum de SNPs, 32 ROH sont partagés par au moins cinq individus (cas ou témoins) (Tableau 18) et 44 ROH partagés sont identifiés chez les individus européens d'EPIC.

Les nombres de ROH identifiés au sein des populations européennes d'EPITHYR et d'EPIC sont similaires. Cependant, on observe une grande hétérogénéité concernant le nombre de ROH identifiés dans les autres populations. Chez les Polynésiens et les Asiatiques, moins de cinq ROH partagés sont identifiés (Tableau 18), tandis que chez les Mélanésiens, 83 ROH partagés sont identifiés avec au moins 65 SNPs homozygotes consécutifs et 67 ROH partagés avec une longueur minimale de 1000kb.

L'association avec le risque de CDT a été testée pour l'ensemble des ROH partagés. Cette analyse montre que six de ces ROH sont associés au risque de CDT, à un seuil p -valeur $\leq 0,05$ (Tableau 19). Quatre ROH (en 1q25.3, 15q23, 2p23 et 11q41.1) sont associés au CDT chez les Européens et deux ROH (en 13q31.3 et 11q14.1) sont associés au CDT chez les Mélanésiens. Chez les Européens, le ROH partagé le plus fortement associé au risque de CDT est situé au **locus 1q25.3** (ROH1, p -valeur = 0,007 ; Tableau 19). Un cas et 14 témoins sont porteurs de ce ROH contenant 161 SNPs homozygotes consécutifs. Au **locus 15q23**, le ROH2 s'étend sur 549 kb et contient 84 SNPs homozygotes consécutifs.

Chez les Européens d'EPITHYR, une région partagée de 549kb est identifiée en **2p23**, (ROH3, p -valeur= 0,03) ainsi qu'une région en **11q24.1** chez les Européens d'EPIC, contenant 10 gènes (ROH4, p -valeur=0,04).

Chez les Mélanésiens, un ROH partagé par 8 témoins est identifié au **locus 13q31.3** (ROH5) et un ROH partagé par 7 témoins au **locus 11q14.1** (ROH6). Ces deux ROH ne sont pas trouvés chez les cas (Tableau 19).

Tableau 18 : Nombre de ROH partagés identifiés dans chaque population.

Populations	ROH ≥ 5 individus	ROH ≥ 500 kb, ≥ 5 individus	ROH partagés (≥ 500 kb, ≥ 5 individus 73 (Polynésiens), 65 (Mélanésiens) ou 55 (Asiatiques) SNPs, même position START-END)	ROH partagés (≥ 1000 kb, ≥ 5 individus, même position START-END)
EPITHYR (Européens)	4 565	205	135	21
EPIC	3 408	458	174	44
EPITHYR EUR + EPIC	8 599	211	126	13
Polynésiens	11 332	66	2	5
Mélanésiens	2 277	338	86	67
Asiatiques	45	18	3	5

Tableau 19 : ROH partagés et associés au risque de CDT dans les différentes populations d'EPITHYR et d'EPIC.

ROH	Population	Chr	Position Start - End (bp) ^a	Taille de la région (en kb)	Nombre de SNPs	Cas/ témoins ^b	Total Cas / Témoins ^c	<i>P</i> ^d	Gènes situés dans le ROH	Commentaire
ROH1	Européens (EPITHYR + EPIC)	1	182,756,286 – 183,551,760	795		01/14	1921 / 2805	0,007	<i>NPL, DHX9, LAMC1, LAMC2, NMNAT2, SMG7, NCF2</i>	Aussi identifié dans EPIC : Chez 0 cas/ 08 témoins ; <i>p</i> =0,11, OR=0,0, IC95%= [0,00-1,34]
ROH2	Européens (EPITHYR)	15	31,890,123 – 32,439,298	549		09/03	1552 /1958	0,04	<i>OTUD7A, CHRNA7</i>	Aussi identifié dans études poolées, EPIC et EPITHYR chez 10 cas/ 06 témoins ; <i>p</i> =0,12, OR=2,44, IC95%= [0,80-8,18]
ROH3	Européens (EPITHYR)	2	29,247,786 – 29,797,217	549		08/02	1552 /1958	0,03	<i>FAM179A, CLIP4, ALK, TOGARAM2,</i>	Non identifié dans EPIC et dans l'étude poolée.
ROH4	Européens (EPIC)	11	123,324,588 – 123,864,216	539		04/01	369 / 847	0,03	<i>GRAMD1B, SNC3B, ZNF202, KIAA1201, OR6M1, TMEM225, OR8D4, OR4D5, OR6T1, OR10S1</i>	Non identifié dans EPIC et dans l'étude poolée.
ROH5	Mélanésiens	13	93,298,523 – 93,990,178	691		08/00	151 /108	0,02	<i>GPC5-AS1, GPC5, GPC6</i>	Non identifié chez les Asiatiques et les Polynésiens.
ROH6	Mélanésiens	11	78,954,142 – 80,125,592	1172		07/00	151 /108	0,04	<i>TENM4, MIR708, MIR5579</i>	Non identifié chez les Asiatiques et les Polynésiens.

^aPositions chromosomiques (NCBI), build 37, hg19 : ^bNombre de cas et témoins porteurs du ROH ; ^cNombre total de cas et témoins de la population ; ^d*p* calculé avec un test exact de Fischer.

6. Analyses de sensibilité

i. Comparaison Nord/Sud des caractéristiques des ROH identifiés chez les Européens

On a observé que les caractéristiques des ROH des Européens d'EPITHYR étaient sensiblement différentes de celles des Européens d'EPIC. Si les Européens d'EPITHYR proviennent majoritairement de France métropolitaine (études CATHY, YongThyr, E3N) et de Biélorussie (Etude Tchernobyl), la population d'EPIC est plus hétérogène. Afin d'étudier si les caractéristiques des ROH identifiés varient entre les différents groupes de population d'origine européenne, nous avons séparé les Européens d'EPITHYR et d'EPIC selon leur origine « Nord » / « Sud » d'après les résultats de l'analyse en composantes principales réalisée sur les données des deux consortia poolés (Figure supplémentaire 6). La séparation Nord/Sud a été faite selon la CP1 : les pays du Nord ont une $CP1 > 0$ et les pays du Sud une $CP1 < 0$. Les participants français d'EPIC sont répartis dans les deux groupes (tableau 20).

Si nous n'observons pas de différences significatives concernant la longueur moyenne des ROH, il est intrigant de souligner que les individus originaires du Sud de l'Europe ont un nombre moyen de ROH plus élevé ($N-ROH_{Nord} = 5,6$ et $N-ROH_{Sud} = 7$) (Tableau supplémentaire 4) mais que les individus originaires du Nord partagent plus de ROH que ceux originaires du Sud (20 ROH partagés versus 2) (Tableau supplémentaire 6).

Nous n'observons aucune différence entre les coefficients de consanguinité des cas et des témoins de chacun des sous-groupes Nord/Sud. Cependant les individus du groupe Nord ont un taux de consanguinité un peu plus élevé que ceux du groupe du Sud (Tableau supplémentaire 7), ce qui peut expliquer le nombre de ROH plus élevé identifié dans cette population par rapport au groupe du Sud.

Aucun de ces ROH partagés n'est significativement associé au risque de CDT.

Tableau 20 : Description des individus des groupes EPITHYR+EPIC NORD et EPITHYR+EPIC SUD selon les études, l'âge le sexe et le type histologique.

		'EPITHYR + EPIC' SUD (N=3759, 1677 cas/ 2082 témoins)				'EPITHYR + EPIC' NORD (N=967, 244 cas/ 723 témoins)			
		Cas		Témoins		Cas		Témoins	
		N=1677	%	N=2082	%	N=244	%	N=723	%
Etude									
	CATHY	459	27,4	534	25,6				
	Cuba	102	6,1	104	5,0				
	Tchernobyl					66	27,0	304	42,0
	E3N	276	16,4	287	13,8				
	Nouvelle Calédonie	22	1,3	69	3,3				
	Polynésie Française			4	0,2				
	Polynésie Française 2								
	YoungThyr	627	37,4	656	31,5				
	Danemark	0	0			35	14,3	71	9,8
	France	36		69	3,3	13	5,3		
	Allemagne	0	0	0	0	71	29,1	157	21,7
	Italie	97	5,8	232	11,1	0	0		
	Pays-Bas	0	0	0	0	13	5,3	33	4,6
	Norvège	0	0	0	0	8	3,3	17	2,3
	Espagne	58	3,4	127	6,1	0	0	0	0
	Suède	0	0	0	0	16	6,5	45	6,2
	Angleterre	0	0	0	0	22	9,0	49	6,8
Age (années)									
	[0-25[	206	12,3	209	10,0	66	27,0	299	41,3
	[25-50[	785	46,8	878	42,2	23	9,5	29	4,0
	≥50	686	40,9	995	47,8	155	63,5	395	54,7
Sexe									
	H	268	16,0	410	19,7	60	24,6	216	29,9
	F	1409	84,0	1672	80,3	184	75,4	507	70,1

Type histologique				
Papillaire	1481	88,3	179	73,4
vésiculaire	158	9,4	44	18,0
NA	38	2,3	21	8,6
Taille du carcinome (pour les CPT)				
< 10 mm	491	33,2	94	52,5
≥ 10 mm	927	62,6	44	24,6
NA	63	4,3	36	22,9

ii. Influence du type de prélèvement à partir duquel est extrait l'ADN sur la caractérisation des ROH

Le type de prélèvement d'ADN peut influencer sur la qualité du génotypage d'un individu. En l'occurrence, de précédentes analyses effectuées sur les données EPITHYR ont montré que les données de génotypage des individus ayant eu des prélèvements d'ADN buccaux par brosse sont de moins bonne qualité que celles obtenues à partir de l'ADN salivaire ou sanguin. Pour l'étude Polynésie Française 1, les échantillons de salive ont été prélevés au moyen de kits Oragene ou par brosse buccale.

Etant donné que 30% des Polynésiens ont été génotypés avec un prélèvement par brosse (164 individus, 61 cas et 103 témoins), j'ai recherché si le type de prélèvement pouvait avoir un impact sur la caractérisation des ROH dans l'étude polynésienne. Pour ce faire et pour ne pas introduire de biais dû à la taille de l'échantillon analysé, j'ai évalué l'impact du type de prélèvement sur un échantillon composé du même nombre de cas et de témoins pour les deux groupes de Polynésiens à comparer, soit 61 cas et 103 témoins pour chaque sous-groupe. Ainsi, j'ai pu vérifier que le type de prélèvement ne modifie pas le nombre ou la longueur des ROH identifiés (Tableau supplémentaire 8).

IV. Discussion

L'objectif de l'étude des régions homozygotes du génome chez les individus d'EPITHYR et d'EPIC était d'identifier de nouveaux facteurs de risque génétiques du CDT agissant selon un mode de transmission récessif et de tirer profit du fait que la population d'EPITHYR est composée de groupes de populations ethniquement diverses.

L'analyse individuelle des SNPs à l'échelle du génome, réalisée avec un modèle récessif dans EPITHYR, a identifié les mêmes loci que ceux précédemment identifiés dans les autres GWAS (Gudmundsson et al, 2017) mais n'a pas permis de mettre en évidence de nouveau signal (Figure supplémentaire 5). Même si on peut s'attendre à un effet plus important des variants récessifs, la faible taille de nos populations limite la puissance de détection de nouveaux signaux avec une GWAS. C'est pour cette raison que j'ai choisi d'utiliser une autre méthode basée sur la définition des ROH et plus fine que l'approche classique des GWAS pour détecter des variants récessifs associés au risque de CDT.

J'ai d'abord décrit les caractéristiques des ROH dans les différents groupes de population étudiés. J'ai ainsi mis en évidence des différences entre les Européens d'une part, et les Polynésiens et Mélanésiens d'autre part et ai montré que ces populations insulaires portent un grand nombre de ROH longs. Il a été montré que la distribution des ROH dans différentes populations dépend de plusieurs facteurs comme notamment des événements démographiques ou la consanguinité (Ceballos et al, 2018), ce qui peut expliquer les différences en termes de profils d'homozygotie observés. Une précédente étude menée sur 1 839 sujets des panels de référence Hapmap et HGDP a démontré que les populations océaniques ont de très longs segments ROH en comparaison à d'autres populations ce qui est en accord avec les estimations démographiques d'un nombre élevé de mariages consanguins observés dans ces populations (Bittles and Black 2010; Pemberton et al. 2012). Ces populations natives et isolées partagent une ascendance commune relativement récente et peuvent héberger des ROH de 2 à 5 Mb en moyenne alors que les Européens, issus d'une population mixte résultant des différentes migrations et de l'occidentalisation, ont en moyenne moins de ROH et des ROH plus courts (Ceballos et al, 2018). L'étude des taux de consanguinité dans chaque groupe ethnique a

confirmé que les Polynésiens et les Mélanésiens ont des niveaux de consanguinité plus élevés que les Européens.

Nous avons identifié 4 ROH associés au CDT chez les Européens et 2 autres ROH associés au CDT chez les Mélanésiens. Plusieurs de ces régions contiennent des gènes d'intérêt pour ce cancer (Tableau 19). En effet au **locus 1q25.3** (ROH1), le gène *DHX9* joue un rôle central dans de nombreux processus cellulaires tels que la régulation de l'expression de gènes et le maintien de la stabilité génomique. *DHX9* code une enzyme impliquée dans la réparation de l'ADN et la transcription et interagit avec des protéines communes du gène *WRN* impliqué dans le syndrome de Werner de transmission récessive et pour lequel des mutations ont été associées à une augmentation de l'instabilité du génome et de risque de cancer, dont le CDT (Chakraborty and Grosse 2010). Ainsi, l'implication de *DHX9* dans plusieurs cancers et leur traitement a été étudié (Lee and Pelletier 2016). Notamment il a été démontré que *DHX9* est impliqué dans différents types de cancer dont les cancers de la prostate, du poumon et du sein (Lai et al, 2011; Schlegel, Starita, and Parvin 2003; Sun et al, 2014). Au même locus, le gène *LAMC1* est impliqué dans la progression de diverses tumeurs malignes, telles que les carcinomes hépatocellulaires (Zhang et al. 2017) et sa surexpression est fortement associée à une faible survie globale des patientes atteinte d'un cancer de l'endomètre (Kunitomi et al, 2020) et des patients atteints d'un cancer de l'estomac (Han, Jiang, and WC. Fan 2021). Aussi, *LAMC1* est impliqué dans le développement de gliomes et peut être utilisé pour le diagnostic et comme cible thérapeutique pour ce cancer (Liu et al. 2019).

Au locus **11q24.1** (ROH4), des études fonctionnelles montrent que le gène *ZNF202* joue un rôle important dans le contrôle du niveau de HDL sanguin (Razzaghi, Santorico, and Ilyas Kamboh 2012; Vrins et al, 2013). L'obésité étant un facteur de risque clairement établi du CDT (Matrone et al. 2020), nous pouvons émettre l'hypothèse d'un rôle de ce gène dans la susceptibilité au CDT impliquant des variants récessifs qui altèrent l'expression de *ZNF202* perturbant ainsi une voie métabolique liée à l'obésité.

D'après la littérature, les quatre ROH significativement associés au risque de CDT aux loci **15q23, 2p23, 13q31.3 et 11q14.1** ne contiennent pas de gène connu *a priori* en lien avec l'étiologie du CDT et demanderaient des études fonctionnelles approfondies pour conclure à un

rôle dans la survenue de ce cancer. De façon générale, les différentes régions d'homozygotie partagées mises en évidence dans notre étude nécessitent d'être explorées dans des études fonctionnelles et/ou d'être répliquées dans des études cas-témoins indépendantes pour confirmer leur rôle dans la prédisposition au CDT.

A ce jour, une seule étude menée dans la population italienne a étudié le lien entre l'homozygotie et le risque de CDT (Thomsen et al. 2016). Dans EPITHYR, les cas européens portent en moyenne moins de ROH ($NROH_{EPITHYR} = 5,8$) que les cas italiens ($NROH_{THOMSEN} = 11,2$) mais la longueur moyenne des ROH est similaire dans les deux études. Dans EPIC, nous observons les mêmes tendances que dans l'étude italienne avec un plus grand nombre de ROH chez les cas que chez les témoins même si la différence est très faible et non significative. Des taux de consanguinité significativement plus élevés chez les cas que chez les témoins ont été observés dans l'étude italienne alors que nous n'observons pas de différence chez les Européens d'EPITHYR ou d'EPIC. Enfin, les ROH associés au risque de CDT que nous avons identifiés sont différents de ceux identifiés par Thomsen et coll. Seize ROH associés au risque de CDT ont été mis en évidence dans l'étude italienne dans lesquelles se situaient des gènes potentiellement impliqués dans le développement de cancers thyroïdiens. Aucun des ROH associés au CDT identifiés dans notre étude ou celle de Thomsen et coll. n'était proche d'un locus de prédisposition mis en évidence dans les GWAS.

A ce jour, du fait du faible nombre d'études menées sur l'identification de ROH, aucune méthodologie standard n'est établie pour définir un ROH, ce qui rend la reproductibilité et la comparaison des résultats entre études difficiles.

La caractérisation des ROH dépend aussi de la puce à ADN utilisée pour le génotypage. Par exemple, la puce OncoArray utilisée dans nos études a été conçue avec un « squelette GWAS » de 250 000 SNPs qui est probablement plus adaptée aux Européens et aux Asiatiques qu'à la population océanienne. Dans cette population, un plus grand nombre de SNPs a été exclu lors du contrôle qualité et la densité des SNPs du génome retenus pour les analyses est donc plus faible que pour les Européens. La méthode employée par le logiciel PLINK que j'ai utilisé n'autorisant aucun SNP manquant dans la fenêtre d'identification des ROH, on peut supposer

que l'exclusion d'un plus grand nombre de SNPs chez les Océaniens mène à l'identification de moins de ROH qu'attendu.

Notre étude a exploré l'hypothèse du rôle de la consanguinité dans le risque de CDT. Bien que nous montrions des taux de consanguinité plus élevés dans les populations insulaires océaniques nos résultats ne sont pas en faveur d'une association entre consanguinité et risque de CDT qui pourrait expliquer en partie l'incidence particulièrement élevée de ce cancer en Nouvelle-Calédonie et en Polynésie française.

Des méthodes alternatives pourraient être utilisées pour approfondir ces analyses. Par exemple la stratégie HBD-GWAS (Génin et al. 2012) implémentée dans les logiciels FSuite (Gazal et al. 2014) et Fantasio (<https://rdr.io/github/genostats/Fantasio/man/Fantasio.html>) permet d'estimer directement le coefficient de consanguinité de chaque individu et d'identifier des ROH partagés uniquement chez les individus considérés comme consanguins à partir de sous-cartes génétiques de SNPs indépendants. Des probabilités *a posteriori* d'être HBD pour chaque individu sont calculées en quantifiant la présence des SNPs de chaque individu au sein de l'ensemble des sous-cartes : un individu ayant un nombre élevé de SNPs présents au sein de plusieurs sous-cartes aura probabilité *a posteriori* d'être HBD élevée. Cette méthode ne nécessite pas de définir une longueur minimale des ROH à identifier et a donc l'avantage de ne pas dépendre de la population étudiée et de permettre d'analyser plusieurs groupes de populations selon les mêmes critères. Un autre avantage de cette méthode réside dans le fait d'exploiter la consanguinité des individus pour détecter les régions HBD, ce qui est un atout pour l'analyse des Mélanésiens et des Polynésiens d'EPITHYR qui ont un taux de consanguinité plus élevé que les Européens étudiés.

Conclusions et perspectives

Les études d'association sur le CDT conduites jusqu'à ce jour ont identifié des variants fréquents avec des effets relativement modestes qui n'expliquent qu'une part de la composante génétique de ce cancer. Au cours de cette thèse, une étude pangénomique m'a permis d'identifier de nouveaux variants de prédisposition au CDT dans des populations d'ascendance européenne et océanienne. J'ai également mis en évidence de nouveaux variants de susceptibilité aux loci 2q35 et 8p12 en utilisant une méthode de cartographie fine de ces régions et j'ai identifié des régions d'homozygotie du génome impliquées dans le risque de CDT.

La réalisation de ces analyses a permis de soulever différentes questions et mis en évidence des contraintes.

Tout d'abord, les effectifs d'EPITHYR sont relativement faibles en comparaison avec les données d'autres consortia s'intéressant à des cancers plus fréquents tels que le cancer du sein (consortium BCAC, <http://bcac.ccge.medschl.cam.ac.uk/>) ou le cancer de la prostate (consortium PRACTICAL, <http://practical.icr.ac.uk/blog/>) qui rassemblent les données génétiques de plus de 100 000 sujets. Toutefois la réunion des données des consortia EPITHYR et EPIC a permis de constituer un échantillon cas-témoin unique avec des données épidémiologiques détaillées et des données génétiques "génomique-entier" associées. C'est également l'étude GWAS qui compte le plus grand nombre de cas de CDT publiée à ce jour. Une des spécificités d'EPITHYR est de disposer de données de populations océaniques, jusqu'ici peu étudiées malgré la forte incidence au CDT observée dans ces populations. Cependant, l'échantillon étudié reste limité ce qui est en partie lié à la petite taille des populations et au manque d'infrastructure sur ces territoires pour recruter des participants dans les études épidémiologiques.

Les résultats obtenus dans le cadre de cette thèse permettent d'ouvrir de nouvelles perspectives de recherche. Tout d'abord, la **réplication des résultats** dans d'autres populations européennes et originaires des îles du Pacifique Sud permettrait de valider le rôle des facteurs de risque génétiques du CDT que nous avons identifiés. Pour avoir un panorama complet des mécanismes biologiques impliqués dans la susceptibilité au CDT, l'identification des voies métaboliques

altérées par les SNPs associés au CDT permet une meilleure compréhension des mécanismes biologiques sous-jacents au développement de la maladie. Ainsi, l'intégration des informations de la biologie systémique provenant d'organismes modèles, de données génomiques d'expression dans des tissus tumoraux et normaux et des GWAS peut aider à identifier les gènes impliqués dans l'étiologie du CDT, ainsi que les voies ou les réseaux fonctionnels dans lesquels ils sont impliqués. En parallèle des travaux présentés dans ce mémoire, cette approche a été utilisée par notre équipe et a fait l'objet d'une publication (Kulkarni et al. 2021)(article disponible dans les Annexes). L'objectif était d'intégrer les données GWAS d'EPITHYR, les données d'annotation des voies biologiques ou des interactions protéine-protéine connues (telles que répertoriées dans les bases de données Kyoto Encyclopedia of Genes and Genomes (KEGG), REACTOME et Gene Ontology (GO) et les données d'expression des gènes du TCGA (Kulkarni et al. 2021). Ces approches ont permis de mettre en évidence huit gènes associés au CDT situés dans l'un des quatre loci de prédisposition connus : 2q35, 8p12, 9q22.3, 14q13.3. Afin d'identifier les voies biologiques impliquées dans le risque de CDT, deux approches ont été utilisées. Une première analyse « par pathway » a d'abord été effectuée avec les annotations des bases de données KEGG, REACTOME et GO. Cette analyse a identifié 21 voies métaboliques de KEGG et 75 voies métaboliques de REACTOME associées au risque de CDT. Ces voies sont impliquées dans le métabolisme du cholestérol, le métabolisme des sucres aminés et des nucléotides, la biosynthèse des stéroïdes et la régulation négative des voies de signalisation ERBB2. Dans un second temps, une approche « réseaux de gènes » ou « *gene-network* » a été employée consistant en l'identification des réseaux de gènes dont les produits sont des protéines interagissant directement ou indirectement entre elles. Nous avons montré que des réseaux de gènes impliqués dans la glycogénolyse, le métabolisme du glycogène, le métabolisme de l'insuline et les voies de transduction du signal associées à la contraction musculaire sont impliqués dans la prédisposition au CDT (Kulkarni et al. 2021). Cependant, dans cette étude, seules les données GWAS de la population européenne d'EPITHYR ont été analysées. Si nous disposions d'un échantillon plus grand pour la population océanienne, nous pourrions appliquer sur celle-ci ces stratégies s'appuyant sur les données de la biologie des systèmes et comparer les voies biologiques et les réseaux de gènes impliqués dans les deux groupes.

Comme dans d'autres maladies multifactorielles, des facteurs environnementaux ou liés au mode de vie, ainsi que des interactions potentielles entre les différents facteurs de risque, génétiques et non génétiques jouent un rôle dans la prédisposition de la maladie. Les données épidémiologiques recueillies dans les études cas-témoins d'EPITHYR et dans la cohorte EPIC vont permettre d'étudier les **interactions gène-environnement** notamment entre les variants identifiés dans le cadre de cette thèse et les facteurs de risque connus liés au mode de vie ou à l'environnement. En particulier les interactions entre ces SNPs et l'obésité semblent importantes à investiguer. En effet, en plus d'être un facteur de risque du CDT clairement établi, la prévalence de l'obésité dans les populations mélanésienne et polynésienne est beaucoup plus élevée que chez les Européens vivant sur le même territoire (Frayon et al. 2020). De plus, nous avons montré que différentes voies métaboliques liées au métabolisme du cholestérol sont associées au risque de CDT.

Contrairement à d'autres cancers présentant aussi des formes familiales, aucun gène majeur à forte pénétrance n'a été à ce jour identifié pour le CDT. Les études GWAS en population générale ont permis de mettre en évidence des SNPs associés à des risques modérés de CDT ($OR < 2,5$) et donc d'un intérêt clinique limité lorsque testés individuellement pour prédire le risque de survenu de la maladie chez les personnes à plus haut risque de développer ce cancer. Cependant, la combinaison des effets de ces SNPs, exprimée en score de risque polygénique ou PRS (pour *Polygenic Risk Score*) (Lewis and Vassos 2020) pourrait avoir une utilité clinique pour améliorer le diagnostic, le pronostic ou encore le choix de la thérapie comme ceci a été suggéré pour plusieurs maladies multifactorielles, notamment d'autres cancers. Dans le cadre des CDT, le PRS le plus récent a été construit à partir des données GWAS de la méta-analyse de Gudmundsson (Gudmundsson et al. 2017). Dans cette étude, le PRS calculé à partir de 10 SNPs est associé à un risque en moyenne sept fois plus élevé pour les individus porteurs des 10 allèles à risque en comparaison aux individus porteurs d'un seul allèle à risque (Liyanarachchi et al. 2020). Pour évaluer le pouvoir prédictif de ce PRS à discriminer les cas et les témoins, les auteurs se sont appuyés sur les données GWAS de 534 cas de CDT et 407 945 témoins de UK biobank. L'AUC obtenu à partir de ce PRS est de 0,69. L'intégration des nouveaux variants identifiés dans EPITHYR dans le PRS pourrait permettre d'améliorer cette prédiction. Un score de prédiction du risque de CDT intégrant des facteurs de risque génétiques et des facteurs liés

au mode de vie, comme cela a été fait par exemple pour le cancer du sein (Zhang et al. 2018), permettrait de mieux déterminer le risque pour chaque patient de développer la maladie. Ce score de risque pourrait par exemple être utile pour déterminer la prise en charge médicale les plus adaptés pour un patient pour lequel un microcarcinome a été détecté.

Bien que mes travaux montrent des effets semblables des variants chez les Européens et les Océaniens d'EPITHYR les facteurs de risque génétiques identifiés dans cette thèse ne suffisent pas à eux seuls à expliquer les différences d'incidence observées au sein des différents groupes ethniques. L'ensemble des pistes discutées ci-dessus font l'objet d'études actuellement menées au sein d'EPITHYR et il semble important de prendre en compte l'ensemble des facteurs génétiques et environnementaux pour mieux comprendre les disparités ethniques et géographiques des CDT.

Références bibliographiques

- Akulevich, Natallia M., Vladimir A. Saenko, Tatiana I. Rogounovitch, Valentina M. Drozd, Eugeny F. Lushnikov, Victor K. Ivanov, Norisato Mitsutake, Ryo Kominami, and Shunichi Yamashita. 2009. "Polymorphisms of DNA Damage Response Genes in Radiation-Related and Sporadic Papillary Thyroid Carcinoma." *Endocrine-Related Cancer* 16(2):491–503.
- Amos, Christopher I., Joe Dennis, Zhaoming Wang, Jinyoung Byun, Fredrick R. Schumacher, Simon A. Gayther, David J. Hunter, and Thomas A. Sellers. 2017. "The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers." *Cancer Epidemiology Biomarkers & Prevention* 26(1):126–35.
- Bachurski, Cindy J., Guan Hu Yang, Tracey A. Currier, Richard M. Gronostajski, and Dihua Hong. 2003. "Nuclear Factor I/Thyroid Transcription Factor 1 Interactions Modulate Surfactant Protein C Transcription." *Molecular and Cellular Biology* 23(24):9014.
- Bacolod, Manny D., Gunter S. Schemmann, Shuang Wang, Richard Shattock, F. Sarah, Zhaoshi Zeng, Jinru Shia, Robert F. Stengel, Norman Gerry, Tomas Kirchhoff, Bert Gold, Michael F. Christman, Kenneth Offit, and L. William. 2015. "The Signatures of Autozygosity among Patients with Colorectal Cancer." *Cancer Research* 68(8):2610–21.
- Bastos, Helder Novais, Mónica Rego Antão, Susana N. Silva, Ana Paula Azevedo, Isabel Manita, Valdemar Teixeira, Julieta Esperança Pina, Octávia Monteiro Gil, Teresa Cruz Ferreira, Edward Limbert, José Rueff, and Jorge Francisco Gaspar. 2009. "Association of Polymorphisms in Genes of the Homologous Recombination DNA Repair Pathway and Thyroid Cancer Risk." <https://Home-Liebertpub-Com.Proxy.Insermbiblio.Inist.Fr/Thy> 19(10):1067–75.
- Belin, Ruth M., Brad C. Astor, Neil R. Powe, and Paul W. Ladenson. 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 Survey (NHANES III)." *The Journal of Clinical Endocrinology & Metabolism* 89(12):6077–86.

- Bener, Abdulbari, Hanadi R. El Ayoubi, Lotfi Chouchane, Awab I. Ali, Aisha Al-Kubaisi, Haya Al-Sulaiti, and Ahmad S. Teebi. 2009. "Impact of Consanguinity on Cancer in a Highly Endogamous Population." *Asian Pacific Journal of Cancer Prevention : APJCP* 10(1):35–40.
- Bevan, S., T. Pal, C. R. Greenberg, H. Green, J. Wixey, G. Bignell, S. A. Narod, W. D. Foulkes, M. R. Stratton, and R. S. Houlston. 2001. "A Comprehensive Analysis of MNG1, TCO1, FPTC, PTEN, TSHR, and TRKA in Familial Nonmedullary Thyroid Cancer: Confirmation of Linkage to TCO1." *The Journal of Clinical Endocrinology & Metabolism* 86(8):3701–4.
- Bignell, G. R., F. Canzian, M. Shayeghi, M. Stark, Y. Y. Shugart, P. Biggs, J. Mangion, R. Hamoudi, J. Rosenblatt, P. Buu, S. Sun, S. S. Stoffer, D. E. Goldgar, G. Romeo, R. S. Houlston, S. A. Narod, M. R. Stratton, and W. D. Foulkes. 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):1123.
- Bittles, A. H. and M. L. Black. 2010. "Colloquium Paper: Consanguinity, Human Evolution, and Complex Diseases." *Proceedings of the National Academy of Sciences of the United States of America* 107(Suppl 1):1779.
- Brindel, Pauline, Françoise Doyon, Frédérique Rachédi, Jean-Louis Boissin, Joseph Sebbag, Larrys Shan, Vaiana Chungue, Frédérique Bost-Bezeaud, Patrice Petitdidier, John Paoaafaite, Joseph Teuri, and Florent de Vathaire. 2009. "Anthropometric Factors in Differentiated Thyroid Cancer in French Polynesia: A Case–Control Study." *Cancer Causes & Control* 20(5):581–90.
- Brindel, Pauline, Françoise Doyon, Frédérique Rachédi, Jean-Louis Boissin, Joseph Sebbag, Larrys Shan, Vaiana Chungue, Laure Yen Kai Sun, Frédérique Bost-Bezeaud, Patrice Petitdidier, John Paoaafaite, Joseph Teuri, and Florent de Vathaire. 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):219–29.
- Browning, Sharon R. and Brian L. Browning. 2007. "Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized

- Haplotype Clustering.” *American Journal of Human Genetics* 81(5):1084.
- Caini, Saverio, Bianca Gibelli, Domenico Palli, Calogero Saieva, Massimiliano Ruscica, and Sara Gandini. 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 and Control* 26(4):511–18.
- Canzian, F., P. Amati, H. R. Harach, J. L. Kraimps, F. Lesueur, J. Barbier, P. Levillain, G. Romeo, and D. Bonneau. 1998. “A Gene Predisposing to Familial Thyroid Tumors with Cell Oxyphilia Maps to Chromosome 19p13.2.” *American Journal of Human Genetics* 63(6):1743.
- Cao, Yijuan, Zengyan Wang, Juan Gu, Fangfang Hu, Yujuan Qi, Qianqian Yin, Qingqing Sun, Guotao Li, and Bin Quan. 2015. “Reproductive Factors but Not Hormonal Factors Associated with Thyroid Cancer Risk: A Systematic Review and Meta-Analysis.” *BioMed Research International* 2015.
- Cardis, Elisabeth, Ausrele Kesminiene, Victor Ivanov, Irina Malakhova, Yoshisada Shibata, Valeryi Khrouch, Vladimir Drozdovitch, Evaldas Maceika, Irina Zvonova, Oleg Vlassov, André Bouville, Guennadi Goulko, Masaharu Hoshi, Alexander Abrosimov, Jadvyga Anoshko, Larisa Astakhova, Sergey Chekin, Evgenyi Demidchik, Rosaria Galanti, Masahiro Ito, Elena Korobova, Evgenyi Lushnikov, Marat Maksioutov, Vladimir Masyakin, Alexander Nerovnia, Vladimir Parshin, Evgenyi Parshkov, Nikolay Pilipstevich, Aldo Pinchera, Semyon Polyakov, Nina Shabeka, Eero Suonio, Vanessa Tenet, Anatoli Tsyb, Shunichi Yamashita, and Dillwyn Williams. 2005. “Risk of Thyroid Cancer After Exposure to ¹³¹I in Childhood.” *JNCI: Journal of the National Cancer Institute* 97(10):724–32.
- Cavaco, Branca M., Pedro F. Batista, Luís G. Sobrinho, and Valeriano Leite. 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 & Metabolism* 93(11):4426–30.
- Cavalli-Sforza, L. Luca. 2005. “The Human Genome Diversity Project: Past, Present and Future.” *Nature Reviews Genetics* 6(4):333–40.
- Ceballos, Francisco C., Peter K. Joshi, David W. Clark, Michèle Ramsay, and James F. Wilson. 2018. “Runs of Homozygosity: Windows into Population History and Trait Architecture.”

- Cetta, F., G. Montalto, M. Gori, MC. Curia, Cama A, and Olschwang S. 2000. "Germline Mutations of the APC Gene in Patients with Familial Adenomatous Polyposis-Associated Thyroid Carcinoma: Results from a European Cooperative Study." *The Journal of Clinical Endocrinology and Metabolism* 85(1):286–92.
- Cetta, Francesco, Gabriele Ugolini, Leonardo Barellini, Serenella Civitelli, and Mario Carmellini. 2011. "FAP Associated Cribriform Morular Variant of PTC: Striking Female Prevalence and Indolent Course [Letter to the Editor]." *Endocrine Journal* 58(9):817–18.
- Chakraborty, P. and F. Grosse. 2010. "WRN Helicase Unwinds Okazaki Fragment-like Hybrids in a Reaction Stimulated by the Human DHX9 Helicase." *Nucleic Acids Research* 38(14):4722–30.
- Chen, Han, Chaolong Wang, Matthew P. Conomos, Adrienne M. Stilp, Zilin Li, Tamar Sofer, Adam A. Szpiro, Wei Chen, John M. Brehm, Juan C. Celedó, Susan Redline, George J. Papanicolaou, Timothy A. Thornton, Cathy C. Laurie, Kenneth Rice, and Xihong Lin. 2016. "Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models." *Am J Hum Genet.* 98(4):653-.
- Cheng, Jiqiu, Jonas Demeulemeester, David C. Wedge, Hans Kristian M. Vollan, Jason J. Pitt, Hege G. Russnes, Bina P. Pandey, Gro Nilsen, Silje Nord, Graham R. Bignell, Kevin P. White, Anne-Lise Børresen-Dale, Peter J. Campbell, Vessela N. Kristensen, Michael R. Stratton, Ole Christian Lingjærde, Yves Moreau, and Peter Van Loo. 2017. "Pan-Cancer Analysis of Homozygous Deletions in Primary Tumours Uncovers Rare Tumour Suppressors." *Nature Communications* 8(1).
- Chiang, Feng-Yu, Che-Wei Wu, Pi-Jung Hsiao, Wen-Rei Kuo, Ka-Wo Lee, Jen-Chih Lin, Yi-Chu Liao, and Suh-Hang Hank Juo. 2008. "Association between Polymorphisms in DNA Base Excision Repair Genes XRCC1, APE1, and ADPRT and Differentiated Thyroid Carcinoma." *Clinical Cancer Research* 14(18):5919–24.
- Cho, Ara, Yoosoo Chang, Jiin Ahn, Hocheol Shin, and Seungho Ryu. 2018. "Cigarette Smoking and Thyroid Cancer Risk: A Cohort Study." *British Journal of Cancer* 119(5):638.
- Cho, Young Ae and Jeongseon Kim. 2014. "Thyroid Cancer Risk and Smoking Status: A Meta-Analysis." *Cancer Causes & Control* 25:9 25(9):1187–95.
- Cipollini, Monica, Susana Pastor, Federica Gemignani, Juan Castell, Sonia Garritano,

- Alessandra Bonotti, Josefina Biarnés, Gisella Figlioli, Cristina Romei, Ricard Marcos, Alfonso Cristaudo, Rossella Elisei, Stefano Landi, and Antonia Velázquez. 2013. “TPO Genetic Variants and Risk of Differentiated Thyroid Carcinoma in Two European Populations.” *International Journal of Cancer* 133(12):2843–51.
- Clarke, Laura, Xiangqun Zheng-Bradley, Richard Smith, Eugene Kulesha, Chunlin Xiao, Iliana Toneva, Brendan Vaughan, Don Preuss, Rasko Leinonen, Martin Shumway, Stephen Sherry, Paul Flicek, and The 1000 Genomes Project Consortium. 2012. “The 1000 Genomes Project: Data Management and Community Access.” *Nature Methods* 9(5):459.
- Clavel-Chapelon, Françoise, Gwenaëlle Guillas, Laura Tondeur, Céline Kernaléguen, and Marie Christine Boutron-Ruault. 2010. “Risk of Differentiated Thyroid Cancer in Relation to Adult Weight, Height and Body Shape over Life: The French E3N Cohort.” *International Journal of Cancer* 126(12):2984–90.
- Cléro, Énora, Christophe Leux, Pauline Brindel, Thérèse Truong, Antoinette Anger, Cécile Teinturier, Ibrahima Diallo, Françoise Doyon, Pascal Guénel, and Florent de Vathaire. 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* 20(11):1285–93.
- Colonna, M., P. Grosclaude, L. Remontet, C. Schvartz, J. Mace-Lesech, M. Velten, A. V. Guizard, B. Tretarre, A. Buemi, P. Arveux, and J. Esteve. 2002. “Incidence of Thyroid Cancer in Adults Recorded by French Cancer Registries (1978–1997).” *European Journal of Cancer* 38(13):1762–68.
- Colonna, Marc, Françoise Borson-Chazot, Patricia Delafosse, Claire Schvartz, and Anne Valérie Guizard. 2020. “Progression of Incidence and Estimate of Net Survival from Papillary Thyroid Cancers Diagnosed between 2008 and 2016 in France.” *Annales d’Endocrinologie* 81(6):530–38.
- Cordina-Duverger, Emilie, Christophe Leux, Monica Neri, Catherine Tcheandjieu, Anne-Valérie Guizard, Claire Schvartz, Thérèse Truong, and Pascal Guénel. 2017. “Hormonal and Reproductive Risk Factors of Papillary Thyroid Cancer: A Population-Based Case-Control Study in France.” *Cancer Epidemiology Biomarkers & Prevention* 48:78–84.
- Czene, Kamila, Paul Lichtenstein, and Kari Hemminki. 2002. “Environmental and Heritable Causes of Cancer among 9.6 Million Individuals in the Swedish Family-Cancer Database.”

International Journal of Cancer 99(2):260–66.

- Diquigiovanni, Chiara, Christian Bergamini, Cecilia Evangelisti, Federica Isidori, Andrea Vettori, Natascia Tiso, Francesco Argenton, Anna Costanzini, Luisa Iommarini, Hima Anbunathan, Uberto Pagotto, Andrea Repaci, Giulia Babbi, Rita Casadio, Giorgio Lenaz, Kerry J. Rhoden, Anna Maria Porcelli, Romana Fato, Anne Bowcock, Marco Seri, Giovanni Romeo, and Elena Bonora. 2018. “Mutant MYO1F Alters the Mitochondrial Network and Induces Tumor Proliferation in Thyroid Cancer.” *International Journal of Cancer* 143(7):1706–19.
- Dossus, Laure, Silvia Franceschi, Carine Biessy, Anne Sophie Navionis, Ruth C. Travis, Elisabete Weiderpass, Augustin Scalbert, Isabelle Romieu, Anne Tjønneland, Anja Olsen, Kim Overvad, Marie Christine Boutron-Ruault, Fabrice Bonnet, Agnès Fournier, Renee T. Fortner, Rudolf Kaaks, Krasimira Aleksandrova, Antonia Trichopoulou, Carlo La Vecchia, Eleni Peppas, Rosario Tumino, Salvatore Panico, Domenico Palli, Claudia Agnoli, Paolo Vineis, H. Bas Bueno-de-Mesquita, Petra H. Peeters, Guri Skeie, Raul Zamora-Ros, María Dolores Chirlaque, Eva Ardanaz, Maria Jose Sánchez, Jose Ramón Quirós, Miren Dorronsoro, Maria Sandström, Lena Maria Nilsson, Julie A. Schmidt, Kay Tee Khaw, Konstantinos K. Tsilidis, Dagfinn Aune, Elio Riboli, and Sabina Rinaldi. 2018. “Adipokines and Inflammation Markers and Risk of Differentiated Thyroid Carcinoma: The EPIC Study.” *International Journal of Cancer* 142(7):1332–42.
- Drozdovitch, Vladimir, André Bouville, Tetuaura Tetuanui, Marc Taquet, Jacques Gardon, Constance Xhaard, Yan Ren, Françoise Doyon, and Florent de Vathaire. 2019. “Behavior and Food Consumption Pattern of the French Polynesian Population in the 1960s –1970s.” *Asian Pacific Journal of Cancer Prevention : APJCP* 20(12):3667.
- E Ron, R A Kleinerman, J D Boice Jr, V A LiVolsi, J T Flannery, and J F Fraumeni Jr. 1987. “A Population-Based Case-Control Study of Thyroid Cancer - PubMed.” *J Natl Cancer Inst* 79:1–12.
- Ellison, LF. and T. Bushnik. 2020. “Changing Trends in Thyroid Cancer Incidence in Canada: A Histologic Examination, 1992 to 2016.” *Health Reports* 31(1):15–25.
- Enewold, Lindsey, Kangmin Zhu, Elaine Ron, Aizen J. Marrogi, Alexander Stojadinovic, George E. Peoples, and Susan S. Devesa. 2009. “Rising Thyroid Cancer Incidence in the United States by Demographic and Tumor Characteristics, 1980–2005.” *Cancer*

Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology 18(3):784.

Epstein, CJ., GM. Martin, AL. Schultz, and AG. Motulsky. 1966. “Werner’s Syndrome a Review of Its Symptomatology, Natural History, Pathologic Features, Genetics and Relationship to the Natural Aging Process.” *Medicine* 45(3):177–221.

Etienne-Manneville, Sandrine. 2004. “Cdc42 - the Centre of Polarity.” *Journal of Cell Science* 117(8):1291–1300.

Franceschi, Silvia, Susan Preston-Martin, Luigino Dal Maso, Eva Negri, Carlo La Vecchia, Wendy J. Mack, Anne McTiernan, Laurence Kolonel, Steven D. Mark, Kiyohiko Mabuchi, Fan Jin, Gun Wingren, Rosaria Galanti, Arne Hallquist, Eystein Glatte, Eiliv Lund, Fabio Levi, Dimitrios Linos, and Elaine Ron. 1999. “A Pooled Analysis of Case±control Studies of Thyroid Cancer. IV. Benign Thyroid Diseases.”

Frayon, Stéphane, Guillaume Wattelez, Emilie Paufigue, Akila Nedjar-Guerre, Christophe Serra-Mallol, and Olivier Galy. 2020. “Overweight in the Pluri-Ethnic Adolescent Population of New Caledonia: Dietary Patterns, Sleep Duration and Screen Time.” *The Lancet Regional Health - Western Pacific* 2.

Frio, T. Rio, A. Bahubeshi, C. Kanellopoulou, N. Hamel, Niedziela M, Sabbaghian N, Pouchet C, Gilbert L, O’Brien PK, Serfas K, Broderick P, Houlston RS, Lesueur F, Bonora E, Muljo S, Schimke RN, Bouron-Dal Soglio D, Arseneau J, Schultz KA, Priest JR, Nguyen VH, Harach HR, Livingston DM, Foulkes WD, and Tischkowitz M. 2011. “DICER1 Mutations in Familial Multinodular Goiter with and without Ovarian Sertoli-Leydig Cell Tumors.” *JAMA* 305(1):68–77.

Galanti, M. R., A. Ekbom, L. Grimelius, and J. Yuen. 1997. “Parental Cancer and Risk of Papillary and Follicular Thyroid Carcinoma.” *British Journal of Cancer* 75(3):451.

Gazal, Steven, Mourad Sahbatou, Marie-Claude Babron, Emmanuelle Génin, and Anne-Louise Leutenegger. 2014. “FSuite: Exploiting Inbreeding in Dense SNP Chip and Exome Data.” *Bioinformatics* 30(13):1940–41.

Ge, Minghua, Meng Shi, Changming An, Wenjun Yang, Xilin Nie, Jian Zhang, Zheng Lv, Jinliang Li, Liqing Zhou, Zhongli Du, and Ming Yang. 2016. “Functional Evaluation of TERT-CLPTM1L Genetic Variants Associated with Susceptibility of Papillary Thyroid

Carcinoma.” *Scientific Reports* 6.

- Génin, Emmanuelle, Mourad Sahbatou, Steven Gazal, Marie-Claude Babron, Hervé Perdry, and Anne-Louise Leutenegger. 2012. “Could Inbred Cases Identified in GWAS Data Succeed in Detecting Rare Recessive Variants Where Affected Sib-Pairs Have Failed?” *Human Heredity* 74(3–4):142–52.
- Goldgar, D. E., D. F. Easton, L. A. Cannon-Albright, and M. H. Skolnick. 1994. “Systematic Population-Based Assessment of Cancer Risk in First-Degree Relatives of Cancer Proband.” *Journal of the National Cancer Institute* 86(21):1600–1608.
- Grossman, RF., SH. Tu, QY. Duh, AE. Siperstein, F. Novosolov, and OH. Clark. 1995. “Familial Nonmedullary Thyroid Cancer. An Emerging Entity That Warrants Aggressive Treatment.” *Archives of Surgery (Chicago, Ill. : 1960)* 130(8):892–99.
- Gudmundsson, Julius, Patrick Sulem, Daniel F. Gudbjartsson, Jon G. Jonasson, Gisli Masson, Huiling He, Aslaug Jonasdottir, Asgeir Sigurdsson, Simon N. Stacey, Hrefna Johannsdottir, Hafdis Th Helgadóttir, Wei Li, Rebecca Nagy, Matthew D. Ringel, Richard T. Kloos, Marieke C. H. de Visser, Theo S. Plantinga, Martin den Heijer, Esperanza Aguillo, Angeles Panadero, Enrique Prats, Almudena Garcia-Castaño, Ana De Juan, Fernando Rivera, G. Bragi Walters, Hjordis Bjarnason, Laufey Tryggvadottir, Gudmundur I. Eyjolfsson, Unnur S. Bjornsdottir, Hilma Holm, Isleifur Olafsson, Kristleifur Kristjansson, Hoskuldur Kristvinsson, Olafur T Magnusson, Gudmar Thorleifsson, Jeffrey R. Gulcher, Augustine Kong, Lambertus ALM Kiemeney, Thorvaldur Jonsson, Hannes Hjartarson, Jose I. Mayordomo, Romana T. Netea-Maier, Albert de la Chapelle, Jon Hrafnkelsson, Unnur Thorsteinsdottir, Thorunn Rafnar, and Kari Stefansson. 2012. “Discovery of Common Variants Associated with Low TSH Levels and Thyroid Cancer Risk.” *Nature Genetics* 44(3):319–22.
- Gudmundsson, Julius, Patrick Sulem, Daniel F. Gudbjartsson, Jon G. Jonasson, Asgeir Sigurdsson, Jon T. Bergthorsson, Huiling He, Thorarinn Blondal, Frank Geller, Margret Jakobsdottir, Droplaug N. Magnusdottir, Sigurborg Matthiasdottir, Simon N. Stacey, Oskar B. Skarphedinsson, Hafdis Helgadóttir, Wei Li, Rebecca Nagy, Esperanza Aguillo, Eduardo Faure, Enrique Prats, Berta Saez, Mariano Martinez, Gudmundur I. Eyjolfsson, Unnur S. Bjornsdottir, Hilma Holm, Kristleifur Kristjansson, Michael L. Frigge, Hoskuldur Kristvinsson, Jeffrey R. Gulcher, Thorvaldur Jonsson, Thorunn Rafnar, Hannes

- Hjartarsson, Jose I. Mayordomo, Albert De La Chapelle, Jon Hrafnkelsson, Unnur Thorsteinsdottir, Augustine Kong, and Kari Stefansson. 2009. "Common Variants on 9q22.33 and 14q13.3 Predispose to Thyroid Cancer in European Populations." *Nature Genetics* 41(4):460–64.
- Gudmundsson, Julius, Gudmar Thorleifsson, Jon K. Sigurdsson, Lilja Stefansdottir, Jon G. Jonasson, Sigurjon A. Gudjonsson, Daniel F. Gudbjartsson, Gisli Masson, Hrefna Johannsdottir, Gisli H. Halldorsson, Simon N. Stacey, Hannes Helgason, Patrick Sulem, Leigha Senter, Huiling He, Sandya Liyanarachchi, Matthew D. Ringel, Esperanza Aguillo, Angeles Panadero, Enrique Prats, Almudena Garcia-Castanõ, Ana De Juan, Fernando Rivera, Li Xu, Lambertus A. Kiemeny, Gudmundur I. Eyjolfsson, Olof Sigurdardottir, Isleifur Olafsson, Hoskuldur Kristvinsson, Romana T. Netea-Maier, Thorvaldur Jonsson, Jose I. Mayordomo, Theo S. Plantinga, Hannes Hjartarson, Jon Hrafnkelsson, Erich M. Sturgis, Unnur Thorsteinsdottir, Thorunn Rafnar, Albert De La Chapelle, and Kari Stefansson. 2017. "A Genome-Wide Association Study Yields Five Novel Thyroid Cancer Risk Loci." *Nature Communications* 8:6–13.
- Guignard, Romain, Thérèse Truong, Yannick Rougier, Dominique Baron-Dubourdieu, and Pascal Guénel. 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):1140–49.
- Guilhen, Ana C. T., Natassia E. Bufalo, Elaine C. Morari, Janaina L. Leite, Ligia V. M. Assumpcao, Alfio J. A. Tincani, and Laura S. Ward. 2009. "Role of the N-Acetyltransferase 2 Detoxification System in Thyroid Cancer Susceptibility." *Clinical Cancer Research* 15(1):406–12.
- Han, Buhm and Eleazar Eskin. 2011. "Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-Wide Association Studies." *American Journal of Human Genetics* 88(5):586.
- Han, ZR., XL. Jiang, and WC. Fan. 2021. "LAMC1 Is Related to the Poor Prognosis of Patients with Gastric Cancer and Facilitates Cancer Cell Malignancies." *Neoplasma* 68(4):711–18.
- He, Huiling, Agnieszka Bronisz, Sandya Liyanarachchi, Rebecca Nagy, Wei Li, Yungui Huang, Keiko Akagi, Motoyasu Saji, Dorota Kula, Anna Wojcicka, Nikhil Sebastian, Bernard Wen, Zbigniew Puch, Michal Kalembe, Elzbieta Stachlewska, Malgorzata

- Czetwertynska, Joanna Dlugosinska, Kinga Dymecka, Rafal Ploski, Marek Krawczyk, Patrick J. Morrison, Matthew D. Ringel, Richard T. Kloos, Krystian Jazdzewski, David E. Symer, Veronica J. Vieland, Michael Ostrowski, Barbara Jarzab, and Albert de la Chapelle. 2013. "SRGAP1 Is a Candidate Gene for Papillary Thyroid Carcinoma Susceptibility." *The Journal of Clinical Endocrinology and Metabolism* 98(5):E973.
- He, Huiling, Wei Li, Sandya Liyanarachchi, Jaroslaw Jendrzewski, Mukund Srinivas, Ramana V Davuluri, Rebecca Nagy, and Albert de la Chapelle. 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):E164.
- He, Huiling, Wei Li, Sandya Liyanarachchi, Yanqiang Wang, Lianbo Yu, Luke K. Genutis, Sophia Maharry, John E. Phay, Rulong Shen, Pamela Brock, and Albert De La Chapelle. 2018. "The Role of NRG1 in the Predisposition to Papillary Thyroid Carcinoma." *J Clin Endocrinol Metab* 103:1369–79.
- He, Huiling, Rebecca Nagy, Sandya Liyanarachchi, Hong Jiao, Wei Li, Saul Suster, Juha Kere, and Albert de la Chapelle. 2009. "A Susceptibility Locus for Papillary Thyroid Carcinoma on Chromosome 8q24." *Cancer Research* 69(2):625.
- Hemminki, K. and C. Eng. 2004. "Clinical Genetic Counselling for Familial Cancers Requires Reliable Data on Familial Cancer Risks and General Action Plans." *Journal of Medical Genetics* 41(11):801.
- Hemminki, K. and P. Vaittinen. 1999. "Familial Cancers in a Nationwide Family Cancer Database: Age Distribution and Prevalence." *European Journal of Cancer* 35(7):1109–17.
- Hemminki, Kari and Chuanhui Dong. 2000. "FAMILIAL RELATIONSHIPS IN THYROID CANCER BY HISTO-PATHOLOGICAL TYPE." *J. Cancer* 85:201–5.
- Hemminki, Kari, Charis Eng, and Bowang Chen. 2005. "Familial Risks for Nonmedullary Thyroid Cancer." *Familial Risks for Nonmedullary Thyroid Cancer. Journal of Clinical Endocrinology and Metabolism*, 90(10), 57.
- Hemminki, Kari, Jan Sundquist, and Justo Lorenzo Bermejo. 2008. "Familial Risks for Cancer as the Basis for Evidence-Based Clinical Referral and Counseling." *The Oncologist* 13(3):239–47.
- Hemminki, Kari, Kristina Sundquist, Jan Sundquist, Asta Försti, Akseli Hemminki, and Xinjun

- Li. 2021. "Familial Risks and Proportions Describing Population Landscape of Familial Cancer." *Cancers* 13(17):4385.
- Hernández, A., N. Xamena, J. Surrallés, P. Galofré, A. Velázquez, Creus A, and Marcos R. 2008. "Role of GST and NAT2 Polymorphisms in Thyroid Cancer." *Journal of Endocrinological Investigation* 31(11):1025–31.
- Herzig, AF., T. Nutile, MC. Babron, M. Ciullo, C. Bellenguez, and AL. Leutenegger. 2018. "Strategies for Phasing and Imputation in a Population Isolate." *Genetic Epidemiology* 42(2):201–13.
- Hill, D. Ashley, J. Ivanovich, John R. Priest, Christina A. Gurnett, Louis P. Dehner, David Desruisseau, Jason A. Jarzembowski, Kathryn A. Wikenheiser-Brokamp, Brian K. Suarez, Alison J. Whelan, Gretchen Williams, Dawn Bracamontes, Yoav Messinger, and Paul J. Goodfellow. 2009. "Germline DICER1 Mutations in Familial Pleuropulmonary Blastoma." *Science (New York, N.Y.)* 325(5943):965.
- Ho, Tang, Guojun Li, Jiachun Lu, Chong Zhao, Qingyi Wei, and Erich M. Sturgis. 2009. "Association of XRCC1 Polymorphisms and Risk of Differentiated Thyroid Carcinoma: A Case–Control Analysis." <https://Home-Liebertpub-Com.Proxy.Insermbiblio.Inist.Fr/Thy> 19(2):129–35.
- Horn-Ross, Pamela L., Alison J. Canchola, Huiyan Ma, Peggy Reynolds, and Leslie Bernstein. 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):1751.
- Horn-Ross, PL., Morris. JS, M. Lee, West. DW, Whittemore. AS, McDougall IR, Nowels K, Stewart SL, Spate VL, Shiau AC, and Krone MR. 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):979–85.
- Howie, Bryan, Jonathan Marchini, and Matthew Stephens. 2011. "Genotype Imputation with Thousands of Genomes." *G3 (Bethesda)* 1(6):457–7.
- Hrafinkelsson, J., H. TULINIUS, J. G. JONASSON, and H. SIGVALDASON. 2001. "Familial

- Non-Medullary Thyroid Cancer in Iceland.” *Journal of Medical Genetics* 38(3):189.
- Iribarren, Carlos, Tmirah Haselkorn, Irene S. Tekawa, and Gary D. Friedman. 2001. “Cohort Study of Thyroid Cancer in a San Francisco Bay Area Population.” *International Journal of Cancer* 93(5):745–50.
- Jajodia, Ajay, Harpreet Kaur, Kalpana Kumari, Neha Kanojia, Meenal Gupta, Ruchi Baghel, Mamta Sood, Sanjeev Jain, Rakesh K. Chadda, and Ritushree Kukreti. 2016. “Evaluation of Genetic Association of Neurodevelopment and Neuroimmunological Genes with Antipsychotic Treatment Response in Schizophrenia in Indian Populations.” *Molecular Genetics and Genomic Medicine* 4(1):18–27.
- Jendrzewski, Jaroslaw, Andrew Thomas, Sandya Liyanarachchi, Andrew Eiterman, Jerneja Tomsic, Huiling He, Hanna S. Radomska, Wei Li, Rebecca Nagy, Krzysztof Sworczak, and Albert de la Chapelle. 2015. “PTCSC3 Is Involved in Papillary Thyroid Carcinoma Development by Modulating S100A4 Gene Expression.” *The Journal of Clinical Endocrinology and Metabolism* 100(10):E1370.
- Kaaks, R., N. Slimani, and E. Riboli. 1997. “Pilot Phase Studies on the Accuracy of Dietary Intake Measurements in the EPIC Project: Overall Evaluation of Results. European Prospective Investigation into Cancer and Nutrition.” *International Journal of Epidemiology* 26 Suppl 1(SUPPL. 1).
- Kabat, Geoffrey C., Mimi Y. Kim, Jean Wactawski-Wende, and Thomas E. Rohan. 2012. “Smoking and Alcohol Consumption in Relation to Risk of Thyroid Cancer in Postmenopausal Women.” *Cancer Epidemiology* 36(4):335–40.
- Kanehisa, Minoru and Susumu Goto. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28(1):27.
- Khan, Nicholas E., Andrew J. Bauer, Kris Ann P. Schultz, Leslie Doros, Rosamma M. Decastro, Alexander Ling, Maya B. Lodish, Laura A. Harney, Ron G. Kase, Ann G. Carr, Christopher T. Rossi, Amanda Field, Anne K. Harris, Gretchen M. Williams, Louis P. Dehner, Yoav H. Messinger, D. Ashley Hill, and Douglas R. Stewart. 2017. “Quantification of Thyroid Cancer and Multinodular Goiter Risk in the DICER1 Syndrome: A Family-Based Cohort Study.” *The Journal of Clinical Endocrinology and Metabolism* 102(5):1614.
- Kichaev, Gleb, Megan Roytman, Ruth Johnson, Eleazar Eskin, Sara Lindström, Peter Kraft,

- and Bogdan Pasaniuc. 2017. “Improved Methods for Multi-Trait Fine Mapping of Pleiotropic Risk Loci.” *Bioinformatics* 33(2):248–55.
- Kilfoy, Briseis A., Tongzhang Zheng, Theodore R. Holford, Xuesong Han, Mary H. Ward, Andreas Sjodin, Yaquun Zhang, Yana Bai, Cairong Zhu, Grace L. Guo, Nathaniel Rothman, and Yawei Zhang. 2009. “International Patterns and Trends in Thyroid Cancer Incidence, 1973–2002.” *Cancer Causes & Control : CCC* 20(5):525.
- Kim, Jina, Jessica E. Gosnell, and Sanziana A. Roman. 2019. “Geographic Influences in the Global Rise of Thyroid Cancer.” *Nature Reviews Endocrinology* 2019 16:1 16(1):17–29.
- Kitahara, Cari M., Martha S. Linet, Laura E. Beane Freeman, David P. Check, Timothy R. Church, Yikyung Park, Mark P. Purdue, Catherine Schairer, and Amy Berrington de González. 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):1615.
- Kitahara, Cari M. and Julie A. Sosa. 2016. “The Changing Incidence of Thyroid Cancer.” *Nature Reviews Endocrinology* 2016 12:11 12(11):646–53.
- Kitahara, Cari M. and Julie A. Sosa. 2020. “Understanding the Ever-Changing Incidence of Thyroid Cancer.” *Nature Reviews. Endocrinology* 16(11):1.
- Kitahara, CM., ML. McCullough, S. Franceschi, S. Rinaldi, Wolk A, Neta G, Olov Adami H, Anderson K, Andreotti G, Beane Freeman LE, Bernstein L, Buring JE, Clavel-Chapelon F, De Roo LA, Gao YT, Gaziano JM, Giles GG, Håkansson N, Horn-Ross PL, Kirsh VA, Linet MS, MacInnis RJ, Orsini N, Park Y, Patel AV, Purdue MP, Riboli E, Robien K, Rohan T, Sandler DP, Schairer C, Schneider AB, Sesso HD, Shu XO, Singh PN, van den Brandt PA, Ward E, Weiderpass E, White E, Xiang YB, Zeleniuch-Jacquotte A, Zheng W, Hartge P, and Berrington de González A. 2016. “Anthropometric Factors and Thyroid Cancer Risk by Histological Subtype: Pooled Analysis of 22 Prospective Studies.” *Thyroid : Official Journal of the American Thyroid Association* 26(2):306–18.
- Köhler, Aleksandra, Bowang Chen, Federica Gemignani, Rossella Elisei, Cristina Romei, Gisella Figlioli, Monica Cipollini, Alfonso Cristaudo, Franco Bambi, Per Hoffmann, Stefan Herms, Michał Kalemba, Dorota Kula, Shelley Harris, Peter Broderick, Richard Houlston, Susana Pastor, Ricard Marcos, Antonia Velázquez, Barbara Jarzab, Kari Hemminki, Stefano Landi, and Asta Försti. 2013. “Genome-Wide Association Study on

- Differentiated Thyroid Cancer.” *The Journal of Clinical Endocrinology & Metabolism* 98(10):E1674–81.
- Kolonel, LN., JH. Hankin, LR. Wilkens, FH. Fukunaga, and MW. Hinds. 1990. “An Epidemiologic Study of Thyroid Cancer in Hawaii.” *Cancer Causes & Control: CCC* 1(3):223–34.
- Kreiger, N. and R. Parkes. 2000. “Cigarette Smoking and the Risk of Thyroid Cancer.” *European Journal of Cancer* 36(15):1969–73.
- Kulkarni, Om, Pierre-Emmanuel Sugier, Julie Guibon, Anne Boland-Augé, Christine Lonjou, Delphine Bacq-Daian, Robert Olaso, Carole Rubino, Vincent Souchard, Frédérique Rachedi, Juan Jesus Lence-Anta, Rosa Maria Ortiz, Constance Xhaard, Pierre Laurent-Puig, Claire Mulot, Anne-Valérie Guizard, Claire Schvartz, Marie-Christine Boutron-Ruault, Evgenia Ostroumova, Ausrele Kesminiene, Jean-François Deleuze, Pascal Guénel, Florent De Vathaire, Thérèse Truong, and Fabienne Lesueur. 2021. “Gene Network and Biological Pathways Associated with Susceptibility to Differentiated Thyroid Carcinoma.” *Scientific Reports* 11(1):8932.
- Kunitomi, Haruko, Yusuke Kobayashi, Ren Chin Wu, Takashi Takeda, Eiichiro Tominaga, Kouji Banno, and Daisuke Aoki. 2020. “LAMC1 Is a Prognostic Factor and a Potential Therapeutic Target in Endometrial Cancer.” *Journal of Gynecologic Oncology* 31(2).
- Lai, Yu Heng, Jessica Cheng, Dongmei Cheng, Mattie E. Feasel, Kyle D. Beste, Junmin Peng, Asma Nusrat, and Carlos S. Moreno. 2011. “SOX4 Interacts with Plakoglobin in a Wnt3a-Dependent Manner in Prostate Cancer Cells.” *BMC Cell Biology* 12.
- Landa, I. and M. Robledo. 2011. “Association Studies in Thyroid Cancer Susceptibility: Are We on the Right Track?” *Journal of Molecular Endocrinology* 47(1).
- Larson, Nicholas B., Cecilia Berardi, Paul A. Decker, Christina L. Wassel, Phillip S. Kirsch, James S. Pankow, Michele M. Sale, Mariza de Andrade, Hugues Sicotte, Weihong Tang, Naomi Q. Hanson, Michael Y. Tsai, Kent D. Taylor, and Suzette J. Bielinski. 2015. “Trans-Ethnic Meta-Analysis Identifies Common and Rare Variants Associated with Hepatocyte Growth Factor Levels in the Multi-Ethnic Study of Atherosclerosis (MESA).” *Annals of Human Genetics* 79(4):264.
- Lauby-Secretan, Béatrice, Chiara Scoccianti, Dana Loomis, Yann Grosse, Franca Bianchini, and Kurt Straif. 2016. “Body Fatness and Cancer - Viewpoint of the IARC Working

- Group.” *N Engl J Med* 375(8):794–98.
- Lauper, Julia M., Alison Krause, Thomas L. Vaughan, Raymond J. Monnat, and Jr. 2013. “Spectrum and Risk of Neoplasia in Werner Syndrome: A Systematic Review.” *PLoS ONE* 8(4).
- Lee, Sun, Yong-Min Ahn, Jee-Youn Kim, Young-Eun Cho, and Jae-Hoon Park. 2018. “Downregulation of NOP53 Ribosome Biogenesis Factor Leads to Abnormal Nuclear Division and Chromosomal Instability in Human Cervical Cancer Cells.” *Pathology & Oncology Research*.
- Lee, Teresa and Jerry Pelletier. 2016. “The Biology of DHX9 and Its Potential as a Therapeutic Target.” *Oncotarget* 7(27):42716–39.
- Lence-Anta, Juan J, Constance Xhaard, Rosa M. Ortiz, Haoiinda Kassim, Celia M. Pereda, Silvia Turcios, Milagros Velasco, Mae Chappe, Idalmis Infante, Marlene Bustillo, Anabel García, Enora Clero D-F, Stephane Maillard, Sirced Salazar, Regla Rodriguez, and Florent De Vathaire. 2014. “E-Mail Clinical Thyroidology / Original Paper Environmental, Lifestyle, and Anthropometric Risk Factors for Differentiated Thyroid Cancer in Cuba: A Case-Control Study.” *Eur Thyroid J* 3:189–96.
- Lence-Anta, Juan J., Constance Xhaard, Rosa M. Ortiz, Haoiinda Kassim, Celia M. Pereda, Silvia Turcios, Milagros Velasco, Mae Chappe, Idalmis Infante, Marlene Bustillo, Anabel García, Enora Clero, Stephane Maillard, Sirced Salazar, Regla Rodriguez, and Florent de Vathaire. 2014. “Environmental, Lifestyle, and Anthropometric Risk Factors for Differentiated Thyroid Cancer in Cuba: A Case-Control Study.” *European Thyroid Journal*.
- Lencz, T., C. Lambert, P. DeRosse, K. E. Burdick, T. V. Morgan, J. M. Kane, R. Kucherlapati, and A. K. Malhotra. 2007. “Runs of Homozygosity Reveal Highly Penetrant Recessive Loci in Schizophrenia.” *Proceedings of the National Academy of Sciences* 104(50):19942–47.
- Lencz, Todd, Christophe Lambert, Pamela DeRosse, Katherine E. Burdick, T. Vance Morgan, John M. Kane, Raju Kucherlapati, and Anil K. Malhotra. 2007. *Runs of Homozygosity Reveal Highly Penetrant Recessive Loci in Schizophrenia*.
- Lesueur, F., M. Stark, T. Tocco, H. Ayadi, M. J. Delisle, D. E. Goldgar, M. Schlumberger, G. Romeo, and F. Canzian. 1999. “Genetic Heterogeneity in Familial Nonmedullary Thyroid

- Carcinoma: Exclusion of Linkage to RET, MNG1, and TCO in 56 Families.” *The Journal of Clinical Endocrinology & Metabolism* 84(6):2157–62.
- Lettre, Guillaume, Christoph Lange, and Joel N. Hirschhorn. 2007. “Genetic Model Testing and Statistical Power in Population-Based Association Studies of Quantitative Traits.” *Genetic Epidemiology* 31(4):358–62.
- Leux, Christophe, Thérèse Truong, Claire Petit, Dominique Baron-Dubourdieu, and Pascal Guénel. 2012. “Family History of Malignant and Benign Thyroid Diseases and Risk of Thyroid Cancer: A Population-Based Case–Control Study in New Caledonia.” *Cancer Causes & Control* 23:5 23(5):745–55.
- Lewis, Cathryn M. and Evangelos Vassos. 2020. “Polygenic Risk Scores: From Research Tools to Clinical Instruments.” *Genome Medicine* 2020 12:1 12(1):1–11.
- Li, Yafang, Jinyoung Byun, Guoshuai Cai, Xiangjun Xiao, Younghun Han, Olivier Cornelis, James E. Dinulos, Joe Dennis, Douglas Easton, Ivan Gorlov, Michael F. Seldin, and Christopher I. Amos. 2016. “FastPop: A Rapid Principal Component Derived Method to Infer Intercontinental Ancestry Using Genetic Data.” *BMC Bioinformatics* 122.
- Li, Ying, Cheng Xiang, Na Shen, Lingyan Deng, Xia Luo, Peihong Yuan, Zhi Ji, Jiaoyuan Li, and Liming Cheng. 2019. “Functional Polymorphisms on Chromosome 5p15.33 Disturb Telomere Biology and Confer the Risk of Non-Small Cell Lung Cancer in Chinese Population.” *Molecular Carcinogenesis* 58(6):913–21.
- Li, Yun R. and Brendan J. Keating. 2014. “Trans-Ethnic Genome-Wide Association Studies: Advantages and Challenges of Mapping in Diverse Populations.” *Genome Medicine* 6(10).
- Liaw, D., DJ. Marsh, J. Li, PL. Dahia, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M, Eng C, and Parsons R. 1997. “Germline Mutations of the PTEN Gene in Cowden Disease, an Inherited Breast and Thyroid Cancer Syndrome.” *Nature Genetics* 16(1):64–67.
- Lim, Hyeyeun, Susan S. Devesa, Julie A. Sosa, David Check, and Cari M. Kitahara. 2017. “Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974-2013.” *JAMA* 317(13):1338–48.
- Liu, Jian, Dingyang Liu, Zhuanyi Yang, and Zhiquan Yang. 2019. “High LAMC1 Expression in Glioma Is Associated with Poor Prognosis.” *OncoTargets and Therapy* 12:4253.
- Liyanarachchi, Sandya, Julius Gudmundsson, Egil Ferkingstad, Huiling He, Jon G. Jonasson,

- Vinicius Tragante, Folkert W. Asselbergs, Li Xu, Lambertus A. Kiemeney, Romana T. Netea-Maier, Jose I. Mayordomo, Theo S. Plantinga, Hannes Hjartarson, Jon Hrafnkelsson, Erich M. Sturgis, Pamela Brock, Fadi Nabhan, Gudmar Thorleifsson, Matthew D. Ringel, Kari Stefansson, and Albert de la Chapelle. 2020. “Assessing Thyroid Cancer Risk Using Polygenic Risk Scores.” *Proceedings of the National Academy of Sciences of the United States of America* 117(11):5997.
- Liyanarachchi, Sandya, Anna Wojcicka, Wei Li, Malgorzata Czetwertynska, Elzbieta Stachlewska, Rebecca Nagy, Kevin Hoag, Bernard Wen, Rafal Ploski, Matthew D. Ringel, Izabella Kozłowicz-Gudzinska, Wojciech Gierlikowski, Krystian Jazdzewski, Huiling He, and Albert de la Chapelle. 2013. “Cumulative Risk Impact of Five Genetic Variants Associated with Papillary Thyroid Carcinoma.” *Thyroid* 23(12):1532.
- Mack, Wendy J., Susan Preston-Martin, Leslie Bernstein, and Dajun Qian. 2002. “Lifestyle and Other Risk Factors for Thyroid Cancer in Los Angeles County Females.” *Annals of Epidemiology* 12(6):395–401.
- Malchoff, Carl D. and Diana M. Malchoff. 2017. “Familial Nonmedullary Thyroid Carcinoma:” [Http://Dx.Doi.Org.Proxy.Insermbiblio.Inist.Fr/10.1177/107327480601300204](http://Dx.Doi.Org.Proxy.Insermbiblio.Inist.Fr/10.1177/107327480601300204) 13(2):106–10.
- Malchoff, Carl D., Mansoor Sarfarazi, Beatriz Tendler, Faripour Forouhar, Giles Whalen, Vijay Joshi, Andrew Arnold, and Diana M. Malchoff. 2000. “Papillary Thyroid Carcinoma Associated with Papillary Renal Neoplasia: Genetic Linkage Analysis of a Distinct Heritable Tumor Syndrome.” *The Journal of Clinical Endocrinology & Metabolism* 85(5):1758–64.
- Mancikova, Veronika, Raquel Cruz, Lucía Inglada-Pérez, Ceres Fernández-Rozadilla, Iñigo Landa, José Cameselle-Teijeiro, Catuxa Celeiro, Susana Pastor, Antonia Velázquez, Ricard Marcos, Victor Andía, Cristina Álvarez-Escolá, Amparo Meoro, Francesca Schiavi, Giuseppe Opocher, Inés Quintela, Juan Ansede-Bermejo, Clara Ruiz-Ponte, Pilar Santisteban, Mercedes Robledo, and Angel Carracedo. 2015. “Thyroid Cancer GWAS Identifies 10q26.12 and 6q14.1 as Novel Susceptibility Loci and Reveals Genetic Heterogeneity among Populations.” *International Journal of Cancer* 137(8).
- Mannathazhathu, Arathy S., Preethi S. George, Sreekala Sudhakaran, Durga Vasudevan,

- Jagathnath Krishna KM, Christopher Booth, and Aleyamma Mathew. 2019. "Reproductive Factors and Thyroid Cancer Risk: Meta-Analysis." *Head & Neck* 41(12):4199–4208.
- Marshman, E. 2003. "Insulin-like Growth Factor Binding Protein 5 and Apoptosis in Mammary Epithelial Cells." *Journal of Cell Science* 116(4):675–82.
- Matrone, A., F. Ferrari, F. Santini, and R. Elisei. 2020. "Obesity as a Risk Factor for Thyroid Cancer." *Current Opinion in Endocrinology, Diabetes, and Obesity* 27(5):358–63.
- McKay, J. D., J. Williamson, F. Lesueur, M. Stark, A. Duffield, F. Canzian, G. Romeo, and L. Hoffman. 1999. "At Least Three Genes Account for Familial Papillary Thyroid Carcinoma: TCO and MNG1 Excluded as Susceptibility Loci from a Large Tasmanian Family." *European Journal of Endocrinology* 141(2):122–25.
- McKay, James D., Fabienne Lesueur, Laurence Jonard, Alessandro Pastore, Jan Williamson, Linda Hoffman, John Burgess, Anne Duffield, Mauro Papotti, Markus Stark, Hagay Sobol, Béatrice Maes, Arnaud Murat, Helena Kääriäinen, Mireille Bertholon-Grégoire, Michele Zini, Mary Anne Rossing, Marie-Elisabeth Toubert, Françoise Bonichon, Marie Cavarec, Anne-Marie Bernard, Andrée Boneu, Frédéric Leprat, Oskar Haas, Christine Lasset, Martin Schlumberger, Federico Canzian, David E. Goldgar, and Giovanni Romeo. 2001. "Localization of a Susceptibility Gene for Familial Nonmedullary Thyroid Carcinoma to Chromosome 2q21." *American Journal of Human Genetics* 69(2):440.
- McQuillan, Ruth, Anne Louise Leutenegger, Rehab Abdel-Rahman, Christopher S. Franklin, Marijana Pericic, Lovorka Barac-Lauc, Nina Smolej-Narancic, Branka Janicijevic, Ozren Polasek, Albert Tenesa, Andrew K. MacLeod, Susan M. Farrington, Pavao Rudan, Caroline Hayward, Veronique Vitart, Igor Rudan, Sarah H. Wild, Malcolm G. Dunlop, Alan F. Wright, Harry Campbell, and James F. Wilson. 2008. "Runs of Homozygosity in European Populations." *American Journal of Human Genetics* 83(3):359–72.
- Meinhold, Cari L., Elaine Ron, Sara J. Schonfeld, Bruce H. Alexander, D. Michal Freedman, Martha S. Linet, and Amy Berrington de González. 2010. "Nonradiation Risk Factors for Thyroid Cancer in the US Radiologic Technologists Study." *American Journal of Epidemiology* 171(2):242.
- Memon, Anjum, Mohamed Darif, Khalid Al-Saleh, and Ajitha Suresh. 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):82–89.

Michailidou, Kyriaki, Per Hall, Anna Gonzalez-Neira, Maya Ghoussaini, Joe Dennis, Roger L. Milne, Marjanka K. Schmidt, Jenny Chang-Claude, Stig E. Bojesen, Manjeet K. Bolla, Qin Wang, Ed Dicks, Andrew Lee, Clare Turnbull, Nazneen Rahman, Olivia Fletcher, Julian Peto, Lorna Gibson, Isabel dos Santos Silva, Heli Nevanlinna, Taru A. Muraanen, Kristiina Aittomäki, Carl Blomqvist, Kamila Czene, Astrid Irwanto, Jianjun Liu, Quinten Waisfisz, Hanne Meijers-Heijboer, Muriel Adank, Rob B. van der Luit, Rebecca Hein, Norbert Dahmen, Lars Beckman, Alfons Meindl, Rita K. Schmutzler, Bertram Müller-Myhsok, Peter Lichtner, John L. Hopper, Melissa C. Southey, Enes Makalic, Daniel F. Schmidt, Andre G. Uitterlinden, Albert Hofman, David J. Hunter, Stephen J. Chanock, Daniel Vincent, François Bacot, Daniel C. Tessier, Sander Canisius, Lodewyk F. A Wessels, Christopher A. Haiman, Mitul Shah, Robert Luben, Judith Brown, Craig Luccarini, Nils Schoof, Keith Humphreys, Jingmei Li, Børge G. Nordestgaard, Sune F. Nielsen, Henrik Flyger, Fergus J. Couch, Xianshu Wang, Celine Vachon, Kristen N. Stevens, Diether Lambrechts, Matthieu Moisse, Robert Paridaens, Marie-Rose Christiaens, Anja Rudolph, Stefan Nickels, Dieter Flesch-Janys, Nichola Johnson, Zoe Aitken, Kirsimari Aaltonen, Tuomas Heikkinen, Annegien Broeks, Laura J. Van, C. Ellen van der Schoot, Pascal Guénel, Thérèse Truong, Pierre Laurent-Puig, Florence Menegaux, Frederik Marme, Andreas Schneeweiss, Christof Sohn, Barbara Burwinkel, M. Pilar Zamora, Jose Ignacio Arias Perez, Guillermo Pita, M. Rosario Alonso, Angela Cox, Ian W. Brock, Simon S. Cross, Malcolm W. R Reed, Elinor J. Sawyer, Ian Tomlinson, Michael J. Kerin, Nicola Miller, Brian E. Henderson, Fredrick Schumacher, Loic Le Marchand, Irene L. Andrulis, Julia A. Knight, Gord Glendon, Anna Marie Mulligan, Annika Lindblom, Sara Margolin, Maartje J. Hooning, Antoinette Hollestelle, Ans M. W van den Ouweland, Agnes Jager, Quang M. Bui, Jennifer Stone, Gillian S. Dite, Carmel Apicella, Helen Tsimiklis, Graham G. Giles, Gianluca Severi, Laura Baglietto, Peter A. Fasching, Lothar Haeberle, Arif B. Ekici, Matthias W. Beckmann, Hermann Brenner, Heiko Müller, Volker Arndt, Christa Stegmaier, Anthony Swerdlow, Alan Ashworth, Nick Orr, Michael Jones, Jonine Figueroa, Jolanta Lissowska, Louise Brinton, Mark S. Goldberg, France Labrèche, Martine Dumont, Robert Winqvist, Katri Pylkäs, Arja Jukkola-Vuorinen, Mervi Grip, Hiltrud Brauch, Ute Hamann, Thomas Brüning, Paolo Radice, Paolo Peterlongo,

- Siranoush Manoukian, Bernardo Bonanni, Peter Devilee, Rob A. E M Tollenaar, Caroline Seynaeve, Christi J. van Asperen, Anna Jakubowska, Jan Lubinski, Katarzyna Jaworska, Katarzyna Durda, Arto Mannermaa, Vesa Kataja, Veli-Matti Kosma, Jaana M. Hartikainen, Natalia V Bogdanova, Natalia N. Antonenkova, Thilo Dörk, Vessela N. Kristensen, Hoda Anton-Culver, Susan Slager, Amanda E. Toland, Stephen Edge, Florentia Fostira, Daehee Kang, Keun-Young Yoo, Dong-Young Noh, Keitaro Matsuo, Hidemi Ito, Hiroji Iwata, Aiko Sueta, Anna H. Wu, Chiu-Chen Tseng, David Van Den Berg, Daniel O. Stram, Xiao-Ou Shu, Wei Lu, Yu-Tang Gao, Hui Cai, Soo Hwang Teo, Cheng Har Yip, Sze Yee Phuah, Belinda K. Cornes, Mikael Hartman, Hui Miao, Wei Yen Lim, Jen-Hwei Sng, Kenneth Muir, Artitaya Lophatananon, Sarah Stewart-Brown, Pornthep Siriwanarangsarn, Chen-Yang Shen, Chia-Ni Hsiung, Pei-Ei Wu, Shian-Ling Ding, Suleeporn Sangrajang, Valerie Gaborieau, Paul Brennan, James McKay, William J. Blot, Lisa B. Signorello, Qiuyin Cai, Wei Zheng, Sandra Deming-Halverson, Martha Shrubsole, Jirong Long, Jacques Simard, Montse Garcia-Closas, Paul D. P. Pharoah, Georgia Chenevix-Trench, Alison M. Dunning, Javier Benitez, and Douglas F. Easton. 2013. "Large-Scale Genotyping Identifies 41 New Loci Associated with Breast Cancer Risk." *Nature Genetics* 45(4):353–61.
- Moses, Willieford, Julie Weng, and Electron Kebebew. 2011. "Prevalence, Clinicopathologic Features, and Somatic Genetic Mutation Profile in Familial Versus Sporadic Nonmedullary Thyroid Cancer." *Thyroid* 21(4):367.
- Muftuoglu, Meltem, Junko Oshima, Cayetano von Kobbe, Wen-Hsing Cheng, Dru F. Leisritz, and Vilhelm A. Bohr. 2008. "The Clinical Characteristics of Werner Syndrome: Molecular and Biochemical Diagnosis." *Human Genetics* 124(4):369.
- Nagano, Jun, Kiyohiko Mabuchi, Yasuhiko Yoshimoto, Yuzo Hayashi, Nobuo Tsuda, Charles Land, and Kazunori Kodama. 2007. "A Case-Control Study in Hiroshima and Nagasaki Examining Non-Radiation Risk Factors for Thyroid Cancer." *Journal of Epidemiology* 17(3):76.
- Nelson, Heidi D., Miranda Pappas, Bernadette Zakher, Jennifer Priest Mitchell, Leila Okinaka-Hu, and Rongwei Fu. 2014. "Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer in Women: A Systematic Review to Update the U.S. Preventive Services Task Force Recommendation." *Annals of Internal Medicine* 160(4):255–66.

- Nettore, Immacolata Cristina, Annamaria Colao, and Paolo Emidio Macchia. 2018. "Nutritional and Environmental Factors in Thyroid Carcinogenesis." *Int J Environ Res Public Health* 15(8).
- Ngeow, Joanne, Jessica Mester, Lisa A. Rybicki, Ying Ni, Mira Milas, and Charis Eng. 2011. "Incidence and Clinical Characteristics of Thyroid Cancer in Prospective Series of Individuals with Cowden and Cowden-Like Syndrome Characterized by Germline PTEN, SDH, or KLLN Alterations." *The Journal of Clinical Endocrinology and Metabolism* 96(12):E2063.
- O'Connell, Jared, Deepti Gurdasani, Olivier Delaneau, Nicola Pirastu, Sheila Ulivi, Massimiliano Cocca, Michela Traglia, Jie Huang, Jennifer E. Huffman, Igor Rudan, Ruth McQuillan, Ross M. Fraser, Harry Campbell, Ozren Polasek, Gershim Asiki, Kenneth Ekoru, Caroline Hayward, Alan F. Wright, Veronique Vitart, Pau Navarro, Jean-Francois Zagury, James F. Wilson, Daniela Toniolo, Paolo Gasparini, Nicole Soranzo, Manjinder S. Sandhu, and Jonathan Marchini. 2014. "A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness." *PLoS Genetics* 10(4).
- Orloff, Mohammed S., Li Zhang, Gurkan Bebek, and Charis Eng. 2012. "Integrative Genomic Analysis Reveals Extended Germline Homozygosity with Lung Cancer Risk in the PLCO Cohort." *PLoS ONE* 7(2).
- Orois, Aida, Sudheer K. Gara, Mireia Mora, Irene Halperin, Sandra Martínez, Rocio Alfayate, Electron Kebebew, and Josep Oriola. 2019. "NOP53 as A Candidate Modifier Locus for Familial Non-Medullary Thyroid Cancer." *Genes* 10(11).
- Paquet, Melanie, Nitin Shivappa, James R. Hébert, Dominique Baron-Dubourdieu, Marie-Christine Boutron-Ruault, Pascal Guénel, and Thérèse Truong. 2020. "Dietary Inflammatory Index and Differentiated Thyroid Carcinoma Risk: A Population-Based Case-Control Study in New Caledonia." *American Journal of Epidemiology* 189(2):95–107.
- Pemberton, Trevor J., Devin Absher, Marcus W. Feldman, Richard M. Myers, Noah A. Rosenberg, and Jun Z. Li. 2012. "Genomic Patterns of Homozygosity in Worldwide Human Populations." *American Journal of Human Genetics* 91(2):275–92.
- Pilarski, Robert. 2009. "Cowden Syndrome: A Critical Review of the Clinical Literature." *Journal of Genetic Counseling* 18(1):13–27.

- Pilarski, Robert, Randall Burt, Wendy Kohlman, Lana Pho, Kristen M. Shannon, and Elizabeth Swisher. 2013. "Cowden Syndrome and the PTEN Hamartoma Tumor Syndrome: Systematic Review and Revised Diagnostic Criteria." *JNCI: Journal of the National Cancer Institute* 105(21):1607–16.
- Pooley, Karen A., Stig E. Bojesen, Maren Weischer, Sune F. Nielsen, Deborah Thompson, Ali Amin Al Olama, Kyriaki Michailidou, Jonathan P. Tyrer, Sara Benlloch, Judith Brown, Tina Audley, Robert Luben, K. T. Khaw, David E. Neal, Freddie C. Hamdy, Jenny L. Donovan, Zsofia Kote-Jarai, Caroline Baynes, Mitul Shah, Manjeet K. Bolla, Qin Wang, Joe Dennis, Ed Dicks, Rongxi Yang, Anja Rudolph, Joellen Schildkraut, Jenny Chang-Claude, Barbara Burwinkel, Georgia Chenevix-Trench, Paul D. P. Pharoah, Andrew Berchuck, Rosalind A. Eeles, Douglas F. Easton, Alison M. Dunning, and Børge G. Nordestgaard. 2013. "A Genome-Wide Association Scan (GWAS) for Mean Telomere Length within the COGS Project: Identified Loci Show Little Association with Hormone-Related Cancer Risk." *Human Molecular Genetics* 22(24):5056.
- Porcu, Eleonora, Marco Medici, Giorgio Pistis, Claudia B. Volpato, Scott G. Wilson, Anne R. Cappola, Steffan D. Bos, Joris Deelen, Martin den Heijer, Rachel M. Freathy, Jari Lahti, Chunyu Liu, Lorna M. Lopez, Ilja M. Nolte, Jeffrey R. O'Connell, Toshiko Tanaka, Stella Trompet, Alice Arnold, Stefania Bandinelli, Marian Beekman, Stefan Böhringer, Suzanne J. Brown, Brendan M. Buckley, Clara Camaschella, Anton J. M. de Craen, Gail Davies, Marieke C. H. de Visser, Ian Ford, Tom Forsen, Timothy M. Frayling, Laura Fugazzola, Martin Gögele, Andrew T. Hattersley, Ad R. Hermus, Albert Hofman, Jeanine J. Houwing-Duistermaat, Richard A. Jensen, Eero Kajantie, Margreet Kloppenburg, Ee M. Lim, Corrado Masciullo, Stefano Mariotti, Cosetta Minelli, Braxton D. Mitchell, Ramaiah Nagaraja, Romana T. Netea-Maier, Aarno Palotie, Luca Persani, Maria G. Piras, Bruce M. Psaty, Katri Räikkönen, J. Brent Richards, Fernando Rivadeneira, Cinzia Sala, Mona M. Sabra, Naveed Sattar, Beverley M. Shields, Nicole Soranzo, John M. Starr, David J. Stott, Fred C. G. J. Sweep, Gianluca Usala, Melanie M. van der Klauw, Diana van Heemst, Alies van Mullem, Sita H. Vermeulen, W. Edward Visser, John P. Walsh, Rudi G. J. Westendorp, Elisabeth Widen, Guangju Zhai, Francesco Cucca, Ian J. Deary, Johan G. Eriksson, Luigi Ferrucci, Caroline S. Fox, J. Wouter Jukema, Lambertus A. Kiemeny, Peter P. Pramstaller, David Schlessinger, Alan R. Shuldiner, Eline P. Slagboom, André G.

- Uitterlinden, Bijay Vaidya, Theo J. Visser, Bruce H. R. Wolffenbuttel, Ingrid Meulenbelt, Jerome I. Rotter, Tim D. Spector, Andrew A. Hicks, Daniela Toniolo, Serena Sanna, Robin P. Peeters, and Silvia Naitza. 2013. "A Meta-Analysis of Thyroid-Related Traits Reveals Novel Loci and Gender-Specific Differences in the Regulation of Thyroid Function." *PLoS Genetics* 9(2).
- Prentice, H. Kato, K. Yoshimoto, and M. Mason. 1982. "Radiation Exposure and Thyroid Cancer Incidence among Hiroshima and Nagasaki Residents." *Epidemiology*.
- Razzaghi, Hamid, Stephanie A. Santorico, and M. Ilyas Kamboh. 2012. "Population-Based Resequencing of LIPG and ZNF202 Genes in Subjects with Extreme HDL Levels." *Frontiers in Genetics* 3(JUN).
- Riboli, E. 1992. "Nutrition and Cancer: Background and Rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC)." *Annals of Oncology* 3:783–91.
- Riboli, E., KJ Hunt, N. Slimani, P. Ferrari, T. Norat, M. Fahey, UR Charrondière, B. Hémon, C. Casagrande, J. Vignat, K. Overvad, A. Tjønneland, F. Clavel-Chapelon, A. Thiébaud, J. Wahrendorf, H. Boeing, D. Trichopoulos, A. Trichopoulou, P. Vineis, D. Palli, HB Bueno-de-Mesquita, PHM Peeters, E. Lund, D. Engeset, CA González, A. Barricarte, G. Berglund, G. Hallmans, NE Day, TJ Key, R. Kaaks, and R. Saracci. 2002. "European Prospective Investigation into Cancer and Nutrition (EPIC): Study Populations and Data Collection." *Public Health Nutrition* 5(6b):1113–24.
- Rode, Line, Børge G. Nordestgaard, and Stig E. Bojesen. 2016. "Long Telomeres and Cancer Risk among 95 568 Individuals from the General Population." *International Journal of Epidemiology* 45(5):1634–43.
- Rossing, Cushing KL, Voigt LF, Wicklund KG, and Daling JR. 2000. "Risk of Papillary Thyroid Cancer in Women in Relation to Smoking and Alcohol Consumption." *Epidemiology (Cambridge, Mass.)* 11(1):49–54.
- Schlegel, Brian P., Lea M. Starita, and Jeffrey D. Parvin. 2003. "Overexpression of a Protein Fragment of Rna Helicase a Causes Inhibition of Endogenous Brca1 Function and Defects in Ploidy and Cytokinesis in Mammary Epithelial Cells." *Oncogene* 22(7):983–91.
- Sharif, Kassem, Shmuel Tiosano, Abdulla Watad, Doron Comaneshter, Arnon D. Cohen, Yehuda Shoenfeld, and Howard Amital. 2018. "The Link between Schizophrenia and Hypothyroidism: A Population-Based Study." *Immunologic Research* 66(6):663–67.

- Shibata, Akiko, Shigehira Saji, Kenji Kamiya, and Seiji Yasumura. 2021. "Statistical Data Trend in Cancer Incidence and Mortality in Fukushima From 2008 Through 2015." *J Epidemiol*.
- Silvera, Stephanie A. Navarro, Anthony B. Miller, and Thomas E. Rohan. 2005. "Risk Factors for Thyroid Cancer: A Prospective Cohort Study." *International Journal of Cancer* 116(3):433–38.
- Sippel, Rebecca S., Nadine R. Caron, and Orlo H. Clark. 2007. "An Evidence-Based Approach to Familial Nonmedullary Thyroid Cancer: Screening, Clinical Management, and Follow-Up." *World Journal of Surgery* 2007 31:5 31(5):924–33.
- Smigielski, Elizabeth M., Karl Sirotkin, Minghong Ward, and Stephen T. Sherry. 2000. "DbSNP: A Database of Single Nucleotide Polymorphisms." *Nucleic Acids Research* 28(1):352.
- Soldin, Offie P., Bethany E. Goughenour, Sappho Z. Gilbert, Helain J. Landy, and Steven J. Soldin. 2009. "Thyroid Hormone Levels Associated with Active and Passive Cigarette Smoking." *Thyroid* 19(8):817.
- Son, Ho-Young, Yul Hwangbo, Seong-Keun Yoo, Sun-Wha Im, San Duk Yang, Soo-Jung Kwak, Min Seon Park, Soo Heon Kwak, Sun Wook Cho, Jun Sun Ryu, Jeongseon Kim, Yuh-Seog Jung, Tae Hyun Kim, Su-jin Kim, Kyu Eun Lee, Do Joon Park, Nam Han Cho, Joohon Sung, Jeong-Sun Seo, Eun Kyung Lee, Young Joo Park, and Jong-Il Kim. 2017. "Genome-Wide Association and Expression Quantitative Trait Loci Studies Identify Multiple Susceptibility Loci for Thyroid Cancer." *Nature Communications* 8:15966.
- Stezhko, Valentin A., Elena E. Buglova, Larissa I. Danilova, Valentina M. Drozd, Nikolaj A. Krysenko, Nadia R. Lesnikova, Victor F. Minenko, Vladislav A. Ostapenko, Sergey V. Petrenko, Olga N. Polyanskaya, Valery A. Rzhetski, Mykola D. Tronko, Olga O. Bobylyova, Tetyana I. Bogdanova, Ovsy V. Ephstein, Iryna A. Kairo, Olexander V. Kostin, Ilya A. Likhtarev, Valentin V. Markov, Valery A. Oliynik, Viktor M. Shpak, Valeriy P. Tereshchenko, Galina A. Zamotayeva, Gilbert W. Beebe, Andre C. Bouville, Aaron B. Brill, John D. Burch, Daniel J. Fink, Ellen Greenebaum, Geoffrey R. Howe, Nickolas K. Luckyanov, Ihor J. Masnyk, Robert J. McConnell, Jacob Robbins, Terry L. Thomas, Paul G. Voillequé, and Lydia B. Zablotska. 2004. "A Cohort Study of Thyroid Cancer and Other Thyroid Diseases after the Chornobyl Accident: Objectives, Design and

- Methods.” *Radiation Research* 161(4):481–92.
- Stratakis, CA, NA. Courcoutsakis, A. Abati, A. Filie, Doppman JL, Carney JA, and Shawker T. 1997. “Thyroid Gland Abnormalities in Patients with the Syndrome of Spotty Skin Pigmentation, Myxomas, Endocrine Overactivity, and Schwannomas (Carney Complex).” *The Journal of Clinical Endocrinology and Metabolism* 82(7):2037–43.
- Sturgis, Erich M., Chong Zhao, Rong Zheng, and Qingyi Wei. 2005. “Radiation Response Genotype and Risk of Differentiated Thyroid Cancer: A Case-Control Analysis.” *The Laryngoscope* 115(6):938–45.
- Sun, Zhifu, Liang Wang, Bruce W. Eckloff, Bo Deng, Yi Wang, Jason A. Wampfler, Jinsung Jang, Eric D. Wieben, Jin Jen, Ming You, and Ping Yang. 2014. “Conserved Recurrent Gene Mutations Correlate with Pathway Deregulation and Clinical Outcomes of Lung Adenocarcinoma in Never-Smokers.” *BMC Medical Genomics* 7(1).
- Takahashi, Meiko, Vladimir A. Saenko, Tatiana I. Rogounovitch, Takahisa Kawaguchi, Valentina M. Drozd, Hisako Takigawa-Imamura, Natallia M. Akulevich, Chanavee Ratanajaraya, Norisato Mitsutake, Noboru Takamura, Larisa I. Danilova, Maxim L. Lushchik, Yuri E. Demidchik, Simon Heath, Ryo Yamada, Mark Lathrop, Fumihiko Matsuda, and Shunichi Yamashita. 2010. “The FOXE1 Locus Is a Major Genetic Determinant for Radiation-Related Thyroid Carcinoma in Chernobyl.” *Hum Mol Genet.* 9(12):2516.
- Tcheandjieu, Catherine, Emilie Cordina-Duverger, Claire Mulot, Dominique Baron-Dubourdieu, Anne-Valérie Guizard, Claire Schvartz, Pierre Laurent-Puig, Pascal Guénel, and Thérèse Truong. 2020. “Role of GSTM1 and GSTT1 Genotypes in Differentiated Thyroid Cancer and Interaction with Lifestyle Factors: Results from Case-Control Studies in France and New Caledonia.” *PLoS ONE* 15(1).
- Tcheandjieu, Catherine, Fabienne Lesueur, Marie Sanchez, Dominique Baron-Dubourdieu, Anne-Valerie Guizard, Claire Mulot, Pierre Laurent-Puig, Claire Schvartz, Therese Truong, and Pascal Guenel. 2016. “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.” *International Journal of Cancer* 139(3):617–27.
- Tchounwou, Paul B., Clement G. Yedjou, Anita K. Patlolla, and Dwayne J. Sutton. 2012.

“Heavy Metals Toxicity and the Environment.” *EXS* 101:133.

- Teumer, Alexander, Rajesh Rawal, Georg Homuth, Florian Ernst, Margit Heier, Matthias Evert, Frank Dombrowski, Uwe Völker, Matthias Nauck, Dörte Radke, Till Ittermann, Reiner Biffar, Angela Döring, Christian Gieger, Norman Klopp, H. Erich Wichmann, Henri Wallaschofski, Christa Meisinger, and Henry Völzke. 2011. “Genome-Wide Association Study Identifies Four Genetic Loci Associated with Thyroid Volume and Goiter Risk.” *American Journal of Human Genetics* 88(5):664.
- Thomsen, Hauke, Bowang Chen, Gisella Figlioli, Rossella Elisei, Cristina Romei, Monica Cipollini, Alfonso Cristaudo, Franco Bambi, Per Hoffmann, Stefan Herms, Stefano Landi, Kari Hemminki, Federica Gemignani, and Asta Försti. 2016. “Runs of Homozygosity and Inbreeding in Thyroid Cancer.” *BMC Cancer* 16(1):227.
- Thomsen, Hauke, Miguel Inacio Da Silva Filho, Andrea Woltmann, Robert Johansson, Jorunn E. Eyfjörd, Ute Hamann, Jonas Manjer, Kerstin Enquist-Olsson, Roger Henriksson, Stefan Herms, Per Hoffmann, Bowang Chen, Stefanie Huhn, Kari Hemminki, Per Lenner, and Asta Försti. 2015. “Inbreeding and Homozygosity in Breast Cancer Survival.” *Scientific Reports* 5.
- Tran, Thi-Van-Trinh, Cari M. Kitahara, Florent de Vathaire, Marie-Christine Boutron-Ruault, and Neige Journy. 2020. “Thyroid Dysfunction and Cancer Incidence: A Systematic Review and Meta-Analysis.” *Endocrine-Related Cancer* 27(4):245–59.
- Truong, Thérèse, Dominique Baron-Dubourdieu, Yannick Rougier, and Pascal Guénel. 2010. “Role of Dietary Iodine and Cruciferous Vegetables in Thyroid Cancer: A Countrywide Case-Control Study in New Caledonia.” *Cancer Causes & Control* 21(8):1183.
- Truong, Thérèse, Fabienne Lesueur, Pierre-Emmanuel Sugier, Julie Guibon, Constance Xhaard, Mojgan Karimi, Om Kulkarni, Elise A. Lucotte, Delphine Bacq-Daian, Anne Boland-Auge, Claire Mulot, Pierre Laurent-Puig, Claire Schvartz, Anne-Valérie Guizard, Yan Ren, Elisabeth Adjadj, Frédérique Rachédi, Francoise Borson-Chazot, Rosa Maria Ortiz, Juan J. Lence-Anta, Celia María Pereda, Daniel F. Comiskey, Huiling He, Sandya Liyanarachchi, Albert de la Chapelle, Rossella Elisei, Federica Gemignani, Hauke Thomsen, Asta Forsti, Anthony F. Herzig, Anne-Louise Leutenegger, Carole Rubino, Evgenia Ostroumova, Ausrele Kesminiene, Marie-Christine Boutron-Ruault, Jean-François Deleuze, Pascal Guénel, and Florent de Vathaire. 2021. “Multiethnic Genome-

- Wide Association Study of Differentiated Thyroid Cancer in the EPITHYR Consortium.” *International Journal of Cancer* 148(12):2935–46.
- Truong, Thérèse, Laurent Orsi, Dominique Dubourdieu, Yannick Rougier, Denis Hémon, and Pascal Guénel. 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):1056–65.
- Truong, Thérèse, Yannick Rougier, Dominique Dubourdieu, Chantal Guihenneuc-Jouyaux, Laurent Orsi, Denis Hémon, and Pascal Guénel. 2007. “Time Trends and Geographic Variations for Thyroid Cancer in New Caledonia, a Very High Incidence Area (1985–1999).” *European Journal of Cancer Prevention* 16(1):62–70.
- Tsilchorozidou, Dr Tasoula, Eleni Vafiadou, John G. Yovos, Giovanni Romeo, James McKay, Fabienne Lesueur, and Elena Bonora. 2006. “A Greek Family with a Follicular Variant of Familial Papillary Thyroid Carcinoma: TCO, MNG1, FPTC/PRN, and NMTC1 Excluded as Susceptibility Loci.” <https://Home.Liebertpub.Com/Thy> 15(12):1349–54.
- Turcios, Silvia, Juan J. Lence-Anta, Jose-Luis Santana, Celia M. Pereda, Milagros Velasco, Mae Chappe, Idalmis Infante, Marlene Bustillo, Anabel García, Enora Clero, Stephane Maillard, Regla Rodriguez, Constance Xhaard, Yan Ren, Carole Rubino, Rosa M. Ortiz, and Florent de Vathaire. 2015. “Thyroid Volume and Its Relation to Anthropometric Measures in a Healthy Cuban Population.” *European Thyroid Journal* 4(1):55.
- Vaccarella, Dal Maso L, Laversanne M, Bray F, Plummer M, and Franceschi S. 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):1127–36.
- La Vecchia, Carlo, Matteo Malvezzi, Cristina Bosetti, Werner Garavello, Paola Bertuccio, Fabio Levi, and Eva Negri. 2014. “Thyroid Cancer Mortality and Incidence: A Global Overview.” *International Journal of Cancer* 136(9):2187–95.
- Vecchia, La, Ron E, Franceschi S, Dal Maso L, Mark SD, Chatenoud L, Braga C, Preston-Martin S, McTiernan A, Kolonel L, Mabuchi K, Jin F, Wingren G, Galanti MR, Hallquist A, Lund E, Levi F, Linos D, and Negri E. 1999. “A Pooled Analysis of Case-Control Studies of Thyroid Cancer. III. Oral Contraceptives, Menopausal Replacement Therapy

- and Other Female Hormones.” *Cancer Causes & Control* : CCC 10(2):157–66.
- Vigneri and Malandrino. 2015. “The Changing Epidemiology of Thyroid Cancer: Why Is Incidence Increasing?” *Current Opinion in Oncology* 27(1):1–7.
- Vigneri, R., P. Malandrino, F. Gianì, M. Russo, and P. Vigneri. 2017. “Heavy Metals in the Volcanic Environment and Thyroid Cancer.” *Molecular and Cellular Endocrinology* 457:73–80.
- Vrins, Carlos L. J., Ruud Out, Peter van Santbrink, André van der Zee, Tokameh Mahmoudi, Martine Groenendijk, Louis M. Havekes, Theo J. C. van Berkel, Ko Willems van Dijk, and Erik A. L. Biessen. 2013. “Znf202 Affects High Density Lipoprotein Cholesterol Levels and Promotes Hepatosteatosis in Hyperlipidemic Mice.” *PLoS ONE* 8(2).
- Wang, Xu, Hui-Xiang Chua, Peng Chen, Rick Twee-Hee Ong, Xueling Sim, Weihua Zhang, Fumihiko Takeuchi, Xuanyao Liu, Chiea-Chuen Khor, Wan-Ting Tay, Ching-Yu Cheng, Chen Suo, Jianjun Liu, Tin Aung, Kee-Seng Chia, Jaspal S. Kooner, John C. Chambers, Tien-Yin Wong, E. Shyong Tai, Norihiro Kato, and Yik-Ying Teo. 2013. “Comparing Methods for Performing Trans-Ethnic Meta-Analysis of Genome-Wide Association Studies.” *Human Molecular Genetics* 22(11):2303–11.
- Weber, Frank and Charis Eng. 2005. “Editorial: Germline Variants within RET: Clinical Utility or Scientific Playtoy?” *The Journal of Clinical Endocrinology & Metabolism* 90(11):6334–36.
- Willer, Cristen J., Yun Li, and Gonçalo R. Abecasis. 2010. “METAL: Fast and Efficient Meta-Analysis of Genomewide Association Scans.” *Bioinformatics* 26(17):2190.
- Wiltshire, J., DrakeThomas M., DrakeThomas, M., UttleyLesley, and P. Balasubramanian. 2016. “Systematic Review of Trends in the Incidence Rates of Thyroid Cancer.” <https://Home-Liebertpub-Com.Proxy.Insermbiblio.Inist.Fr/Thy> 26(11):1541–52.
- Wyszynski, Asaf, Chi-chen Hong, Kristin Lam, Kyriaki Michailidou, Christian Lytle, Song Yao, Yali Zhang, Manjeet K. Bolla, Qin Wang, Joe Dennis, John L. Hopper, Melissa C. Southey, Marjanka K. Schmidt, Annegien Broeks, Kenneth Muir, Artitaya Lophatananon, Peter A. Fasching, Matthias W. Beckmann, Julian Peto, Alison M. Dunning, Georgia Chenevix-trench, Per Hall, and Michael D. Cole. 2018. “An Intergenic Risk Locus Containing an Enhancer Deletion in 2q35 Modulates Breast Cancer Risk by Deregulating IGFBP5 Expression.” *Hum Mol Genet.* 25(17):3863–76.

- Xhaard, C., Y. Ren, E. Cléro, S. Maillard, Brindel P, Rachedi F, Boissin JL, Sebbag J, Shan L, Bost-Bezeaud F, Petitdidier P, Drozdovitch V, Doyon F, Rubino C, and de Vathaire F. 2014. "Differentiated Thyroid Carcinoma Risk Factors in French Polynesia." *Asian Pacific Journal of Cancer Prevention : APJCP* 15(6):2675–80.
- Xhaard, Constance, Carole Rubino, Enora Cléro, Stéphane Maillard, Yan Ren, Françoise Borson-Chazot, Geneviève Sassolas, Claire Schvartz, Marc Colonna, Brigitte Lacour, Arlette Danzon, Michel Velten, Antoine Buemi, Laurent Bailly, Eugénia Mariné Barjoan, Martin Schlumberger, Jacques Orgiazzi, Elisabeth Adjadj, and Florent de Vathaire. 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):1007–17.
- Xhaard, Constance, Florent De Vathaire, Enora Cléro, Stéphane Maillard, Yan Ren, Françoise Borson-Chazot, Geneviève Sassolas, Claire Schvartz, Marc Colonna, Brigitte Lacour, Arlette Danzon, Michel Velten, Emilie Marrer, Laurent Bailly, Mariné Barjoan, Martin Schlumberger, Jacques Orgiazzi, Elisabeth Adjadj, and Carole Rubino. 2015. "Original Contribution Anthropometric Risk Factors for Differentiated Thyroid Cancer in Young Men and Women From Eastern France: A Case-Control Study." *Am J Epidemiol.* 182(3):202.
- Xiong, Yin, Lisa Zhang, Alisha K. Holloway, Xiaolin Wu, Ling Su, and Electron Kebebew. 2011. "MiR-886-3p Regulates Cell Proliferation and Migration, and Is Dysregulated in Familial Non-Medullary Thyroid Cancer." *PLoS ONE* 6(10).
- Xu, Li, Guojun Li, Qingyi Wei, Adel K. El-Naggar, and Erich M. Sturgis. 2012. "Family History of Cancer and Risk of Sporadic Differentiated Thyroid Carcinoma." *Cancer* 118(5):1228.
- Yang, Samantha Peiling and Joanne Ngeow. 2016. "Familial Non-Medullary Thyroid Cancer: Unraveling the Genetic Maze." *Endocrine-Related Cancer* 23(12):R577–95.
- Ye, Feng, Hongwei Gao, Lin Xiao, Zhixiang Zuo, Yueping Liu, Qi Zhao, Huijiao Chen, Weiyi Feng, Bo Fu, Linyong Sun, Xiaolin Jiang, Du He, He Jiang, Mei Yang, Li Li, Fei Chen, Xiang Liu, Shuang Li, Zhihui Li, Yong Jiang, Liang Cheng, and Hong Bu. 2019. "Whole Exome and Target Sequencing Identifies MAP2K5 as Novel Susceptibility Gene for Familial Non-Medullary Thyroid Carcinoma." *International Journal of Cancer*

144(6):1321–30.

- Yu, CE, J. Oshima, YH Fu, EM Wijsman, F Hisama, Alisch R, Matthews S, Nakura J, Miki T, Ouais S, Martin GM, Mulligan J, and Schellenberg GD. 1996. “Positional Cloning of the Werner’s Syndrome Gene.” *Science (New York, N.Y.)* 272(5259):258–62.
- Zablotska, Lydia, Eldar Nadyrov, Alexander Rozhko, Zhihong Gong, Olga Polyanskaya, Robert McConnell, Patrick O’Kane, Alina Brenner, Mark P. Little, Evgenia Ostroumova, Andre Bouville, Vladimir Drozdovitch, Viktor Minenko, Yuri Demidchik, Alexander Nerovnya, Vassilina Yauseyenko, Irina Savasteeva, Sergey Nikonovich, Kiyohiko Mabuchi, and Maureen Hatch. 2015. “Analysis of Thyroid Malignant Pathological Findings Identified during Three Rounds of Screening (1997-2008) of a Belarusian Cohort of Children and Adolescents Exposed to Radioiodines after the Chernobyl Accident.” *Cancer* 121(3):457.
- Zamora-Ros, Raul, Sabina Rinaldi, Carine Biessy, Anne Tjønneland, Jytte Halkjær, Agnes Fournier, Marie-Christine Boutron-Ruault, Sylvie Mesrine, Kaja Tikk, René T. Fortner, Heiner Boeing, Jana Förster, Antonia Trichopoulou, Dimitrios Trichopoulos, Eleni-Maria Papatesta, Giovanna Masala, Giovanna Tagliabue, Salvatore Panico, Rosario Tumino, Silvia Polidoro, Petra H. M. Peeters, H. B(as. Bueno-de-Mesquita, Elisabete Weiderpass, Eiliv Lund, Marcial Argüelles, Antonio Agudo, Esther Molina-Montes, Carmen Navarro, Aurelio Barricarte, Nerea Larrañaga, Jonas Manjer, Martin Almquist, Maria Sandström, Joakim Hennings, Konstantinos K. Tsilidis, Julie A. Schmidt, Kay-Thee Khaw, Nicholas J. Wareham, Isabelle Romieu, Graham Byrnes, Marc J. Gunter, Elio Riboli, and Silvia Franceschi. 2015. “Reproductive and Menstrual Factors and Risk of Differentiated Thyroid Carcinoma: The EPIC Study.” *International Journal of Cancer* 136(5):1218–27.
- Zayed, Ayman A., Justin Z. Amarin, Abdallah T. Al-Ani, Tareq L. Altell, Sultan S. Abdelhamid, Murad M. Qirem, Suhub M. Fahmawi, Sanad M. Elshebli, Khair M. Hamo, Louay Y. Zaghlol, Michael E. Tierney, Jana A. Zayed, and Alireza Haghighi. 2021. “Association of Parental Consanguinity With Papillary Thyroid Carcinoma: A Case-Control Study.” *The Journal of Clinical Endocrinology & Metabolism* 106(7):e2769–74.
- Zhang, Xuehong, Megan Rice, Shelley S. Tworoger, Bernard A. Rosner, A. Heather Eliassen, Rulla M. Tamimi, Amit D. Joshi, Sara Lindstrom, Jing Qian, Graham A. Colditz, Walter C. Willett, Peter Kraft, and Susan E. Hankinson. 2018. “Addition of a Polygenic Risk

- Score, Mammographic Density, and Endogenous Hormones to Existing Breast Cancer Risk Prediction Models: A Nested Case–Control Study.” *PLOS Medicine* 15(9):e1002644.
- Zhang, Yaojun, Shaoyan Xi, Jinbin Chen, Dongsheng Zhou, Hengjun Gao, Zhongguo Zhou, Li Xu, and Minshan Chen. 2017. “Overexpression of LAMC1 Predicts Poor Prognosis and Enhances Tumor Cell Invasion and Migration in Hepatocellular Carcinoma.” *Journal of Cancer* 8(15):2992.
- Zidane, M., Y. Ren, C. Xhaard, A. Leufroy, Côte S, Dewailly E, Noël L, Guérin T, Bouisset P, Bernagout S, Paoafaite J, Iltis J, Taquet M, Suhas E, Rachédi F, Boissin JL, Sebbag J, Shan L, Bost-Bezeaud F, Petitdidier P, Rubino C, Gardon J, and de Vathaire F. 2019. “Non-Essential Trace Elements Dietary Exposure in French Polynesia: Intake Assessment, Nail Bio Monitoring and Thyroid Cancer Risk.” *Asian Pacific Journal of Cancer Prevention : APJCP* 20(2):355–67.
- Zimmermann, Michael B. and Valeria Galetti. 2015. “Iodine Intake as a Risk Factor for Thyroid Cancer: A Comprehensive Review of Animal and Human Studies.” *Thyroid Research* 8(1).

Annexes

Tableau supplémentaire 1 : Polymorphismes associés au risque de CDT dans les études GWAS et dans la méta-analyse européenne.

Locus	SNP	Gène(s)	Etude ^a	N Cas/ N Témoins ^b	A_ris que ^c	ORa	IC à 95 % a	p-valeur ^a	Rep ^c
1p13.3	rs4915076	VAV3	Son et al, 2017 (6)	1085/8884	T	1,33	NA	8,47 x 10 ⁻⁰⁸	
1q42.2	rs4649295	PCNXL2	Son et al, 2017 (6)	1085/8884	T	1,43	NA	6,00 x 10 ⁻⁰⁸	
1q42.2	rs12129938	PCNXL2	Gudmundsson et al, 2017 (7)	3001/287550	A	1.32	(1,20- 1,43)	4,00 x 10 ⁻¹¹	
2q35	rs966423*	DIRC3	Gudmundsson et al, 2012 (3)	1150/41448	C	1,34	(1,22 - 1,47)	1,30 x 10 ⁻⁰⁹	
2q35	rs6759952*	DIRC3	Köhler et al, 2013 (4)	2131/2068	T	1,30	(1,18 - 1,43)	7,30 x 10 ⁻⁰⁸	
2q35	rs12990503	DIRC3	Son et al, 2017 (6)	1085/8884	G	1,34	NA	3,55 x 10 ⁻⁰⁹	
2q35	rs11693806*	DIRC3	Gudmundsson et al, 2017 (7)	3001/287550	C	1.43	(1,33- 1,54)	1,50 x 10 ⁻²⁴	
3q35	rs7617304	RARRES1	Köhler et al, 2013 (4)	2139/2068	A	1,25	(1,12 - 1,39)	4,00 x 10 ⁻⁰⁵	
3p14.2	rs9858271	FHIT	Son et al, 2017 (6)	1085/8884	G	1,26	NA	6,82 x 10 ⁻⁰⁷	
3q26.2	rs6793295	TERC, LRRC34	Gudmundsson et al, 2017 (7)	3001/287550	T	1.23	(1,15- 1,33)	2,70 x 10 ⁻⁰⁸	
4q21.1	rs1874564	SEPT11	Son et al, 2017 (6)	1085/8884	G	1,31	NA	2,04 x 10 ⁻⁰⁷	
5p15.33	rs10069690	TERT	Gudmundsson et al, 2017 (7)	3001/287550	T	1.2	(1,12- 1,29)	3,20 x 10 ⁻⁰⁷	
5q22.1	rs73227498	NREP	Gudmundsson et al, 2017 (7)	3001/287550	A	1.37	(1,23- 1,49)	3,00 x 10 ⁻¹⁰	
6q14	rs4075570	HTR1B	Mancikova et al, 2015 (5)	502/398	A	0,68	(0,51 - 0,80)	9,60 x 10 ⁻⁰⁵	

7q21	rs7800391	<i>IMMP2</i>	Köhler et al, 2013 (4)	2134/2043	T	1,25	(1,14 - 1,38)	5,7 x 10 ⁻⁰⁶	
7q21	rs10238549	<i>IMMP2</i>	Köhler et al, 2013 (4)	2142/2059	C	1,27	(1,15 - 1,40)	4,10 x 10 ⁻⁶	
8p12	rs2439302	<i>NRG1</i>	Gudmundsson et al, 2012 (3)	1134/5625	G	1,36	(1,23 - 1,50)	2,00 x 10 ⁻⁰⁹	
8p12	rs6996585	<i>NRG1</i>	Son et al, 2017 (6)	1085/8884	G	1,39	NA	1,08 x 10 ⁻¹⁰	
8p12	rs12542743	<i>NRG1</i>	Son et al, 2017 (6)	1085/8884	C	1,36	NA	4,61 x 10 ⁻¹⁰	
8p12	rs2439302	<i>NRG1</i>	Son et al, 2017 (6)	1085/8884	G	1,37	NA	1,42 x 10 ⁻⁰⁹	
8p12	rs11693806*	<i>NRG1</i>	Gudmundsson et al, 2017 (7)	3001/287550	G	1,32	(1,23- 1,41)	1,50 x 10 ⁻¹⁷	
9q22.33	rs965513	<i>FOXE1</i>	Gudmundsson et al, 2009 (1)	962/38923	A	1,75	(1,59 - 1,94)	1,10x10 ⁻²⁷	2, 3, 4, 5
9q22.33	rs7028661	<i>FOXE1</i>	Mancikova et al, 2015 (5)	2409/1819	A	1,64	(1,49 - 1,81)	1,00 x 10 ⁻²²	
9q22.33	rs10122541	<i>FOXE1</i>	Mancikova et al, 2015 (5)	2409/1819	C	1,54	(1,40 - 1,70)	1,10 x 10 ⁻¹⁷	
9q22.33	rs7037324	<i>FOXE1</i>	Mancikova et al, 2015 (5)	2409/1819	A	1,54	(1,39 - 1,70)	1,20 x 10 ⁻¹⁷	
9q22.33	rs72753537	<i>FOXE1</i>	Son et al, 2017 (6)	1085/8884	C	1,41	NA	7,67 x 10 ⁻⁰⁶	
9q22.33	rs1588635*	<i>FOXE1</i>	Gudmundsson et al, 2017 (7)	3001/287550	A	1,69	(1,59- 1,82)	2,00 x 10 ⁻⁵⁸	
9q34	rs1078500	<i>SNAPC4</i>	Köhler et al, 2013 (4)	2126/2053	T	1,23	(1,12 - 1,36)	3,5 x 10 ⁻⁰⁵	
10q24.33	rs7902587	<i>OBFC1</i>	Gudmundsson et al, 2017 (7)	3001/287550	T	1,41	(1,27- 1,56)	5,40 x 10 ⁻¹¹	
10q26	rs2997312	<i>WDR11-AS1</i>	Mancikova et al, 2015 (5)	1998/1464	A	1,35	(1,16 - 1,57)	1,20x10 ⁻⁰⁴	
10q26	rs125467	<i>WDR11-AS1</i>	Mancikova et al, 2015 (5)	2409/1817	C	1,38	(1,18 - 1,62)	5,90x10 ⁻⁰⁵	
12q14.3	rs11175834	<i>MSRB3</i>	Son et al, 2017 (6)	1085/8884	T	1,37	NA	4,26 x 10 ⁻⁰⁸	
14q13.3	rs944289	<i>NKX2-1</i>	Gudmundsson et al, 2009 (1)	1006/38345	T	1,37	(1,24 - 1,52)	2,00 x 10 ⁻⁰⁹	4, 5

14q13.3	rs116909374	<i>MBIP, NKX2-1</i>	Gudmundsson et al, 2012 (3)	1138/5340	T	2,09	(1,68 - 2,60)	4,60 x 10 ⁻¹¹ 7
14q13.3	rs34081947	<i>NKX2-1</i>	Son et al, 2017 (6)	1,085/8,884	T	1,27	NA	1,19 x 10 ⁻⁰⁷
14q13.3	rs944289	<i>NKX2-1</i>	Son et al, 2017 (6)	1,085/8,884	T	1,25	NA	1,39 x 10 ⁻⁰⁶
14q13.3	rs368187	<i>NKX2-1</i>	Gudmundsson et al, 2017 (7)	3,001/287,550	G	1.39	(1,30- 1,47)	5,10 x 10 ⁻²³
15q22.33	rs2289261	<i>SMAD3</i>	Gudmundsson et al, 2017 (7)	3,001/287,550	C	1.23	(1,15- 1,32)	3,10 x 10 ⁻⁰⁹
19p13.2	rs7248104	<i>INSR</i>	Son et al, 2017 (6)	1,085/8,884	A	1,22	NA	2,00 x10 ⁻⁰⁵

(1) Gudmundsson et al, 2009 ; (2) Takahashi et al, 2009 ; (3) Gudmundsson et al, 2012 ; (4) Köhler et al, 2013 ; (5) Mancikova et al, 2015. (6) Son et al, 2017 ; (7) Gudmundsson et al, 2017. ^a étude ayant identifié le SNP ; ^b N cas /N Témoins : nombre de cas/nombre de témoins dans l'étude initiale ; ^cA_risque : allèle à risque ; ^dRep : réplification (études ayant répliqué l'association avec le SNP) ; *SNPs situés sur le même locus et sont en déséquilibre de liaison.

Tableau supplémentaire 2 : Résultats des analyses de réplication des tops SNPs chez les Européens aux loci ayant au moins 2 SNPs associés au CDT (p -valeur $< 10^{-05}$) dans l'analyse poolée.

locus	# of SNPs with $p < 10^{-5}$	Lead SNP	position (hg19)	OA EA	Discovery set					Italy (649 cases/431 controls)					Replication sets					Combining Italy + Ohio					Combining discovery set + replication sets				
					EPITHYR Europeans (1554 cases/1973 controls)																								
					Imputation info score	EAF	OR	95%CI	p (Score) (Wald)	EAF	OR	95%CI	p	EAF	OR	95%CI	p	EAF	OR	95%CI	p	p-hetero geneity	OR	95%CI	p	Direct ion	p-hetero g eneity		
1p31.3	51	rs334729	61,608,473	C G	0.96	0.95	0.53	0.41 0.66	4.5×10^{-8} 8.7×10^{-8}	0.96	0.69	0.46 1.05	0.08	0.95	0.86	0.70 1.05	0.17	0.82	0.69 0.99	0.03	0.36	0.70	0.60 0.80	7.6×10^{-7}	---	8.6×10^{-1}			
3q12.1	7	rs4264763	98,842,761	A T	0.85	0.17	0.71	0.61 0.83	4.9×10^{-6} 8.8×10^{-6}	0.20	1.04	0.84 1.29	0.71	0.17	0.89	0.79 1.01	0.10	0.93	0.83 1.03	0.18	0.22	0.85	0.78 0.92	2.0×10^{-4}	→+	9.4×10^{-1}			
3q21.2	4	rs73199524	124,548,860	C T	0.96	0.23	0.78	0.69 0.88	5.9×10^{-6} 5.6×10^{-5}	0.22	0.96	0.78 1.18	0.70	0.25	1.00	0.98 1.03	0.96	1.00	0.97 1.03	0.89	0.67	0.99	0.96 1.02	0.46	→+	3.5×10^{-4}			
5q31.1	7	rs55840596	133,893,101	G A	0.99	0.13	0.72	0.62 0.84	6.6×10^{-6} 1.9×10^{-5}	0.13	0.99	0.77 1.28	0.95	0.12	0.98	0.89 1.08	0.77	0.98	0.89 1.07	0.66	0.93	0.90	0.83 0.97	9.4×10^{-4}	---	2.8×10^{-1}			
7q36.1	16	rs2727572	151,298,246	C T	0.96	0.45	0.80	0.72 0.88	3.4×10^{-6} 1.3×10^{-5}	0.43	0.86	0.72 1.03	0.10	0.47	0.97	0.90 1.05	0.58	0.95	0.89 1.02	0.19	0.23	0.90	0.85 0.95	3.7×10^{-4}	---	8.5×10^{-1}			
16q23.2	29	rs17767383	79,744,401	A G	geno	0.68	0.79	0.71 0.88	5.9×10^{-6} 1.5×10^{-5}	0.72	0.73	0.61 0.89	1×10^{-5}	0.69	0.86	0.78 0.94	3×10^{-5}	0.83	0.76 0.90	2.2×10^{-5}	0.15	0.81	0.76 0.87	9.7×10^{-9}	---	0.29			
17q21.33	10	rs9903280	48,412,296	T A	0.95	0.02	2.20	1.51 3.21	6.1×10^{-6} 4.6×10^{-5}	0.03	1.19	0.74 1.90	0.49	0.02	0.84	0.62 1.13	0.31	0.93	0.72 1.19	0.55	0.22	1.21	0.98 1.49	0.08	→+	4.5×10^{-4}			
18q12.3	7	rs12709703	37,909,846	G A	0.98	0.08	1.56	1.30 1.88	1.1×10^{-6} 2.9×10^{-5}	0.08	0.72	0.51 1.00	0.05	0.07	1.18	0.99 1.40	0.09	1.06	0.91 1.23	0.48	9.5×10^{-3}	1.24	1.10 1.40	4.2×10^{-4}	→+	2.4×10^{-4}			
19p12	2	rs10415826	22,248,317	A T	0.97	0.20	1.35	1.19 1.53	2.2×10^{-6} 5.0×10^{-5}	0.20	1.11	0.90 1.38	0.34	0.20	1.16	1.03 1.30	0.02	1.15	1.04 1.27	7.9×10^{-3}	0.73	1.22	1.13 1.32	8.7×10^{-7}	+++	0.15			
20q13.2	4	rs7260863	54,938,429	C T	0.95	0.01	3.07	1.84 5.11	6.0×10^{-6} 1.8×10^{-5}	0.02	1.05	0.57 1.93	0.89	0.01	0.59	0.33 1.05	0.10	0.77	0.51 1.18	0.23	0.18	1.34	0.97 1.85	0.07	→+	9.8×10^{-5}			

OA: Other allele

EAF: Effect allele frequency among controls

OR: Odds ratio (log additive)

95%CI : 95% confidence interval

p : p-value

Direction: direction of effect in studies included in the meta-analysis (EPITHYR/Italy/Ohio)

Allèle						Etude GWAS				Européens			Polynésiens			Mélanésiens		
Locus	SNPs	Position (hg19)	info score ^d	Ref	Risque	Study (year), population	EAFe	OR ^f	P ^g	EAFe	OR ^f	P ^g	EAFe	OR ^f	P ^g	EAFe	OR ^f	P ^g
2q35	rs1549738* ^c	218 118 722	> 0,98	G	A	Son (2017), Coréens	0,61	1,28	2,9x10 ⁻⁴	0,14	1,03	0,70	0,86	0,99	0,97	0,94	0,36	0,03
	rs6759952 ^a	218 271 719		C	T	Kohler(2013), Italiens	0,42	1,32	9,8x10 ⁻⁶	0,44	1,26	2,9x10 ⁻¹⁰	0,90	1,16	0,58	0,98	0,35	0,20
	rs11693806* ^a	218 292 158	> 0,98	G	C	Gudmundsson (2017), Islandais	0,28	1,49	3,6x10 ⁻¹⁴	0,30	1,43	1,1x10 ⁻¹⁰	0,83	1,49	0,09	0,92	1,10	0,77
	rs12990503 ^c	218 294 217		C	G	Son (2017), Coréens	0,68	1,32	1,8x10 ⁻⁴	0,29	0,67	3,0x10 ⁻¹⁰	0,83	1,46	0,09	0,92	1,10	0,75
	rs1382435 ^a	218 296 003		C	T	Kohler (2013), Italiens	-	1,51	7,2x10 ⁻⁶	0,35	1,32	1,2x10 ⁻⁷	0,88	1,22	0,43	0,98	0,27	0,10
	rs966423 ^b	218 310 340		T	C	Gudmundsson (2012), Islandais	0,49	1,26	3,8x10 ⁻⁴	0,43	1,27	2,6x10 ⁻⁶	0,89	1,13	0,64	0,98	0,35	0,20
						Gudmundsson (2017), Islandais	0,44	1,32	3,2x10 ⁻⁸									
						Son (2017), Coréens	-	1,24	8,1x10 ⁻³									
8p12	rs12542743 ^c	32 318 355		T	C	Son (2017), Coréens	0,25	1,42	1,1x10 ⁻⁶	0,45	1,22	5,9x10 ⁻⁵	0,63	1,05	0,77	0,65	1,26	0,22
	rs6996585 ^c	32 400 803		A	G	Son (2017), Coréens	0,30	1,48	1,2x10 ⁻⁷	0,44	1,21	5,9x10 ⁻⁸	0,54	1,09	0,57	0,63	1,16	0,42
	rs2466076 ^a	32 432 796		T	G	Gudmundsson (2017), Islandais	0,21	1,32	1,5x10 ⁻¹⁷	0,48	1,21	1,0x10 ⁻⁴	0,31	1,10	0,56	0,40	1,05	0,79
	rs2439302 ^b	32 432 369		C	G	Gudmundsson	0,45	1,41	1,3x10 ⁻⁶	0,47	1,20	2,6x10 ⁻⁴	0,30	1,10	0,55	0,40	1,05	0,79

	(2012), Islandais				
	Gudmundsson (2017), Islandais				
	Son (2017), Coréens				
	0,46	1,33	6,0x10 ⁻⁹		
	0,27	1,37	1,4x10 ⁻⁹		

Tableau supplémentaire 3 : Résultats d'association des précédents SNP GWAS identifiés aux locus 2q35 et 8p12 chez les Européens, les Mélanésiens et les Polynésiens.

* SNP imputé

^a SNP GWAS associé au risque de CDT chez les Européens

^b SNP GWAS associé au risque de CDT chez les Asiatiques et les Européens

^c SNP GWAS associé au risque de CDT chez les Asiatiques

^d Qualité d'imputation du variant, valeur de 0 à 1

^e Fréquence allele à risque

^f Odds ratios ajusté sur l'âge, les sexe et l'étude

^g P-valeur

Tableau supplémentaire 4 : Caractéristiques globales des ROH identifiés dans les groupes Nord et SUD d'EPITHYR et d'EPIC poolés.

	EUR (EPITHYR + EPIC) ^a (N=4726, 1921 ca/ 2805 témoins)			EPITHYR + EPIC SUD ^b (N=3759, 1677 ca/ 2082 témoins)			EPITHYR + EPIC NORD (N=967, 244 cas/ 723 témoins)		
- Nombre moyen de ROH identifiés / personne	5,8			5,7			7,3		
- Longueur moyenne cumulée des ROH/ personne (kb)	11 946			11 634			14 666		
	Cas	Témoins	P diff ^d	Cas	Témoins	P diff	Cas	Témoins	P diff
-Nombre moyen de ROH / personne	5,7	6,0	5,7x10 ⁻⁵	5,6	5,7	0,10	7,2	7,3	0,48
-Longueur moyenne de ROH (kb) / personne	1 817	1 861	0,19	1 803	1 846	0,25	1 835	1 877	0,48
-Longueur moyenne cumulée totale de ROH / personne (kb)	11 415	12 209	0,06	11 356	11 858	0,35	13 855	14 939	0,24

^a 316 203 SNPs ; ^b 320 695 SNPs ; ^c 314 078 SNPs ; ^d test de student

Tableau supplémentaire 5 : Association entre le nombre de ROH ou la longueur totale ROH et le risque de CDT dans les groupes NORD et SUD des données européennes EPIC et EPITHYR poolées.

	EUR (EPITHYR + EPIC) ^a (N=4726, 1921 cas/ 2805 témoins)					EPITHYR + EPIC SUD ^b (N=3759, 1677 cas/ 2082 témoins)					EPITHYR + EPIC NORD ^c (N=967, 244 cas/ 723 témoins)				
Nombre de ROH / individu ^a	Cas	Témoins	OR ^d	95 % IC	P	Cas	Témoins	OR ^b	95 % IC	P	Cas	Témoins	OR ^b	95 % IC	P
<4	503	646	1,00	Ref	Ref	400	490	1,00	Ref	Ref	24	72	1,00	Ref	Ref
4-6	609	793	1,06	0,90-1,25	0,47	573	634	1,13	0,95-1,35	0,16	54	166	0,90	0,51-1,61	0,72
6-8	406	655	0,93	0,78-1,11	0,42	378	510	0,95	0,78-1,15	0,58	171	62	1,12	0,64-2,00	0,69
≥8	403	711	0,98	0,82-1,18	0,88	326	448	0,96	0,79-1,18	0,73	104	314	1,06	0,62-1,87	0,83
Longueur totale de ROH /individu (Mb)															
<5,8	566	718	1,00	Ref	Ref	469	562	1,00	Ref	Ref	24	90	1,00	Ref	Ref
5,8-8,9	539	700	1,05	0,89-1,23	0,55	553	494	1,10	0,92-1,31	0,29	57	158	1,36	0,79-2,41	0,27
8,9-13	409	654	0,90	0,76-1,07	0,25	372	493	0,94	0,78-1,13	0,49	65	170	1,47	0,86-2,60	0,17
≥13	407	733	0,93	0,77-1,10	0,40	342	474	0,93	0,77-1,13	0,47	98	305	1,35	0,79-2,35	0,28

^a set de 316 203 SNPs ; ^b set de 320 695 SNPs ; ^c set de 314 078 SNPs ; ^d Modèle logistique sous R ajusté sur l'âge, le sexe et les 10 premières CPs.

Tableau supplémentaire 6 : Nombre de ROH partagés au sein des groupes NORD et SUD des données européennes d'EPIC et d'EPITHYR poolées.

	ROH ≥ 5 individus	ROH ≥ 500 kb, ≥ 5 individus	ROH communs (≥ 500 kb, ≥ 5 individus, 76 SNPs, même pos START-END)	ROH communs (≥ 1000 kb, ≥ 5 individus, même pos START-END)
EUR 'EPITHYR + EPIC' (316 203 SNPs)	6 678	233	84	14
EUR 'EPITHYR + EPIC' SUD (320 695 SNPs)	10 847	88	27	2
EUR 'EPITHYR + EPIC' NORD (314 078 SNPs)	4220	243	67	20

Tableau supplémentaire 7: Comparaison des coefficients de consanguinité (Froh) chez les cas et témoins des groupes NORD et SUD des données européennes d'EPIC et EPITHYR poolées.

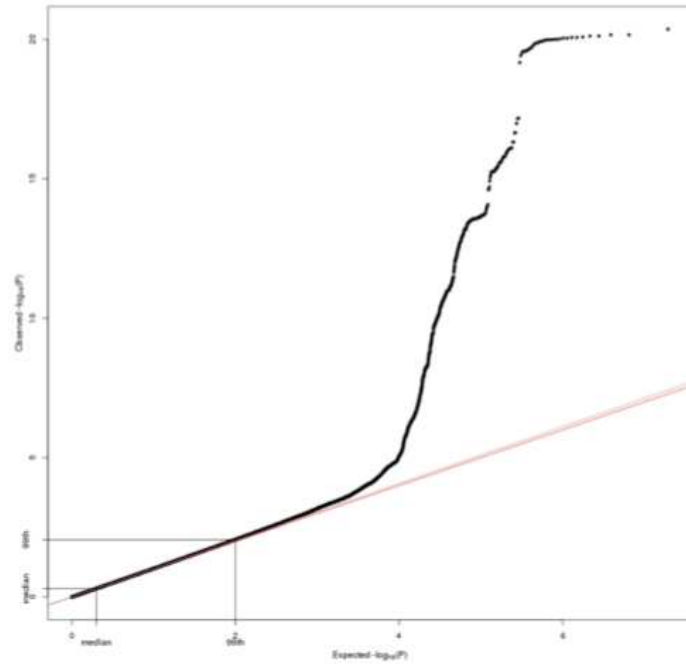
	Coefficient de consanguinité (Froh)		
	Cas	Témoins	p^a
Européens EPITHYR	0,0098	0,0098	0,75
Européens EPIC	0,011	0,011	0,69
EPITHYR EPIC poolés	0,0095	0,0099	0,12
EPITHYR + EPIC poolés groupe Nord	0,0078	0,0073	0,96
EPITHYR + EPIC poolés groupe Sud	0,0060	0,0063	0,29

Tableau supplémentaire 8 : Comparaison et caractéristiques des ROHs partagés chez les individus des études Polynésie Française selon le type de prélèvement d'ADN et selon différents seuils d'exclusion de SNPS avec nombre de cas et témoins similaires dans les deux groupes.

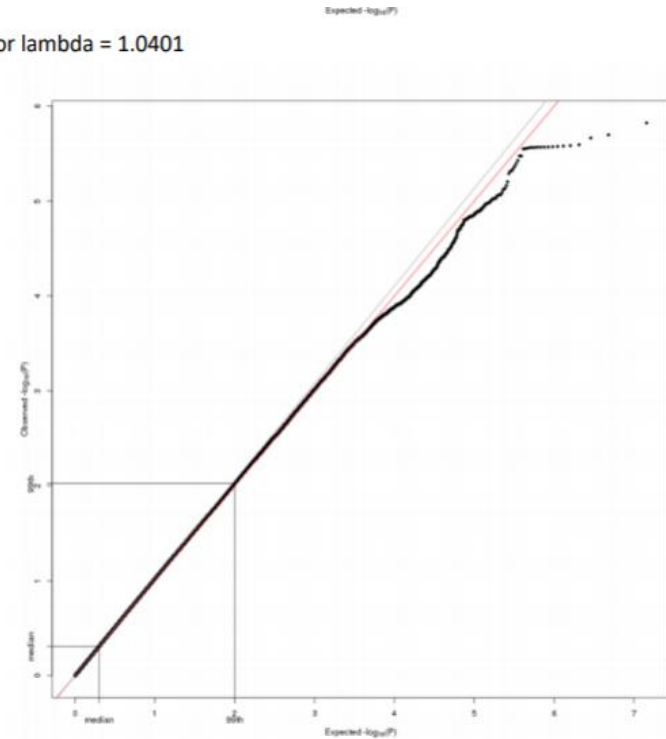
			POLYNESIENS – ADN salive (N=164, 61 cas/103 témoins)					POLYNESIENS – ADN brochette (N=164, 61 cas/103 témoins)				
ind			N SNPs	Nb ROH partagés	Longueur moyenne (kb)	Nombre moyen de SNPs	Nombre moyen d'individus	N SNPs	Nb ROH partagés	Longueur moyenne (kb)	Nombre moyen de SNPs	Nombre moyen d'individus
Pas de filtre			358 126	113	186,7	43,8	4,7	358 126	140	207,5	41,2	4,2
Filtrés selon la FAM	FAM 1%	all	352 423	102	211,3	49,6	4,8	352 412	130	188,9	38,2	3,9
Filtrés selon le DL	$r^2 > 0,8$	all	255 013	16	177,2	29,3	2,9	246 892	47	196,2	36,0	2,4
Filtrés selon FAM et DL	FAM 1% & $r^2 > 0,8$	all	250 951	12	155,2	28	2,4	242 819	40	190,1	34,1	2,1

Figure supplémentaire 1 : Diagrammes quantile-quantile des p -valeurs de GWAS chez (A) les Européens, (B) les Océaniens, (C) Analyse combinant les deux populations.

(A) Inflation factor lambda = 1.0211



(B) Inflation factor lambda = 1.0401



(C) Inflation factor $\lambda = 1.0776$

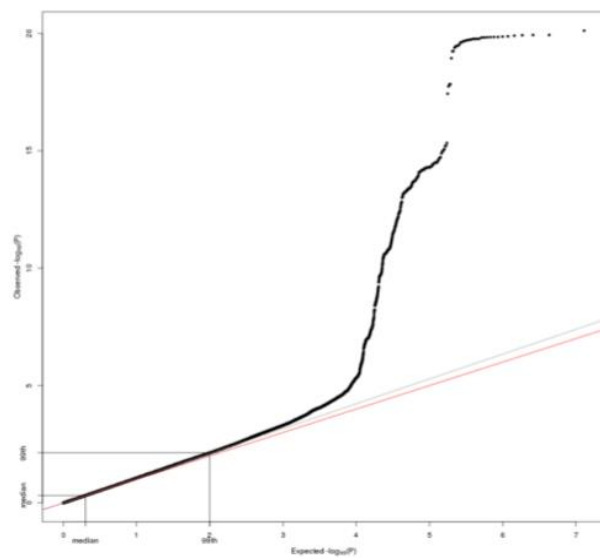
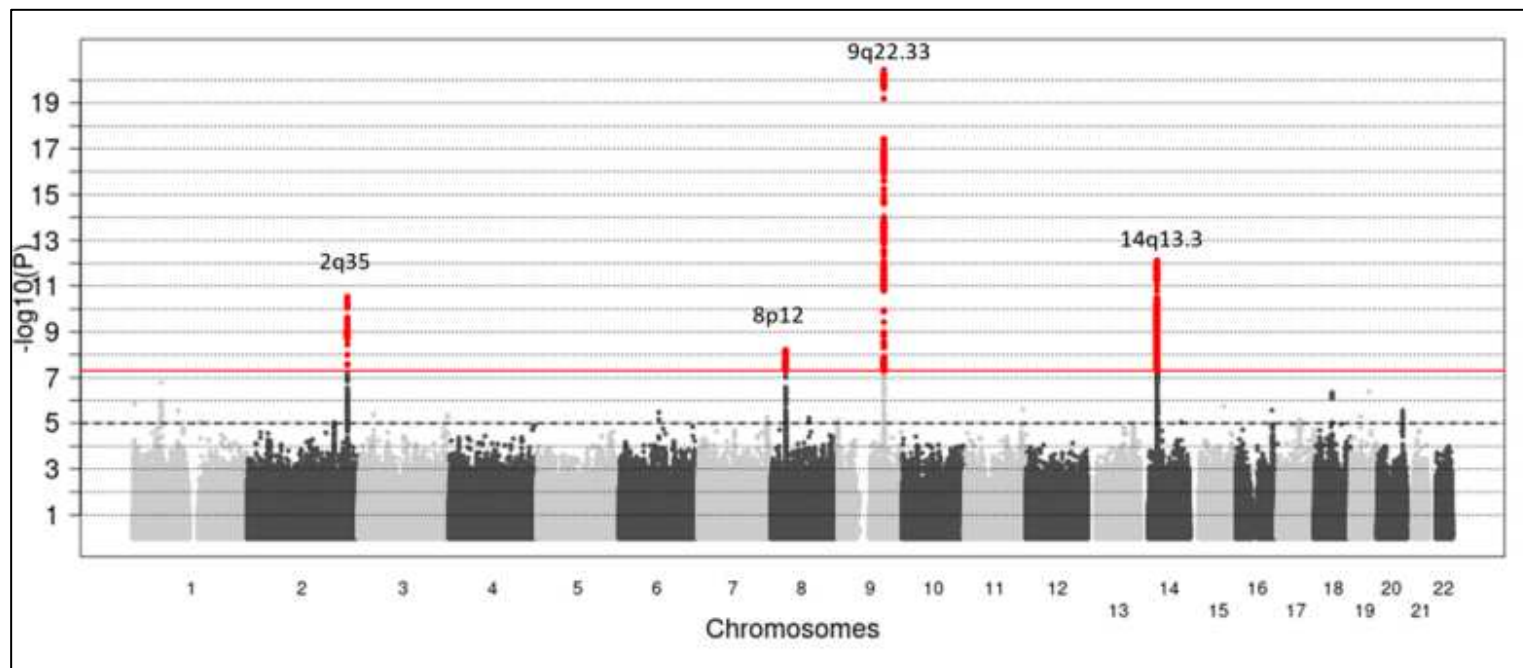
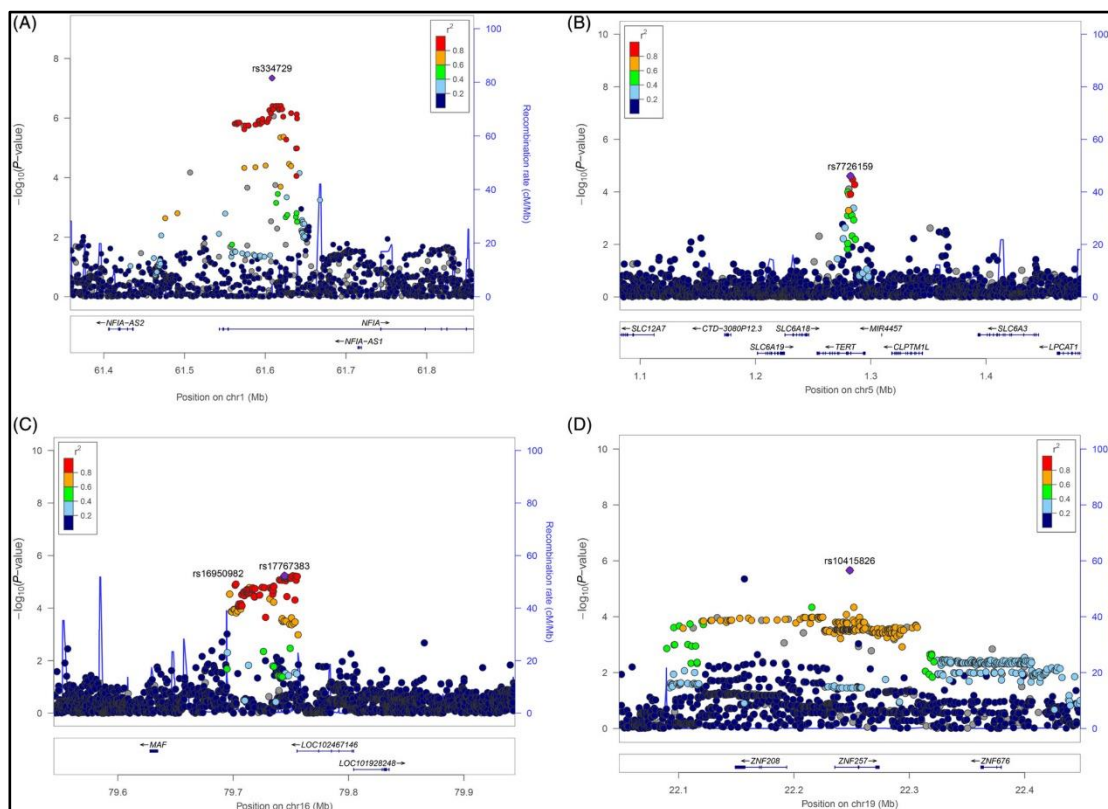


Figure supplémentaire 2 : Manhattan plot des résultats de l'analyse GWAS menée chez les Européens de CPT.



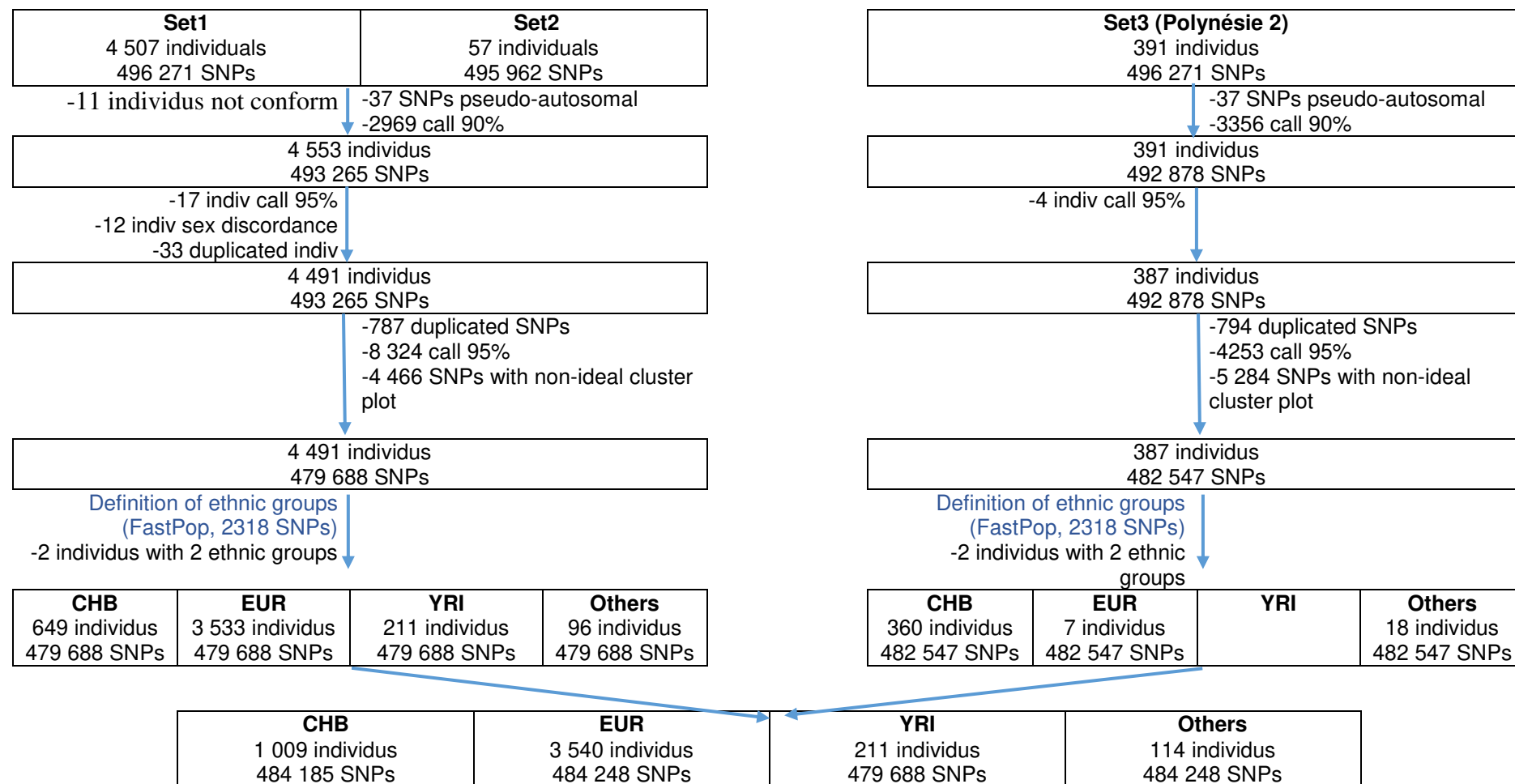
Résultats de 9 673 063 SNPs provenant de 1554 cas et 1973 témoins européens. L'axe des abscisses représente l'emplacement chromosomique et l'axe des ordonnées représente les $-\log_{10}(p\text{-valeurs})$ du modèle logistique mixte (test de score). Les lignes horizontales en pointillés indiquent le seuil de suggestivité ($p\text{-valeur} = 10^{-5}$), tandis que les lignes rouges indiquent un seuil de significativité ($p\text{-valeur} = 5 \times 10^{-8}$). Les loci déjà connus sont indiqués en noir et le nouveau locus identifié est indiqué en rouge.

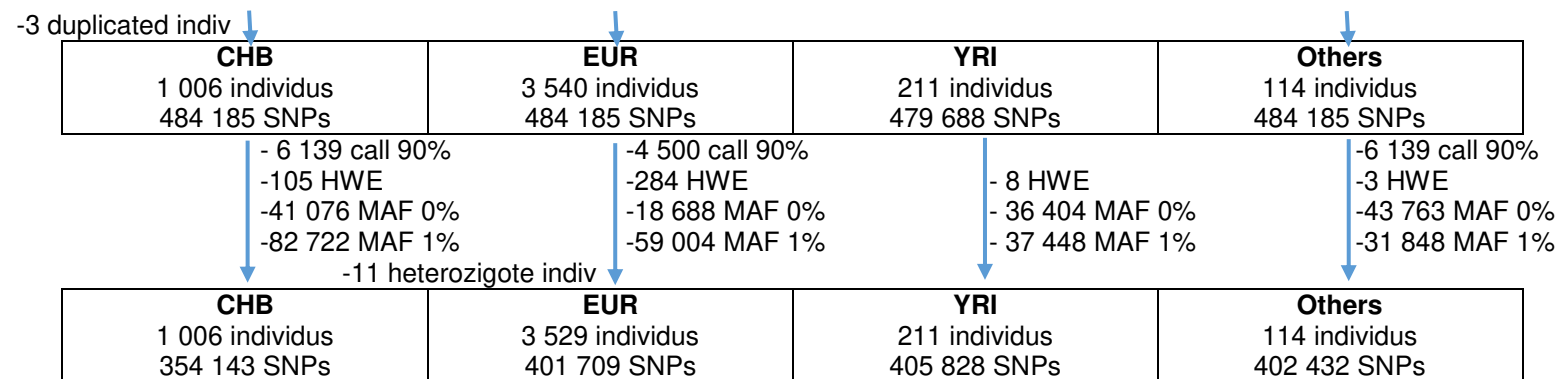
Figure supplémentaire 3 : Locuszoom des résultats d'association GWAS chez les Européens.



En (A) 1p13.3, (B) 5p15.33, (C) 16q23.2 et (D) 19p12. L'axe des abscisses représente la position des SNPs et du gène sur le chromosome, l'axe des ordonnées représente les $-\log_{10}$ p-valeurs et le taux de recombinaison. La couleur de chaque SNP reflète sa corrélation (r^2) avec le SNP le plus associé dans la région (en violet). La corrélation entre les SNPs varie de forte (en rouge) à faible (en bleu).

Figure supplémentaire 4 : Diagramme du contrôle qualité réalisé sur les données EPITHYR après ajout des individus de Polynésie Française 2.





Finding some individuals with problems in histological data (Elimination of these individuals), MAF0 :

CHB	EUR	YRI	Others
1 001 individus 436 865 SNPs	3 525 individus 460 713 SNPs	209 individus 443 276 SNPs	Without modifications

Figure supplémentaire 5 : Manhattan plot des p -valeurs d'associations obtenues avec un modèle logistique mixte chez les Européens d'EPITHYR (modèle récessif).

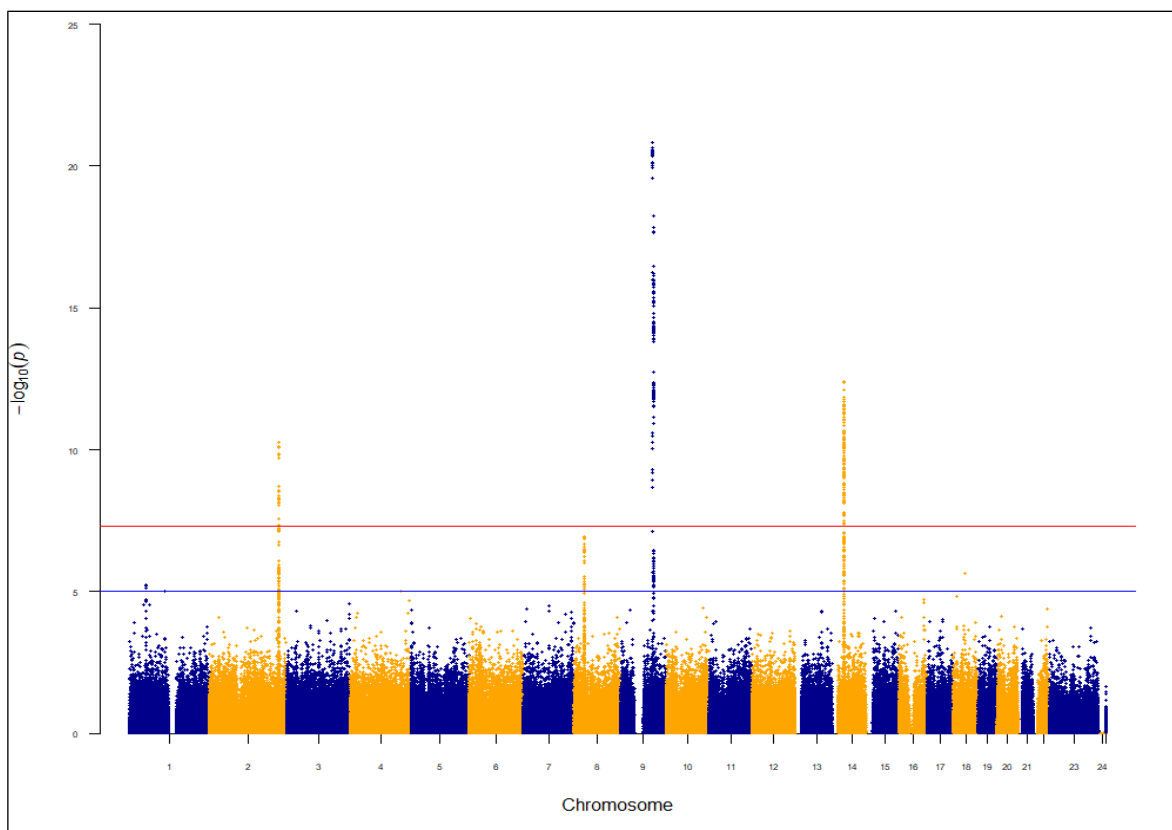
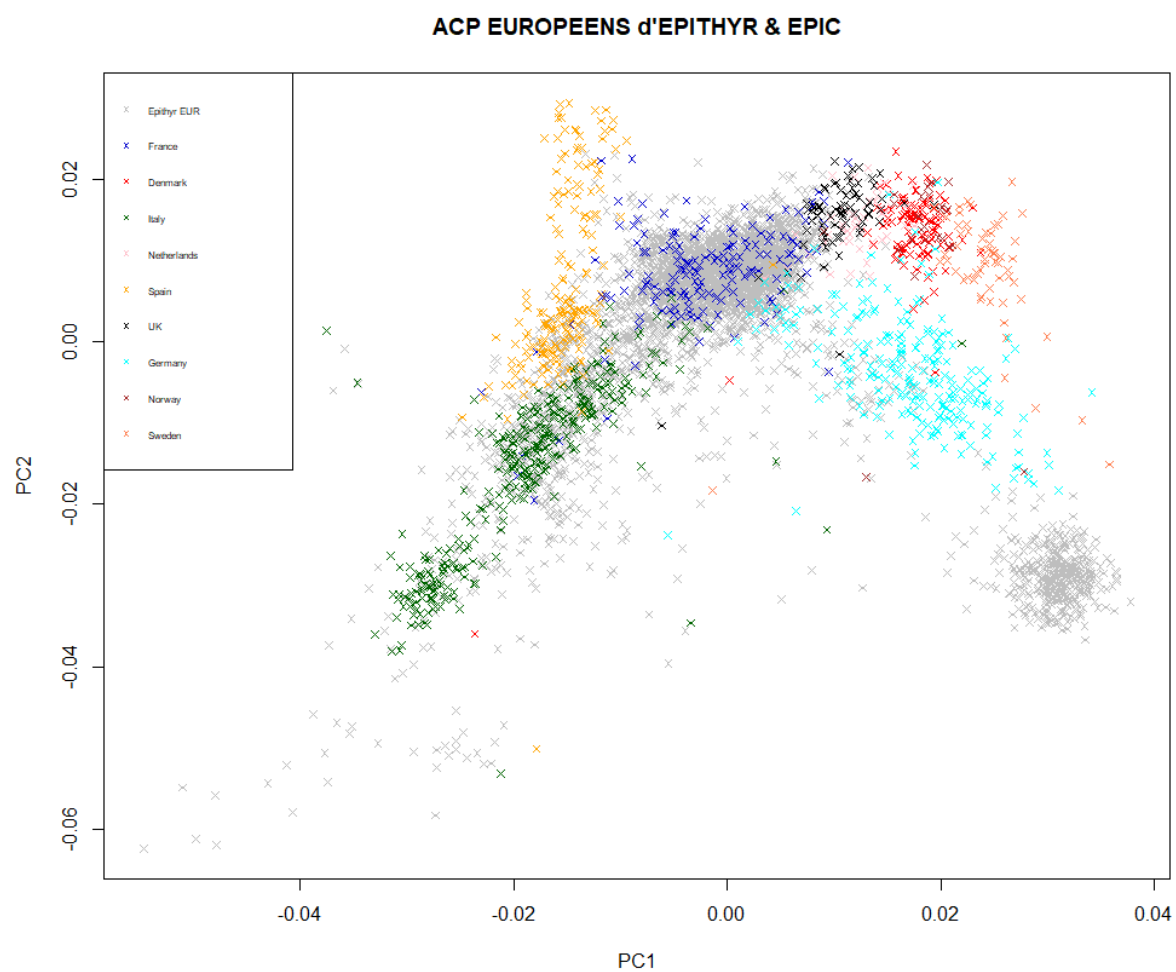


Figure supplémentaire 6 : Analyse en composantes principales des individus d'EPITHYR et d'EPIC



Multiethnic genome-wide association study of differentiated thyroid cancer in the EPITHYR consortium

Thérèse Truong¹  | Fabienne Lesueur²  | Pierre-Emmanuel Sugier¹ | Julie Guibon^{1,2} | Constance Xhaard^{3,4} | Mojgan Karimi¹ | Om Kulkarni² | Elise A. Lucotte¹ | Delphine Bacq-Daian⁵ | Anne Boland-Auge⁵ | Claire Mulot⁶ | Pierre Laurent-Puig⁶ | Claire Schvartz⁷ | Anne-Valérie Guizard^{8,9} | Yan Ren³ | Elisabeth Adjadj³ | Frédérique Rachédi¹⁰ | Francoise Borson-Chazot¹¹ | Rosa Maria Ortiz¹² | Juan J. Lence-Anta¹² | Celia María Pereda¹² | Daniel F. Comiskey Jr¹³ | Huiling He¹³ | Sandya Liyanarachchi¹³ | Albert de la Chapelle^{13†} | Rossella Elisei¹⁴ | Federica Gemignani¹⁵ | Hauke Thomsen^{16,17} | Asta Forsti^{16,18,19} | Anthony F. Herzig²⁰ | Anne-Louise Leutenegger²¹ | Carole Rubino³ | Evgenia Ostroumova²² | Ausrele Kesminiene²² | Marie-Christine Boutron-Ruault¹  | Jean-François Deleuze⁵ | Pascal Guénel¹ | Florent de Vathaire³

¹University Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, Team "Exposome and Heredity", Villejuif, France

²Inserm, U900, Institut Curie, PSL University, Mines ParisTech, Paris, France

³University Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, Team "Epidemiology of radiations", Villejuif, France

⁴University of Lorraine, INSERM CIC 1433, Nancy CHRU, Inserm U1116, FCRIN, INI-CRCT, Nancy, France

⁵Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, Evry, France

⁶Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université Paris Descartes, Université Paris Diderot, EPIGENETEC, Paris, France

⁷Registre des Cancers Thyroïdiens, Institut GODINOT, Reims, France

⁸Registre Général des tumeurs du Calvados, Centre François Baclesse, Caen, France

⁹Inserm U1086 -UCN "ANTICIPE", Caen, France

¹⁰Endocrinology Unit, Territorial Hospital Taaone, Papeete, Tahiti, French Polynesia

¹¹Fédération d'endocrinologie, Hôpital Louis Pradel, Hospices Civils de Lyon, EA 7425, Université Lyon 1, Lyon, France

¹²Institute of Oncology and Radiobiology, La Havana, Cuba

¹³Human Cancer Genetics Program and Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio

¹⁴Endocrine Unit, University Hospital of Pisa, Pisa, Italy

¹⁵Department of Biology, University of Pisa, Pisa, Italy

¹⁶Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

¹⁷GeneWerk GmbH, Heidelberg, Germany

¹⁸Hopp Children's Cancer Center (KITZ), Heidelberg, Germany

¹⁹Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany

Abbreviations: DTC, differentiated thyroid carcinoma; FTC, follicular thyroid carcinoma; GWAS, genome-wide association study; LD, linkage disequilibrium; MAF, minor allele frequency; OR, odds ratio; PTC, papillary thyroid carcinoma; SNP, single-nucleotide polymorphism; TSH, thyroid-stimulating hormone.

Thérèse Truong and Fabienne Lesueur contributed equally as co-first authors.

Pascal Guénel and Florent de Vathaire contributed equally as co-last authors.

† Deceased.

²⁰Inserm, U1078, GGB, Université de Bretagne Occidentale, EFS, Brest, France

²¹Université de Paris, NeuroDiderot, Inserm, U1141, Paris, France

²²International Agency for Research on Cancer, Lyon, France

Correspondence

Thérèse Truong, University Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, Team "Exposome and Heredity", Villejuif, France. Email: therese.truong@inserm.fr

Funding information

Agence Nationale de la Recherche, Grant/Award Number: ANR-10-COHO-0006; Fondation ARC pour la Recherche sur le Cancer, Grant/Award Number: PGA120150202302; Institut National Du Cancer, Grant/Award Number: 9533

Abstract

Incidence of differentiated thyroid carcinoma (DTC) varies considerably between ethnic groups, with particularly high incidence rates in Pacific Islanders. DTC is one of the cancers with the highest familial risk suggesting a major role of genetic risk factors, but only few susceptibility loci were identified so far. In order to assess the contribution of known DTC susceptibility loci and to identify new ones, we conducted a multiethnic genome-wide association study (GWAS) in individuals of European ancestry and of Oceanian ancestry from Pacific Islands. Our study included 1554 cases/1973 controls of European ancestry and 301 cases/348 controls of Oceanian ancestry from seven population-based case-control studies participating to the EPITHYR consortium. All participants were genotyped using the OncoArray-500K Beadchip (Illumina). We confirmed the association with the known DTC susceptibility loci at 2q35, 8p12, 9q22.33 and 14q13.3 in the European ancestry population and suggested two novel signals at 1p31.3 and 16q23.2, which were associated with thyroid-stimulating hormone levels in previous GWAS. We additionally replicated an association with 5p15.33 reported previously in Chinese and European populations. Except at 1p31.3, all associations were in the same direction in the population of Oceanian ancestry. We also observed that the frequencies of risk alleles at 2q35, 5p15.33 and 16q23.2 were significantly higher in Oceanians than in Europeans. However, additional GWAS and epidemiological studies in Oceanian populations are needed to fully understand the highest incidence observed in these populations.

KEYWORDS

case-control study, genome-wide association study, thyroid cancer

1 | INTRODUCTION

The incidence of thyroid cancer was estimated to be 567 233 new cases worldwide in 2018, representing 3.1% of all cancers.¹ Papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC) are the most frequent types of differentiated thyroid carcinomas (DTC) accounting for about 90% of all thyroid cancers. The incidence of DTC is characterized by 3 to 4 times higher rates in women than in men and wide geographic and ethnic variations.^{1,2} The highest incidence rates in 2018 were observed in North America, Pacific Islands and Eastern Asia.¹ In some multi-ethnic populations such as in New Caledonia or New Zealand, higher rates were reported in Pacific Islanders than in other ethnic groups.^{3,4} The causes underlying these geographic and ethnic variations are still unknown and a role for environmental and genetic factors is suspected. During the last decades, the incidence of DTC has increased in most high-resource countries.⁵

Part of the increase in incidence has been attributed to changes in medical screening practices that enhance the detection of micro-

What's new?

Differentiated thyroid carcinoma (DTC) has a strong genetic component, and Pacific Islanders have a particularly high incidence of the disease. Here, the authors conducted a genome-wide association study (GWAS) in a mixed population including individuals of European and Pacific Islander ancestry. The study drew from the EPITHYR consortium, a collection of 2000 DTC cases and age-matched controls. The analysis confirmed 5 previously known susceptibility loci and identified 2 new ones. Three of these loci were more prevalent among Pacific Islanders. More studies are needed to fully account for the rate of DTC in that population.

PTC (≤ 1 cm in diameter) that are of minimal clinical relevance. It has also been suggested that changes in environmental and lifestyle risk factors may contribute to this increase.⁶ Apart from exposure to

ionizing radiation during childhood, a well-established risk factor for DTC,⁷ increased DTC risk has also been consistently associated with obesity.^{8–11} Other risk factors such as deficient or excess iodine intake¹² or some reproductive and menstrual factors¹³ are suspected to play a role in the etiology of DTC. By contrast, several studies have reported an inverse association between DTC risk and cigarette smoking,¹⁴ alcohol intake¹⁵ or oral contraceptive pills.^{16,17}

DTC is also one of the cancers with the highest familial risk,¹⁸ suggesting a major role of genetic risk factors. However, only a few susceptibility loci have been identified thus far by genome-wide association studies (GWAS),^{19–26} with limited sample size. All previous GWAS, except one, were conducted in European descent populations and highlighted four susceptibility loci at 2q35, 9q22.33, 8p12 and 14q13.3 that were subsequently widely replicated in diverse populations.^{27–31} One of these GWAS was conducted in Belarus and focused on radiation-related PTC by including individuals aged under 18 years old at the time of the Chernobyl accident. Our study highlighted no other signal than the one at locus 9q22.33. One GWAS was conducted in a population of Asian ancestry in Korea and, in addition to the same four previously reported loci identified in populations of European ancestry, reported six other genomic regions at 1p13.3, 1q42.2, 3p14.2, 4q21.1, 12q14.3 and 19q13.2.²⁶ More recently, a meta-analysis of five GWAS that included 3001 cases and 287 550 controls of European ancestry²⁵ reported five new susceptibility loci at 1q42.2, 3q26.2, 5q22.1, 10q24.33 and 15q22.33. However, to our knowledge, findings from this meta-analysis have not been replicated in an independent sample.

In order to assess the contribution of known DTC susceptibility loci and to identify new ones in individuals of European ancestry and in individuals of Oceanian ancestry from Pacific Islands, we conducted a multiethnic GWAS in unexplored case-control studies from the EPITHYR consortium.

2 | METHODS

2.1 | Discovery samples

We conducted a GWAS using data from seven population-based case-control studies participating to the EPITHYR consortium. Of these studies, three were conducted in Metropolitan France (CATHY, YOUNG-Thyr and E3N studies), two in South Pacific Islands (French Polynesia and New Caledonia), one in Cuba and one in Gomel region of Belarus contaminated after the Chernobyl fall-out. The study designs are briefly outlined in Supplementary Methods and were described in detail previously.^{9,11,16,28,32,33} All studies in EPITHYR provided information on ethnicity, personal and familial history of thyroid disease, weight, height, dietary habits, alcohol intake, tobacco smoking, residential and occupational histories, and for women, menstrual and reproductive factors and exogenous hormone use. A total of 2572 cases and 3219 controls are included in EPITHYR, of which 2080 cases and 2473 controls provided a saliva or a blood sample (Supplementary Table 1).

2.2 | OncoArray SNPs selection

All EPITHYR subjects with a saliva or blood sample were genotyped at the Centre National de Recherche en Génomique Humaine (CNRGH/CEA, Evry, France) using the Infinium OncoArray-500K BeadChip (Illumina), on an Illumina automated high-throughput platform. This array, which was previously described in details³⁴ contains 499 170 SNPs with a genome-wide backbone of about 250 000 tag SNPs designed such that the large majority of common variants could be accurately imputed. Additional SNPs include dense coverage across known genetic variants associated with breast, colorectal, lung, ovarian and prostate cancers, SNPs covering ancestry, putative rare susceptibility variants from sequencing studies and other variants of potential biological relevance. As the OncoArray was not originally designed to include thyroid cancer susceptibility variants, we incorporated 13 759 additional custom markers based on prior evidence of association, candidate submissions, fine-mapping of known susceptibility regions and in genes involved in relevant biological pathways.

2.3 | Quality control of genotyping data

Details of the genotyping calling for OncoArray are described in more details elsewhere.³⁴ Cluster files generated using 56 284 samples from the GAME-ON consortium (available at http://consortia.ccge.medschl.cam.ac.uk/oncoarray/onco_v2c.zip) were applied to our dataset.

A total of 4553 individuals were genotyped. We excluded samples with a call rate < 95% ($n = 17$), with a sex discordant call ($n = 12$) and duplicates within and across studies ($n = 20$). Additionally, we excluded pairs of individuals that were identified as duplicates using the genetic data but with different epidemiological data ($n = 10$).

Ancestry was computed using the Fastpop program,³⁵ which derived principal component analysis (PCA) scores from 2318 informative markers and a subset of ~47 000 individuals from the GAME-ON consortium. These scores were applied to our dataset. Individuals of European ancestry and Asian ancestry were defined as individuals with respectively estimated proportion of European ancestry >0.8 and an estimated proportion of Asian ancestry >0.4 with reference to the HapMap populations, based on the first two principal components (Supplementary Figure 1). Individuals that were not considered to be of European or Asian ancestries were excluded from the analysis. Of the remaining 4494 individuals, 3538 were of European ancestry and 649 of Asian Ancestry. Additional individuals with extreme heterozygosity (4.89 SD from the mean of the ethnicity) ($n = 11$) were excluded and the final dataset comprised 3527 individuals of European ancestry (1554 cases and 1973 controls) and 649 individuals of Asian ancestry (301 cases and 348 controls).

Of the 512 929 SNPs that were successfully manufactured, we received 493 266 SNPs with a call rate > 90%. We excluded SNPs with a call rate < 95% by study ($n = 8327$), duplicate SNPs ($n = 814$), monomorphic SNPs ($n = 13 832$), SNPs for which the cluster plots were judged to be not ideal ($n = 4083$).

Of the remaining 466 210 SNPs, we additionally excluded SNPs not in HWE in each ethnic group ($P < 10^{-7}$ in controls, $P < 10^{-12}$ in

cases) ($n = 563$ in the study of European ancestry individuals and $n = 417$ in the study of Asian ancestry individuals), monomorphic SNPs ($n = 5210$ in European ancestry individuals and $n = 37\,513$ in Asian ancestry individuals) as well as non-monomorphic SNPs with $MAF < 1\%$ ($n = 58\,912$ in European ancestry individuals and $n = 75\,177$ Asian ancestry individuals). We retained for the analysis 401 525 SNPs among Europeans and 353 103 SNPs among Asians (Supplementary Figure 2).

2.4 | Imputation

We performed the imputation of missing SNPs in each population (European and Asian) independently using the October 2014 (version 3) release of the 1000 Genomes Project data set as the reference panel. We excluded from imputation (but not from association analysis) SNPs with a call $< 98\%$ in each ethnic group, SNPs that could not be linked to the 1000 Genomes project reference as well as ambiguous SNPs (AT/GC).

For the European population, we additionally excluded SNPs for which the difference statistic $(|p_1 - p_2| - 0.01)^2 / ((p_1 + p_2)(2 - p_1 - p_2))$ was > 0.008 , where p_1 and p_2 are, respectively, the frequencies in our data set and in the 1000 Genomes Project. Therefore, imputation in Europeans was based on 362 025 genotyped SNPs and a two-stage procedure was performed, using SHAPEIT2 (using the default parameters (“–burn 10 –prune 10 –main 50”) to derive phased genotypes and

IMPUTE2 to perform imputation using 5 Mb nonoverlapping intervals. The number of reference haplotypes to use as templates when imputing missing genotypes was increased to 800 (“–k_hap 800”), and the buffer region was increased to 500 kb (“–buffer 500”). We excluded imputed SNPs with an imputation $r^2 < 0.3$ and SNPs with an $MAF < 1\%$ in our population leaving a total of 9 673 063 SNPs.

For the Asian population, we imputed the nongenotyped data with IMPUTE2 but without pre-phasing in SHAPEIT, which is computationally more demanding but improves imputation accuracy.³⁶ Because our “Asian” population was closer to Oceanian populations (Supplementary Figure 3), we added to the 1000 Genomes project reference panel, genomes from 26 Oceanian individuals from the Simons Diversity Project that were available publicly (https://sharehost.hms.harvard.edu/genetics/reich_lab/sgdp/phased_data/). Imputation was based on 314 954 SNPs. We excluded SNPs with an imputation $r^2 < 0.3$ and SNPs with an $MAF < 1\%$ in our population leaving a total of 7 179 638 SNPs.

2.5 | Replication samples

In the replication stage, we used summary statistics from two independent studies conducted, respectively, in Ohio (USA)²⁵ and Italy,²² totaling 2270 cases and 2125 controls (Supplementary Methods) of European ancestry. Due to the lack of replication sample, we did not attempt to replicate our findings from the Oceanian population.

TABLE 1 Demographic characteristics of the study populations by ethnic group

	European (N = 3527)				Oceanian (N = 649)			
	Cases N = 1554	%	Controls N = 1973	%	Cases N = 301	%	Controls N = 348	%
Study								
Cathy	451	28.9	534	27.1	1	0.3	0	
Young-Thyr	637	41.0	661	33.5	3	1.0	0	
E3N	277	17.8	287	14.5	0		0	
Chernobyl	66	4.2	304	15.4	0		0	
Cuba	102	6.6	113	5.7	0		0	
New Caledonia	21	1.3	70	3.5	183	60.8	161	46.3
French Polynesia	0	0.0	4	0.2	114	37.9	181	52.0
Age (years)								
<25	272	17.5	513	26.0	0		0	
[25-50[	763	49.1	887	45.2	202	67.1	255	73.3
≥50	519	33.4	573	29.0	99	32.9	93	26.7
Sex								
Women	1279	82.3	1519	77.0	278	92.4	318	91.4
Men	275	17.7	454	23.0	23	7.6	30	8.6
Histology								
Follicular	137	8.8			46	16.4		
Papillary	1417	91.2			253	83.9		
<10 mm	506	35.7			136	53.6		
≥10 mm	904	63.8			107	42.1		

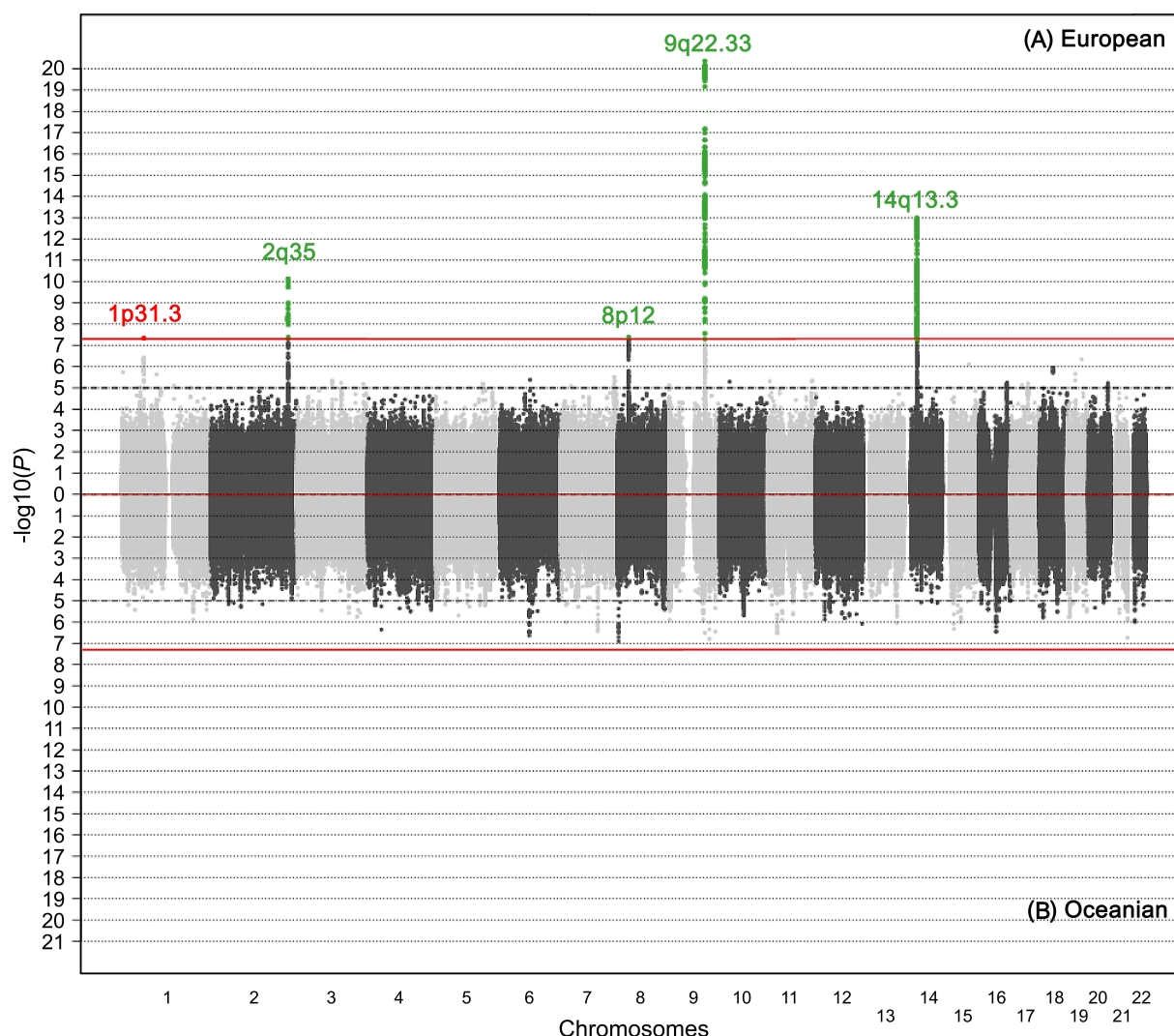


FIGURE 1 Miami plot of the DTC GWAS results stratified by ethnic group. A, European: 9 673 063 SNPs from 1554 cases and 1973 controls. B, Oceanian: 7 179 638 SNPs from 301 cases and 348 controls. The x-axis represents chromosomal location and the y-axis represents $-\log_{10}(P)$ value from a logistic mixed model (score test). Horizontal dashed lines indicate a genome-wide suggestiveness threshold at $P = 10^{-5}$ while red lines indicate genome-wide significance threshold at $P = 5 \times 10^{-8}$. Red points denote SNPs that were significantly associated with DTC. Previously published loci are indicated in green and the new locus identified in this analysis is indicated in red [Color figure can be viewed at wileyonlinelibrary.com]

2.6 | Statistical analysis

In the discovery stage, odds ratios (ORs) were calculated using logistic mixed models with the genetic relationship matrix as random effect to account for population structure and relatedness.³⁷ All analyses were adjusted for age, sex and study. We conducted separate analysis in Europeans and Oceanians, as well as a pooled analysis combining both populations. We also performed stratified analyses by sex, age group (<50, ≥50 years-old), histology (papillary/follicular carcinoma) and size of carcinoma for PTCs (<10 mm/≥10 mm). We used the GASTON R package (<https://cran.r-project.org/web/packages/gaston/index.html>) to perform the logistic mixed models. For genome-wide analyses, P values for each genetic variant were derived from a score test using a logistic mixed model as this test is less computationally demanding than a Wald test but asymptotically equivalent. A Wald test was then

performed for a subset of variants of interest in order to get the effect size estimate. The traditional Bonferroni threshold for GWAS at $P < 5 \times 10^{-8}$ was used to assess significance.

In the replication phase, we performed a fixed-effects meta-analysis with inverse-variance weighting as implemented in METAL software.³⁸ Heterogeneity of ORs across the studies and across the stratification groups was assessed using the Cochran's Q test.

3 | RESULTS

3.1 | Genome-wide analysis

A total of 1554 cases and 1973 controls of European ancestry and 301 cases and 348 controls of Oceanian ancestry from seven case-

TABLE 2 Association with the lead SNP at each significant locus by ethnic group

Locus	Lead SNP	Position (hg19)	OA	EA	EAF	OR	(95% CI)	P	EAF	OR	(95% CI)	P
1p31.3	rs334729 ^a	61 608 473	C	G	0.95	0.53	(0.41-0.66)	8.7×10^{-8} ^b	0.86	1.02	(0.72-1.44)	.92
2q35	rs3821098	218 292 141	T	C	0.72	0.70	(0.62-0.78)	1.0×10^{-10}	0.10	0.57	(0.41-0.81)	1.4×10^{-3}
8p12	rs142450470 ^a	32 365 296	CT	C	0.35	1.33	(1.19-1.47)	1.5×10^{-7} ^c	0.22	1.18	(1.11-1.26)	.23
9q22.33	rs1588635	100 537 802	A	C	0.61	0.61	(0.55-0.68)	2.1×10^{-21}	0.72	0.75	(0.58-0.97)	.03
14q13.3	rs368187	36 532 576	C	G	0.63	1.47	(1.32-1.63)	3.8×10^{-13}	0.31	1.51	(1.16-1.96)	2.3×10^{-3}
	rs116909374	36 738 361	T	C	0.94	0.43	(0.33-0.55)	1.6×10^{-10}	0.99	0.24	(0.05-1.17)	.08

Note: Lead SNP, SNP with the smallest *P* value in the European GWAS; OA, other allele; EA, effect allele; EAF, effect allele frequency in controls; OR, odds ratio derived from the log-additive logistic mixed model; 95% CI, 95% confidence interval. *P*, *P* value given by the Wald test.

^aImputed SNP, all imputed SNPs have imputation $r^2 > 0.90$ in both ethnic groups.

^bThe *P* value given by the score test was 4.5×10^{-8} .

^cThe *P* value given by the score test was 4.2×10^{-8} .

control studies were included in the analyses (Supplementary Table 1). The study sample included more than 80% women and most of the participants were younger than 50 years old in both ethnic groups. European cases mostly developed PTC, with a higher proportion of tumors larger than 10 mm, while papillary microcarcinomas were more frequent in Oceanian cases (Table 1). We performed a GWAS separately in each ethnic group and identified significant variants at loci 1p31.3, 2q35, 8p12, 9q22.33 and 14q13.3 in Europeans, and none in Oceanians (Figure 1). There was no indication of genomic inflation in both GWAS ($\lambda = 1.02$ in Europeans and $\lambda = 1.04$ in Oceanians, Supplementary Figure 4), suggesting no confounding by cryptic population structure. In the pooled analysis combining both populations (inflation factor $\lambda = 1.08$), no additional loci were highlighted and the association at locus 1p31.3 became nonsignificant (Supplementary Figure 5). We also conducted an analysis restricted to PTC cases in Europeans and observed a strengthened association at locus 8p12, but no additional signal was evidenced (Supplementary Figure 6). For each significant locus, we reported the association of the SNP with the smallest *P* value in the genomic region (lead SNP) according to the European GWAS in Table 2. Except at 1p31.3, the associations with these SNPs were in the same direction in Oceanians.

3.1.1 | Known loci

DTC susceptibility loci at 2q35, 8p12, 9q22.33 and 14q13.3 have been highlighted in previous GWAS conducted in Europeans. Lead SNPs at 2q35, 9q22.33 and 14q13.3 in the present study were the same or highly correlated with those reported in the published meta-analysis of GWAS on DTC²⁵ (Supplementary Table 2). At 8p12, the SNP rs142450470 in the present study (OR = 1.33, 5×10^{-7} for DTC, OR = 1.36, $P = 1.8 \times 10^{-8}$ for PTC in Europeans) was moderately correlated with rs2439302 highlighted by previous GWAS^{19,21} ($r^2 = 0.45$, OR = 1.19, $P = 4.8 \times 10^{-4}$ in our study) (Supplementary Table 2). The conditional analysis of rs142450470 on rs2439302 gave an OR = 1.32 with $P = 1.3 \times 10^{-4}$, whereas the conditional analysis of rs2439302 on rs142450470 provided an OR = 1.01, $P = .97$, suggesting a single signal at this locus.

Stratified analyses by histology showed that the association with rs1588635 at 9q22.33 was stronger for PTC, in particular of size >10 mm, while the association with rs142450470 at 8p12 was specific to PTC (Supplementary Figures 6).

3.1.2 | Replication of the novel locus at 1p31.3

The lead SNP at 1p31.3, rs334729 lies in intron 3 of the *NFIA* gene (Figure 2A). We conducted a replication analysis of rs334729 using data from two independent case-control studies conducted in Ohio and Italy and the association was nominally replicated ($P = .03$) and in the same direction as in EPITHYR, but the heterogeneity between the three studies was significant (P -heterogeneity = 8.6×10^{-3})

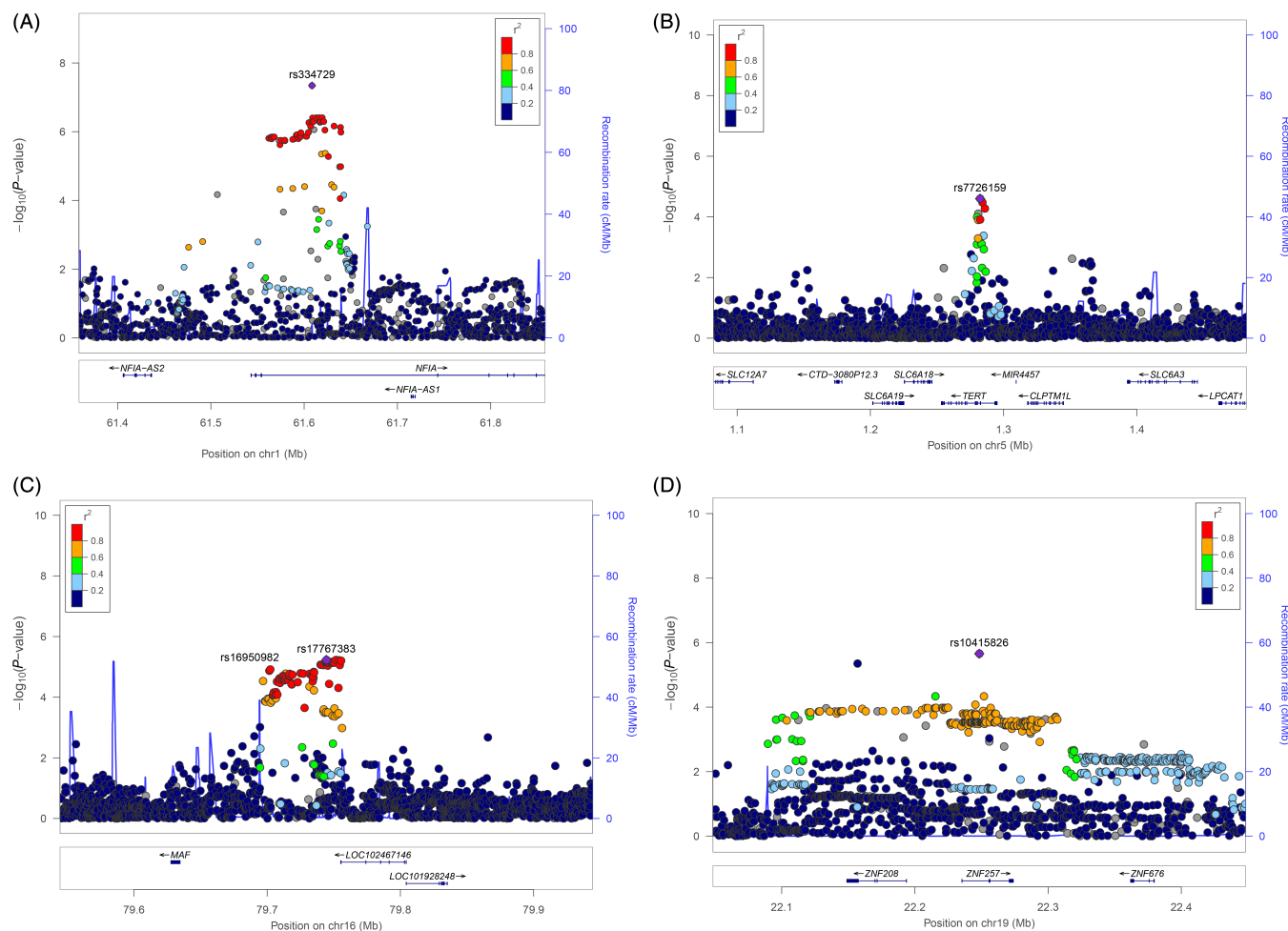


FIGURE 2 Regional plot of the DTC GWAS in Europeans. At (A) 1p13.3, (B) 5p15.33, (C) 16q23.2 and (D) 19p12. The x-axis represents the position of SNPs and gene on the chromosome, the y-axis represents the $-\log_{10}(P\text{-value})$ and the recombination rate. The color of each SNP spot reflects its correlation (r^2) with the most associated SNP in the region (in purple). The correlation between SNP is ranged from high (in red) to low (in blue) [Color figure can be viewed at wileyonlinelibrary.com]

(Supplementary Table 3). Rs334729[C] is highly correlated with rs334725[C] ($r^2 = 0.93$), which was previously associated with DTC risk in a meta-analysis of four studies ($OR = 1.30$, $P = 6.6 \times 10^{-3}$).²¹ We combined the results from these four studies with our findings in Europeans ($OR = 1.89$, $P = 1.4 \times 10^{-7}$ for rs334725[C]) in a meta-analysis and found an $OR = 1.52$, $P = 6.1 \times 10^{-8}$ ($P\text{-heterogeneity} = 0.25$) (Supplementary Table 4), which was at the limit of the significance level.

3.1.3 | Replication of suggestive loci

All chromosomal regions with at least two SNPs associated with DTC at suggestive level $P\text{-value} < 10^{-5}$ in the EPITHYR GWAS conducted in Europeans (Supplementary Table 3) or in the EPITHYR pooled GWAS (Supplementary Table 5) were considered for replication. In addition to the known loci at 2q35, 8p12, 9q22.33, 14q13.3 and the novel locus at 1p31.3, we identified 9 loci in Europeans and 11 loci in the pooled analysis corresponding to these criteria. The lead SNPs at each locus of interest were

analyzed in the replication sample, and only variants at 5p15.33, 16q23.2 and 19p12 were associated with DTC at a nominal $P\text{-value} < .05$ (Supplementary Tables 3 and 5). For these SNPs, we report the results of the combined analysis of the three studies in Europeans and the association in Oceanians from EPITHYR in Table 3.

Only the variants at 16q23.2 (rs16950982 and rs17767383) reached the genome-wide $P\text{-value}$ threshold of 5×10^{-8} in the meta-analysis. Association with rs16950982[G] was in the same direction in both ethnic groups, while we observed opposite effects for the effect allele [A] of rs17767383 (Table 3). These two variants were in strong linkage disequilibrium (LD) in Europeans ($r^2 = 0.88$), but only in weak LD in Oceanians ($r^2 = 0.37$). In the analysis stratified by sex, we observed a more striking association in men ($OR = 1.60$, $P = 4.7 \times 10^{-5}$) than in women ($OR = 1.18$, $P = 2 \times 10^{-3}$) for rs16950982 ($P\text{-heterogeneity} = 0.02$) (Supplementary Figure 7).

At 5p15.33, we nominally replicated results for rs7726159 ($P = .02$, Supplementary Table 5), but the meta-analysis combining the discovery and replication phases analysis did not reach the GWAS

TABLE 3 Replication analysis of suggestive loci

Locus	Lead SNP	OA	EA	Study	European EAF	OR	(95% CI)	P	Oceanian EAF	OR	(95% CI)	P
5p15.33	rs7726159	C	A	EPITHYR	0.39	1.26	(1.13-1.40)	2.0×10^{-5}	0.70	1.42	(1.11-1.82)	5.6×10^{-3}
				Italy	0.37	1.20	(1.00-1.43)	7.9×10^{-2}				
				Ohio	0.35	1.09	(0.98-1.20)	0.13				
				Meta-analysis		1.17	(1.09-1.25)	4.9×10^{-6}				
				<i>P</i> -het = .14								
16q23.2	rs17767383	G	A	EPITHYR	0.32	1.27	(1.14-1.41)	1.5×10^{-5}	0.34	0.93	(0.74-1.18)	.56
				Italy	0.28	1.37	(1.12-1.64)	1.3×10^{-3}				
				Ohio	0.31	1.16	(1.06-1.28)	3.1×10^{-3}				
				Meta-analysis		1.23	(1.15-1.32)	1.7×10^{-9}				
				<i>P</i> -het = .29								
16q23.2	rs16950982	A	G	EPITHYR	0.37	1.26	(1.13-1.40)	2.4×10^{-5}	0.58	1.15	(0.92-1.45)	.22
				Italy	0.30	1.31	(1.09-1.57)	5.0×10^{-3}				
				Ohio	0.32	1.17	(1.06-1.29)	2.7×10^{-3}				
				Meta-analysis		1.22	(1.14-1.31)	4.7×10^{-9}				
				<i>P</i> -het = .46								
19p12	rs10415826	A	T	EPITHYR	0.20	1.35	(1.19-1.53)	5.0×10^{-6}	0.20	1.33	(0.99-1.79)	.06
				Italy	0.20	1.11	(0.90-1.38)	0.34				
				Ohio	0.20	1.16	(1.03-1.30)	1.8×10^{-2}				
				Meta-analysis		1.22	(1.13-1.32)	8.7×10^{-7}				
				<i>P</i> -het = .15								

Note: Lead SNP, SNP with the smallest *P* value in the European GWAS or pooled analysis; OA, other allele; EA, effect allele; EAF, effect allele frequency in controls; OR, odds ratio derived from a log-additive logistic mixed model; 95% CI, 95% confidence interval. *P*, *P* value (Wald test). *P*-het, *P*-heterogeneity (Cochran Q test).

significance threshold ($P = 4.9 \times 10^{-6}$) (Table 3). We found that the associations in both ethnic groups were in the same direction but the allele frequency differed greatly, with a higher frequency of the risk allele [A] in the Oceanian ancestry population.

At 19p12, we replicated the association with rs10415826 with $P = 7.9 \times 10^{-3}$ (Supplementary Table 3) and the meta-analysis gave an OR = 1.22 ($P = 8.7 \times 10^{-7}$) in Europeans. Allele frequencies and direction of association were similar in Oceanians (Table 3).

3.2 | Replication study of hits identified in a previously published meta-analysis of GWAS on DTC

The meta-analysis of DTC GWAS conducted by Gudmundsson et al²⁵ identified five susceptibility loci at 1q42.2, 3q26.2, 5q22.1, 10q24.33 and 15q22.33, but the authors could not replicate these results in an independent sample. Here, we conducted a replication analysis of the lead SNPs at these loci and observed association with DTC in the same direction in Europeans for all SNPs; for SNPs at 1q42.2, 5q22.1 and 10q24 the P values were $<.001$. In Oceanians, associations with these SNPs were all in the opposite direction and none were significant (Supplementary Table 6).

4 | DISCUSSION

EPITHYR constitutes a new consortium to study DTC risk factors; it gathers over 2000 DTC cases and matched population controls of various geographical origin, with detailed epidemiological and genetic data. Although lifestyle risk factors have been extensively analyzed in individual studies of EPITHYR,^{9-11,16,17,39-42} this is the first analysis of genome-wide data in this consortium. Although the power of our study was limited for variants with low minor allele frequencies or low OR (Supplementary Table 7), we were able to confirm the association with the known DTC susceptibility loci at 2q35, 8p12, 9q22.33 and 14q13.3 in the European ancestry population. We additionally nominally replicated variants at 1p31.3 (rs334729, rs334725), 5p15.33 (rs7726159), 16q23.2 (rs16950982) and 19p12 (rs10415826).

The locus 1p31.3 was previously reported in a GWAS on thyroid-stimulating hormone (TSH) levels^{21,43} but not in GWAS on DTC. In our study, the most significant SNP at this locus, rs334729[C] is associated with increased risk of DTC and is in LD with rs334725[C], which was found to be associated with decreased TSH levels in a GWAS conducted in an Islandic population ($\beta = -0.13$, $P = 1.0 \times 10^{-20}$).²¹ In that study, rs334725[C] was additionally found to be associated with DTC risk in a secondary analysis increased risk (OR = 1.30, $P = 6.6 \times 10^{-3}$), in accordance with our findings. These variants are located in the *NF1A* gene, which encodes a member of the NF1 (nuclear factor 1) family of transcription factors, and NF1 isoforms were shown to interact with thyroid transcription factor 1 (TTF-1) to regulate cell function in experimental studies.⁴⁴

At 16q23.2, we reported an increased risk of DTC with rs16950982[G]. This variant is correlated with rs3813582[C]

($r^2 = 0.88$), which was previously associated with decreased TSH levels ($\beta = -0.082$, $P = 8.45 \times 10^{-18}$) with a male-specific effect⁴³ also observed in our study on DTC. Rs16950982[G] is in LD with rs17767419[T] ($r^2 = 0.88$), which was associated with increased thyroid volume ($\beta = 0.068$, $P = 9.4 \times 10^{-15}$).⁴⁵ These variants are located in an intergenic region close to the proto-oncogene *MAF* encoding a transcription factor (Figure 2C). According to GTEx Portal (<https://www.gtexportal.org/home/>), rs16950982[G] and rs17767383[A] are significantly associated with lower expression of *MAFTRR* in many tissues including thyroid tissue (respectively $\beta = -0.55$, $P = 2.8 \times 10^{-36}$ and $\beta = -0.60$, $P = 2.7 \times 10^{-43}$) and this gene is suspected to have a tumor-suppressor function.⁴⁶

Interestingly, at loci 1p31.3 and 16q23.2, the alleles of the SNPs conferring an increased risk of DTC (respectively, rs334729 and rs16950982) have also been associated with decreased TSH levels in a previous GWAS.^{21,43} These results are concordant with those of previous epidemiological studies that measured thyroid hormone levels in blood samples of DTC cases and their matched controls.⁴⁷

At 5p15.33, we highlighted the SNP rs7726159, which is located in intron 3 of *TERT* encoding the telomerase reverse transcriptase playing a key role in the maintenance of the telomeres (Figure 2B). In the previous meta-analysis of DTC GWAS conducted in an European population,²⁵ the best hit reported at 5p15.33 was for rs10069690[T] (OR = 1.20, $P = 3.2 \times 10^{-7}$), which is moderately correlated with our lead SNP rs7726159 ($r^2 = 0.46$). Interestingly, another study conducted in a Chinese population⁴⁸ and analyzing 15 haplotype-tagging SNPs in the *TERT-CLPTM1* region reported an increased risk of PTC with rs2736100[G], which is highly correlated with rs7726159[A] in Asians ($r^2 = 0.92$ in 1000 genomes CHB population), but only moderately correlated in European populations ($r^2 = 0.43$ in 1000 genomes CEU population). rs7726159 was shown to enhance *TERT* transcription⁴⁹ and the risk allele [A] of rs7726159 was previously associated with increased telomere length in European⁵⁰ and Asian⁴⁹ populations, as well as with risk of multiple cancers including breast and ovarian cancers.^{50,51} This is in line with a Danish study in 95 568 individuals reporting that genetic determinants of long telomeres were associated with increased risk of cancer. In our study, it was hypothesized that long telomeres could lead to the development of cancers by promoting the survival of precancerous cells.⁵¹

At 19p12, the lead SNP in our study, rs10415826, is located in intron 1 of *ZNF257* (Figure 2D). Little is known about the function of this gene encoding a transcription factor from the zinc finger family, and to our knowledge, no association with other traits has been reported so far with this variant.

Our study is the first GWAS conducted in an Oceanian population from French Polynesia and New Caledonia, two territories where the incidence of DTC is among the highest. Except at 1p31.3, all associations we highlighted in Europeans were in the same direction in the Oceanian ancestry population. However, our study was limited by the sample size and by the lack of replication sample in this population. We used the Oceanian population to strengthen the association with variants that may be associated with DTC in both populations. Interestingly, we observed that the frequencies of the risk alleles at 2q35,

5p15.33 and 16q23.2 were significantly higher in Oceanians than in Europeans (Supplementary Table 8). Additional GWAS and epidemiological studies in Oceanian populations are needed to fully understand the highest incidence observed in these populations.

ACKNOWLEDGEMENTS

We thank Dr Yannick Rougier and Dr Dominique Baron-Dubourdieu for providing pathological reports for the NC study, as well as Dr Sylvie Laumond, Dr Jean-Paul Grangeon (Direction des affaires sanitaires et sociales de Nouvelle-Calédonie) and the country's provincial health authorities (DPASS Sud, DPASS Nord, DPASS Iles Loyauté) for support during data collection in the NC study. We thank Milagros Velasco, Mae Chappe and Idalmis Infante (Institute of Oncology and Radiobiology, La Havana, Cuba) and Silvia Turcios (National Institute of Endocrinology, La Havana, Cuba) for helping in the collection of the data in the Cuban Study. We thank John Paoaafaite and Joseph Teuri who contacted and interviewed cases and controls for the study. Finally, we also thank P. Morales, J. Iltis, P. Giraud, P. Didiergeorge, M. Brisard, G. Soubiran, B. Caillou, P. Dupire, J. Ienfa, G. de Clermont, N. Cerf, B. Oddo, M. Bambridge, C. Baron, A. Mouchard-Rachet, O. Simonet, D. Lamarque, J. Vabret, J. Delacre, M.P. Darquier and J. Leninger, for their help in the collection of the cases or in the organization of study in French Polynesia.

We would like to thank the Association Centre de Regroupement Informatique et Statistique en Anatomie Pathologique en Provence-Alpes-Côte d'Azur (CRISAP PACA), as well as Dr Arlette Danzon, Dr Geneviève Sasolas, Dr Marc Christophe Sattonnet, Dr Marc Colonna, Dr Brigitte Lacour, Dr Michel Velten, Dr Enora Clero, Dr Stéphane Maillard, Dr Laurent Bailly, Dr Eugénia Mariné Barjoan, Dr Jean-Luc Lassalle, Dr Z Hafdi-Nejjari, Dr P Delafosse, Dr Elisabeth Adjadj, Kami-Marie Moreau, Cyrielle Orenes, Laurianne Sarrazin, Stéphanie Bonnay, Frédérique Chatelain, Maryse Barouh, Evelyne Rapp, Julie Festraëts, Julie Valbousquet, Yusuf Atilgan, Jean Chappellet, Lallia Bedhouche, Florent Dayet and Ziyani Fami, for their help in the collection of cases, the organization and the management of the Young-Thyr study. We acknowledge Stefano Landi for the Italian GWAS data and Subhayan Chattopadhyay and Yasmeen Niazi (Division of Molecular Genetic Epidemiology, German Cancer Research Center—DKFZ) for technical assistance in these data analysis. We are grateful to Dr Hervé Perdry for his help in using the GASTON package.

The EPITHYR GWAS was supported by Institut National du Cancer (grant number 9533) and Fondation ARC (grant number PGA120150202302). The E3N cohort received support from the MGEN, Gustave Roussy and Ligue contre le cancer for its setup and maintenance. The E3N cohort was also supported by a state grant from the Agence Nationale pour la Recherche (ANR) (grant number ANR-10-COHO-0006) within the Investissement d'Avenir program and from the French Ministry of Higher Education, Research and Innovation (MESRI, grant number 2102 918823). The other participating studies were funded by Ligue Nationale Contre le Cancer; ANR; the Direction Générale de la Santé; the Agence Française de Sécurité Sanitaire de l'alimentation; de l'environnement et du travail (ANSES); CHILDTHYR EEC program; and the Fondation de France. JG was the recipient of a PhD fellowship from Région Ile-de-France, part of OK

was the recipient of a post-doc fellowship from Electricité de France (conseil scientifique de Radioprotection d'EDF, grant EP 2019-01).

CONFLICT OF INTEREST

The other authors declare no conflict of interest.

DISCLAIMER

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ETHICS STATEMENT

Participants from all studies provided written informed consent and each study was approved by their governing ethics committee.

ORCID

Thérèse Truong  <https://orcid.org/0000-0002-2943-6786>

Fabienne Lesueur  <https://orcid.org/0000-0001-7404-4549>

Marie-Christine Boutron-Ruault  <https://orcid.org/0000-0002-5956-5693>

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre L, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin France*. 2018;68(6):394-424.
2. Magreni A, Bann DV, Schubart JR, Goldenberg D. The effects of race and ethnicity on thyroid cancer incidence. *JAMA Otolaryngol Head Neck Surg*. 2015;141(4):319-323.
3. Truong T, Rougier Y, Dubourdieu D, et al. Time trends and geographic variations for thyroid cancer in New Caledonia, a very high incidence area (1985–1999). *Eur J Cancer Prev*. 2007;16(1):62-70.
4. Meredith I, Sarfati D, Atkinson J, Blakely T. Thyroid cancer in Pacific women in New Zealand. *N Z Med J*. 2014;127(1395):52-62.
5. Wiltshire JJ, Drake TM, Uttley L, Balasubramanian SP. Systematic Review of Trends in the Incidence Rates of Thyroid Cancer. *Thyroid*. 2016;26(11):1541-1552.
6. Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. *Nat Rev Endocrinol*. 2016;12(11):646-653.
7. Iglesias ML, Schmidt A, Al Ghuzlan A, et al. Radiation exposure and thyroid cancer: a review. *Arch Endocrinol Metab*. 2017;61(2):180-187.
8. Wang H. Obesity and risk of thyroid cancer: evidence from a meta-analysis of 21 observational studies. *Med Sci Monit*. 2015;21:283-291.
9. 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*. 2007;166(10):1140-1149.
10. Brindel P, Doyon F, Rachédi F, et al. Anthropometric factors in differentiated thyroid cancer in French Polynesia: a case-control study. *Cancer Causes Control*. 2009;20(5):581-590.

11. Clavel-Chapelon F, Guillas G, Tondeur L, Kernaeguen C, Boutron-Ruault MC. Risk of differentiated thyroid cancer in relation to adult weight, height and body shape over life: the French E3N cohort. *Int J Cancer*. 2010;126(12):2984-2990.
12. Lee JH, Hwang Y, Song RY, et al. Relationship between iodine levels and papillary thyroid carcinoma: A systematic review and meta-analysis. *Head Neck*. 2017;39(8):1711-1718.
13. Caini S, Gibelli B, Palli D, Saieva C, Ruscica M, Gandini S. 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*. 2015;26(4):511-518.
14. Cho YA, Kim J. Thyroid cancer risk and smoking status: a meta-analysis. *Cancer Causes Control*. 2014;25(9):1187-1195.
15. Hong S-H, Myung S-K, Kim H. Alcohol intake and risk of thyroid cancer: a meta-analysis of observational studies. *Cancer Res Treat*. 2016;49(2):534-547.
16. Cordina-Duverger E, Truong T, Guénel P. Hormonal and reproductive risk factors of papillary thyroid cancer: a population-based case-control study in France. *Cancer Epidemiol*. 2017;48:78-84.
17. Xhaard C, Rubino C, Cléro E, et al. Menstrual and reproductive factors in the risk of differentiated thyroid carcinoma in young women in France: a population-based case-control study. *Am J Epidemiol*. 2014;180(10):1007-1017.
18. Hemminki K, Li X. Familial risk of cancer by site and histopathology. *Int J Cancer*. 2003;103(2002):105-109.
19. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet*. 2009;41(4):460-464.
20. Takahashi M, Saenko V, Rogounovitch TI, et al. The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Hum Mol Genet*. 2010;19(12):2516-2523.
21. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat Genet*. 2012;44(3):319-322.
22. Köhler A, Chen B, Gemignani F, et al. Genome-wide association study on differentiated thyroid cancer. *J Clin Endocrinol Metab*. 2013;98(10):E1674-E1681.
23. Figlioli G, Chen B, Elisei R, et al. Novel genetic variants in differentiated thyroid cancer and assessment of the cumulative risk. *Sci Rep*. 2015;5:8922.
24. Mancikova V, Cruz R, Inglada-Pérez L, et al. Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations. *Int J Cancer*. 2015;137(8):1870-1878.
25. Gudmundsson J, Thorleifsson G, Sigurdsson JK, et al. A genome-wide association study yields five novel thyroid cancer risk loci. *Nat Commun*. 2017;8:14517.
26. Son H-Y, Hwangbo Y, Yoo S-K, et al. Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer. *Nat Commun*. 2017;8:15966.
27. Tcheandjieu C, Lesueur F, Sanchez M, et al. 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(3):617-627.
28. Lence-Anta JJ, Xhaard C, Ortiz RM, et al. Environmental, lifestyle, and anthropometric risk factors for differentiated thyroid cancer in Cuba: a case-control study. *Eur Thyroid J*. 2014;3:189-196.
29. Wang Y-L, Feng S-H, Guo S-C, et al. Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population. *J Med Genet*. 2013;50(10):689-695.
30. Maillard S, Damiola F, Cléro E, 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*. 2015;10(4):1-14.
31. Pereda C, Lesueur F, Pertesi M, 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*. 2015;16:22.
32. Xhaard C, De Vathaire F, Cléro E, et al. Anthropometric risk factors for differentiated thyroid cancer in young men and women from Eastern France: A case-control study. *Am J Epidemiol*. 2015;182(3):202-214.
33. Damiola F, Byrnes G, Moissonnier M, et al. Contribution of ATM and FOXE1 (TTF2) to risk of papillary thyroid carcinoma in Belarusian children exposed to radiation. *Int J Cancer*. 2014;134(7):1659-1668.
34. Amos CI, Dennis J, Wang Z, et al. The oncoarray consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomarkers Prev*. 2017;26(1):126-135.
35. Li Y, Byun J, Cai G, et al. FastPop: a rapid principal component derived method to infer intercontinental ancestry using genetic data. *BMC Bioinform*. 2016;17:122.
36. Roshara NR, Horn K, Kirsten H, et al. Comparing performance of modern genotype imputation methods in different ethnicities. *Sci Rep*. 2016;6:34386.
37. Chen H, Wang C, Conomos MP, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am J Hum Genet*. 2016;98(4):653-666.
38. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.
39. 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*. 2005;161(11):1056-1065.
40. 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*. 2010;21(8):1183-1192.
41. Cléro É, Leux C, Brindel P, et al. 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*. 2010;20(11):1285-1293.
42. Cardis E, Kesminiene A, Ivanov V, et al. Risk of thyroid cancer after exposure to 131I in childhood. *J Natl Cancer Inst*. 2005;97(10):724-732.
43. Porcu E, Medici M, Pistis G, et al. A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PLoS Genet*. 2013;9(2):e1003266.
44. Bachurski CJ, Yang GH, Currier TA, Gronostajski RM, Hong D. Nuclear factor I/thyroid transcription factor 1 interactions modulate surfactant protein C transcription. *Mol Cell Biol*. 2003;23(24):9014-9024.
45. Teumer A, Rawal R, Homuth G, et al. Genome-wide association study identifies four genetic loci associated with thyroid volume and goiter risk. *Am J Hum Genet*. 2011;88(5):664-673.
46. Cheng J, Demeulemeester J, Wedge DC, et al. Pan-cancer analysis of homozygous deletions in primary tumours uncovers rare tumour suppressors. *Nat Commun*. 2017;8(1):1-14.
47. Rinaldi S, Plummer M, Biessy C, et al. Thyroid-stimulating hormone, thyroglobulin, and thyroid hormones and risk of differentiated thyroid carcinoma: The EPIC study. *J Natl Cancer Inst*. 2014;106(6):dju097.
48. Ge M, Shi M, An C, et al. Functional evaluation of TERT-CLPTM1L genetic variants associated with susceptibility of papillary thyroid carcinoma. *Sci Rep*. 2016;6(May):1-7.

49. Li Y, Xiang C, Shen N, et al. Functional polymorphisms on chromosome 5p15.33 disturb telomere biology and confer the risk of non-small cell lung cancer in Chinese population. *Mol Carcinog.* 2019;58(6):913-921.
50. Pooley KA, Bojesen SE, Weischer M, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: Identified loci show little association with hormone-related cancer risk. *Hum Mol Genet.* 2013;22(24):5056-5064.
51. Rode L, Nordestgaard BG, Bojesen SE. Long telomeres and cancer risk among 95 568 individuals from the general population. *Int J Epidemiol.* 2016;45(5):1634-1643.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Truong T, Lesueur F, Sugier P-E, et al. Multiethnic genome-wide association study of differentiated thyroid cancer in the EPITHYR consortium. *Int. J. Cancer.* 2021;148:2935–2946. <https://doi.org/10.1002/ijc.33488>

Research Paper

Fine-mapping of two differentiated thyroid carcinoma susceptibility loci at 2q35 and 8p12 in Europeans, Melanesians and Polynesians

Julie Guibon^{1,2}, Pierre-Emmanuel Sugier¹, Om Kulkarni³, Mojgan Karimi¹, Delphine Bacq-Daian³, Céline Besse³, Anne Boland³, Elisabeth Adjadj⁴, Frédérique Rachédi⁵, Carole Rubino⁴, Constance Xhaard^{4,6}, Claire Mulo⁷, Pierre Laurent-Puig⁷, Anne-Valérie Guizard^{8,9}, Claire Schvartz¹⁰, Rosa Maria Ortiz¹¹, Yan Ren⁴, Evgenia Ostroumova¹², Jean-François Deleuze³, Marie-Christine Boutron-Ruault¹, Ausrele Kesminiene¹², Florent De Vathaire⁴, Pascal Guénel¹, Fabienne Lesueur^{2,*} and Thérèse Truong^{1,*}

¹University Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, Exposome and Heredity Team, Villejuif, France

²Inserm, U900, Institut Curie, PSL University, Mines ParisTech, Paris, France

³University Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, Evry, France

⁴University Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, Epidemiology of Radiations Team, Villejuif, France

⁵Endocrinology Unit, Territorial Hospital Taane, Papeete, French Polynesia

⁶University of Lorraine, INSERM CIC 1433, Nancy CHRU, Inserm U1116, FCRIN, INI-CRCT, Nancy, France

⁷Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université Paris Descartes, Université Paris Diderot, EPIGENETEC, Paris, France

⁸Registre Général des Tumeurs du Calvados, Centre François Baclesse, Caen, France

⁹Inserm U1086 -UCN "ANTICIPE", Caen, France

¹⁰Registre des Cancers Thyroïdiens, Institut Godinot, Reims, France

¹¹Institute of Oncology and Radiobiology, La Havana, Cuba

¹²International Agency for Research on Cancer, Lyon, France

*These authors contributed equally to this work

Correspondence to: Thérèse Truong, email: Therese.truong@inserm.fr

Keywords: thyroid cancer; cancer genetics; case-control study; fine-mapping study; single nucleotide polymorphism

Received: October 14, 2020

Accepted: January 26, 2021

Published: March 02, 2021

Copyright: © 2021 Guibon et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Differentiated thyroid carcinoma (DTC) incidence is characterized by wide ethnic and geographic variations, with high incidence rates observed in Oceanian populations. Genome-wide association studies (GWAS) identified mainly four DTC susceptibility loci at 9q22.33, 14q13.3, 2q35 and 8p12. Here we performed fine-mapping of the 2q35 and 8p12 loci in the population of the EPITHYR consortium that includes Europeans, Melanesians and Polynesians to identify likely causal variants for DTC risk. We conducted a colocalization analysis using eQTLs data to determine the SNPs with the highest probability of causality.

At 2q35, we highlighted rs16857609 located in *DIRC3*. This SNP has a high probability of causality in the three populations, and a significant association in Europeans (OR = 1.4, $p = 1.9 \times 10^{-10}$). It is also associated with expression of *DIRC3* and of the nearby gene *IGFBP5* in thyroid tumour cells. At 8p12, we identified rs7844425 which was significantly associated with DTC in Europeans (OR = 1.32, $p = 7.6 \times 10^{-8}$) and rs2439304, which was highlighted by the colocalization analysis but only moderately associated with DTC in our dataset (OR = 1.2, $p = 0.001$). These SNPs are linked to the expression of *NRG1* in thyroid tissue.

Hence, our study identified novel variants at 2q35 and 8p12 to be prioritized for further functional studies.

INTRODUCTION

Thyroid cancer (TC) is the most common endocrine malignancy. Its incidence has been rising over the past few decades and varies considerably around the world. Elevated incidence was reported in some Pacific islands such as French Polynesia and New-Caledonia, two French territories, with estimated incidence rates in 2018 being 27.7/100,000 and 25.6/100,000 person-years (PY) in women and 6.5/100 000 and 7.6/100 000 PY in men [1], respectively. Ethnic differences in incidence have also been reported in New-Caledonia with higher rates among Melanesians than in other ethnic groups [2]. In metropolitan France, the incidence rates were 20.9/100,000 in women and 5.7/100,000 PY in men in 2018 [1].

Papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC) are the most frequent subtypes of differentiated thyroid carcinoma (DTC), and represent about 90% of all TC. Apart from exposure to ionizing radiation during childhood which is the only well-established risk factor, DTC risk has been consistently associated with obesity [3]. Other risk factors such as deficiency or excess of iodine levels [4] or some hormonal and reproductive factors are also suspected [5]. Moreover, familial risk for TC is among the highest of all cancers [6], suggesting a role of genetic risk factors. Six previous genome-wide association studies (GWAS) of DTC conducted in European populations [7–12] and a GWAS conducted in the Asian population [13] allowed the identification of several DTC susceptibility loci, with the strongest associations being reported at 9q22.33, 14q13.3, 2q35 and 8p12. Recently, a meta-analysis including 3,001 DTC cases and 287,500 controls of European descent confirmed these findings and yielded five additional loci associated to DTC at 1q42.2, 3q26.2, 5q22.1, 10q24.33, 15q22.33 that were not replicated yet [12]. We recently conducted a GWAS in EPITHYR that confirmed the susceptibility loci at 9q22.33, 14q13.3, 2q35 and 8p12, and suggested novel signals at 1p31.3 and 16q23.2 [14].

Fine-mapping studies have already been conducted to identify potential causal variants at loci 9q22.33 and 14q13.3 [15, 16]. These variants were subsequently widely replicated, and functional studies were performed to support their role in the aetiology of DTC. To our knowledge, no fine-mapping analyses have been performed so far at 2q35, 8p12 or other DTC susceptibility loci.

Rs966423 at 2q35 and rs2439302 at 8p12 were first identified in GWAS to investigate genetic factors associated with thyroid stimulating hormone (TSH) levels. In the study conducted in Iceland on 27,758 individuals

with available circulating TSH measurements [9], the identified SNPs were also found to be associated with DTC risk [9]. Subsequently, these variants were replicated in other populations of European [9, 12] and Asian [13] ancestry.

At locus 2q35, several other variants were also highlighted by GWAS to be associated with DTC in populations of European ancestry. Rs6759952 was reported in GWAS conducted on 3648 DTC cases and 4,224 controls from Italy [10], while a recent meta-analysis of GWAS reported the strongest association with variant rs11693806 at this locus [12]. In European populations, both rs11693806 and rs6759952 were moderately correlated with rs966423 (with $r^2 = 0.47$ and $r^2 = 0.74$, respectively) [10, 12]. In Korea, a GWAS identified two additional SNPs at 2q35, namely rs12990503 and rs1549738, associated with DTC risk [13]. According to 1000Genomes, rs12990503 was highly correlated ($r^2 = 0.97$) with rs11693806 in East Asian populations while rs1549738 was moderately correlated with the GWAS SNP rs966423 ($r^2 = 0.41$). Most of the risk variants at this locus are located within the *DIRC3* (*disrupted renal cancer 3*) gene. Although this gene is presumed to have a tumour suppressor activity [10], no functional studies have been performed so far to validate this hypothesis. In 2012, Gudmundsson *et al.* hypothesized that *DIRC3* variants by reducing TSH levels, may lead to reduced differentiation of the thyroid epithelium, and consequent predisposition to malignant transformation [9].

At locus 8p12, only rs2439302 and rs2466076 have been associated with DTC risk in European populations, and the two SNPs were reported as strongly correlated ($r^2 = 0.94$ in Icelandic population) [12]. The Korean GWAS highlighted two other DTC risk variants, namely rs6996585 and rs12542743 [13], with rs6996585 being moderately correlated with the GWAS SNP rs2439302 ($r^2 = 0.67$ in Asian population of 1000Genomes). All associated variants at this locus were located within the *NRG1* gene, which encodes neuregulin 1, a signalling protein that mediates cell-cell interactions and plays a role in the development of multiple organ systems.

Altogether, several DTC risk variants have been identified at 2q35 and 8p12 but few of them have been further explored with fine-mapping or functional studies to identify the causal ones. Here, our goal was to perform a fine-mapping analysis of these two loci using genotypic data from the EPITHYR GWAS involving Europeans, Polynesians and Melanesians, three populations with distinguishable genetic architecture, and functional annotations to highlight the potential causal variants.

RESULTS

We used data from EPITHYR consortium in which GWAS data was available for 1,855 cases and 2,321 controls of Europeans, Polynesians and Melanesians ancestries. The distribution of the study participants by study, age groups, sex and TC histology is presented in Table 1 for each ethnic group. Women represented more than 82% of cases in each ethnic group. As expected, the majority of cases were PTC (> 82% of cases in each ethnic group). The number of SNPs included in the current analysis at loci 2q35 and 8p12 are provided for each ethnic group in Supplementary Table 1.

Locus 2q35

Fine-mapping

In Europeans, among the 2,682 analysed SNPs at 2q35, 42 SNPs were associated with DTC risk at $p < 5 \times 10^{-8}$ (Supplementary Table 2). We replicated the associations from previous GWAS with SNPs rs6579952 (OR = 1.26, $p = 2.9 \times 10^{-10}$), rs11693806 (OR = 1.43, $p = 1.1 \times 10^{-10}$), rs1382435 (OR = 1.32, $p = 1.2 \times 10^{-7}$) and rs966423 (OR = 1.27, $p = 2.6 \times 10^{-6}$) (Supplementary Table 3). However, in our sample, the most significant association was observed for the genotyped SNP rs57481445 (OR = 1.43, $p = 8.6 \times 10^{-11}$) (Figure 1A and Supplementary Table 2). In Europeans, this variant was strongly correlated with rs11693806 ($r^2 = 0.98$) and 38 of the most significantly associated SNPs (Supplementary Table 4). No significant additional signal was observed after conditioning the association test on the top SNP rs57481445 (Figure 1B), suggesting that any residual association signal due to some additional independent SNPs was unlikely within the region.

In Melanesians and Polynesians, no SNP reached the standard genome-wide significance P -value threshold of 5×10^{-8} (Figure 1C and 1D) and the four previously reported associated SNPs in European GWAS, rs6579952, rs11693806, rs1382435 and rs966423, were not replicated here (Supplementary Table 3). Unlike the European population, very high effect allele frequencies were observed for these SNPs. The most significant SNPs in this locus were rs7575155 in Melanesians (OR = 2.42, $p = 1.1 \times 10^{-2}$) and rs60862033 (OR = 6.3, $p = 1.8 \times 10^{-3}$) in Polynesians (Figure 1C and 1D).

The LD patterns at 2q35 in the three ethnic groups were slightly similar with LD blocks of stronger LD between SNPs in Melanesians and Polynesians (Supplementary Figure 1). The identified SNPs rs57481445, rs16857609 and rs3821098 were located within the same LD block in the three populations.

In the pooled analysis, the most significant association at locus 2q35 was observed for the genotyped SNP rs3821098 (OR = 1.44, $p = 1.6 \times 10^{-12}$) (Supplementary Figure 2A) which is in strongly correlated

to rs57481445 ($r^2 > 0.87$ in the three populations). After conditioning on the top SNP rs3821098, no additional SNP was highlighted (Supplementary Figure 2B). Of note, although only imputed SNPs with an imputation quality score above 0.3 were kept in the analyses, all highlighted variants had info scores above 0.9.

Functional data

The colocalization analysis using eQTL data (Supplementary Figure 3A) highlighted two intronic *DIRC3* variants, namely rs16857609 (PP of causality = 0.41) and rs12990503 (PP of causality = 0.39) (Table 2 and Supplementary Figure 3B). These two SNPs were highly correlated to rs57481445 ($r^2 > 0.9$) and also strongly associated with the expression of *DIRC3* and *IGFBP5* (insulin growth factor binding protein 5) in thyroid tumour cells, according to TCGA data [17].

In Melanesians and Polynesians, the colocalization analysis also gave the highest PP of causality values for the same two SNPs identified in Europeans: rs16857609 with a $PP_{\text{Melanesians}} = 0.22$ and $PP_{\text{Polynesians}} = 0.32$, and rs12990503 with a $PP_{\text{Melanesians}} = 0.39$ and $PP_{\text{Polynesians}} = 0.35$ (Supplementary Figure 3C and 3D). Rs16857609 and rs12990503 were strongly correlated in the three ethnic groups with $r^2 > 0.88$.

Rs57481445 and rs16857609, very close to each other (< 240 bases) and located within the same LD block in the three populations (Supplementary Figure 1), are mostly located within the same functional annotation sites. They are both located within sites of histone modification, notably sites of epigenetic acetylation and methylation on histone H3 protein subunit. They are also located within markers of enhancer chromatin state and actively transcribed chromatin states, which are active states associated with genes' expression level in adipose tissue and adipocytes. Furthermore, rs16857609 is located within sites of DNase involving in three different cellular types (ovary, fetal heart and pluripotent cells). Rs12990503 is located within markers of chromatin state in adipocytes which are low or weak transcription site of gene.

Stratified analyses

We performed subgroup analyses on rs16857609 in strong LD with the top SNP rs57481445 in the three ethnic groups because it was the most interesting candidate SNP with regards to its functional annotation data. Stratified analyses were performed by study, sex, age group, TC histology and size of papillary carcinoma (Figure 3A). Heterogeneity was observed when stratifying by tumour histology (with $I^2 = 99.2\%$, $p < 10^{-4}$). We found that carriers of the risk allele [T] have a higher risk of developing PTC rather than FTC. No heterogeneity was detected in Melanesians and Polynesians. Although association was non-significant in these two populations,

Table 1: Characteristics of subjects from EPITHYR included in the analyses by ethnic group

	Europeans				Polynesians				Melanesians			
	Cases		Controls		Cases		Controls		Cases		Controls	
	<i>N</i> (1554)	%	<i>N</i> (1973)	%	<i>N</i> (146)	%	<i>N</i> (242)	%	<i>N</i> (155)	%	<i>N</i> (106)	%
Study												
CATHY	451	28.9	534	27.1	1	0.7						
Cuba	102	6.6	113	5.7								
Chernobyl	66	4.2	304	15.4								
E3N	277	17.8	287	14.5								
New Caledonia	21	1.3	70	3.5	28	19.2	55	22.7	155	100	106	100
French Polynesia			4	0.2	114	78.1	181	74.8				
YOUNG-thyr	637	41.0	661	33.5	3	2.0	6	2.5				
Age (years)												
< 25	272	17.5	513	26.0	0	0	0	0	0	0	0	0
25–50	763	49.1	887	45.2	113	77.4	186	73.3	89	57.4	69	65.1
≥ 50	519	33.4	573	29.0	33	22.6	56	26.7	66	42.6	37	34.9
Sex												
Men	275	17.7	454	23.0	14	9.6	22	8.6	9	5.8	8	7.5
Women	1279	82.3	1519	77.0	132	92.4	220	91.4	146	94.2	98	92.5
Histology												
Papillary	1417	91.2			121	82.9			132	85.2		
Follicular	137	8.8			25	17.1			23	14.8		
Size of carcinoma (only papillary)												
< 10 mm	506	32.6			67	45.9			69	44.5		
≥ 10 mm	904	58.2			45	30.8			62	40.0		
Missing	144	9.2			34	23.3			24	15.5		

ORs associated with this allele were in the same direction in the three populations. Interestingly, we observed that the risk allele [T] was more common in Oceanians (allele frequency = 0.92 and 0.82 in Melanesians and Polynesians, respectively) than in Europeans (allele frequency = 0.28).

Locus 8p12

Fine-mapping

At locus 8p12, we replicated the previously reported associations with SNPs rs2466076 (OR = 1.21, $p = 1.0 \times 10^{-4}$) and rs2439302 (OR = 1.20, $p = 2.6 \times 10^{-4}$) in Europeans. Furthermore, rs6996585 and rs1254274, identified in previous GWAS in Asians, were also associated with DTC risk in Europeans from EPITHYR (OR = 1.21, $p = 1.6 \times 10^{-4}$, and OR = 1.22, $p = 9.8 \times 10^{-5}$ respectively) (Supplementary Table 3). However, none of these associations reached a genome-wide significance level ($p < 5 \times 10^{-8}$). In Melanesians and in Polynesians, none of these SNPs were significant but the OR estimates for these four previously reported

SNPs were in the same direction than for the Europeans (Supplementary Table 3).

In Europeans, the strongest association was observed for the genotyped SNP rs28406305 (OR = 1.32, $p = 5.9 \times 10^{-8}$) and 37 SNPs were associated with DTC risk at $p < 1 \times 10^{-7}$ (Supplementary Table 5). In our European sample, this variant was moderately correlated with the SNP rs6996585 highlighted in a previous GWAS [13] with pairwise $r^2 = 0.35$. After conditioning the association test on the lead SNP rs28406305 (Figure 2A), no additional significant signal was observed (Figure 2B) suggesting the presence of only one causal variant within the region.

In Melanesians and Polynesians, no SNP was significantly associated with DTC (Figure 2C and 2D). The most significant association was observed for rs6989422 in Melanesians (OR = 2.06, $p = 4.1 \times 10^{-3}$) and for rs58964893 (OR = 1.71, $p = 2.5 \times 10^{-3}$) in Polynesians. In these populations, these two SNPs were not correlated with rs28406305 which was highlighted in Europeans ($r^2 < 0.01$).

The LD patterns at 8p12 in the three ethnic groups were similar, with larger LD blocks observed in

Table 2: Results of colocalization analyses for SNPs with posterior probability of causality > 0.10 at loci 2q35 and 8p12, in Europeans, Polynesians, Melanesians and all individuals

Locus	SNP	EA	Europeans				Polynesians				Melanesians				eQTL	
			EAF	OR	p	PP	EAF	OR	p	PP	EAF	OR	p	PP	P ^(b)	associated gene
2q35	rs16857609	T	0.28	1.42	3.7×10^{-10}	0.41	0.82	1.42	0.12	0.32	0.92	1.05	0.88	0.22	1.9×10^{-20} 2.8×10^{-19}	<i>DIRC3</i> <i>IGFBP5</i>
	rs12990503	C	0.28	1.41	3.4×10^{-10}	0.39	0.17	1.48	0.35	0.35	0.08	0.90	0.75	0.39	1.9×10^{-20} 1.2×10^{-19}	<i>DIRC3</i> <i>IGFBP5</i>
8p12	rs2439304 ^(a)	A	0.50	1.16	1.7×10^{-3}	0.65	1.11	1.11	0.52	0.52	0.41	1.03	0.84	0.35	3.1×10^{-39}	<i>NRG1</i>

EA: effect allele; EAF: effect allele frequency; OR: odds ratio; p: p-value; PP: posterior probability of causality. PP were derived from the colocalization analysis (using eQTL p-values obtained from PancanQTL database). (a) Imputed SNP, the quality score of imputation was equal to 1 for the 3 ethnic groups. (b) P-values of eQTLs of thyroid tumor cells from PancanQTL database.

Melanesians (Supplementary Figure 4). However, the SNPs rs2439304 and rs28406305 were located in the same LD block in Melanesians but not in Europeans nor in Polynesians.

In the pooled analysis, the most significant association was observed for the genotyped SNP rs28406305 (OR = 1.30, $p = 4.1 \times 10^{-8}$). After conditioning the analysis on the top SNP rs28406305, no additional SNP was highlighted (Supplementary Figures 2C and 2D).

Functional data

Using eQTL data from PancanQTL, the colocalization analysis gave the highest PP for rs2439304 in the three ethnic groups (PP_{Europeans} = 0.65, PP_{Melanesians} = 0.52 and PP_{Polynesians} = 0.35) (Table 2 and Supplementary Figure 5). In these three groups, this SNP was strongly correlated with the GWAS SNP rs2439302 ($r^2 = 1$), but moderately correlated with the top SNPs ($r^2 = 0.50$ with rs28406305 in Europeans, $r^2 = 0.40$ with rs6989422 in Melanesians and $r^2 = 0.59$ with rs58964893 in Polynesians). Furthermore, this SNP was not significantly related to DTC (in Europeans: OR = 1.20, $p = 2.6 \times 10^{-4}$; in Polynesians: OR = 1.10, $p = 0.55$; in Melanesians: OR = 1.05, $p = 0.79$). Also, rs2439304 is an eQTL strongly associated with the expression of *NRG1* in thyroid tumor cells with $p = 3.1 \times 10^{-39}$ in PancanQTL. Rs2439304 is located in the promoter region of *NRG1*. It alters many enhancer sites of histone modifications and markers of chromatin states in several tissues. This SNP is notably located within two markers of actively transcribed chromatin states, which are active states associated with genes' expression level in adipocytes. Rs2439304 is also located within enhancer site of epigenetic methylation to the DNA packaging protein Histone H3 in adipocytes.

Rs28406305, which is the top SNP in Europeans, is located within two sites of epigenetic methylation to the DNA packaging protein Histone H3 H3K4me3 in adipocytes and liver cells. Among the top SNPs that are in the same LD block than rs28406305, rs7844425 ($r^2 = 0.96$, OR = 1.32, $p = 7.6 \times 10^{-8}$) was more likely to have a deleterious impact based on the probability score of being likely a regulatory variant calculated with RegulomeDB and the results of Haploreg V4. This SNP is located within

seven sites of epigenetic methylation to H3K4me3 in several tissues, within two sites of epigenetic acetylation to the DNA packaging Histone H3 H3K27ac, and within several markers of chromatin states in different tissues. Importantly, rs7844425 is an eQTL strongly associated with the expression of *NRG1* in thyroid tissue, with $p = 2.3 \times 10^{-17}$ in PancanQTL. However, this SNP was not highlighted in the colocalization analysis, with very low PP in the three ethnic groups (PP_{Europeans} = 8.8×10^{-28} , PP_{Melanesians} = 3.4×10^{-27} and PP_{Polynesians} = 3.4×10^{-27}).

Stratified analyses

We performed subgroup analyses for rs2439304, which had the highest PP of causality based on eQTL data, by stratifying by study, sex, age group, histology and size of the thyroid carcinoma in Europeans (Figure 3B). No heterogeneity by subgroup was observed. However, while comparing the results by ethnic groups, we observed a significant heterogeneity between these groups ($I^2 = 92.3\%$, $p < 10^{-4}$). This variant was imputed with high imputation quality in all populations (info score > 0.99).

We also performed subgroups analyses for rs7844425 which was the most functionally interesting SNP associated to DTC based on the results of Regulome DB and Haploreg V4. In Europeans we observed a significant heterogeneity by histological type of DTC ($I^2 > 81\%$, $p = 0.02$), with a positive association observed only for PTC (Figure 4).

DISCUSSION

The few GWAS conducted so far allowed the identification of several DTC susceptibility loci with the strongest associations located on 2q35, 9q22.33, 8p12 and 14q13.3. Published fine-mapping studies have been conducted at loci 9q22.22 and 14q13.3 [15, 16], including our previous study investigating loci 9q22.33 and 14q.13.3 in Melanesians [15], but the present study is, to our knowledge, the first study that investigates loci 2q35 and 8p12 in details. Moreover, investigating these genomic regions not only in Europeans but also in Melanesians and Polynesians from New-Caledonia and French Polynesia is

of particular relevance given that these populations have one of the highest incidence rates of DTC in the world. Although high parity and obesity have been associated with DTC risk in Melanesians [18, 19] and could explain a part of the elevated incidence, the remaining reasons are yet unknown. Therefore, deciphering the genetic architecture of DTC in these under-studied populations could shed new light on the susceptibility to this endocrine-related cancer.

The trans-ethnic aspect is an asset of the EPITHYR study. Indeed, it was showed that trans-ethnic meta-analyses that combine results of the same trait across genetically diverse populations can aid fine-mapping by capitalizing on ethnic differences in LD patterns [20]. The LD patterns at 2q35 and 8p12 in the three ethnic groups were slightly similar although LD blocks were larger in Melanesians and Polynesians than in Europeans (Supplementary Figures 1 and 4). However, the limited sample size of the Polynesian and Melanesian populations led to a lack of power to detect associations. In order to gain statistical power, we chose to conduct a pooled

analysis using LMM, so that the differences in LD patterns between the individuals were taken into account by adjusting for the genetic relationship matrix as random effect.

For the two investigated loci, association between DTC risk and the previously reported GWAS SNPs were replicated in Europeans (Supplementary Table 3). The PCA performed in Polynesians and in Melanesians using the HGDP reference panel showed that Polynesians are genetically close to East Asians (Supplementary Figure 6). Therefore, despite the lack of power, we could have expected to replicate the associations previously highlighted in Asians, among Polynesians from EPITHYR and to a lesser extend in Melanesians. At locus 8p12, although the results were not statistically significant, the reported ORs in these two populations were in the same direction than those previously reported in the Asian GWAS. However, at locus 2q35, the previously associated SNPs in Asian GWAS were in the same direction in Polynesians but not in Melanesians (Supplementary Table 3).

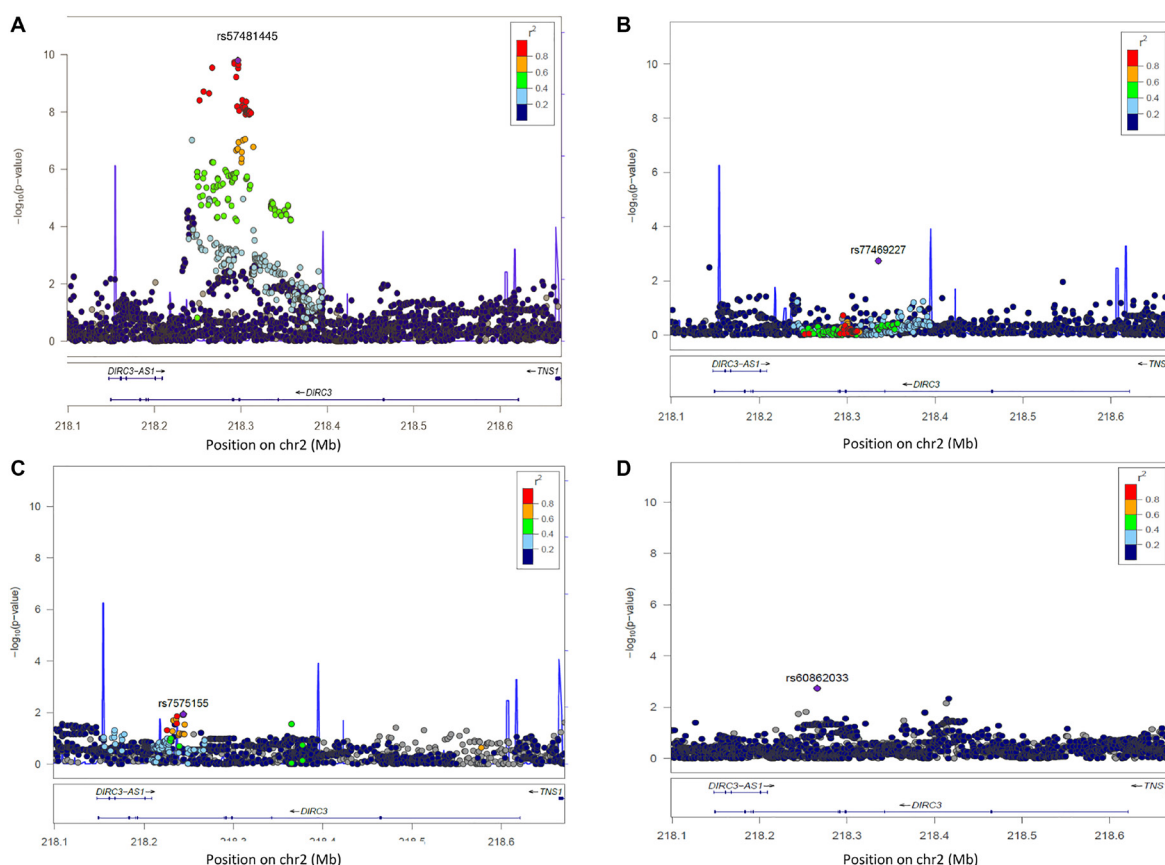


Figure 1: Regional plot of single-SNP association and conditional regression at locus 2q35 in Europeans, Melanesians and Polynesians. On Y axis, are the $-\log_{10}(p\text{-value})$; on X axis, the position of SNPs and genes on the chromosome. The color of each SNP spot reflects its r^2 with the most associated SNP (in purple). The correlation between SNP is ranged from high (in red) to low (in blue). Grey dots correspond to missing variants (imputed SNPs with no genotype with probability > 0.9). (A) Plot of $-\log_{10}(p\text{-values})$ of SNP association results at locus 2q35 in Europeans. (B) Plot of $-\log_{10}(p\text{-values})$ of SNP association results at locus 2q35 in Europeans, after conditioning the test on SNP rs57481445. (C) Plot of $-\log_{10}(p\text{-values})$ of SNP association results at locus 2q35 in Melanesians. (D) Plot of $-\log_{10} p\text{-values}$ of SNP association results at locus 2q35 in Polynesians.

In Europeans, the conditional analysis showed that the association of DTC risk with the top SNP rs57481445 was correlated with other SNPs within the region, suggesting that only one independent signal lies at 2q35. Our analyses pinpointed rs16857609 as a possible causal SNP for this region in Europeans, Melanesians and Polynesians. This SNP was strongly associated with the expression of the two nearby genes *DIRC3* and *IGFBP5* in thyroid tumour cells, according to TCGA data. Insulin-like growth factors (IGFs) play a pivotal role in tissue homeostasis, regulating cell proliferation, differentiation and migration during development and also in adulthood. Since it was shown that *IGFBP5* is overexpressed in PTC, it could have a proliferative effect on thyroid tumour cells [21]. We found that rs16857609 is located within different methylation sites involving histone H3 in many tissues. Also, this SNP is located within many regulatory elements (chromatin state linked to an enhancer, histone modifications in DNA methylation or acetylation) that

control genes expression in adipose nuclei tissue and in adipocyte stem cells. Since obesity has been consistently associated with an increased risk of TC in epidemiological studies [3], these observations support the hypothesis that rs16857609 could play a role in the aetiology of DTC via complex obesity-associated pathways. The frequency of the risk allele [T] is much higher in Oceanians than in Europeans and this could explain a part of the highest incidence of DTC observed in the former populations if the association is confirmed. Interestingly, rs16857609 has been associated with increased breast cancer risk in Europeans [22]. Since *IGFBP5* is a physiological regulator of epithelial cell survival in the mammary gland [23], it was proposed that a deregulation of *IGFBP5* expression induced by variants located upstream of *DIRC3* [24] could lead to the development of a breast tumour. Altogether, the link between rs16857609, breast cancer risk and *IGFBP5* on one side, and the link between TC and *IGFBP5* on the other suggests that a deregulation of *IGFBP5* expression

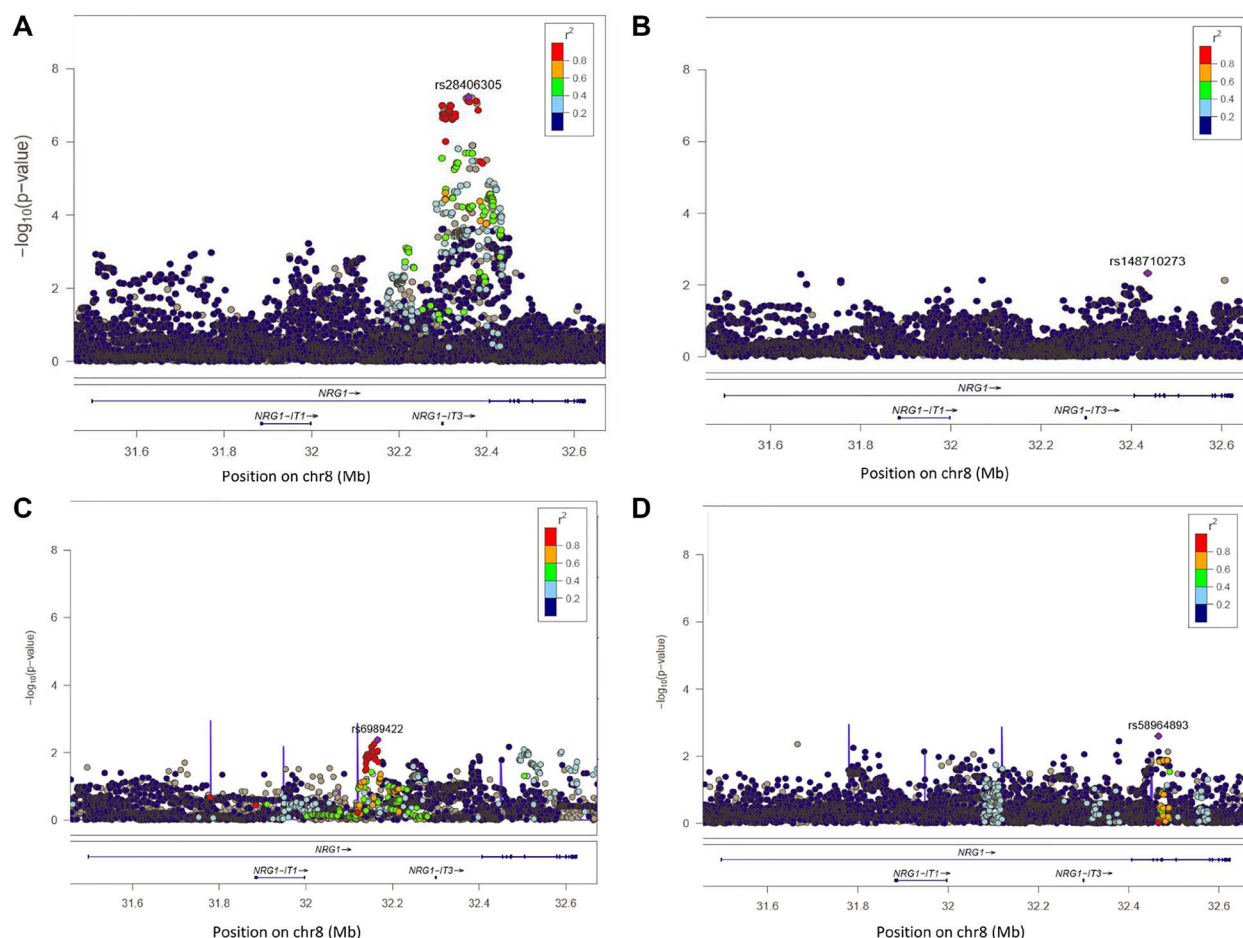


Figure 2: Regional plot of single-SNP association and conditional regression at locus 8p12 in Europeans, Melanesians and Polynesians. On Y axis, are the $-\log_{10}(p\text{-value})$; on X axis, the position of SNPs and genes on the chromosome. The color of each SNP spot reflects its r^2 with the most associated SNP (in purple). The correlation between SNP is ranged from high (in red) to low (in blue). Grey dots correspond to missing variants (imputed SNPs with no genotype with probability > 0.9). (A) Plot of $-\log_{10}(p\text{-values})$ of SNP association results at locus 8p12 in Europeans. (B) Plot of $-\log_{10}(p\text{-values})$ of SNP association result rs28406305 at locus 8p12 in Europeans, after conditioning the test on SNP rs28406305. (C) Plot of $-\log_{10}(p\text{-values})$ of SNP association results at locus 8p12 in Melanesians. (D) Plot of $-\log_{10}(p\text{-values})$ of SNP association results at locus 8p12 in Polynesians.

may also lead to the proliferation of thyroid cells and eventually to PTC development.

The conditional analysis at locus 8p12 in Europeans revealed a single association signal. Among the most associated SNPs, we reported rs7844425 as the variant which is the most likely to have a functional deleterious impact. This SNP is located in the upstream transcript region of the major isoforms of the *NRG1* gene expressed in human thyroid tissue and thyroid cell lines, namely NM_004495, NM_013958 and NM_001160008. The gene encodes a membrane glycoprotein that mediates cell-cell signalling and plays a critical role in the growth and development of multiple organ systems. Rs7844425

is predicted to alter conserved elements involved in the regulation of *NRG1* expression. To our knowledge, no previous GWAs highlighted rs7844425 or other SNPs from our study located in the same LD block. Interestingly, this SNP is strongly correlated with rs17716295 in Europeans ($r^2 = 0.93$) which was associated with antipsychotic treatment response in schizophrenia patients [25] and it was recognized that disturbances in the balance of thyroid hormones contribute to mental status deregulation recognized that thyroid dysfunction is associated with psychiatric disorders, with a higher proportion of hypothyroidism among patients with schizophrenia [26]. We found that rs7844425 is only moderately correlated

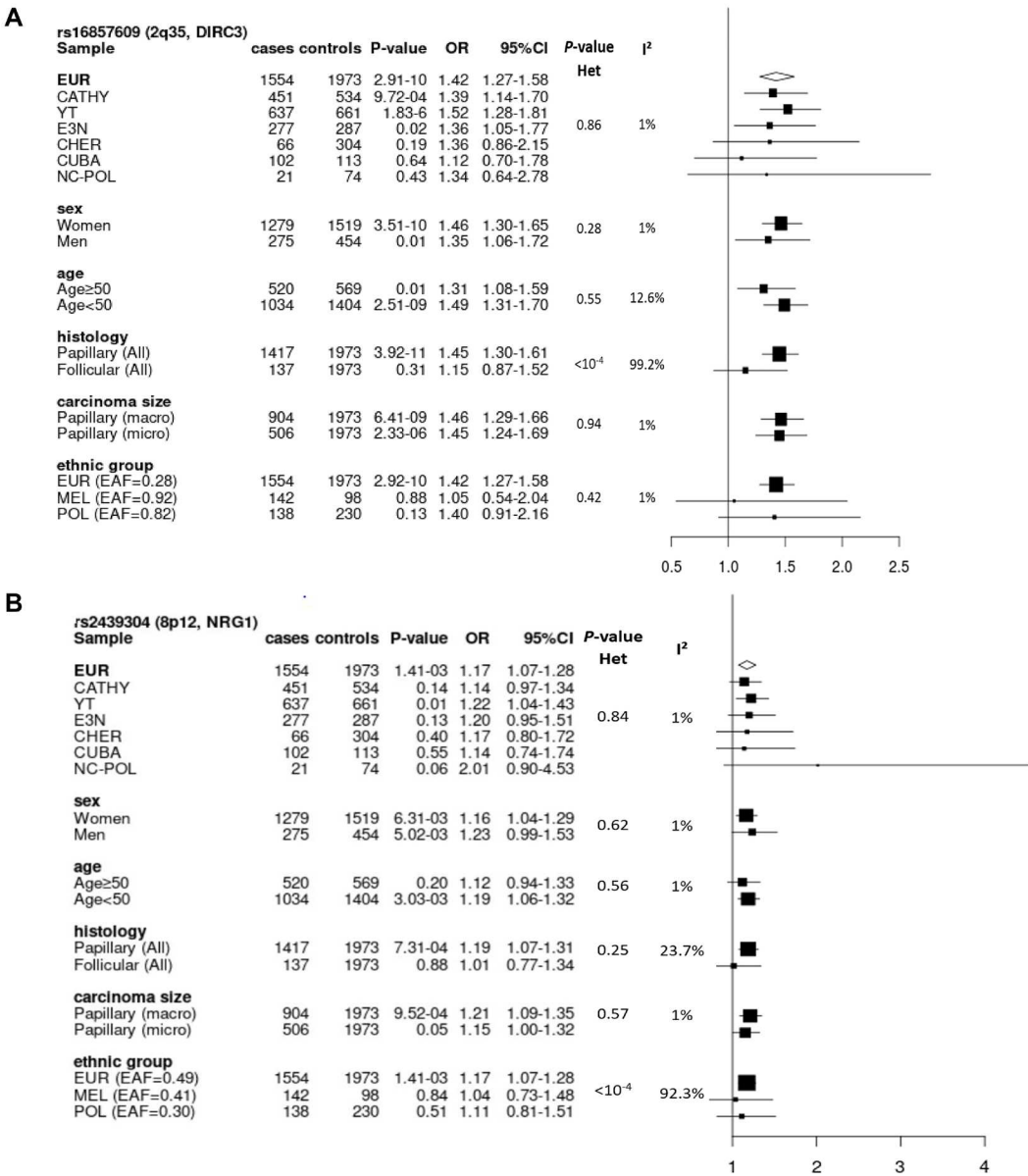


Figure 3: Forest plots of subgroup analyses for rs16857609 (A) and rs2439304 (B), by study, sex, age group, histology and size of carcinoma in Europeans, and by ethnic group in the entire dataset. (A) Result for the genotyped SNP rs16857609 in *DIRC3* at 2q35 (Effect allele (T) frequency in Europeans: 0.28). (B) Result for the imputed SNP rs2439304 in *NRG1* at 8p12 (Effect allele (A) frequency in Europeans: 0.49). EAF: effect allele frequency; EUR: Europeans; MEL: Melanians; POL: Polynesians; Het: heterogeneity.

with rs2439304 that we also highlighted at the locus 8p12 ($r^2 = 0.51$ in European). Although rs2439304 was not significantly associated with DTC in our populations ($OR = 1.2, p = 0.001$ in Europeans), it had the highest PP of causality in the three ethnic groups using eQTL data at 8p12. Rs2439304 is strongly correlated with rs2439302 which was reported in a previous GWAS on DTC risk [27] ($r^2 = 1$). Both SNPs are located within a 32kb LD block in intron 1 of *NRG1*. By performing haplotype analysis of this LD block, He *et al.* proposed that the impact of several risk alleles on transcriptional regulation is probably combinatorial in dictating *NRG1* expression and in conferring susceptibility to DTC. Furthermore, by using CHIP-seq experiment and luciferase assay in primary thyroid cell culture, they showed that the risk allele of rs2439304 and the risk allele of rs2439302 increase expression of all three major *NRG1* isoforms when comparing with the wild type allele. This result is in line with the eQTL analysis conducted by Son *et al.* suggesting that the increased expression of *NRG1* in thyroid tissue might influence the development or progression of TC [13] but do not support results by Gudmundsson *et al.* who reported that carriers of rs2439302 risk allele have a lower expression of *NRG1* [9].

For both SNPs, rs7844425 and rs16857609, we performed analyses restricted to PTC as sensitivity tests but results did not differed significantly from the overall analysis so we chose not to present these results (data not shown).

In summary, we confirmed the previously associated SNPs in Europeans and in Asians at loci 2q35 and 8p12, and our fine-mapping approach identified new potential causal SNPs at each locus (rs16857609 at locus 2q35, rs7844425 and rs2439304 at locus 8p12) to be prioritized for functional studies.

MATERIALS AND METHODS

Study populations and genotyping

We used GWAS data from the EPITHYR consortium that included subjects from seven case-control studies: three were conducted in Metropolitan France (CATHY [28], YOUNG-thyr [29] and E3N [30] studies), two in South Pacific Islands [19, 31], one in Cuba [32] and one in the Gomel region of Belarus, affected by the Chernobyl accident [33]. Study protocols were approved by the local ethic committees and written informed consent was obtained from all participants. Genotyping was conducted using the Infinium Oncoarray-500k Beadchip (Illumina) [34] comprising about 500,000 SNPs to which we incorporated 13,759 additional custom markers involved in TC or thyroid hormones metabolism. Details of the participating studies [19, 28–32, 35] and quality controls (QC) of the genotyping were described in detail previously [14]. QC and ethnicity definition were similar to those described by the OncoArray consortium [34].

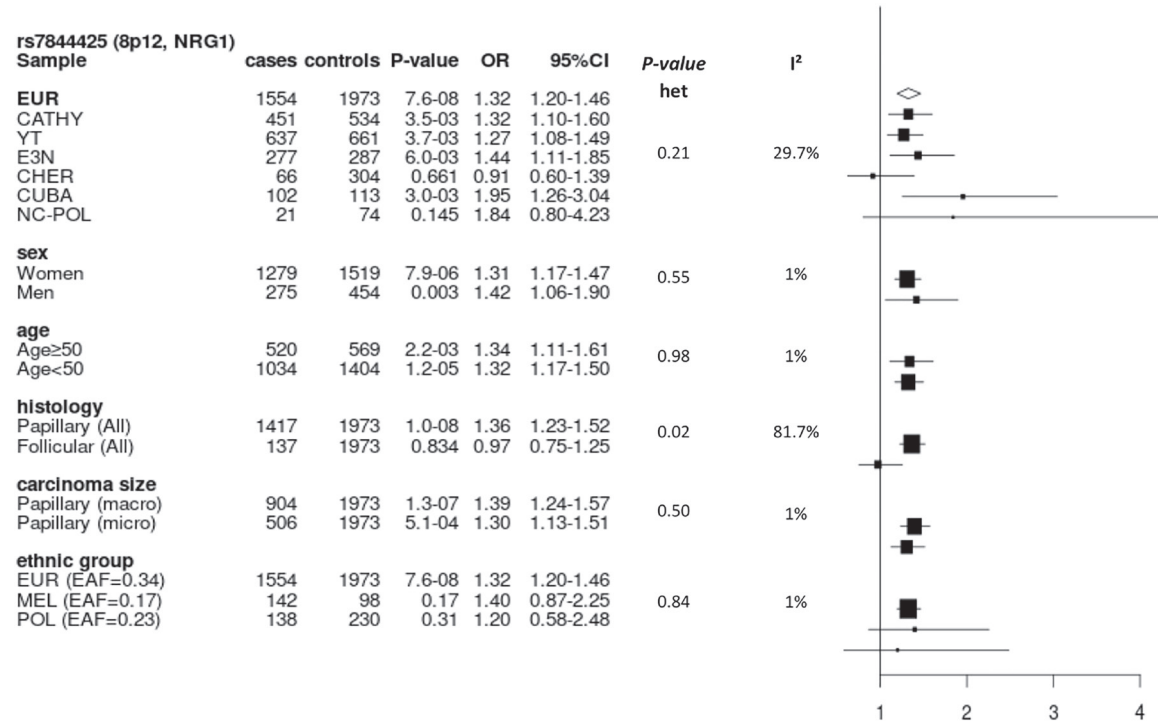


Figure 4: Forest plot of subgroup analyses for rs7844425 stratifying on study, sex, age group, histology and size of carcinoma in Europeans, and stratifying on population group in the entire dataset. Result for the imputed SNP rs7844425 in *NRG1* at 8p12 (Effect allele (A) frequency in Europeans: 0.34). EAF: effect allele frequency; EUR: Europeans; MEL: Melanesians; POL: Polynesians; Het: heterogeneity.

After QC, we analysed data from 3,527 Europeans (1,554 cases/1,973 controls) and 649 Asians (301 cases/348 controls). Asian participants came mostly from Polynesia and New-Caledonia so we renamed this group as Oceanians. In order to compare this population to other reference panel, we conducted a principal component analysis (PCA) using the Human Genome Diversity Panel (HGDP) [36]. We observed that Polynesians were closest to East Asian population whereas Melanesians were closest to the Oceanian population (which was from Bougainville). Based on this PCA (Supplementary Figure 6), we separated Oceanian participants into the Melanesian (155 cases/106 controls) and Polynesian subgroups (146 cases/242 controls) in order to reduce the heterogeneity in this population.

SNP selection

At locus 2q35

For the current study, we defined a mapping interval of 162,680bp at 2q35 (218,231,649–218,394,329; NCBI build 37 assembly) that included SNPs in the same linkage disequilibrium (LD) block as the two GWAS SNPs, rs966423 and rs6759952 (*i.e.* SNPs with $r^2 > 0.2$) according to the 1000 Genomes Project (phase 1 CEU) [37]. We first catalogued 299 SNPs correlated with one of the two GWAS SNPs at $r^2 > 0.2$ and with a minor allele frequency (MAF) greater than 0.02 in Europeans according to the 1000 Genomes Project. We added a set of SNPs to tag the remaining SNPs of the region with $MAF > 0.02$ and $r^2 > 0.8$ by forcing these 299 SNPs and SNPs that were already on the chip. A total of 398 unique tag SNPs were identified of which 226 SNPs were not in the initial design of the chip and were therefore added for genotyping.

As the mapping region was entirely included within the *DIRC3* gene, we chose to extend the region to analyse to 50 kb upstream and downstream from the gene boundaries (572 kb, positions 218,098,746–218,671,316; NCBI build 37 assembly).

Finally, 452 SNPs were genotyped for this region, and 441 of them passed QC in Europeans and in Oceanians (Melanesians and Polynesians). These genotypic data were used to impute the non-genotyped SNPs using IMPUTEv2 [37] and the 1000 Genomes Project (phase 3) as the reference panel [37]. For the imputation in Oceanian individuals, we added the 26 Oceanians from Simons Genome Diversity Panel (SGDP, phase 3) [38] to the 1000 Genomes reference panel. All monomorphic imputed SNPs were excluded and only imputed SNPs with an imputation quality score above 0.3 were kept in the analyses. Hence, the analysis of the 2q35 locus was based on 452 genotyped SNPs and 2,230 imputed SNPs in Europeans, 441 genotyped SNPs and 2053 imputed SNPs in Polynesians, and 441 genotyped SNPs and 1,484 imputed SNPs in Melanesians.

At locus 8p12

Similarly, at 8p12, the region to be enriched was defined based on the LD block containing rs2439302 (positions 32,158,293–32,641,727; NCBI build 37 assembly). We added 492 SNPs to the design of the chip in this region.

As this region, which included the three previously identified SNPs (rs2439302, rs6996585 and rs2466076), is located within the *NGR1* gene, we extended the analysis to all SNPs located 50kb upstream and downstream from this gene boundaries (1,215 kb; 31,456,820–32,672,558; NCBI build 37 assembly).

Thus, the analysis of the 8p12 locus was based on 552 genotyped SNPs and 5,941 imputed SNPs in Europeans, 497 genotyped SNPs and 5,316 imputed SNPs in Polynesians, and 497 genotyped and 3,728 imputed SNPs in Melanesians.

Statistical analyses

Imputed SNPs were analysed as expected genotype count (gene dosage). Odds ratios (OR) and 95% confidence interval (CI) testing the single marker association with DTC were calculated using the logistic mixed model (LMM) with the genetic relationship matrix as random effect to account for population structure and cryptic relatedness [39]. All analyses were adjusted for age, sex and study and were stratified by ethnic group. A pooled analysis was also performed with Europeans, Melanesians and Polynesians using the corresponding genetic relationship matrix as a random effect in the LMM. These analyses were performed using Gaston R package [40]. LD between variants was calculated from the imputed data with PLINK (v1.9), using the genotype with a probability higher than 0.9. Otherwise, genotypes were missing for the SNP.

For some variants of interest, we also conducted stratified analysis by study, age group, sex and histology. Heterogeneity of odds ratios across the stratification groups was assessed by using Cochran Q test and I^2 index [41].

For each locus, we first performed conditional analysis in order to test for residual association after accounting for a lead SNP, which permits to make an assumption about the number of independent signals within the region.

Functional annotation

To assess the possible functional effect of the associated SNPs, several annotation databases with different features were interrogated. We were mainly interested in features involved in regulatory aspects of gene expression including gene promoters, enhancers, transcription factor binding sites (TFBS), untranslated

regions (UTR), DNase I hypersensitive sites (DHSs) and histone modifications.

We annotated gene promoter regions and UTR using Gene Elements of GenCode [42], and enhancer elements using FANTOM5 [43] and Roadmap collections [44]. Enhancers are short regions of DNA playing a key role in the transcription of genes by increasing the likelihood that transcription of a particular gene will occur. We also used annotations of super-enhancers available in the catalogue by Hnisz et al. [45]. These elements consist of large clusters of transcriptional enhancers driving expression of genes that define cell identity. We used TFBS database [46] for annotations of TFBS and retrieved annotation of DNase I hypersensitive sites (DHSs) from three sources: Roadmap and maps of human DHS sites as defined by Thurman et al. [47] and by Maurano et al [48]. DHSs are regions of chromatin functionally related to transcriptional activity and are sensitive to cleavage by the DNase I enzyme. Finally, we used chromatin state annotation from the Roadmap database.

To investigate noncoding variants, we also used the following freely available tools, RegulomeDB [49] (<https://regulomedb.org/regulome-search/>) and Haploreg V4 (<http://www.broadinstitute.org/mammals/haploreg>). These tools predict whether a given SNP may alter the regulatory protein binding sites, chromatin structure and histone modifications.

In addition, expression Quantitative Trait Locus (eQTL) information was used to perform a colocalization analysis within each genomic region using the Coloc R package [50]. This Bayesian method allows to assess whether two related traits share a causal variant by calculating a posterior probability of causality (PP) for each SNP. For this analysis, we used results of the association test between each SNP and DTC risk and results of the association test between each SNP and gene expression in thyroid tumour cells that are publicly available in the PancanQTL database [17]. PancanQTL obtained the gene expression profiles from The Cancer Genome Atlas (TCGA) (<http://cancergenome.nih.gov/abouttcga>). In Europeans, among the 2,682 tested variants at locus 2q35, 564 were cis-eQTLs, that is, an eQTL that acts in (locally) to a gene. At locus 8p12, among the 6,493 tested variants, 736 were cis-eQTLs.

Abbreviations

DHS: DNase I hypersensitive sites; DTC: differentiated thyroid carcinoma; eQTL: expression quantitative trait locus; FTC: follicular thyroid carcinoma; GWAS: genome-wide association study; LD: linkage disequilibrium; LMM: logistic mixed model; MAF: minor allele frequency; OR: odds ratio; PCA: principal component analysis; PP: posterior probability; PTC: papillary thyroid carcinoma; PY: person-year; QC: quality control; SNP: single nucleotide polymorphism; TC: thyroid cancer; TFBS: transcription factor binding site;

TSH: thyroid stimulating hormone; UTR: untranslated region.

Author contributions

Project supervision: FL and TT. Selection of add-ons on the Oncoarray chip: TT, FL, CX, CR, FDV. Quality controls and data management: by JG, MK. Genotyping was supervised and carried out by DB, CB, AB and JFD. Bioinformatics analyses were performed by JG, MK, OK, PES. Statistical analyses were carried out by JG, PES. Subjects recruitment, biological material collection and handling was organized and carried out by FDV, PG, CS, AVG, CM, PLP, MCB, AK, EO, RMO, CX, YR, CR, FR, EA, TT. JG, FL and TT drafted the manuscript. All authors reviewed the manuscript and approved the final version of the paper.

CONFLICTS OF INTEREST

Authors have no conflicts of interest to declare.

FUNDING

This work was supported by the Institut National du Cancer (Inca grant 9533), the Fondation ARC pour la Recherche sur le Cancer (ARC grant PGA120150202302), the Centre National de Recherche en Génomique Humaine, CEA, Electricité de France (conseil scientifique de Radioprotection d'EDF, grant EP 2019-01), Fondation de France (Grant 2016-70074) and Plan Cancer (Project ThyGenRad, Grant ENV201415). The E3N cohort received support from the MGEN, Gustave Roussy and Ligue contre le cancer for its set up and maintenance. The E3N cohort was also supported by a state grant from the Agence Nationale pour la Recherche (ANR, grant number ANR-10-COHO-0006) within the Investissement d'Avenir program; and from Ministère de l'enseignement supérieur, de la recherche et de l'innovation (MESRI, grant number 2102 918823). JG was the recipient of a PhD fellowship from Région Ile-de-France.

REFERENCES

1. IARC. Thyroid fact sheet, Globocan 2018. 2018. <https://gco.iarc.fr/today/data/factsheets/cancers/>.
2. 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.* 2007; 16:62–70. <https://doi.org/10.1097/01.ccej.0000236244.32995.e1>.
3. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body fatness and cancer - Viewpoint of the IARC working group. *N Engl J Med.* 2016; 375: 794–8. <https://doi.org/10.1056/NEJMsrl606602>. [PubMed]

4. Lee JH, Hwang Y, Song RY, Yi JW, Yu HW, Kim SJ, Chai YJ, Choi JY, Lee KE, Park SK. Relationship between iodine levels and papillary thyroid carcinoma: A systematic review and meta-analysis. *Head Neck*. 2017; 39:1711–8. <https://doi.org/10.1002/hed.24797>. [PubMed]
5. Caini S, Gibelli B, Palli D, Saieva C, Ruscica M, Gandini S. 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*. 2015; 26:511–8. <https://doi.org/10.1007/s10552-015-0546-z>. [PubMed]
6. Hemminki K, Eng C, Chen B. Familial Risks for Nonmedullary Thyroid Cancer. *J Clin Endocrinol Metab*. 2005; 90:57. <https://doi.org/10.1210/jc.2005-0935>. [PubMed]
7. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M, Magnusdottir DN, Matthiasdottir S, Stacey SN, et al. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet*. 2009; 41:460–4. <https://doi.org/10.1038/ng.339>. [PubMed]
8. 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, Demidchik YE, et al. The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Hum Mol Genet*. 2010; 9:2516. <https://doi.org/10.1093/hmg/ddq123>. [PubMed]
9. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Masson G, He H, Jonassdottir A, Sigurdsson A, Stacey SN, Johannsdottir H, Th Helgadóttir H, Li W, Nagy R, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat Genet*. 2012; 44:319–22. <https://doi.org/10.1038/ng.1046>. [PubMed]
10. Köhler A, Chen B, Gemignani F, Elisei R, Romei C, Figlioli G, Cipollini M, Cristaudo A, Bambi F, Hoffmann P, Herms S, Kalembe M, Kula D, et al. Genome-Wide Association Study on Differentiated Thyroid Cancer. *J Clin Endocrinol Metab*. 2013; 98:E1674–81. <https://doi.org/10.1210/jc.2013-1941>. [PubMed]
11. Figlioli G, Köhler A, Chen B, Elisei R, Romei C, Cipollini M, Cristaudo A, Bambi F, Paolicchi E, Hoffmann P, Herms S, Kalembe M, Kula D, et al. Novel Genome-Wide Association Study-Based Candidate Loci for Differentiated Thyroid Cancer Risk. *Sci Rep*. 2014; 5:1–7. <https://doi.org/10.1210/jc.2014-1734>. [PubMed]
12. Gudmundsson J, Thorleifsson G, Sigurdsson JK, Stefansdottir L, Jonasson JG, Gudjonsson SA, Gudbjartsson DF, Masson G, Johannsdottir H, Halldorsson GH, Stacey SN, Helgason H, Sulem P, et al. A genome-wide association study yields five novel thyroid cancer risk loci. *Nat Commun*. 2017; 8:6–13. <https://doi.org/10.1038/ncomms14517>. [PubMed]
13. Son HY, Hwangbo Y, Yoo SK, Im SW, Yang SD, Kwak SJ, Park MS, Kwak SH, Cho SW, Ryu JS, Kim J, Jung YS, Kim TH, et al. Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer. *Nat Commun*. 2017; 8:15966. <https://doi.org/10.1038/ncomms15966>.
14. Truong T, Lesueur F, Sugier P, Guibon J, Xhaard C, Karimi M, Kulkarni O, Lucotte E, Guizard A, Ren Y, Adjad E. Multiethnic genome-wide association study of differentiated thyroid cancer in the EPITHYR consortium. *Int J Cancer*. 2021 Feb 2. <https://doi.org/10.1002/ijc.33488>. [Epub ahead of print].
15. Tcheandjieu C, Lesueur F, Sanchez M, Baron-Dubourdieu D, Guizard AV, Mulot C, Laurent-Puig P, Schvartz C, Truong T, 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; 139:617–27. <https://doi.org/10.1002/ijc.30088>. [PubMed]
16. Jendrzewski J, Liyanarachchi S, Eiterman A, Thomas A, He H, Nagy R, Senter L, Sworzak K, de la Chapelle A. Fine mapping of 14q13 reveals novel variants associated with different histological subtypes of papillary thyroid carcinoma. *Int J Cancer*. 2019; 144:503–12. <https://doi.org/10.1002/ijc.31933>. [PubMed]
17. Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, Feng J, Liu R, Diao L, Guo AY, Miao X, Han L. PancanQTL: Systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res*. 2018; 46:D971–6. <https://doi.org/10.1093/nar/gkx861>. [PubMed]
18. 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*. 2005; 161:1056–65. <https://doi.org/10.1093/aje/kwi136>. [PubMed]
19. 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*. 2007; 166:1140–9. <https://doi.org/10.1093/aje/kwm204>. [PubMed]
20. Li YR, Keating BJ. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Med*. 2014; 6:91. <https://doi.org/10.1186/s13073-014-0091-5>. [PubMed]
21. Stolf BS, Carvalho AF, Martins WK, Runza FB, Brun M, Hirata R, Neves EJ, Soares FA, Postigo-Dias J, Kowalski LP, Reis LFL. Differential expression of IGFBP-5 and two human ESTs in thyroid glands with goiter, adenoma and papillary or follicular carcinomas. *Cancer Lett*. 2003; 191:193–202. [https://doi.org/10.1016/S0304-3835\(02\)00679-1](https://doi.org/10.1016/S0304-3835(02)00679-1).
22. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013; 45:353–61. <https://doi.org/10.1038/ng.2563>. [PubMed]

23. Marshman E. Insulin-like growth factor binding protein 5 and apoptosis in mammary epithelial cells. *J Cell Sci.* 2003; 116:675–82. <https://doi.org/10.1242/jcs.00263>. [PubMed]
24. Wyszynski A, Hong C, Lam K, Michailidou K, Lytle C, Yao S, Zhang Y, Bolla MK, Wang Q, Dennis J, Hopper JL, Southey MC, Schmidt MK, et al. An intergenic risk locus containing an enhancer deletion in 2q35 modulates breast cancer risk by deregulating IGFBP5 expression. *Hum Mol Genet.* 2018; 25:3863–76. <https://doi.org/10.1093/hmg/ddw223>. [PubMed]
25. Jajodia A, Kaur H, Kumari K, Kanojia N, Gupta M, Baghel R, Sood M, Jain S, Chadda RK, Kukreti R. Evaluation of genetic association of neurodevelopment and neuroimmunological genes with antipsychotic treatment response in schizophrenia in indian populations. *Mol Genet Genomic Med.* 2016; 4:18–27. <https://doi.org/10.1002/mgg3.169>. [PubMed]
26. Sharif K, Tiosano S, Watad A, Comaneshter D, Cohen AD, Shoenfeld Y, Amital H. The link between schizophrenia and hypothyroidism: a population-based study. *Immunol Res.* 2018; 66:663–6. <https://doi.org/10.1007/s12026-018-9030-7>. [PubMed]
27. He H, Li W, Liyanarachchi S, Wang Y, Yu L, Genutis LK, Maharry S, Phay JE, Shen R, Brock P, De La Chapelle A. The Role of NRG1 in the Predisposition to Papillary Thyroid Carcinoma. *J Clin Endocrinol Metab.* 2018; 103:1369–79. <https://doi.org/10.1210/jc.2017-01798>. [PubMed]
28. Cordina-Duverger E, Leux C, Neri M, Tcheandjeu C, Guizard AV, Schvartz C, Truong T, Guénel P. Hormonal and reproductive risk factors of papillary thyroid cancer: A population-based case-control study in France. *Cancer Epidemiol Biomarkers Prev.* 2017; 48:78–84. <https://doi.org/10.1016/j.canep.2017.04.001>. [PubMed]
29. Xhaard C, De Vathaire F, Cléro E, Maillard S, Ren Y, Borson-Chazot F, Sassolas G, Schvartz C, Colonna M, Lacour B, Danzon A, Velten M, Marrer E, et al. Original Contribution Anthropometric Risk Factors for Differentiated Thyroid Cancer in Young Men and Women From Eastern France: A Case-Control Study. *Am J Epidemiol.* 2015; 182:202. <https://doi.org/10.1093/aje/kwv048>. [PubMed]
30. Clavel-Chapelon F, Guillas G, Tondeur L, Kernaeguen C, Boutron-Ruault MC. Risk of differentiated thyroid cancer in relation to adult weight, height and body shape over life: The French E3N cohort. *Int J Cancer.* 2010; 126:2984–90. <https://doi.org/10.1002/ijc.25066>. [PubMed]
31. Brindel P, Doyon F, Rachédi F, Boissin JL, Sebbag J, Shan L, Chungue V, Bost-Bezeaud F, Petitdidier P, Paoaafaite J, Teuri J, de Vathaire F. Anthropometric factors in differentiated thyroid cancer in French Polynesia: a case-control study. *Cancer Causes Control.* 2009; 20:581–90. <https://doi.org/10.1007/s10552-008-9266-y>. [PubMed]
32. Lence-Anta JJ, Xhaard C, Ortiz RM, Kassim H, Pereda CM, Turcios S, Velasco M, Chappe M, Infante I, Bustillo M, García A, Clero DF, Maillard S, et al. Environmental, Lifestyle, and Anthropometric Risk Factors for Differentiated Thyroid Cancer in Cuba: A Case-Control Study. *Eur Thyroid J.* 2014; 3:189–96. <https://doi.org/10.1159/000362928>. [PubMed]
33. Stezhko VA, Buglova EE, Danilova LI, Drozd VM, Krysenko NA, Lesnikova NR, Minenko VF, Ostapenko VA, Petrenko SV, Polyanskaya ON, Rzhetski VA, Tronko MD, Boblyova OO, et al. A Cohort Study of Thyroid Cancer and Other Thyroid Diseases after the Chornobyl Accident: Objectives, Design and Methods. *Radiat Res.* 2004; 161:481–92. <https://doi.org/10.1667/3148>. [PubMed]
34. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, Hunter DJ, Sellers TA. The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev.* 2017; 26:126–35. <https://doi.org/10.1038/ncomms5234.SUM01>.
35. Lonjou C, Damiola F, Moissonnier M, Durand G, Malakhova I, Masyakin V, Le Calvez-Kelm F, Cardis E, Byrnes G, Kesminiene A, Lesueur F. Investigation of DNA repair-related SNPs underlying susceptibility to papillary thyroid carcinoma reveals MGMT as a novel candidate gene in Belarusian children exposed to radiation. *BMC Cancer.* 2017; 17:328. <https://doi.org/10.1186/s12885-017-3314-5>. [PubMed]
36. Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. *Nat Rev Genet.* 2005; 6:333–40. <https://doi.org/10.1038/nrg1596>. [PubMed]
37. Howie B, Marchini J, Stephens M. Genotype Imputation with Thousands of Genomes. *G3 (Bethesda).* 2011; 1:457–7. <https://doi.org/10.1534/g3.111.001198>. [PubMed]
38. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature.* 2016; 538:201–6. <https://doi.org/10.1038/nature18964>. [PubMed]
39. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, Szpiro AA, Chen W, Brehm JM, Celedó JC, Redline S, Papanicolaou GJ, Thornton TA, et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet.* 2016; 98:653. <https://doi.org/10.1016/j.ajhg.2016.02.012>. [PubMed]
40. Perdry H, Dandine-Roulland C, Banddyopadhyay D, Kettner L, Gaston. Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. CRAN. 2018.
41. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002; 21:1539–58. <https://doi.org/10.1002/sim.1186>. [PubMed]
42. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, et al. GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* 2012; 22:1760–74. <https://doi.org/10.1101/gr.135350.111>. [PubMed]

43. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014; 507:455–61. <https://doi.org/10.1038/nature12787>. [PubMed]
44. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518:317–29. <https://doi.org/10.1038/nature14248>. [PubMed]
45. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke H, Young RA. Transcriptional super-enhancers connected to cell identity and disease. *Cell*. 2013; 155:934–47. <https://doi.org/10.1016/j.cell.2013.09.053>. [PubMed]
46. Yang JH, Li JH, Jiang S, Zhou H, Qu LH. ChIPBase: A database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res*. 2013; 41:D177–87. <https://doi.org/10.1093/nar/gks1060>. [PubMed]
47. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012; 489:75–82. <https://doi.org/10.1038/nature11232>. [PubMed]
48. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* (80-). AAAS. 2012; 337:1190–5. <https://doi.org/10.1126/science.1222794>. [PubMed]
49. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012; 22:1790–7. <https://doi.org/10.1101/gr.137323.112>. [PubMed]
50. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian Test for Colocalisation Between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet*. 2014; 10:e100.



OPEN

Gene network and biological pathways associated with susceptibility to differentiated thyroid carcinoma

Om Kulkarni¹, Pierre-Emmanuel Sugier², Julie Guibon^{1,2}, Anne Boland-Augé³, Christine Lonjou¹, Delphine Bacq-Daian³, Robert Olaso³, Carole Rubino², Vincent Souchard², Frédérique Rachedi⁴, Juan Jesus Lence-Anta⁵, Rosa Maria Ortiz⁵, Constance Xhaard^{2,6}, Pierre Laurent-Puig⁷, Claire Mulo⁷, Anne-Valérie Guizard^{8,9}, Claire Schvartz¹⁰, Marie-Christine Boutron-Ruault², Evgenia Ostroumova¹¹, Ausrele Kesminiene¹¹, Jean-François Deleuze³, Pascal Guénel², Florent De Vathaire², Thérèse Truong^{2,12} & Fabienne Lesueur^{1,12}✉

Variants identified in earlier genome-wide association studies (GWAS) on differentiated thyroid carcinoma (DTC) explain about 10% of the overall estimated genetic contribution and could not provide complete insights into biological mechanisms involved in DTC susceptibility. Integrating systems biology information from model organisms, genome-wide expression data from tumor and matched normal tissue and GWAS data could help identifying DTC-associated genes, and pathways or functional networks in which they are involved. We performed data mining of GWAS data of the EPITHYR consortium (1551 cases and 1957 controls) using various pathways and protein–protein interaction (PPI) annotation databases and gene expression data from The Cancer Genome Atlas. We identified eight DTC-associated genes at known loci 2q35 (*DIRC3*), 8p12 (*NRG1*), 9q22 (*FOXO1*, *TRMO*, *HEMGN*, *ANP32B*, *NANS*) and 14q13 (*MBIP*). Using the EW_dmGWAS approach we found that gene networks related to glycogenolysis, glycogen metabolism, insulin metabolism and signal transduction pathways associated with muscle contraction were overrepresented with association signals (false discovery rate adjusted p-value < 0.05). Additionally, suggestive association of 21 KEGG and 75 REACTOME pathways with DTC indicate a link between DTC susceptibility and functions related to metabolism of cholesterol, amino sugar and nucleotide sugar metabolism, steroid biosynthesis, and downregulation of ERBB2 signaling pathways. Together, our results provide novel insights into biological mechanisms contributing to DTC risk.

Differentiated thyroid carcinoma (DTC) is the most common type of endocrine cancer and accounts for 98% of all cases of thyroid cancer. It originates from epithelial follicular cells of the thyroid and includes three histological types, namely papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), and Hürthle cell carcinoma¹, with PTC representing about 85% of all thyroid malignancies². DTC incidence varies considerably around the world with age-standardized incidence rates of 10.2 per 100,000 person-years in women and

¹Inserm, U900, Institut Curie, PSL University, Mines ParisTech, 75248 Paris, France. ²Université Paris-Saclay, UVSQ, Gustave Roussy, Inserm, CESP, 94807 Villejuif, France. ³Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057 Evry, France. ⁴Centre Hospitalier Territorial de Polynésie Française, CHTPF, Piraie, Tahiti, 98713 Papeete, French Polynesia. ⁵Instituto Nacional de Oncología y de Radiobiología, INOR, La Habana, Cuba. ⁶University of Lorraine, INSERM CIC 1433, Nancy CHRU, Inserm U1116, FCRIN, INI-CRCT, 54000 Nancy, France. ⁷Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université Paris Descartes, Université Paris Diderot, EPIGENETEC, 75006 Paris, France. ⁸Registre Général des Tumeurs du Calvados, Centre François Baclesse, 14000 Caen, France. ⁹Inserm U1086-UCNB, Cancers and Prevention, 14000 Caen, France. ¹⁰Registre des Cancers Thyroïdiens, Institut Jean Godinot, 51100 Reims, France. ¹¹Environment and Radiation Section, International Agency for Research on Cancer, 69008 Lyon, France. ¹²These authors contributed equally: Thérèse Truong and Fabienne Lesueur. ✉email: fabienne.lesueur@curie.fr

3.1 per 100,000 person-years in men in 2018³. In most countries, DTC incidence has increased at a faster rate than most other malignancies during the last few decades. It is now the 5th most frequent cancer in women, whereas it was ranked 14th 20 years ago^{3,4}. The causes underlying geographic, ethnic and temporal variations are still unknown. It could be explained by environmental and genetic factors, as well as changes in screening practices. In particular, some have attributed the increase in DTC incidence to improved diagnosis that leads to the detection of small tumors of minimal clinical relevance (microcarcinomas)⁴, whereas others argue that more sensitive diagnostic procedures cannot completely explain this increase of DTC rates⁵. The only well-established environmental risk factor for DTC is exposure to ionizing radiation during childhood and adolescence⁶ but it does not appear to have contributed importantly to these trends⁷. Anthropometric factors such as excess weight, tall height and large body size have also been consistently associated with risk of DTC^{8–14}. In particular, a large meta-analysis showed that increase of weight, body mass index (BMI), waist or hip circumference and waist-to-hip ratio are associated with a greater risk of PTC, FTC and anaplastic thyroid cancer¹⁵. Because DTC occurs more frequently in women than in men, it was also suspected to be associated with hormonal and reproductive factors among women⁷. Thyroid cancer is also characterized by having one of the highest familial risk of any cancer supporting heritable predisposition¹⁶. In spite of such a high familial risk, few chromosomal loci have been implicated in DTC so far. Genome-wide association studies (GWAS)^{17–23} including ours²⁴ identified mainly four DTC susceptibility loci at 9q22, 14q13, 2q35 and 8p12, which were replicated in different populations. However, the identified single nucleotide polymorphism (SNPs) were shown to account for only about 10% of the DTC familial risk, emphasizing that much remains to be discovered. Furthermore, all published studies examined genetic associations with DTC at the individual SNP or gene level. Data mining of GWAS data at a higher level of complexity using systems biology is still an under-explored topic. Of the seven GWAS performed for DTC, only one of the published datasets was additionally analyzed using pathway identification methods²¹. None of the studies employed protein–protein interaction (PPI) network-based methods to explore links between associated genes, and only two of them used expression quantitative trait loci (eQTL) data to identify potential causal regulatory sequence variants at DTC associated loci^{21,23}. However, such approaches have been successful in identifying new susceptibility alleles for other complex traits. For instance, analysis of GWAS data using pathway-based enrichment methods successfully identified IL12/IL23 pathways associated with Crohn disease, involving genes that were subsequently identified as susceptibility genes only through meta-analysis of several GWAS²⁵. Integrative analyses of GWAS data, eQTL and PPI networks also provided valuable biological insights in some complex diseases, such as Alzheimer disease²⁶ and asthma²⁷.

Here we re-analyzed the genome-wide genotyping data from seven case–control studies on DTC from the EPITHYR consortium using protein–protein interaction databases, various resources for pathway maps, as well as available eQTL data on DTC from The Cancer Genome Atlas (TCGA) to annotate SNPs and to identify biological mechanisms contributing to DTC susceptibility.

Results

Data set and results of the standard SNP-level analysis. We used GWAS data from the EPITHYR consortium²⁴ that included subjects from case-control studies conducted in Metropolitan France (CATHY¹¹, YOUNG-thyr¹³ and E3N¹² studies), South Pacific Islands (Polynesia⁹ and New Caledonia⁸), Cuba¹⁴ and the Gomel region of Belarus, affected by the Chernobyl accident²⁸. Characteristics of the study participants of European ancestry included in the analyses are described in Table 1.

In the SNP-level analysis, 258 SNPs reached the standard genome-wide significance P -value threshold of 5×10^{-8} . All SNPs were located in the known DTC susceptibility loci at 2q35, 8p12, 9q22.33 and 14q13.3 (Supplementary Figure 1A). No additional signal was evidenced when the analysis was restricted to PTC cases only (Supplementary Figure 1B).

Gene-level analysis. According to GENCODE release 28, the analyzed SNPs were mapped to 19,120 protein-coding genes that were next used in the gene-based association test from VEGAS2²⁹. This analysis identified eight genes associated with DTC with a false discovery rate adjusted p -value (P_{FDR}) < 0.05 , namely, *DIRC3*, *NRG1*, *FOXE1*, *TRMO*, *HEMGN*, *ANP32B*, *NANS* and *MBIP*, all of them being located at known DTC susceptibility loci (Table 2). The analysis restricted to PTC cases identified *TRIM14* at 9q22.33 in addition to these eight genes (Supplementary Table S1).

To get more insight in the genetic mechanisms of DTC, we interrogated whether SNPs in or nearby the associated genes were acting as cis-eQTLs (defined as a SNP within 1 Mb from the gene transcriptional start site) using transcriptome data from 497 DTC cases from TCGA available through the PanCanQTL project³⁰. We identified a number of cis-eQTL for *DIRC3*, *IGFBP5*, *NRG1*, *TRMO* and *NANS* (Table 2), indicating that SNPs at the associated loci could alter the regulation of the expression of these five genes.

Pathway-level analysis. To clarify which biological pathways are involved in the etiology of DTC, we next used Vegas2Pathway which uses gene-based p -values from VEGAS2 and pathway definitions from Kyoto Encyclopedia of Genes and Genomes (KEGG)^{31–33}, Reactome³⁴ and Gene Ontology (GO)³⁵ (Table 3). Out of 380 KEGG pathways, 361 were tagged by SNPs from our dataset. Of those, 21 pathways were associated with DTC risk with $P_{EMP} < 0.05$, with the top pathway being linked to cholesterol metabolism; however, none of the highlighted pathways were significant after correction for multiple testing (Supplementary Table S2). Only four of the 21 highlighted pathways involved one of the eight genes identified in the gene-level analysis, namely ‘Messenger RNA biogenesis’ (*ANP32B*), ‘Amino sugar and nucleotide sugar metabolism’ (*NANS*), ‘EGFR tyrosine kinase inhibitor resistance’ (*NRG1*) and ‘Transfer RNA biogenesis’ (*TRMO*) and the three latter pathways were not associated anymore with DTC after excluding SNPs tagging these candidate genes.

Study	Cases		Controls	
	N = 1551	%	N = 1957	%
CATHY	450	29.0	533	27.2
Cuba	102	6.6	103	5.3
Chernobyl	66	4.3	304	15.5
E3N	276	17.8	287	14.7
New Caledonia	21	1.4	68	3.5
French Polynesia	0	0	4	0.2
YOUNG-Thyr	636	41.0	658	33.6
Age (years)				
[0–10]	5	0.3	43	2.2
[10–20]	115	7.4	301	15.4
[20–30]	378	24.4	425	21.7
[30–40]	335	21.6	382	19.5
[40–50]	199	12.8	239	12.2
≥ 50	519	33.5	567	29.0
Mean age [range]	40.6 [7–83]	–	37.0 [5–80]	–
Sex				
Female	1276	82.3	1508	77.1
Male	275	17.7	449	22.9
Histology				
Papillary	1414	91.2	–	–
Follicular	137	8.8	–	–

Table 1. Characteristics of participants of the seven EPITHYR case–control studies used in the gene-, pathway- and network-level analyses.

Locus	Gene	Gene P_{EMP}^a	Gene P_{FDR}^b	#SNPs (N) ^c	Top SNP	OR _{per allele} ^d	95%CI	$P_{per allele}$	Cis-eQTL (N) ^e	eGene ^f
2q35	<i>DIRC3</i>	1.00×10^{-7}	0.0038	451	rs16857611	1.42	1.28–1.58	1.25×10^{-10}	223	<i>DIRC3</i> , <i>IGFBP5</i>
8p12	<i>NRG1</i>	2.00×10^{-6}	0.0063	523	rs28406305	1.34	1.21–1.48	3.19×10^{-8}	197	<i>NRG1</i>
9q22.33	<i>FOXE1</i>	1.00×10^{-7}	0.0038	138	rs10739513	1.60	1.44–1.79	1.88×10^{-17}	87	<i>TRMO</i>
9q22.33	<i>TRMO</i>	1.00×10^{-7}	0.0038	92	rs7046645	1.58	1.42–1.77	9.85×10^{-17}	64	<i>TRMO</i>
9q22.33	<i>HEMGN</i>	1.00×10^{-7}	0.0038	61	rs7037324	1.49	1.35–1.65	8.03×10^{-15}	31	<i>TRMO</i> , <i>NANS</i>
9q22.33	<i>ANP32B</i>	1.00×10^{-7}	0.0038	39	rs56145417	1.30	1.18–1.43	1.91×10^{-7}	4	<i>TRMO</i> , <i>NANS</i>
9q22.33	<i>NANS</i>	4.00×10^{-6}	0.0095	20	rs7870926	1.30	1.18–1.43	2.08×10^{-7}	2	<i>TRMO</i> , <i>NANS</i>
14q13.3	<i>MBIP</i>	3.00×10^{-6}	0.0082	47	rs116909374	2.14	1.66–2.76	4.88×10^{-9}	0	None

Table 2. Genes associated with DTC risk, SNPs and eQTL within or in the vicinity of these genes, and effect of eQTL on the expression of genes in cis. ^aEmpirical p-value of the association test at the gene level. ^bp-value of the association test with DTC risk at the gene level, after FDR correction. ^cNumber of analyzed SNP within the gene or at ± 50 kb from the gene boundaries. ^dPer allele Odds Ratio (OR) for the top SNP at the gene locus. ^enumber of eQTL at the gene locus. ^fGene whose expression is affected by the cis-eQTL.

Out of 2020 Reactome pathways, 1698 included SNPs from our dataset. Of those, 75 definitions were associated with DTC risk with $P_{EMP} < 0.05$. After excluding SNPs tagging the eight candidate genes, the 16 pathways involving *NRG1* were not associated anymore with DTC (Supplementary Table S3). Gene Ontology (GO) definitions related to biological processes, molecular functions and cellular components were also investigated. Associated definitions with $P_{EMP} < 0.05$ are listed in (Supplementary Table S4).

To assess similarities between pathways associated with DTC at $P_{EMP} < 0.05$ identified with KEGG, Reactome and GO, we performed pairwise comparisons between definitions of the three databases. Pairs of pathways with Jaccard Index > 0.1 are shown in Supplementary Table S5. We found that 16 of the top KEGG pathways showed some similarity with some Reactome pathways, and 40, 4 and 6 KEGG pathways showed some similarity with GO biological processes, cell components and molecular functions, respectively, confirming the inter-feature dependencies of the pathways highlighted with the three pathway databases.

Database	Definitions (N)	Definition tagged with oncoarray SNPs (N)	Definitions with $P_{EMP}^a < 0.05$ (N)	Source	Version
KEGG (HSA and BRITE definitions)	380	361	21	https://www.genome.jp/kegg-bin/get_htext?hsa00001.keg	v88 (Oct, 2018)
REACTOME	2020	1698	75	https://reactome.org/download/current/ReactomePathways.txt	v66 (Sep 2018)
GO biological process	5214	5203	253	GO database: http://purl.obolibrary.org/obo/go/go-basic.obo GO annotations: https://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2go.gz	(Oct 2018)
GO cellular component	655	652	38		
GO molecular function	1060	1059	31		

Table 3. Pathway definitions used in the pathway-level analysis. ^aEmpirical p-value of the association test with DTC risk at the pathway level.

Co-analysis of thyroid carcinoma gene expression and GWAS data. To gain a deeper understanding of the genetic architecture of DTC, we then combined TCGA genomic expression data from 59 PTC/normal tissue sample pairs with PPI networks and EPITHYR GWAS data using the EW_dmGWAS approach³⁶. The 19,129 genes containing OncoArray SNPs were involved in 4524 subnetworks describing binary interactions (that is direct PPI) and in 6590 subnetworks describing co-complex interactions. Among the 19,129 genes, 16,386 were differentially expressed between normal and tumor tissue. This information was used by the algorithm to assign edge weights to the nodes of the subnetworks to rank them for downstream gene enrichment analysis. Hence, the top 1% subnetworks contributing to DTC susceptibility involved 72 genes with binary interactions and 143 genes with in co-complex interactions. Using Reactome pathway definitions, we found that five pathways were significantly enriched, including ‘Glycogen breakdown (glycogenolysis)’ ($P_{FDR} = 7.9 \times 10^{-3}$), ‘Glycogen metabolism’ ($P_{FDR} = 2.5 \times 10^{-2}$) and two pathways related to muscle contraction when binary interactions annotations were considered (Table 4). Furthermore, we found 47 Reactome pathways significantly enriched when co-complex interactions annotations were considered (Table 4). Using GO definitions, we found 14 biological processes, 12 cellular components and 4 molecular functions associated with DTC (Fig. 1) while with KEGG definitions, only the ‘Ribosome’ ($P_{FDR} = 2.7 \times 10^{-43}$) and ‘starch and sucrose metabolism’ ($P_{FDR} = 4.6 \times 10^{-2}$) pathways were significantly enriched.

Discussion

Incorporating gene network and pathway classification tools in GWAS data analysis can point toward significantly overrepresented molecular pathways, which had not been picked up in traditional single-SNP analysis due to the stringent genome-wide significance level and to the limited power of some case–control studies to identify low-risk alleles. To our knowledge, this is the first study on DTC susceptibility where integrative analyses of GWAS data, gene expression data in tumor, and biological pathways or physical PPI network data were performed to gain biological insights in the disease. Data mining of the EPITHYR GWAS data using several systems biology annotation tools and various analysis strategies has allowed to identify high confidence candidate pathways for subsequent analyses to be further explored to understand the underlying mechanisms of DTC carcinogenesis. Indeed, although the EPITHYR GWAS is one of the GWAS with the largest number of DTC cases reported so far (1551 cases and 1957 controls of European ancestry), new findings from the classical per-SNP analysis were limited and the eight candidate genes (*DIRC3*, *NRG1*, *FOXE1*, *TRMO*, *HEMGN*, *ANP32B*, *NANS* and *MBIP*) identified in the gene-level analysis were all located in the well characterized DTC susceptibility loci 2q35, 8p12, 9q22.33, and 14q13.3²². Moreover, a functional link between these candidate genes could not clearly be established at this point.

SNPs in the nuclear long noncoding RNA *DIRC3* (*disrupted in renal cancer 3*) have been associated with both thyroid stimulating hormone level and DTC risk¹⁹, and it was shown that *DIRC3*, playing a role in tumor invasion and multifocality, represents a potential prognostic factor for PTC³⁷. Interestingly, the top SNP for *DIRC3*, rs16857611, is an eQTL which downregulates the expression of *DIRC3* and the expression of its neighboring tumor suppressor gene *IGFBP5* whose product belongs to a family of proteins which interacts with insulin-like growth factors (IGFs) involved in regulation of vital processes such as cell proliferation, differentiation and apoptosis. In melanoma, it was shown that *DIRC3* activates expression of *IGFBP5* through modulating chromatin structure and suppressing SOX10 binding to putative regulatory elements³⁸, suggesting that the two genes at the 2q35 could represent potential therapeutic targets for both melanoma and DTC.

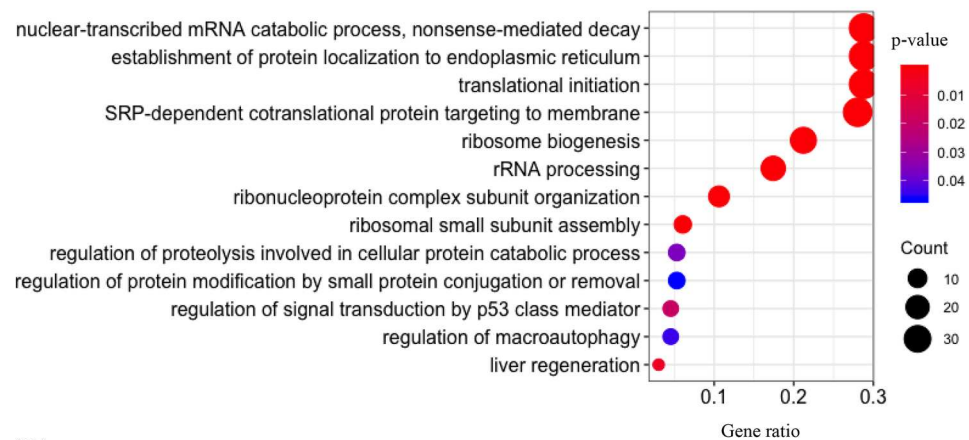
NRG1 encodes the membrane glycoprotein Neuregulin 1, which acts on the erb-b2 receptor tyrosine kinase (ERBB) family of tyrosine kinase receptors. It is the major HER3 ligand, which promotes its engagement with HER2 kinase and the subsequent transphosphorylation of HER3. It is involved in regulation of MAPK and AKT signaling pathways which are involved in thyroid carcinoma cells proliferation and survival³⁹. *FOXE1*, is a thyroid-specific transcription factor essential for thyroid gland development and maintenance of the differentiated state. In vitro studies in thyroid cancer cell lines revealed that *FOXE1* modulates cell migration, suggesting a role in epithelial-to-mesenchymal transition⁴⁰. *HEMGN*, also known as *EDAG-1* (*Embryonic develop-associated gene 1*) is upregulated in thyroid carcinoma tissues and cells, and it has been proposed to regulate the proliferation and apoptosis of cells via PI3K/Akt signaling pathway⁴¹.

ANP32B (*Acidic Nuclear Phosphoprotein 32 Family Member B*) is a multifunctional protein working as a cell cycle progression factor as well as an anti-apoptotic protein is involved in hepatocellular carcinoma⁴². The gene

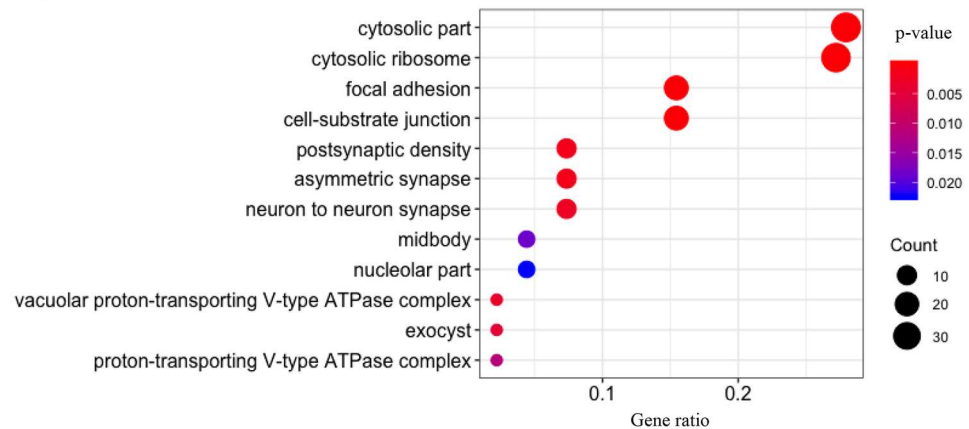
Interactions type	Enriched reactome pathway	Genes (N)	P_{EMP}^a	P_{FDR}^b
Binary	Striated muscle contraction	7	2.2×10^{-10}	7.3×10^{-08}
	Muscle contraction	9	3.6×10^{-07}	5.9×10^{-05}
	Glycogen breakdown (glycogenolysis)	3	7.3×10^{-05}	7.9×10^{-03}
	Glycogen metabolism	3	3.2×10^{-04}	2.5×10^{-02}
	The role of GTSE1 in G2/M progression after G2 checkpoint	4	3.9×10^{-04}	2.5×10^{-02}
Co-complex	Peptide chain elongation	37	4.4×10^{-53}	1.2×10^{-50}
	Viral mRNA translation	37	4.4×10^{-53}	1.2×10^{-50}
	Formation of a pool of free 40S subunits	38	1.6×10^{-52}	3.0×10^{-50}
	Eukaryotic translation elongation	37	3.4×10^{-52}	3.1×10^{-50}
	Selenocysteine synthesis	37	3.4×10^{-52}	3.1×10^{-50}
	Eukaryotic translation termination	37	3.4×10^{-52}	3.1×10^{-50}
	Nonsense mediated decay (NMD) independent of the exon junction complex (EJC)	37	9.1×10^{-52}	7.2×10^{-50}
	L13a-mediated translational silencing of ceruloplasmin expression	38	1.3×10^{-50}	8.7×10^{-49}
	GTP hydrolysis and joining of the 60S ribosomal subunit	38	1.9×10^{-50}	1.2×10^{-48}
	Nonsense-mediated decay (NMD)	38	6.3×10^{-50}	3.1×10^{-48}
	Nonsense mediated decay (NMD) enhanced by the exon junction complex (EJC)	38	6.3×10^{-50}	3.1×10^{-48}
	Eukaryotic translation initiation	38	2.9×10^{-49}	1.2×10^{-47}
	Cap-dependent translation initiation	38	2.9×10^{-49}	1.2×10^{-47}
	SRP-dependent cotranslational protein targeting to membrane	37	1.5×10^{-48}	5.9×10^{-47}
	Selenoamino acid metabolism	37	1.5×10^{-47}	5.3×10^{-46}
	Influenza viral RNA transcription and replication	38	4.0×10^{-47}	1.4×10^{-45}
	Major pathway of rRNA processing in the nucleolus and cytosol	41	5.8×10^{-47}	1.9×10^{-44}
	Regulation of expression of SLITs and ROBOs	40	8.3×10^{-46}	2.5×10^{-44}
	Influenza life cycle	38	9.3×10^{-46}	2.7×10^{-44}
	rRNA processing in the nucleus and cytosol	41	6.2×10^{-45}	1.7×10^{-43}
	Influenza Infection	38	2.2×10^{-44}	5.8×10^{-43}
	rRNA processing	41	5.8×10^{-44}	1.5×10^{-42}
	Signaling by ROBO receptors	40	3.2×10^{-41}	7.8×10^{-40}
	Translation	40	5.9×10^{-36}	1.6×10^{-34}
	Infectious disease	41	1.9×10^{-32}	4.2×10^{-31}
	Metabolism of amino acids and derivatives	39	1.9×10^{-30}	4.1×10^{-29}
	Formation of the ternary complex, and subsequently, the 43S complex	19	5.1×10^{-26}	1.0×10^{-24}
	Translation initiation complex formation	19	9.4×10^{-25}	1.8×10^{-23}
	Ribosomal scanning and start codon recognition	19	9.4×10^{-25}	1.8×10^{-23}
	Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	19	1.4×10^{-24}	2.5×10^{-23}
	TCR signaling	6	1.1×10^{-3}	2.0×10^{-2}
	Regulation of mRNA stability by proteins that bind AU-rich elements	5	1.7×10^{-3}	3.0×10^{-2}
	FBXL7 down-regulates AURKA during mitotic entry and in early mitosis	4	1.9×10^{-3}	3.2×10^{-2}
	Insulin receptor recycling	3	2.1×10^{-3}	3.3×10^{-2}
	Regulation of RUNX3 expression and activity	4	2.1×10^{-3}	3.3×10^{-2}
	Insulin processing	3	2.3×10^{-3}	3.5×10^{-2}
	Stabilization of p53	4	2.4×10^{-3}	3.5×10^{-2}
	Iron uptake and transport	4	2.5×10^{-3}	3.7×10^{-2}
	Downstream TCR signaling	5	2.8×10^{-3}	3.9×10^{-2}
	G2/M transition	7	3.0×10^{-3}	4.2×10^{-2}
	Mitotic G2-G2/M phases	7	3.2×10^{-3}	4.2×10^{-2}
	rRNA modification in the nucleus and cytosol	4	3.2×10^{-3}	4.2×10^{-2}
	Transferrin endocytosis and recycling	3	3.4×10^{-3}	4.4×10^{-2}
	Cilium assembly	7	3.5×10^{-3}	4.4×10^{-2}
	ROS, RNS production in phagocytes	3	3.8×10^{-3}	4.6×10^{-2}
	p53-dependent G1 DNA damage response	4	4.0×10^{-3}	4.7×10^{-2}
	p53-dependent G1/S DNA damage checkpoint	4	4.0×10^{-3}	4.7×10^{-2}

Table 4. Reactome pathways enriched with genes involved in the top 1% subnetworks obtained when considering binary and co-complex interactions. ^aEmpirical p-value of the association test with DTC risk at the pathway level. ^bp-value of the association test with DTC risk at the pathway level, after FDR correction.

(A) Enriched GO Biological Processes



(B) Enriched GO Cellular Components



(C) Enriched GO Molecular Functions

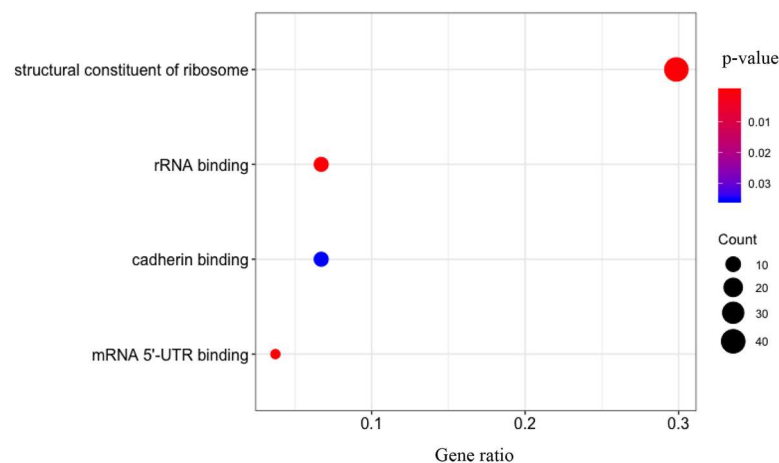


Figure 1. GO enrichment analysis using high throughput co-complex interaction annotations for (A) biological processes, (B) cellular components, (C) molecular functions. For each plot, Y-axis represents a significant GO definition, and X-axis represents the counts of enriched genes (Gene ratio). The gradient of color represents the different *p*-values, and size of the dot represents the count number of genes in each GO term.

product of *MBIP* regulates the JNK pathway which is involved in intracellular signaling of thyroid and other human cancers⁴³. A role for *TRMO* encoding a tRNA methyltransferase involved in tRNA processing, and for *NANS* involved in sialic acid synthesis process in tumorigenesis has not been evidenced so far although variation in the expression of the two genes has been observed in thyroid carcinoma according to TCGA transcriptomic data, suggesting that further studies on these candidates should be pursued.

Network and pathway tools were developed for computational gene prioritization to make use of functional information from gene and protein databases to gain more insights in disease-related biological mechanisms. They are, therefore, biased toward the well-studied genes; interactions and pathways and SNPs in non-coding genes (lncRNA, miRNA, and snRNA) and in intergenic regions are omitted. Here, we assigned SNPs that lie within 50 kb on either side of a gene's coding sequence boundaries to compute its association *p* value which is used by pathways and networks centric approaches. With this gene definition, only 1951 (0.4%) OncoArray SNPs that passed QC were not linked to a gene.

Our study also illustrates that the alternative representation of the same biological pathway (e.g. in KEGG, Reactome and GO) may influence the results of the statistical enrichment analysis and that pathway-centric approaches employed to interpret -omics data rely on the choice of the pathway databases used. This is because pathways are often described at varying level of detail, with diverse data types and with vaguely defined boundaries. In particular, KEGG includes pathway maps such as for metabolism, genetic, and environmental information processing, while Reactome is based on biological reactions (binding, activation, translocation, degradation) and GO is a hierarchy of terms representing biological processes, molecular functions and cellular components. We chose to use these three databases that differ in the average number of pathways they contain, the average number of proteins per pathway, the types of biochemical interactions they incorporate, and the subcategories that they provide (e.g. signal transduction, genetic interaction, and metabolic) to gain a comprehensive overview of pathway landscapes altered in DTC. Reassuringly, we found that, although limited in number, similar pathways named differently across databases were associated with DTC with comparable *p*-values.

We also found that the EW_dmGWAS approach combining association, differential gene co-expression profile and functional interaction analyses was more informative than the standard pathway-based approaches to prioritize gene sets. The integrative analysis showed that genes involved in 'muscle contraction', 'glycogen' and 'insulin' related pathways play a role in the etiology of DTC. Using this approach, KEGG definitions, GO biological processes and GO molecular functions were also significantly enriched for ribosome-related pathways, and GO cellular components were enriched for several nervous system related terms.

Although top ranked pathways highlighted in the standard pathway analyses with VEGAS2 did not achieve statistical significance, some are in line with those evidenced with EW_dmGWAS or play a role in the development of other carcinomas, and therefore could help prioritizing the best candidates for therapeutic intervention. For instance, VEGAS2 analyses suggested involvement of cholesterol homeostasis pathways in DTC and indicate that MAPK pathway, involved in melanoma and other cancer types^{44,45}, and steroid biosynthesis related pathways, involved in prostate cancer⁴⁵ could also be altered in DTC. Other top ranked pathways related to NCAM1, a neural cell adhesion molecule shown to be involved in development of the nervous system as well as in cancer metastasis⁴⁶ or to ERBB2 and other growth factors acting in thyroid tumorigenesis were also evidenced.

Gene-networks and pathways highlighted in this study were identified using the European subset of EPITHYR only, due to the limited sample size and heterogeneity in the population structure in other ethnic groups. Since allele frequencies of SNPs and DTC risk associated to them may vary from one population to another, pathway- or network-guided GWAS analysis in larger non-European samples will be useful to confirm the association with biological functions identified in Europeans and also to identify new ones. The major advantage of approaches such as EW_dmGWAS or similar approaches like the weighted gene co-expression network analysis (WGCNA)^{47,48} is their capability to perform biologically relevant dimension reduction as a result of the analysis. However, they use results of transcriptomic data analysis which reflect the inherent complexity of multiple biological processes. Moreover, data generated from different platforms also lead to noise and error generated by variations in experiment also affect the accuracy to distinct different samples. Further improvement of the algorithms is therefore needed to facilitate identification of causal hub genes involved in molecular mechanisms that could be used as therapeutic targets of the disease. Building methods using multitype data such as gene expression data, transcriptomic data and protein data will help to identify more accurate and reliable pathways as biological markers of disease. Alternatively, deep learning models may be used to jointly learn features from different type of omics data and then predict the key genes forming the modules, as such multi-task methods have been proposed for image classification in other complex diseases⁴⁹.

To summarize, the strongest associations were found for gene sets acting in insulin resistance, amino sugar and nucleotide sugar metabolism-related pathways, which trigger weight gain, overweight or obesity reported to be positively associated with in thyroid cancer risk⁵⁰. In EPITHYR, data on weight and height are available for all participants, and association between anthropometric factors and DTC risk was investigated separately in all studies^{8–10,12–14}. In all studies, weight, height and BMI were positively associated with DTC risk. High body surface area was also investigated in three of the studies^{10,13,14}, and it was also found to increase DTC risk. These results support the relevance of the above-mentioned pathways in DTC susceptibility. Genes sets acting in signaling pathways involved in muscle contraction, were also evidence in the EW_dmGWAS analysis. Interestingly, a recent GO term and KEGG pathway enrichment analysis performed on mRNA microarray datasets for human thyroid carcinomas and adenomas indicated that some biological functions of genes that were differentially expressed in the tumors included protein binding, cardiac muscle cell potential involved in contraction⁵¹, indicating that these functions play a role in both thyroid cancer development and progression. Hence, translating EPITHYR GWAS data into biologically relevant pathways and gene sets expands our knowledge on the potential mechanisms underlying DTC carcinogenesis, and provides evidence for the future development of clinically relevant of multigenic predictors for identifying individuals at high risk. Further population, clinical and laboratory research

is needed to confirm our findings. Strategies to accelerate functional biological follow-up may include replication of the findings in other populations, fine-mapping, experimental studies such as metabolomic analyses to fully understand the biology and functional nature of the loci involved in signal transduction pathways associated with muscle contraction, glycogenolysis, and insulin metabolism in DTC susceptibility.

Materials and methods

Study participants. Study participants consisted of DTC cases and cancer-free controls of European descent originated from metropolitan France, New Caledonia, French Polynesia, Cuba and Gomel region of Belarus contaminated after the Chernobyl fallout, and who had been enrolled in one of the seven case-control studies from the EPITHYR consortium, with available blood or saliva DNA sample. After quality controls of the genotyping data (see next paragraph), 1551 cases and 1957 controls were used for all further analyses (Table 1). The study designs have been described in detail previously^{8,11–13,52–54}. All studies provided information on histology of the tumor, ethnicity, personal and familial history of thyroid disease, menstrual and reproductive factors, exogenous hormone use, weight, height, dietary habits and residential and occupational histories. DTC cases with missing histology and individuals related at the first, second and third degree according to their genotypic data (i.e. 19 individuals, data not shown) were excluded.

Participants from all studies provided written informed consent. The present study was performed in compliance with the Helsinki Declaration and to the reference methodology from the National Committees for personal data protection in medical research.

CATHY, YOUNG-Thyr, E3N and New Caledonian studies were approved by the French ethics committee “Comité de Protection des Personnes” and the French data protection authority “Commission Nationale de l’Informatique et des Libertés” (CNIL). The French Polynesian study was approved by the Ethical committee of French Polynesia and the CNIL. The Cuban study was approved by the Clinical Research Ethics Committee of the National Institute of Oncology, Havana, Cuba. The Chernobyl study was approved by the International Agency for Research on Cancer ethics committee and the Belarus Coordinating Council for Studies of the Medical Consequences of the Chernobyl Accident.

Genotyping data. All individuals were genotyped at the Centre National de Recherche en Génomique Humaine (CNRGH/CEA) with the Infinium OncoArray beadchip (Illumina) designed to target over 530,000 SNPs across the genome⁵⁵. For the purpose of EPITHYR studies, the beadchip was augmented with 13,759 SNPs known or suspected to be involved in DTC susceptibility or in thyroid hormone metabolism²⁴. Standard genotyping array QC steps were applied to filter out SNPs which were either duplicate SNPs (814 SNPs) pseudo autosomal SNPs (37 SNPs), monomorphic SNPs (5210 SNPs) or SNPs deviating from Hardy Weinberg Equilibrium (HWE), i.e. applying HWE p-value thresholds of 10^{-7} for controls and 10^{-12} for cases, as performed by the OncoArray consortium in other studies⁵⁵ (563 SNPs). In addition, SNPs with call rate per study < 95% (8327 SNPs) or showing cluster plot discordancy (4083 SNPs) were also discarded. This left 460,437 SNPs, of which 458,486 were located within or at ± 50 kb of a protein coding gene. In total, 3508 individuals with European descent (1551 cases and 1957 controls) as identified using ancestry markers and standard procedures described by the OncoArray consortium⁵⁵ were used for all further analyses.

SNP-level analysis. SNPs were tested individually with the assumption of an additive genetic model, using an unconditional logistic regression model adjusted for age (age at diagnosis for cases and age at inclusion for controls), sex, study and the first ten principal components to correct for population stratification. Analyses were performed with PLINK software v1.9⁵⁶.

Gene-level analysis. Gene-level analyses were performed using VEGAS2v02²⁹. As we described in another work, VEGAS2 “performs gene-based tests based on association test from single variant analyses and accounts for linkage disequilibrium (LD) between SNPs and number of SNPs tested to avoid an increase in false positive results due to genes with multiple, highly correlated markers”⁵⁷. Following the same strategy as what we reported previously, “we considered a SNP to belong to a gene if located within 50 kb on either side of the gene’s transcribed region, which we found to be a good balance between incorporating short-range regulatory variants while maintaining the specificity of the result for a specific gene, as variants associated with neighboring genes can influence the test statistic for a gene of interest”⁵⁷. All SNPs were provided to the tool which assigns SNPs to genes and calculates gene-based empirical association p-values. The results shown were obtained using EPITHYR European controls as reference dataset for LD calculation. For SNP annotation, the latest GENCODE28 definitions mapped to hg19 were downloaded (ftp://ftp.ebi.ac.uk/pub/databases/genCODE/Gencode_human/release_28/GRCh37_mapping/genCODE.v28lift37.annotation.gff3.gz). Only protein coding definitions ($N = 20,298$) were used for the gene-based association tests.

Multiple testing was taken into account by using the Benjamini and Hochberg’s procedure to compute the FDR, with a statistical significance threshold of 0.05. The same gene level analysis was repeated for after excluding 137 FTC cases, using all the same parameters.

Annotation of eQTLs. We used the PancanQTL database³⁰ (bioinfo.life.hust.edu.cn/PancanQTL/) to search for eQTL within or nearby genes associated with DTC risk in EPITHYR. This database provides access to eQTL-based analysis of genotype and expression data of 9196 tumor samples in 33 cancer types obtained from TCGA. For the present study, we downloaded the cis-eQTL identified in the 497 DTC samples analyzed in the PancanQTL project (<https://portal.gdc.cancer.gov/>).

Pathway-level analysis. Pathway-level analysis were performed using Vegas2Pathway⁵⁸, which accounts for LD between the tested markers, and corrects for gene and pathway sizes. This test uses the VEGAS2 output and external pathway definitions. Here we used the reference biological pathway annotation databases KEGG, considering the HSA and BRITE hierarchies^{31–33}, GO³⁵ and Reactome³⁴. The latest definitions for each database were downloaded. Number of definitions and number of OncoArray SNPs tagging genes involved in these definitions are provided in Table 3. The GO terms were subjected to filtering on basis of pathway size by only considering definitions with number of genes between 10 and 400. In addition, to further reduce the number of overlapping GO term definitions, a similarity measure (Jaccard index⁵⁹) was calculated for each pair of GO terms. Two terms were considered "highly similar" if their Jaccard index was > 0.85, in that case only the largest set was kept. Internally, VEGAS2Pathway only considers pathway definitions having a minimum of five genes. The statistical test used by VEGAS2Pathway is similar to the test used by VEGAS2 but considering a gene set as a pathway definition. For each pathway-based test, an FDR correction with a statistical significance threshold of 0.05 was applied to correct for multiple testing.

Gene network analysis. We used the EW_dmGWAS algorithm to investigate joined association signals beyond single markers³⁶. EW_dmGWAS first annotates sets of genes using PPI networks as described in the HINT database (HINTDB), which collates interactions from BioGRID, MINT, iRefWeb, DIP, IntActa, HPRD, MIPS and the PDB⁶⁰. Interactions are defined as either: *binary*, that is a direct biophysical interaction between two proteins, or *co-complex* associations, which means co-membership in a group, without implying direct pairwise interaction. These definitions of interactions, also called "co-complex categories" are divided into either literature-curated or deduced from high throughput experiments. Literature curated definitions include interaction data from thousands of small-scale studies focused at validating a single or a few specific hypotheses, while high throughput experiments produce large-scale interaction maps. Here we considered the high throughput definitions, for both binary ($N = 47,427$) and co-complex ($N = 102,807$) sets of interactions. EW_dmGWAS also uses data on condition-specific differential gene co-expression profiles to assign edge weights to the nodes of the PPI networks to prioritize gene sets (also called modules or subnetworks) for downstream gene enrichment analysis. Here we were interested in prioritizing genes that are differentially expressed in thyroid tumor tissue versus adjacent normal tissue samples. We used expression data from TCGA and selected tumor/normal tissue sample pairs for PTC cases available through the TCGA firebrowse portal (<http://firebrowse.org/>)⁶¹. Entire dataset in the file "illumina_hiseq_rnaseqv2-RSEM_genes_normalized" was downloaded, and the TCGA barcodes were used to find matching tumor and healthy tissue by parsing the 'sample' field (https://docs.gdc.cancer.gov/Encyclopedia/pages/TCGA_Barcode/). The sample field has values with range 01–09 for tumor types, and range 10–19 for normal types, using this criteria we found 59 matching tumor and normal sample pairs. In brief, EW_dmGWAS integrates GWAS signals and gene expression profiles to extract subnetworks from a background PPI network. Node weights are derived from GWAS signals and edge weights are derived from gene expression profiles. Modules are ranked according to their score which is a combination of node weight and edge weight.

EW_dmGWAS was executed for each set of binary and co-complex interactions listed in HINTDB using gene-level association test p-values. For each category of interactions, only the top 1% modules were considered for use in gene enrichment analysis. Reactome gene enrichment analysis was performed with the R package ReactomePA⁶² and KEGG and GO gene enrichment analyses were performed with R package clusterProfiler⁶³. Specifically, the functions *enrichPathway*, *enrichKEGG* and *enrichGO* were used. For GO annotations, a pre-processing step was necessary using the *simplify* function from clusterProfiler in order to remove redundant GO terms in the enrichment analysis.

Disclaimer. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Received: 18 January 2021; Accepted: 9 April 2021

Published online: 26 April 2021

References

- Lloyd, R. V., O. R., Kloppel, G. & Rosai, J. WHO classification of tumours of endocrine organs. *WHO Classification of Tumours*, 4th ed, Vol 10 (2017).
- Pacini, F. *et al.* European consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. *Eur. J. Endocrinol.* **154**, 787–803. <https://doi.org/10.1530/eje.1.02158> (2006).
- Ferlay, J. *et al.* Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **144**, 1941–1953. <https://doi.org/10.1002/ijc.31937> (2019).
- La Vecchia, C. *et al.* Thyroid cancer mortality and incidence: A global overview. *Int. J. Cancer* **136**, 2187–2195. <https://doi.org/10.1002/ijc.29251> (2015).
- Pellegriti, G., Frasca, F., Regalbuto, C., Squatrito, S. & Vigneri, R. Worldwide increasing incidence of thyroid cancer: Update on epidemiology and risk factors. *J. Cancer Epidemiol.* **2013**, 965212. <https://doi.org/10.1155/2013/965212> (2013).
- Veiga, L. H. *et al.* Thyroid cancer after childhood exposure to external radiation: An updated pooled analysis of 12 studies. *Radiat. Res.* **185**, 473–484. <https://doi.org/10.1667/RR14213.1> (2016).
- Ito, Y., Nikiforov, Y. E., Schlumberger, M. & Vigneri, R. Increasing incidence of thyroid cancer: Controversies explored. *Nat. Rev. Endocrinol.* **9**, 178–184. <https://doi.org/10.1038/nrendo.2012.257> (2013).

8. Guignard, R., Truong, T., Rougier, Y., Baron-Dubourdieu, D. & Guenel, 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.* **166**, 1140–1149. <https://doi.org/10.1093/aje/kwm204> (2007).
9. Brindel, P. *et al.* Anthropometric factors in differentiated thyroid cancer in French Polynesia: A case-control study. *Cancer Causes Control* **20**, 581–590. <https://doi.org/10.1007/s10552-008-9266-y> (2009).
10. Clero, E. *et al.* 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* **20**, 1285–1293. <https://doi.org/10.1089/thy.2009.0456> (2010).
11. Cordina-Duverger, E. *et al.* Hormonal and reproductive risk factors of papillary thyroid cancer: A population-based case-control study in France. *Cancer Epidemiol.* **48**, 78–84. <https://doi.org/10.1016/j.canep.2017.04.001> (2017).
12. Clavel-Chapelon, F., Guillas, G., Tondeur, L., Kernaléguen, C. & Boutron-Ruault, M. C. Risk of differentiated thyroid cancer in relation to adult weight, height and body shape over life: The French E3N cohort. *Int. J. Cancer* **126**, 2984–2990. <https://doi.org/10.1002/ijc.25066> (2010).
13. Khaard, C. *et al.* Anthropometric risk factors for differentiated thyroid cancer in young men and women from Eastern France: A case-control study. *Am. J. Epidemiol.* **182**, 202–214. <https://doi.org/10.1093/aje/kwv048> (2015).
14. Lence-Anta, J. J. *et al.* Environmental, lifestyle, and anthropometric risk factors for differentiated thyroid cancer in Cuba: A case-control study. *Eur. Thyroid J.* **3**, 189–196. <https://doi.org/10.1159/000362928> (2014).
15. Schmid, D., Ricci, C., Behrens, G. & Leitzmann, M. F. Adiposity and risk of thyroid cancer: A systematic review and meta-analysis. *Obes. Rev.* **16**, 1042–1054. <https://doi.org/10.1111/obr.12321> (2015).
16. Hemminki, K. & Li, X. Familial risk of cancer by site and histopathology. *Int. J. Cancer* **103**, 105–109. <https://doi.org/10.1002/ijc.10764> (2003).
17. Gudmundsson, J. *et al.* Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat. Genet.* **41**, 460–464. <https://doi.org/10.1038/ng.339> (2009).
18. Takahashi, M. *et al.* The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Hum. Mol. Genet.* **19**, 2516–2523. <https://doi.org/10.1093/hmg/ddq123> (2010).
19. Gudmundsson, J. *et al.* Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat. Genet.* **44**, 319–322. <https://doi.org/10.1038/ng.1046> (2012).
20. Kohler, A. *et al.* Genome-wide association study on differentiated thyroid cancer. *J. Clin. Endocrinol. Metab.* **98**, E1674–E1681. <https://doi.org/10.1210/jc.2013-1941> (2013).
21. Son, H. Y. *et al.* Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer. *Nat. Commun.* **8**, 15966. <https://doi.org/10.1038/ncomms15966> (2017).
22. Gudmundsson, J. *et al.* A genome-wide association study yields five novel thyroid cancer risk loci. *Nat. Commun.* **8**, 14517. <https://doi.org/10.1038/ncomms14517> (2017).
23. Figlioli, G. *et al.* Novel genetic variants in differentiated thyroid cancer and assessment of the cumulative risk. *Sci. Rep.* **5**, 8922. <https://doi.org/10.1038/srep08922> (2015).
24. Truong, T. *et al.* Multiethnic genome-wide association study of differentiated thyroid cancer in the EPITHYR consortium. *Int. J. Cancer* <https://doi.org/10.1002/ijc.33488> (2021).
25. Wang, K. *et al.* Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn disease. *Am. J. Hum. Genet.* **84**, 399–405. <https://doi.org/10.1016/j.ajhg.2009.01.026> (2009).
26. Cong, W. *et al.* Genome-wide network-based pathway analysis of CSF t-tau/Abeta1-42 ratio in the ADNI cohort. *BMC Genomics* **18**, 421. <https://doi.org/10.1186/s12864-017-3798-z> (2017).
27. Huang, Y. T., Liang, L., Moffatt, M. F., Cookson, W. O. & Lin, X. iGWAS: Integrative genome-wide association studies of genetic and genomic data for disease susceptibility using mediation analysis. *Genet. Epidemiol.* **39**, 347–356. <https://doi.org/10.1002/gepi.21905> (2015).
28. Stezhko, V. A. *et al.* A cohort study of thyroid cancer and other thyroid diseases after the Chernobyl accident: Objectives, design and methods. *Radiat. Res.* **161**, 481–492. <https://doi.org/10.1667/3148> (2004).
29. Mishra, A. & Macgregor, S. VEGAS2: Software for more flexible gene-based testing. *Twin Res. Hum. Genet.* **18**, 86–91. <https://doi.org/10.1017/thg.2014.79> (2015).
30. Gong, J. *et al.* PanCanQTL: Systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.* **46**, D971–D976. <https://doi.org/10.1093/nar/gkx861> (2018).
31. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
32. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
33. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551. <https://doi.org/10.1093/nar/gkaa970> (2021).
34. Fabregat, A. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **46**, D649–D655. <https://doi.org/10.1093/nar/gkx1132> (2018).
35. The Gene Ontology, C. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338. <https://doi.org/10.1093/nar/gkw1108> (2017).
36. Wang, Q., Yu, H., Zhao, Z. & Jia, P. EW_dmGWAS: Edge-weighted dense module search for genome-wide association studies and gene expression profiles. *Bioinformatics* **31**, 2591–2594. <https://doi.org/10.1093/bioinformatics/btv150> (2015).
37. Wei, W. J. *et al.* Clinical significance of papillary thyroid cancer risk loci identified by genome-wide association studies. *Cancer Genet.* **208**, 68–75. <https://doi.org/10.1016/j.cancergen.2015.01.004> (2015).
38. Coe, E. A. *et al.* The MITF-SOX10 regulated long non-coding RNA DIRC3 is a melanoma tumour suppressor. *PLoS Genet.* **15**, e1008501. <https://doi.org/10.1371/journal.pgen.1008501> (2019).
39. Montero-Conde, C. *et al.* Relief of feedback inhibition of HER3 transcription by RAF and MEK inhibitors attenuates their antitumor effects in BRAF-mutant thyroid carcinomas. *Cancer Discov.* **3**, 520–533. <https://doi.org/10.1158/2159-8290.CD-12-0531> (2013).
40. Morillo-Bernal, J., Fernandez, L. P. & Santisteban, P. FOXE1 regulates migration and invasion in thyroid cancer cells and targets ZEB1. *Endocr. Relat. Cancer* **27**, 137–151. <https://doi.org/10.1530/ERC-19-0156> (2020).
41. Xu, Y. & Lv, S. X. The effect of JAK2 knockout on inhibition of liver tumor growth by inducing apoptosis, autophagy and anti-proliferation via STATs and PI3K/AKT signaling pathways. *Biomed. Pharmacother.* **84**, 1202–1212. <https://doi.org/10.1016/j.biopha.2016.09.040> (2016).
42. Ohno, Y. *et al.* Downregulation of ANP32B exerts anti-apoptotic effects in hepatocellular carcinoma. *PLoS ONE* **12**, e0177343. <https://doi.org/10.1371/journal.pone.0177343> (2017).
43. Saenko, V. A. & Rogounovitch, T. I. Genetic polymorphism predisposing to differentiated thyroid cancer: A review of major findings of the genome-wide association studies. *Endocrinol. Metab.* **33**, 164–174. <https://doi.org/10.3803/EnM.2018.33.2.164> (2018).
44. Solus, J. F. & Kraft, S. Ras, Raf, and MAP kinase in melanoma. *Adv. Anat. Pathol.* **20**, 217–226. <https://doi.org/10.1097/PAP.0b013e3182976c94> (2013).
45. Sharifi, N. & Auchus, R. J. Steroid biosynthesis and prostate cancer. *Steroids* **77**, 719–726. <https://doi.org/10.1016/j.steroids.2012.03.015> (2012).

46. Li, S., Sun, Y. & Gao, D. Role of the nervous system in cancer metastasis. *Oncol. Lett.* **5**, 1101–1111. <https://doi.org/10.3892/ol.2013.1168> (2013).
47. Chen, J. *et al.* Genetic regulatory subnetworks and key regulating genes in rat hippocampus perturbed by prenatal malnutrition: Implications for major brain disorders. *Aging* **12**, 8434–8458. <https://doi.org/10.18632/aging.103150> (2020).
48. Li, H. *et al.* Co-expression network analysis identified hub genes critical to triglyceride and free fatty acid metabolism as key regulators of age-related vascular dysfunction in mice. *Aging* **11**, 7620–7638. <https://doi.org/10.18632/aging.102275> (2019).
49. Liu, M. *et al.* A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *Neuroimage* **208**, 116459. <https://doi.org/10.1016/j.neuroimage.2019.116459> (2020).
50. Lauby-Secretan, B. *et al.* Body fatness and cancer-viewpoint of the IARC Working Group. *N. Engl. J. Med.* **375**, 794–798. <https://doi.org/10.1056/NEJMsrl606602> (2016).
51. Zhang, B., Chen, Z., Wang, Y., Fan, G. & He, X. Integrated bioinformatics analysis for the identification of key genes and signaling pathways in thyroid carcinoma. *Exp. Ther. Med.* **21**, 298. <https://doi.org/10.3892/etm.2021.9729> (2021).
52. Cardis, E. *et al.* Risk of cancer after low doses of ionising radiation: Retrospective cohort study in 15 countries. *BMJ* **331**, 77. <https://doi.org/10.1136/bmj.38499.599861.E0> (2005).
53. Pereda, C. M. *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.* **16**, 22. <https://doi.org/10.1186/s12863-015-0180-5> (2015).
54. Maillard, S. *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* **10**, e0123700. <https://doi.org/10.1371/journal.pone.0123700> (2015).
55. Amos, C. I. *et al.* The OncoArray Consortium: A network for understanding the genetic architecture of common cancers. *Cancer Epidemiol. Biomarkers Prev.* **26**, 126–135. <https://doi.org/10.1158/1055-9965.EPI-16-0106> (2017).
56. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575. <https://doi.org/10.1086/519795> (2007).
57. Lonjou, C. *et al.* Gene- and pathway-level analyses of iCOGS variants highlight novel signaling pathways underlying familial breast cancer susceptibility. *Int. J. Cancer*. <https://doi.org/10.1002/ijc.33457> (2020).
58. Mishra, A. & MacGregor, S. A novel approach for pathway analysis of GWAS data highlights role of BMP signaling and muscle cell differentiation in colorectal cancer susceptibility. *Twin Res. Hum. Genet.* **20**, 1–9. <https://doi.org/10.1017/thg.2016.100> (2017).
59. Jaccard, P. The distribution of the flora in the Alpine Zone. 1. *New Phytol.* **11**, 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x> (1912).
60. Das, J. & Yu, H. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst. Biol.* **6**, 92. <https://doi.org/10.1186/1752-0509-6-92> (2012).
61. Cancer Genome Atlas Research, N. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690. <https://doi.org/10.1016/j.cell.2014.09.050> (2014).
62. Yu, G. & He, Q. Y. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479. <https://doi.org/10.1039/c5mb00663e> (2016).
63. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).

Acknowledgements

We are most grateful to all the subjects who participated in the EPITHYR studies. We wish to thank Héctor Climente-González for helpful discussions on the specificity of the various databases and technical help for the use of VEGAS.

Author contributions

Project coordination: F.L., T.T. Quality controls and data management were done by P.E.S., J.G., C.L., C.R., C.X. Genotyping was supervised and carried out by: A.B.A., D.B.D., R.O. and J.F.D. Bioinformatics and statistical analyses were performed by O.K., P.E.S. under the supervision of F.L. and T.T. Subjects recruitment, biological material collection and handling were organized and carried out by: C.R., C.X., C.R.V.S., F.R., J.J.L.A., R.M.O., P.L.P., C.M., A.V.G., C.S., M.C.B.R., E.O., A.K., P.G., F.D.V. O.K., F.L., P.E.S. and T.T. drafted the manuscript. All authors reviewed the manuscript and approved the final version of the paper.

Funding

This work was supported by the Institut National du Cancer (INCa grant 9533), the Fondation ARC pour la Recherche sur le Cancer (ARC grant PGA120150202302), the Centre National de Recherche en Génomique Humaine, CEA, Electricité de France (conseil scientifique de Radioprotection d'EDF, grant EP 2019-01), Fondation de France (Grant 2016-70074) and Plan Cancer (Project ThyGenRad, Grant ENV201415). J.G. was the recipient of a PhD fellowship from Région Ile-de-France. The E3N cohort received support from the MGEN, Gustave Roussy and Ligue contre le cancer for its set up and its maintenance. The cohort was also supported by a state grant ANR (grant number ANR-10-COHO-0006) from the Agence Nationale pour la Recherche (ANR) within the Investissement d'Avenir program and from Ministère de l'enseignement supérieur, de la recherche et de l'innovation (MESRI, grant number 2102 918823).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-88253-0>.

Correspondence and requests for materials should be addressed to F.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Role of DNA Repair Variants and Diagnostic Radiology Exams in Differentiated Thyroid Cancer Risk: A Pooled Analysis of Two Case-Control Studies



Monia Zidane^{1,2,3}, Thérèse Truong^{1,2,3}, Fabienne Lesueur⁴, Constance Xhaard^{5,6}, Emilie Cordina-Duverger^{1,2,3}, Anne Boland⁷, Hélène Blanché^{8,9}, Catherine Ory¹⁰, Sylvie Chevillard¹⁰, Jean-François Deleuze^{7,8,9}, Vincent Soucard^{1,2,3}, Yan Ren^{1,2,3}, Mohammed Zakarya Zemmach^{1,2,3}, Sandra Canale², Françoise Borson-Chazot¹¹, Claire Schvartz¹², Eugènia Mariné Barjoan¹³, Anne-Valérie Guizard^{14,15}, Pierre Laurent-Puig¹⁶, Claire Mulot¹⁶, Julie Guibon^{1,2,3,4}, Mojgan Karimi^{1,2,3}, Martin Schlumberger^{2,3}, Elizabeth Adjadj¹, Carole Rubino^{1,2,3}, Pascal Guenel^{1,2,3}, Jean-Baptiste Cazier¹⁷, and Florent de Vathaire^{1,2,3}

ABSTRACT

Background: Given the increased use and diversity of diagnostic procedures, it is important to understand genetic susceptibility to radiation-induced thyroid cancer.

Methods: On the basis of self-declared diagnostic radiology examination records in addition to existing literature, we estimated the radiation dose delivered to the thyroid gland from diagnostic procedures during childhood and adulthood in two case-control studies conducted in France. A total of 1,071 differentiated thyroid cancer (DTC) cases and 1,188 controls from the combined studies were genotyped using a custom-made Illumina OncoArray DNA chip. We focused our analysis on variants in genes involved in DNA damage response and repair pathways, representing a total of 5,817 SNPs in 571 genes. We estimated the OR per milli-Gray (OR/mGy) of the radiation dose delivered to the thyroid gland using condi-

tional logistic regression. We then used an unconditional logistic regression model to assess the association between DNA repair gene variants and DTC risk. We performed a meta-analysis of the two studies.

Results: The OR/mGy was 1.02 (95% confidence interval, 1.00–1.03). We found significant associations between DTC and rs7164173 in CHD2 ($P = 5.79 \times 10^{-5}$), rs6067822 in NFATc2 ($P = 9.26 \times 10^{-5}$), rs1059394 and rs699517 both in ENOSF1/THYS, rs12702628 in RPA3, and an interaction between rs7068306 in MGMT and thyroid radiation doses ($P = 3.40 \times 10^{-4}$).

Conclusions: Our results suggest a role for variants in CDH2, NFATc2, ENOSF1/THYS, RPA3, and MGMT in DTC risk.

Impact: CDH2, NFATc2, ENOSF1/THYS, and RPA3 have not previously been shown to be associated with DTC risk.

¹INSERM, Centre for Research in Epidemiology and Population Health (CESP), 94800 Villejuif, France. ²Université Paris-Sud Orsay, Villejuif, France. ³Gustave Roussy, Villejuif, France. ⁴Inserm, U900, Institut Curie, Université PSL, Mines ParisTech, Paris, France. ⁵INSERM Centre d'Investigation Clinique CIC-P 1433, CHRU Nancy, France. ⁶INSERM U1116, FCRIN INI-CRCT, Lorraine Université, Nancy, France. ⁷Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, Evry, France. ⁸Fondation Jean Dausset-CEPH (Centre Etude du Polymorphisme Humain), Paris, France. ⁹Laboratory of Excellence GENMED (Medical Genomics). ¹⁰CEA, Direction de la Recherche Fondamentale, Institut de Biologie François Jacob, IRCM, SREIT, Laboratoire de Cancérologie Expérimentale (LCE), Université Paris-Saclay, Fontenay-aux-Roses, France. ¹¹Rhône-Alpes Thyroid Cancer Registry, Cancer Research Center of Lyon (UMR INSERM 1052, CNRS 5286), RTH Laennec Faculty of Medicine, University of Lyon, Lyon. ¹²Thyroid Cancer Registry of Champagne-Ardenne, Institut Jean Godinot, Reims, Cancer Registry of Isère, Meylan. ¹³Public Health Department, University Hospital Nice, Nice. ¹⁴Registre Général des Tumeurs du Calvados, Centre François Baclesse, Caen, France. ¹⁵U1086 INSERM-UCN "ANTICIPE," Caen, France. ¹⁶Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université Paris Descartes, Université Paris Diderot, EPIGENETEC, Paris, France. ¹⁷Institute of Cancer and Genomic Sciences, Centre for Computational Biology, University of Birmingham, Birmingham, UK.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Florent de Vathaire, UMR 1018, INSERM, 114 rue Edouard Vaillant, Villejuif 94805, France. E-mail: florent.devathaire@gustaveroussy.fr

Cancer Epidemiol Biomarkers Prev 2021;30:1208–17

doi: 10.1158/1055-9965.EPI-20-1142

©2021 American Association for Cancer Research.

Introduction

Differentiated thyroid cancer (DTC) is the most common cancer of the endocrine system (1). The registered incidence of DTC has recently been increasing, especially in developed countries mostly due to increased screening (2). Exposure to thyroid ionizing radiation (IR) during childhood was the first risk factor linked with DTC, an association that has been well characterized over several decades (3–8).

In a pooled study performed on doses higher than ≥ 1 Gy (9, 10), the excess of DTC relative risk per Gray (ERR/Gy) delivered to the thyroid gland during childhood ranged between 5 and 10. When restricting the pooled analysis to doses lower than 0.2 Gy, however, the ERR/Gy was slightly higher (11). A higher magnitude of excess risk for low-dose exposure during childhood was also observed in a cohort study of Japanese atomic bomb survivors (12). Nevertheless, some uncertainties remain about how the recent increase in the use of diagnostic procedures involving irradiation, such as CT, might affect DTC incidence (13, 14).

Despite being more subject to recall bias than cohort studies, case-control studies may be more helpful for improving understanding of thyroid low-dose radiation risk because case-control studies utilize a high number of cases and permit a more detailed investigation of a high number of potential risk cofactors, which are costlier to investigate in cohort studies. Because DTC risk after low radiation dose exposure is also likely modulated by individual susceptibility, investigating the association between DNA variants and DTC risk after low radiation dose could also facilitate an improved understanding of this risk. IR

exposure causes DNA damage as well as other physical and chemical biomolecule alterations. DNA double-strand breaks (DSB) are a characteristic cellular response to IR exposure (15). Exposure to low-dose IR (from levels 50 mGy and lower) from X-ray examinations has been demonstrated to increase the risk of chromosomal translocation (16), suggesting the possibility of long-term adverse health effects.

DSBs trigger the DNA repair system, but the effect of these DSBs varies with cumulative exposures and individual susceptibility. To our knowledge, only one published study has investigated an interaction between genetic background and the radiation-related DTC risk after diagnostic radiation exposure (17). This study identified 24 SNPs with a *P* value of variant-RI doses interaction was <0.05 for all types of DTC cases (papillary and follicular). Currently, few genes including *FOXO1*, *ATM*, and *LIG1* have been associated with DTC risk after IR exposure, but most published studies of these genes were conducted in highly irradiated populations. Moreover, these identified genes were associated with sporadic DTC (18).

Given the increasing use of diagnostic radiology exams and the proliferation of radiology procedure types, it is important to explore the side effects of multiple low radiation exposures, such as potential effects on DTC development, especially for patients exposed at a young age.

In this study, we investigated the association between diagnostic low-dose IR exposures and DTC risk. We quantified the doses delivered to the thyroid gland during childhood and adulthood in two French case-control studies. We also investigated the role of genetic variants located in 604 DNA damage signaling and repair genes in modifying the effect of exposure to diagnostic procedure radiation on DTC risk.

Materials and Methods

Case-control studies

Our study included cases and controls from two studies conducted in distinct regions from metropolitan France.

The CATHY study (19) was a case-control study conducted in Marne, Ardennes, and Calvados, three French administrative areas covered by a cancer registry. Eligible cases were all patients with DTC ages 25 years and older that were diagnosed between 2002 and 2007 and resided in these areas. Controls were recruited by selecting and contacting telephone numbers at random, then frequency-matching the individuals to patients by sex and 5-year age groups; controls were randomly selected among healthy individuals residing in the same study areas at the time of the matched cases diagnoses. From the 621 cases and 706 controls initially included in the CATHY study, saliva DNA samples (Oragene DNA Testing Kit) were available for 583 cases and 643 controls.

The Young-Thyr study (20) is a case-control study conducted in eastern regions of France: Alsace, Champagne-Ardenne, Corse, Franche-Comté, Lorraine, Rhône-Alpes, and Provence-Alpes-Côte d'Azur. Eligible cases were all patients with DTC diagnosed with DTC between January 1, 2002, and December 31, 2006 that were younger than 35 years of age with their main residence in a region of eastern France. Controls were selected from the general population and individually matched to a single patient of the same sex, year of birth (within 1 year), and region of residence during the year when the patient was diagnosed with cancer. From the 805 cases and the 876 controls included in the Young-Thyr study, DNA from saliva (Oragene) was available for 715 cases and 692 controls.

Participants from both studies answered similar questionnaires that were collaboratively developed. Information on a personal and familial history of thyroid disease, hormonal and reproductive factors, weight, height, dietary habits, tobacco smoking, and medical history was collected during in-person interviews. Participants from both studies gave their informed written consent before being interviewed by trained interviewers. Both studies were approved by the Ethics Committee and the French National Commission for Data Protection (CNIL).

Clinical and epidemiological data from both studies were harmonized and aggregated in one database. During this process, 30 subjects were excluded because they were duplicates in the two studies.

Radiologic procedure history

In the questionnaire, participants were asked to report their history of diagnostic radiology examinations by providing the year and the reason for each examination. The exams considered were conventional radiography, CT scans, and nuclear medicine. Examinations linked to investigating thyroid pathology were excluded and were not considered in the dose estimation or any analysis. Examinations carried out during childhood (before the age of 20 years) and adulthood (over 20 years) were analyzed separately. All procedures reported by participants were classified by period (before 1980, 1980–1989, 1990, and after) to take into account the development of radiologic procedure technology over this time.

Thyroid radiation dose estimates

To estimate the absorbed radiation doses administered to the thyroid gland for different types of diagnostic radiology examinations, we reviewed papers published up to July 2018 that included absorbed dose estimates to the thyroid. From 1970 to 2002, only one study provided estimates of absorbed doses to organs, including the thyroid gland (21), and was used to calculate dose estimates from conventional radiography.

If sex-specific data were available, a mean dose was used. Because the available data about childhood exams were very limited, specific doses were available only for children under the age of 2 years old without consideration of the period of exposure. Children older than 2 years were assumed to receive the same doses as adults. All doses per exam that were equal to or less than 0.01 mGy were considered to be null because of the magnitude of daily natural irradiation in France. **Table 1** summarizes the different values used per diagnostic procedure type (conventional radiography, CT scan, nuclear medicine), per period (before 1980, 1980–1989, 1990, and after), and per age at examination (≤ 2 years old, >2 years).

Genotyping and quality control

DNA was extracted from saliva samples using a semiautomated salt precipitation method. All subjects were genotyped using the Infinium OncoArray-500K BeadChip (Illumina). This array has been previously described in detail (22) and contains 499,170 SNPs. We incorporated 13,759 additional custom markers based on prior evidence of association in genes involved in relevant biological pathways, such as thyroid function.

The quality control process was applied to the data of each study separately. A threshold of 5% was applied for a missing call rate per SNP and individual. The Hardy-Weinberg equilibrium (HWE) was also assessed per SNP among controls, using an exact test with a *P* value threshold at 10^{-5} . Only SNPs with a minor allele frequency (MAF) above 5% were included in the analyses.

Table 1. References used for thyroid radiation dose estimates per diagnostic procedure type.

Exams types	Childhood exposure age < 2	Doses in mGy			References
		Childhood exposure age > 2 years and adulthood exams			
		Before 1980	1980–1989	1990 and after	
Conventional Radiography					
Skull	– ^a	3.6	3.6	0.42	Chang et al. 2017
Dental (panoramic)	– ^a	0.56	0.07	0.02	Chang et al. 2017
Paranasal sinuses	– ^a	0.02	0.04	0.04	Chang et al. 2017
Intravenous urography	0.38 ^b	0.38	0.21	0.21	Inskip et al. 1995
Chest	0.13	0.16	0.31	0.19	Almén and Mattsson 1995/Chang et al. 2017
Spine	2.8 ^b	2.8	4.5	0.99	Chang et al. 2017
CT Scan					
Skull	3	0.43	0.49	0.64	Lee et al. 2007/Chang et al. 2017
Chest	5.9	15	20.5	22	Mabille et al. 2008/Chang et al. 2017
Spine	13.5	17	24	25.5	Mabille et al. 2008/Chang et al. 2017
Abdomen	0.17	0.28	0.38	0.41	Mabille et al. 2008/Chang et al. 2017
Nuclear medicine					
Bone scan	1.1 ^b	1.1	0.96	1.1	Chang et al. 2017
Brain scan	– ^a	1.3	3.6	9	Chang et al. 2017
Lung scan (perfusion/ventilation)	– ^a	0.2	0.25	0.25	Chang et al. 2017
Heart scan	– ^a	11	17	12	Chang et al. 2017
Liver scan	– ^a	0.15	0.15	0.15	Chang et al. 2017
Renal scan	– ^a	0.17	0.31	0.17	Chang et al. 2017

^aNo exposed subject at this age.^bNo available data for children, adult data used.

The FastPop R package was used to infer intercontinental ancestry with a principal component-derived method (23). Only individuals identified as European were kept in the genetic analyses. A total of 443 cases and 532 controls from CATHY and 629 cases and 656 controls from Young-Thyr were included in the analysis of SNPs.

SNPs of DNA damage signaling and repair genes according to the Gene Ontology database (24, 25) were examined. We initially selected exonic and intronic variants in 604 genes. After the quality control steps, a total of 5,817 SNPs in 571 genes were retained for the analysis (all genes are listed in the Supplementary Table S1).

Statistical methods

We used a conditional logistic regression model to estimate the OR of DTC per mGy (OR/mGy) of radiation dose received by the thyroid gland. Because in our study, thyroid doses were very low and cell killing not likely to play signification, and because our goal was not to define the best model for thyroid cancer risk after low dose but rather the best model to investigate the interaction between thyroid radiation and genetic factors, we only investigate models with one parameter for the thyroid radiation dose. In this way, we compared a linear model [$OR = Cst(1 + \alpha \text{ dose})$] to a “constant” model without a radiation dose ($OR = Cst$) and other models with one parameter, such as quadratic ($OR = 1 + \alpha \text{ dose}^2$) and exponential [$OR = 1 + e(\alpha \text{ dose})$], rather to traditional dose–response models used in radiation epidemiology (26). These comparisons were done using Akaike Information Criterion (AIC; ref. 27). Similarly, we investigated the potential role of a dose–response modifier by introducing a term for interaction in the models. The analysis of the radiation dose was performed by using the PECAN module of Epiwin software. To assess the association between SNPs and DTC risk, ORs and 95% confidence intervals (95% CI) were derived from unconditional logistic regression models, assuming a log-additive model because the stratum was not considered in this step.

Analyses were first conducted in the Young-Thyr study to filter our SNPs and limit the number of tests to be performed later; only the 100 SNPs with the lowest *P* values from this step were tested in the CATHY study. We also conducted a meta-analysis combining results from both studies; overall *z*-statistics and *P* value for these 100 top SNPs were then calculated from a weighted sum of the individual statistics. Weights were proportional to the square root of the number of individuals examined in each sample and selected such that the squared weights summed to 1.0, as implemented in the METAL program (28). This step aims to improve signal detection in the genetic analysis by avoiding a large number of tests, and to substantiate homologous signals in both studies since our population is relatively limited.

To correct for multiple tests in the main effect analysis, the FDR procedure (29) was calculated for each result of the meta-analysis. We investigated the interaction between SNPs and radiation doses using the likelihood ratio test, comparing models with and without the interaction term following the same steps described above.

All analyses were adjusted for age (5-year period), sex, thyroid cancer family history (yes/no), goiter history (yes/no), radiotherapy history (yes/no), BMI (continuous), height (continuous), educational level (primary school, secondary school, bachelors or higher), smoking status (never/former smoker/smoker), number of pregnancies longer than seven months in women (three classes), and study (CATHY/Young-Thyr) only in the pooled analysis. Analyses of genetic variants were additionally adjusted for the first three principal components to take into account the population stratification. Subgroup analysis was also performed separately for papillary histologic type cases and according to tumor size (microcarcinoma and macrocarcinoma).

The log-linear model poses fewer convergence problems than the linear model, for that reason the association and interaction tests between radiation doses and DNA variants were performed using a log-linear model. Since we did not apply any cut-off *P* values for SNPs

selected to be included in the meta-analysis step, in this study we chose to only show our best five results from each analysis.

Association analyses and meta-analyses were performed using R 3.6 and METAL (28). SNPs were annotated using the VEP tool (30), and gene regions were plotted with LocusZoom software (31).

Results

Demographic characteristics of the pooled study population (Young-Thyr and CATHY) are described in Table 2. More than 75% of our population were women and nonsmokers. Papillary thyroid cancer type was the most frequent type of thyroid cancer observed, with a mean age at diagnosis younger than 40 years old. However, the mean age at diagnosis was lower in the Young-Thyr study than in the CATHY study, as shown in Supplementary Table S2 (27 years vs. 51 years).

Diagnostic radiation and thyroid cancer risk

Thyroid doses determined by diagnostic radiology examination estimates

Estimated IR doses administered to the thyroid during childhood were lower than those given after 20 years of age. Among children, the most frequent exam was panoramic dental radiography, which was reported in 971 individuals. Nuclear medicine exams were rarely reported among children; the most frequent exam scan was bone scans reported in 14 children (Table 3). During adulthood, the most frequent conventional radiography exam was chest radiography, which was seen in 30.8% of adults. Skull CT scan was the most common CT exam during adulthood. Nuclear medicine exams were also relatively rarely reported among adults (Table 3). The median thyroid doses delivered during childhood were lower than doses administered during adulthood. The maximum childhood estimated thyroid dose was 194.2 mGy among the whole population, the mean among cases was 2.8 and 2.0 mGy among controls. A total of 1,672 subjects reported not having received any thyroid-irradiating diagnostic procedure during childhood. The estimated maximum dose delivered to the thyroid during adulthood in a diagnostic radiology procedure was 260.0 mGy, and the mean dose was 5.6 and 5.5 mGy among cases and controls, respectively (Supplementary Tables S3 and S4). A total of 1,466 subjects did not report experiencing any thyroid-irradiating diagnostic procedures during adulthood. Detailed information on diagnostic radiology exams and thyroid doses in cases and controls in each study are shown in Supplementary Tables S2, S3, and S4. Characteristics and thyroid doses among genotyped subgroups from both studies are presented in Table 2. The DTC risk in relation to the radiation doses in this population is detailed in Supplementary Table S5.

Differentiated thyroid cancer risk in relation to thyroid radiation doses

Only estimated radiation dose received by the thyroid gland during childhood was found to be associated with an increased DTC risk, unlike doses received during adulthood. In the linear model, the DTC OR per mGy administered to the thyroid was equal to 1.02 (95% CI, 1.00–1.03). When restricting the analysis to papillary thyroid carcinoma, the OR per mGy was very similar. The fit of the model was improved by the addition of a quadratic term to take into account a possible higher effect of high doses, although not significant ($P = 0.06$). No significant difference ($P > 0.05$) was found between the dose responses for men and for women. No interaction was apparent

between the thyroid radiation dose and obese status, age at DTC occurrence, study, number of pregnancies, female sex, smoking status, height, or the personal radiotherapy history. In a log-linear model, the OR for a dose of 1 mGy was 1.01 (95% CI, 1.01–1.04; Table 4). Results from a subanalysis by tumor size, shown in Table 4, yielded similar DTC risk in relation to thyroid IR doses to these from the whole population analysis.

Differentiated thyroid cancer risk in relation to DNA damage repair and signaling gene variants

Association analysis between DTC risk and the 5,817 variants in DNA damage response genes from the Young-Thyr study including all histologic types yielded only two SNPs with a P value lower than 5×10^{-4} : rs4962347 (chromosome 10, *C10orf90*) and rs16983787 (chromosome 10, *SMC6*). None of the top 100 variants with the lowest P values reached the threshold P value of 5×10^{-4} in the CATHY study. In the meta-analysis, results of the association test with DTC risk showed a significantly reduced risk for carriers of the minor allele [T]; both SNPs, rs6067822 in *NFATc2* and rs7164173 in *CHD2*, had a P value of 5.79×10^{-5} and 9.26×10^{-5} respectively, under an additive model of inheritance. Associations with both variants were in the same direction in CATHY and Young-Thyr (Table 5A). Other variants in *NFATc2* present among the top 100 SNPs associated with DTC in the Young-Thyr study are described in Supplementary Fig. S1. Rs7164173 was the only SNP in *CHD2* among the top 100 SNPs. Rs105939 and rs699517 in *ENOSF1/THYS* were in our best results, but statistical significance was not achieved ($\text{OR}_{\text{meta-analysis}} = 1.30$, $P_{\text{value meta-analysis}} = 7.81 \times 10^{-4}$). Only one variant in *RPA3* (rs12702628 in 7p21.3) was found to be significantly associated with DTC risk.

When the analysis was restricted to only papillary carcinoma histologic types, only rs7164173 was significant ($P_{\text{value meta-analysis}} = 9.33 \times 10^{-5}$; Table 5B). In the *NFATc2* region, two SNPs, rs77971457 and rs6067822, had respective $\text{OR}_{\text{meta-analysis}}$ values of 1.45 and 0.77 but did not reach the significance threshold. rs1290997 and rs1950764 in *RAD51B* had $\text{OR}_{\text{meta-analysis}}$ values of 0.74 and 0.60, respectively (Table 5B). The polymorphism rs1950764 had a $P_{\text{value meta-analysis}} > 5 \times 10^{-4}$. Detailed results for SNPs in *RAD51B* are shown in Supplementary Fig. S2. Results from the best SNPs in the macrocarcinoma analysis (Table 5C) yielded two SNPs in the *MLK7-AS1* gene (rs72625242, rs16861406) in 2q31.1 in addition to two other SNPs in chromosome 2 (*PMSD1* and *SMC6*) and the *C10orf90* gene in chromosome 10. In the microcarcinoma analysis, a significant result in 2q31.1 was found (rs75361806, *TLK1*; Table 5D). The best results from this later analysis also yielded three SNPs in the *MACROD2* gene (20p12.1).

Differentiated thyroid cancer risk related to the interaction between thyroid doses and variants of DNA damage repair and signaling genes

We investigated the interaction between variants in DNA damage response genes and thyroid exposure to radiation in DTC risk. In the meta-analysis, the SNP-radiation interaction term yielded a $P_{\text{value meta-analysis}} = 3.4 \times 10^{-4}$ and an $\text{OR}_{\text{interaction}} = 0.65$ for rs7068306 in *MGMT* (Table 6). A second *MGMT* SNP, rs7087131, was also among the top 100 SNPs identified in the Young-Thyr study. Results on the modifying effect of rs7068306 genotypes on DTC risk according to diagnostic radiation exposure study separately in CATHY and Young-Thyr are presented in Supplementary Table S6.

Zidane et al.

Table 2. Clinical and demographic description of the population included in DTC risk in relation to thyroid radiation doses analysis and genetic analysis.

	Whole population		The population included in the genetic analysis	
	Cases (n = 1,393)	Controls (n = 1,580)	Cases (n = 1,072)	Controls (n = 1,188)
Gender (%)				
Men	300 (21.5)	398 (25.2)	228 (21.3)	302 (25.4)
Women	1,093 (78.5)	1,182 (74.8)	844 (78.7)	886 (74.6)
Age at diagnosis/reference year (years)				
Mean	37.58	37.21	37.4 (15.0)	37.6 (15.2)
Median [Min, Max]	32.0 [9.00, 83.0]	32.0 [9.00, 83.0]	32.0 [11.0, 83.0]	32.0 [9.00, 80.0]
Smoking status (%)				
Nonsmoker	739 (53.1)	768 (48.6)	232 (52.4)	271 (50.9)
Former smoker	304 (21.8)	306 (19.4)	140 (31.6)	149 (28.0)
Smoker	350 (25.1)	506 (32.0)	71 (16.0)	112 (21.1)
Alcohol consumption (number of alcohol glasses per week, %)				
No consumption	778 (55.9)	763 (48.3)	582 (54.3)	559 (47.1)
Less than 10	504 (36.2)	673 (42.6)	403 (37.6)	521 (43.9)
More than 10	110 (7.9)	144 (9.1)	87 (8.1)	108 (9.1)
BMI (kg/m ²)				
Mean (SD)	24.6 (5.21)	24.1 (4.88)	24.7 (5.36)	24.2 (4.89)
Median [Min, Max]	23.5 [11.0, 51.8]	23.1 [13.4, 49.1]	23.6 [11.0, 51.8]	23.1 [13.4, 48.7]
Goitre history (%)				
No	899 (64.5)	1,468 (92.9)	695 (64.8)	1,102 (92.8)
Yes	494 (35.5)	112 (7.1)	377 (35.2)	86 (7.2)
Thyroid cancer family history (%)				
No	1,542 (97.6)	1,322 (94.9)	1,023 (95.4)	1,158 (97.5)
Yes	38 (2.4)	71 (5.1)	49 (4.6)	30 (2.5)
Radiotherapy history (%)				
No	1,558 (98.6)	1,354 (97.2)	1,043 (97.3)	1,168 (98.3)
Yes	29 (2.8)	22 (1.3)	19 (1.7)	20 (1.6)
Educational level (%)				
Bachelor's degree or higher	765 (54.9)	982 (62.2)	618 (57.6)	743 (62.5)
Secondary school	418 (30.0)	406 (25.7)	315 (29.4)	311 (26.2)
Primary school level	210 (15.1)	192 (12.2)	139 (13.0)	134 (11.3)
Childhood thyroid radiation doses in mGy (%)				
0	1,021 (73.3)	1,161 (73.5)	777 (72.5)	876 (73.7)
[0,1]	112 (8.0)	147 (9.3)	89 (8.3)	104 (8.8)
[1,5]	139 (10.0)	150 (9.5)	116 (10.8)	113 (9.5)
[5,10]	25 (1.8)	45 (2.8)	18 (1.7)	36 (3.0)
[10,20]	49 (3.5)	57 (3.6)	35 (3.3)	42 (3.5)
>20	46 (3.3)	20 (1.3)	37 (3.5)	17 (1.4)
Mean (SD)	2.79 (10.3)	1.98 (6.18)	2.89 (11.1)	2.11 (6.62)
Median [Min, Max]	0 [0, 194]	0 [0, 80.9]	0 [0, 194]	0 [0, 80.9]
Adulthood thyroid radiation doses in mGy (%)				
0	669 (48.1)	797 (50.4)	523 (48.8)	591 (49.7)
[0,1]	247 (17.7)	283 (17.9)	187 (17.4)	220 (18.5)
[1,5]	245 (17.6)	262 (16.6)	182 (17.0)	193 (16.2)
[5,10]	59 (4.2)	61 (3.9)	46 (4.3)	46 (3.9)
[10,20]	53 (3.8)	43 (2.7)	44 (4.1)	35 (2.9)
>20	119 (8.5)	134 (8.5)	90 (8.4)	103 (8.7)
Mean (SD)	5.63 (17.3)	5.50 (17.9)	5.28 (16.6)	5.54 (17.2)
Median [Min, Max]	0 [0, 245]	0.02 [0, 262]	0.02 [0, 245]	0.02 [0, 157]
Number of children (%)				
0 (and men)	666 (47.8)	829 (52.5)	514 (47.9)	622 (52.4)
1-3	648 (46.6)	529 (43.8)	542 (50.6)	542 (45.6)
4-6	67 (4.8)	44 (3.4)	10 (0.9)	20 (1.7)
>6	11 (0.4)	4 (0.3)	6 (0.6)	4 (0.3)
Study				
Young-Thyr (%)	805 (57.8)	876 (55.4)	629 (58.7)	656 (55.2)
Cathy (%)	587 (42.2)	704 (44.6)	443 (41.3)	532 (44.8)
Thyroid cancer histology (%)				
Follicular	140 (10.1)		109 (10.2)	
Papillary	1243 (89.2)		955 (89.1)	
Missing	10 (0.7)		8 (0.7)	

(Continued on the following page)

Table 2. Clinical and demographic description of the population included in DTC risk in relation to thyroid radiation doses analysis and genetic analysis. (Cont'd)

	Whole population		The population included in the genetic analysis	
	Cases (n = 1,393)	Controls (n = 1,580)	Cases (n = 1,072)	Controls (n = 1,188)
Thyroid cancer size in mm (%)				
<10	431 (31.0)		341 (31.8)	
>10	957 (68.8)		729 (68.0)	
Missing	4 (0.28)		2 (0.2)	

Discussion

Our main objective was to investigate the potential genetic susceptibility to DTC after exposure to low IR doses from diagnostic procedures performed during childhood. In our study, the estimated OR/mGy for thyroid radiation doses received during childhood [OR/mGy = 1.02; 95% CI, 1.00–1.03] was similar to the OR/mGy values of the pooled DTC study by Veiga and colleagues (10) when considering analyses restricted to thyroid radiation doses lower than 100 mGy (OR/mGy = 1.010; 95% CI, 1.004–1.018) or to thyroid radiation doses lower than 200 mGy (OR/mGy = 1.011; (95% CI, 1.007–1.02; ref. 11). Compared with the latest publication on the US radiologic technologist cohort, including 90,305 radiologic technologists in the United States who were followed during 1983–1998, the values of our coefficients for thyroid radiation doses received during childhood are slightly higher than the coefficients estimated in the general analysis and the papillary carcinoma subanalysis (32). Our estimates are also slightly higher than estimates in both the most recent analysis among Japanese atomic-bomb survivors (12) and the IARC study of the highly contaminated area of Belarus following the Chernobyl nuclear accident (8).

The genetic analyses showed an association between DTC risk and SNPs: rs6067822 in intron 1 of *NFATc2* (20q11.22), rs7164173 in intron 3 of *CHD2* (15q21.3), rs1059394 (3 prime UTR variant), and rs699517 (3 prime UTR variant) both in *ENOSF1/THYS* (18p11.32),

and rs12702628 intron variant in *RPA3* (7p21.3). To our knowledge, none of these genes have previously been reported to be associated with DTC risk. The effect direction of these SNPs was similar in the Young-Thyr and CATHY studies and reached the significance threshold for the meta-analysis; furthermore, *NFATc2* and *CHD2* genes have been implicated in the response to ionizing and ultraviolet radiation *in vitro* studies (33, 34). Moreover, cells with mutations in *CHD2* were found to be defective in their ability to repair DNA damage, especially after ionizing and ultraviolet radiation exposure (34). Genes of the *NFAT* family encode for transcription factors, which have a role in cancer progression, and also are important for responding to ionizing and ultraviolet radiation exposure (35–37); these genes were previously found to be associated with other types of cancer risk, such as melanoma, breast, colorectal, and oral cancers (38–41). rs6067822 in *NFATc2* and rs7164173 in *CHD2* showed a similar stabilizing effect in general and papillary subgroup analyses. Our results from the papillary thyroid carcinoma subanalysis yielded a significant association of rs1290997 in *RAD51B* with DTC risk. Another variant in this gene, rs1950764, was among our top five SNPs but did not reach the significance threshold. These findings suggest a role for *RAD51B* in the etiology of radiation-related DTC. Interestingly, the *RAD51B* gene was involved in DNA damage repair after ultraviolet irradiation exposure (42, 43) and associated with breast cancer risk (44, 45). We report this result with great caution because 525 SNPs in this gene were tested in the first step of the analysis (Supplementary Table S1).

Table 3. Number of subjects by age and exam type.

	Cases n (%) having at least one listed exam		Controls n (%) having at least one listed exam	
	<20 years	≥20 years	<20 years	≥20 years
Conventional radiography				
Skull	89 (6.4)	93 (6.7)	94 (5.9)	82 (5.2)
Sinus	67 (4.8)	124 (8.9)	77 (4.9)	133 (8.4)
Dental	416 (29.9)	260 (18.7)	555 (35.1)	309 (19.6)
Spine	180 (12.9)	277 (19.9)	199 (12.6)	296 (18.7)
Chest	159 (11.4)	451 (32.4)	198 (12.5)	465 (29.4)
Intravenous urography	27 (1.9)	94 (6.7)	31 (2.0)	74 (4.7)
CT scan				
Abdomen	7 (0.5)	47 (3.4)	2 (0.1)	33 (2.1)
Chest	8 (0.6)	46 (3.3)	0 (0)	33 (2.1)
Skull	53 (3.8)	137 (9.8)	47 (3.0)	117 (7.4)
Spine	31 (2.22)	122 (8.75)	29 (1.83)	160 (10.12)
Nuclear medicine				
Cerebral	1 (0.1)	3 (0.2)	0 (0)	9 (0.6)
Heart	1 (0.1)	2 (0.1)	1 (0.1)	7 (0.4)
Lung	1 (0.1)	3 (0.2)	0 (0)	7 (0.4)
Bones	8 (0.6)	43 (3.1)	6 (0.4)	52 (3.3)
Hepatic	0 (0)	2 (0.1)	0 (0)	2 (0.1)
Renal	4 (0.3)	5 (0.4)	5 (0.3)	13 (0.8)

Table 4. Dose-response analysis for thyroid radiation dose: whole population and subgroups.

	Whole population: adjusted analysis			Only micro-carcinomas cases (431 cases and 1580 controls)			Excluding micro-carcinomas cases (962 cases and 1580 controls)		
	Coefficient (95% CI) for 1 mGy	Deviance	P	Coefficient (95% CI) for 1 mGy	Deviance	P	Coefficient (95% CI) for 1 mGy	Deviance	P
No radiation dose		3966.7							
The radiation dose received during childhood									
Linear	0.017 (0.00060, 0.035)	3959.5	0.007	0.024 (−0.0024, 0.051)	2370.9	0.008	0.015 (−0.0041, 0.035)	3029.4	0.04
Quadratic	0.00053 (−0.00047−0.0011)	3957.9	0.003	0.00078 (−0.00011, 0.0017)	2363.6	0.003	0.00048 (−0.00015, 0.00079)	3025.2	0.02
Exponential	0.022 (0.0081−0.036)	3058.4	0.004	0.026 (0.010, 0.042)	2361.8	0.003	0.021 (0.0049−0.037)	3024	0.03
The radiation dose received during adulthood									
Linear	8.7 10 ^{−5} (−6.0 10 ^{−4} , 7.7 10 ^{−4})	3966.7	>0.9	−8.6 10 ^{−4} (−5.8 10 ^{−3} , 4.0 10 ^{−3})	2370.8	>0.9	−2.7 × 10 ^{−5} (−5.4 × 10 ^{−3} , 5.3 × 10 ^{−3})	3029.4	>0.9
Quadratic	−9.1 10 ^{−6} (−2.6 10 ^{−5} , 7.4 10 ^{−6})	3966.2	>0.5	−6.0 10 ^{−6} (−3.1 10 ^{−5} , 1.9 10 ^{−5})	2370.7	>0.9	−1.5 × 10 ^{−5} (−1.5 × 10 ^{−5} , 1.5 × 10 ^{−5})	3028.5	0.4
Exponential	1.2 10 ^{−4} (−0.0089, 0.0091)	3966.7	>0.9	−1.6 10 ^{−3} (−0.014, 0.011)	2370.8	>0.9	−3.4 × 10 ^{−5} (−0.011, 0.011)	3029.4	>0.9

^aStratified on age, sex, and study, and adjusted on thyroid cancer family history, radiotherapy history, BMI, height, smoking habits, and number of pregnancies.

An association was found between two SNPs (rs1059394 and rs699517) falling within the *ENOSF1/THYS* gene region. These SNPs are noncoding transcript exon variants, and this region is known to be associated with ovarian and endometrial cancers (46, 47).

The polymorphism rs12702628 in replication protein A3 gene (*RPA3*), which was associated with DTC risk in our study, has previously been found to be involved in tumorigenesis, especially in gastric cancer (48).

The macrocarcinoma and microcarcinoma analyses both yielded significant results in 2q31.1. However, since these are subgroups results, they should be taken with caution, especially for the microcarcinoma analysis due to the low number of cases.

The minor allele of *MGMT* rs7068306 was found to interact with thyroid IR doses during childhood in reducing the risk of DTC after exposure. The first study to report an association between an *MGMT* variant and radiation-related DTC risk was performed in Belarusian children exposed to radiation (49). The polymorphism rs2296675, associated with DTC risk in the Belarusian population, is located 94 kb from rs7068306 in intron 4 of the *MGMT* region. A correlation coefficient of $r^2 = 0.005$ was found between rs7068306 and rs2296675. In a study of a Caucasian population, Sandler and colleagues found an interaction between other *MGMT* variants: rs1762444, rs4750763, and rs12219606, and exposure to radiation from diagnostic procedures (17). These SNPs are located in intron 2 and are 244kb, 26kb, and 15kb from rs7068306, respectively. The r^2 values between these three SNPs and rs7068306 are 0.009 and 0.044 (Supplementary Table S7). The r^2 values between rs7068306 and the SNPs tested in studies by Lonjou and Sandler were estimated in a European reference population of 1,000 Genomes project (50). Results from this study were obtained after estimation and quantification of thyroid doses, based on a combination of self-reported radiologic procedure history and literature doses separately on childhood and adulthood observations. The Sandler and colleagues study used a different procedure since no doses were taken into account; only categories of diagnostic procedure exposure based on self-reported history were evaluated. Taken together, results from these different studies highlight the role of *MGMT* in modulating the effect of IR on cells of the thyroid.

The expression of *CLIP2*, a gene located on 7q11.22, has been found to be strongly linked with previous radiation exposure in thyroid cancer (51). In our study, only 22 SNPs located on chromosome 7 were tested (Supplementary Table S1), of which, only two were near the *CLIP2* gene region. None of these were among the 100 best SNPs in our study, because the low numbers of concerned SNPs prevented us from confirming any results about locus 7q11.22.

Limitations of the study

Investigating the thyroid radiation dose-response from self-reported radiation examinations carried out in case-control studies is sensitive to recall bias. This is particularly true when addressing public health issues highly debated by the public, such as the health effects of irradiation. As Hallquist and Jansson previously showed, in case-control studies investigating the impact of self-reported diagnostic radiology procedures on DTC risk, this bias produces under-reporting of examinations by both cases and controls, but significantly more by controls; furthermore, this bias could cause an overestimate of the IR effect (52). We cannot exclude such a bias in our study, which could result in an overestimation of the radiation dose effect. However, in the Young-Thyr study, we previously showed that despite patients with DTC being more likely than controls to believe that the consequences of the Chernobyl accident were responsible for DTC occurrences, self-reported vegetable consumption during the 2 months

Role of DNA Repair Variants in Thyroid Cancer

Table 5. SNPs with the lowest *P* values associated with DTC risk in the meta-analysis among all DTC types cases (A), among only papillary DTC cases (B), macrocarcinoma cases (C), and microcarcinoma cases (D).

A/SNPs with lowest <i>P</i> values associated with DTC risk in the meta-analysis including all histologic types												
SNP	Locus	RA	EA	EAF	Gene	Young-Thyr		CATHY		Meta-analysis		
						OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	FDR <i>P</i> value
rs12702628	7p21.3	A	G	0.4	<i>RPA3</i>	1.25	1.03×10^{-2}	1.22	4.37×10^{-2}	1.24	1.12×10^{-3}	0.02
rs7164173	15q26.1	G	T	0.1	<i>CHD2</i>	0.7	1.05×10^{-3}	0.69	1.06×10^{-2}	0.68	9.26×10^{-5}	0.004
rs1059394	18p11.32	C	T	0.3	<i>ENOSF1/THYS</i>	1.34	2.91×10^{-3}	1.19	8.97×10^{-2}	1.3	7.81×10^{-4}	0.01
rs699517	18p11.32	C	T	0.3	<i>ENOSF1/THYS</i>	1.34	2.91×10^{-3}	1.19	8.97×10^{-2}	1.3	7.81×10^{-4}	0.01
rs6067822	20q13.2	G	T	0.3	<i>NFATc2</i>	0.64	1.6×10^{-3}	0.7	1.13×10^{-3}	0.64	5.79×10^{-5}	0.005
B/SNPs with lowest <i>P</i> values associated with papillary thyroid carcinoma risk in the meta-analysis (563 cases from Young-Thyr and 392 cases from CATHY)												
SNP	Locus	RA	EA	EAF	Gene	Young-Thyr		CATHY		Meta-analysis		
						OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	FDR <i>P</i> value
rs1950764	14q24.1	G	A	0.2	<i>RAD51B</i>	0.56	2.87×10^{-3}	0.66	7.27×10^{-2}	0.60	6.08×10^{-4}	0.01
rs1290997	14q24.1	G	T	0.2	<i>RAD51B</i>	0.77	2.01×10^{-3}	0.70	7.18×10^{-2}	0.74	4.36×10^{-4}	0.01
rs7164173	15q26.1	G	T	0.3	<i>CHD2</i>	0.69	3.24×10^{-3}	0.68	1.02×10^{-2}	0.68	9.33×10^{-5}	0.009
rs77971457	20q13.2	G	T	0.1	<i>NFATc2</i>	1.44	1.09×10^{-3}	1.47	1.76×10^{-2}	1.45	5.06×10^{-4}	0.01
rs6067822	20q13.2	G	T	0.3	<i>NFATc2</i>	0.81	3.47×10^{-3}	0.72	4.63×10^{-2}	0.77	5.59×10^{-4}	0.01
C/SNPs with lowest <i>p</i> -values associated with large thyroid carcinoma (>10 mm) risk in the meta-analysis (501 cases from Young-Thyr and 237 cases from CATHY)												
SNP	Locus	RA	EA	EAF	Gene	Young-Thyr		CATHY		Meta-analysis		
						OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	FDR <i>P</i> value
rs72625242	2q31.1	C	T	0.22	<i>MLK7-AS1</i>	1.92	6.27×10^{-4}	0.99	1	1.92	6.25×10^{-4}	0.004
rs17619600	2q37.1	T	C	0.15	<i>PMSD1</i>	1.85	1.20×10^{-3}	1.01	0.99	1.85	1.20×10^{-3}	0.004
rs7578070	2p24.2	A	G	0.41	<i>SMC6</i>	1.53	1.86×10^{-3}	1.03	0.99	1.53	1.86×10^{-3}	0.004
rs16861406	2q31.1	A	G	0.21	<i>MLK7-AS1</i>	1.91	2.56×10^{-3}	1.009	1	1.91	2.56×10^{-3}	0.005
rs5023821	10q26; 2	T	C	0.25	<i>C10orf90</i>	0.62	1.14×10^{-3}	1.026	0.99	0.62	1.14×10^{-3}	0.004
D/SNPs with lowest <i>P</i> values associated with small thyroid carcinoma (<10 mm) risk in the meta-analysis (128 cases from Young-Thyr and 206 cases from CATHY)												
SNP	Locus	RA	EA	EAF	Gene	Young-Thyr		CATHY		Meta-analysis		
						OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	FDR <i>P</i> value
rs75361806	2q31.1	A	G	0.05	<i>TLK1</i>	25.55	1.12×10^{-2}	1.02	1	25.54	1.12×10^{-2}	0.03
rs73234091	20p12.1	T	C	0.05	<i>MACROD2</i>	30.89	6.99×10^{-3}	1.007	1	30.88	6.99×10^{-3}	0.03
rs17757541	18q21.33	C	G	0.06	<i>BCL2</i>	7.18	1.22×10^{-2}	1.005	1	7.18	1.12×10^{-2}	0.01
rs116851051	20p12.1	T	C	0.1	<i>MACROD2</i>	28.39	1.12×10^{-2}	0.99	1	28.38	1.12×10^{-2}	0.03
rs2327965	20p12.1	A	G	0.05	<i>MACROD2</i>	27.11	1.12×10^{-2}	1.016	1	27.10	1.12×10^{-2}	0.03

Abbreviations: EA, effect allele; EAF, effect allele frequency; RA, reference allele.

^aAdjusted on age, sex, childhood IR thyroid doses, thyroid cancer family history, radiotherapy history, goiter history, BMI, height, educational level, smoking status, number of pregnancies, and three first genetic components.

after the Chernobyl accident, known to be a major source of radioactive contamination, was correlated with the status of participants, cases or controls, but not with their beliefs (53). Thus, exposure could be overreported in patients, particularly when they believe that this exposure contributed to their disease. In this type of study, case and control self-declarations are often impacted by recall bias, especially in

terms of the number of radio diagnostic exams and images in this study. Despite these caveats, case-control studies could help to improve understanding of the risks associated with low-dose IR radiation by utilizing a large number of cases and permitting more detailed investigations on several potential risk factors, which are costlier to investigate in cohort studies.

Table 6. Five SNPs with the lowest *P* values in SNPs—diagnostic radiation thyroid doses interaction term from the meta-analysis.

SNP	Locus	Reference allele	Effect allele	EAF	Gene	Young-Thyr		CATHY		Meta-analysis		
						OR ^a	<i>P</i> value	OR ^a	<i>P</i> value	OR	<i>P</i> value	FDR correction
rs10779613	1q41	T	C	0.4	<i>SMYD2</i>	0.60	1.19×10^{-3}	0.94	6.08×10^{-1}	0.80	5.42×10^{-3}	0.1
rs17514740	7p11.2	G	C	0.4	<i>EGFR</i>	1.41	9.39×10^{-3}	1.19	1.84×10^{-1}	1.30	4.64×10^{-3}	0.1
rs7068306	10q26.3	C	G	0.3	<i>MGMT</i>	0.63	1.91×10^{-3}	0.70	5.85×10^{-2}	0.65	3.40×10^{-4}	0.03
rs10402248	19p13.3	C	T	0.4	<i>MUM1</i>	0.65	4.79×10^{-3}	0.85	2.54×10^{-1}	0.75	4.03×10^{-3}	0.1
rs6066138	20q13.12	G	A	0.2	<i>EYA2</i>	1.52	9.73×10^{-3}	1.22	9.01×10^{-2}	1.32	2.20×10^{-3}	0.1

Abbreviations: EAF, effect allele frequency; OR, odds ratio for the interaction between effect allele and radiation dose, considered as a continuous variable in an exponential dose-response model.

^aAdjusted on age, sex, thyroid cancer family history, radiotherapy history, goiter history, BMI, height, educational level, smoking status, number of pregnancies, and three first genetic components.

For some types of exams, there are no available thyroid dose estimates available in current literature, especially during childhood. This disadvantage could decrease the precision of our estimates; consequently, to harmonize our analyses we considered the same doses for all children older than 2 years. The simplest way to account for age variation of radiation dose exposure (0–2 years compared with older) in our analysis, has probably a limited effect because the reference doses were the same for cases and controls. Thyroid dose estimates were based on self-declarations of diagnostic exam history. In this data, dates for each exam are not precise, causing the impossibility to impose a lagging period and raising the problem of reverse causation. Despite the absence of specific dates for each procedure, we only kept data from radiologic procedures that were not carried out for diagnosing thyroid cancer. There is some heterogeneity in demographic and cancer characteristics between our two studies: patients from the Young-Thyr study were younger at cancer diagnosis and inclusion in the study compared with cases from CATHY. Furthermore, diagnostic radiology technologies that were available in childhood differ substantially between the participants of the CATHY and Young-Thyr studies; in addition, the availability and frequency of these examinations differ (Supplementary Tables S2–S4). To better discern signals from the genetic analysis and to increase the statistical power, we chose to perform a meta-analysis to substantiate homologous signals in both studies by keeping a limited number of SNPs in this step to avoid a high number of tests.

Despite the small size of our study populations, our results suggest a role for *NFATc2*, *CHD2*, *RPA3*, and *ENOSF1/THYS* that has not previously been associated with DTC risk and confirm the previously demonstrated role of *RAD51B* and *MGMT*. However, functional and validation studies are needed to better understand and confirm the biological role of these variants.

Conclusion

We present evidence that DTC risk is significantly increased in higher thyroid radiation doses received during childhood from radio diagnostic examinations. In addition, we demonstrated an interaction between this thyroid radiation dose and the minor allele of rs706830 located in *MGMT*. Our results also suggest an association between polymorphisms in *NFATc2*, *CHD2*, *ENOSF1/THYS*, *RPA3*, and *RAD51B*. Taken together, despite the limited size of both populations studied, these results underline the role of DNA repair pathways in DTC risk. More studies of larger populations are needed to validate

and characterize the interaction between DNA repair genes and exposure to low IR doses.

Authors' Disclosures

No disclosures were reported.

Authors' Contributions

M. Zidane: Conceptualization, formal analysis, methodology, writing—original draft. **T. Truong:** Investigation. **F. Lesueur:** Validation, writing—review and editing. **C. Xhaard:** Investigation. **E. Cordina-Duverger:** Investigation. **A. Boland:** Data curation, genotyping. **H. Blanché:** Data curation, genotyping. **C. Ory:** Writing—review and editing. **S. Chevillard:** Writing—review and editing. **J.-F. Deleuze:** Data curation, genotyping. **V. Souchart:** Investigation. **Y. Ren:** Investigation. **M.Z. Zemmach:** Data curation. **S. Canale:** Data curation. **F. Borson-Chazot:** Investigation. **C. Schvartz:** Investigation. **E. Mariné Barjoan:** Investigation, writing—review and editing. **A.-V. Guizard:** Investigation. **P. Laurent-Puig:** Investigation. **C. Mulot:** Data curation, investigation. **J. Guibon:** Data curation. **M. Karimi:** Data curation, methodology. **M. Schlumberger:** Investigation, methodology. **E. Adjadj:** Investigation. **C. Rubino:** Investigation, writing—review and editing. **P. Guenel:** Supervision, investigation, methodology, writing—review and editing. **J.-B. Cazier:** Formal analysis, supervision, funding acquisition, investigation, methodology, writing—review and editing. **F. de Vathaire:** Formal analysis, supervision, funding acquisition, investigation, methodology, writing—review and editing.

Acknowledgments

F. de Vathaire received the following grants as principal investigator of Young-Thyr study: Fondation de France (Grant No. 65391), L'Alliance pour les sciences de la vie et de la santé, "AVIESAN" (Grants Nos. ENV201416 and 2007005070), the Ligue Nationale Contre le Cancer (LNCC; Grants Nos. EPDAC6036, EPDTP6004, and R10127LL). T. Truong received the following grants for CATHY and Young-Thyr studies genotyping Institut National du Cancer (Grant No. 9533), the ARC foundation (Grant No. PGA12015020302). M. Zidane performed this work during her PhD funded by Ecole des Hautes Etudes en Santé Publique (EHESP) and the ARC Foundation (Grant No. ARCD042020070002532). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would like to thank the CEPH-Biobank team, Informatique et Statistique en Anatomie Pathologique en Provence-Alpes-Côte d'Azur (CRISAP PACA) team, as well as Dr. C. Sattouet, Dr. J.L. Lassalle, Dr. Z. Hafdi-Nejari, Dr. P. Delafosse, Ms. Kami-Marie Moreau, Ms. Cyrielle Orenes, Ms. Laurianne Sarrazin, Ms. Stéphanie Bonnay, Ms. Frédérique Chatelain, Ms. Maryse Barouh, Ms. Evelynne Rapp, Ms. Julie Festræts, Ms. Julie Valbousquet, Mr. Yusuf Atilgan, Mr. Jean Chappellet, Ms. Lallia Bedhouche, Mr. Florent Dayet and Ms. Ziyen Fami for their involvement in Young-Thyr study.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 3, 2020; revised November 12, 2020; accepted March 29, 2021; published first April 7, 2021.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;66:7–30.
2. Vanderpump MPJ. The epidemiology of thyroid disease. *Br Med Bull* 2011;99:39–51.
3. Stepanenko VF, Voillequé PG, Gavrilin YI, Khrouch VT, Shinkarev SM, Orlov MY, et al. Estimating individual thyroid doses for a case-control study of childhood thyroid cancer in Bryansk Oblast, Russia. *Radiat Prot Dosimetry* 2004;108:143–60.
4. Modan B, Ron E, Werner A. Thyroid cancer following scalp irradiation. *Radiology* 1977;123:741–4.
5. Thomas GA, Bethel JA, Galpine A, Mathieson W, Krznaric M, Unger K. Integrating research on thyroid cancer after chernobyl - The Chernobyl Tissue Bank. *Clin Oncol* 2011;23:276–81.
6. Inskip PD, Sigurdson AJ, Veiga L, Bhatti P, Ronckers C, Rajaraman P, et al. Radiation-related new primary solid cancers in the childhood cancer survivor study: comparative radiation dose response and modification of treatment effects. *Int J Radiat Oncol* 2016;94:800–7.
7. Memon A, Godward S, Williams D, Siddique I, Al-Saleh K. Dental x-rays and the risk of thyroid cancer: a case-control study. *Acta Oncol* 2010;49:447–53.
8. Cardis E, Kesminiene A, Ivanov V, Malakhova I, Shibata Y, Khrouch V, et al. Risk of thyroid cancer after exposure to 131I in childhood. *J Natl Cancer Inst* 2005;97:724–32.
9. Ron E, Lubin JH, Shore RE, Mabuchi K, Modan B, Pottern LM, et al. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat Res* 1995;141:259.
10. Veiga LHSS, Lubin JH, Anderson H, De Vathaire F, Tucker M, Bhatti P, et al. A pooled analysis of thyroid cancer incidence following radiotherapy for childhood cancer. *Radiat Res* 2012;178:365–76.
11. Lubin JH, Adams MJ, Shore R, Holmberg E, Schneider AB, Hawkins MM, et al. Thyroid cancer following childhood low-dose radiation exposure:

Role of DNA Repair Variants in Thyroid Cancer

- a pooled analysis of nine cohorts. *J Clin Endocrinol Metab* 2017;102:2575–83.
12. Furukawa K, Preston D, Funamoto S, Yonehara S, Ito M, Tokuoka S, et al. Long-term trend of thyroid cancer risk among Japanese atomic-bomb survivors: 60 years after exposure. *Int J Cancer* 2013;132:1222–6.
 13. Leung AM. Low-dose childhood radiation effects to the thyroid follow a linear dose–response trend and persist even 45+ years after exposure. *Clin Thyroidol* 2017;29:235–6.
 14. Mazonakis M, Tzedakis A, Damilakis J, Gourtsoyannis N. Thyroid dose from common head and neck CT examinations in children: is there an excess risk for thyroid cancer induction? *Eur Radiol* 2007;17:1352–7.
 15. Mahaney BL, Meek K, Lees-Miller SP. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 2009;417:639–50.
 16. Bhatti P, Yong LC, Doody MM, Preston DL, Kampa DM, Ramsey MJ, et al. Diagnostic X-ray examinations and increased chromosome translocations: evidence from three studies NIH Public Access. *Radiat Env Biophys* 2010;49:685–92.
 17. Sandler JE, Huang H, Zhao N, Wu W, Liu F, Ma S, et al. Germline variants in DNA repair genes, diagnostic radiation, and risk of thyroid cancer. *Cancer Epidemiol Biomarkers Prev* 2018;27:285–94.
 18. Zidane M, Cazier J-B, Chevillard S, Ory C, Schlumberger M, Dupuy C, et al. Genetic susceptibility to radiation-related differentiated thyroid cancers: a systematic review of literature *Endocr Relat Cancer* 2019;26:R583–96.
 19. Cordina-Duverger E, Leux C, Neri M, Tcheandjieu C, Guizard A-V, Schwartz C, et al. Hormonal and reproductive risk factors of papillary thyroid cancer: a population-based case-control study in France. *Cancer Epidemiol* 2017;48:78–84.
 20. Xhaard C, Rubino C, Cléro E, Maillard S, Ren Y, Borson-Chazot F, et al. Menstrual and reproductive factors in the risk of differentiated thyroid carcinoma in young women in France: a population-based case-control study. *Am J Epidemiol* 2014;180:1007–17.
 21. Chang L, Miller D, Lee C, Melo D, Villoing D, Drozdovitch V, et al. Thyroid radiation dose to patients from diagnostic radiology procedures over eight decades: 1930–2010. *Health Phys* 2017;113:458–73.
 22. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The oncoarray consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomarkers Prev* 2017;26:126–35.
 23. Li Y, Byun J, Cai G, Xiao X, Han Y, Cornelis O, et al. FastPop: a rapid principal component derived method to infer intercontinental ancestry using genetic data. *BMC Bioinformatics* 2016;17:122.
 24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;25:25–9.
 25. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* 2019;47:D330–8.
 26. Berrington de Gonzalez A, Daniels RD, Cardis E, Cullings HM, Gilbert E, Hauptmann M, et al. Epidemiological studies of low-dose ionizing radiation and cancer: rationale and framework for the monograph and overview of eligible studies. *J Natl Cancer Inst Monogr* 2020;2020:97–113.
 27. Akaike H, Petrov BN, Csaki F. Information theory and an extension of the maximum likelihood principle BT - selected papers of Hirotugu Akaike. *Second Int Symp Inf Theory* 1973;267–81.
 28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.
 29. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
 30. Hunt SE, McLaren W, Gil L. Ensembl variation resources. *Database* 2018;2018:1–12.
 31. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinform Appl NOTE* 2010;26:2336–7.
 32. Little MP, Lim H, Friesen MC, Preston DL, Doody MM, Sigurdson AJ, et al. Assessment of thyroid cancer risk associated with radiation dose from personal diagnostic examinations in a cohort study of US radiologic technologists, followed 1983–2014. *BMJ Open* 2018;8:e021536.
 33. Nagarajan P, Onami TM, Rajagopalan S, Kania S, Donnell R, Venkatachalam S. Role of chromodomain helicase DNA-binding protein 2 in DNA damage response signaling and tumorigenesis. *Oncogene* 2009;28:1053–62.
 34. Rajagopalan S, Nepa J, Venkatachalam S. Chromodomain helicase DNA-binding protein 2 affects the repair of X-ray and UV-Induced DNA damage. *Environ Mol Mutagen* 2012;53:44–50.
 35. Beals CR, Sheridan CM, Turck CW, Gardner P, Crabtree GR. Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* 1997;275:1930–3.
 36. Flockhart RJ, Diffey BL, Farr PM, Lloyd J, Reynolds NJ. NFAT regulates induction of COX-2 and apoptosis of keratinocytes in response to ultraviolet radiation exposure. *FASEB J* 2008;22:4218–27.
 37. Glover D, Little JB, Lavin MF, Gueven N. Low dose ionizing radiation-induced activation of connexin 43 expression. *Int J Radiat Biol* 2003;79:955–64.
 38. Perotti V, Baldassari P, Molla A, Vegetti C, Bersani I, Maurichi A, et al. NFATc2 is an intrinsic regulator of melanoma dedifferentiation. *Oncogene* 2016;35:2862–72.
 39. Kim GC, Kwon HK, Lee CG, Verma R, Rudra D, Kim T, et al. Upregulation of Ets1 expression by NFATc2 and NFKB1/RELA promotes breast cancer cell invasiveness. *Oncogenesis* 2018;7:91.
 40. Damani Shah H, Saranath D, Pradhan S. Single nucleotide polymorphisms in transcription factor genes associated with susceptibility to oral cancer. *J Cell Biochem* 2020;121:1050–60.
 41. Lang T, Ding X, Kong L, Zhou X, Zhang Z, Ju H, et al. NFATC2 is a novel therapeutic target for colorectal cancer stem cells. *Onco Targets Ther* 2018;11:6911–24.
 42. Havre PA, Rice MC, Noe M, Kmiec EB. The human REC2/RAD51B gene acts as a DNA damage sensor by inducing G1 delay and hypersensitivity to ultraviolet irradiation. *Cancer Res* 1998;58:4733–9.
 43. Takata M, Sasaki MS, Sonoda E, Fukushima T, Morrison C, Albala JS, et al. The Rad51 paralog Rad51B promotes homologous recombinational repair. *Mol Cell Biol* 2000;20:6476–82.
 44. Lee PS, Fang J, Jessop L, Myers T, Raj P, Hu N, et al. Rad51b activity and cell cycle regulation in response to DNA damage in breast cancer cell lines. *Breast Cancer Basic Clin Res* 2014;8:135–44.
 45. Peltari LM, Khan S, Vuorela M, Kiiski JI, Vilske S, Nevanlinna V, et al. RAD51B in familial breast cancer. *PLoS One* 2016;11:e0153788.
 46. Xu WH, Long JR, Zheng W, Ruan ZX, Cai Q, Cheng JR, et al. Association of thymidylate synthase gene with endometrial cancer risk in a Chinese population. *Cancer Epidemiol Biomarkers Prev* 2009;18:579–84.
 47. Kelemen L, Earp M, Fridley B, Chenevix-Trench G, Fasching P, Beckmann M, et al. rs495139 in the TYMS-ENOSF1 region and risk of ovarian carcinoma of mucinous histology. *Int J Mol Sci* 2018;19:2473.
 48. Dai Z, Wang S, Zhang W, Yang Y. Elevated expression of RPA3 is involved in gastric cancer tumorigenesis and associated with poor patient survival. *Dig Dis Sci* 2017;62:2369–75.
 49. Lonjou C, Damiola F, Moissonnier M, Durand G, Malakhova I, Masyakin V, et al. Investigation of DNA repair-related SNPs underlying susceptibility to papillary thyroid carcinoma reveals MGMT as a novel candidate gene in Belarusian children exposed to radiation. *BMC Cancer* 2017;17:328.
 50. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. *Nature* 2015;526:68–74.
 51. Heß J, Thomas G, Braselmann H, Bauer V, Bogdanova T, Wienberg J, et al. Gain of chromosome band 7q11 in papillary thyroid carcinomas of young patients is associated with exposure to low-dose irradiation. *Proc Natl Acad Sci U S A* 2011;108:9595–600.
 52. Hallquist A, Jansson P. Self-reported diagnostic X-ray investigation and data from medical records in case-control studies on thyroid cancer: evidence of recall bias? *Eur J Cancer Prev* 2005;14:271–6.
 53. Xhaard C, Dumas A, Soucard V, Ren Y, Borson-Chazot F, Sassolas G, et al. Are dietary reports in a case-control study on thyroid cancer biased by risk perception of Chernobyl fallout? *Rev Epidemiol Sante Publique* 2017;65:301–8.

Titre : Étude des facteurs de risque génétiques des cancers différenciés de la thyroïde dans une population multi-ethnique : cartographie fine des régions 2q35 et 8p12 et caractérisation des régions d'homozygotie associées à ces cancers

Mots clés : cancer différencié de la thyroïde, épidémiologie génétique, étude pangénomique, cartographie fine, région d'homozygotie

Résumé : L'incidence des cancers différenciés de la thyroïde (CDT) est caractérisée par de fortes variations ethniques et géographiques. Les facteurs de risque génétiques encore mal connus pour les CDT pourraient expliquer en partie ces variations.

L'objectif de cette thèse est de mieux caractériser les facteurs de risque génétiques des CDT avec dans un premier temps une analyse GWAS conduite chez les Européens et Océaniens dans les études cas-témoins du consortium EPITHYR. Des analyses de cartographie fine (ou *fine-mapping*) des régions 2q35 et 8p12 ont ensuite été menées afin d'identifier les variants potentiellement causaux. Enfin, nous avons caractérisé les régions d'homozygotie (runs of homozygosity, ROH) du génome dans les populations d'EPITHYR et EPIC afin d'identifier des gènes de transmission autosomique récessive impliqués dans le CDT.

Dans l'analyse GWAS classique, nous avons confirmé le rôle des quatre loci de prédisposition connus aux CDT situés en 2q35, 8p12, 9q22.33 et 14q13.3 et identifié un nouveau locus en 1p31.3 chez les Européens d'EPITHYR. Aucune association significative n'a été identifiée chez les Océaniens. L'étude de cartographie fine met en évidence les SNPs rs16857609 en 2q35 et rs2439304 en 8p12 comme étant les variants les plus susceptibles d'avoir un impact fonctionnel délétère dans ces régions. Par ailleurs, l'analyse ROH, a permis de mettre en évidence deux régions homozygotes aux loci 1q25.1 et 11q24.1 associées au risque de CDT. Ces régions contiennent les gènes *DHX*, *LAMC1* et *ZNF202* impliqués dans la tumorigénèse d'autres cancers communs.

L'ensemble de ces résultats permettra de mieux évaluer la contribution des facteurs de risque génétiques et des facteurs de risque environnementaux dans le risque de CDT et dans les différences d'incidence observées entre populations.

Title: Genetic risk factors of differentiated thyroid cancer in a multi-ethnic population: fine-mapping of loci 2q35 and 8p12, and characterization of runs of homozygosity associated with this cancer

Keywords: differentiated thyroid cancer, genetic epidemiology, genome-wide association study, fine-mapping, runs of homozygosity

Abstract: The incidence of differentiated thyroid cancer (DTC) is characterized by wide ethnic and geographical variations worldwide. Genetic risk factors of DTC are still poorly known and may explain part of these variations.

The objective of this thesis was to characterize the genetic risk factors of DTC in a multiethnic population by conducting with a GWAS in Europeans and Oceanians enrolled in the case-control studies of the EPITHYR consortium. We next conducted fine-mapping analyses of two DTC susceptibility loci at 2q35 and 8p12 in order to identify causal variants. Finally, we aimed at characterizing regions of homozygosity (runs of homozygosity, ROH) in EPITHYR and EPIC studies to identify DTC susceptibility genes following an autosomal recessive mode of inheritance.

In the classical GWAS analysis, we replicated the four known DTC susceptibility loci at 2q35, 8p12, 9q22.33, 14q13.3, and identified a new locus at 1p31.3 in the

European of EPITHYR. No significant association was identified in Oceanians. The fine mapping analysis revealed rs16857609 at 2q35 and rs2439304 at 8p12 as the variants most likely to have a deleterious functional effect in these regions. In the ROH analysis, we identified two homozygous regions associated with DTC risk at loci 1q25.1 and 11q24.1 in which several genes, including *DHX*, *LAMC1* and *ZNF202* have been found to be involved in tumorigenesis or the etiology of other common cancers.

Our results provide novel insights into the inherited genetic basis of DTC and will help to better understand the mechanisms involved in thyroid carcinogenesis. Identification of population-specific SNPs may also increase our understanding of the genetic architecture underlying DTC in different ethnic populations and may help to understand the large ethnic variations observed in the incidence of DTC.

