



HAL
open science

Cancers secondaires chez les patients ayant survécu à un cancer durant l'enfance et Identification de variants rares associés

Claire Ducos

► **To cite this version:**

Claire Ducos. Cancers secondaires chez les patients ayant survécu à un cancer durant l'enfance et Identification de variants rares associés. Bio-informatique [q-bio.QM]. Université Paris-Saclay, 2024. Français. NNT: 2024UPASR013 . tel-04819340

HAL Id: tel-04819340

<https://theses.hal.science/tel-04819340v1>

Submitted on 4 Dec 2024

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.

Cancers secondaires chez les patients ayant survécu à un cancer durant l'enfance et Identification de variants rares associés

*Secondary cancer among childhood cancer survivors and Identification of
associated rare variants*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 570, Santé Publique (EDSP)
Spécialité de doctorat : Génétique épidémiologique et statistique
Graduate School : Santé Publique. Référent : Faculté de médecine

Thèse préparée dans l'unité de recherche **CESP (Université Paris-Saclay, UVSQ, Inserm)**,
sous la direction de **Simone BENHAMOU**, Directrice de Recherche
et la co-direction de **Nadia HADDY**, Chargée de Recherche.

Thèse soutenue à Paris-Saclay, le 24 septembre 2024, par

Claire DUCOS

Composition du Jury

Membres du jury avec voix délibérative

Véronique MINARD

PU-PH, université Paris-Saclay

Présidente

Emmanuelle GENIN

Directrice de recherche, Université Bretagne Occidentale

Rapporteur & Examinatrice

Jacqueline CLAVEL

Directrice de recherche, Université Paris cité

Rapporteur & Examinatrice

Nicolas SERVANT

PhD, Université Paris Sciences &
Lettres

Examineur

Titre : Cancers secondaires chez les patients ayant survécu à un cancer durant l'enfance et Identification de variants rares associés.

Mots clés : Seconds cancers ; Variants rares; Survivants à un cancer de l'enfant; Séquençage d'exome; Radiothérapie

Résumé : Il est estimé qu'un enfant sur 440 développera un cancer avant sa majorité en France. Depuis quelques décennies, grâce à une amélioration des traitements et de la prise en charge, la survie à 5 ans des patients guéris d'un cancer pédiatrique atteint plus de 80 %. Cependant, la fréquence et la diversité des pathologies iatrogènes tardives ont également augmenté. Les seconds cancers font partie des effets ayant le plus d'impact sur la mortalité et la morbidité de cette population croissante. Bien que le traitement reçu lors du cancer pédiatrique soit un facteur de risque reconnu, ce dernier ne suffit pas à expliquer complètement le risque de cancers secondaires suggérant l'existence d'une composante génétique modulant ce risque.

Dans un premier temps, nous avons cherché à mieux comprendre les facteurs de risque de l'un des seconds cancers les plus fréquents, celui de la thyroïde, en étudiant l'impact des radiations reçues par les organes lymphoïdes. Nous avons ainsi mis en évidence une association entre ce risque et l'irradiation de la rate et du thymus, deux organes

clés du système immunitaire. Ensuite, afin de mieux comprendre l'impact génétique sur le risque de second cancer, nous avons réalisé une revue systématique de la littérature qui a permis de recenser l'ensemble des variants, gènes et voies biologiques déjà associés au risque des différents types de seconds cancers. Finalement, une étude cas-témoins nichée dans la cohorte French Childhood Cancer Survivors Study (FCCSS) a permis de montrer que différents gènes, porteurs de variants rares, tels que *APOBEC3F* et *RNASEL*, étaient associés au risque de seconds cancers. Nous avons également mis en évidence une association entre le risque de second cancer du sein avec le gène *FANCM*, et de second cancer de la thyroïde avec en particulier, les gènes *IFI16* et *HINFP*.

Ainsi, les résultats obtenus au cours de cette thèse fournissent des orientations prometteuses pour contribuer à identifier les patients les plus à risque de développer un second cancer. Nos résultats après validation pourraient être considérés pour adapter le suivi à long terme des survivants mais également le traitement des patients atteints d'un cancer pédiatrique afin de minimiser le risque de survenu d'un second cancer à l'âge adulte.

Title: Secondary cancer among childhood cancer survivors and Identification of associated rare variants

Keywords: Seconds malignant neoplasm; Rare Variants; Childhood Cancer Survivors; Whole exome sequencing; Radiotherapy

Abstract: One in 440 children in France is expected to develop cancer before reaching the age of majority. In recent decades, thanks to improvements in treatment and patient care, the 5-year survival rate for children diagnosed with cancer has reached over 80 %. However, the frequency and diversity of late iatrogenic events have also increased. Second cancers are one of the most significant adverse effects on mortality and morbidity in this growing population. Although the treatment received for pediatric cancer is a well-known risk factor, it is not sufficient to fully explain the risk of secondary cancers, suggesting the existence of a genetic component modulating this risk.

Initially, we aimed to better understand the risk factors for one of the most frequent second cancers, thyroid cancer, by investigating the impact of radiation received by lymphoid organs. We have identified an association between this risk and irradiation of the spleen and thymus, two key immune system organs. Then, to better understand the genetic impact on the

risk of second cancers, we conducted a systematic review of the literature, compiling all the variants, genes and biological pathways already associated with the risk of various types of second cancers. Finally, a case-control study nested in the French Childhood Cancer Survivors Study (FCCSS) cohort showed that various genes carrying rare variants, such as *APOBEC3F* and *RNASEL*, were associated with the risk of second cancers. We also demonstrated an association between the risk of second breast cancer with the *FANCM* gene, and second thyroid cancer with, notably, the *IFI16* and *HINFP* genes.

The results obtained in this thesis provide promising directions that could contribute to the identification of patients at the highest risk of second cancers. Once validated, our results could be applied to adapt the long-term follow-up of survivors and the treatment of pediatric cancer patients to minimize the risk of second cancers in adulthood.

Remerciements

Tout d'abord, je tiens à remercier mes directrices de thèse Docteure Nadia Haddy et Docteure Simone Benhamou, pour leur confiance et leur soutien tout au long de ces quatre années de travail. Merci beaucoup pour votre aide qui m'a été très importante dans la rédaction de cette thèse.

Je tiens également à remercier tous les membres de mon jury, pour qui j'ai beaucoup d'estime et devant qui je suis fière de présenter mon travail de thèse. Au Docteure Emmanuelle Génin et au Docteure Jacqueline Clavel pour avoir accepté d'évaluer mon manuscrit de thèse, ainsi qu'au Professeure. Véronique Minard et au Docteur Nicolas Servant pour avoir bien voulu être examinateurs.

Ce travail a été rendu possible grâce aux diverses collaborations qui m'ont guidé dans la bonne direction. Je tiens à exprimer ma gratitude particulière au Pr Filippo Rosselli pour son immense expertise, sa disponibilité et sa gentillesse, qui ont grandement enrichi ma compréhension et fait progresser mes recherches de thèse. Je voudrais aussi adresser mes remerciements à l'équipe du Docteur Emanuelle Génin, et plus particulièrement à Gaëlle Marenne et Thomas Ludwig qui m'ont été d'une grande aide sur la partie méthodologie de cette thèse.

Je tiens également à remercier le Docteur Florent de Vathaire et toute l'équipe d'Épidémiologie des radiations ainsi que l'équipe Oncostat pour leur accueil durant ces années de thèse. Je souhaite adresser des remerciements particuliers à Rodrigue, Brice, Françoise, Giao et Laurie, qui ont toujours su répondre à mes questions et me conseiller. Un immense merci à Boris, toujours disponible et à l'écoute, surtout en cette fin de thèse. Ces années de thèse auraient été beaucoup moins agréables sans les discussions avec David, Thibault, Stefania et Daniel qui sont devenus bien plus que de simples collègues de bureau.

J'aimerais également adresser des remerciements plus personnels à Naila, un de mes plus grands soutiens au cours de ces quatre années. Initialement simples co-doctorantes, nous sommes aujourd'hui de véritables amies. Nous avons tout partagé, jusqu'à soutenir (presque) le même jour, et je sais que nos années de thèse s'arrêteront ce jour-là, mais pas notre amitié !

Enfin, à mes parents et à ma petite sœur pour leur soutien et leur amour infaillibles. Merci de m'avoir toujours poussée à faire face aux différentes épreuves de la vie. Je vous serai éternellement reconnaissante et je ne serais jamais arrivée jusqu'ici sans vous.

Productions Scientifiques

Article publié

- Claire Ducos, Naïla Aba, Filippo Rosselli, Brice Fresneau, Baraah Al Ahmad-Nachar, Monia Zidane, Florent de Vathaire, Simone Benhamou, Nadia Haddy, Genetic risk of second malignant neoplasm after childhood cancer treatment: A systematic review. *Cancer Epidemiology, Biomarkers and Prevention*. Publié en ligne le 27 mai 2024. doi:10.1158/1055-9965.EPI-24-0010.

Article en cours de soumission

- Claire Ducos, Brice Fresneau, Filippo Rosselli, Giao Vu-Bezin, Boris Schwartz, Rodrigue Allodji, Gaelle Marenne, Thomas E.Ludwig, Thibault Dayris, Anne Boland, Jean-François Deleuze, Helene Blanché, Julie Nys, Chiraz El-Fayech, Carole Rubino, Ibrahima Diallo, Florent de Vathaire, Simone Benhamou, Nadia Haddy, Rare Variants Associated with Pediatric Cancer Treatment-Related Second Malignant Neoplasm Risk. *Soumis fin juillet 2024 à Journal of Clinical Oncology – Precision Oncology*.

Article en préparation

- Claire Ducos, Boris Schwartz, Giao Vu-Bezin, Carole Rubino, Florent de Vathaire, Simone Benhamou, Nadia Haddy, Radiation Exposure to Lymphoid Organs During Pediatric Cancer Radiotherapy and the Risk of Secondary Thyroid Cancer.

Autres publications

- Naïla Aba, Claire Ducos, Eric Morel, Chiraz El Fayech, Brice Fresneau, Florent de Vathaire, Gwénaél Le Teuff, Nadia Haddy, Influence of Genetic Biomarkers on Cardiac Diseases in Childhood Cancer Survivors: a Systematic Review. *The Pharmacogenomics Journal (soumis)*.

Communications à des conférences nationales et internationales

- Claire Ducos, Brice Fresneau, Filippo Rosselli, Giao Vu-Bezin, Boris Schwartz, Rodrigue Allodji, Gaelle Marenne, Thomas E.Ludwig, Thibault Dayris, Anne Boland, Jean-François Deleuze, Helene Blanché, Julie Nys, Chiraz El-Fayech, Carole Rubino, Ibrahima Diallo, Florent de Vathaire, Simone Benhamou, Nadia Haddy, Rare Variants Associated with Pediatric Cancer Treatment-Related Second Malignant Neoplasm Risk. Conférence francophone d'Épidémiologie CLINique (EPICLIN), Mai 2023, Nancy, France. Poster.
- Claire Ducos, Brice Fresneau, Filippo Rosselli, Giao Vu-Bezin, Boris Schwartz, Rodrigue Allodji, Gaelle Marenne, Thomas E.Ludwig, Thibault Dayris, Anne Boland, Jean-François Deleuze, Helene Blanché, Julie Nys, Chiraz El-Fayech, Carole Rubino, Ibrahima Diallo, Florent de Vathaire, Simone Benhamou, Nadia Haddy, Rare Variants Associated with Pediatric Cancer Treatment-Related Second Malignant Neoplasm Risk ISORED, Mai 2023, Sitges, Espagne. Poster.

Autres communications

- C. Ducos, Identification of rare variants involved in the risk of second cancer following childhood cancer treatment Club Bio-informatique de l'Institut Gustave Roussy, Décembre 2021-2022-2023. Communication orale.
- C. Ducos, Identification of rare variants involved in the risk of second cancer following childhood cancer treatment. Réunion du Comité de suivi à long terme, Décembre 2021, visioconférence. Communication orale.
- C. Ducos, Identification of rare variants involved in the risk of second cancer following childhood cancer treatment. Journée des Jeunes Chercheurs de l'ARC, Octobre 2023. Poster.
- C. Ducos, Rare Variants Associated with Pediatric Cancer Treatment-Related Second Malignant Neoplasm Risk. Réunion dans le cadre d'une collaboration FCCSS/CCSS avec Lindsay Morton, juillet 2024. Communication orale.

Table des matières

PRODUCTIONS SCIENTIFIQUES	I
TABLE DES MATIERES	III
LISTE DES FIGURES.....	VII
LISTE DES TABLEAUX.....	VIII
LISTE DES ABRÉVIATIONS.....	IX
CHAPITRE 1 – INTRODUCTION ET OBJECTIFS.....	1
I. LES CANCERS PEDIATRIQUES	2
1. ÉPIDEMIOLOGIE DES CANCERS PEDIATRIQUES	2
a. Incidence.....	2
b. Taux de survie	4
2. TRAITEMENTS DES CANCERS PEDIATRIQUES	6
3. COMPLICATIONS APRES TRAITEMENT POUR UN CANCER PEDIATRIQUE.....	7
a. Les différents effets iatrogènes.....	7
b. Les facteurs de risque	9
II. LES SECONDS CANCERS.....	9
1. ÉPIDEMIOLOGIE.....	9
2. FACTEURS DE RISQUE DE SECOND CANCER.....	11
a. La radiothérapie.....	12
b. La chimiothérapie	14
c. La génétique	14
III. HYPOTHESE DE RECHERCHE ET OBJECTIFS	15
1. OBJECTIFS DE LA THESE	15
2. PLAN DES CHAPITRES DE LA THESE	16
CHAPITRE 2 - DONNEES DES ETUDES – LES COHORTES FRENCH CHILDHOOD CANCER SURVIVORS STUDY (FCCSS) ET BRITISH CHILDHOOD CANCER SURVIVORS STUDY (BCCSS)	19
I. LA COHORTE « FRENCH CHILDHOOD CANCER SURVIVORS STUDY » (FCCSS).....	20
1. DEFINITION DE LA COHORTE.....	20
2. LES DONNEES DE LA FCCSS	21
a. Radiothérapie	22
b. Chimiothérapie	22

3.	SUIVI ET IDENTIFICATION DES SECONDS CANCERS DANS LA FCCSS	23
4.	LES DONNEES BIOLOGIQUES	24
II.	LA COHORTE « BRITISH CHILDHOOD CANCER SURVIVOR STUDY » (BCCSS)	25
CHAPITRE 3 - METHODES BIO-INFORMATIQUES APPLIQUEES A DES DONNEES EXOMIQUES ET METHODES		
STATISTIQUES POUR L'ETUDE DES VARIANTS RARES		
I.	INTRODUCTION	28
II.	METHODES BIO-INFORMATIQUES	29
1.	PREPARATION DES LIBRAIRIES ET SEQUENÇAGE	29
2.	TRAITEMENT DES DONNEES BRUTES DE SEQUENÇAGE	31
a.	<i>Contrôle qualité des fichiers de séquençage.....</i>	<i>31</i>
b.	<i>Alignement sur un génome de référence.....</i>	<i>31</i>
3.	DETECTION DES VARIANTS	34
a.	<i>Création des fichiers gVCF</i>	<i>Erreur ! Signet non défini.</i>
b.	<i>Etapas post variant calling.....</i>	<i>35</i>
c.	<i>VQSR</i>	<i>36</i>
4.	CONTROLE QUALITE	36
a.	<i>Contrôle qualité des variants.....</i>	<i>Erreur ! Signet non défini.</i>
b.	<i>Contrôle qualité des individus.....</i>	<i>37</i>
5.	ANNOTATION ET IDENTIFICATION DES VARIANTS	39
a.	<i>La localisation</i>	<i>Erreur ! Signet non défini.</i>
b.	<i>Impact du variant</i>	<i>39</i>
c.	<i>La Fréquence Allélique Mineure (MAF).....</i>	<i>40</i>
III.	METHODES STATISTIQUES	41
1.	VARIANTS COMMUNS VS VARIANTS RARES.....	41
2.	LES TESTS D'ASSOCIATION – VARIANTS RARES	42
3.	LES TESTS D'ASSOCIATION PAR AGREGATION	44
a.	<i>Les tests de fardeau</i>	<i>44</i>
b.	<i>Méthodes de décomposition de la variance</i>	<i>46</i>
c.	<i>Les méthodes Combinées (SKAT-O)</i>	<i>48</i>
d.	<i>Puissance des méthodes pour variants rares.....</i>	<i>48</i>
4.	SELECTION DES GENES	49
CHAPITRE 4 - RADIATIONS REÇUES AUX ORGANES LYMPHOÏDES DURANT LA RADIOTHERAPIE POUR UN CANCER DE		
L'ENFANT ET RISQUE DE SECOND CANCER DE LA THYROÏDE		
I.	CONTEXTE	54
II.	METHODOLOGIE DE L'ETUDE.....	57

1.	DONNEES DE L'ETUDE ET DESIGN DE L'ETUDE	57
2.	VARIABLES ETUDIÉES	57
3.	ANALYSES STATISTIQUES	58
III.	RESULTATS	59
IV.	DISCUSSION	63
CHAPITRE 5 – RISQUE GENETIQUE DE SECOND CANCER APRES TRAITEMENT POUR UN CANCER PEDIATRIQUE : UNE REVUE SYSTEMATIQUE DE LA LITTERATURE		
67		
I.	INTRODUCTION	68
II.	MATERIEL ET METHODES.....	68
1.	STRATEGIE DE RECHERCHE	69
2.	PROCEDURE DE SELECTION	70
3.	CONTROLE QUALITE DES ETUDES.....	70
III.	RESULTATS	73
1.	DESCRIPTION DES ARTICLES INCLUS.....	76
2.	LES ÉTUDES GÈNES-CANDIDATS.....	82
a.	<i>Etudes gènes-candidat sur le risque de seconds cancers.....</i>	<i>82</i>
b.	<i>Etudes gènes-candidat sur le risque de seconds cancers spécifiques</i>	<i>83</i>
3.	LES ETUDES PANGENOMIQUES (GWAS)	88
4.	LES ETUDES DE SEQUENÇAGE HAUT DEBIT (NGS – WGS & WES)	92
a.	<i>Etudes NGS sur le risque de seconds cancers.....</i>	<i>92</i>
b.	<i>Etudes NGS sur le risque de seconds cancers spécifiques</i>	<i>96</i>
III.	DISCUSSION.....	98
CHAPITRE 6 - IDENTIFICATION DES VARIANTES RARES ASSOCIES AU SECOND CANCER, UNE ETUDE CAS-TEMOIN NICHEE DANS LA COHORTE FCCSS		
104		
I.	CONTEXTE DE L'ETUDE	105
II.	METHODOLOGIE.....	107
1.	DESIGN DE L'ETUDE.....	107
2.	DONNEES DE SEQUENÇAGE ET TRAITEMENT BIO-INFORMATIQUE DES DONNEES	107
a.	<i>Données génétique</i>	<i>107</i>
b.	<i>Sélection des gènes des voies 3R</i>	<i>108</i>
c.	<i>Pipeline d'analyses.....</i>	<i>110</i>
3.	ANALYSES STATISTIQUES	113
III.	RESULTATS	115

1.	CARACTERISTIQUES CLINIQUES	115
2.	VARIANT IDENTIFIES PAR LE SEQUENÇAGE D'EXOME.....	115
3.	RISQUE ASSOCIE A UN CANCER SECONDAIRE (TOUS TYPE DE SECONDS CANCERS).....	118
4.	SECOND CANCER DE LA THYROÏDE	119
5.	SECOND CANCER DU SEIN	120
6.	ANALYSES STRATIFIEES SELON LA DOSE DE RADIATION REÇUE	121
IV.	DISCUSSION	125
	CHAPITRE 7 – DISCUSSION GENERALE	131
I.	RESUME DES PRINCIPAUX RESULTATS.....	132
II.	IMPLICATIONS CLINIQUES ET PERSPECTIVES	135
1.	VALIDATION DES RESULTATS	135
2.	IMPORTANCE DES COLLABORATIONS INTERNATIONALES	136
3.	CREATION DE BANQUES BIOLOGIQUES.....	136
4.	ETUDES FONCTIONNELLES.....	136
5.	IMPLICATION CLINIQUES DES RESULTATS POUR UNE MEDECINE PERSONNALISEE	137
	CONCLUSION	138
	ANNEXES	175
	ANNEXE A – 1. STRATEGIE DE RECHERCHE PICOS- PUBMED	176
	ANNEXE A – 2. STRATEGIE DE RECHERCHE PICOS- SCOPUS	178
	ANNEXE A – 3. STRATEGIE DE RECHERCHE PICOS- WOS.....	179
	ANNEXE B – PATIENTS INCLUS DANS L'ETUDE ET PORTEURS D'UN SYNDROME GENETIQUE	180
	PUBLICATIONS.....	183

Liste des figures

FIGURE 1 - INCIDENCE DES CANCERS DE L'ENFANT SELON L'AGE ET LE SEXE EN FRANCE METROPOLITAINE DE 2000-2004 ERREUR ! SIGNET NON DEFINI.	
FIGURE 2 - REPARTITION DES TYPES DE CANCER PAR GROUPE D'AGE SUR LA PERIODE 2001–2010.	3
FIGURE 3 - SURVIE PONDEREE PAR PAYS SELON LE TYPE DE PREMIER CANCER POUR LES ENFANTS EUROPEENS DIAGNOSTIQUES AVEC UN CANCER ENTRE 2000 ET 2007.	4
FIGURE 4 - TAUX DE SURVIE A 5 ANS CHEZ LES PATIENTS GUERIS D'UN CANCER PEDIATRIQUE DE 2004 A 2014	5
FIGURE 5 - ENSEMBLE DES EFFETS A LONG-TERME RETROUVES CHEZ LES SURVIVANTS D'UN CANCER PEDIATRIQUE	8
FIGURE 6 - INCIDENCE CUMULEE DES SECONDS CANCERS DANS LA COHORTE CCSS CHEZ LES PATIENTS DE PLUS DE 40 ANS.	10
FIGURE 7 - INCIDENCE CUMULEE DU SECOND CANCER DU SEIN DANS LA COHORTE DCOG LATER.....	11
FIGURE 8 - ENSEMBLE DES FACTEURS DE RISQUE DE SECOND CANCER	12
FIGURE 9 – <i>RELATION DOSE-EFFET SUR LE RISQUE DE SECONDS CANCERS APRES IRRADIATION DURANT LA RADIOTHERAPIE.</i>	13
FIGURE 10 - ILLUSTRATION DES FANTOMES ANTHROPOMETRIQUES POUR ESTIMER LA DOSE REÇUE PAR DIFFERENTS ORGANES DES PATIENTS DE LA FCCSS 22	
FIGURE 11 - ETAPES D'UN SEQUENÇAGE ILLUMINA.....	30
FIGURE 12 - SCHEMATISATION DE L'ALIGNEMENT DES READS SUR LE GENOME DE REFERENCE	32
FIGURE 13 - SUPPRESSION DES DUPLICATS.....	33
FIGURE 14 - FONCTIONNEMENT DE LA DETECTION DES VARIANTS.	35
FIGURE 15 - LOCALISATION POSSIBLES DES VARIANTS.....	39
FIGURE 16 - IMPACT PREDIT DU VARIANT CLASSE EN QUATRE CATEGORIES PAR VEP.....	40
FIGURE 17 - VARIANTS COMMUNS VS VARIANTS COMMUNS	42
FIGURE 18 - TAILLES D'ECHANTILLONS NECESSAIRES POUR DETECTER UNE ASSOCIATION ENTRE UN ALLELE AVEC UN RISQUE SPECIFIQUE ET UN PHENOTYPE D'INTERET	43
FIGURE 19 - PRINCIPAUX ORGANES LYMPHOÏDES DU SYSTEME IMMUNITAIRE.....	56
FIGURE 20 – PROCESSUS DE SELECTION DES ETUDES POUR INCLUSION DANS LA REVUE SYSTEMATIQUE.	75
FIGURE 21 - SELECTION DES GENES ET DES VARIANTS INCLUS DANS LES ANALYSES STATISTIQUES.	109
FIGURE 22 - RESULTATS DE LA RECHERCHE DE GENES DANS LES TROIS BASES DE DONNEES	109
FIGURE 23 - SUCCESSION DES OUTILS UTILISES LORS DES DIFFERENTES ETAPES LORS DE L'ANALYSE DES DONNEES EXOMIQUES.	110
FIGURE 24 - CONTROLE QUALITE DES READS - SUPPRESSION DES ADAPTATEURS.	111
FIGURE 25 - STRATIFICATION DE LA POPULATION	112
FIGURE 26 - RESULTATS DES TESTS SPECIFIQUES AUX VARIANTS RARES - SKAT, TEST DE FARDEAU ET SKAT-O.....	117

Liste des tableaux

TABLEAU 1 - CARACTERISTIQUES DES PATIENTS INCLUS DANS LA FCCSS.	21
TABLEAU 2 - DESCRIPTION DES MOLECULES DE CHIMIOTHERAPIE DANS LA COHORTE FCCSS.	23
TABLEAU 3 - AVANTAGES ET INCONVENIENTS DES METHODES STATISTIQUES SPECIFIQUES A L'ETUDE DES VARIANTS RARES.	48
TABLEAU 4 - CORRELATION ENTRE LES DOSES REÇUES AUX DIFFERENTS ORGANES D'INTERET	59
TABLEAU 5 - CARACTERISTIQUES CLINIQUES DES CAS ET DES TEMOINS DE L'ETUDE	60
TABLEAU 6 - CARACTERISTIQUES DE TRAITEMENT DES CAS ET DES TEMOINS DE L'ETUDE	61
TABLEAU 7 - ANALYSES MULTIVARIEES POUR LES TROIS ORGANES D'INTERET ET LE SCORE COMPOSITE D'IRRADIATION.	62
TABLEAU 8 - TERMES PICOS UTILISES POUR L'INCLUSION DES ARTICLES A LA REVUE SYSTEMATIQUE.	69
TABLEAU 9 - EVALUATION DU RISQUE DE BIAS DES DIX-HUIT ETUDES INCLUSES A L'AIDE DE L'OUTIL Q-GENIE (SOHANI ET AL., 2015).	72
TABLEAU 10 - RAISON D'EXCLUSION DES ARTICLES APRES LECTURE DU TEXTE	74
TABLEAU 11 - CARACTERISTIQUES DES ETUDES INCLUSES DANS LA REVUE SYSTEMATIQUE.	78
TABLEAU 12 - RESULTATS DES ETUDES GENE-CANDIDATS INCLUSES DANS LA REVUE SYSTEMATIQUE.	85
TABLEAU 13 - RESULTATS DES ETUDES PANGENOMIQUES INCLUSES DANS LA REVUE SYSTEMATIQUE.	91
TABLEAU 14 - RESULTATS DES ETUDES WES ET WGS INCLUSES DANS LA REVUE SYSTEMATIQUE.	94
TABLEAU 15 - CARACTERISTIQUES DES PATIENTS INCLUS.	116
TABLEAU 16 - RESULTATS DES ANALYSES POUR TOUS TYPES DE SECONDS CANCERS APRES UN CANCER PEDIATRIQUE.	118
TABLEAU 17 - RESULTATS DES ANALYSES POUR L'ETUDE DU RISQUE DE CANCER SECONDAIRE DE LA THYROÏDE.	119
TABLEAU 18 - RESULTATS DES ANALYSES POUR L'ETUDE DU RISQUE DE CANCER SECONDAIRE DU SEIN.	120
TABLEAU 19 - RESULTATS DES ANALYSES STRATIFIEES SELON LA DOSE REÇUE AUX DIFFERENTS ORGANES.	124

Liste des abréviations

BCCSS – British Childhood Cancer Survivors Study

CNRGH – Centre National de Recherche en Génomique Humaine

FCCSS – French Childhood Cancer Survivors Study

GWAS – Genome Wide Association Study (études pangénomiques)

IC – Intervalle de Confiance

NGS – Next Generation Sequencing

OR – Odds Ratio

RR – Risque Relatif

SC – Second Cancer

SIR – Ratio standardisé d'incidence

SMR – Taux de mortalité standardisé

SNC – Système Nerveux Central

SNDS – Système National des Données de Santé

WES – Whole Exome Sequencing

WGS – Whole Genome Sequencing

CHAPITRE 1

—

Introduction et objectifs

I. Les cancers pédiatriques

1. Épidémiologie des cancers pédiatriques

a. Incidence

Les données de 153 registres de cancers de 62 pays ont permis d'estimer qu'environ 400 000 cancers ont été diagnostiqués chez les enfants âgés de 0 à 19 ans entre 2001 et 2010.

Le cancer chez l'enfant est une maladie rare qui touche environ 1800 enfants de moins de 15 ans chaque année en France métropolitaine. La surveillance de ces cancers est assurée par le Registre National des Cancers de l'Enfant (RNCE), qui regroupe deux registres certifiés : le Registre national des hémopathies malignes de l'enfant (RNHE) et le Registre national des tumeurs solides de l'enfant (RTNSE), respectivement créés en 1990 et 2000. Entre 2014 et 2020, ces registres ont estimé une incidence annuelle standardisée à 166,6 cas par million avant l'âge de 15 ans, avec un sex-ratio garçons/filles de 1,1. On estime ainsi qu'environ 1 nouveau-né sur 425 sera atteint d'un cancer avant ses 15 ans.

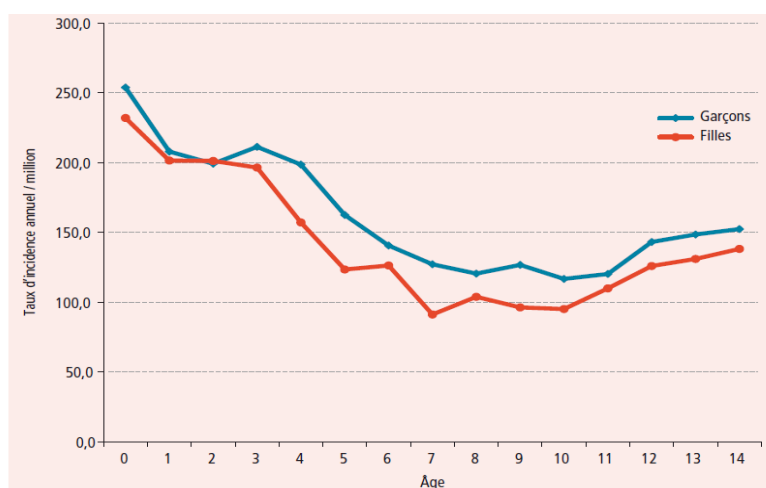


Figure 1 - Incidence des cancers de l'enfant selon l'âge et le sexe en France métropolitaine de 2000 à 2004. Source : Lacour B, Guyot Goubin A, Guissou S, Bellec S, Desandes E, Clavel J *Bulletin Épidémiologique Hebdomadaire*, 2010, n°. 49-50, p. 497-500 (Lacour et al., 2010).

Les cancers pédiatriques forment un groupe hétérogène de cancers, presque tous génétiquement et histologiquement distincts des cancers de l'adulte (Zahnreich & Schmidberger, 2021). Cette particularité souligne l'importance, reconnue depuis plusieurs décennies, d'étudier cette population séparément afin d'adapter le traitement et le suivi de manière spécifique.

Chez les enfants âgés de 0 à 4 ans, la leucémie représente le type de cancers le plus fréquent (36,1 % des cas) ; cette proportion diminue à 15,4 % chez ceux âgés de 15 à 19 ans. En revanche, les lymphomes sont moins fréquents chez les 0-4 ans (5,3 % des cas) et représentent un des principaux types de cancers chez les 15-19 ans (22,5 % des cas). Les tumeurs du système nerveux central (SNC) sont le deuxième type de tumeurs le plus fréquent chez les enfants de 0 à 4 ans (17,2 % des cas), de 5 à 9 ans (26,3 % des cas) et de 10 à 14 ans (20,0 % des cas) (Figure 2).

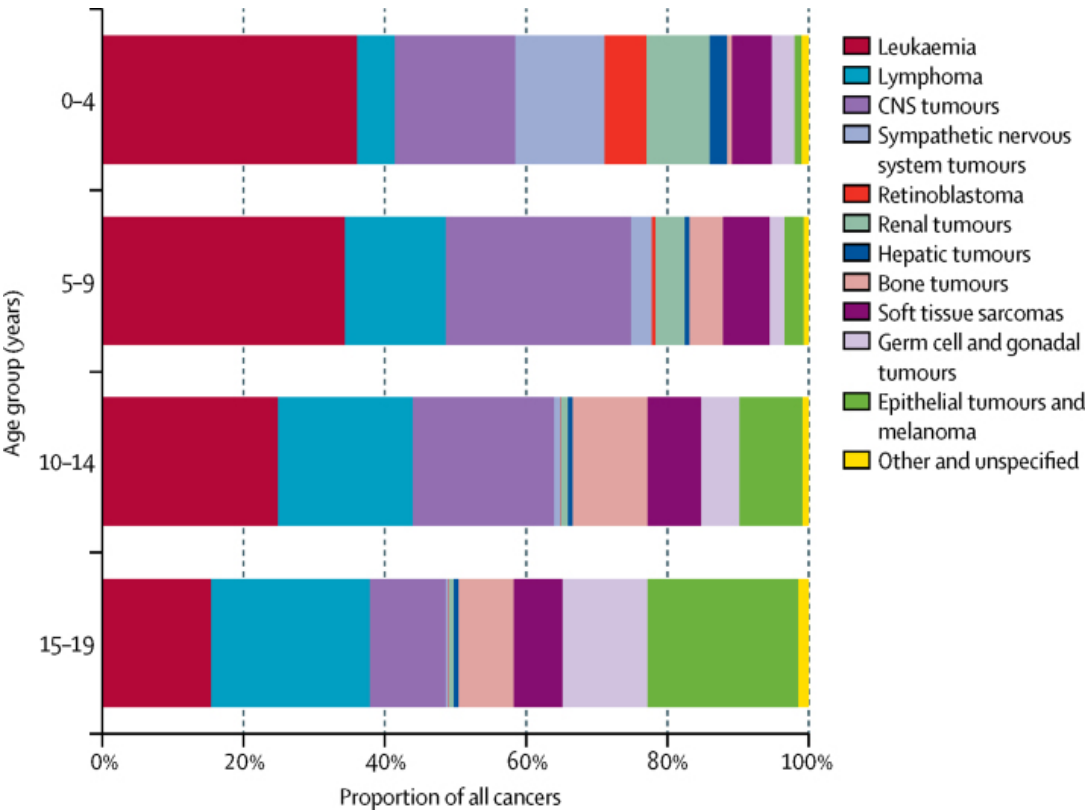


Figure 2 - Répartition des types de cancer par groupe d'âge sur la période 2001–2010. Source : (Steliarova-Foucher et al., 2017).

b. Taux de survie

Depuis quelques décennies, on observe une amélioration de la survie des patients guéris d'un cancer pédiatrique. En effet, en Europe, le taux de survie à 5 ans après un cancer pédiatrique pour tous types de tumeurs était de 54 % dans les années 1970-1980 (Magnani et al., 2006) et de 75 % au début des années 2000 (Gatta et al., 2014) (Figure 3).

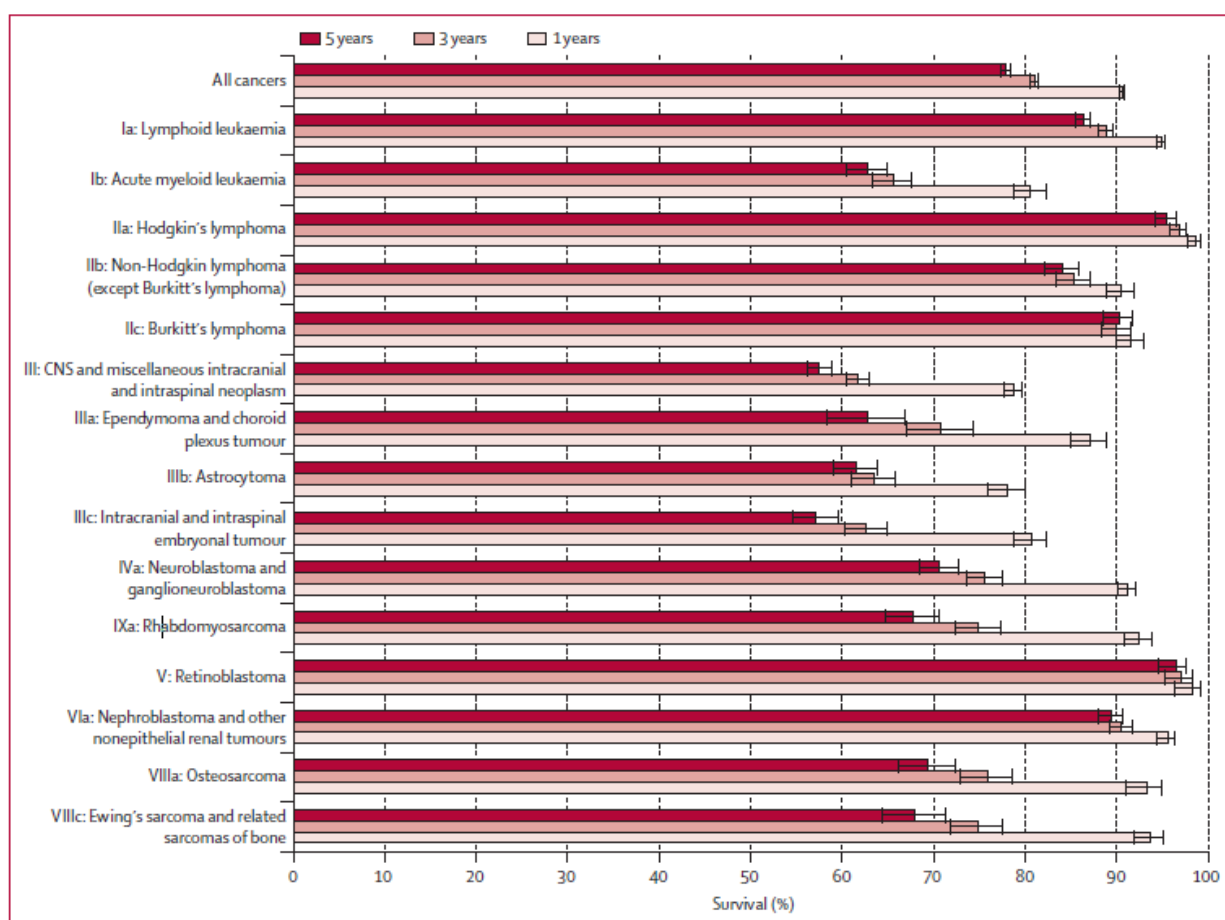


Figure 3 - Survie pondérée par pays selon le type de premier cancer pour les enfants européens diagnostiqués avec un cancer entre 2000 et 2007. Source : (Gatta et al., 2014). Les données incluent 57 956 cas. Les barres d'erreur représentent les intervalles de confiance à 95 %.

Pour tous les cancers combinés, la survie à 5 ans s'est améliorée de manière linéaire et significative au fil du temps, passant de 78 % (IC 95 % : 78 à 79) en 2004-2006 à 81 % (IC 95 % : 81 à 82) en 2010-2014 (Botta et al., 2022) (Figure 4).

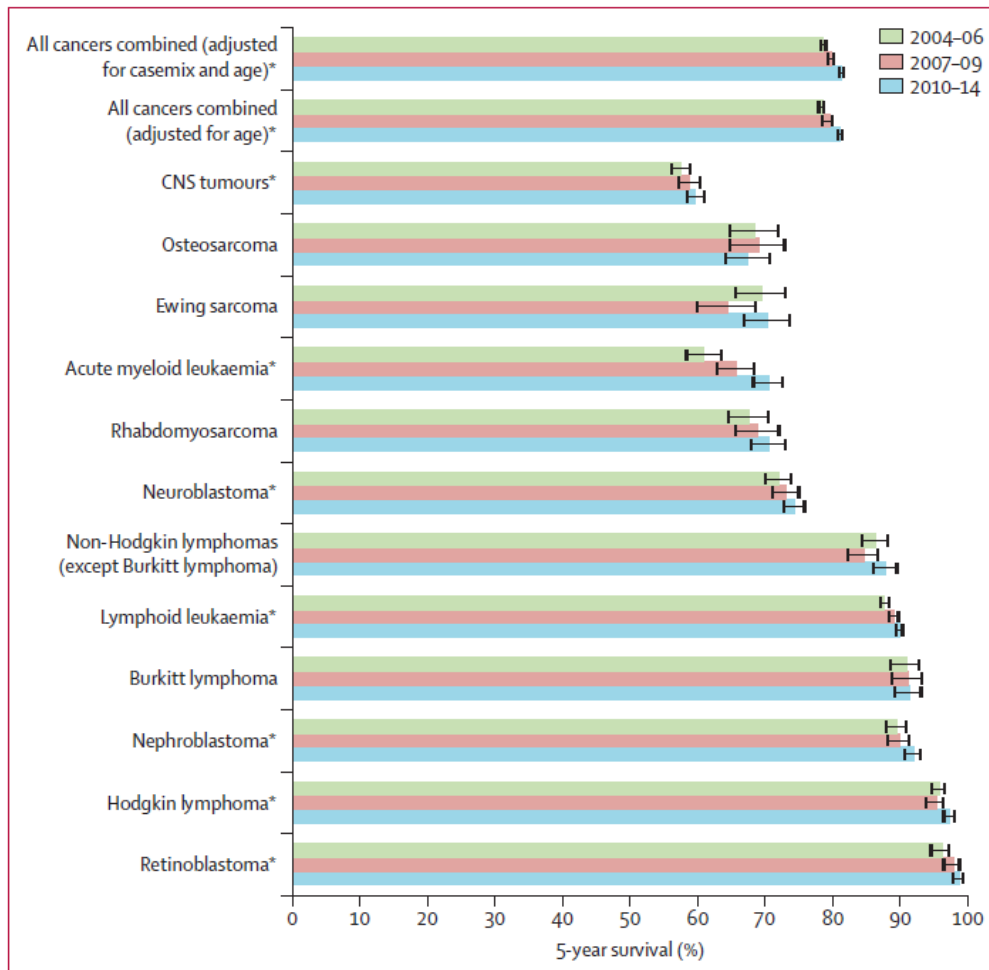


Figure 4 - Taux de survie à 5 ans chez les patients guéris d'un cancer pédiatrique de 2004 à 2014.
 Source : (Botta et al., 2022)

Cependant, la survie à 5 ans varie selon les types de cancers pédiatriques les plus fréquemment rencontrés (Botta et al., 2022; Lacour et al., 2014) : les taux de survie sont très élevés chez les enfants atteints de lymphome ou de rétinoblastome et plus faibles chez ceux atteints de tumeur du SNC, d'ostéosarcome, ou de sarcome d'Ewing. Pour ces derniers, peu de progrès thérapeutiques ont été réalisés au cours des deux dernières décennies, en grande partie à cause de la résistance aux médicaments et au nombre limité d'essais cliniques pour ces tumeurs rares et complexes (Butler et al., 2021).

2. Traitements des cancers pédiatriques

La survie des cancers pédiatriques a été considérablement améliorée grâce aux avancées dans les techniques d'imagerie, une meilleure compréhension de la biologie du cancer et l'amélioration des traitements.

Le traitement des différents types de cancers pédiatriques varie considérablement en termes de modalités, d'approche et d'intensité. En général, la plupart des enfants atteints de cancer sont traités par chirurgie, chimiothérapie, radiothérapie, une combinaison de deux ou plusieurs de ces modalités selon des protocoles standardisés (Meadows, 2003). Ces traitements évoluent progressivement vers des modalités plus ciblées, dans l'objectif de réduire la toxicité à long terme tout en améliorant la survie.

Ces traitements étaient d'abord chirurgicaux et locaux. Puis, l'introduction de la radiothérapie, a permis de minimiser le risque de dissémination. Ils ont ensuite été complétés par la chimiothérapie.

À l'origine, les champs de radiothérapie étaient conçus à partir de radiographies en deux dimensions et des repères anatomiques, ce qui entraînait l'utilisation de larges marges autour du site de la tumeur, incluant parfois des organes sains, et donc un risque élevé de dommages aux tissus environnants (Purdy, 2008). Cela a maintenant bien évolué grâce à deux facteurs importants : premièrement, le fractionnement des doses consistant à administrer des doses relativement faibles plusieurs fois sur plusieurs semaines ou mois ; deuxièmement, la radiothérapie par modulation d'intensité (IRMT) ou bien la protonthérapie qui sont des traitements de haute précision permettant de cibler les tumeurs, et certaines zones au sein des tumeurs (Connell & Hellman, 2009; Dixon et al., 2018). Divers facteurs influencent les résultats cliniques, notamment la source de radiation, la dose cumulée, le volume irradié, le fractionnement des doses et des facteurs démographiques tels que le sexe et l'âge lors de l'exposition. L'exposition aux radiations peut entraîner des effets secondaires à long terme dont les détails sont donnés dans la suite de ce chapitre.

L'utilisation de la chimiothérapie a été largement répandue à partir des années 1970, notamment dans le traitement des leucémies et de certains lymphomes (Shen & Terezakis, 2021). Aujourd'hui, la chimiothérapie tient un rôle important dans le traitement des cancers de l'enfant ; elle constitue l'un des traitements de première ligne de certains cancers et est utilisée en complément de la chirurgie ou de la radiothérapie dans d'autres cancers. En général, différentes molécules de chimiothérapies sont combinées pour cibler divers mécanismes d'action et obtenir un effet synergique. Une partie importante des patients sont maintenant traités avec des agents alkylants, des anthracyclines, des épipodophyllotoxines ou des médicaments à base de platine (Turcotte et al., 2017).

Bien que la radiothérapie et la chimiothérapie dans les stratégies de traitement aient permis une meilleure survie des patients, ces traitements peuvent entraîner des effets secondaires à long terme, tels que des cardiomyopathies ou l'apparition de seconds cancers (Lipshultz et al., 2014; Teepen et al., 2017).

3. Complications après traitement pour un cancer pédiatrique

Comme cela a été dit au début de ce chapitre, on observe une augmentation du nombre de patients survivant à un cancer pédiatrique. En contrepartie, on constate également une hausse significative des effets iatrogènes à long terme chez ces patients, dont certains peuvent avoir un impact majeur sur la mortalité et la morbidité. Ainsi, 60 % à 90 % des patients survivant développeront au moins une maladie chronique et 20 % à 80 % présenteront des complications graves ou potentiellement mortelles à l'âge adulte (Bhakta et al., 2017; Oeffinger Kevin C. et al., 2006).

a. Les différents effets iatrogènes

Afin d'étudier ces effets iatrogènes, plusieurs cohortes ont été mises en place, principalement en Europe et aux États-Unis, parmi lesquelles la « French Childhood Cancer Survivors » (FCCSS), la « COHort of the PEdiatric cancer Registries » (COHOPER), la « British Childhood Cancer Survivors Study » (BCCSS), la « Childhood

Cancer Survivor Study » (CCSS), la « St-Jude Lifetime cohort » (SJLIFE) et la DCOG-LATER. Toutes ces études incluent des patients guéris d'un cancer pédiatrique et suivis sur le long terme. Elles ont permis de mener de nombreux travaux démontrant que les effets iatrogènes sont nombreux et divers, et impactent plus ou moins la vie des patients.

Ainsi, les patients guéris d'un cancer pédiatrique sont confrontés à de nombreux problèmes de santé à long terme (illustrés sur la Figure 5) qui touchent leur croissance (Demoor-Goldschmidt et al., 2020), leur développement, leur fertilité (Green et al., 2009; Servitzoglou et al., 2015), leur santé mentale (de Laage et al., 2016), leur éducation (Dumas et al., 2016) et leur emploi (Dumas et al., 2016). Les traitements reçus lors du premier cancer peuvent également entraîner, entre autres, des problèmes cardiaques (Armstrong et al., 2013; Guldner et al., 2006), endocriniens (Chemaitilly & Hudson, 2014; de Vathaire et al., 2012), pulmonaires (Dietz et al., 2016) et neurologiques (El-Fayech et al., 2017; Haddy et al., 2011), ainsi que plusieurs types de seconds cancers (Allodji et al., 2019, 2021; Bhatti et al., 2010; Ghosh et al., 2021; Henderson et al., 2012), détaillés dans la partie suivante.

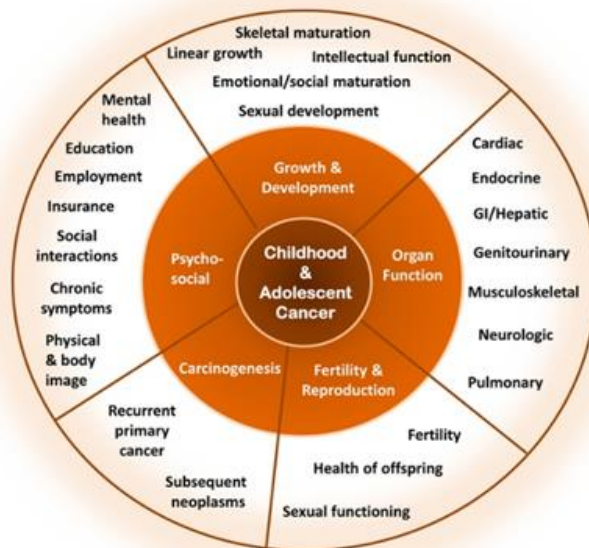


Figure 5 - Ensemble des effets à long-terme retrouvés chez les survivants d'un cancer pédiatrique.
Source : (Robison & Hudson, 2014)

b. Les facteurs de risque

Il a été mis en évidence plusieurs facteurs influençant le risque d'effets à long terme, que l'on peut regrouper en trois grandes classes : la localisation du cancer, les facteurs liés au traitement (type de chimiothérapie, dose de radiothérapie, combinaison de traitements, etc.) et les caractéristiques cliniques du patient (sexe, problèmes de santé, âge au moment du diagnostic et du traitement, durée depuis le diagnostic ou le traitement, antécédents familiaux, statut socio-économique et facteurs génétiques) (Dixon et al., 2018).

La partie suivante se concentrera sur les seconds cancers qui constituent le sujet principal de cette thèse.

II. Les seconds cancers

1. Épidémiologie

Les seconds cancers figurent parmi les effets secondaires les plus graves après le traitement d'un cancer pédiatrique. Ils représentent, après la récurrence ou la progression du cancer initial, la principale cause de surmortalité tardive (Fidler et al., 2016). Ces cancers secondaires sont aussi la cause principale de mortalité liée au traitement des cancers pédiatriques avec un taux de mortalité standardisé (SMR) de 15,2 ; [IC 95 % : 13,9-16,6] (Armstrong et al., 2009). Des études antérieures ont montré que le risque de second cancer à l'âge adulte est 6 à 10 fois plus important que celui de la population générale (Friedman et al., 2010; Teepen et al., 2017; Varan & Kebudi, 2011) et ce risque augmente avec la durée de suivi des patients. En effet, le taux d'incidence cumulé à 30 ans pour tous types de seconds cancers confondus atteint 20,5 % [IC 95 % : 19,1 %–21,8 %] (Friedman et al., 2010) et est de 34,6% à 40 ans [IC 95 % : 28,7-40,6] (Figure 6).

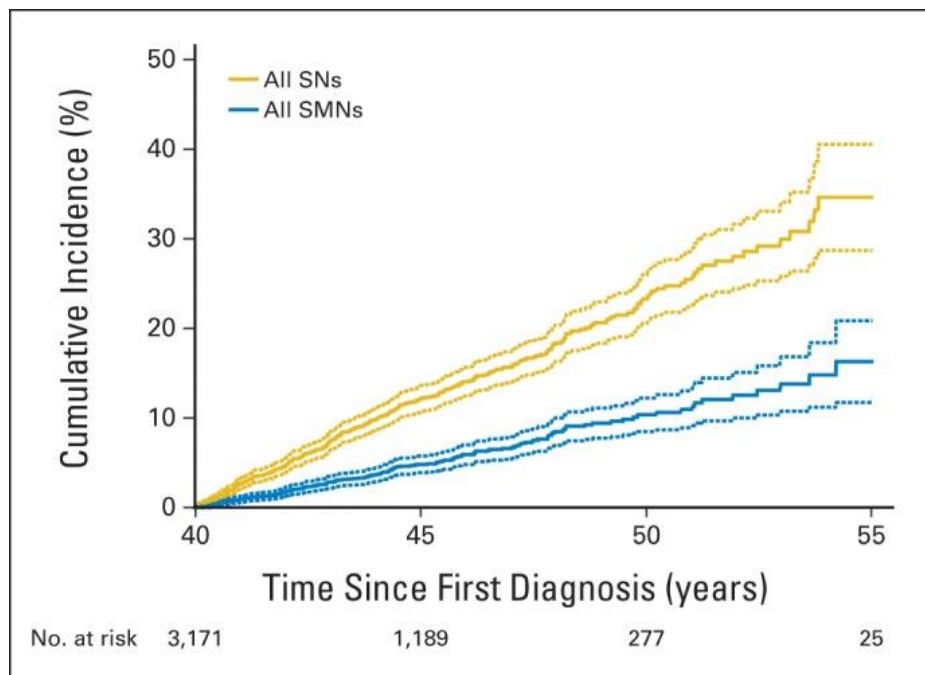


Figure 6 - Incidence cumulée des seconds cancers dans la cohorte CCSS chez les patients de plus de 40 ans. Source : (Turcotte et al., 2015). Les lignes pleines représentent les valeurs de l'incidence cumulée calculée ; les lignes pointillées représentent les intervalles de confiance à 95 %.

L'incidence ainsi que le type de second cancer dépendent principalement de trois facteurs incluant le type de premier cancer, le type de traitement reçu et finalement la génétique qui sera abordée plus largement dans la suite du manuscrit.

Les cancers du sein et de la thyroïde sont les seconds cancers les plus couramment retrouvés (Turcotte et al., 2017).

L'incidence cumulée d'un second cancer du sein est de 8 % à l'âge de 45 ans dans une étude incluant plus de 11 000 survivantes (Figure 7). Cela représente un risque six fois supérieur (Ratio d'incidence standardisé [SIR] : 6,6; IC 95 % : 6,1-7,2) par rapport à celui de la population (Henderson et al., 2022). On le retrouve principalement chez les patientes ayant été traitées avec de la radiothérapie au niveau de la poitrine, le plus souvent pour le traitement d'un lymphome hodgkinien (Schaapveld et al., 2015). Les patientes n'ayant reçu aucune irradiation ont toutefois un risque 4 fois plus important que la population générale (Henderson et al., 2016).

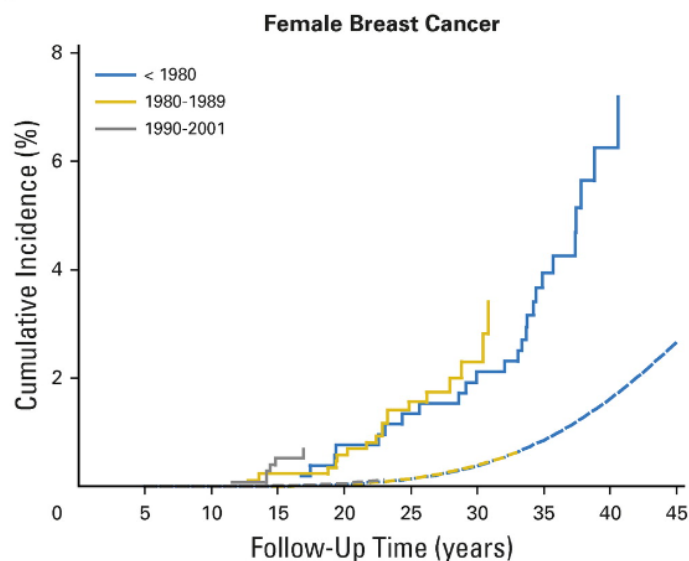


Figure 7 - Incidence cumulée du second cancer du sein dans la cohorte DCOG LATER. *Source : (Teepen et al., 2017). Les lignes pleines correspondent à l'incidence au sein de la cohorte et celles en pointillées l'incidence attendue dans la population générale.*

Le risque de second cancer de la thyroïde est lui aussi particulièrement important chez les survivants d'un cancer pédiatrique avec un risque plus de dix fois supérieur à celui de la population générale (SIR :10,5 ; 95 % IC : 9,1–12)(Turcotte et al., 2017).

2. Facteurs de risque de second cancer

Les seconds cancers pouvant survenir à la suite du traitement d'un cancer pédiatrique sont classifiés en fonction de l'exposition à certains traitements formant ainsi deux principaux types : les hémopathies incluant la leucémie aigüe myéloïde et les syndromes myélodysplasiques, et les seconds cancers solides.

Le risque de leucémie est principalement accru durant les dix premières années suivant le traitement du cancer de l'enfance (Nottage et al., 2011) et sont fortement associées aux types de chimiothérapie reçue lors du premier cancer, tels que les agents alkylants et les inhibiteurs de topoisomérase (Thirman & Larson, 1996). En revanche, le risque de cancer solide, constituant 80 % des seconds cancers présents chez les survivants, est légèrement augmenté durant les dix premières années après le diagnostic et reste élevé des décennies après le diagnostic de cancer de l'enfance (Turcotte et al., 2017).

Parmi les facteurs de risque de second cancer, on retrouve notamment l'âge auquel a été diagnostiqué le cancer pédiatrique, le type de premier cancer ou encore le sexe (Figure 8). Un des principaux facteurs largement étudiés est le traitement reçu lors du cancer pédiatrique. La suite de ce chapitre s'intéresse donc au risque de seconds cancers associé à la radiothérapie et à la chimiothérapie.

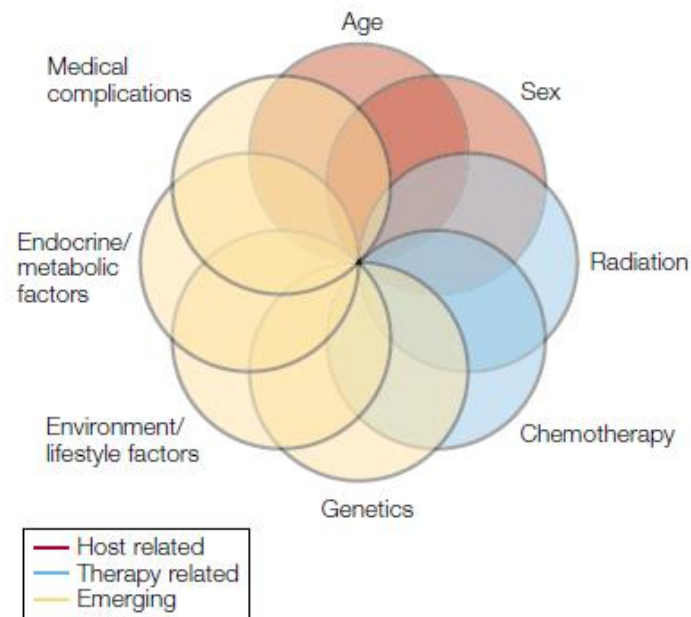


Figure 8 - Ensemble des facteurs de risque de second cancer. Source : (Bhatia & Sklar, 2002)

a. La radiothérapie

La radiothérapie, bien qu'étant un traitement efficace contre le cancer, est reconnue comme un important facteur de risque dans le développement de cancers secondaires après le traitement d'un cancer pédiatrique. En effet, la radiothérapie utilise des rayonnements ionisants pouvant endommager les cellules normales entourant la tumeur et entraîner des mutations si les cellules ne parviennent pas à réparer correctement ces dommages. Les patients de jeune âge sont d'autant plus sensibles aux effets de la radiothérapie car leurs cellules se divisent plus rapidement et leurs tissus sont encore en développement.

Différentes études ont permis d'établir des relations dose-effet pour divers types de

seconds cancers, notamment le cancer du sein, de la thyroïde, le cancer colorectal, les sarcomes, les tumeurs du système nerveux central et le cancer de la peau.

La littérature scientifique a mis en évidence une relation généralement linéaire entre la dose de radiation reçue et le risque de développer la plupart des seconds cancers solides (Figure 9).

Cependant, le cancer de la thyroïde présente une exception notable à cette relation linéaire. Pour ce cancer, le risque augmente avec des doses allant jusqu'à 15-20 Gy, puis diminue paradoxalement lorsque la dose continue d'augmenter. L'hypothèse privilégiée pour expliquer ce phénomène est que des doses plus élevées de radiation conduisent à la mort des cellules thyroïdiennes plutôt qu'à leur transformation cancéreuse (Inskip et al., 2016). Cette observation souligne la complexité des mécanismes biologiques impliqués dans le développement des seconds cancers.

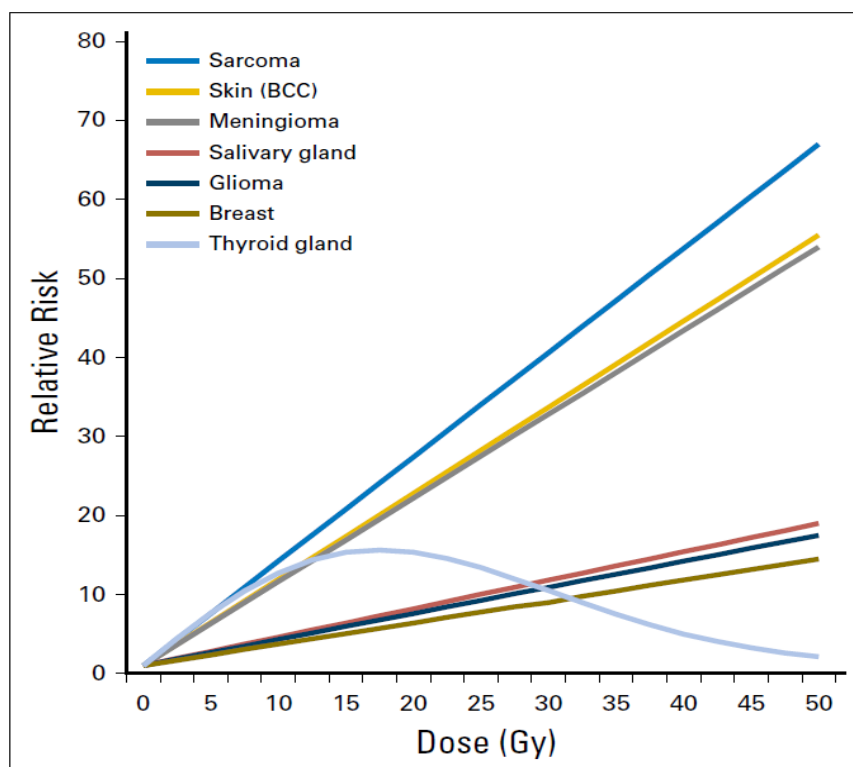


Figure 9 – Relation dose-effet sur le risque de seconds cancers après irradiation durant la radiothérapie. Source : (Inskip et al., 2016).

b. La chimiothérapie

Plusieurs études ont également étudié le risque de seconds cancers associé aux différentes molécules de chimiothérapie reçues lors du traitement du cancer pédiatrique, tels que les agents alkylants ou les anthracyclines (Teepen et al., 2017; Turcotte et al., 2019).

Les agents alkylants sont également associés à un risque accru de sarcomes chez les survivants de cancers pédiatriques (Teepen et al., 2017), de cancer du poumon (Travis et al., 2002) et de l'estomac (Morton et al., 2013) chez les survivants d'un lymphome hodgkinien. Des études récentes ont également mis en évidence un lien entre le risque de second cancer de la thyroïde et l'administration d'agents alkylants chez les patients n'ayant pas ou peu reçu de radiothérapie (de Vathaire, Haddy, et al., 2015; Veiga et al., 2012).

Une exposition importante à des doses de chimiothérapie à base de platine supérieures à 750 mg/m^2 a aussi été associée à une augmentation du risque de tous types de cancers secondaires (RR = 2,7).

Les anthracyclines ont également été identifiées comme augmentant spécifiquement le risque de cancer du sein en fonction de la dose reçue (RR = 1,3 / 100 mg/m^2) (Henderson et al., 2016) ainsi que les sarcomes (Henderson et al., 2012).

c. La génétique

Les données disponibles montrent donc clairement que la chimiothérapie et la radiothérapie jouent un rôle important dans l'apparition de cancers secondaires. Cependant, cette relation varie d'un individu à l'autre, ce qui suggère que des variations génétiques peuvent influencer la sensibilité aux traitements génotoxiques.

Les avancées génomiques des dernières décennies ont permis d'améliorer notre compréhension de la génétique des cancers. De manière générale, l'impact génétique peut être attribué à des variants extrêmement rares et très pénétrants, souvent liés à

des syndromes de susceptibilité au cancer, jusqu'à des variants génétiques plus communs associés à un risque de cancer dans la population générale. En effet, certains syndromes de prédisposition au cancer, bien caractérisés avec une forte pénétrance, comme le syndrome de Li-Fraumeni, peuvent, à l'instar des traitements, augmenter le risque de second cancer suggérant une possible interaction entre les traitements anti-cancéreux et les prédispositions génétiques des patients (Henderson et al., 2016; Teepen et al., 2017).

Cependant, les connaissances sur les variants génétiques en dehors des gènes liés aux syndromes génétiques et associés aux cancers secondaires restent encore très limitées. La majorité des recherches antérieures se sont concentrées sur des variants génétiques dans les mécanismes de détection et de réparation des dommages à l'ADN comme modificateurs potentiels des risques de seconds cancers liés aux traitements (Bhatia, 2015). Toutefois, ces études sont limitées par de petites tailles d'échantillons, ou des données d'exposition aux traitements insuffisantes.

III. Hypothèse de recherche et objectifs

1. Objectifs de la thèse

L'objectif global de cette thèse est de contribuer à une meilleure connaissance des facteurs de risque de seconds cancers, d'identifier les variants génétiques rares associés à ce risque et d'évaluer les interactions entre les variants et les traitements. Ces travaux ont été réalisés à partir des données et des échantillons biologiques collectés dans le cadre de la cohorte French Childhood Cancer Survivors Study (FCCSS), ainsi que des données de la British Childhood Cancer Survivors Study (BCCSS).

Les trois objectifs principaux de la thèse sont les suivants :

- Étudier le risque de seconds cancers de la thyroïde chez les survivants d'un cancer de l'enfance, lié à l'exposition aux radiations reçues aux organes

lymphoïdes dans deux cohortes (FCCSS et BCCSS).

- Réaliser une revue de littérature sur les facteurs génétiques associés aux seconds cancers chez les survivants à un cancer dans l'enfance.
- Identifier des variants rares associés aux seconds cancers dans une étude cas-témoins nichée dans la cohorte FCCSS.

2. Plan des chapitres de la thèse

Pour répondre aux objectifs, cette thèse est structurée en sept chapitres.

Le chapitre 1, constituant une introduction au sujet de cette thèse, décrit dans un premier temps l'épidémiologie des cancers pédiatriques, leur incidence ainsi que les risques à long terme pour les patients guéris d'un premier cancer. Ensuite, ce chapitre s'intéresse à l'un des aspects iatrogènes à long terme les plus importants pour cette population, à savoir les seconds cancers, en se concentrant principalement sur leurs principaux facteurs de risque.

Le chapitre 2 décrit les données de la cohorte FCCSS, incluant les informations cliniques, les détails des traitements reçus ainsi que la banque biologique associée à cette cohorte, qui ont été utilisés pour les deux études de cette thèse (Chapitres 4 & 6). Ce chapitre présente également, de manière plus concise, la cohorte BCCSS dont les données cliniques et de traitements ont été utilisées conjointement avec celles de la FCCSS pour la première étude de cette thèse (chapitre 4).

Le chapitre 3 détaille l'ensemble des méthodes utilisées dans les différentes études de cette thèse. Ce chapitre est divisé en trois grandes parties, la première visant à décrire les différentes méthodes d'études d'association génétiques existantes. La seconde partie décrit les différents traitements bio-informatiques appliqués à des données de séquençage d'exome. La dernière partie détaille les méthodes statistiques spécifiques à l'étude des variants rares.

Le chapitre 4 présente une étude évaluant les facteurs de risque des cancers de la

thyroïde, l'un des cancers solides les plus courants parmi les survivants de cancers pédiatriques. Cette étude cas-témoin, incluant des patients des cohortes FCCSS et BCCSS, vise à déterminer l'impact de l'irradiation des organes lymphoïdes, tels que la rate, le thymus et la moelle osseuse, sur le risque de développer un second cancer de la thyroïde.

Le chapitre 5 décrit les résultats de la revue systématique de la littérature sur les études d'association génétique évaluant le risque de seconds cancers chez les patients ayant été traités pour un cancer pédiatrique. Cette dernière permet ainsi de regrouper les gènes, variants et voies biologiques associés au risque de seconds cancers dans les études antérieures. Les résultats sont détaillés en fonction des différentes méthodes d'association et de séquençage utilisées ainsi que selon les sous-types de seconds cancers étudiés.

Le chapitre 6 présente les résultats de l'étude d'association génétique. Cette étude cas-témoins nichée dans la FCCSS et basée sur des données de séquençage de l'exome, a pour objectif d'identifier des variants rares associés au risque de second cancer chez les patients guéris d'un cancer pédiatrique. Nous avons particulièrement examiné les variants présents dans trois voies biologiques clés : la réplication, la recombinaison et la réparation de l'ADN.

Le chapitre 7 fournit une discussion générale des résultats de cette thèse, en résumant les principales découvertes et en mettant en évidence leur contribution à la compréhension des seconds cancers chez les survivants de cancers pédiatriques. Il aborde également les implications cliniques de ces résultats et propose des pistes pour des recherches futures.

CHAPITRE 2

-

Données des études – Les cohortes French Childhood Cancer Survivors Study (FCCSS) et British Childhood Cancer Survivors Study (BCCSS)

Ce chapitre est consacré à la description générale de la cohorte French Childhood Cancer Survivors Study (FCCSS) utilisée pour répondre aux différents objectifs de la thèse. La cohorte British Childhood Cancer Survivors Study a aussi été utilisée pour l'étude en chapitre 4, cette dernière sera donc également présentée plus brièvement dans ce chapitre. Les données plus spécifiques aux différents objectifs sont détaillées dans les chapitres correspondants.

I. La cohorte « French Childhood Cancer Survivors Study » (FCCSS)

1. Définition de la cohorte

La cohorte « French Childhood Cancer Survivors Study » (FCCSS, www.fccss.fr) comprend 7 670 patients ayant survécu au moins 5 ans après avoir été traités avant leur 20 ans pour un cancer solide ou un lymphome entre 1942 et 2000. La FCCSS est l'extension de la partie française de la cohorte rétrospective EURO2K qui incluait initialement des patients traités entre 1942 et 1985 dans 5 centres spécialisés en oncologie en France (*Gustave Roussy, Villejuif; Institut Curie, Paris; Centre Claudius Regaud, Toulouse; Institut Jean Godinot, Reims; Centre Antoine Lacassagne, Nice*) et son extension, correspondant aux patients traités entre 1986 et 2000. L'étude FCCSS a été approuvée par le comité d'éthique de l'Institut National de la Santé et de la Recherche Médicale (INSERM) ainsi que par la Commission Nationale de l'Informatique et des Libertés (CNIL).

L'objectif de cette cohorte est d'étudier les effets à long terme liés au cancer de l'enfant et à ses traitements dans tous leurs aspects – médicaux, sociaux, économiques et psychologiques – afin d'en améliorer leur gestion et leur prévention. Cette cohorte a conduit à de nombreuses études sur les divers effets iatrogènes pouvant survenir après le traitement d'un cancer pédiatrique, parmi lesquels les pathologies cardiaques (Haddy et al., 2016) ou cérébrovasculaires (El-Fayech et al., 2017), le diabète (de Vathaire et al., 2012) et les seconds cancers : leucémies (Allodji et al., 2021; Haddy et al., 2006), cancers de la thyroïde (de Vathaire, Haddy, et al., 2015), des reins (de Vathaire, Scwhartz, et al., 2015), du sein (Haddy et al., 2010), ou bien les cancers digestifs (Tukenova et al., 2012).

2. Les données de la FCCSS

Les données démographiques, les caractéristiques cliniques ainsi que les informations détaillées sur les traitements délivrés pour le cancer pédiatrique ont été recueillies à partir des dossiers médicaux des patients. Les caractéristiques des patients inclus peuvent être trouvés dans le Tableau 1.

Tableau 1 - Caractéristiques des patients inclus dans la FCCSS.

Caractéristiques	n = 7670	%
Sexe		
Homme	4201	54,8
Femme	3469	45,2
Centres		
Institut Gustave Roussy	5504	71,8
Institut Curie	1917	25
Centre Claudius Regaud	159	2,07
Institut Jean Godinot	75	0,98
Centre Antoine Lacassagne	15	0,2
Année du premier cancer		
<1970	658	8,58
1970–1979	1549	20,2
1980–1989	2484	32,4
>=1990	2979	38,8
Age au premier cancer		
0–1	1727	22,5
2–4	1783	23,2
5–9	1736	22,6
10–14	1609	21
≥15	815	10,6
Type de premier cancer		
Lymphome	1279	16,7
Tumeurs du système nerveux central	1138	14,8
Tumeur du rein	1128	14,7
Neuroblastome	1035	13,5
Sarcome des tissus mous	819	10,7
Ostéosarcomes	701	9,14
Rétinoblastome	619	8,07
Autres types de cancers solides	512	6,68
Tumeurs gonadiques/tumeurs des cellules germinales	439	5,72
Traitement		
Pas de RT et pas de CT	905	11,8
RT	1089	14,2
CT	2569	33,5
RT et CT	3107	40,5
Suivi		
Médiane (min-max)	24,4 (5,0-67,8)	
Statut Vital en 2022		
Vivants	6471	84,4
Décédés	1199	15,6

Abréviations : CT, Chimiothérapie; n, nombre de patients inclus dans la FCCS, RT, Radiothérapie..

a. Radiothérapie

Pour chaque patient ayant reçu un traitement par radiothérapie, la dose de rayonnement ionisant reçue par la plupart des organes a été reconstituée par les physiciens de l'équipe. Pour cela, de nombreux paramètres ont été recueillis et intégrés pour estimer les doses d'irradiation et les histogrammes dose-volume, c'est-à-dire les doses reçues par différents pourcentages de volumes d'organes concernés. Pour cette estimation des doses sur l'ensemble des organes du corps, que ce soit dans le champ de traitement, à proximité ou à distance, chaque patient a été apparié à un fantôme anthropomorphe voxelisé en prenant en compte la position de traitement (Figure 10) (Veres et al., 2014).

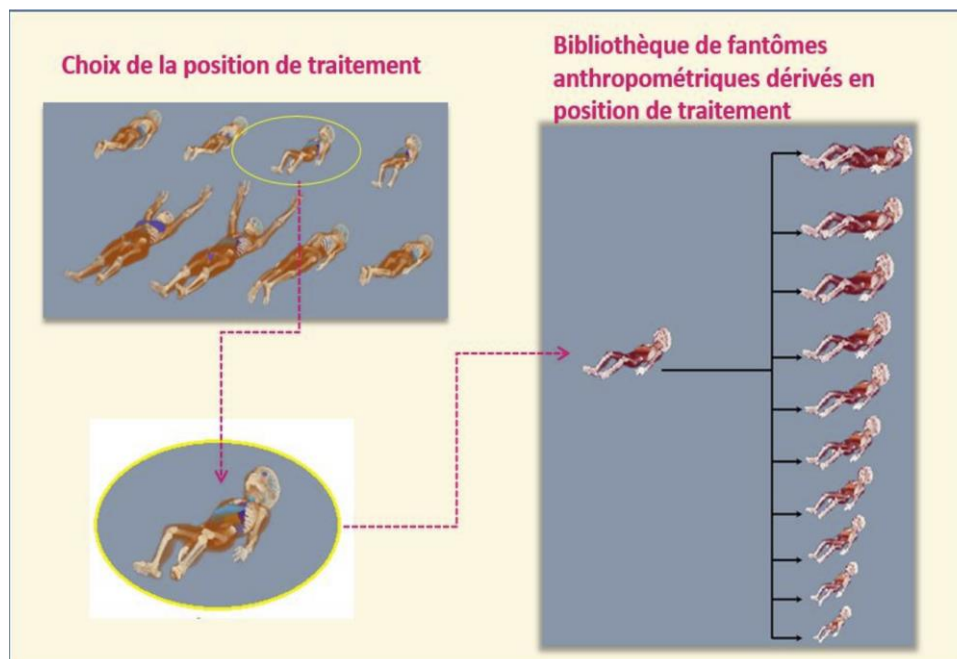


Figure 10 - Illustration des fantômes anthropométriques pour estimer la dose reçue par différents organes des patients de la FCCSS.

b. Chimiothérapie

Les données sur les traitements par chimiothérapie ont été extraites des dossiers médicaux. Des données qualitatives binaires (oui/non) ont été recueillies, ainsi que les doses cumulées de chaque produit cytotoxique (calculées en grammes par mètre carré

de surface corporelle). Les molécules de chimiothérapie sont classées en fonction de leurs mécanismes d'action connus dans la cellule : les agents alkylants, épipodophyllotoxines, anthracyclines, vinca-alcaloïdes, les sels de platine, et les antimétabolites. La description des médicaments de chimiothérapie dans la cohorte FCCSS est présentée dans le Tableau 2.

Tableau 2 - Description des molécules de chimiothérapie dans la cohorte FCCSS.

Classes de chimiothérapie	Nom de la molécule	Cohorte (n=7670)	
		N	(%)
Antimétabolites	Methotrexate	1135	(14,8)
	Cytarabine	757	(9,9)
	Thioguanine	191	(2,5)
	Hydroxyurea	230	(3,0)
	Autres	184	(2,4)
Agents alkylants	Cyclophosphamide	2813	(36,7)
	Carboplatine	1108	(14,4)
	Cisplatine	1052	(13,7)
	Ifosfamide	945	(12,3)
	Procarbazine	715	(9,3)
	Melphalan	351	(4,6)
	Lomustine-CCNU	298	(3,9)
	Caryolisine	287	(3,7)
	Carmustine-BCNU	245	(3,2)
	Busulfan	228	(3,0)
	Autres	214	(2,7)
	Dacarbazine	172	(2,2)
	Inhibiteurs de topoisomérase II	Actinomycine	2162
Etoposide		1585	(20,7)
Autres		304	(3,9)
Anthracyclines	Adriamycine	2485	(32,4)
	Epiadriamycine	231	(3,0)
	Autres	154	(2,0)
Vinca-alcaloïdes	Vincristine	4098	(53,4)
	Vinblastine	487	(6,3)
	Autres	166	(2,2)
Enzymes Cytotoxique	Asparginase	315	(4,1)
Inhibiteurs de Topoisomérase I	Irinotecan	29	(0,4)
Inhibiteurs de tyrosine kinase	Imatinib	11	(0,1)
Autres / Non-spécifié	Bleomycine	550	(7,2)
	Autres	38	(0,5)

3. Suivi et identification des seconds cancers dans la FCCSS

Les patients de la FCCSS sont suivis prospectivement jusqu'à aujourd'hui, grâce à l'accès aux dossiers médicaux. Entre 2005 et 2010, un auto-questionnaire a été envoyé

aux patients dont les adresses étaient connues ou avaient été trouvées par la Caisse Nationale d'Assurance Maladie (CNAM) et ayant consenti à la recherche. Depuis 2012, un programme de suivi à long terme (LTFU) a été mis en place à l'Institut Gustave Roussy et à l'Institut Curie. Ensuite, grâce à un décret en Conseil d'État la cohorte FCCSS est appariée aux données du Système National de Données de Santé (SNDS) depuis 2006.

Les pathologies secondaires auto-déclarées, font l'objet d'une validation par contact des médecins traitants, des services médicaux spécialisés et des laboratoires de biologie et d'anatomopathologie.

4. Les données biologiques

Une banque biologique d'ADN constitutionnel a été constituée, au sein de la cohorte FCCSS. À ce jour, un total de 2774 personnes ont eu un échantillon de salive et/ou de sang prélevé et stocké dans la bio-banque nommée Bio_F. Parmi ces 2774 patients, 331 ont développé au moins un second cancer validé.

Les prélèvements salivaires ont été réalisés à l'aide de kits Oragene™, conçus pour la collecte de l'ADN salivaire, et envoyés au Centre d'Étude du Polymorphisme Humain (CEPH) pour l'extraction de l'ADN. Les prélèvements sanguins, quant à eux, ont été effectués soit dans un laboratoire d'analyses médicales, soit lors des consultations de suivi à long terme à l'Institut Gustave Roussy (IGR). Ces échantillons ont ensuite été transférés au Généthon, où l'isolation des lymphocytes et l'établissement des lignées de lymphoblastes ont été réalisés, ainsi que l'extraction de l'ADN et du sérum, avec un stockage des échantillons et des produits dérivés.

Ces lignées lymphoblastiques, correspondent à des cultures de cellules lymphocytaires transformées qui ont acquis la capacité de se diviser indéfiniment en laboratoire permettant ainsi d'avoir une source d'ADN stable et reproductible pour les analyses génétiques telles que le séquençage d'exome ou de génome.

Une fois les lignées établies, un contrôle de qualité est effectué pour s'assurer de l'absence de contamination par des mycoplasmes. L'extraction de l'ADN de ces lignées est réalisée avec un automate Autopure LS (Qiagen), utilisant la méthode de « *salting-out* ». Cette technique, basée sur la précipitation de l'ADN par des sels solubles, permet d'obtenir de l'ADN purifié de haute qualité, prêt pour le séquençage.

II. La cohorte « British Childhood Cancer Survivor Study » (BCCSS)

La cohorte British Childhood Cancer Survivors Study (BCCSS) comprend 17 981 individus qui ont été diagnostiqués avec un cancer pédiatrique avant l'âge de 15 ans entre 1940 et 1991 en Grande-Bretagne et qui ont survécu au moins 5 ans après leur diagnostic (Hawkins et al., 2008). Les participants ont été identifiés à l'aide du National Registry of Childhood Tumours (NRCT), un registre national qui contient des données de population sur les cancers pédiatriques diagnostiqués dans toute la Grande-Bretagne depuis 1962 (Stiller et al., 1995). L'étude BCCSS a été approuvée par un Comité d'éthique de la recherche multicentrique ainsi que par chaque Comité d'Éthique de la Recherche Local en Grande-Bretagne.

Cette cohorte, tout comme la FCCSS, a pour but d'étudier les effets à long terme chez les survivants de cancers pédiatriques, notamment les seconds cancers et les pathologies cardiaques.

Récemment, la cohorte a été étendue pour inclure 16 509 individus supplémentaires diagnostiqués entre 1992 et 2006, répondant aux mêmes critères que la cohorte initiale.

La BCCSS est connectée aux registres centraux du NHS (NHSCR ; National Health Service Central Register), qui regroupent les registres nationaux des cancers et des décès. Cela permet d'identifier les survivants ayant développé un second cancer. Une

plus petite proportion de ces seconds cancers a également été détectée via des questionnaires envoyés aux patients. Tous les nouveaux cas de cancer ont été confirmés par des analyses histopathologiques ou des rapports radiologiques (Taylor et al., 2009).

CHAPITRE 3

-

Méthodes Bio-informatiques appliquées à des données exomiques et Méthodes Statistiques pour l'étude des variants rares

Dans le cadre de cette thèse, nous avons réalisé une étude d'association génétique à partir de données de séquençage d'exomes réalisé au CNRGH (Centre National de Recherche en Génomique Humaine). Ce chapitre décrit les approches méthodologiques adoptées dans cette recherche, en se concentrant sur les aspects bio-informatiques et statistiques.

I. Introduction

Les études d'association permettant d'étudier le risque génétique associé à divers phénotypes ont évolué dans le temps en fonction des technologies disponibles pour l'analyse des variants.

Les premières études d'association génétique ont considéré un nombre restreint de variants au sein d'un nombre limité de gènes sélectionnés *a priori* (études gènes candidats). Les gènes candidats sont généralement choisis en fonction de trois critères principaux : leur pertinence biologique, leur impact fonctionnel et une fréquence suffisamment élevée des variants dans la population générale. Cette approche est cependant limitée par sa dépendance aux connaissances existantes sur la biologie, connue ou théorique, de la maladie. En effet, bien que ces études gènes candidats soient efficaces pour analyser les associations génétiques, elles présentent un biais inhérent au processus de sélection initiale des gènes et à l'accent mis sur les régions codantes des gènes étudiés.

Des puces pangénomiques ont par la suite été développées permettant d'analyser plusieurs centaines de milliers de variants génétiques répartis sur le génome afin d'identifier des associations entre des variants fréquents (généralement définis par une fréquence au moins égale à 5 % dans la population générale) et des phénotypes d'intérêt (Genome-Wide Association Study ; GWAS). En 2019, plus de 5700 GWAS avaient été réalisées sur plus de 3000 traits (Uffelmann et al., 2021) ; ces études ont permis d'identifier un grand nombre de nouveaux variants associés à divers phénotypes. Cependant, en raison de leur focus sur les variants communs, les GWAS sont moins efficaces pour identifier des variants rares ou ultra-rares, potentiellement importants dans les maladies multifactorielles.

Pour pallier cette limitation, des technologies avancées telles que le séquençage de l'exome (WES) et du génome complet (WGS) ont été développées et sont désormais

utilisées. Entre le début des années 2000 et 2022, le prix du séquençage d'un génome humain est passé de 100 000\$ à 1000\$. Cette diminution significative du prix permet actuellement de réaliser des études sur des variants rares ou ultra-rares.

Dans ce chapitre, nous détaillerons l'ensemble des méthodes bio-informatiques appliquées ainsi que les différentes méthodes statistiques utilisées pour l'analyse des variants rares.

II. Méthodes bio-informatiques

1. Préparation des librairies et séquençage

Dans un premier temps, l'ADN est fragmenté, le rendant ainsi plus manipulable. Des adaptateurs sont ensuite ajoutés à chaque fragment d'ADN permettant de préparer les banques d'ADN pour le séquençage en marquant chaque fragment avec les séquences nécessaires à l'amplification et à la reconnaissance sur la plateforme de séquençage. La phase de capture est ensuite réalisée. Pour cela, il existe diverses techniques selon le fournisseur de kits de capture et la plateforme de séquençage utilisée. Dans le cadre de notre étude, la technique de capture en phase liquide avec le kit « *Agilent SureSelect All Exon V6* » a été utilisée par le CNRGH, ciblant les exons à l'aide de sondes oligonucléotiques.

Enfin, trois principaux contrôles qualité sont effectués avant de pouvoir passer à l'étape de séquençage des fragments d'ADN. Le contrôle qualité après fragmentation permet de vérifier que la fragmentation de l'ADN a été effectuée de manière uniforme. Puis, le contrôle qualité après préparation des banques d'ADN permet de vérifier que les adaptateurs ont bien été ajoutés à chaque fragment d'ADN et enfin, le contrôle qualité après la capture permet de s'assurer que les régions cibles – les exons – ont bien été capturées.

Dans le cadre de cette étude, un séquençage d'exome dit « *paired-end* » a été réalisé

sur la plateforme Illumina HiSeq4000 du CNRGH. Il repose sur une méthode de séquençage par synthèse. Le processus de séquençage est illustré sur la Figure 11.

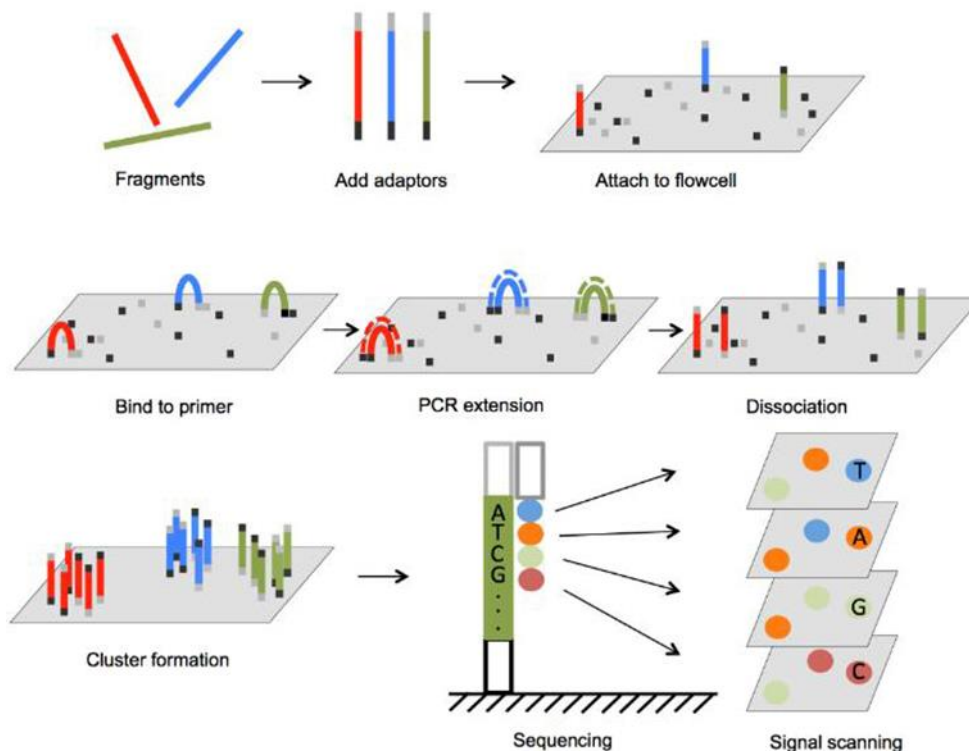


Figure 11 - Etapes d'un séquençage Illumina. Source : (Lu et al. 2016).

Les fragments d'ADN sont attachés à des flow cells pouvant retenir les fragments grâce à des séquences oligonucléotiques complémentaires aux adaptateurs. Ces fragments attachés sont amplifiés par PCR bridge comme indiqué sur le schéma, afin de créer plusieurs milliers de copies identiques de chaque fragment créant ainsi un cluster. Sont ensuite ajoutés aux flow cells : des amorces, des ADN polymérases et des nucléotides, chaque type de nucléotide étant identifié par un marqueur fluorescent différent. Les amorces s'attachent alors sur les fragments d'ADN et sont étendues par incorporation de nouveaux nucléotides, à l'aide de l'ADN polymérase. À chaque incorporation, la base ayant été ajoutée est déterminée en fonction de la longueur d'onde de son marqueur fluorescent. Le processus peut ainsi se répéter, avec une nouvelle incorporation de nucléotide, jusqu'à ce que le fragment d'ADN soit totalement séquençé.

Ce séquençage Illumina permet ainsi de lire les fragments d'ADN, d'une longueur de 150 paires de bases, qui sont stockés dans des fichiers au format FASTQ. Chacun de ces fichiers contient donc l'ensemble des fragments séquençés appelés des « reads ». En résumé, ce format est un format textuel, contenant la séquence de chaque *read* ainsi que le score de confiance de chaque base.

Afin d'analyser ces données brutes de séquençage, des pipelines d'analyses ont été développés à l'aide d'un gestionnaire de *workflows Snakemake* (64) et du gestionnaire de paquet et d'environnement Conda. Cela permet ainsi de garantir une bonne reproductibilité des données obtenues, en respectant les principes FAIR (Findability, Accessibility, Interoperability, and Reusability). Ces pipelines, incluant plusieurs opérations successives décrites dans les parties suivantes, permettent de passer des données brutes de séquençage, des fichiers au format FASTQ à des données directement utilisables pour les analyses d'association génétique des variants rares (cf Partie III.2 - Les tests d'association par agrégation). Les fichiers ne contiennent ainsi que des variants de bonne qualité, chacun annoté avec toutes les informations nécessaires aux analyses statistiques.

2. Traitement des données brutes de séquençage

a. Contrôle qualité des fichiers de séquençage

Le traitement des *reads* et le contrôle qualité avant l'alignement sur un génome de référence sont des étapes cruciales pour garantir la fiabilité et la précision des analyses ultérieures. Ces étapes permettent d'éliminer les séquences de faible qualité, les erreurs de séquençage et les contaminants potentiels, ce qui réduit le nombre de faux positifs et minimise les artefacts lors de l'alignement.

b. Alignement sur un génome de référence

Une fois la qualité des *reads* vérifiée, l'étape suivante consiste à aligner l'ensemble des fragments nucléotidiques de chaque individu sur un génome de référence, afin de détecter les variations par rapport à ce génome de référence. Cela peut être représenté graphiquement comme une superposition des *reads* sous les gènes correspondants (Figure 12).

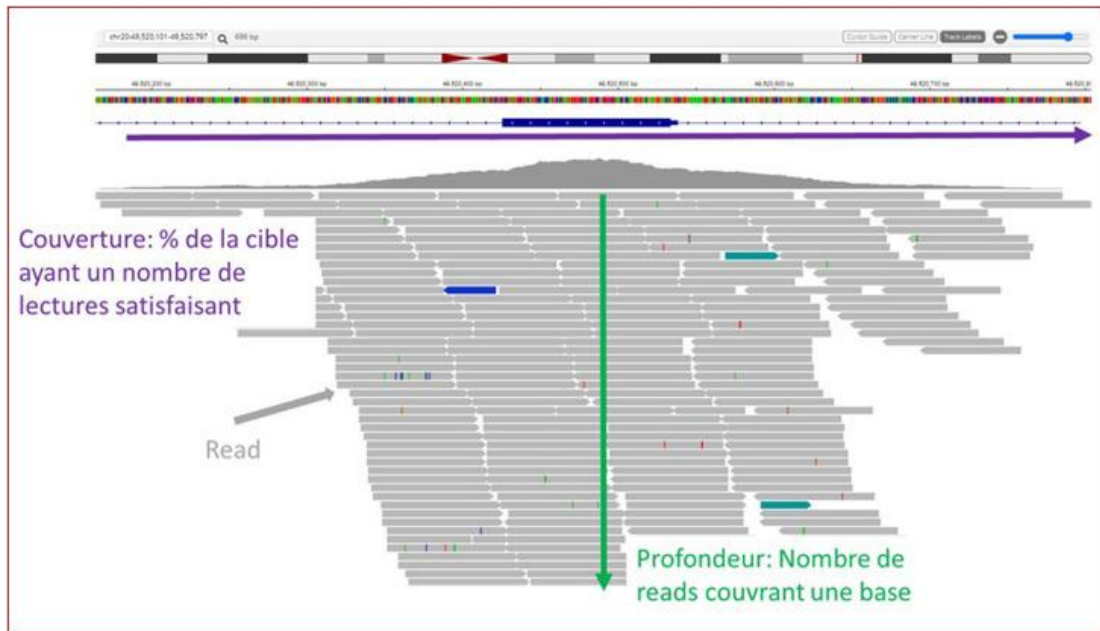


Figure 12 - Schématisation de l'alignement des *reads* sur le génome de référence. Source : (Georget et al.,2023). *Read* et *lecture* sont deux termes équivalents.

Cette superposition des *reads* sous chaque base correspond à la notion de profondeur et permet d'évaluer la couverture et donc la fiabilité des données. Une profondeur suffisante garantit que les variants détectés sont fiables et non des artefacts dus à une couverture insuffisante. En général, pour assurer une bonne fiabilité des données, une couverture d'environ 30X est recherchée lors du séquençage, ce qui signifie que chaque base de l'exome a été lue en moyenne 30 fois.

La version du génome de référence la plus couramment utilisée pour l'alignement et la détection des variants est la version GRCh38, publiée en décembre 2013 par le *Genome Reference Consortium*. Cette version intègre des améliorations significatives par rapport à sa version précédente (GRCh37), notamment une meilleure représentation des régions complexes et des corrections de séquences (67) et représente un outil essentiel pour les études de génomique humaine. Cela en fait un outil essentiel pour l'alignement des *reads* et la détection des variants.

Avant la détection des variants, il est important d'effectuer des traitements post-

alignement sur les fichiers obtenus (au format BAM/SAM). En effet, lors de l'alignement, des erreurs peuvent survenir et pourraient être considérées à tort comme des variants, générant ainsi un fort taux de faux positifs. Les différents traitements décrits ci-dessous ont tous été inclus dans le pipeline à la suite de l'outil d'alignement.

Suppression des duplicats

La présence de duplicats lors de l'alignement des fragments d'ADN sur un génome de référence peut significativement affecter la précision des analyses. Les duplicats, qu'ils soient générés lors de l'amplification par PCR ou détectés optiquement pendant le séquençage, peuvent fausser le calcul du score de confiance attribué à un variant et conduire à des erreurs dans l'attribution des génotypes. Une sur-représentation artificielle de certains fragments peut surestimer la fréquence d'un variant comme cela est illustré sur la Figure 13.

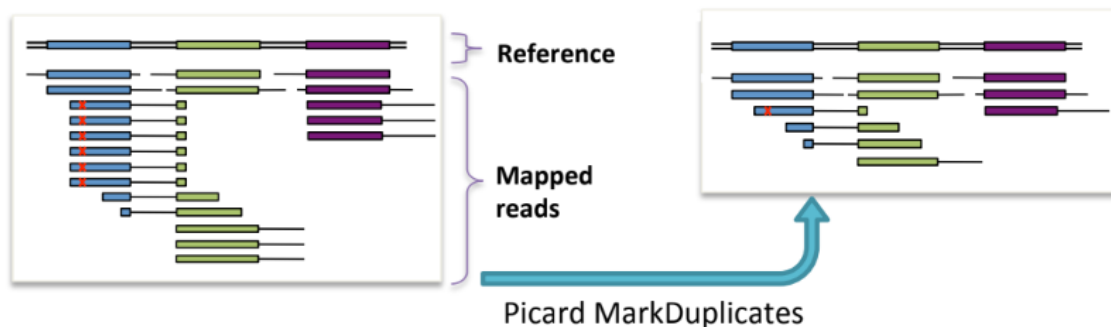


Figure 13 - Suppression des duplicats. Source : (GATK).

Réalignement des insertions/délétions (INDELS)

Lors de l'alignement initial des séquences brutes sur un génome de référence, les indels peuvent représenter un défi particulier. Les indels sont des variations génétiques où une petite séquence d'ADN est soit insérée (insertion) soit supprimée (délétion) par rapport au génome de référence. En raison de leur taille variable, les indels peuvent ne pas être correctement alignés lors de la première étape d'alignement, surtout s'ils sont de tailles différentes de celles attendues par rapport au génome de référence.

Le réalignement spécifique des régions contenant des indels est essentiel pour optimiser la précision de l'alignement génomique. Les algorithmes de réalignement sont conçus pour détecter et corriger ces variations, facilitant ainsi une meilleure concordance entre les séquences d'échantillon et le génome de référence. Cette étape est intégrée lors du *variant calling* avec l'outil *HaplotypeCaller* (présenté dans la partie suivante), où elle joue un rôle crucial dans la réduction des risques de faux positifs lors de l'identification des variants génétiques.

Recalibration des scores de qualité des bases (BQSR)

Dans la partie concernant les fichiers FASTQ, nous avons évoqué les scores de qualité attribués à chacune des bases séquencées, exprimant la confiance du séquenceur dans chaque appel de base. Ces scores présents dans les fichiers FASTQ sont reportés dans les fichiers d'alignement, mais il est nécessaire de les recalibrer pour améliorer la qualité des données. Pour cela, nous avons inclus au pipeline l'outil GATK BaseRecalibrator qui modifie les valeurs de ces scores de qualité bruts en utilisant un algorithme de recalibration qui prend en compte des covariables telles que le cycle de séquençage de la machine, le contexte de séquence ou encore la position dans le *read* qui ont une influence sur le signal de séquençage et la qualité des bases séquencées. Ainsi, en corrigeant ces scores, cela permet d'améliorer la fiabilité des appels de variants en indiquant plus précisément la confiance accordée à chaque observation de base.

3. Détection des variants

a. Création des fichiers gVCF

Cette étape consiste à détecter des variations potentielles dans les séquences par rapport au génome de référence. Il existe de nombreux outils, chacun utilisant sa propre méthode, mais le principe général reste le même. À partir des fichiers d'alignement, ces outils évaluent les séquences pour identifier les positions où des variants peuvent être présents. Une présentation simplifiée de ce processus est donnée à la Figure 14. Les variants détectés sont alors enregistrés par l'outil dans un fichier VCF

(*Variant Call Format*) où chaque ligne correspond à un variant. Ce fichier contient des informations détaillées sur chaque variant, telles que la position dans le génome, le type de variant (substitution, insertion, délétion) et des annotations sur la qualité et la profondeur de lecture.

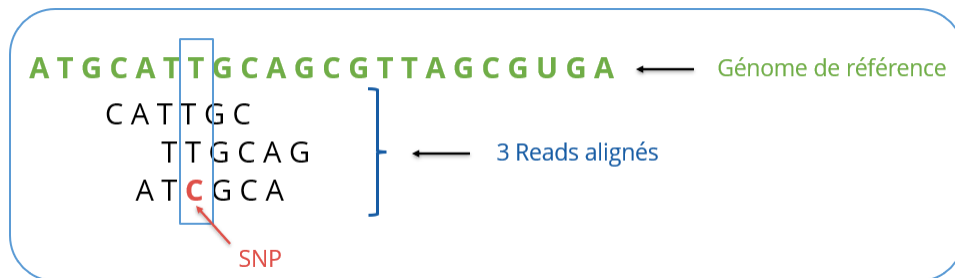


Figure 14 - Fonctionnement de la détection des variants.

L'outil HaplotypeCaller de GATK est le plus adapté dans le cadre d'études d'association génétique avec un design cas-témoin. Il permet de générer un fichier gVCF pour chaque individu contenant les mêmes informations qu'un fichier VCF, mais avec une ligne pour chaque position du génome, que l'individu porte ou non un variant à cette position. Cela permet par la suite d'obtenir les génotypes (homozygote non variant – 0/0, hétérozygote – 0/1 ou homozygote variant – 1/1) fiables, utiles pour la suite des analyses, notamment les prochains contrôles qualité.

b. Etapes post variant calling

L'étape suivant l'obtention des fichiers gVCFs consiste, dans un premier temps, à réunir tous les fichiers au sein d'un fichier unique contenant toutes les positions génomiques (en lignes) pour tous les individus (en colonne). Pour cela, les outils *CombinegVCF* et *GenotypeGVCF* (GATK) sont couramment utilisés. *GenotypeGVCFs* utilise les variants identifiées par *HaplotypeCaller* pour effectuer le génotypage conjoint de tous les individus. Il examine les informations disponibles pour chaque site, incluant les allèles variants et non-variants, à travers tous les échantillons et produit un fichier VCF qui ne contient que les sites ayant au moins un variant détecté chez un individu. Cela permet

d'améliorer la précision et la cohérence des appels de variants.

c. VQSR

Le recalibrage des scores de qualité des variants (VQSR) est une méthode permettant d'améliorer la précision des appels de variants. Cet outil utilise des modèles d'apprentissage automatique et évalue les annotations génomiques telles que la couverture de séquençage, l'orientation des *reads* et d'autres caractéristiques à partir de données de référence comme 1000Genomes ou HapMap permettant ainsi de calculer un nouveau score de qualité. Ce score est utilisé pour filtrer les variants de manière plus précise en minimisant les faux positifs et en maximisant la détection des variants réels. Pour cela, les outils *VariantRecalibrator* et *ApplyVQSR* ont été inclus au pipeline de *variant calling*.

4. Contrôle qualité

Pour assurer la qualité des données génomiques et minimiser les biais potentiels ainsi que la détection de faux positifs, plusieurs filtres de contrôle qualité ont été appliqués dans le cadre de cette thèse.

a. Contrôle qualité des variants

Les critères appliqués aux variants visent à repérer d'éventuelles erreurs de génotypage. Cela passe par l'élimination des variants ayant une profondeur de lecture insuffisante, indicateur d'une faible fiabilité des données, ainsi que ceux ayant une qualité inférieure à un certain seuil. Cette démarche garantit que seuls les variants de haute confiance, et donc moins susceptibles d'être de faux-positifs sont conservés pour la suite des analyses.

Un autre paramètre important à contrôler est le taux d'appel des variants. Ce dernier correspond à la proportion d'individus pour lesquels un génotype a été assigné avec succès pour ce variant particulier. Un taux d'appel élevé indique que le variant a été génotypé de manière fiable pour la plupart des individus de l'étude ; à l'inverse, un

faible taux d'appel peut indiquer des problèmes techniques ou une mauvaise qualité des données de séquençage. En général, les variants dont le taux d'appel est inférieur à 95 % sont éliminés.

Combiner ces différents processus permet de garantir que seuls des variants de haute qualité et robustes sont conservés dans les études d'association génétique.

b. Contrôle qualité des individus

Les contrôles appliqués aux individus permettent de détecter des problèmes comme une mauvaise qualité de l'ADN ou une stratification de la population pouvant biaiser les résultats des études.

Stratification de la population

La stratification de la population est une problématique récurrente des études d'association génétique et survient lorsque les individus comparés sont issus de populations génétiquement hétérogènes. Ce problème est bien connu dans les études d'association pangénomiques (GWAS) sur les variants communs, car il entraîne une inflation du taux de faux positifs de type I. Dans le cas des études sur les variants rares, cette problématique est plus complexe à gérer car ces variants ont une origine plus récente et sont donc davantage regroupés géographiquement (Persyn et al., 2018). Une des solutions courantes consiste à appliquer la méthode utilisée en GWAS, basée sur l'analyse en composantes principales.

Ainsi, à partir de variants communs informatifs et des données génétiques provenant du projet 1000 Genomes (Auton et al., 2015), la stratification génétique peut être analysée au sein d'une population d'étude donnée.

Une présence notable d'individus d'ascendances diverses peut introduire des biais dans l'analyse en raison des différences génétiques substantielles entre les populations d'ascendance africaine et européenne. Une telle hétérogénéité peut en effet polluer le signal génétique, rendant plus difficile l'identification de véritables associations

généétiques. Pour minimiser ces biais potentiels et garantir la robustesse et la validité des résultats, les études d'association génétiques peuvent être restreintes à une seule et unique population, le plus souvent à une population d'ascendance européenne.

Contrôle du taux d'appel

Le taux d'appel d'un individu correspond à la proportion de variants pour lesquels un génotype a été assigné avec succès pour cet individu particulier. Un faible taux d'appel peut indiquer des problèmes avec l'échantillon d'ADN ou des erreurs techniques pendant le processus de séquençage. Les individus ayant un taux inférieur à 95 % sont généralement exclus des analyses.

Détection des individus apparentés

Il est essentiel de vérifier la présence d'individus apparentés dans les études d'association génétique. En partageant une proportion significative de leur génome, ces individus peuvent biaiser les résultats en donnant l'impression que certains allèles sont significativement associés à un phénotype, apparaissant plus fréquemment qu'ils ne le sont réellement dans la population générale.

Contrôle de la consanguinité

La consanguinité augmente la probabilité d'être homozygote pour un variant, ce qui peut fausser les résultats des études d'association génétique.

Contrôle de la contamination

La dernière métrique à prendre en compte pour s'assurer de la bonne qualité des échantillons, est de vérifier le taux d'hétérozygotie de chacun des individus. Un taux élevé peut indiquer une contamination par de l'ADN étranger, ce qui justifie donc l'exclusion de ces individus afin de préserver l'intégrité des données génomique analysées.

5. Annotation et identification des variants

Les études de séquençage, telles que le séquençage complet de l'exome, génèrent un grand volume de variants. Il est essentiel de filtrer ces données pour prioriser les variants les plus pertinents à analyser, présentant un risque plus élevé d'être associés à l'apparition d'un phénotype. Cette étape de filtrage repose sur l'annotation de l'ensemble des variants, prenant en compte des facteurs tels que la fréquence allélique, l'impact prédictif des variants, leur localisation génomique et d'autres informations pertinentes.

a. La localisation

La première information nécessaire est la localisation du variant en fonction de ses coordonnées chromosomiques. Cela permet d'identifier le gène dans lequel il se trouve ainsi que sa position précise au sein d'un élément génétique. Tous les éléments sont présentés dans la Figure 15 ci-dessous.

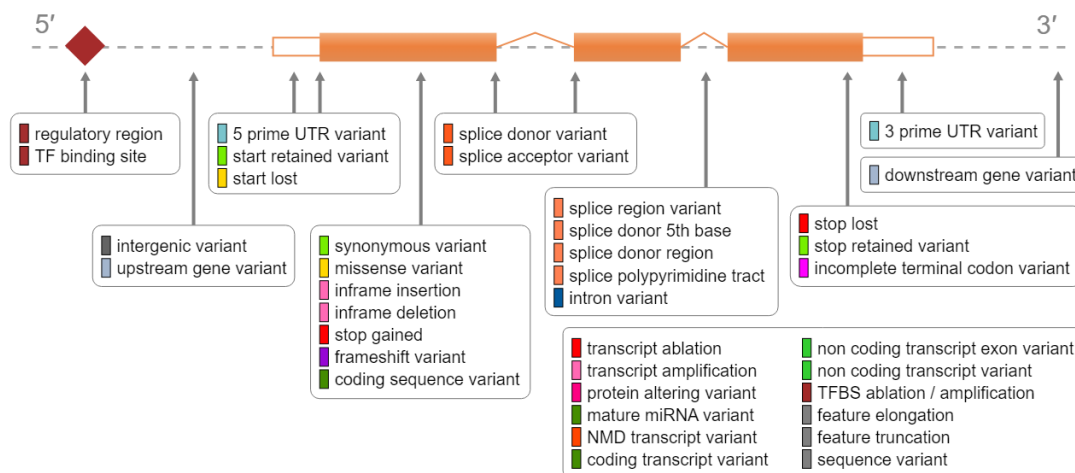


Figure 15 - Localisation possibles des variants Source : (VEP)

b. Impact du variant

Une autre information importante est l'impact du variant sur la protéine produite. Cela permet, en fonction de sa localisation et de son impact sur la protéine, d'ajouter une

information sur la gravité du variant, classée en quatre catégories par l'outil d'annotation VEP (McLaren et al., 2016) qui a été utilisé dans le cadre de cette thèse (Figure 16).

* SO term	SO description	SO accession	Display term	IMPACT
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO:0001893	Transcript ablation	HIGH
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO:0001574	Splice acceptor variant	HIGH
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO:0001575	Splice donor variant	HIGH
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained	HIGH
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift variant	HIGH
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost	HIGH
start_lost	A codon variant that changes at least one base of the canonical start codon	SO:0002012	Start lost	HIGH
transcript_amplification	A feature amplification of a region containing a transcript	SO:0001889	Transcript amplification	HIGH
feature_elongation	A sequence variant that causes the extension of a genomic feature, with regard to the reference sequence	SO:0001907	Feature elongation	HIGH
feature_truncation	A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence	SO:0001906	Feature truncation	HIGH
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO:0001821	Inframe insertion	MODERATE
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	Inframe deletion	MODERATE
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	Missense variant	MODERATE
protein_altering_variant	A sequence_variant which is predicted to change the protein encoded in the coding sequence	SO:0001818	Protein altering variant	MODERATE

Figure 16 - Impact prédit du variant classé en quatre catégories par VEP. Source : <https://www.ensembl.org/info/docs/tools/vep/index.html> (McLaren et al., 2016)

HIGH : Ces variants sont prédits comme ayant un impact élevé sur la fonction de la protéine, souvent en interrompant la séquence codante, comme les mutations non-sens ou les décalages du cadre de lecture. **MODERATE** : Ces variants ont un impact modéré sur la protéine, par exemple les mutations faux-sens, qui changent un acide aminé en un autre. **LOW** : Ces variants ont un faible impact sur la protéine, tels que les mutations synonymes, qui ne changent pas l'acide aminé codé. **MODIFIER** : Ces variants sont généralement situés en dehors des régions codantes et leur impact sur la fonction de la protéine est moins direct. Ils peuvent affecter la régulation génétique, comme les variants dans les régions régulatrices ou les introns.

c. La Fréquence Allélique Mineure (MAF)

Une autre information importante est l'ajout de la fréquence allélique des différents variants dans la population générale permettant ainsi de pouvoir sélectionner les variants rares selon le seuil choisi. Cela est possible grâce aux bases de données de fréquences allélique telles que gnomAD (Genome Aggregation Database) (Karczewski et al., 2020) qui est une vaste ressource compilant des données de séquençage provenant de diverses populations à travers le monde et fournissant des estimations précises de la fréquence des variants génétiques.

Dans les études portant sur les variants rares, un point de départ couramment utilisé pour filtrer les variants est une fréquence allélique mineure (MAF) inférieure à 1 % (Goswami et al., 2021).

III. Méthodes statistiques

1. Variants communs vs Variants rares

On retrouve tout au long du génome des variants génétique que l'on peut catégoriser en deux classes – communs et rares – dépendant de leur fréquence allélique, plus couramment caractérisé par la MAF. Deux principales théories se distinguent autour de ces variants : la théorie « maladies communes – variants communs » et « maladies communes – variants rares » (Schork et al., 2009).

Les variants communs sont présents chez une proportion significative de la population. Par convention, bien qu'on retrouve dans la littérature plusieurs définitions, on estime qu'un variant est commun si sa MAF est supérieure à 1 % (Cirulli & Goldstein, 2010). Ces variants ont été largement étudiés, d'abord par des études de gènes candidats, et plus largement avec l'avènement des études d'association pangénomique (GWAS) (cf Partie I de ce chapitre).

Cependant, il est maintenant de plus en plus question d'héritabilité manquante (Cirulli & Goldstein, 2010; Eichler et al., 2010; Zuk et al., 2014). En effet, ces variants communs n'expliquent en fait qu'une faible proportion de l'héritabilité qui avait été estimé pour ces phénotypes donnant ainsi lieu à une hypothèse alternative à celle retrouvée en GWAS : « maladie commune – variants rares » (Saint Pierre & Génin, 2014), avec une effet plus fort comme cela est représenté sur la Figure 17.

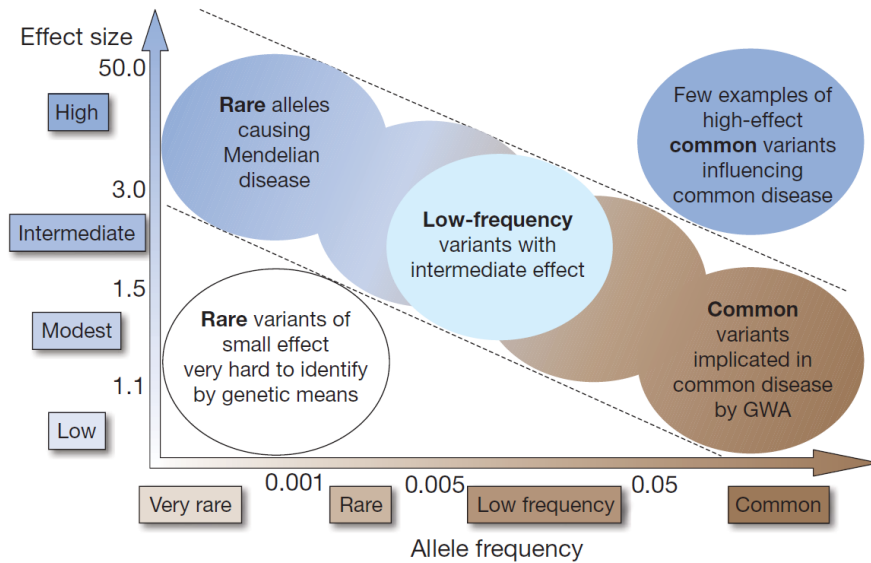


Figure 17 - Variants communs vs Variants communs.Source : (Eichler et al., 2010). Cette figure illustre la relation entre la fréquence des allèles à risque et la force de leur effet. Les variants communs ont une fréquence élevée avec un effet modeste sur le risque de maladie, tandis que les variants rares ont une fréquence faible mais un effet potentiellement plus important. Les études GWAS sont efficaces pour détecter des variants communs, tandis que le séquençage de nouvelle génération (NGS) est utilisé pour identifier les variants rares.

Les études antérieures sur le risque génétique de second cancer après traitement pour un cancer dans l'enfance ont généralement démontré leur association avec des variants communs. Toutefois, certaines publications (Morton et al., 2020; Qin et al., 2020) suggèrent également une implication des variants rares dans ce risque.

Dans cette partie, nous aborderons uniquement la méthodologie statistique concernant les études d'association génétique spécifique aux variants rares.

2. Les tests d'association – variants rares

L'étude d'association des variants rares représente une approche prometteuse qui implique la comparaison des variants identifiés, par le séquençage de type WES ou WGS, chez des cas présentant une trait spécifique étudié et des témoins non-porteurs.

Cependant, les analyses des données générées par ces nouvelles technologies doivent faire face à un défi majeur, à savoir, la faible puissance statistique des tests

d'association simples marqueur en raison de la faible fréquence de ces variants et du grand nombre de tests à effectuer. En effet, les tests d'association pour les variants communs sont inefficaces dans ce contexte, à moins d'utiliser des tailles d'échantillons extrêmement importantes. Par exemple, pour atteindre une puissance de 80 % dans la détection d'un effet avec un risque relatif de 5 et un MAF de 5 %, un échantillon de 1000 individus serait nécessaire. Pour un variant vraiment rare avec une fréquence de 1 %, la taille d'échantillon nécessaire atteint alors les 100 000 individus (Figure 18).

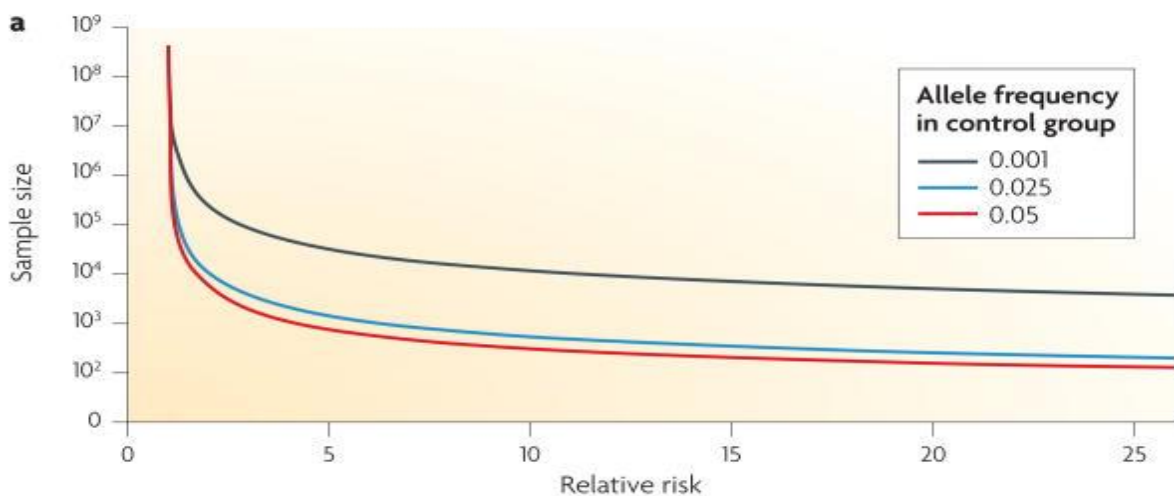


Figure 18 - Tailles d'échantillons nécessaires pour détecter une association entre un allèle avec un risque spécifique et un phénotype d'intérêt. Source : (Bansal et al., 2010).

Pour pallier ces difficultés, plusieurs développements méthodologiques de tests d'association de variants rares proposent d'évaluer l'effet combiné de variants regroupés au sein d'un gène, une région chromosomique ou encore une voie biologique d'intérêt.

L'exemple le plus simple et le plus couramment utilisé consiste à regrouper les variants par gène. La puissance de ces tests d'agrégation ne dépendra donc plus de la fréquence individuelle des variants, mais de la fréquence cumulée des variants au sein de l'unité considérée. Un autre avantage des tests d'agrégation est la réduction du nombre de tests à effectuer diminuant le facteur de la correction pour la multiplicité

des tests (Bansal et al., 2010; Lee et al., 2014).

Un exemple illustratif est celui de l'étude de Li et Leal (B. Li & Leal, 2008) sur un échantillon simulé de 250 cas et 250 témoins avec 10 variants rares, chacun ayant un risque relatif de 2 et une fréquence de 1 %. Lorsque les variants sont testés individuellement, la puissance statistique était de seulement 5 %. En revanche, en les étudiant conjointement, la puissance pour détecter une association entre le phénotype et l'unité génétique atteignait 83 %. Ce gain de puissance démontre l'intérêt majeur de l'utilisation de ce type de tests lors de l'analyse des variants rares.

Dans la partie qui suit, les trois différents types de tests statistiques d'agrégation utilisés lors de nos analyses seront détaillés.

3. Les tests d'association par agrégation

Il existe trois principales catégories de méthodes pour les tests d'association spécifique à l'analyse des variants rares : les tests de fardeau ; les méthodes basées sur la décomposition de la variance (Sequence Kernel Association Test : SKAT) et les méthodes combinées (SKAT optimized : SKAT-O).

a. Les tests de fardeau

Les tests de fardeau consistent dans un premier temps à calculer un score génétique pour chaque individu inclus dans l'étude, en tenant compte des variants rares inclus dans une unité génétique. Cela permet de résumer l'information génétique d'un individu en un unique score.

Il existe plusieurs manières de définir ce score, mais la méthode générale consiste à sommer l'information de tous les variants rares au sein d'une unité pour un individu.

Ensuite, l'association entre ce score et le phénotype étudié est évaluée avec des modèles de régression linéaire ou logistique permettant l'inclusion de covariables aux modèles statistique. Il existe plusieurs variantes aux tests de fardeau dont les

principales sont présentées ci-dessous.

Test de fardeau classique

La définition classique d'un test de fardeau consiste à attribuer à chaque individu un score basé sur le nombre de variants rares pour lesquels l'individu porte au moins un allèle mineur. Plus précisément, pour chaque individu, on compte le nombre de variants pour lesquels l'individu possède au moins un allèle mineur (0 = 0 allèle mineur, 1 = hétérozygote, 2 = homozygote). Ce score est ensuite utilisé pour analyser la distribution de ces variants entre les cas et les témoins.

Cette approche suppose que les effets des variants vont tous dans la même direction et sont additifs : plus le score est élevé, plus le risque d'être atteint de la maladie est supposé augmenter. Cependant, il est important de noter que cette hypothèse d'additivité n'est pas toujours vérifiée.

« Cohort allelic sum test » (CAST) (Morgenthaler & Thilly, 2007)

Cette approche repose sur l'hypothèse que la présence d'un seul variant rare dans un gène est suffisante pour augmenter le risque de maladie, sans considérer un effet additionnel significatif de l'accumulation de plusieurs variants.

Cette méthode permet ainsi de coder les individus d'une manière binaire, ceux portant au moins un variant rare dans l'unité génétique d'intérêt sont codés en tant que 1, tandis que ceux n'en ayant aucun le sont en tant que 0.

La méthode CAST suppose que les effets des variants rares sont homogène et qu'un seul variant est suffisant pour augmenter le risque de développer la maladie. L'accumulation de plusieurs variants dans la même unité génétique n'est pas considérée comme ayant un effet additionnel significatif sur le risque de maladie.

« Combined Multivariate and Collapsing » (B. Li & Leal, 2008)

Dans le test CMC (Combined Multivariate and Collapsing), les variants rares sont classés

en fonction de leur fréquence allélique. Ensuite, un score binaire de type CAST (Cohort Allelic Sum Test) est défini pour chaque catégorie. Cette méthode utilise un modèle multivarié pour tester l'association entre ces catégories de variants et le phénotype d'intérêt. Cette approche permettrait d'optimiser la détection des effets des variants rares tout en tenant compte de leur distribution allélique.

Test « Burden » pondéré (WSS) (Madsen & Browning, 2009)

Dans le test de fardeau pondéré, contrairement au test de fardeau classique qui consiste à sommer simplement les allèles mineurs, un poids est attribué à chaque variant en fonction de sa fréquence. L'objectif de cette pondération est d'attribuer un effet plus important aux variants les plus rares, qui seraient plus susceptibles d'avoir un impact fonctionnel significatif.

Les limites des tests de fardeau

Les tests de fardeau présentent certaines limites par leur hypothèse sous-jacente que tous les variants d'une unité génétique ont des effets similaires agissant dans la même direction, c'est-à-dire soit à risque pour le phénotype étudié, soit protecteurs. Cela réduit leur efficacité lorsque l'unité génétique contient des variants neutres ou aux effets mixtes (Basu & Pan, 2011).

b. Méthodes de décomposition de la variance

Principe de ces méthodes

Ces méthodes, contrairement aux tests de fardeau permettent de prendre en compte un mélange de variants pathogènes, protecteurs, et neutres.

Un mélange de variants pathogènes et protecteurs au sein d'une unité génétique entraîne une surdispersion des données et donc de la variance par rapport à ce qui serait attendu si tous les variants avaient des effets similaires. Ainsi, ces tests ont pour objectif de détecter l'augmentation de la variance.

Les méthodes les plus couramment utilisées sont le « Sequence Kernel Association Test » (SKAT) (Wu et al., 2011) et le test C-alpha (Neale et al., 2011), SKAT étant la méthode de référence dans les études d'association génétique pour variants rares.

Sequence Kernel Association Tests – SKAT

L'idée derrière ces tests va donc être de mesurer la similarité des individus dans une région spécifique du génome pour détecter si les individus qui se ressemblent phénotypiquement présentent également une plus grande similarité génotypique dans cette région.

Il existe plusieurs versions de cette méthode qui a été adaptée à divers scénarios et hypothèses. Parmi les plus courantes, on trouve le SKAT pondéré, assez équivalent sur le principe au fardeau pondéré (plus un variant est rare plus un poids important lui est accordé), mais également le iSKAT permettant de tester des interactions entre variants rares et facteurs environnementaux (Lin et al., 2016).

Les limites de la méthode SKAT

Contrairement aux tests de fardeau, SKAT permet donc de prendre en compte des effets opposés ainsi que des variants neutres. Cette méthode est donc plus robuste face à la présence de variants neutres et à des effets variant dans différentes directions. Toutefois, elle est moins puissante que le test de fardeau lorsqu'une grande proportion des variants rares dans une région sont réellement causaux et influencent le phénotype dans la même direction.

Étant donné que nous ne connaissons pas *a priori* le modèle biologique exact, ni la proportion de variants causaux, ni les seuils de fréquence, il pourrait être tentant de multiplier les tests pour couvrir toutes les possibilités. Toutefois, une alternative plus efficace est le test combiné SKAT-O (SKAT-Optimal) (Lee, Wu, et al., 2012), qui cherche à optimiser la puissance en combinant les avantages des tests de fardeau et des tests de décomposition de la variance comme SKAT.

c. Les méthodes Combinées (SKAT-O)

L'objectif de SKAT-O est d'atteindre une puissance maximale dans les conditions optimales pour les deux types de tests, offrant ainsi une approche plus équilibrée et potentiellement plus puissante pour la détection des variants rares associés aux phénotypes d'intérêt.

Cette approche permet de maximiser la puissance du test en exploitant les conditions les plus favorables entre les deux méthodes, fardeau et SKAT, et offre une solution robuste pour détecter les associations génétiques dans des contextes où les effets des variants peuvent être hétérogènes.

Les avantages et inconvénients de ces trois méthodes sont résumés dans le Tableau 3.

Tableau 3- Avantages et inconvénients des méthodes statistiques spécifiques à l'étude des variants rares. *Tableau adapté à partir de l'article de (B. Li & Leal, 2008).*

	Tests de Fardeau <i>Agrège les variants rares en un score génétique</i>	SKAT <i>Teste la variance des effets génétiques</i>	SKAT-O <i>Combine les deux tests précédents</i>
Puissance	Quand une large proportion des variants sont causaux et vont dans la même direction.	En présence de variants à la fois délétères et protecteurs au sein de l'unité génétique ou si petite proportion de variants causaux.	Combine les statistiques de test des deux tests précédents.
Perte de puissance	En présence de variants à la fois délétères et protecteurs au sein de l'unité génétique ou si petite proportion de variants causaux.	La plupart des variants sont causaux et vont dans le même sens (délétères ou protecteurs)	Légèrement moins puissant que les tests de fardeau ou de variance si leurs hypothèses sont respectées.

d. Puissance des méthodes pour variants rares

La puissance de ces méthodes dépend de plusieurs paramètres.

La taille de l'échantillon joue un rôle crucial : plus l'échantillon est grand, plus la puissance est élevée. La fréquence cumulée et la taille des effets cumulés des variants

au sein d'une unité génétique sont également déterminantes. En GWAS, on examine souvent un seul variant à la fois, ce qui signifie que la puissance dépend de la fréquence du variant et de son effet individuel. Le regroupement de variants, comme cela se fait avec les tests de fardeau et SKAT, peut être avantageux pour augmenter la puissance en cumulant les effets des variants rares. La puissance de ces tests dépend également de la proportion des variants causaux (non neutres) au sein de l'unité génétique. Plus la proportion de variant causaux est grande plus la puissance du test sera importante.

En conséquence, une partie préliminaire essentielle de ces méthodologies consiste à utiliser des outils bio-informatiques pour sélectionner des variants causaux, les plus susceptibles d'avoir un impact sur le phénotype étudié. Cette étape est cruciale pour éviter d'inclure des informations non pertinentes, telles que des variants neutres, qui peuvent diluer le signal et diminuer la puissance des tests.

Il est également courant – à l'instar des études gène-candidats – de sélectionner un nombre limité d'unités génétiques, la plupart du temps des gènes, afin de limiter le nombre de tests statistiques réalisés et donc de gagner en puissance.

e. Sélection des gènes

Comme nous avons pu le voir, une des problématiques majeures des tests spécifiques aux variants est le manque de puissance dans ce type d'étude. En effet, regrouper les variants rares au sein d'unités génétiques permet de réduire le nombre de tests à réaliser et d'augmenter ainsi la puissance statistique. Cependant, étant donné les faibles effectifs dans la plupart des études d'association génétique, il est souvent nécessaire de se restreindre, comme dans les études de gènes candidats, à une liste de gènes d'intérêt. Cette approche s'applique également à notre étude. Pour obtenir des points de comparaison pertinents au vu de la littérature (cf. chapitre 4 – Revue systématique) et pour des raisons biologiques, nous avons choisi de nous concentrer uniquement sur certains gènes impliqués dans trois voies spécifiques : la voie de la réparation de l'ADN, la voie de la réplication de l'ADN et la voie de la recombinaison

de l'ADN.

Les voies de contrôle de la stabilité génétique tiennent une place essentielle dans le maintien de l'intégrité du génome. Parmi ces dernières, on peut notamment citer les voies de la réplication, de la réparation et de la recombinaison ; qui ont un intérêt particulier dans le cadre de notre étude d'association génétique sur le risque de seconds cancers. Compte tenu de l'importance de ces trois voies dans le développement du cancer, il est probable que ces variants puissent modifier le risque de second cancer, en interaction avec la chimiothérapie/radiothérapie reçue lors du premier cancer.

Construction de la liste

Afin de constituer cette liste de gènes impliqués dans les trois voies d'intérêts – Réplication, Réparation et Recombinaison (3R) – une fouille des données a été réalisée dans les trois majeures base de données publiques internationales.

Gene Ontology (GO) (Ashburner et al., 2000; Gene Ontology Consortium et al., 2023)

GO est une base de données qui a pour but de définir un vocabulaire commun de description génique – une ontologie. Ainsi, chaque gène/produit génique est décrit par un ensemble de termes, ces derniers étant en relation les uns avec les autres. La structure de cette base peut se représenter sous forme d'un graphe acyclique orienté et offre une description normalisée des fonctions des gènes, des processus biologiques et des composants cellulaires. Elle est largement utilisée pour l'annotation fonctionnelle des gènes et les analyses de *pathways*.

Molecular Signature Database (MSigDb) (Liberzon et al., 2011; Subramanian et al., 2005)

MSigDb est une large base de données contenant les annotations sur des sets de gènes impliqués dans des voies biochimiques, cascades de signalisation et bien d'autres concepts biologiques. Cette dernière est subdivisée en 8 collections basées sur divers critères tels que les processus biologiques, les voies de signalisation, les réponses aux

traitements et les perturbations génétiques (de C1 à C7 + Hallmark).

Kyoto Encyclopedia of Gene and Genome (KEGG) (Kanehisa & Goto, 2000)

KEGG est une base de données japonaise contenant toutes les informations en ce qui concerne les connaissances actuelles sur les voies moléculaires et génétiques, acquises d'observation expérimentales. Cela explique que cette dernière regroupe moins d'informations que les deux précédentes, car sont présents uniquement les gènes qui ont été observés de manière expérimentale comme appartenant à une voie.

CHAPITRE 4

-

Radiations reçues aux organes lymphoïdes durant la radiothérapie pour un cancer de l'enfant et risque de second cancer de la thyroïde

Ce chapitre est consacré à une étude spécifique sur le risque de second cancer de la thyroïde, l'un des plus fréquents parmi les patients guéris d'un cancer pédiatrique. À partir des données de la cohorte FCCSS et de la cohorte BCCSS, décrites au chapitre 2, le but de cette étude est d'évaluer le risque de cancer secondaire de la thyroïde après traitement par radiothérapie. Plus précisément, nous nous concentrerons sur les effets des radiations ionisantes sur trois organes du système immunitaire : le thymus, la moelle osseuse et la rate. La première partie de ce chapitre décrit l'importance de la prise en compte de ces trois organes dans l'étude du risque de second cancer de la thyroïde. La deuxième partie détaille les méthodes statistiques utilisées, notamment la création d'un score composite d'irradiation. Enfin, les troisième et quatrième parties sont consacrées aux résultats et aux conclusions de cette étude. Un article scientifique est en cours de préparation et sera soumis d'ici la fin août 2024.

I. Contexte

Comme mentionné dans le chapitre 1, l'un des cancers secondaires les plus courants chez les survivants d'un cancer pédiatrique est le cancer de la thyroïde, représentant environ 10 % de tous les seconds cancers (Reulen et al., 2011; Turcotte et al., 2015). Ce cancer secondaire est particulièrement associé à la radiothérapie en raison de la sensibilité élevée des tissus thyroïdiens aux radiations qui est d'autant plus marquée chez les patients de jeune âge (Travis, 2006).

La relation entre la radiothérapie et le risque de cancer de la thyroïde est désormais bien établie et se distingue de celle observée pour les autres seconds cancers solides, pour lesquels une relation linéaire est observée même à des doses élevées de radiation (Figure 9).

On observe une relation linéaire pour une exposition aux radiations jusqu'à 15-20 Gy, puis une décroissance du risque avec l'augmentation de la dose de radiation reçue à la thyroïde, probablement en raison de l'effet de mort cellulaire des tissus thyroïdiens (Inskip et al., 2016).

Une étude a également mis en évidence plusieurs autres facteurs de risque influençant la relation entre le risque de cancer secondaire de la thyroïde et la dose de radiation, parmi lesquels l'âge au moment du premier cancer (avec un risque plus élevé chez les patients traités jeunes), le sexe (avec un risque plus élevé chez les femmes), et le temps écoulé depuis l'exposition aux radiations (avec un risque plus élevé avec une durée de suivi plus longue) (Veiga et al., 2016). Il existe également un risque de cancer secondaire de la thyroïde chez les patients ayant été traités par chimiothérapie. Une étude (Veiga et al., 2012) a mis en évidence un risque accru de cancer de la thyroïde chez les personnes traitées avec des agents alkylants et ayant reçu moins de 20 Gy de radiations. Cependant, le risque attribuable à la chimiothérapie diminue chez les patients traités par radiothérapie à mesure que la dose de radiation à la thyroïde

augmente. Ces résultats soutiennent donc l'hypothèse de la mort cellulaire liée aux fortes doses de radiations lors de la thérapie, qui pourrait masquer l'effet de la chimiothérapie.

Par ailleurs, le système immunitaire joue un rôle crucial dans la surveillance immunitaire, la prévention et la régression des cellules tumorales (Hadden, 2003). Une étude antérieure a mis en évidence un risque élevé de développement d'un second cancer de la thyroïde chez les patients ayant subi une splénectomie, qu'elle soit chirurgicale ou induite par une forte dose de radiations (de Vathaire, Haddy, et al., 2015).

L'objectif de cette étude est d'évaluer l'effet potentiel de l'irradiation d'organes autres que la thyroïde sur le risque d'apparition d'un second cancer de la thyroïde associé au traitement reçu lors du cancer pédiatrique. Pour cela, nous avons choisi de nous intéresser à trois organes lymphoïdes en particulier : le thymus, la rate et la moelle osseuse, qui sont impliqués dans le système immunitaire. On distingue deux types d'organes lymphoïdes selon leurs rôles. On retrouve d'abord les organes lymphoïdes primaires – la moelle osseuse et le thymus – où les lymphocytes, cellules effectrices de la réponse immunitaire adaptative sont produits, se développent et sont sélectionnés. Les organes lymphoïdes secondaires incluent notamment la rate et participent à l'activation des lymphocytes et sont donc le point de départ de la réponse immunitaire (Figure 19).

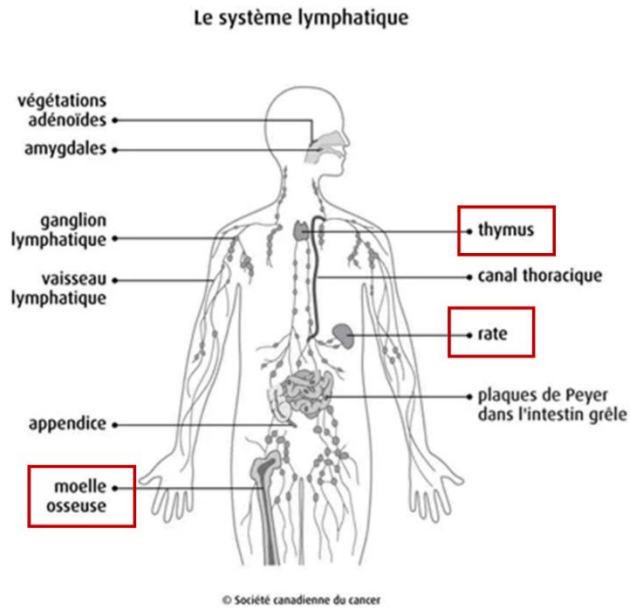


Figure 19 - Principaux organes lymphoïdes du système immunitaire *Les trois organes encadrés en rouge sont les trois organes d'intérêt dans cette étude*

L'hypothèse de notre étude est que l'irradiation des organes principaux du système immunitaire, tels que le thymus, la moelle osseuse et la rate, compromet leur fonctionnalité, voire les rend totalement dysfonctionnels, comme dans le cas d'une splénectomie. Le système immunitaire est également impliqué dans la reconnaissance et la destruction des cellules précancéreuses ; ainsi, une diminution de la fonction immunitaire due à l'irradiation de ces organes pourrait donc permettre à ces cellules précancéreuses de se développer augmentant le risque de second cancer de la thyroïde.

Dès lors, cette étude a pour objectif d'étudier le rôle de l'irradiation de ces organes dans le risque de cancer de la thyroïde tout en contrôlant le rôle direct de l'irradiation sur la thyroïde et celui de la chimiothérapie, facteur de risques bien connus.

II. Méthodologie de l'étude

1. Données de l'étude et design de l'étude

Cette étude repose sur les données de la cohorte FCCSS et de la cohorte BCCSS (chapitre 2). Pour étudier le risque de second cancer de la thyroïde associé à l'irradiation des trois organes d'intérêt, une étude cas-témoins nichée dans ces deux cohortes a été menée.

Les cas, patients ayant développé un cancer de la thyroïde, ont d'abord été sélectionnés dans les deux cohortes. Ces 124 cas ont individuellement été appariés à des témoins (de un à cinq) n'ayant pas développé de second cancer de la thyroïde. L'appariement a été réalisé sur le sexe, le pays (France/Royaume-Uni), la classe d'âge au diagnostic du cancer de l'enfant (0-1 ; 2-5 ; 6-10 ; 11-15 ; >15), l'année de diagnostic du premier cancer (+/- 9 ans), la durée de suivi (le suivi de chaque témoin doit être supérieur ou égale à celui du cas), la classe de dose de radiation à la thyroïde (pas de radiothérapie ; 0-1 ; 2-5 ; 6-20 ; >20) et la chimiothérapie reçue (oui/non). Au total, ce sont donc 582 témoins qui ont été inclus.

2. Variables étudiées

La chimiothérapie a été définie comme covariable binaire (oui/non) et les doses cumulées, en mg/m² d'anthracyclines et d'agents alkylants, ont été calculées avec les formules suivantes (de Baat et al., 2023) :

$$\begin{aligned} & \text{Dose cumulée total d'anthracyclines (mg/m}^2\text{)} \\ & = \text{doxorubicin} + (\text{daunorubicin} \times 0,5) + (\text{epirubicin} \times 0,8) \\ & + (\text{idarubicin} \times 3) + (\text{mitoxantrone} \times 10,5) \end{aligned}$$

Dose équivalente de cyclophosphamide (DEC)

$$\begin{aligned} &= 1,0 \times (\text{dose cumulée de cyclophosphamide (mg/m}^2\text{)}) \\ &+ 0,244 \times (\text{dose cumulée d'ifosfamide (mg/m}^2\text{)}) \\ &+ 0,857 (\text{dose cumulée de procarbazine (mg/m}^2\text{)}) \\ &+ 14,286 (\text{dose cumulée de chlorambucil (mg/m}^2\text{)}) \\ &+ 15,0 (\text{dose cumulée de BCNU (carmustine) (mg/m}^2\text{)}) \\ &+ 16,0 (\text{dose cumulée de CCNU (lomustine) (mg/m}^2\text{)}) \\ &+ 40 (\text{dose cumulée de melphalan (mg/m}^2\text{)}) \\ &+ 50 (\text{dose cumulée de ThioTEPA (thiotepa) (mg/m}^2\text{)}) \\ &+ 100 (\text{dose cumulée de moutarde azoté (mg/m}^2\text{)}) \\ &+ 8,823 (\text{dose cumulée de busulfan (mg/m}^2\text{)}) \end{aligned}$$

La distribution de la dose de radiation reçue par l'ensemble des organes du corps avait été rétrospectivement reconstruite (chapitre 2) pour les patients inclus dans cette étude. Les doses estimées de radiothérapie pour les organes ainsi que les indicateurs des histogrammes dose-volume avaient été ainsi calculés. Dans notre étude, les organes d'intérêt étaient la thyroïde (utilisée pour l'appariement des cas et des témoins), le thymus, la moelle osseuse active et la rate. La dose moyenne de radiation reçue par ces organes a été utilisée dans les analyses.

La splénectomie a été définie de la manière suivante : elle concerne les patients ayant subi une ablation chirurgicale de la rate ou ceux ayant reçu une dose de radiation moyenne supérieure à 40 Gy sur cet organe. Dans la suite des analyses, la dose de radiation reçue à la rate des patients ayant subi une splénectomie a été considérée comme nulle.

3. Analyses statistiques

Les Odds Ratio (OR) et les intervalles de confiance à 95 % (IC 95 %) ont été estimés à l'aide d'une régression logistique conditionnelle. Chaque strate contient un cas et ses témoins. Les données ont été ajustées sur les variables d'appariement lorsqu'elles n'étaient pas exactement les mêmes, c'est-à-dire l'âge au moment du diagnostic, l'époque du traitement et la dose moyenne de radiation à la thyroïde, pour lesquelles un appariement par classe a été effectué.

Un test de corrélation de Pearson entre les différentes doses reçues aux 3 organes d'intérêt (rate, thymus et moelle osseuse), nous avons pu constater une corrélation positive statistiquement significative entre les différents organes (Tableau 4).

Tableau 4 - Corrélation entre les doses reçues aux différents organes d'intérêt

		Dose moyenne à la rate	Dose moyenne à la moelle	Dose moyenne au thymus
Dose moyenne à la rate	p-value	1	0,49	0.28
	Coefficient de corrélation		<.0001	<.0001
Dose moyenne à la moelle	p-value		1	0.45
	Coefficient de corrélation			<.0001
Dose moyenne au thymus	p-value			1
	Coefficient de corrélation			

Les coefficients de corrélation de Pearson ont été calculés pour évaluer la relation linéaire entre les doses d'irradiations reçues par les trois organes lymphoïde d'intérêt. Les valeurs du coefficient varient entre -1 et 1, où 1 indique une corrélation positive parfaite, -1 une corrélation négative parfaite, et 0 aucune corrélation. Les corrélations sont jugées statistiquement significatives si la p-value <0,05

Les doses de radiation aux organes d'intérêt ont été analysées une à une pour éviter les problèmes de colinéarité qui existe entre les doses reçues à chacun des organes (Tableau 4). Trois modèles indépendants ont ainsi été réalisés : le modèle 1 afin d'étudier les doses d'irradiations au thymus, le modèle 2 afin d'étudier les doses reçues à la moelle osseuse et le modèle 3 afin d'étudier les doses à la rate. Toutes les analyses statistiques ont été effectuées avec le logiciel SAS version 9.4 (SAS Institute Inc). Une valeur p bilatérale inférieure à 0,05 était considérée comme statistiquement significative.

III. Résultats

L'étude porte sur 124 cas ayant développé un second cancer de la thyroïde et 586 témoins. Les caractéristiques cliniques des 706 patients de l'étude sont présentées dans le Tableau 5 et les caractéristiques des traitement reçus dans le

Tableau 6.

Tableau 5 - Caractéristiques cliniques des cas et des témoins de l'étude

		Tous les patients (N = 706)		Cas (N = 124)		Témoins (N = 582)	
		N	%	N	%	N	%
Age au diagnostic (en années)	0-1	52	7,4	8	6,5	44	7,6
	2-5	152	21,5	29	23,4	123	21,1
	6-10	218	30,9	39	31,5	179	30,8
	11-15	223	31,6	40	32,3	183	31,4
	15	61	8,6	8	6,5	53	9,1
Médiane (IQR)		8	(4-13)	8	(4-12)	8	(4-13)
Sexe	Homme	295	41,8	53	42,7	242	41,6
	Femme	411	58,2	71	57,3	340	58,4
Années de traitement	< 1980	230	32,6	40	32,3	190	32,6
	1980-1989	200	28,3	33	26,6	167	28,7
	≥1990	276	39,1	51	41,1	225	38,7
Durée de suivi (en années)	<20	69	9,8	56	45,2	13	2,2
	20-29	237	33,6	37	29,8	200	34,4
	30-39	189	26,8	21	16,9	168	28,9
	40-49	150	21,2	7	5,6	143	24,6
	50 et plus	61	8,6	3	2,4	58	10
Médian (IQR)		33	(25-42)	21	(16-30)	36	(27-43)
Age atteint (en années)	<30	123	17,4	61	49,2	62	10,7
	30-39	170	24,1	32	25,8	138	23,7
	40-49	219	31	25	20,2	194	33,3
	50 et plus	194	27,5	6	4,8	188	32,3
Médiane (IQR)		42	(34-51)	30	(24-40)	45	(37-52)
Type de cancer pédiatrique	Lymphome	193	27,3	48	38,7	145	24,9
	Tumeur du SNC	153	21,7	17	13,7	136	23,4
	Néphroblastome	82	11,6	16	12,9	66	11,3
	Sarcome des tissus mous	78	11	11	8,9	67	11,5
	Neuroblastome	53	7,5	12	9,7	41	7
	Ostéosarcome	49	6,9	6	4,8	43	7,4
	Autres	45	6,3	5	4	40	6,9
	Tumeur des Gonades	39	5,5	8	6,5	31	5,3
	Rétinoblastome	14	2	1	0,8	13	2,2

Abréviations : IQR, Intervalle Interquartiles ; SNC, Système Nerveux Central

Tableau 6 - Caractéristiques de traitement des cas et des témoins de l'étude

		Tous les patients (n=706)		Cas (n=124)		Témoins (n=582)	
		N	%	N	%	N	%
Traitement	Chirurgie	30	4,2	1	0,8	29	5,0
	RT sans CT	127	18,0	21	16,9	106	18,2
	CT sans RT	79	11,2	18	14,5	61	10,5
	CT et RT	470	66,6	84	67,7	386	66,3
Radiothérapie	Non	109	15,4	19	15,3	90	15,5
	Oui	597	84,6	105	84,7	492	84,5
Chimiothérapie	Non	157	22,2	22	17,7	135	23,2
	Oui	549	77,8	102	82,3	447	76,8
Agents alkylants	Non	279	39,5	44	35,5	235	40,4
	Oui	427	60,5	80	64,5	347	59,6
Anthracycline	Non	438	62	70	56,5	368	63,2
	Oui	268	38	54	43,5	214	36,8
Dose de radiation à la thyroïde (Gy)	≤0	108	15,3	18	14,5	90	15,5
]0-1]	134	19	23	18,5	111	19,1
]1-5]	110	15,6	19	15,3	91	15,6
]5-20]	244	34,6	42	33,9	202	34,7
	>20	110	15,6	22	17,7	88	15,1
Médiane (IQR)		8	(1-17)	9	(1-17)	8	(1-17)
Dose de radiation à la rate (Gy)	≤0	152	21,5	21	16,9	131	22,5
]0-5]	331	46,9	59	47,6	272	46,7
]5-10]	74	10,5	13	10,5	61	10,5
]10-20]	88	12,5	12	9,7	76	13,1
	>20	61	8,6	19	15,3	42	7,2
Médiane (IQR)		2	(0-10)	4	(1-12)	2	(0-10)
Splénectomie	0	688	97,5	117	94,4	571	98,1
	1	18	2,5	7	5,6	11	1,9
Dose de radiation à la moelle osseuse (Gy)	≤0	117	16,6	20	16,1	97	16,7
]0-5]	342	48,4	58	46,8	284	48,8
]5-10]	177	25,1	34	27,4	143	24,6
]10-20]	70	9,9	12	9,7	58	10
Médiane (IQR)		4	(2-7)	4	(2-8)	4	(2-7)
Dose de radiation au thymus (Gy)	≤0,1	138	19,5	21	16,9	117	20,1
]0,1-5]	261	37	33	26,6	228	39,2
]5-10]	54	7,6	10	8,1	44	7,6
]10-20]	107	15,2	31	25	76	13,1
	>20	146	20,7	29	23,4	117	20,1
Médiane (IQR)		6	(1-20)	12	(2-21)	5	(1-20)

Le score composite varie de 0 à 3, où un score de 0 signifie qu'aucun organe d'intérêt n'a reçu en moyenne, plus de 5 Gy, tandis qu'un score de 3 signifie que chaque organe a reçu au moins 5 Gy. Les trois organes concernés sont le thymus, la rate et la moelle osseuse.

Abréviations : IQR, Intervalle Interquartiles

Le suivi médian est de 21 ans (IQR : 16-30) pour les cas et de 36 ans (27-43) pour les témoins, avec respectivement 31 (25 %) et 168 (29 %) patients ayant un suivi de plus de 30 ans. Les seconds cancers de la thyroïde sont principalement observés chez des patients ayant été traités pour un lymphome (38,7 %), un cancer du système nerveux

central (SNC) (13,7 %) ou un néphroblastome (13,9 %) durant l'enfance.

La majorité des cas et des témoins ont été traités par chimiothérapie et radiothérapie, et 84,6 % des patients ont reçu une radiothérapie dans le cadre du traitement du cancer pédiatrique.

Nous avons observé un risque de cancer de la thyroïde significativement associé à la dose de radiation reçue par le thymus (p-value d'hétérogénéité < 0,0001). Plus précisément, le risque de cancer de la thyroïde le plus élevé est observé chez les patients ayant reçu des doses de radiation au thymus comprises entre 10 et 20 Gy (OR = 7,27 ; IC 95 % : 3,29-16,06). Les résultats suggèrent également un effet dose significatif entre la dose de radiation reçue au thymus et le risque de cancer de la thyroïde (p-tendance <0,0001) (

Tableau 7).

Tableau 7 - Analyses multivariées pour les trois organes d'intérêt et le score composite d'irradiation

	Classes	OR	IC 95%	p-value*	p-het**	P-tendance***
Modèle 1	5-10	3,24	1,22-8,59	0,018	<.,001	0,0001
	10-20	7,27	3,29-16,06	<,0001		
	>20	4,26	1,79-10,15	0,0011		
Modèle 3	0-5	2,69	0,79-9,22	0,1143	0,0137	0,0264
	5-10	2,41	0,61-9,55	0,2112		
	11-20	2,2	0,56-8,62	0,2593		
	>20	6,98	1,82-26,86	0,0047		
Modèle 2	0-5	1,4	0,38-5,12	0,6117	0,8154	0,8110
	5-10	1,7	0,43-6,69	0,4459		
	10-20	1,34	0,31-5,77	0,6988		

Modèle 1 : Dose de radiations reçues au thymus (Gy) ; Modèle 3 : Dose de radiations reçues à la rate (Gy) ; Modèle 3 : Dose de radiation reçue à la moelle osseuse (Gy).

Variables d'ajustement : âge au diagnostic, dose cumulée d'agents alkylants en mg/m², période de traitement, dose moyenne à la thyroïde en Gy.

*p-value : Une p-value inférieure à 0,05 indique une différence significative entre le groupe de référence et les autres classes. La classe de référence correspond à ne pas avoir reçu de radiothérapie aux organes concernés.

**p-het : Une p-hétérogénéité < 0,05 indique une différence significative entre les différentes classes d'irradiation.

***p-tendance : Une p-tendance inférieure à 0,05 indique une tendance linéaire significative entre les différentes classes d'irradiations

Abréviations : IC, Intervalle de Confiance ; OR, Odds Ratio

De même, un risque accru de second cancer de la thyroïde est observé pour les patients ayant reçu des doses plus élevées de radiation à la rate, avec un effet de dose démontré (p tendance = 0,026) (

Tableau 7). On peut aussi observer un risque plus important de second cancer de la thyroïde chez les individus ayant reçu plus de 20 Gy à la rate par rapport à ceux n'ayant pas eu de radiothérapie (OR = 7 ; IC 95 % : 1,82-26,86 ; p = 0,0047).

En revanche, aucune relation significative n'a été observée entre la dose de radiation reçue par la moelle osseuse et le risque de développer un second cancer de la thyroïde (p -tendance = 0,81) (

Tableau 7).

IV. Discussion

Dans cette étude, nous avons évalué l'impact de l'irradiation des organes lymphoïdes sur le risque de développement d'un second cancer de la thyroïde chez les survivants de cancers pédiatriques. Nos résultats mettent en évidence pour la première fois une association significative entre l'irradiation des organes lymphoïdes et l'augmentation du risque de second cancer de la thyroïde, avec des variations notables en fonction des organes irradiés.

Plus précisément, nous avons observé que l'irradiation du thymus est fortement associée à un risque accru de second cancer de la thyroïde, illustrée par une relation dose-réponse significative.

D'autre part, nous avons également observé une tendance similaire concernant le risque de second cancer de la thyroïde en relation avec l'irradiation de la rate, notamment une association plus marquée chez les individus ayant reçu des doses

élevées de radiation. Ce résultat vient confirmer ce qui avait été observé auparavant dans l'étude de Vathaire et al. (2015) dans les mêmes cohortes, mais dans une population plus restreinte (55 cas de cancer de la thyroïde).

Dans cette publication, une association significative entre une splénectomie chirurgicale ou radiologique et le risque de second cancer de la thyroïde avait été mise évidence, association qui se maintenait après ajustement sur la dose de radiation reçue à la thyroïde.

Nous avons donc mis en évidence un rôle spécifique de l'irradiation des organes lymphoïdes secondaires, à savoir la rate et le thymus, dans le risque de cancer de la thyroïde. En revanche, l'irradiation de la moelle osseuse n'a pas montré d'effet notable sur le risque de seconds cancers de la thyroïde. Cette absence d'effet pourrait être due à la répartition diffuse de la moelle osseuse active dans l'organisme. Dans notre étude, qui exclut les patients atteints de leucémie, peu d'individus ont reçu une irradiation totale du corps. Par conséquent, même ceux avec une dose moyenne « élevée » ont conservé une partie de leur moelle osseuse exposée à des doses très faibles. Cette exposition réduite a probablement permis de préserver leur réserve de cellules souches hématopoïétiques, réduisant ainsi les effets délétères de la radiothérapie sur la production sanguine et minimisant le risque de second cancer de la thyroïde.

Récemment, une étude de randomisation mendélienne (Fang et al., 2024) a corroboré nos résultats en mettant en lumière le rôle causal des cellules immunitaires dans le risque de cancer de la thyroïde. Il est également essentiel de considérer que l'exposition à des doses élevées de radiations dans les organes lymphoïdes peut entraîner une suppression immunitaire prolongée. Cette suppression pourrait non seulement augmenter le risque de cancer secondaire, mais aussi modifier la réponse à d'autres thérapies comme la chimiothérapie. Il sera donc crucial dans les futures recherches d'explorer les interactions complexes entre l'irradiation des organes lymphoïdes et l'apparition de seconds cancers.

Dans notre étude, l'association entre l'irradiation du thymus et le risque de second cancer de la thyroïde mérite une attention particulière en raison de leur proximité anatomique, qui favorise une exposition simultanée lors de l'irradiation du thymus. En effet, on observe une tendance relativement similaire entre le risque de cancer de la thyroïde et les doses de radiation reçues au thymus et celles reçues à la thyroïde (Inskip et al., 2016). Par conséquent, il est difficile d'isoler l'effet spécifique de l'irradiation du thymus sur le risque de cancer de la thyroïde, ce qui pourrait conduire à une surestimation de son rôle.

Les résultats de notre étude dans leur ensemble soulignent que l'évaluation du risque de second cancer de la thyroïde ne doit donc pas se limiter à la dose de radiation reçue à la thyroïde elle-même. D'autres facteurs, tels que l'exposition aux agents alkylants, la splénectomie, ou encore la dose de radiation aux organes impliqués dans l'immunité (notamment le thymus et la rate), jouent également un rôle crucial dans le risque de cancer secondaire de la thyroïde. Ces éléments doivent être intégrés dans les modèles de prévision du risque de second cancer afin de mieux comprendre et de prédire les risques à long terme pour les survivants de cancers pédiatriques.

Enfin, un autre facteur important à considérer dans l'évaluation du risque de second cancer de la thyroïde est le facteur génétique, qui pourrait jouer un rôle significatif comme cela a été démontré dans déjà plusieurs études (Gramatges et al., 2019; Richard et al., 2020). Les prochains chapitres de cette thèse se concentrent donc sur ce risque génétique qui pourrait permettre, à terme, de mieux comprendre comment les variants génétiques peuvent interagir avec les traitements pour influencer le développement de cancers secondaires.

CHAPITRE 5

—

Risque génétique de second cancer après traitement pour un cancer pédiatrique : une revue systématique de la littérature

Ce chapitre est consacré à une revue systématique de la littérature. Cette dernière a pour objectif de dresser un état des lieux des variants génétiques, gènes ou voies biologiques ayant été associés au risque de seconds cancers après traitement d'un cancer pédiatrique. Les résultats des 18 études sont présentés selon la technique de séquençage utilisée : les études gènes-candidats, les études pangénomiques et les études utilisant les technologies de séquençage telles que le séquençage d'exome ou de génome.

Ce travail a été publié au sein de la revue « Cancer Epidemiology, Biomarkers and Prevention » en mai 2024.

I. Introduction

Les divers facteurs de risque associés aux seconds cancers, notamment la chimiothérapie et la radiothérapie, ont été décrits dans le chapitre 1. Bien que le traitement reçu lors du cancer pédiatrique soit un facteur de risque reconnu, ce dernier ne suffit pas à expliquer le risque de cancers secondaires et des preuves de variabilité interindividuelle suggèrent une composante génétique modulant ce risque (Armenian & Bhatia, 2009; Bhatia, 2015; Gramatges & Bhatia, 2018).

Depuis le début des années 2000, plusieurs études se sont intéressées au risque génétique associé à la survenue de seconds cancers chez les patients guéris d'un cancer pédiatrique. Les premières études étaient principalement des études candidats et d'association pangénomique afin d'identifier des variants communs associés à ce risque. Avec l'introduction récente des nouvelles technologies de séquençage, telles que le WES ou WGS, l'attention se tourne maintenant également vers l'exploration des variants rares jusqu'à présent très peu étudiés.

Dans ce chapitre, nous examinerons en détail les différents gènes, variants et voies biologiques qui ont été identifiés comme étant liés au risque de seconds cancers chez les patients ayant été traités pour un cancer dans l'enfance.

II. Matériel et Méthodes

Le protocole de cette revue, enregistré sur PROSPERO sous le numéro *CRD42023388965*, a été réalisé conformément aux bonnes pratiques PRISMA (Page et al., 2021) (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*), qui visent à améliorer la qualité et la transparence des revues systématiques et des méta-analyses.

1. Stratégie de recherche

Une recherche systématique d'articles publiés avant juin 2024 a été menée en utilisant trois bases de données : PubMed, Scopus et Web of Sciences.

Les termes utilisés pour la recherche dans ces trois bases de données, présentés plus précisément en annexe A reposent sur la méthode PICOS (Amir-Behghadami & Janati, 2020), couramment utilisée pour définir les éléments clé d'une stratégie/question de recherche. Nous avons inclus toutes les études, impliquant des patients ayant survécu à un cancer pédiatrique (Population), traités par chimiothérapie et/ou radiothérapie (Intervention), sans comparaison spécifique pour cette revue (Comparaison non applicable). L'évènement étudié est le risque génétique de développer un second cancer (Outcome) et les études incluses sont des études cas-témoins et des études de cohorte (Study design). Cette revue se concentre uniquement sur les articles ne traitant pas des syndromes génétiques bien établis (tels que Li-Fraumeni ou la neurofibromatose). Cependant, les résultats des études ayant identifié des variants dans des gènes associés à ces syndromes sont présentés. Ces différentes informations sont décrites dans le Tableau 8.

Tableau 8 - Termes PICOS utilisés pour l'inclusion des articles à la revue systématique.

Terme PICOS	Explication	Application
Population (P)	Groupe de personnes sur lequel porte l'étude	Patients ayant survécu à un cancer pédiatrique
Intervention (I)	Traitement, ou exposition étudiée dans l'étude	Chimiothérapie et/ou radiothérapie
Comparateur (C)	Groupe de référence ou de contrôle utilisé pour comparer les résultats de l'intervention	Non applicable pour la revue
Résultats (O - Outcome)	Résultat mesuré ou observé de l'intervention ou de l'exposition	Risque génétique de SC
Design de l'étude (S - Study Design)	Type d'étude utilisé pour répondre à la question de recherche	Etudes cas-témoins et études de cohorte

2. Procédure de sélection

Tous les articles repérés dans les trois bases de données ont été examinés de manière indépendante par deux investigateurs, en se basant sur le titre et le résumé de chaque étude. Ensuite, les articles retenus ont été analysés par un seul investigateur, en examinant le texte intégral, afin de décider lesquels seraient inclus dans la revue.

Les articles non rédigés en anglais ou traitant exclusivement des gènes de prédisposition au cancer, tels que *RB1*, *NF1* et *TP53* ont été exclus. Les références bibliographiques de chacun des articles inclus ainsi que des revues déjà existantes sur le risque de seconds cancers ont également été examinées pour inclure d'éventuels autres articles pertinents.

3. Contrôle qualité des études

Nous avons utilisé l'échelle Q-GENIE (Sohani et al., 2015) pour évaluer la qualité des articles inclus dans cette revue. Cette échelle permet d'évaluer les biais classiques pris en compte dans l'échelle Newcastle-Ottawa (Wells et al., 2014) ainsi que les biais spécifiques aux études de liaison génétique. L'échelle juge les articles selon 11 critères, attribuant un score de 1 (faible) à 7 (excellent). Le score global permet ensuite de classer l'article en trois catégories : faible, modéré et haute qualité. L'évaluation de la qualité est résumée dans le Tableau 9.

L'échelle Q-GENIE offre une évaluation plus adaptée aux études d'association génétique que l'échelle Newcastle-Ottawa, couramment utilisée dans la mise en place des revues systématique de la littérature, en intégrant des critères supplémentaires spécifiques à ce type de recherche. En plus des critères classiques comme la sélection des participants et la comparabilité des groupes, Q-GENIE évalue la rigueur de la stratégie d'analyse génétique et la pertinence des méthodes statistiques employées. Elle prend en compte le risque de stratification des populations, un biais fréquent dans les études génétiques, ainsi que la validité externe des résultats, essentielle pour leur généralisation à d'autres populations. L'échelle permet aussi de vérifier la puissance

statistique des études, assurant que l'échantillon est suffisant pour détecter les associations, et contrôle les erreurs de génotypage qui peuvent compromettre la fiabilité des résultats. De plus, elle intègre l'évaluation des interactions gène-environnement, souvent négligée dans d'autres échelles, mais cruciale pour une compréhension complète des mécanismes génétiques. Ces éléments font de Q-GENIE une échelle parfaitement adaptée à l'évaluation des études d'association génétique.

Tableau 9 - Evaluation du risque de biais des dix-huit études incluses à l'aide de l'outil Q-Genie (Sohani et al., 2015).

Etude	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	TOTAL
Morton et al.,2020	7	6	7	5	2	1	5	6	5	3	6	53
Qin et al., 2020	7	6	5	3	1	1	5	6	5	2	6	47
Opstal-van Winden, 2019	7	7	7	7	3	3	5	6	6	5	6	62
Wang et al., 2018	7	7	5	4	1	3	5	6	6	5	6	55
Wang et al., 2018 (PRS)	7	7	5	4	1	2	3	5	5	2	5	46
Morton et al., 2017	6	5	6	5	1	4	5	7	6	5	6	56
Best et al., 2011	6	5	6	7	6	2	4	6	6	5	7	60
Richard et al., 2020	7	6	5	3	1	3	4	3	6	3	5	46
Gramatges et al., 2019	6	6	5	3	1	3	4	3	6	3	5	45
Wang et al.,2017	6	7	6	5	1	2	4	5	3	5	7	51
Vodusek et al., 2016	7	6	5	5	1	2	2	3	4	1	5	41
Applebaum et al., 2016	6	5	5	2	3	1	3	5	3	2	5	40
Ma et al., 2012	7	7	6	4	1	2	3	3	4	1	6	44
Erculj et al., 2012	7	6	5	4	1	2	2	5	3	1	5	41
Stanulla et al., 2009	6	4	5	3	2	1	3	3	3	2	4	36
Mertens et al., 2004	7	5	5	4	3	1	4	5	3	2	5	44
Blanco et al., 2002	7	4	4	3	1	1	4	4	3	2	5	38
Woo et al., 2000	6	4	5	3	1	1	2	4	2	3	5	36

Des scores ≤ 35 indiquent des études de basse qualité, des scores >35 et ≤ 45 indiquent des études de qualité modérée, et des scores >45 indiquent des études de bonne qualité (Sohani et al., 2015).

Signification des items: 1 - justification de l'étude, 2 - résultat d'intérêt, 3 - groupe de comparaison, 4 - classification technique de l'exposition, 5 - classification non technique de l'exposition, 6 - autres sources de biais, 7 - taille de l'échantillon et puissance, 8 - planification a priori des analyses, 9 - méthodes statistiques et contrôle des facteurs de confusion, 10 - vérification des hypothèses et des inférences pour les analyses génétiques, 11 - pertinence des conclusions tirées des résultats.

III. Résultats

Au total, 709 articles ont été identifiés dans les trois bases de données. Après la suppression des 163 doublons, 546 références restantes ont été examinées indépendamment par deux évaluateurs, sur titres et résumés. Cela a permis d'identifier 32 études pour une évaluation complète en texte intégral selon nos critères d'inclusion prédéfinis. Suite à la lecture des articles complets, 14 parmi eux (Broniscer et al., 2004; Danuta et al., 2020; Dincer et al., 2015; Ehrhardt et al., 2019; Felix et al., 1998; Gramatges et al., 2014; Kelly et al., 2002; Knapke et al., 2012; Kony et al., 1997; Lavergne et al., 2020; Nichols et al., 2003; Sapkota et al., 2019; Sherborne et al., 2015; Uckunkaya et al., 2020) ne répondaient pas aux critères prédéfinis et ont été exclus. Les articles exclus et la raison de leur exclusion est décrite dans le Tableau 10. **Erreur ! Source du renvoi introuvable.**

Finalement, 18 articles (Applebaum et al., 2017; Best et al., 2011; Blanco et al., 2002; Erčulj et al., 2012; Gramatges et al., 2019; Ma et al., 2012; Mertens et al., 2004; Morton et al., 2017, 2020; Opstal-van Winden et al., 2019; Qin et al., 2020; Richard et al., 2020; Stanulla et al., 2009; Vodusek et al., 2016a; X. Wang et al., 2017; Z. Wang, Liu, et al., 2018; Z. Wang, Wilson, et al., 2018; Woo et al., 2000) ont été retenus pour inclusion dans cette revue systématique. Le processus de sélection détaillé est présenté dans la Figure 20.

Tableau 10 - Raison d'exclusion des articles après lecture du texte

Auteurs, année	Titre	Raison de l'exclusion (PICOS)
Kony et al. 1997	Radiation and genetic factors in the risk of second malignant neoplasms after a first cancer in childhood	OUTCOME (genetic)
Felix et al. 1998	Association of CYP3A4 genotype with treatment-related leukemia	POPULATION
Kelly et al. 2002	Polymorphisms of drug metabolizing enzymes and markers of genotoxicity to identify patients with Hodgkin's lymphoma at risk of treatment-related complications	OUTCOME (SMN)
Nichols et al. 2003	TP53, BRCA1, and BRCA2 tumor suppressor genes are not commonly mutated in survivors of Hodgkin's disease with second primary neoplasms	POPULATION
Broniscer et al. 2004	Second neoplasms in pediatric patients with primary central nervous system tumors: the St. Jude Children's Research Hospital experience	OUTCOME (genetic)
Knapke et al. 2012	Hereditary cancer risk assessment in a pediatric oncology follow-up clinic	OUTCOME (genetic)
Gramatges et al. 2014	Telomere content and risk of second malignant neoplasm in survivors of childhood cancer: A report from the childhood cancer survivor study	OUTCOME (genetic)
Dincer et al. 2015	DNA Repair Gene Polymorphisms and Their Relation with DNA Damage, DNA Repair, and Total Antioxidant Capacity in Childhood Acute Lymphoblastic Leukemia Survivors	OUTCOME (genetic)
Sherborn et al. 2017	Somatic and germline TP53 alterations in second malignant neoplasms from pediatric cancer survivors	OUTCOME (genetic)
Sapkota et al. 2019	Genome-Wide Association Study in Irradiated Childhood Cancer Survivors Identifies HTR2A for Subsequent Basal Cell Carcinoma	OUTCOME (SMN)
Ehrhardt et al. 2019	Subsequent breast cancer in female childhood cancer survivors in the St Jude Lifetime Cohort Study (SJLIFE)	OUTCOME (genetic)
Danuta et al. 2020	Molecular karyotyping and gene expression analysis in childhood cancer patients	OUTCOME (genetic)
Uckunkaya et al. 2020	Patterns of Prior and Subsequent Neoplasms in Children and Adolescents with Soft Tissue Sarcomas	OUTCOME (genetic)
Lavergne et al. 2020	Germline MUTYH Mutation in a Pediatric Cancer Survivor Developing a Secondary Malignancy	POPULATION

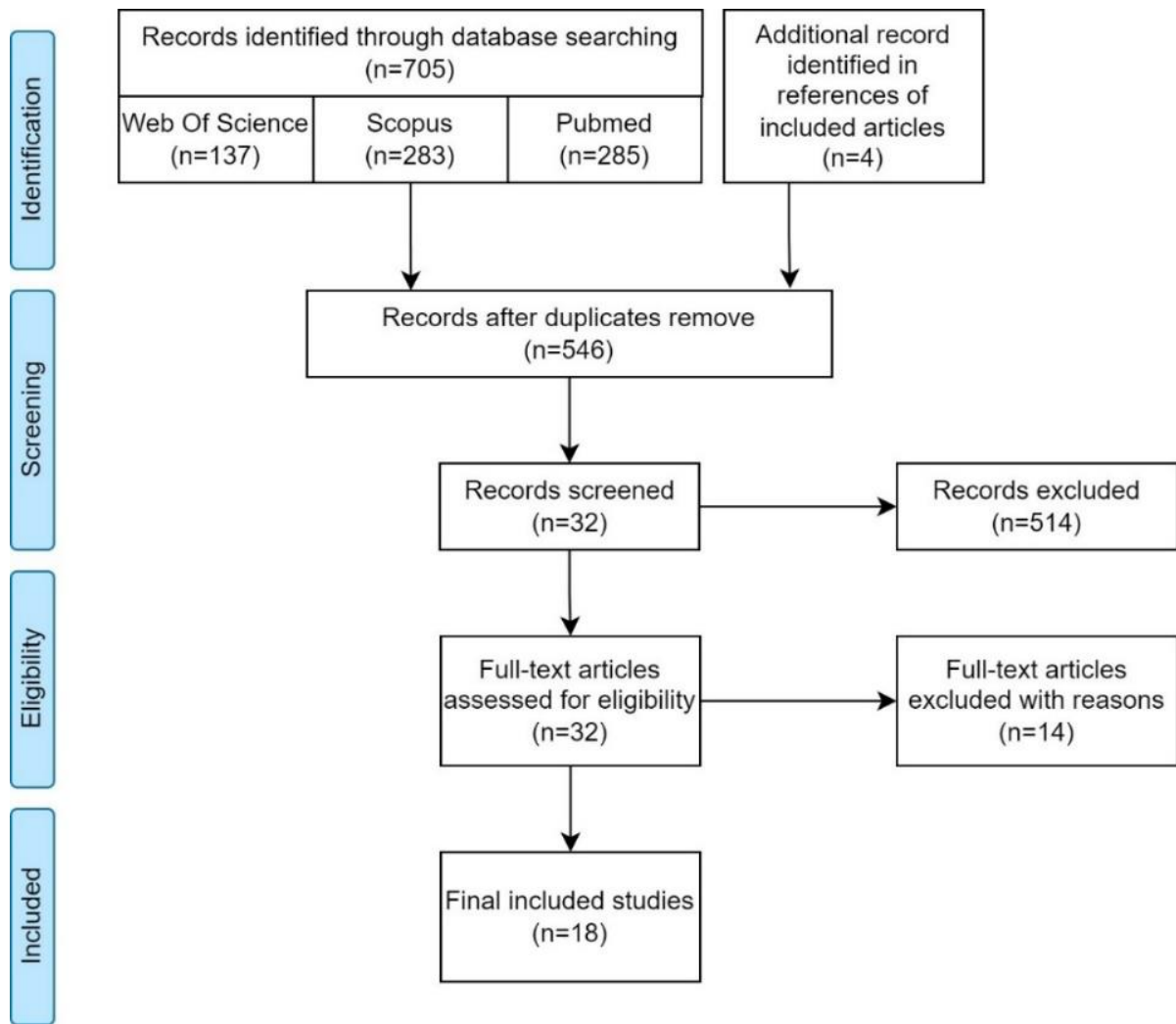


Figure 20 – Processus de sélection des études pour inclusion dans la revue systématique.

1. Description des articles inclus

Les caractéristiques des 18 articles inclus sont présentées dans le Tableau 11. Une grande partie des études s'appuient sur deux grandes cohortes américaines de patients ayant été traités pour un cancer pédiatrique : la *Childhood Cancer Survivors Study* (CCSS) et la *St-Jude Lifetime* (SJLIFE) (Best et al., 2011; Gramatges et al., 2019; Mertens et al., 2004; Morton et al., 2017, 2020; Opstal-van Winden et al., 2019; Qin et al., 2020; Richard et al., 2020; Z. Wang, Liu, et al., 2018; Z. Wang, Wilson, et al., 2018). Le risque de second cancer est évalué après traitement quel que soit le type de premier cancer dans 9 études, pour un lymphome Hodgkinien dans quatre études (Best et al., 2011; Ma et al., 2012; Mertens et al., 2004; Opstal-van Winden et al., 2019), pour une leucémie lymphoblastique aiguë (Blanco et al., 2002; Erčulj et al., 2012; Stanulla et al., 2009; Woo et al., 2000) dans quatre études et pour un neuroblastome dans une étude (Applebaum et al., 2017). En ce qui concerne les seconds cancers, les études évaluent le risque globalement (Applebaum et al., 2017; Best et al., 2011; Erčulj et al., 2012; Mertens et al., 2004; Qin et al., 2020; Stanulla et al., 2009; Z. Wang, Wilson, et al., 2018) et/ou par type de cancer, notamment du sein (Ma et al., 2012; Morton et al., 2017; Opstal-van Winden et al., 2019; Z. Wang, Liu, et al., 2018), de la thyroïde (Gramatges et al., 2019; Richard et al., 2020; Vodusek et al., 2016a), du système nerveux central (X. Wang et al., 2017), de leucémie myéloïde aiguë liée au traitement (t-AML, *treatment-related Acute Myeloid Leukemia*) (Blanco et al., 2002; Woo et al., 2000) et de cancers connus comme étant liés à la radiothérapie (Morton et al., 2020).

Deux études ont été incluses bien que leur schéma ne corresponde pas directement à nos critères d'inclusion. Ces deux articles (Ma et al., 2012) ; (Opstal-van Winden et al., 2019) présentent le risque génétique de seconds cancers chez les patients guéris d'un cancer avant 36 et 40 ans respectivement. Cependant, des analyses stratifiées selon l'âge au premier cancer (<20 ans / ≥20 ans) sont également présentées, ce qui explique l'inclusion de ces deux articles dans notre revue.

La stratification de la population est un facteur de confusion important dans les études d'association génétique (cf. chapitre 3). Huit études ont mené des analyses exclusivement sur des populations de descendance européenne ou ont stratifié leurs données en fonction de cette descendance. De plus, trois études ont ajusté leurs analyses pour tenir compte de l'origine des patients inclus.

Dans la suite de ce chapitre, les résultats sont présentés en fonction de l'approche utilisée (les approches sont détaillées dans le chapitre 3). Onze articles concernent des études gènes-candidats (Applebaum et al., 2017; Blanco et al., 2002; Erčulj et al., 2012; Gramatges et al., 2019; Mertens et al., 2004; Morton et al., 2017; Richard et al., 2020; Stanulla et al., 2009; Vodusek et al., 2016a; X. Wang et al., 2017; Woo et al., 2000), trois articles des études d'association pangénomiques (Best et al., 2011; Morton et al., 2017; Opstal-van Winden et al., 2019) et quatre articles des études de séquençage à haut débit (séquençage de génome ou d'exome) (Morton et al., 2020; Qin et al., 2020; Z. Wang, Liu, et al., 2018; Z. Wang, Wilson, et al., 2018). Les résultats des études sont présentés respectivement dans les parties 2, 3 et 4 suivantes du chapitre.

Tableau 11 - Caractéristiques des études incluses dans la revue systématique.

Article	Méthode de séquençage	ADN	Design de l'étude Cohorte Test statistique	Pays, origine	Durée de suivi	Cancers pédiatriques	Année de traitement	Age au diagnostic du 1er cancer	Seconds cancers
Richard et al., PLoS One, 2020	Gène-candidat	Sang, Salive, Cellules buccales	Cas-témoins CCSS Cas : 110 Témoins : 4956 Modèle de Cox (HR)	USA Ascendance européenne et africaine	NA	Tous types	1970 - 1986	Moyenne (ET) Cases : 9 (5,5) Controls : 7,9 (5,9)	Thyroïde
Gramatges et al., CEBP, 2019	Gène-candidat	NA	Cohorte CCSS SC : 118 Pas de SC : 5206 Modèle de Cox (HR)	EUA Ascendance européenne et africaine	NA	Tous types	1970 - 1986	Année: n (%) Cas < 15 : 96 (81,4) ≥ 15 : 22 (18,6) Témoins < 15 : 4251 (81,7) ≥ 15 : 955 (18,3)	Thyroïde
Wang et al., JCO, 2017	Gène-candidat	Sang, Salive	Cas-témoins appariée COG Cas : 82 Témoins : 228 Validation : 25/54 Régression logistique (OR)	EUA Appariement sur l'ascendance	Médiane Cas : 13,2 Témoins : 13,2	Tous types	NA	Médiane (min-max) Cas : 3,5 (0-21) Témoins : 5 (0-20)	SNC
Applebaum et al., Eur J Cancer, 2017	Gène-candidat	NA	Cas-témoins appariée COG SC : 31 Pas de SC : 309 Régression logistique (OR)	EUA Ascendance européenne et africaine	Médiane (min-max) Cas : 41 mois (1-223)	Neuroblastome	1990 - 2006	Médiane Cas : 27,5 mois Témoins : 18 mois	Tous types
Vodusek et al., Radiol Oncol, 2016	Gène-candidat	Biopsie	Cas-témoins appariée Patients suivis à Lubjana Cas : 24 Témoins : 24 Modèle de Cox (RR)	Slovenie	Médiane (min-max) Cas : 19,6 (9,0-3,6) Témoins : 18,8 (12,80-27,6)	Tous types	1960 - 2006	Médiane (min-max) Cas : 12,7 (7,1-6,7) Témoins : 12,9 (5-15,4)	Thyroïde

Article	Méthode de séquençage	ADN	Design de l'étude Cohorte Test statistique	Pays, origine	Durée de suivi	Cancers pédiatriques	Année de traitement	Age au diagnostic du 1er cancer	Seconds cancers
Erčulj et al., J Cancer Res Clin Oncol, 2012	Gène-candidat	Moelle osseuse	Cas-témoins appariée Patients suivis à Ljubljana SC : 20 Pas de SC : 39 Modèle de Cox (RR)	Slovenie	Médiane (min-max) Cas : 9,5 (2-34) Témoins : 28 (10-37)	LAL	1971 - 2001	Médiane (min-max) Cas : 5 (1,3-12,3)	Tous types
Ma et al., Blood, 2012 ^a	Gène-candidat	Sang	Cas-témoins Étude de cohorte nationale sur le cancer du sein SC : 232 (< 20: 101) Pas de SC : 461 (< 20 : 140) Régression logistique (RR)	Grande-Bretagne (Validation : Pays-Bas) Ascendance européenne	NA	LH	1963 - 2003	Année : n (%) Cas : 0-19: 68(48,6) Validation : 0-19: 33(35,9) Témoins : 0 -19 : 102(33) Validation : 0-19 : 38(25)	Sein
Stanulla et al., Blood, 2009	Gène-candidat	Sang, Moelle osseuse	Cohorte Cohorte ALL-BFM SC : 105 Pas de SC : 814 Test exact du χ^2 (OR)	Allemagne	Médiane (jusqu'au SC) Année : n (%) < 3 : 33 (31,4) ≥ 3 - < 10 : 56 (53,3) ≥ 10 : 16 (15,2)	LAL	> 1979	Année : n (%) < 1-6 : 72 (68,6) ≥ 6 : 33 (31,4)	Tous types
Mertens et al., Cancer, 2004	Gène-candidat	Cellules buccales	Cohorte CCSS SC: 133 Pas de SC : 523 Modèle de Cox (RR)	EUA	Année : n (%) 5-10 : 3 (1) 11-15 : 73 (11) 6-20 : 197 (30) 21-25 : 195 (30) < 25 : 182 (28)	LH (traité par radiothérapie)	1970 - 1986	Année : n (%) 0-5 : 23 (3) 6-10 : 84 (13) 11-15 : 291 (45) 16-20 : 252 (39)	Tous types
Blanco et al., Pharmacogenetics, 2002	Gène-candidat	Sang, Moelle osseuse	Cas-témoins SJCRH, POG, DFW Cas : 53 Témoins : 224 Test exact du χ^2 (OR)	EUA	Temps médian d'apparition de LM : 3,5 (0,83-12,4)	ALL	NA	Médiane (min-max) Cas (gène MLL réarrangé) : 8,6 (17-18,5) Cas (MLL non réarrangé) : 8,1 (2,6-17,3) Témoins : 7,2 (0,1-18,8)	LM

Article	Méthode de séquençage	ADN	Design de l'étude Cohorte Test statistique	Pays, origine	Durée de suivi	Cancers pédiatriques	Année de traitement	Age au diagnostic du 1er cancer	Seconds cancers
Woo et al., Leukemia, 2000	Gène-candidat	Sang, Moelle osseuse	Cas-témoins SJCRH, DFW, CHOP, POG SC : 57 Pas de SC : 245 Test exact du χ^2 (OR)	EUA	Les témoins sont suivis pendant au moins 2,5 ans après le diagnostic de LAL.	LAL	NA	Médiane (quartiles) Cas : 7,1 (3,5-11,1) Témoins : 5,5 (3,4-10,2)	LM
Opstal-van Winden et al., Blood, 2019 ^a	GWAS ^c	Sang	Cas-témoins CCSS, cohorte britannique de LH, cohorte néerlandaise LH SC : 327 (≤ 20 : 250) Pas de SC : 491 (316) Régression logistique (OR)	EUA, Pays-Bas, Grande-Bretagne Ajustement sur l'ascendance et l'ethnicité	Médiane (min-max) Cas : 24 (9 -46) Témoins : 30 (9-49)	LH	1964-1999	Médiane (min-max) Cas : 19 (10-40) Témoins : 22 (6-40)	Sein
Morton et al., J Natl Cancer Inst, 2017	GWAS	CCSS: Sang, Salive, Cellules buccales SJLIFE: Sang	Cohorte CCSS and SJLIFE SC: 207 Pas de SC : 2 774 Modèle de Cox (HR)	EUA Ascendance européenne	Age médian au moment du cancer du sein 39,2 (3,1-43,2) Âge médian au dernier suivi (quartiles) 32,9 (27,7-39,3)	Tous types	CCSS : 1970-1986 SJLIFE : 1962-2005	Médiane (quartiles) SC : 15,6 (13,5-18,0) Pas de SC : 6,1 (2,9-12,7)	Sein
Best et al., Nat Med, 2011	GWAS	Sang, Salive	Case-Control CCSS SC : 158 (96+62) Pas de SC : 153 (82+71) Test exact du χ^2 (HR)	EUA Ascendance européenne	Durée moyenne (ET) avant le second cancer ^b Cas : 20 ans (5,8) Médiane (min-max) ^b Témoins : 32 (27-38)	LH	1970-1986	Moyenne (ET) ^b Cas : 15,9 (3,0) Témoins : 15,6 (2,8)	Tous types

Article	Méthode de séquençage	ADN	Design de l'étude Cohorte Test statistique	Pays, origine	Durée de suivi	Cancers pédiatriques	Année de traitement	Age au diagnostic du 1er cancer	Seconds cancers
Morton et al., JCO Precis. Oncol, 2020	WES	Sang, Salive, Cellules buccales	Cas-Témoins CCSS Cas : 1 108 Témoins : 3 883 Régression logistique conditionnelle (OR)	EUA Ascendance européenne	Moyenne SC : 24,7 ans Pas de SC : 32 ans	Tous types	1970-1986	Année : n (%) Cas : <5 : 274 (24.7) / 5-9 : 205 (18.5) 10-14 : 299 (27) / ≥15 : 330 (29.8) Témoins : <5 : 1 678 (43.2) / 5-9 : 879 (22.6) 10-14 : 738 (19) / ≥15 : 588 (15.1)	SC liés à la radiothérapie
Qin et al., JCO, 2020	WGS + WES	Sang, Salive, Cellules buccales	Cohorte SJLIFE SC : 495 Pas de SC : 3 907 Modèle de Cox (RR)	EUA Ascendance européenne	Médiane (min-max) Cohorte : 22,3 (14,2-32,1) Cas : 26,5 (19,1-33,9)	Tous types	>1962	Médiane (min-max) 6,3 (2,8-12,5)	Tous types
Wang et al., JCO, 2018	WGS + WES	Sang	Cohorte SJLIFE SC : 492 Pas de SC : 2 514 Modèle de Cox (RR)	EUA	Médiane (min-max) Cohorte : 28 (6,1-54,7) SC : 25,6 (18,6-32,7)	Tous types	>1962	Médiane (min-max) 7,1 (0,1-23,6)	Tous Types
Wang et al., Clin Cancer Res, 2018	Étude de séquençage	Sang	Cas-Témoins SJLIFE, only female survivors Cas : 47 Témoins : 1 086 Régression logistique (OR)	EUA Ascendance européenne	Médiane (min-max) Cas : 31,2 (13,4-51,2) Témoins : 28,1 (5,6-53,5)	Tous types	>1962	Médiane (min-max) SC : 15,2 (3,1-21,2) Pas de SC : 6,1 (0,01-22,7)	Sein

a Seuls les résultats pour les patients de moins de 20 ans sont présentés pour ces études.

b Caractéristiques de la cohorte de découverte

c GWAS réalisée à l'aide d'une puce Illumina iSelect Array personnalisée, comprenant 211 155 variants, spécialement développée pour le projet « *European Collaborative Oncological Gene-Environment Study* » (iCOGS array).

Abréviation : BFM, *Berlin-Frankfurt-Münster* ; CCSS, *Childhood Cancer Survivor Study* ; CHOP, *Children's Hospital of Philadelphia* ; COG, *Children's Oncology Group* ; DFW, *Children's Medical Center/Southwestern University* ; ET, *Ecart-Type* ; EUA, *Etats-Unis d'Amérique* ; GWAS, *étude pangénomique* ; LAL, *Leucémie aigüe lymphoblastique* ; LH, *Lymphome Hodgkinien* ; LM, *Leucémie Myéloïde* ; SC, *Second Cancer* ; MLL, *mixed lineage leukemia* ; POG, *Paediatric Oncology Group* ; SJCRH, *St Jude Children's Research Hospital* ; SNC, *Système Nerveux Central* ; WGS, *Whole Genome Sequencing* ; WES, *Whole Exome Sequencing*

2. Les études gènes-candidats

Au total, onze articles ont évalué l'association génétique du risque de seconds cancers par des études de type gène candidats. Parmi ceux-ci, quatre se sont intéressés au risque de cancers secondaires dans son ensemble tandis que sept se sont concentrés sur des types spécifiques. Les résultats sont présentés dans le Tableau 12.

Ces études se sont concentrées sur des gènes impliqués dans le métabolisme des médicaments, comme *GSTM1*, *GSTT1*, *TPMT*, *CYP1A1* et *NQO1*, ainsi que sur des gènes impliqués dans la réparation de l'ADN, tels que *XRCC1*, *XRCC3* et *NBN* ou sur des gènes connus pour leurs effets oncogéniques.

a. Etudes gènes-candidat sur le risque de seconds cancers

Une analyse des gènes *GSTM1*, *GSTT1* et *XRCC1* a été réalisée pour estimer le risque de seconds cancers après traitement pour un lymphome hodgkinien (Mertens et al., 2004). Ces gènes ont été sélectionnés du fait de leurs variations fréquentes dans la population et de leur activité pouvant affecter la sensibilité aux seconds cancers induits par la radiothérapie. Les gènes GST jouent un rôle crucial dans la détoxification des composés oxydatifs pouvant survenir après traitement par chimiothérapie et/ou radiothérapie. Le gène *XRCC1* est impliqué dans la réparation des cassures simples de l'ADN causées par les radiations. Cette étude suggère un rôle modeste du génotype nul *GSTT1* dans l'augmentation du risque de développer un second cancer globalement (Risque Relatif [RR] = 1,5 ; intervalle de confiance à 95 % [IC 95 %] : 1,0-2,3) et spécifiquement dans le champ de radiation (RR = 1,4 ; IC 95 % : 0,9-2,1).

Dans une sous-analyse chez des patients ayant été traités pour un neuroblastome, une étude a évalué 66 variants génétiques dans 47 gènes précédemment associés au développement de cancer secondaire dans une revue incluant les patients ayant eu un cancer à tout âge (Applebaum et al., 2017). Les résultats indiquent un risque élevé de second cancer associé au variant rs861539 du gène *XRCC3* (Odds Ratio [OR] = 2,04 ; IC 95 % : 1,19-3,46 ; P = 0,006) et un risque réduit de second cancer associé au variant

rs17036651 du gène *MSH2* (OR = 0,26 ; IC 95 % : 0,09-0,81 ; P = 0,008).

Pour étudier le risque de développer un second cancer après une leucémie, une autre étude a porté sur onze variants dans cinq gènes de réparation de l'ADN (Erčulj et al., 2012). Cette étude montre un risque significativement accru de développer un second cancer associé au variant rs709816 (RR = 4,36 ; IC 95 % : 1,2-16,0 ; P = 0,026) du gène *NBN*, et un risque diminué associé au variant rs1799794 (RR = 0,20 ; IC 95 % : 0,04-0,99 ; P = 0,049) du gène *XRCC3*.

b. Etudes gènes-candidat sur le risque de seconds cancers spécifiques

Second cancer du Système Nerveux Central (SNC)

Une étude a évalué l'association entre 129 variants, associés au cancer du SNC chez les adultes, situés dans 46 gènes et le risque de développer ce type de cancer en tant que second cancer après traitement d'un cancer pédiatrique (X. Wang et al., 2017). L'étude a porté sur le SNC globalement ainsi que sur certains sous-types spécifiques tels que le méningiome et le gliome. L'association la plus importante concerne le variant rs15869 du gène *BRCA2* (OR = 2,5 ; IC 95 % : 1,5-4,2 ; P < 0,001). Une tendance similaire a été observée pour ce variant concernant le risque de gliome et de méningiome. L'étude a également mis en évidence une association entre le variant rs1805389 du gène *LIG4* et le risque de développer un second cancer du SNC globalement, et spécifiquement de gliome, et une association entre les variants rs1673041 du gène *POLD1* et rs25489 du gène *XRCC1* et le risque de SNC globalement et spécifiquement de méningiome. Enfin, des associations spécifiques ont été détectées entre le variant rs1801270 du gène *CDKN1A* et le risque de méningiome et entre le variant rs1805389 du gène *TP53* et le risque de gliome.

Leucémie secondaire

Deux études (Blanco et al., 2002; Woo et al., 2000) ont examiné le risque de leucémie myéloïde aiguë en tant que second cancer et des variations dans des gènes impliqués

dans le métabolisme des médicaments et la détoxification (*CYP3A5*, *CYP3A4*, *NQO1*, *GSTM1*, *GSTT1*), en raison de l'association bien documentée dans la littérature entre ce type de seconds cancer et la chimiothérapie reçue lors du traitement du cancer pédiatrique. Aucune association n'a été trouvée avec les variants étudiés ; cependant, une étude (Woo et al., 2000) a montré une prévalence significativement plus élevée du double génotype nul *GSTM1/GSTT1* chez les patients masculins atteints d'une leucémie secondaire.

Tableau 12 - Résultats des études gène-candidats incluses dans la revue systématique.

Etude	Gènes d'intérêt	Type de SC	Sous-analyses	Gènes	Variants	Allele	OR IC 95 %	p-value	Correction – multiplicité des tests	Variables d'appariement / d'ajustement
Richard et al., 2020	60 variants dans/ou à proximité de 14 gènes associés à la gestion des télomères	Thyroïde	Tous les individus	POT1	rs58722976	G	HR = 6,1 [2,4-15,5]	0,0001	Correction de Bonferroni	Ajustement : sexe, décennie de naissance, âge au diagnostic du 1 ^{er} cancer, type du cancer pédiatrique, exposition aux traitements, nodules thyroïdiens.
			Ajusté sur l'ascendance (Européenne et Africaine)	POT1	rs58722976	G	HR = 8,0 [2,3-27,2]	0,0009		
			Seulement sur les individus d'ascendance européenne	POT1	rs58722976	G	HR = 18,9 [3,5-101,7]	0,0006		
Gramatges et al., 2019	9 variants précédemment identifiés comme associés à la longueur du télomère des leucocytes	Thyroïde	-	No association					FDR	Ajustement : année de naissance, type du cancer pédiatrique, âge au diagnostic du 1 ^{er} cancer, sexe, exposition aux traitements.
Wang et al., 2017	129 variants dans 46 gènes impliqués dans les tumeurs primaires du SNC chez l'adulte	SNC	-	BRCA	rs15869	C	2,5 [1,5-4,2]	<0,001	NA	Appariement : type du cancer pédiatrique, année du diagnostic, origine ethnique, durée de suivi Ajustement : âge au diagnostic du 1 ^{er} cancer, sexe, irradiation crâniale
				LIG4	rs1805389	A	3,7 [1,4-9,9]	0,008		
				TP53	rs8079544	G	4,0 [1,1-15,2]	0,04		
				PHLDB1	rs498872	G	1,8 [1,0-3,3]	0,06		
				CCDC26	rs4295627	A	1,7 [0,9-3,2]	0,06		
				XRCC1	rs25489	G	3,6 [1,1-11,5]	0,03		
				POLD1	rs1673041	A	1,9 [1,1-3,1]	0,01		
				ERCC1	rs11615	A	1,6 [1,0-2,5]	0,04		
	XRCC5	rs828699	A	1,5 [1,0-2,2]	-					

Etude	Gènes d'intérêt	Type de SC	Sous-analyses	Gènes	Variants	Allele	OR IC 95 %	p-value	Correction – multiplicité des tests	Variables d'appariement / d'ajustement
Wang et al., 2017 (suite)	129 variants dans 46 gènes impliqués dans les tumeurs primaires du SNC chez l'adulte	Gliome	-	TP53	rs2909430	G	7,5 [1,5-36,9]	0,01	NA	Appariement : type du cancer pédiatrique, année du diagnostic, origine ethnique, durée de suivi Ajustement : âge au diagnostic du 1 ^{er} cancer, sexe, irradiation crâniale
				LIG4	rs1805389	A	19,7 [2,5-156,4]	0,005		
				BRCA2	rs15869	C	3,7 [1,3-10,2]	0,01		
				TPMT	rs2518471	G	3,9 [0,9-16,7]	0,07		
		Meningiome		BRCA2	rs15869	C	2,4 [1,2-4,9]	0,01		
				XRCC5	rs828699	A	1,7 [0,9-3,1]	0,07		
				XRCC1	rs25489	G	6,9 [1,1-43,6]	0,04		
				CDKN1A	rs1801270	C	9,6 [1,1-85,4]	0,04		
				POLD1	rs1673041	A	2,0 [1,0 – 3,8]	0,04		
Applebaum et al., 2017	66 variants dans 47 gènes précédemment associés au développement de SC	Tous types	-	XRCC3	rs861539	T	2,04 [1,19-3,46]	0,006	NA	Appariement : groupe de risque, ascendance génétique
				MSH2	rs17036651	G	0,26 [0,09-0,81]	0,008		
Vodusek et al., 2016	Loci dans les gènes liés à la défense contre les agents antioxydants	Thyroïde	-	No association					NA	Appariement : Sexe, âge au diagnostic du 1 ^{er} cancer, sexe, type du cancer pédiatrique, traitement (en particulier la RT)
Erčulj et al., 2012	11 polymorphismes de réparation de l'ADN	Tous types	-	NBN	rs709816	G	RR = 4,36 [1,2-16,0]	0,026	NA	Appariement : radiothérapie crânienne, administration d'épipodophyllotoxines, doses cumulées d'anthracyclines, sexe et âge au diagnostic du 1 ^{er} cancer
				XRCC3	rs1799794	G	RR = 0,20 [0,04-0,99]	0,049		

Etude	Gènes d'intérêt	Type de SC	Sous-analyses	Gènes	Variants	Allele	OR IC 95 %	p-value	Correction – multiplicité des tests	Variation d'appariement / d'ajustement
Ma et al., 2012	14 variants associés au risque de cancer du sein	Sein	< 20 ans	FGFR2	rs1219648	G	1,70 [1,16-2,50]	-	Correction de Bonferroni	Appariement : âge et date au diagnostic du 1 ^{er} cancer, région, durée de suivi
Stanulla et al., 2009	TPMT	Tous types	-	No association	NA	NA				
Mertens et al., 2004	GSTM1, GSTT1, XRCC1	Tous types	SC au sein du champ d'irradiation	GSTM1 / GSTT1	GSTM1 - / GSTT1 +		RR = 1,4 [0,91-2,12]	-	NA	Ajustement : âge au diagnostic et statut tabagique
Blanco et al., 2002	CYP3A4, CYP3A5, NQO1	LM	-	No association					NA	NA
Woo et al., 2000	GSTM1, GSTT1	LM	Homme uniquement	GSTM1 / GSTT1	Génotype double nul	-	-	0,036	NA	Ajustement : ethnicité

Abréviations : FDR, *False Discovery Rate* ; HR, Hazard Ratio ; LM, Leucémie Myéloïde ; OR, Odds Ratio ; RR, Risque Relatif ; RT, Radiothérapie ; SC, Second Cancer ; SNC, Système Nerveux Central

Second cancer de la thyroïde

Deux études ont évalué le risque de second cancer de la thyroïde, associé aux variants des gènes impliqués dans le contrôle de la longueur des télomères (Gramatges et al., 2019; Richard et al., 2020).

Dans une étude précédente Gramatges et al. avait montré que les patients ayant développé un second cancer de la thyroïde avaient des télomères plus courts (Gramatges et al., 2014). Des télomères trop courts peuvent entraîner des fusions chromosomiques et une instabilité génomique, augmentant ainsi le risque de cancer. Les traitements comme la chimiothérapie et les radiations ionisantes peuvent endommager les télomères, rendant les patients traités pour un premier cancer plus vulnérables à de nouveaux cancers. Ces résultats justifient l'étude en particulier des gènes liés à la gestion de la longueur des télomères comme candidats pour de nouvelles recherches.

Dans ce sens, une des deux études incluses a étudié 9 variants précédemment identifiés comme étant liés à la longueur des télomères (Gramatges et al., 2019). Aucune association entre ces variants et le risque de développer un second cancer de la thyroïde n'a été observée. En revanche, une étude ultérieure (Richard et al., 2020) sur 60 variants situés à proximité ou dans des gènes associés à la gestion des télomères, a révélé une association significative entre le variant rs58722976 du gène *POT1* et le risque de second cancer de la thyroïde (Hazard Ratio [HR] = 6,1 ; IC 95 % : 2,4-15,5 ; P = 0,0001).

Enfin, une dernière étude sur le risque de second cancer de la thyroïde a étudié 6 variants situés dans des gènes antioxydants (*SOD2*, *CAT*, *CPI*, *GSTP1*, *GSTM1*, *GSTT1*) et n'a montré aucune association (Vodusek et al., 2016a).

3. Les études pangénomiques (GWAS)

A partir d'une étude cas-témoins, trois variants ont été identifiés comme étant

significativement associés au risque de développer un second cancer parmi les patients d'ascendance européenne ayant été traités par radiothérapie pour un cancer pédiatrique (Best et al., 2011). Les résultats sont présentés dans le Tableau 13.

Deux variants en déséquilibre de liaison (LD), impliqués dans le contrôle du gène *PRDM1*, sont associés à un risque élevé de second cancer (OR = 15,78 ; IC 95 % : 3,31–75,18 et OR = 11,4 ; IC 95 % : 3,23–40,25 ; respectivement) chez les patients homozygotes pour les variants rs4946728 et rs1040411. Ces deux variants ont été validés dans une cohorte de validation. Le gène *PRDM1*, impliqué dans des processus tels que la différenciation, la prolifération et l'apoptose, a précédemment été reconnu comme un suppresseur de tumeur dans les lymphocytes B (Mandelbaum et al., 2010). Bien que le troisième variant (rs8083533), situé dans la région intronique du gène *TAF4B*, ait été significativement associé au risque de second cancer (OR = 9,65 ; IC 95 % : 3,04–30,65), ce dernier n'a pas pu être répliqué.

Dans une deuxième étude menée dans la cohorte CCSS (Morton et al., 2017), trois variants ont été identifiés comme associés au risque de développer un second cancer du sein chez les patientes d'ascendance européenne ayant reçu une radiothérapie au niveau du sein lors du traitement du cancer pédiatrique.

Le premier variant, rs4342822, situé à proximité du gène *PROX1*, est associé à un risque élevé de second cancer du sein chez les patientes ayant reçu une dose de radiation supérieure à 10 Gy (RR = 1,92 ; IC 95 % : 1,49–2,44 ; $P = 7,09 \times 10^{-9}$). Cette association est d'autant plus intéressante que le facteur de transcription *PROX1*, impliqué dans la différenciation cellulaire et la migration, a déjà été associé au cancer du sein ainsi qu'à divers types de cancers (Elsir et al., 2012).

Dans ce même groupe, un autre variant, rs74949440 dans le gène *TAGLN*, est également associé à un risque de second cancer du sein (RR = 2,59 ; IC 95 % : 1,62–4,16 ; $P = 8,4 \times 10^{-8}$). Ce gène est par ailleurs surexprimé dans le cancer du sein triple négatif (Rao et al., 2015). Ces résultats viennent ainsi appuyer l'importance de

l'interaction entre la génétique et le traitement reçu lors du premier cancer. La présence de ces variants génétiques chez les patientes ayant reçu des doses élevées de radiothérapie semble accroître le risque de second cancer du sein, potentiellement en favorisant un phénotype pro-prolifératif et pro-invasif des cellules tumorales à la suite de leur exposition aux radiations ionisantes. Par ailleurs, les patientes ayant reçu des doses de radiation de moins de 10 Gy au niveau du sein ont montré une association significative entre le variant rs17020562 situé dans la région intronique du gène *RPS6KC1* et le risque du cancer du sein chez les patientes ayant reçu (RR = 44,52 ; IC 95 % : 15,06-131,62 ; P = 5,00 x 10⁻⁹) (Morton et al., 2017).

Cette étude a également tenté de reproduire les résultats de Best *et al.*, sans réussir à obtenir des résultats similaires. Cette disparité pourrait être attribuée à des différences dans les critères de sélection des individus dans les deux études et aux tailles d'échantillon relativement modestes.

Tableau 13 - Résultats des études pangénomiques incluses dans la revue systématique.

Etude	Type de SC	Sous-analyses	Gènes	Variants	Allèle	OR IC 95 %	p-value	Correction – Multiplicité des tests	Variables d'ajustement
Opstal-van Winden et al., 2019 ^a	Sein	Âge au traitement du LH ≤ 20 ans	SPR incluant 9 variants qui ont montré une interaction statistiquement significative avec le cancer du sein induit par la RT		Tertile inférieur	REF	P trend = 0,027	NA	SPR d'une étude précédente, âge au diagnostic du LH, pays et ethnies
					Tertile intermédiaire	1,3 [0,7-2,2]	0,3722		
					Tertile supérieur	1,6 [1,0-2,8]	0,071		
Morton et al., 2017	Sein	< 10 Gy au sein	RPS6KC1 (variant rare)	rs17020562	C	HR = 44,52 [15,06-131,62]	5,00 × 10 ⁻⁹	Permutation	Cohorte, agent alkylant ou chimiothérapie à base de platine, composante principale pour prendre en compte la stratification de la population
		≥ 10 Gy au sein	PROX1	rs4342822	G	HR = 1,92 [1,49-2,44]	7,09 × 10 ⁻⁹		
			TAGLN (variant rare)	rs74949440	T	HR = 2,59 [1,62-4,16]	8,40 × 10 ⁻⁸		
Best et al., 2011	Tous types	-	PRDM1	rs4946728	C	OR = 3,32 [2,25-4,90]	5,99 × 10 ⁻¹⁰	Permutation	Âge au diagnostic du LH, sexe, année au diagnostic du LH, irradiation des gonades chez les femmes et chimiothérapie alkylante
				rs1040411	T	OR = 2,39 [1,7-3,30]	1,18 × 10 ⁻⁷		
				TAF4B	rs808533	T	OR = 1,86 [1,32-2,62]		

a.GWAS réalisée à l'aide d'une puce Illumina iSelect Array personnalisée, comprenant 211 155 variants, spécialement développée pour le projet « *European Collaborative Oncological Gene-Environment Study* » (iCOGS array).

Abréviations : IC, Intervalle de Confiance ; HR, Hazard Ratio ; LH, Lymphome Hodgkinien ; OR, Odds Ratio ; RT, Radiothérapie ; SC, Second Cancer ; SPR, Score Polygénique de Risque.

4. Les études de séquençage haut débit (NGS – WGS & WES)

Les études de GWAS, basées sur le génotypage via des puces prédéfinies avec un panel de variants génétiques, ne sont généralement pas efficaces pour identifier des variants rares ou ultra-rares, qui pourraient pourtant avoir un rôle important dans le risque de seconds cancers. Les progrès récents, comme le séquençage de l'exome (WES) et le séquençage du génome entier (WGS), offrent maintenant de nouvelles possibilités pour étudier l'impact des variants génétiques rares sur les maladies telles que les seconds cancers.

Les résultats correspondants aux études WES et WGS sont présentés dans le Tableau 14.

a. Etudes NGS sur le risque de seconds cancers

La première étude de séquençage du génome entier (Whole Genome Sequencing, WGS) a été menée par Wang et al. au sein de la cohorte SJLIFE (Z. Wang, Wilson, et al., 2018). Les auteurs ont évalué le risque de développer un second cancer chez les survivants d'un cancer pédiatrique porteurs de mutations considérées comme pathogènes ou "probablement" pathogènes dans une liste de 60 gènes de prédisposition au cancer, créée spécialement pour l'étude. Cette liste comprend des gènes connus pour augmenter le risque de cancer lorsqu'ils présentent une mutation. Ces mutations sont héritées de manière dominante sur les chromosomes et peuvent entraîner un risque de cancer allant de modéré à élevé. Les auteurs ont ainsi montré que les porteurs de variants dans ces 60 gènes étaient plus susceptibles de développer un second cancer (RR = 1,8 ; IC 95 % : 1,2-2,6 ; P = 0,003) ou plusieurs seconds cancer (RR = 3,1 ; IC 95 % : 1,5-56,3 ; P = 0,002). De plus, après stratification en fonction de l'exposition aux radiations, les survivants n'en ayant pas reçu et porteurs de variants pathogènes/présumés pathogènes sont plus susceptibles de développer un second cancer (RR = 4,7 ; IC 95 % : 2,4-9,3) ou plusieurs seconds cancers (RR = 18,6 ; IC 95 % : 3,5-99,3).

Dans la même cohorte Qin et al. (Qin et al., 2020) ont étudié le risque de seconds cancers associé à des mutations dans une liste comprenant 127 gènes appartenant aux 6 voies majeures de réparation de l'ADN. L'étude a révélé un risque accru de développer des cancers secondaires chez les porteurs de variants dans les voies de la recombinaison homologue (RR = 1,5 ; IC 95 % : 1,1-2,2 ; P = 0,017), de la réparation des mésappariements (RR = 2,1 ; IC 95 % : 1,2-3,7 ; P = 0,014), ou de la réparation par excision de nucléotides (RR = 1,6 ; IC 95 % : 1,1-2,5 ; P = 0,027).

Tableau 14 - Résultats des études WES et WGS incluses dans la revue systématique.

Etude	Gènes d'intérêt	Sous-analyses	Type de SC	Gène/Voie biologique	RR [IC 95 %]	p-value	Correction – Multiplicité des tests	Covariables d'appariement / d'ajustement
Morton et al., 2020	476 gènes impliqués dans la réponse aux dommages de l'ADN ou dans les syndromes cliniques de sensibilité aux radiations	RT-SC hors -champs de l'irradiation excl, BCC	RT-SC	HRR pathway	OR = 2,6 [1,7-3,9]	4,79×10 ⁻⁵	Bonferroni	Appariement : sexe, type de cancer pédiatrique, âge au diagnostic, survie, dose de radiation à l'emplacement du SC
				FANCM	OR = 11,2 [4,1-30,8]	9,64×10 ⁻⁵		
	Variants rares de haute qualité classés comme délétères pour les protéines à l'aide de SnpEff et ClinVar	RT-SC	RT-SC	EXO1	OR = 5,9 [3,6-9,8]	3,31×10 ⁻⁵		
				NEIL3	OR = 0,1 [0,0-1,0]	6,14×10 ⁻⁴		
	RT-SC excl, BCC sans RT	RT-SC	TP53	∞	9,99×10 ⁻⁸			
Qin et al., 2020	127 gènes dans 6 grandes voies de réparation de l'ADN Mutations pathogènes/probablement pathogènes		Tous types	MP dans RRH ^a	1,5 [1,1-2,2]	0,051	FDR	Ajustement : âge atteint, sexe et ascendance déterminés par l'ACP et les variables cliniques
				MP dans REN ^b	1,6 [1,1-2,5]	0,054		
				MP dans MMR ^c	2,1 [1,2-3,7]	0,051		
			Sein	MP dans RRH ^a	3,7 [1,8-7,7]	0,0016		
				MP dans FA ^d	3,7 [1,3-10,6]	0,034		
			Sein	MP dans RRH ^a	4,4 [1,6-12,4]	0,02		
				MP dans FA ^d	5,5 [1,5-14,2]	0,03		
			Sein	MP dans RRH ^a	4,4 [1,7-11,4]	0,01		
				MP dans FA ^d	4,2 [1,4-12,32]	0,03		
			Sarcome	MP dans RRH ^a	14,9 [4,0-38,0]	0,001		

Etude	Gènes d'intérêt	Sous-analyses	Type de SC	Gène/Voie biologique	RR [IC 95 %]	p-value	Correction – Multiplicité des tests	Covariables d'appariement / d'ajustement
Wang et al., 2018	Gènes SJCPG60 ^e : associations établies avec un risque de cancer monogénique hérité de manière autosomique dominante avec une pénétrance modérée à élevée	Patients traités par RT	Tous type	Porteurs de variants au sein de la liste de gènes SJCPG60	1,8 [1,2- 2,6]	0,003	Correction de Holm	Ajustement : sexe, âge diagnostic du 1 ^{er} cancer et traitement
			Sein		9,4 [4,8-18,2]	< 0,001		
			Sarcome		10,9 [4,7-25,2]	< 0,001		
			Plusieurs SC		3,1 [1,5-6,3]	0,002		
	Mutations pathogènes/probablement pathogènes	Patients non traités par RT	Tous types		4,7 [2,4-9,3]	-		
			Sein		7,7 [2,4-44,4]	-		
			Carcinome cutané		11,0 [2,9-41,4]	-		
			Plusieurs SC		18,6 [3,5-99,3]	-		
Wang et al., 2018	SPR : 170 loci associés au cancer du sein	RT à la poitrine	Sein	Porteurs de Variants	21,8 [7,1-66,8]	<0,001	NA	Ajustement : âge au diagnostic du 1 ^{er} cancer, RT, doses cumulées d'agents alkylants et d'anthracyclines, les vecteurs propres des génotypes et l'âge atteint
				SPR Q5	2,7 [1,0-7,3]	0,05		
	Mutations pathogènes/probablement pathogènes	Pas de RT		Porteurs de variants	10,3 [3,6-29,8]	<0,001		
				SPR Q5	3 [1,1-8,1]	0,03		
				Porteurs de variants	42,8 [9,9-183,8]	<0,001		

Abréviations : BCC, carcinome basocellulaire ; FA, voie de l'anémie de Fanconi ; FDR, *False Discovery Rate* ; IC, Intervalle de Confiance ; MMR, voie de la réparation des mésappariements ; MP, Mutation Pathogène ; OR, Odds Ratio ; REN, voie de la réparation par excision de nucléotides ; RR, Risque Relatif ; RRH, voie de la réparation par recombinaison homologue ; RT, Radiothérapie; RT-SC, Second Cancer associés à la Radiothérapie ; SC, Second Cancer ; SPR, Score Polygénique de Risque

^a PM dans la voir RRH (n=42) : *BABAM1, BARD1, BLM, BRCC3, EME1, EME2, EXO1, FAM175A, GEN1, H2AFX, MDC1, MRE11A, MUS81, NBN, POLQ, RAD50, RAD51AP1, RAD51B, RAD51D, RAD52, RAD54B, RAD54L, RBBP8, RECQL, REV1, RTEL1, SHFM1, SLX1A, SLX1B, TP53BP1, UIMC1, XRCC2, XRCC3, BRCA1, BRCA2, BRIP1, PALB2, RAD51, RAD51C, SLX4, RPA1, RPA2,*

^b PM dans la voie REN (n=33) : *CCNH, CDK7, CETN2, DDB1, DDB2, ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, ERCC6, ERCC8, GTF2H1, GTF2H2, GTF2H3, GTF2H4, GTF2H5, LIG1, MMS19, MNAT1, POLK, RAD23A, RAD23B, RPA3, RPA4, UVSSA, XAB2, XPA, XPC, POLD1, POLE, RPA1, RPA2REN*

^c PM dans la voie FA (n=20) : *C19orf40, C1orf86, FAN1, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, BRCA1, BRCA2, BRIP1, PALB2, RAD51, RAD51C, SLX4,*

^d PM dans la voie MMR (n=12) : *MLH1, MLH3, MSH2, MSH3, MSH4, MSH5, MSH6, PMS1, PMS2, PMS2L3, POLD1, POLE,*

^e SJCPG60 : *ALK, APC, BAP1, BMPR1A, BRAF, BRCA1, BRCA2, CBL, CDC73, CDH1, CDK4, CDKN1C, CDKN2A, CEBPA, DICER1, EPCAM, FH, GATA2, HRAS, KRAS, MAP2K1, MAP2K2, MAX, MEN1, MLH1, MSH2, MSH6, NF1, NF2, NRAS, PALB2, PAX5, PHOX2B, PMS2, PRKAR1, PTCH1, PTEN, PTPN11, RAF1, RB1, RET, RUNX1, SDHA, SDHAF2, SDHB, SDHC, SDHD, SHOC2, SMAD4, SMARCA4, SMARCB1, SOS1, STK11, SUFU, TMEM127, TP53, TSC1, TSC2, VHL, WT1,*

Morton et al. (Morton et al., 2020) ont également mené une étude de séquençage de l'exome chez des patients de la cohorte CCSS pour examiner le risque de second cancer associé à des variations rares dans 476 gènes impliqués dans la réponse aux dommages de l'ADN et la sensibilité aux radiations. Une association entre les variants de la voie de la recombinaison homologues et le risque de développer un second cancer, hors du champ de la radiothérapie reçue pour le premier cancer (OR = 2,6, IC 95 % : 1,7-3,9, P = $4,79 \times 10^{-5}$).

Parmi les gènes appartenant à cette voie, deux associations ont été observées. La première concerne le gène *FANCM* (OR = 11,2, IC 95 % : 4,1-30,8, P = $9,64 \times 10^{-5}$). De précédents travaux ont déjà établi un lien avec le risque cancer, mais également avec la toxicité liée à la chimiothérapie chez les porteurs de variants de ce gène (Catucci et al., 2018). Ainsi, ces résultats renforcent l'importance de ce gène dans le contexte de la toxicité liée au traitement du cancer pédiatrique. La deuxième association concerne le gène *EXO1*, impliqué dans de multiples voies de la réparation (OR = 9,6 ; IC 95 % : 2,8-33,4 ; P = $5,05 \times 10^{-3}$). Dans une précédente étude (Bolderson et al., 2010), ce gène a été lié à une susceptibilité accrue aux radiations, renforçant ainsi les conclusions de Morton et al. selon lesquelles le gène *EXO1* présente une tendance à être associé au risque de seconds cancers induits par la radiothérapie.

b. Etudes NGS sur le risque de seconds cancers spécifiques

Second cancer du sein

Wang et al (Z. Wang, Wilson, et al., 2018) ont également réalisé des analyses par sous-type de cancer, stratifiées sur la dose de radiation reçue. Ils ont mis en évidence un risque accru de second cancer du sein chez les patientes porteuses de mutations pathogènes ou probablement pathogènes dans la liste de gènes préétablie discutée dans la partie précédente (RR = 9,4 ; IC 95 % : 2,9-41). Les patientes n'ayant pas reçu de radiothérapie et celles ayant reçu de la radiothérapie ont un risque plus important du cancer de sein (RR = 7,7 ; IC 95 % : 2,4-44,4 et RR = 13,9 ; IC 95 % : 6,0-32,2),

respectivement).

Dans une autre étude, les mêmes auteurs (Z. Wang, Liu, et al., 2018) ont évalué le risque de cancer du sein à l'aide d'un score polygénique de risque (PRS) incluant 170 loci associés au risque de cancer du sein. Ils ont mis en évidence un risque plus élevé de cancer secondaire du sein parmi les patientes dont le PRS est dans le plus haut quintile comparé à celles dont le PRS est dans le plus faible (RR = 2,7 ; IC 95 % : 1,0-7,3). Ce résultat est également retrouvé lorsque seules les patientes ayant reçu de la radiothérapie sont incluses (RR = 3 ; IC 95 % : 1,1-8,1).

Finalement, Qin et al. (Qin et al., 2020) en se concentrant sur les 6 majeures voies de la réparation et le cancer du sein, ont pu démontrer que les porteuses de variants pathogéniques dans la voie de recombinaison homologue ont un risque de second cancer du sein significativement augmenté (RR = 3,7 ; IC 95 % : 1,8-7,7). Le risque est d'autant plus important chez les patientes ayant reçu des doses de radiation ≥ 20 Gy au niveau de la poitrine (RR = 4,4 ; IC 95 % : 1,6-12,4), ou bien chez celles ayant reçu une dose cumulée d'anthracyclines ≥ 98 mg/m² (RR = 4,4 ; IC 95 % : 1,7-11,4). Des résultats similaires ont été également observés pour les patientes porteuses de mutations dans la voie de l'Anémie de Fanconi (RR = 3,7 ; IC 95 % : 1,3-10,6), avec également un risque accru chez les patientes ayant reçu des doses de radiation ≥ 20 Gy au niveau de la poitrine (RR = 5,5 ; IC 95 % : 1,5-14,2), ou bien chez celles ayant une dose cumulée d'anthracyclines ≥ 98 mg/m² (RR = 4,2 ; IC 95 % : 1,4-12,3).

Sarcome secondaire

Dans la même étude que précédemment, Qin et al. (Qin et al., 2020) ont également pu observer que les porteurs de variants dans les gènes de la voie de la recombinaison et ayant reçu une dose ≥ 104.5 mg/m² d'agents alkylants (RR = 14,9 ; IC 95 % : 4,0-38,0).

V. Discussion

Cette revue systématique a permis de rassembler toutes les études ayant, à ce jour, réalisé une analyse d'association génétique sur le risque de développer un cancer secondaire après un traitement pour un cancer pédiatrique. Nous avons ainsi pu dresser un état des lieux des différents variants génétiques, gènes et voies biologiques associés à ce risque.

Une revue préalable réalisée par Bhatia et al.(Bhatia, 2015) avait évalué le risque de second cancer chez les adultes et les enfants survivants de cancer. Toutefois, les spécificités des cancers pédiatriques et de leurs traitements nécessitent une étude distincte pour cette population. De plus, les articles inclus précédemment ne couvraient pas les recherches récentes utilisant les technologies avancées comme le séquençage nouvelle génération. Ainsi, ce travail représente la première revue exhaustive consacrée au risque génétique de seconds cancers après un cancer pédiatrique intégrant les avancées les plus récentes dans ce domaine de recherche. Cette revue offre une synthèse exhaustive des avancées dans les études génétiques des seconds cancers, débutant par les recherches sur les gènes candidats, elle aborde ensuite les GWAS et enfin se termine par les études basées sur le séquençage de nouvelle génération. Elle met donc en évidence les progrès les plus récents dans ce domaine de recherche.

Cette étude souligne qu'avant l'avènement du séquençage haut débit, les recherches étaient principalement basées sur les études gènes candidats, en se concentrant essentiellement sur deux voies biologiques : le métabolisme des médicaments et la réparation de l'ADN. Les gènes impliqués dans le métabolisme des médicaments comme *GSTM1*, *GSTT1*, *TPMT*, *CYP1A1* et *NQO1* ainsi que ceux de la réparation de l'ADN, tels que *XRCC1*, *XRCC3* et *NBN* ont été étudiés. Ces choix ont été motivés par l'importance de ces voies dans le traitement des cancers pédiatriques et par les résultats des recherches antérieures montrant leur association avec le risque de seconds cancers. Cependant, malgré la capacité des études sur les gènes candidats à

détecter des associations avec des gènes pertinents, elles restent limitées à cause de la sélection initiale, masquant potentiellement d'autres gènes d'intérêt pour le risque génétique de second cancer.

Avec les progrès technologiques notamment l'introduction des puces de génotypage, les études pangénomiques sont devenues plus accessibles. Ces études ne se concentraient plus spécifiquement sur une sélection des gènes particuliers, mais examinaient un large ensemble de variants répartis dans tout le génome.

Ainsi, dans cette revue, trois GWAS (Best et al., 2011; Morton et al., 2017; Opstal-van Winden et al., 2019) ont été considérées, dont deux ont révélé de nouveaux variants dans deux gènes, *PRDM1* et *PROX1*, associés respectivement à tous types de seconds cancers et au cancer secondaire du sein. Bien que les GWAS, présentent de nombreux avantages par rapport aux études de gènes candidats, elles sont toutefois limitées dans la détection de variants rares. En effet, elles se concentrent uniquement sur les variants communs, définis comme ceux retrouvés dans au moins 5 % d'une population donnée.

Les avancées récentes, comme le séquençage d'exome et le séquençage du génome entier, permettent aux chercheurs d'explorer les variants génétiques rares associés aux maladies. Cependant, la puissance statistique de ces études peut être limitée en raison de la taille restreinte des échantillons.

Notre revue systématique a inclus quatre études utilisant le séquençage du génome et de l'exome (WES/WGS) sur les cohortes américaines CCSS et SJLIFE. Ces travaux ont permis d'identifier de nouvelles associations génétiques avec le risque de second cancer en se focalisant sur une sélection restreinte de gènes, afin d'atteindre la puissance statistique nécessaire. Parmi ces résultats importants, deux gènes étaient notamment associés au risque de second cancer : le gène *FANCM*, associé au risque de second cancer hors champ de radiothérapie, et le gène *EXO1*, impliqué dans la réparation des mésappariements de l'ADN. Ce dernier a été identifié comme étant associé à une susceptibilité élevée de la toxicité liée à la chimiothérapie et au

développement du cancer. Ceci souligne l'importance des mécanismes de réparation de l'ADN dans la survenue du second cancer.

La littérature dans ce domaine est globalement peu abondante et limitée par la variabilité dans la conception des études, les méthodes d'identification des variants est la localisation des cancers primaires et secondaires. Certaines études se sont concentrées sur les seconds cancers radio-induits, tels que les cancers du sein et de la thyroïde, tandis que d'autres ont évalué tout type de cancers secondaires. En effet, cette revue de la littérature a révélé une grande hétérogénéité méthodologique entre les études, tant au niveau de la méthode de séquençage utilisée, que des hypothèses initiales concernant la sélection des gènes à étudier.

Les études WES/WGS manquent souvent de puissance statistique pour détecter des associations entre les variants génétiques rares, leur interaction avec le traitement et le risque de second cancer. En raison de cette puissance statistique insuffisante, seule une étude, parmi les 18 incluses, a évalué ces interactions gène-traitement (26), tandis que les autres ont réalisé des analyses stratifiées, selon la dose de chimiothérapie ou de radiothérapie. La dose de radiothérapie était souvent résumée à la dose reçue par la partie d'organe située dans le champ de l'irradiation (par exemple, la dose reçue au niveau du thorax pour le cancer du sein). Ce manque de puissance soulève des préoccupations quant à la capacité de ces études à saisir pleinement les interactions complexes entre les prédispositions génétiques et la réponse aux traitements. Ceci peut avoir des implications importantes pour le développement de stratégies thérapeutiques personnalisées chez les survivants du cancer pédiatrique.

De nombreuses études n'ont pas pu être répliquées ou validées, soulignant ainsi l'importance de futures recherches pour confirmer ces résultats. Par conséquent, la plupart des résultats ne peuvent pas encore être directement appliqués en pratique clinique pour adapter le traitement et le suivi des enfants atteints de cancer.

Une autre limitation des études génétiques est la stratification de la population, qui

peut conduire à la détection de faux positifs. La plupart des articles inclus se sont concentrés sur des individus d'ascendance européenne ce qui rend les résultats non généralisables aux populations non européennes. La problématique des tests multiples représente un enjeu majeur dans les études d'association génétique, en raison du grand nombre de variants génétiques testés. Différentes méthodes de correction existent pour les prendre en compte, comme la correction de Bonferroni, les méthodes de permutation ou FDR. Dans cette revue, la majorité des études GWAS et WES/WGS ont pris en compte les tests multiples. Cependant, la plupart des études gènes-candidats ne l'ont pas fait.

Compte tenu du jeune âge relatif de la population étudiée, il est primordial de considérer que l'effet des variants génétiques rares et communs sur le risque de second cancer pourrait évoluer avec l'âge et la durée du suivi. De plus, une grande partie des études incluses dans cette revue se base sur des échantillons d'ADN prélevés chez des survivants de cancer pédiatrique encore en vie au moment de l'étude, ce qui introduit un biais de survie potentiel.

A l'avenir, les collaborations internationales sont essentielles pour étudier le risque génétique des seconds cancers après un traitement d'un cancer pédiatrique. Ces collaborations permettraient d'atteindre des tailles de population suffisantes pour réaliser des études de séquençage d'exome ou de génome entier (WES ou WGS) avec une puissance statistique adéquate. Par exemple, la cohorte *PanCareSurFup*, une large cohorte européenne fournissant désormais des preuves solides sur l'épidémiologie des effets tardifs des cancers pédiatriques et de leurs traitements, devrait constituer des bio-banques pour des études génétiques approfondies. De plus, les collaborations internationales permettraient la réplification des résultats d'association sur des cohortes indépendantes.

Avec le nombre croissant d'études génétiques à grande échelle, l'investigation des interactions entre les effets génétiques et les doses de radiothérapie ou de chimiothérapie devrait devenir un objectif important et atteignable. Cela permettrait

d'établir des directives plus précises et fiables pour la prise en charge des patients nouvellement diagnostiqués avec un cancer pédiatrique, ainsi que pour le suivi de ces derniers. De plus, les tests génétiques permettent une approche de suivi personnalisée pour les individus précédemment traités, en tenant compte de leurs prédispositions génétiques et de leurs traitements passés.

Cette revue systématique représente un outil important pour de futures recherches sur le risque de second cancer associé aux variants génétiques et aux gènes. En compilant de manière exhaustive tous les variants et gènes associés, à ce jour, associés au risque de cancers secondaires chez les survivants de cancers pédiatriques mais cela permet l'avancement de notre compréhension des bases génétiques et ce risque important pour les patients guéris d'un cancer pédiatrique.

CHAPITRE 6

-

Identification des variants rares associés au second cancer, une étude cas-témoin nichée dans la cohorte FCCSS

Ce chapitre est consacré au projet principal de cette thèse dont l'objectif est d'identifier des variants rares associés au risque de second cancer après traitement pour un cancer pédiatrique dans la cohorte FCCSS. Les résultats sont présentés pour l'ensemble des types de seconds cancers et pour les cancers secondaires les plus fréquents : ceux du sein et de la thyroïde.

Ce travail a été présenté aux congrès EPICLIN et ISORED en 2023 et sera soumis pour publication dans une revue internationale en juillet 2024

I. Contexte de l'étude

Grâce aux progrès thérapeutiques et à l'amélioration des méthodes de diagnostic, le taux de survie à 5 ans des enfants atteints de cancer a augmenté et atteint désormais 80 % dans la plupart des pays européens (Botta et al., 2022). Le nombre de patients ayant survécu à un cancer pédiatrique est estimé à un demi-million en 2020 en Europe (Vassal et al., 2016).

Cependant, deux tiers de ces survivants souffrent de complications ultérieures liées à leur cancer primaire et à son traitement, comme cela a été discuté dans le premier chapitre. Parmi ces complications, le risque de développer un deuxième cancer est trois à cinq fois supérieur à celui de la population générale (Mertens et al., 2008).

Plusieurs études ont mis en évidence une relation entre la survenue de seconds cancers et le traitement reçu pour le premier cancer, en particulier la radiothérapie et la chimiothérapie comme cela a été détaillé dans le chapitre 1. Une relation dose-réponse entre le risque de cancers secondaires et la dose de radiation reçue a été établie pour plusieurs types de seconds cancers tels que le cancer du sein, de la thyroïde ou encore des sarcomes (Inskip et al., 2016). La plupart des molécules de chimiothérapie, telles que les agents alkylants et les anthracyclines, ont également été associées au risque de plusieurs types de seconds cancers (Teepen et al., 2017; Turcotte et al., 2019).

Bien que les traitements anti-cancéreux soient un facteur de risque reconnu de cancers secondaires, il a été démontré qu'il existe une variabilité interindividuelle dans la réponse au traitement, suggérant une possible composante génétique (Armenian & Bhatia, 2009).

Les systèmes de réparation de l'ADN jouent un rôle essentiel dans la prévention du cancer en maintenant l'intégrité génétique, en détectant et en réparant les dommages de l'ADN causés par divers agents cytotoxiques. Les rayonnements ionisants peuvent provoquer de graves lésions de l'ADN qui, si elles ne sont pas réparées efficacement,

peuvent entraîner des mutations et des anomalies chromosomiques, augmentant le risque de survenue de cancers (Hoeijmakers, 2001). Les variations génétiques dans les gènes de réparation de l'ADN pourraient modifier le risque de second cancer en cas de traitement combiné avec la chimiothérapie ou la radiothérapie. Ceci représente un défi important dans la compréhension du risque de second cancer chez les survivants traités pour un cancer dès l'enfance.

Dans le chapitre précédent nous avons évalué plusieurs études d'association de type GWAS (Best et al., 2011; Morton et al., 2017; Opstal-van Winden et al., 2019) et de type gènes candidats (Applebaum et al., 2017; Blanco et al., 2002; Erčulj et al., 2012; Gramatges et al., 2019; Ma et al., 2012; Mertens et al., 2004; Richard et al., 2020; Stanulla et al., 2009; Vodusek et al., 2016a; X. Wang et al., 2017; Woo et al., 2000) qui ont permis d'identifier des variants communs principalement situés dans les gènes de la réparation de l'ADN ou du métabolisme des xénobiotiques associés au risque de développer un second cancer après traitement d'un cancer pédiatrique.

La susceptibilité génétique au risque de cancers secondaires peut également être attribuée à des variants rares à forte pénétrance (Eichler et al., 2010). Quatre études américaines récentes (Morton et al., 2020; Qin et al., 2020; Z. Wang, Liu, et al., 2018; Z. Wang, Wilson, et al., 2018) utilisant le séquençage du génome entier (WGS) ou le séquençage de l'exome entier (WES) se sont concentrées sur des voies, en particulier les voies de réparation de l'ADN et les gènes de prédisposition au cancer. Des analyses stratifiées sur les traitements reçus (Morton et al., 2017; Qin et al., 2020; Z. Wang, Liu, et al., 2018) ont suggéré un lien entre la génétique et le traitement reçu pour le premier cancer, en particulier la radiothérapie.

Toutefois, la littérature actuelle sur ce sujet reste limitée, avec très peu d'études d'association génétique portant sur le risque de cancers secondaires après un cancer de l'enfance, et encore moins d'études WES ou WGS permettant une exploration élargie à un plus grand nombre de gènes, en particulier des variants rares.

Nous avons mené une étude cas témoins nichée dans la cohorte FCCSS afin d'identifier des variants génétiques rares dans les gènes impliqués dans la réplication, la recombinaison et la réparation, associés au risque de seconds cancers. Nous avons également exploré l'impact du traitement reçu sur le risque associé à ces variants.

II. Méthodologie

1. Design de l'étude

Une étude cas témoins a été mise en place au sein de la cohorte FCCSS décrite dans le chapitre 2. Seuls les patients ayant eu un prélèvement de sang ont été considérés. 163 cas (patients ayant été traités pour un cancer pédiatrique et ayant ensuite développé un second cancer) ont d'abord été sélectionnés. Ces cas ont ensuite été appariés individuellement et aléatoirement à des témoins (patients ayant été traités pour un premier cancer mais n'ayant pas développé de second cancer) sur le sexe, l'âge au diagnostic du premier cancer (+/- 3 ans), l'année du diagnostic du premier cancer (+/- 3 ans), le type du premier cancer et la durée du suivi (le suivi de chaque témoin devait être au moins aussi long que celui du cas correspondant). Au total, 450 patients ont été inclus dans cette étude : 163 cas appariés à 287 témoins.

2. Données de séquençage et traitement bio-informatique des données

a. Données génétique

Le séquençage de l'exome des 450 sujets a été réalisé au CEA-CNRRGH à partir de l'ADN extrait des lignées lymphoblastiques établies au Généthon. Le contrôle qualité des ADN a été réalisé au CEPH-Biobanque (Fondation Jean Dausset-CEPH, Paris).

La préparation des bibliothèques a été réalisée en utilisant la technologie de capture « All Exon V6 » de SureSelect (Agilent Technologies, CA, USA). Ces bibliothèques ont ensuite été séquencées en utilisant le kit de séquençage HiSeq 4000 (Illumina, CA, USA) afin d'obtenir une couverture moyenne d'au moins 70X par échantillon (méthode

détaillée dans le chapitre 3). Dans cette étude, nous nous sommes intéressés aux variants rares dans les gènes impliqués dans la réplication, la recombinaison et la réparation de l'ADN (3R).

b. Sélection des gènes des voies 3R

Les gènes des 3R ont été sélectionnés à partir de 3 principales bases de données publiques internationales : la Gene Ontology (GO) (Ashburner et al., 2000; Gene Ontology Consortium et al., 2023), la Molecular Signature Database (MSigDb) (Liberzon et al., 2011) et la Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000). L'intersection des trois bases a conduit à une liste de 1644 gènes. La Figure 22 présente le nombre de gènes identifiés dans chacune des bases. Un nombre plus faible de gènes a été identifié dans la base KEGG car elle ne contient que des gènes annotés de manière expérimentale. Cela explique également que tous les gènes des 3R identifiés dans KEGG sont également retrouvés dans les deux autres bases (Figure 22).

En collaboration avec un biologiste spécialisé dans ces voies (Filippo Rosselli), un score a été attribué chaque fois qu'un gène était identifié dans l'une des trois voies par l'une des trois bases de données. Ce score varie de 1 à 9 ; le score minimal de 1 indique que le gène a été identifié par une seule base de données et est associé à une seule voie tandis que le score maximal de 9 indique qu'un gène apparaît dans toutes les bases et est associé aux trois voies biologiques. Pour notre étude, les gènes ayant obtenu un score de 2 ou plus ont été retenus pour les analyses, soit 969 gènes sur les 1,644 gènes initialement identifiés (Figure 21).

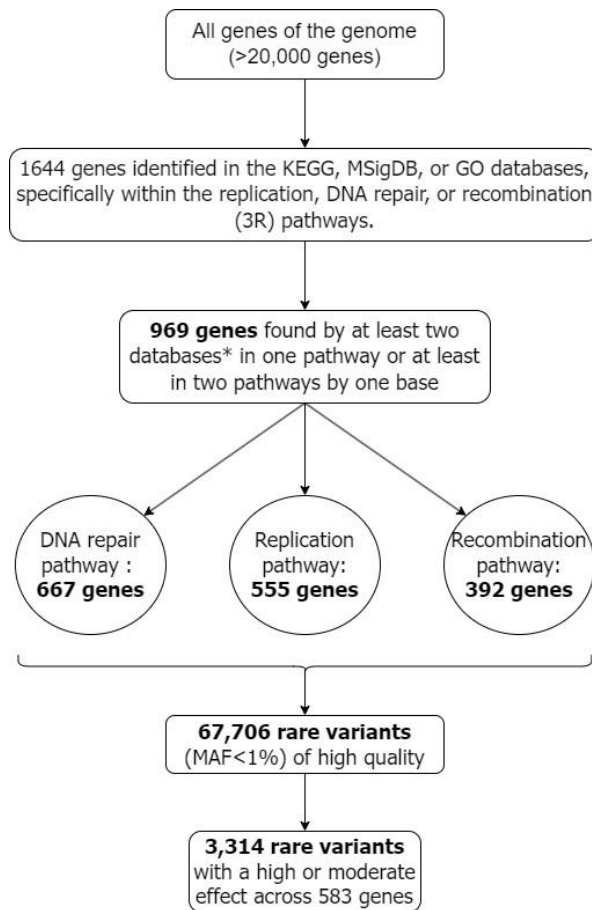


Figure 21 - Sélection des gènes et des variants inclus dans les analyses statistiques.

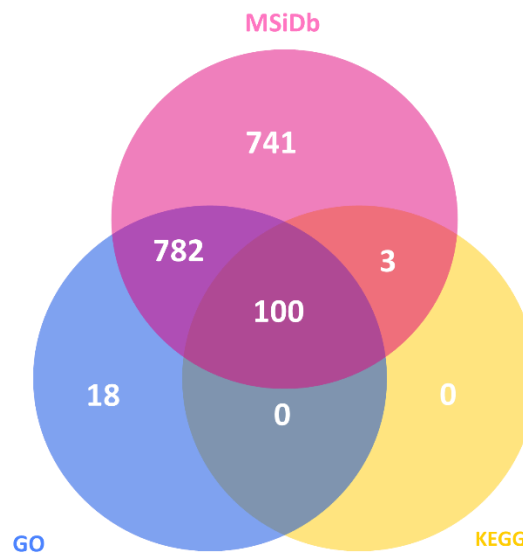


Figure 22 - Résultats de la recherche de gènes dans les trois bases de données - Intersection des résultats retrouvés par les trois bases. GO, KEGG et MSigDb
 Abréviations : GO, Gene Ontology ; KEGG, Kyoto Encyclopedia of Genes and Genomes ; MSigDb, Molecular Signature Database

c. Pipeline d'analyses

Afin d'analyser les données brutes de séquençage, différents pipeline d'analyses successifs ont été développés à l'aide du gestionnaire de workflows Snakemake (Mölder et al., 2021) et du gestionnaire de paquet et d'environnement Conda. Cette approche assure une bonne reproductibilité des données obtenues, en respectant les principes FAIR (Findability, Accessibility, Interoperability, and Reusability).

Ces pipelines comprenant plusieurs étapes bio-informatiques et les outils utilisés pour chacune sont résumés dans la Figure 23 et décrites en détails dans le chapitre 3. Elles sont brièvement expliquées ci-dessous. L'objectif de ces dernières est de transformer les données brutes de séquençage stockées dans des fichiers au format FASTQ, réparties en 450 fichiers (un par individu), en données directement utilisables pour les analyses d'association génétique. Ces analyses ne porteront que sur des variants de bonne qualité, chacun étant annotés avec toutes les informations nécessaires pour pouvoir appliquer les filtres essentiels pour la suite de l'étude.

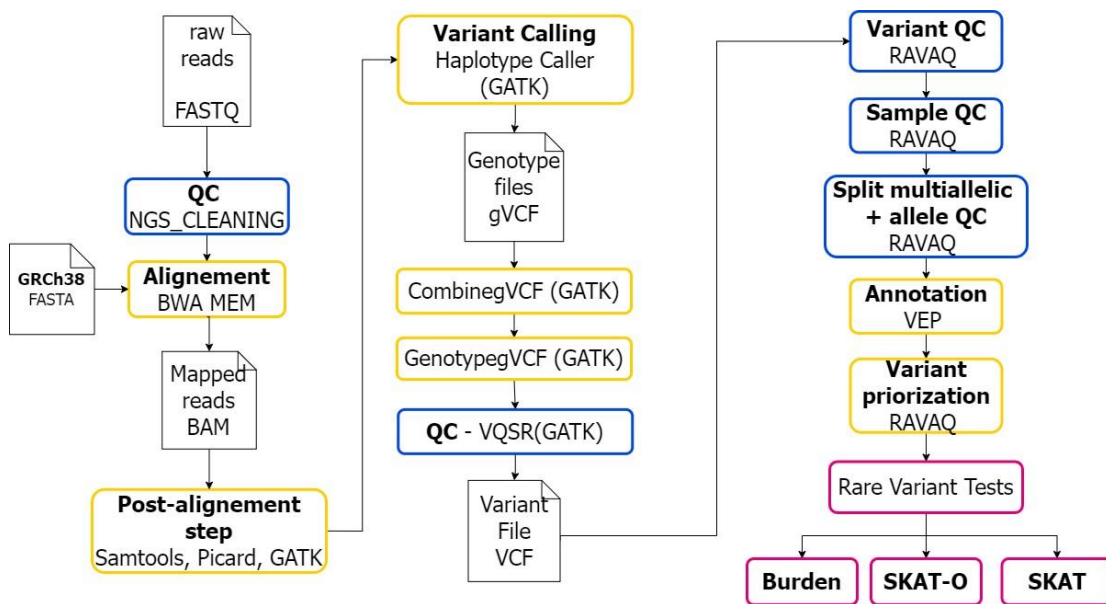


Figure 23- Succession des outils utilisés lors des différentes étapes lors de l'analyse des données exomiques.

Abréviations : QC, Contrôle qualité.

Contrôle qualité des données brutes de séquençage

Le contrôle qualité des données brutes de séquençage a été réalisé à l'aide des outils FastQC et FastqScreen (Wingett & Andrews, 2018). Ces outils ont permis de vérifier la qualité des *reads*, de détecter de potentielles contaminations et de supprimer les adaptateurs de séquençage. Comme illustré sur la Figure 24, les *reads* contenus dans les fichiers FASTQ ne présentent plus d'adaptateurs après leur traitement par le pipeline.

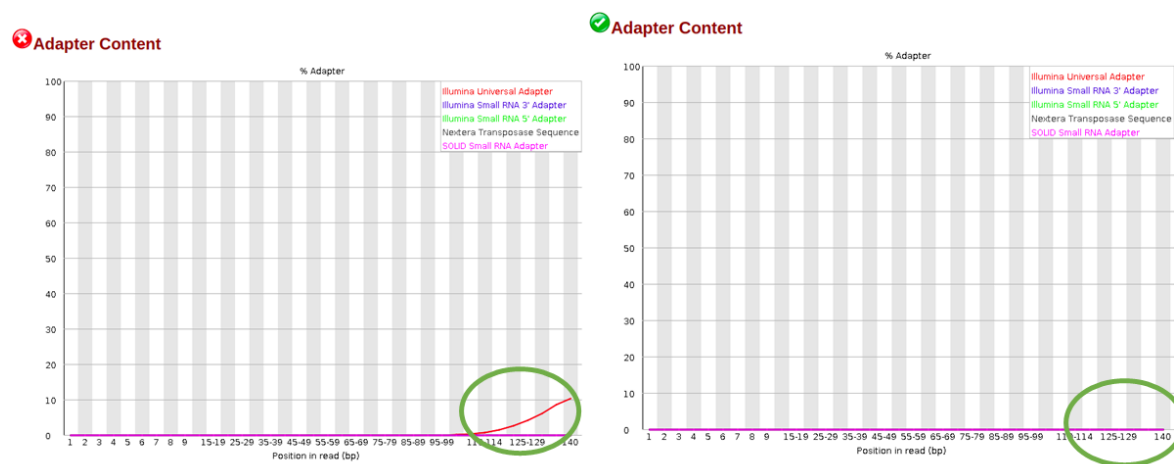


Figure 24 - Contrôle qualité des reads - Suppression des adaptateurs. Ce schéma illustre la suppression des adaptateurs présents à l'extrémité de chaque reads, comparant leur présence avant (à gauche) et après (à droite) l'application du pipeline.

Alignement

Dans notre étude, nous avons utilisé la version GRCH38 pour l'alignement des *reads* sur le génome de référence et l'algorithme BWA-MEM (H. Li & Durbin, 2009) qui est couramment employé pour ce type de données.

Variant Calling

Pour le variant calling, nous avons utilisé l'outil *HaplotypeCaller* de GATK permettant de générer un fichier gVCF pour chaque individu inclus dans notre étude. Ce fichier contient les mêmes informations qu'un fichier VCF, mais a la particularité d'avoir une ligne pour chaque position du génome.

Contrôle qualité

Pour assurer la qualité des données génomiques et réduire les biais potentiels ainsi que la détection de faux positifs, plusieurs filtres de contrôle qualité ont été appliqués à nos données.

Concernant les variants, seuls ceux soutenus par une profondeur d'au moins 10 *reads*, avec une qualité minimale de 20, satisfaisant les critères de qualité définis par le filtre VQSR et avec le taux d'appel supérieur à 95 % ont été considérés.

Concernant les individus de l'étude, la stratification de la population a été analysée à partir de variants communs informatifs et des données génétiques provenant du projet 1000 Genomes (Auton et al., 2015). Les résultats de cette analyse, présentés à la Figure 25, montrent qu'une proportion significative des participants, cas et témoins, est d'ascendance africaine. Afin d'éviter de potentiels biais dus à cette mixité de population, nous avons décidé de restreindre notre population d'étude aux individus d'ascendance européenne, en accord avec la littérature des différentes études d'associations génétiques comme décrites dans le chapitre 4.

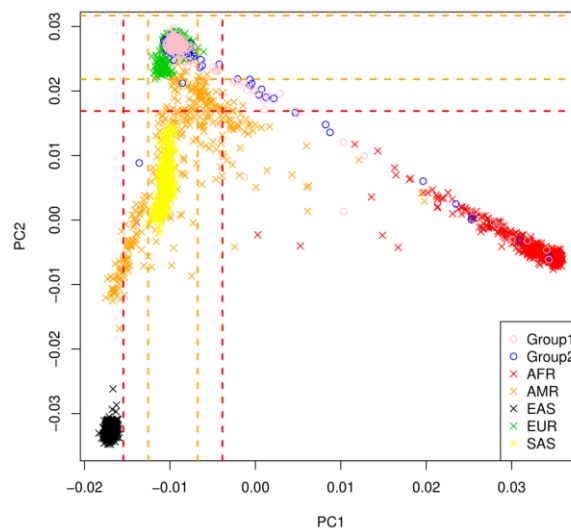


Figure 25 - Stratification de la population

Analyse en Composantes Principales (ACP) réalisée à partir de nos données (cercles) et des données de 1000G (croix) pour déterminer l'ascendance de nos individus révèle deux groupes bien distincts. La majorité de la population est d'ascendance européenne (représentée par les cercles roses et bleus sur les croix vertes), tandis qu'une proportion non négligeable d'individus d'ascendance africaine se situe en dehors des limites délimitées par les pointillés rouges

Par la suite seuls les individus ayant un taux d'appel (pourcentage des données génotypique valides) inférieur à 95 %, identifiés comme apparentés, présentant des preuves de consanguinités ou dont l'échantillon est contaminé, sont exclus de la suite des analyses ultérieures.

Annotation et identification des variants

Pour annoter les variants, nous avons utilisé l'outil VEP (Variant Effect Predictor), qui enrichit les variants avec diverses informations en s'appuyant sur plusieurs bases de données.

Pour notre étude, nous avons choisi de ne sélectionner que les gènes ayant un impact prédit "HIGH" ou "MODERATE" avec une fréquence allélique mineure (MAF) inférieure à 1 %.

3. Analyses statistiques

Nous avons limité les analyses statistiques aux individus d'ascendance européenne en utilisant une analyse en composantes principales (PCA) où les variants communs ont été comparés aux données du projet 1000 Genomes (1000G) (Auton et al., 2015) pour prendre en compte la stratification potentielle de la population.

Ainsi, nous avons inclus 395 individus d'origine européenne dont 141 cas et 254 témoins qui ont passé avec succès les critères de contrôle qualité décrits ci-dessus.

Nous avons également mené des analyses séparées pour les seconds cancers du sein et de la thyroïde, les plus fréquents dans notre étude, avec respectivement 48 et 69 cas. Nous avons également réalisé une analyse de sensibilité incluant uniquement 381 patients sans aucun syndrome de prédisposition génétique connu. Pour les sous-analyses concernant le cancer du sein, seuls les patients de sexe féminin ont été inclus.

Tous les variants rares ayant passé les différents contrôles de qualité ont été regroupés dans les gènes de la liste préétablie des 3R. Trois types de tests d'association adaptés

à l'analyse des variants rares tels que décrits dans le chapitre 3, ont été effectués : un test SKAT (Wu et al., 2011) , un test de fardeau classique (B. Li & Leal, 2008) et un test SKAT-O (Lee, Emond, et al., 2012). Seuls les gènes dont la *p-value* était inférieure à 0,05 pour au moins deux des trois méthodes ont été retenus pour la suite des analyses. Cette approche, qui prend en compte les intersections de plusieurs tests, permettrait de réduire les faux positifs (Martínez-Bueno & Alarcón-Riquelme, 2019).

En raison de l'exclusion de patients lors du contrôle de qualité, il n'a pas été possible de maintenir l'appariement initial. Par conséquent, les analyses statistiques ont été ajustées pour tenir compte des variables d'appariement.

Le risque de second cancer associé aux gènes des voies 3R a été évalué à l'aide d'une analyse de régression logistique. Le modèle statistique a été ajusté sur les variables suivantes : le sexe, l'âge et l'année au moment du diagnostic du premier cancer, le type du premier cancer, la durée du suivi, et les trois premières composantes principales de l'ACP pour tenir compte de la stratification génétique.

Nous avons effectué des analyses similaires pour les cancers secondaires du sein et de la thyroïde. Pour chacune de ces trois analyses, nous avons également étudié l'impact du traitement sur la modification de ce risque. Pour cela, nous avons défini des catégories de doses de radiation en fonction de la dose moyenne reçue par chaque organe d'intérêt pour tous les patients inclus dans chacune des analyses : la dose à la moelle osseuse pour tous type de seconds cancers ; la dose reçue au sein pour le cancer du sein et la dose reçue à la thyroïde pour le cancer de la thyroïde.

III. Résultats

1. Caractéristiques cliniques

Les caractéristiques des 141 cas de seconds cancers et des 254 témoins sont présentées dans le Tableau 15. L'âge médian au moment du diagnostic du premier cancer était de 8 ans (IQR : 4-12), et la durée médiane du suivi était de 25 ans (IQR : 17-34) pour les cas comme pour les témoins.

Les types de cancers pédiatriques les plus fréquents dans notre étude sont les lymphomes (29 %), les néphroblastomes (17 %), les tumeurs osseuses (15 %) et les sarcomes des tissus mous (STS) (12 %). Parmi les 141 cas, les seconds cancers les plus fréquents sont les cancers de la thyroïde (41 %), du sein (31 %), les ostéosarcomes (11 %), les cancers du rein et la vessie (10 %), et le cancer colorectal (7 %). Un syndrome de prédisposition génétique était présent chez 14 patients, dont dix cas (Annexe B).

2. Variant identifiés par le séquençage d'exome

Un total de 1 215 288 variants dans les 969 gènes sélectionnés ont été identifiés, chacun étant présent chez au moins un des 395 patients. Après exclusion des variants communs (définis comme ceux ayant une MAF > 1 %), 67 706 variants rares ont été identifiés. En se limitant aux gènes des voies 3R, 3 314 variants ayant un effet élevé ou modéré dans 583 gènes ont été identifiés (Figure 21).

Tableau 15 - Caractéristiques des patients inclus.

Caractéristiques	Cas n (%) N = 141	Témoins n (%) N = 254	Tous les patients n (%) N = 395
Age au premier cancer, années			
<5	35 (24,8)	62 (24,4)	97 (24,6)
[5-9]	44 (31,2)	76 (29,9)	120 (30,4)
[10-14]	40 (28,4)	77 (30,3)	117 (29,6)
≥15	22 (15,6)	39 (15,4)	61 (15,4)
Sexe			
Homme	46 (32,6)	88 (34,6)	134 (33,9)
Femme	95 (67,4)	166 (65,4)	261 (66,1)
Cancer pédiatrique			
Lymphome	41 (29,1)	75 (29,5)	116 (29,4)
Rein	22 (15,6)	44 (17,3)	66 (16,7)
Ostéosarcome	22 (15,6)	37 (14,6)	59 (14,9)
Sarcome des tissus mous	16 (11,3)	32 (12,6)	48 (12,2)
SNC*	13 (9,2)	25 (9,8)	38 (9,6)
Autres	27 (19,2)	41 (16,2)	68 (17,2)
Année de diagnostic du cancer pédiatrique			
<1970	28 (19,9)	52 (20,5)	80 (20,3)
[1970-1979]	45 (31,9)	77 (30,3)	122 (30,9)
[1980-1989]	42 (29,8)	86 (33,9)	128 (32,4)
≥1990	26 (18,4)	39 (15,4)	65 (16,5)
Temps écoulé entre le diagnostic du premier cancer et le second/dernier suivi, en années			
<15	21 (14,9)	46 (18,1)	67 (17)
[15-24]	48 (34)	81 (31,9)	129 (32,7)
[25-34]	39 (27,7)	67 (26,4)	106 (26,8)
≥35	33 (23,4)	60 (23,6)	93 (23,5)
Second cancer			
Thyroïde	58 (41,1)	-	58 (41,1)
Sein	44 (31,2)	-	44 (31,2)
Os	15 (10,6)	-	15 (10,6)
Rein et vessie	14 (9,9)	-	14 (9,9)
Colorectal	10 (7,1)	-	10 (7,1)
Radiothérapie			
Oui	130 (92,9)	175 (68,9)	305 (77,2)
Non	11 (7,8)	77 (30,3)	88 (22,3)
Dose moyenne de radiation, Gy (IC*)			
Moelle osseuse	5 (4,4-5,6)	3,9 (3,4-4,3)	4,3 (3,9-4,7)
Sein	12,4 (9,4-15,4)	4,7 (3,3-6,0)	8,11 (6,5-9,8)
Thyroïde	8 (6,3-97,0)	6,5 (5,0-7,9)	7,1 (6-8,2)
Chimiothérapie			
Oui	129 (91,5)	208 (81,9)	337 (85,3)
Non	12 (8,5)	44 (17,3)	56 (14,2)
Anthracycline			
Oui	52 (37,9)	108 (42,5)	160 (40,5)
Non	89 (63,1)	146 (57,5)	235 (59,5)
Médiane, mg/m ² (IQR*)	285,3 (160,5-403,6)	294,6 (199,2-379,1)	293,3 (179,8-382)
Agents alkylant			
Oui	79 (56,0)	143 (56,0)	222 (56,0)
Non	62 (44,0)	111 (44,0)	173 (44,0)
Médiane, mg/m ² (IQR*)	6771,8 (401,5-13240,4)	7524,3 (3595-13358,6)	8304,6 (5033,3-13772,6)

* Abréviations : IC, Intervalle de confiance ; IQR, Intervalle Interquartile ; SNC, Système Nerveux Central

Les résultats de l'identification des gènes associés aux cancers secondaires par les trois tests — SKAT, le test de fardeau et le test combiné SKAT-O — sont présentés dans la Figure 26. Huit gènes ont été identifiés comme étant associés à un cancer secondaire par ces trois approches, tandis que 12 gènes supplémentaires ont été identifiés par au moins deux de ces méthodes (Figure 26A). Ces 20 gènes ont été sélectionnés pour des analyses de régression. En ce qui concerne les cancers secondaires du sein et de la thyroïde, dix-huit et vingt-deux gènes, respectivement, ont été considérés pour ces analyses (Figure 26 B et C).

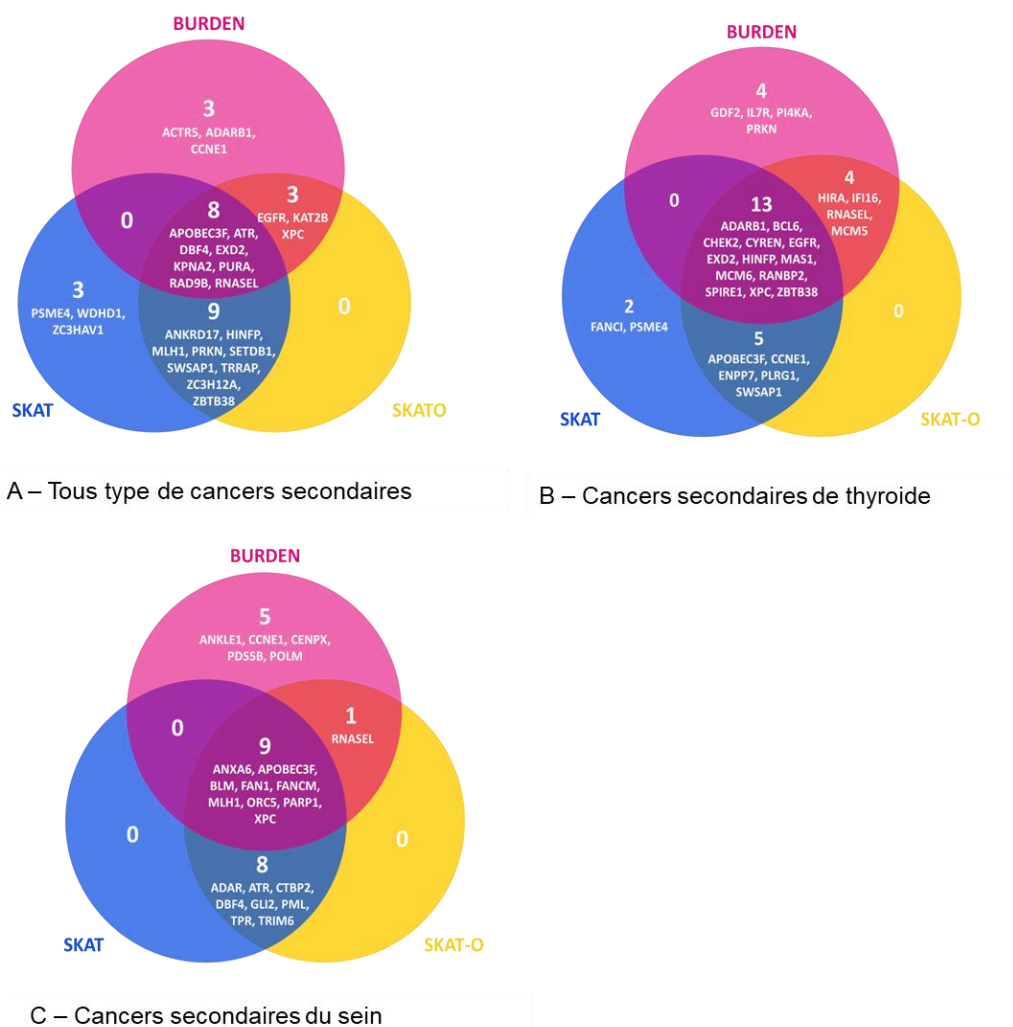


Figure 26 - Résultats des tests spécifiques aux variants rares - SKAT, test de Fardeau et SKAT-O.

Au total 20 gènes sont inclus dans la suites des analyses pour l'ensemble des seconds cancers, 22 pour les seconds cancers de la thyroïde et 18 pour les seconds cancers du sein.

3. Risque associé à un cancer secondaire (tous type de seconds cancers)

Nous avons identifié sept gènes associés à un risque accru de cancers secondaires (Tableau 16) : deux gènes impliqués dans la réplication, *APOBEC3F* (OR=4,8 ; IC 95 % : 1,28-22,95 ; p=0,027) et *DBF4* (OR=4,10 ; IC 95 % : 1,31-14,04 ; p=0,016), un gène impliqué dans les trois voies, *ATR* (OR=3,13, IC 95 % : 1,17-8,95 ; p=0,026), un gène dans la réparation de l'ADN et la réplication, *EGFR* (OR=10,67 ; IC 95 % : 1,6-209,6 ; p=0,034) et trois gènes dans la réparation de l'ADN, *RNASEL* (OR=5,43, IC à 95 % : 1,76-20,44, p=0,0055), *HINFP* (OR=3,41 ; IC 95 % : 1,06-12,03 ; p=0,043), et *XPC* (OR=4,82 ; IC 95 % : 1,49-18,57 ; p=0,012). Le nombre de variants identifiés dans chaque gène et le nombre de cas et de témoins porteurs d'au moins un variant sont présentés dans le Tableau 16.

Tableau 16 - Résultats des analyses pour tous types de seconds cancers après un cancer pédiatrique.

Gènes	Tests spécifique aux variants rares			Analyses multivariées			n variants** (cas / témoins)
	SKATO	SKAT	BURDEN	Pvalue	OR	IC 95 %	
ANKRD17*	0,05	0,039	NA	0,429	1,94	(0,35-10,93)	4 (3/3)
APOBEC3F*	0,0084	0,0084	0,035	0,027	4,8	(1,28-22,95)	2 (7/3)
ATR*	0,021	0,021	0,016	0,026	3,13	(1,17-8,95)	12 (11/7)
DBF4*	0,021	0,032	0,01	0,016	4,1	(1,35-14,04)	5 (10/5)
EGFR*	0,018	NA	0,0073	0,034	10,67	(1,6-209,6)	4 (5/1)
EXD2	0,0028	0,0028	0,0028	0,980	INF	NA	3 (5/0)
HINFP*	0,015	0,0093	NA	0,043	3,41	(1,06-12,03)	5 (8/5)
KAT2B	0,033	NA	0,021	0,983	0	NA	6 (0/9)
KPNA2	0,022	0,022	0,022	0,98	INF	NA	2 (3/0)
MLH1	0,039	0,018	NA	0,231	1,88	(0,66-5,4)	8 (4/7)
PRKN	0,035	0,032	NA	0,098	4,23	(0,85-30,92)	4 (5/2)
PURA	0,022	0,022	0,022	0,980	INF	NA	2 (2/0)
RAD9B	0,022	0,022	0,022	0,976	INF	NA	3 (2/0)
RNASEL*	0,0012	0,0064	0,00063	0,0055	5,43	(1,76-20,44)	10 (8/4)
SETDB1	0,044	0,036	NA	0,121	4	(0,74-30,01)	5 (4/2)
SWSAP1	0,045	0,045	NA	0,122	6,27	(0,75-131,56)	3 (3/1)
TRRAP	0,05	0,037	NA	0,169	2,38	(0,68-8,63)	10 (6/5)
XPC*	0,021	NA	0,0057	0,012	4,82	(1,49-18,57)	11 (9/4)
ZBTB38	0,045	0,045	NA	0,17	5,26	(0,6-112,44)	3 (1/3)
ZC3H12A	0,007	0,0064	NA	0,063	4,11	(0,97-21,21)	6 (6/3)

Abréviations : IC, Intervalle de confiance ; OR, Odds Ratio.

* p-value significative < 0,05

** n = nombre de variants rares présents au sein du gène étudié. Par exemple, on retrouve chez les 395 individus inclus dans l'étude 12 variants rares au sein du gène *ATR* : 11 cas sont porteurs d'au moins un variant rare et 7 témoins le sont également.

4. Second cancer de la thyroïde

Le gène *HINFP* est également associé au risque de second cancer de la thyroïde chez les patients traités pour un cancer pédiatrique (OR=5,43 ; IC à 95 % : 1,26-23,5 ; p=0,022). Une augmentation significative du risque a également été observée pour quatre autres gènes : *HIRA* (OR=8,46 ; IC à 95 % : 1,06-89,81 ; p=0,047) et *IFI16* (OR=3,49 ; IC à 95 % : 1,28-9,32 ; p=0,013) impliqués dans la réplication. *MCM5* (OR=5,7 ; IC 95 % : 1,36-23,87 ; p=0,015) et *RANBP2* (OR=8,68 ; IC 95 % : 1,98-40,2 ; p=0,004) impliqués dans la réplication et la réparation de l'ADN.

Les résultats sont présentés dans le Tableau 17.

Tableau 17 – Résultats des analyses pour l'étude du risque de cancer secondaire de la thyroïde.

Gènes	Tests spécifique aux variants rares			Analyses multivariées			n variants** (cas / témoins)
	SKATO	SKAT	BURDEN	OR	IC 95 %	Pvalue	
BCL6*	0,049	0,049	0,049	45,33	(3,07-1479,93)	0,0086	4 (2/1)
EGFR*	0,0035	0,0079	0,0016	32,53	(3,62-806,66)	0,0066	4 (4/1)
HINFP*	0,00068	0,00035	0,022	5,34	(1,26-23,5)	0,022	5 (5/5)
MAS1*	0,049	0,049	0,049	23,98	(1,72-638,95)	0,022	3 (2/1)
RANBP2*	0,0063	0,0082	0,0066	8,68	(1,98-40,2)	0,004	8 (3/5)
ZBTB38*	0,0064	0,0064	0,017	132,13	(9,62-4654,71)	0,00088	3 (3/1)
ADARB1	0,045	0,028	0,038	2,93	(0,76-10,98)	0,11	6 (5/7)
APOBEC3F	0,030	0,030	NA	3,58	(0,61-21,12)	0,14	2 (3/3)
CCNE1	0,030	0,030	NA	3,04	(0,48-20,41)	0,23	4 (3/3)
CHEK2	0,049	0,049	0,049	7,75	(0,63-185,96)	0,12	3 (2/1)
ENPP7	0,043	0,026	NA	1,84	(0,45-6,45)	0,36	9 (4/10)
EXD2	0,018	0,018	0,018	INF	NA	0,98	3 (2/0)
MCM6	0,049	0,049	0,030	2,45	(0,56-9,68)	0,21	7 (3/7)
PLRG1	0,036	0,030	NA	5,04	(0,78-33,82)	0,081	4 (3/3)
SPIRE1	0,049	0,049	0,049	6,8	(0,58-157,55)	0,14	3 (2/1)
SWSAP1	0,021	0,021	NA	6,25	(0,53-145,13)	0,16	3 (2/1)
XPC	0,033	0,033	0,011	4	(0,83-19,81)	0,078	11 (4/4)

Abréviations : IC, Intervalle de confiance ; OR, Odds Ratio.

* p-value significative < 0,05

** n = nombre de variants rares présents au sein du gène étudié. Par exemple, on retrouve chez les 395 individus inclus dans l'étude, 8 variants rares au sein du gène *RANBP2* : 3 cas sont porteurs d'au moins un variant rare et 5 témoins le sont également.

5. Second cancer du sein

Les résultats présentés dans le Tableau 18.

Huit gènes sont significativement associés au risque de second cancer du sein. Parmi eux, *BLM* (OR=5,21 ; IC 95 % : 1,5-19,11 ; p=0,0096) et *FANCM* (OR=4,65 ; IC 95 % : 1,06-21,56 ; p=0,041) sont associés aux trois voies d'intérêt.

D'autres gènes sont également associés à ce risque : *FAN1* (OR=4,27 ; IC 95 % : 1,26-14,26 ; p=0,018) et *PARP1* (OR=7,79 ; IC 95 % : 1,15-22,04 ; p=0,034) impliqués dans la réparation de l'ADN et la recombinaison, *CTBP2* (OR=9,13 ; IC 95 % : 1,73-51,47 ; p=0,009), *ORC5* (OR=18,41 ; IC 95 % : 1,9-418,21 ; p=0,02) et *ANXA6* (OR=16,63 ; IC à 95 % : 2,35-177,44 ; p=0,008) impliqués dans la réplication et *PML* (OR=5,02 ; IC 95 % : 1,01-27,82 ; p=0,048), impliqué dans la réparation de l'ADN.

Tableau 18 - Résultats des analyses pour l'étude du risque de cancer secondaire du sein.

Gene	Tests spécifique aux variants rares			Analyses multivariées			n variants (n cas / n témoins)
	SKATO	SKAT	BURDEN	OR	IC 95 %	pvalue	
ANXA6*	0,010	0,010	0,010	16,63	(2,35-177,44)	0,008	6(4/2)
APOBEC3F*	0,016	0,016	0,016	22,62	(2,06-781,04)	0,026	1 (3/1)
BLM*	0,0093	0,016	0,0050	5,21	(1,5-19,11)	0,0096	8 (7/6)
CTBP2*	0,033	0,021	NA	9,13	(1,73-51,47)	0,0089	6 (4/4)
FAN1*	0,017	0,017	0,016	4,27	(1,26-14,26)	0,018	7 (7/9)
FANCM*	0,022	0,034	0,015	4,65	(1,06-21,56)	0,041	9 (5/6)
ORC5*	0,016	0,016	0,016	18,41	(1,9-418,21)	0,020	2 (3/1)
PARP1*	0,0077	0,013	0,0041	4,79	(1,15-22,04)	0,034	5 (6/4)
PML*	0,017	0,0097	NA	4,79	(1,15-22,04)	0,034	5 (6/4)
ADAR	0,026	0,026	NA	7,69	(0,57-202,97)	0,14	2 (2/1)
ATR	0,028	0,018	NA	1,82	(0,38-8,04)	0,43	7 (4/5)
DBF4	0,027	0,021	NA	4,75	(0,66-39,25)	0,12	3 (3/3)
GLI2	0,031	0,025	NA	3,56	(0,60-19,82)	0,141	7 (3/5)
MLH1	0,021	0,012	0,049	3,34	(0,72-14,84)	0,11	7 (4/5)
RNASEL	0,018	0,0076	0,011	4,79	(1,15-22,04)	0,034	5 (6/4)
TPR	0,012	NA	NA	2,09	(0,60-6,90)	0,23	8 (6/10)
TRIM6	0,044	0,024	NA	1,44	(0,24-762)	0,67	3 (3/4)
XPC	0,014	0,049	0,0073	6,67	(1,02-58,53)	0,055	5 (4/2)

Abréviations : IC, Intervalle de confiance ; OR, Odds Ratio.

* p-value significative < 0,05

** n = nombre de variants rares présents au sein du gène étudié. Par exemple, on retrouve chez les 395 individus inclus dans l'étude, 9 variants rares au sein du gène *FANCM* : 5 cas sont porteurs d'au moins un variant rare et 6 témoins le sont également.

6. Analyses stratifiées selon la dose de radiation reçue

Les résultats des analyses stratifiées selon la dose de radiation sont présentés dans le

Gène	Type de SC	Groupe de patients ayant reçu de faibles doses de RT			Groupe de patients ayant reçus de fortes doses de RT		
		OR	IC 95 %	P-value	OR	IC 95%	P-value
ADAR	Sein	10,93	(0,31-440,87)	0,14	∞	NA	0,9976
ADARB1	Thyroïde	4,36	(0,45-36,97)	0,172	∞	NA	0,996
ANXA6	Sein	9,02	(0,47-308,11)	0,155	∞	NA	0,994
ANKRD17	Tous type	1,84	(0,05-68,61)	0,720	0,75	(0,07-6,69)	0,80
APOBEC3F	Tous type	3,29	(0,45-29,65)	0,240	3,16	(0,37-68,45)	0,34
	Sein	0	NA	0,993	∞	NA	0,997
ATR	Thyroïde	2,56	(0,08-87,23)	0,560	4,47	(0,35-60,66)	0,230
	Tous type	3,8	(0,6-29,65)	0,142	5,14	(1,1-37,8)	0,058
BCL6	Sein	0	NA	0,997	∞	NA	0,996
	Thyroïde	∞	NA	0,996	∞	NA	0,99
BLM	Sein*	6,39	(1,46-29,07)	0,013	∞	NA	0,998
CCNE1	Thyroïde	5,14	(0,12-218,56)	0,349	∞	NA	0,997
CHECK2	Thyroïde	∞	NA	0,998	0,51	(0,01-17,17)	0,679
CTBP2	Sein*	32,23	(3,11-470,09)	0,0053	1,47	(0,01-458,56)	0,8954
CYREN	Thyroïde	16,43	(0,44-681,34)	0,103	∞	NA	0,998
DBF4	Tous type	4,61	(1-26,11)	0,059	3,4	(0,037-85,24)	0,34
	Sein	3,55	(0,28-39,66)	0,295	∞	NA	0,996
EGFR	Tous type	2,58	(0,09-72,59)	0,525	∞	NA	0,99
	Thyroïde	9,87	(0,27-394,87)	0,177	∞	NA	0,993
ENPP7	Thyroïde	2,15	(0,25-13,48)	0,432	2,65	(0,26-23,32)	0,373
EXD2	Tous type	∞	NA	0,995	∞	NA	0,98
	Thyroïde	∞	NA	0,996	∞	NA	NA
FAN1	Sein*	7,41	(1,49-40,75)	0,016	0,42	(0-37,17)	0,7279
FANCM	Sein*	7,38	(0,64-95,66)	0,100	39,73	(1,3-3449,35)	0,054
GLI2	Sein	2,51	(0,1-35,72)	0,502	0,19	(0-27,86)	0,4944
HINFP	Tous type	4,76	(0,73-31,86)	0,094	2,51	(0,45-19,79)	0,32
	Thyroïde*	11,83	(1,47-131,78)	0,024	6,51	(0,48-118,61)	0,162
HIRA	Thyroïde	0	NA	0,997	∞	NA	0,996
IFI16	Thyroïde*	9,27	(2,7-33,77)	0,00047	0,25	(0,01-2,86)	0,315
KAT2B	Tous type	0	NA	0,993	∞	NA	0,99
KPNA2	Tous type	NA	NA	NA	∞	NA	0,988
MAS1	Thyroïde *	35,32	(0,97-1628,67)	0,037	∞	NA	1
MCM5	Thyroïde	8,91	(0,82-89,41)	0,057	15,29	(0,89-614,57)	0,083
MCM6	Thyroïde	1,09	(0,05-9,25)	0,945	2,01	(0,21-29,98)	0,557
MLH1	Tous type	2,31	(0,62-8,53)	0,202	0,5	(0,05-5,21)	0,55
	Sein**	1,14	(0,05-9,91)	0,0917	88,74	(1,22-13212,6)	0,0427
ORC5	Sein*	17,62	(1,39-434,99)	0,031	∞	NA	0,997
PARP1	Sein*	8,9	(1,23-72,62)	0,030	∞	NA	0,996
PLRG1	Thyroïde	0	NA	0,997	7,61	(0,63-198,43)	0,136
PML	Sein*	10,63	(1,38-108,86)	0,026	0,28	(0-14,27)	0,509
PRKN	Tous type	∞	NA	0,988	3,66	(0,31-89,66)	0,33
PURA	Tous type	∞	NA	0,996	∞	NA	0,99
RAD9B	Tous type	NA	NA	NA	∞	NA	0,99
RANBP2	Thyroïde *	32,33	(3,84-348,64)	0,0017	∞	NA	0,994
RNASEL	Tous type	3,68	(0,87-19,23)	0,089	∞	NA	0,99
	Sein	1,81	(0,06-30,24)	0,685	∞	NA	0,994
	Thyroïde	6,16	(0,9-43,13)	0,057	∞	NA	0,998
SETDB1	Tous type	2,77	(0,23-32,09)	0,395	∞	NA	0,987
SPIRE1	Thyroïde	3,52	(0,12-105,03)	0,412	∞	NA	0,998
SWSAP1	Tous type	8,35	(0,91-182,44)	0,083	∞	NA	
	Thyroïde	8,78	(0,66-240,97)	0,116	NA	NA	NA
TRRAP	Tous type	1,36	(0,25-6,45)	0,702	∞	NA	0,99
TPR	Sein	0,73	(0,03-5,56)	0,787	28,1	(0,8-2546,38)	0,0859
TRIM6	Sein	2,49	(0,08-43,59)	0,536	35,57	(0,39-5406,75)	0,120
XPC	Tous type*	9,26	(1,62-77,92)	0,019	4,44	(0,53-96,46)	0,22
	Sein	9,6	(0,28-3747,69)	0,166	1,01	(0-390,88)	0,9961
	Thyroïde	5,76	(0,6-50,72)	0,106	∞	NA	0,996
ZBTB28	Tous type	3,54	(0,27-85,98)	0,341	∞	NA	0,99
	Thyroïde *	141,2	(7,05-6423,3)	0,0024	∞	NA	0,997
ZC3H12A	Tous type	5,39	(0,39-140,2)	0,221	∞	NA	0,99

* Association significative pour le sous-groupe ayant reçu de faibles doses de radiation

* Association significative pour le sous-groupe ayant reçu de fortes doses de radiation

. Bien que chaque sous-groupe soit de taille réduite, certains résultats peuvent être soulignés. Concernant l'association entre le risque de cancers secondaire et le gène *XPC*, un risque accru a été observé chez les patients ayant reçu de faibles doses d'irradiation à la moelle osseuse (OR=9,26 ; IC 95 % : 1,62-77,92).

Dans l'analyse stratifiée du risque de second cancer du sein, chez les individus ayant reçu des doses de radiation au sein inférieures à 4,4 Gy, le risque était associé aux gènes suivants : *CTPBP2* (OR=32,23 ; IC 95 % : 3,11-470,09), *BLM* (OR=6,39 ; IC 95 % : 1,46-29,07), *FAN1* (OR=7,41 ; IC 95 % : 1,49-40,75), *PARP1* (OR=8,9 ; IC 95 % : 1,38-108,86) et *PML* (OR=10,63 ; IC 95 % : 1,38-108,86).

Enfin, trois gènes ont montré des associations potentielles avec un risque accru de second cancer de la thyroïde chez les individus exposés à de faibles doses de radiation thyroïdienne : *HINFP* (OR=11,83 ; IC 95 % : 1,47-131,78), *IFI16* (OR=9,27 ; IC 95 % : 2,7-33,77) et *RANBP2* (OR=32,33 ; IC 95 % : 3,84-348,64).

Des analyses stratifiées selon les doses reçues au niveau du thymus et de la rate ont également été réalisées, mais aucun résultat concernant les gènes préalablement étudiés n'a été obtenu.

Tableau 19 – Résultats des analyses stratifiées selon la dose reçue aux différents organes

Gène	Type de SC	Groupe de patients ayant reçu de faibles doses de RT			Groupe de patients ayant reçus de fortes de RT		
		OR	IC 95 %	P-value	OR	IC 95%	P-value
ADAR	Sein	10,93	(0,31-440,87)	0,14	∞	NA	0,9976
ADARB1	Thyroïde	4,36	(0,45-36,97)	0,172	∞	NA	0,996
ANXA6	Sein	9,02	(0,47-308,11)	0,155	∞	NA	0,994
ANKRD17	Tous type	1,84	(0,05-68,61)	0,720	0,75	(0,07-6,69)	0,80
APOBEC3F	Tous type	3,29	(0,45-29,65)	0,240	3,16	(0,37-68,45)	0,34
	Sein	0	NA	0,993	∞	NA	0,997
ATR	Thyroïde	2,56	(0,08-87,23)	0,560	4,47	(0,35-60,66)	0,230
	Tous type	3,8	(0,6-29,65)	0,142	5,14	(1,1-37,8)	0,058
BCL6	Sein	0	NA	0,997	∞	NA	0,996
	Thyroïde	∞	NA	0,996	∞	NA	0,99
BLM	Sein*	6,39	(1,46-29,07)	0,013	∞	NA	0,998
CCNE1	Thyroïde	5,14	(0,12-218,56)	0,349	∞	NA	0,997
CHECK2	Thyroïde	∞	NA	0,998	0,51	(0,01-17,17)	0,679
CTBP2	Sein*	32,23	(3,11-470,09)	0,0053	1,47	(0,01-458,56)	0,8954
CYREN	Thyroïde	16,43	(0,44-681,34)	0,103	∞	NA	0,998
DBF4	Tous type	4,61	(1-26,11)	0,059	3,4	(0,037-85,24)	0,34
	Sein	3,55	(0,28-39,66)	0,295	∞	NA	0,996
EGFR	Tous type	2,58	(0,09-72,59)	0,525	∞	NA	0,99
	Thyroïde	9,87	(0,27-394,87)	0,177	∞	NA	0,993
ENPP7	Thyroïde	2,15	(0,25-13,48)	0,432	2,65	(0,26-23,32)	0,373
EXD2	Tous type	∞	NA	0,995	∞	NA	0,98
	Thyroïde	∞	NA	0,996	∞	NA	NA
FAN1	Sein*	7,41	(1,49-40,75)	0,016	0,42	(0-37,17)	0,7279
FANCM	Sein*	7,38	(0,64-95,66)	0,100	39,73	(1,3-3449,35)	0,054
GLI2	Sein	2,51	(0,1-35,72)	0,502	0,19	(0-27,86)	0,4944
HINFP	Tous type	4,76	(0,73-31,86)	0,094	2,51	(0,45-19,79)	0,32
	Thyroïde*	11,83	(1,47-131,78)	0,024	6,51	(0,48-118,61)	0,162
HIRA	Thyroïde	0	NA	0,997	∞	NA	0,996
IFI16	Thyroïde*	9,27	(2,7-33,77)	0,00047	0,25	(0,01-2,86)	0,315
KAT2B	Tous type	0	NA	0,993	∞	NA	0,99
KPNA2	Tous type	NA	NA	NA	∞	NA	0,988
MAS1	Thyroïde *	35,32	(0,97-1628,67)	0,037	∞	NA	1
MCM5	Thyroïde	8,91	(0,82-89,41)	0,057	15,29	(0,89-614,57)	0,083
MCM6	Thyroïde	1,09	(0,05-9,25)	0,945	2,01	(0,21-29,98)	0,557
MLH1	Tous type	2,31	(0,62-8,53)	0,202	0,5	(0,05-5,21)	0,55
	Sein**	1,14	(0,05-9,91)	0,0917	88,74	(1,22-13212,6)	0,0427
ORC5	Sein*	17,62	(1,39-434,99)	0,031	∞	NA	0,997
PARP1	Sein*	8,9	(1,23-72,62)	0,030	∞	NA	0,996
PLRG1	Thyroïde	0	NA	0,997	7,61	(0,63-198,43)	0,136
PML	Sein*	10,63	(1,38-108,86)	0,026	0,28	(0-14,27)	0,509
PRKN	Tous type	∞	NA	0,988	3,66	(0,31-89,66)	0,33
PURA	Tous type	∞	NA	0,996	∞	NA	0,99
RAD9B	Tous type	NA	NA	NA	∞	NA	0,99
RANBP2	Thyroïde *	32,33	(3,84-348,64)	0,0017	∞	NA	0,994
RNASEL	Tous type	3,68	(0,87-19,23)	0,089	∞	NA	0,99
	Sein	1,81	(0,06-30,24)	0,685	∞	NA	0,994
	Thyroïde	6,16	(0,9-43,13)	0,057	∞	NA	0,998
SETDB1	Tous type	2,77	(0,23-32,09)	0,395	∞	NA	0,987
SPIRE1	Thyroïde	3,52	(0,12-105,03)	0,412	∞	NA	0,998
SWSAP1	Tous type	8,35	(0,91-182,44)	0,083	∞	NA	NA
	Thyroïde	8,78	(0,66-240,97)	0,116	NA	NA	NA
TRRAP	Tous type	1,36	(0,25-6,45)	0,702	∞	NA	0,99
TPR	Sein	0,73	(0,03-5,56)	0,787	28,1	(0,8-2546,38)	0,0859
TRIM6	Sein	2,49	(0,08-43,59)	0,536	35,57	(0,39-5406,75)	0,120
XPC	Tous type*	9,26	(1,62-77,92)	0,019	4,44	(0,53-96,46)	0,22
	Sein	9,6	(0,28-3747,69)	0,166	1,01	(0-390,88)	0,9961
	Thyroïde	5,76	(0,6-50,72)	0,106	∞	NA	0,996
ZBTB28	Tous type	3,54	(0,27-85,98)	0,341	∞	NA	0,99
	Thyroïde *	141,2	(7,05-6423,3)	0,0024	∞	NA	0,997
ZC3H12A	Tous type	5,39	(0,39-140,2)	0,221	∞	NA	0,99

* Association significative pour le sous-groupe ayant reçu de faibles doses de radiation

* Association significative pour le sous-groupe ayant reçu de fortes doses de radiation

IV. Discussion

Dans cette étude cas-témoins nichée dans la FCCSS, nous avons étudié l'impact de variants rares dans 969 gènes impliqués dans la réparation de l'ADN, la réplication et la recombinaison de l'ADN. Nous avons identifié 7 gènes significativement associés au risque de seconds cancers (notamment les gènes *APOBEC3F* et *RNASEL*), 8 gènes associés au risque de second cancer du sein (notamment le gène *FANCM*) et 6 gènes associés au risque de second cancer de la thyroïde (notamment les gènes *HINFP* et *IFI16*).

Le gène *RNASEL*, connu pour son rôle dans l'immunité et impliqué dans plusieurs types de cancer, est associé à un risque cinq fois plus élevé de second cancer dans notre étude. Ce résultat est pertinent compte tenu de son rôle établi dans la carcinogenèse et de son association avec divers types de cancers dans la population générale (Bartsch et al., 2005; Carpten et al., 2002; Datta et al., 2015; Rökman et al., 2002). Les variants rares identifiés, rs151296858 et rs74315364, tous deux présents exclusivement chez 3 cas, sont en déséquilibre de liaison complet (LD) (Rökman et al., 2002). Le variant rs74315364 correspond à un variant introduisant un codon stop ; il a été auparavant associé aux cancers de la prostate (Carpten et al., 2002; Datta et al., 2015) et du pancréas (Bartsch et al., 2005). Toutes ces études soulignent l'importance de ce gène et des variants rares au sein de ce dernier dans le développement de cancer et la nécessité de poursuivre les recherches pour confirmer leur implication dans le risque de seconds cancers chez les personnes ayant survécu à un cancer pédiatrique.

De même, le gène *APOBEC3F*, appartenant à la famille des cytidines désaminases ayant une activité antivirale (Ying et al., 2007), a été associé au risque de cancer, notamment du sein dans la population générale (Xuan et al., 2013). Dans notre étude, seuls deux variants ont été retrouvés chez sept cas et trois témoins. Le variant rs146924719 était présent chez tous les cas, soulignant son importance. Bien que nos résultats reposent sur de petits effectifs, ils suggèrent une association potentielle entre le gène *APOBEC3F*

et le risque de second cancer du sein. Ces résultats prometteurs n'ont pas été observés dans les études antérieures WES et WGS. Cela pourrait s'expliquer par le fait que notre étude est la seule à évaluer le risque de seconds cancers en utilisant des tests spécifiques pour des variants rares tels que SKAT, SKATO et Burden. De plus, les différences méthodologiques entre les études, notamment dans les techniques de séquençage utilisées, ou la sélection initiale des gènes d'intérêt, peuvent également contribuer à ces divergences.

Notre étude confirme également l'association entre le gène *FANCM* et le risque de second cancer du sein retrouvé dans l'étude de Morton et al. (Morton et al., 2020). Les auteurs ont montré que le gène *FANCM* était fortement associé au risque de seconds cancers liés à la radiothérapie, incluant le cancer du sein. Cette association suggère que la perte de *FANCM* entraîne une diminution de l'efficacité de la réparation par recombinaison homologue (HRR).

Ce gène, qui joue un rôle crucial dans la voie de l'Anémie de Fanconi en participant à la réparation des liaisons inter-brins de l'ADN, a déjà été associé au cancer du sein héréditaire (Neidhardt et al., 2017; Peterlongo et al., 2015; Schubert et al., 2019), au cancer du sein triple négatif (Figlioli et al., 2019; Kiiski et al., 2014) et au cancer de l'ovaire (Dicks et al., 2017; Lopes et al., 2019). Dans notre étude, nous avons également observé un risque accru de cancer du sein chez les patientes exposées à de fortes doses de radiations au niveau du sein, alors qu'aucune association n'a été trouvée chez celles ayant reçu de faibles doses.

La taille limitée de notre étude ne nous a pas permis d'explorer les interactions entre les gènes et la chimiothérapie. Les résultats d'études antérieurs (Catucci et al., 2018; Morton et al., 2020) montrent que les individus présentant des mutations homozygotes de perte de fonction dans ce gène ont une prédisposition accrue au cancer et une toxicité accrue de la chimiothérapie. Nous avons également identifié plusieurs gènes de prédisposition au cancer bien connus, notamment *BLM*, associé au syndrome de Bloom, et *PML*, impliqué dans de nombreux processus cellulaires tels que la

suppression des tumeurs, la réponse aux dommages de l'ADN, la régulation transcriptionnelle, la sénescence et l'apoptose (Vogiatzoglou et al., 2021).

Des associations intéressantes entre certains gènes et le risque de second cancer de la thyroïde ont également été observées. Le gène *HINFP* a été associé à un risque accru de seconds cancers, et également de second cancer de la thyroïde. Un variant rare au sein de ce gène, rs61753080, est présent chez cinq cas, dont quatre ont développé un second cancer de la thyroïde. Il a également été associé au risque de cancer agressif de la prostate (Zhuo & al., 2015).

Le gène *IFI16* a également été associé à un risque accru de second cancer de la thyroïde chez les patients ayant reçu de fortes doses de radiations, avec un OR passant de 3,5 à 9,27. Ce gène pourrait avoir un impact majeur en raison de ses interactions directes avec TP53, un gène de prédisposition au cancer bien connu (Kwak et al., 2003).

Limitations de l'étude

Bien que notre étude apporte de nouvelles informations sur les variants rares associés aux cancers secondaires chez les survivants d'un cancer pédiatrique, certaines limites doivent être considérées.

Premièrement, la taille de notre population peut être insuffisante pour détecter des associations significatives à l'échelle de l'exome. Par ailleurs, la nécessité de prendre en compte la stratification de la population a limité notre analyse aux individus d'origine européenne, limitant ainsi la généralisation de nos résultats et réduisant davantage la taille de la population étudiée.

Malgré cela, nos résultats suggèrent une relation potentielle entre la radiothérapie reçue et des variants génétiques spécifiques. Une validation externe serait bénéfique pour renforcer nos résultats, mais aucun échantillon externe n'était disponible pour cette analyse.

Notre méthodologie repose sur l'utilisation de trois tests adaptés aux variants rares

(SKAT, SKAT-O et Fardeau) et sur leur combinaison pour minimiser le taux de faux positifs. Néanmoins, cette approche combinée présente une limitation potentielle : certains gènes, dont l'association avec le risque de second cancer aurait pu être observée s'ils avaient été identifiés par un seul de ces tests, pourraient avoir été manqués. Dans les futures études, il serait intéressant d'examiner de manière plus approfondie les résultats spécifiques à chaque test, notamment lorsque des échantillons plus larges seront disponibles, afin d'accroître la puissance statistique permettant ainsi la possibilité d'identifier d'autres associations potentielles entre les variants génétiques et le risque de second cancer.

Perspectives et conclusion

Cette étude apporte de nouvelles connaissances sur le risque de seconds cancers, et sur les risques de seconds cancers du sein et de la thyroïde, chez les patients guéris d'un cancer pédiatrique, en identifiant des variants génétiques rares dans les gènes impliqués dans les 3R. Elle a également confirmé le rôle du gène *FANCM* dans la survenue du cancer du sein. Enfin, elle souligne l'importance de tenir compte à la fois du contexte génétique et des antécédents de traitement des survivants d'un cancer pédiatrique lors de l'évaluation de leur risque de développer un cancer secondaire.

Les futures recherches nécessitent de réaliser des études de validation dans des cohortes indépendantes afin de confirmer ces résultats, ainsi que l'exploration de l'impact biologique des variants génétiques identifiés par des études fonctionnelles. Des collaborations internationales devront être établies afin de disposer de populations de tailles suffisantes garantissant une puissance adéquate pour des études d'association génétique utilisant des données de séquençage et pour étudier les effets modificateurs de la radiothérapie et de la chimiothérapie sur le risque génétique.

S'ils sont répliqués de façon robuste, nos résultats présenteront un intérêt majeur pour une application clinique. Cela pourrait conduire à une meilleure identification des patients présentant un risque élevé de second cancer, permettant un traitement

personnalisé en tenant compte de la balance bénéfice/risque. De plus, cela pourrait faciliter la mise en œuvre de programmes personnalisés de surveillance afin de réduire la morbidité et la mortalité à long terme chez les patients survivants d'un cancer pédiatrique.

CHAPITRE 7

—

Discussion générale

Ce chapitre commence par une synthèse des principales conclusions de la thèse. Il aborde ensuite les implications cliniques potentielles, tant pour les patients atteints de cancer pédiatrique que pour le suivi des patients déjà traités. Enfin, il propose des pistes pour les recherches futures.

I. Résumé des principaux résultats

Chez les patients ayant déjà été traités pour un cancer pédiatrique, les seconds cancers sont la première cause de mortalité et de morbidité. Plusieurs facteurs de risque ont déjà été étudiés pour les différents types de seconds cancers. Ainsi, plusieurs études ont mis en évidence l'impact significatif du traitement reçu, en particulier la radiothérapie, sur l'ensemble des cancers solides, dont les plus fréquents sont les seconds cancers du sein et de la thyroïde.

Cependant, ces facteurs de risque ne suffisent pas à expliquer entièrement le risque de second cancer dans cette population croissante, et une variabilité interindividuelle persiste, incitant à examiner les prédispositions génétiques chez les survivants.

Ainsi, au cours de cette thèse, nous avons dans un premier temps mené une étude cas-témoins nichée dans deux cohortes de patients ayant été traités pour un cancer pédiatrique, la FCCSS et la BCCSS, afin d'étudier le risque de second cancer de la thyroïde après irradiation des organes lymphoïde notamment la moelle osseuse, la rate et le thymus. Nos résultats montrent un risque accru de cancer de la thyroïde après irradiation des organes du système immunitaire. En analysant chaque organe individuellement, nous avons établi une association entre l'irradiation du thymus et de la rate et le risque de second cancer de la thyroïde, avec un effet de dose notable. Ces résultats mettent en évidence l'impact des radiations sur les organes lymphoïdes, plutôt que sur la thyroïde elle-même mais également l'importance du système immunitaire dans le développement de seconds cancers.

Par la suite, afin d'étudier le risque de second cancer lié à des facteurs génétiques et leur rôle modificateur sur les effets des traitements, nous avons d'abord réalisé une revue systématique de la littérature. Cette revue avait pour objectif d'identifier l'ensemble des variants, gènes et voies biologiques associés au risque de développer un second cancer après un traitement pour un cancer pédiatrique. Nous avons analysé

18 études sur les associations génétiques, incluant des recherches sur des gènes candidats, des études d'association pangénomique et des analyses de séquençage de l'exome et du génome. Ces travaux se concentraient principalement sur les gènes impliqués dans les voies de réparation de l'ADN et le métabolisme des médicaments.

En compilant de manière exhaustive tous les variants et gènes associés, nous avons pu établir un état des lieux des connaissances actuelles concernant le risque de seconds cancers chez les survivants de cancers pédiatriques. Cependant, cette revue a également révélé qu'il existe un nombre limité d'études sur le risque génétique de seconds cancers, avec des résultats souvent hétérogènes et rarement répliqués.

Dans un second temps, toujours dans le but de mieux comprendre le risque de seconds cancers attribuable aux variants génétiques, nous avons mené une étude d'association génétique axée sur la recherche de variants rares associés au risque de seconds cancers, tous types confondus, avec une attention particulière portée sur les cancers du sein et de la thyroïde. Pour cela, nous avons utilisé le séquençage d'exome pour explorer le risque de second cancer après traitement d'un cancer pédiatrique. Cette approche a conduit à l'identification de plusieurs gènes jusqu'alors inconnus en lien avec le risque de développer un second cancer.

Plus précisément, nous avons mis en évidence 7 gènes qui présentent une association significative avec le risque global de seconds cancers, 8 gènes associés au risque de second cancer du sein, et 6 gènes liés au risque de second cancer de la thyroïde. En plus des gènes identifiés, notre étude a montré que certains variants rares pourraient jouer un rôle dans le risque de second cancer. Ces variants, bien que nécessitant d'être répliqués, représentent des cibles potentielles pour des recherches futures.

En examinant les gènes associés au risque de second cancer du sein, nous avons pu confirmer les résultats de l'étude menée par Morton et al. (2020) montrant une association entre le gène *FANCM* et le risque de second cancer du sein. Nous avons également montré que des gènes déjà connus dans la littérature pour leur impact sur

le risque de cancer dans la population générale, tels que *BLM* et *PML*, sont aussi associés au risque de second cancer du sein.

Nous avons observé une association entre le gène *APOBEC3F* et le risque de seconds cancers. Il convient de souligner que le variant rare rs146924719 a été identifié chez tous les patients porteurs de variants associés à ce gène. Bien que ces résultats soient préliminaires en raison de la taille limitée de notre échantillon et des contraintes méthodologiques de l'étude, qui ne permettent pas d'effectuer une analyse approfondie des variants rares de manière individuelle, ils laissent entrevoir une association potentielle avec le risque de second cancer.

Par ailleurs, nous avons également identifié une association significative entre le gène *RNASEL* et le risque de second cancer, avec un intérêt particulier pour le variant rare rs74315364. Ce variant, qui introduit un codon STOP dans la séquence protéique, a déjà été associé à divers types de cancers, ce qui suggère qu'il pourrait également jouer un rôle important dans le risque de second cancer.

En résumé, les résultats de cette thèse ont montré un risque accru de cancer de la thyroïde après irradiation des organes lymphoïdes chez des patients ayant survécu à un cancer pédiatrique, soulignant le rôle potentiel du système immunitaire dans le développement de ces cancers secondaires. En parallèle, la littérature génétique révèle que des variantes rares et des gènes impliqués dans la réparation de l'ADN et le métabolisme des médicaments pourraient également influencer ce risque. Ainsi, pour mieux comprendre l'interaction entre les facteurs tels que la radiothérapie et les facteurs génétiques, une étude d'association génétique a été réalisée. Elle a permis d'identifier plusieurs gènes et variants rares associés au risque de seconds cancers, notamment ceux du sein et de la thyroïde. Ces résultats, bien que préliminaires, ouvrent de nouvelles perspectives pour explorer la contribution génétique dans la variabilité du risque, et soulignent l'importance d'intégrer les données épidémiologiques, particulièrement concernant le traitement reçu lors du cancer pédiatrique, et génétiques pour mieux prédire et prévenir les seconds cancers chez les patients traités

pour un cancer pédiatrique.

II. Implications cliniques et perspectives

1. Validation des résultats

Les futures recherches doivent se concentrer sur plusieurs axes pour approfondir nos résultats et les confirmer.

Tout d'abord, il est essentiel de réaliser des études de validation dans des cohortes indépendantes pour tester la robustesse de nos résultats et garantir leur généralisation à d'autres populations indépendantes. Une telle approche, essentielle dans toute étude d'association génétique, permettra de renforcer nos résultats et d'assurer qu'ils puissent être appliqués de manière pertinente dans un contexte clinique.

Le deuxième axe de recherche à développer concerne la poursuite des validations des résultats issus des nombreuses études d'association génétique précédemment publiées et compilées dans notre revue systématique. Un grand nombre de ces études ont mis en avant des associations potentielles entre des variants génétiques et le risque de second cancer, mais une grande partie de ces résultats n'a pas encore été validée de manière indépendante. Or, avant de pouvoir envisager une quelconque application clinique, ces résultats doivent être rigoureusement testés et confirmés.

Cela implique de réévaluer les résultats antérieurs avec des échantillons plus larges et mieux caractérisés, tout en prenant en compte les avancées méthodologiques récentes dans le domaine de la génomique. Par exemple, les technologies de séquençage à haut débit et les outils d'analyse bio-informatique plus sophistiqués permettent désormais d'analyser des volumes de données bien plus importants et d'identifier des interactions complexes entre les gènes et les traitements anticancéreux reçus. Ces efforts de validation, bien qu'exigeants en termes de temps et de ressources, sont essentiels pour garantir la reproductibilité et la fiabilité des découvertes dans un contexte clinique. Cela

souligne également l'importance des collaborations entre équipes de recherche et la constitution de banques biologiques adaptées.

2. Importance des collaborations internationales

En effet, les études d'association génétique nécessitent souvent des échantillons de très grande taille pour être suffisamment puissantes sur le plan statistique, surtout lorsqu'il s'agit d'étudier des variants rares ou des interactions complexes entre gènes et environnement. La constitution de cohortes de grande envergure, à travers des partenariats internationaux, permettrait non seulement d'augmenter la taille des échantillons, mais aussi de diversifier les populations étudiées, ce qui est essentiel pour la généralisation des résultats. Ces collaborations internationales permettront également d'étudier de manière plus approfondie les interactions entre gènes et traitements, notamment la radiothérapie et la chimiothérapie, qui sont des facteurs cruciaux dans l'apparition des seconds cancers. En explorant comment les variants génétiques influencent la réponse aux traitements anticancéreux, nous pourrions identifier des sous-groupes de patients qui présentent un risque plus élevé de développer des complications à long terme, comme des cancers secondaires.

3. Création de banques biologiques

La création de banques biologiques, regroupant des échantillons de tissus, de sang et d'autres matériaux biologiques, constitue un élément clé pour soutenir ces recherches. Ces banques ne se contentent pas de préserver des échantillons pour des analyses futures, elles fournissent également une ressource précieuse pour réaliser des études de réplication et de plus grande envergure à mesure que leur contenu s'enrichit.

Actuellement, la banque biologique majeure de la FCCS est en cours de séquençage, ce qui facilitera la réplication de nos résultats.

4. Etudes fonctionnelles

En parallèle à ces études de validation, un troisième axe majeur concerne les

investigations fonctionnelles sur les variants rares identifiés au cours de l'étude. Pour comprendre l'impact de ces variants rares sur la biologie des cellules et sur le développement de cancers secondaires, des études fonctionnelles in vitro et in vivo seront nécessaires. Ces études permettront de déterminer si ces variants altèrent réellement les processus cellulaires mais pourront aussi permettre de mieux comprendre si certains variants rares augmentent spécifiquement la susceptibilité des individus à des effets indésirables des traitements anticancéreux, tels que la radiothérapie ou la chimiothérapie ouvrant ainsi la voie à des ajustements thérapeutiques individualisés, où les patients porteurs de ces variants bénéficieraient de thérapies alternatives ou de doses ajustées pour minimiser le risque de cancer secondaire tout en maintenant l'efficacité du traitement contre le cancer primaire.

5. Implication cliniques des résultats pour une médecine personnalisée

Les résultats obtenus dans le cadre de ces recherches, s'ils sont confirmés et validés de manière robuste, pourraient avoir des implications cliniques majeures. Une des premières applications cliniques serait la possibilité d'identifier les patients à risque élevé de développer un second cancer, permettant ainsi de personnaliser les traitements en fonction des profils génétiques individuels. Cela pourrait mener à une optimisation des traitements en prenant en compte la balance bénéfice/risque pour chaque patient. Par exemple, les patients porteurs de certains variants génétiques pourraient bénéficier de traitements moins agressifs, réduisant ainsi leur risque de complications à long terme tout en maintenant l'efficacité du traitement.

De plus, des stratégies de surveillance plus précises pourraient être mises en place pour les patients identifiés comme étant à haut risque. Ces patients pourraient faire l'objet de suivis plus fréquents, permettant une détection précoce des cancers secondaires. Cela pourrait avoir un impact direct sur la survie des patients, car les cancers détectés à un stade précoce sont souvent plus faciles à traiter.

Enfin, à plus long terme, ces recherches pourraient guider le développement de

thérapies ciblées pour neutraliser les effets des mutations spécifiques identifiées. Cela permettrait d'améliorer non seulement le traitement initial, mais aussi de diminuer le risque de complications à long terme, incluant les cancers secondaires.

Conclusion

En résumé, les futures recherches devront s'appuyer sur des validations rigoureuses, des études fonctionnelles approfondies et des collaborations internationales pour confirmer les résultats obtenus et les traduire en applications cliniques. Les implications potentielles de ces recherches sont vastes et pourraient transformer la manière dont nous traitons et surveillons les patients atteints de cancers pédiatriques et d'autres types de cancers. Une médecine véritablement personnalisée, basée sur les profils génétiques individuels, pourrait émerger de ces efforts, améliorant la qualité de vie et les perspectives de survie des patients à long terme.

Bibliographie

- Allodji, R. S., Haddy, N., Vu-Bezin, G., Dumas, A., Fresneau, B., Mansouri, I., Demoor-Goldschmidt, C., El-Fayech, C., Pacquement, H., Munzer, M., Bondiau, P.-Y., Berchery, D., Oberlin, O., Rubino, C., Diallo, I., & de Vathaire, F. (2019). Risk of subsequent colorectal cancers after a solid tumor in childhood: Effects of radiation therapy and chemotherapy. *Pediatric Blood & Cancer*, *66*(2), e27495. <https://doi.org/10.1002/pbc.27495>
- Allodji, R. S., Tucker, M. A., Hawkins, M. M., Le Deley, M.-C., Veres, C., Weathers, R., Howell, R., Winter, D., Haddy, N., Rubino, C., Diallo, I., Little, M. P., Morton, L. M., & de Vathaire, F. (2021). Role of radiotherapy and chemotherapy in the risk of leukemia after childhood cancer: An international pooled analysis. *International Journal of Cancer*, *148*(9), 2079-2089. <https://doi.org/10.1002/ijc.33361>
- Amir-Behghadami, M., & Janati, A. (2020). Population, Intervention, Comparison, Outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. *Emergency Medicine Journal: EMJ*, *37*(6), 387. <https://doi.org/10.1136/emered-2020-209567>
- Applebaum, M. A., Vaksman, Z., Lee, S. M., Hungate, E. A., Henderson, T. O., London, W. B., Pinto, N., Volchenbom, S. L., Park, J. R., Naranjo, A., Hero, B., Pearson, A. D., Stranger, B. E., Cohn, S. L., & Diskin, S. J. (2017). *Neuroblastoma survivors are at increased risk for second malignancies: A report from the International Neuroblastoma Risk Group Project*. <https://doi.org/10.1016/j.ejca.2016.11.022>
- Armenian, S. H., & Bhatia, S. (2009). Chronic Health Conditions in Childhood Cancer

Survivors: Is it All Treatment-Related—or Do Genetics Play a Role? *Journal of General Internal Medicine*, 24(Suppl 2), 395. <https://doi.org/10.1007/s11606-009-0995-8>

Armstrong, G. T., Liu, Q., Yasui, Y., Neglia, J. P., Leisenring, W., Robison, L. L., & Mertens, A. C. (2009). Late Mortality Among 5-Year Survivors of Childhood Cancer: A Summary From the Childhood Cancer Survivor Study. *Journal of Clinical Oncology*, 27(14), 2328-2338. <https://doi.org/10.1200/JCO.2008.21.1425>

Armstrong, G. T., Oeffinger, K. C., Chen, Y., Kawashima, T., Yasui, Y., Leisenring, W., Stovall, M., Chow, E. J., Sklar, C. A., Mulrooney, D. A., Mertens, A. C., Border, W., Durand, J.-B., Robison, L. L., & Meacham, L. R. (2013). Modifiable risk factors and major cardiac events among adult survivors of childhood cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 31(29), 3673-3680. <https://doi.org/10.1200/JCO.2013.49.3205>

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1), 25-29. <https://doi.org/10.1038/75556>

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E.

- S., Lee, C., ... National Eye Institute, N. (2015). A global reference for human genetic variation. *Nature*, 526(7571), Article 7571. <https://doi.org/10.1038/nature15393>
- Bansal, V., Libiger, O., Torkamani, A., & Schork, N. J. (2010). Statistical Analysis Strategies for Association Studies Involving Rare Variants. *Nature reviews. Genetics*, 11(11), 773-785. <https://doi.org/10.1038/nrg2867>
- Bartsch, D. K., Fendrich, V., Slater, E. P., Sina-Frey, M., Rieder, H., Greenhalf, W., Chaloupka, B., Hahn, S. A., Neoptolemos, J. P., & Kress, R. (2005). RNASEL germline variants are associated with pancreatic cancer. *International Journal of Cancer*, 117(5), 718-722. <https://doi.org/10.1002/ijc.21254>
- Best, T., Li, D., Skol, A. D., Kirchhoff, T., Jackson, S. A., Yasui, Y., Bhatia, S., Strong, L. C., Domchek, S. M., Nathanson, K. L., Olopade, O. I., Huang, R. S., Mack, T. M., Conti, D. V., Offit, K., Cozen, W., Robison, L. L., & Onel, K. (2011). Variants at 6q21 implicate PRDM1 in the etiology of therapy-induced second malignancies after Hodgkin's lymphoma. *Nature Medicine*, 17(8), 941-943. <https://doi.org/10.1038/nm.2407>
- Bhakta, N., Liu, Q., Ness, K. K., Baassiri, M., Eissa, H., Yeo, F., Chemaitilly, W., Ehrhardt, M. J., Bass, J., Bishop, M. W., Shelton, K., Lu, L., Huang, S., Li, Z., Caron, E., Lanctot, J., Howell, C., Folse, T., Joshi, V., ... Robison, L. L. (2017). The Cumulative Burden of Surviving Childhood Cancer: An Initial Report from the St. Jude Lifetime Cohort Study. *Lancet (London, England)*, 390(10112), 2569-2582. [https://doi.org/10.1016/S0140-6736\(17\)31610-0](https://doi.org/10.1016/S0140-6736(17)31610-0)

- Bhatia, S. (2015). *Genetic variation as a modifier of association between therapeutic exposure and subsequent malignant neoplasms in cancer survivors*.
<https://doi.org/10.1002/cncr.29096>
- Bhatia, S., & Sklar, C. (2002). Second cancers in survivors of childhood cancer. *Nature Reviews Cancer*, 2(2), 124-132. <https://doi.org/10.1038/nrc722>
- Bhatti, P., Veiga, L. H. S., Ronckers, C. M., Sigurdson, A. J., Stovall, M., Smith, S. A., Weathers, R., Leisenring, W., Mertens, A. C., Hammond, S., Friedman, D. L., Neglia, J. P., Meadows, A. T., Donaldson, S. S., Sklar, C. A., Robison, L. L., & Inskip, P. D. (2010). Risk of Second Primary Thyroid Cancer after Radiotherapy for a Childhood Cancer in a Large Cohort Study: An Update from the Childhood Cancer Survivor Study. *Radiation research*, 174(6), 741-752.
<https://doi.org/10.1667/RR2240.1>
- Blanco, J. G., Edick, M. J., Hancock, M. L., Winick, N. J., Dervieux, T., Amylon, M. D., Bash, R. O., Behm, F. G., Camitta, B. M., Pui, C.-H., Raimondi, S. C., & Relling, M. V. (2002). Genetic polymorphisms in CYP3A5, CYP3A4 and NQO1 in children who developed therapy-related myeloid malignancies. *Pharmacogenetics and Genomics*, 12(8), 605.
- Bolderson, E., Tomimatsu, N., Richard, D. J., Boucher, D., Kumar, R., Pandita, T. K., Burma, S., & Khanna, K. K. (2010). Phosphorylation of Exo1 modulates homologous recombination repair of DNA double-strand breaks. *Nucleic Acids Research*, 38(6), 1821-1831. <https://doi.org/10.1093/nar/gkp1164>
- Botta, L., Gatta, G., Capocaccia, R., Stiller, C., Cañete, A., Dal Maso, L., Innos, K., Mihor,

- A., Erdmann, F., Spix, C., Lacour, B., Marcos-Gragera, R., Murray, D., Rossi, S., & EURO CARE-6 Working Group. (2022). Long-term survival and cure fraction estimates for childhood cancer in Europe (EURO CARE-6): Results from a population-based study. *The Lancet. Oncology*, 23(12), 1525-1536. [https://doi.org/10.1016/S1470-2045\(22\)00637-4](https://doi.org/10.1016/S1470-2045(22)00637-4)
- Broniscer, A., Ke, W., Fuller, C. E., Wu, J., Gajjar, A., & Kun, L. E. (2004). *Second Neoplasms in Pediatric Patients with Primary Central Nervous System Tumors: The St. Jude Children's Research Hospital Experience*. <https://doi.org/10.1002/cncr.20253>
- Butler, E., Ludwig, K., Pacent, H. L., Klesse, L. J., Watt, T. C., & Laetsch, T. W. (2021). Recent progress in the treatment of cancer in children. *CA: A Cancer Journal for Clinicians*, 71(4), 315-332. <https://doi.org/10.3322/caac.21665>
- Carpten, J., Nupponen, N., Isaacs, S., Sood, R., Robbins, C., Xu, J., Faruque, M., Moses, T., Ewing, C., Gillanders, E., Hu, P., Bujnovszky, P., Makalowska, I., Baffoe-Bonnie, A., Faith, D., Smith, J., Stephan, D., Wiley, K., Brownstein, M., ... Trent, J. (2002). Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. *Nature Genetics*, 30(2), 181-184. <https://doi.org/10.1038/ng823>
- Catucci, I., Osorio, A., Arver, B., Neidhardt, G., Bogliolo, M., Zanardi, F., Riboni, M., Minardi, S., Pujol, R., Azzollini, J., Peissel, B., Manoukian, S., De Vecchi, G., Casola, S., Hauke, J., Richters, L., Rhiem, K., Schmutzler, R. K., Wallander, K., ... Peterlongo, P. (2018). Individuals with FANCM biallelic mutations do not develop Fanconi anemia, but show risk for breast cancer, chemotherapy toxicity and may display chromosome fragility. *Genetics in Medicine: Official Journal of the American*

College of Medical Genetics, 20(4), 452-457.

<https://doi.org/10.1038/gim.2017.123>

Chemaitilly, W., & Hudson, M. M. (2014). *Update on endocrine and metabolic therapy-related late effects observed in survivors of childhood neoplasia.*

<https://doi.org/10.1097/MED.0000000000000029>

Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, 11(6),

415-425. <https://doi.org/10.1038/nrg2779>

Connell, P. P., & Hellman, S. (2009). Advances in Radiotherapy and Implications for the Next Century: A Historical Perspective. *Cancer Research*, 69(2), 383-392.

<https://doi.org/10.1158/0008-5472.CAN-07-6871>

Danuta, G., Tobias, M., Marcus, D., Miriam, E., Nergiz, K., Olesja, S., Steffen, R., Tanja, Z., Christian, M., Thomas, H., Peter, S.-K., Heather, C., Johanna, M., Alicia, P., Heidi, R., Claudia, S., Thomas, H., Dirk, P., Manuela, M., & Heinz, S. (2020). *Molecular karyotyping and gene expression analysis in childhood cancer patients.*

<https://doi.org/10.1007/s00109-020-01937-4>

Datta, A., Mazumder, M. H. H., Chowdhury, A. S., & Hasan, M. A. (2015). Functional and Structural Consequences of Damaging Single Nucleotide Polymorphisms in

Human Prostate Cancer Predisposition Gene RNASEL. *BioMed Research International*, 2015, 271458. <https://doi.org/10.1155/2015/271458>

de Baat, E. C., Feijen, E. A. M., Reulen, R. C., Allodji, R. S., Bagnasco, F., Bardi, E., Belle, F.

N., Byrne, J., van Dalen, E. C., Debiche, G., Diallo, I., Grabow, D., Hjorth, L.,

- Jankovic, M., Kuehni, C. E., Levitt, G., Llanas, D., Loonen, J., Zaletel, L. Z., ... Kremer, L. C. M. (2023). Risk Factors for Heart Failure Among Pan-European Childhood Cancer Survivors: A PanCareSurFup and ProCardio Cohort and Nested Case-Control Study. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 41(1), 96-106. <https://doi.org/10.1200/JCO.21.02944>
- de Laage, A., Allodji, R., Dauchy, S., Rivollet, S., Fayeche, C., Fresneau, B., & Oberlin, O. (2016). Screening for psychological distress in very long-term adult survivors of childhood cancer. *Pediatric Hematology and Oncology*, 33(5), 295-313. <https://doi.org/10.1080/08880018.2016.1204400>
- Demoor-Goldschmidt, C., Allodji, R. S., Journy, N., Rubino, C., Zrafi, W. S., Debiche, G., Llanas, D., Veres, C., Thomas-Teinturier, C., Pacquement, H., Vu-Bezin, G., Fresneau, B., Berchery, D., Bolle, S., Diallo, I., Haddy, N., & de Vathaire, F. (2020). Risk Factors for Small Adult Height in Childhood Cancer Survivors. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 38(16), 1785-1796. <https://doi.org/10.1200/JCO.19.02361>
- de Vathaire, F., El-Fayeche, C., Ben Ayed, F. F., Haddy, N., Guibout, C., Winter, D., Thomas-Teinturier, C., Veres, C., Jackson, A., Pacquement, H., Schlumberger, M., Hawkins, M., Diallo, I., & Oberlin, O. (2012). Radiation dose to the pancreas and risk of diabetes mellitus in childhood cancer survivors: A retrospective cohort study. *The Lancet. Oncology*, 13(10), 1002-1010. [https://doi.org/10.1016/S1470-2045\(12\)70323-6](https://doi.org/10.1016/S1470-2045(12)70323-6)

de Vathaire, F., Haddy, N., Allodji, R. S., Hawkins, M., Guibout, C., El-Fayech, C., Teinturier, C., Oberlin, O., Pacquement, H., Diop, F., Kalhouche, A., Benadjaoud, M., Winter, D., Jackson, A., Bezin Mai-Quynh, G., Benabdennebi, A., Llanas, D., Veres, C., Munzer, M., ... Rubino, C. (2015). Thyroid Radiation Dose and Other Risk Factors of Thyroid Carcinoma Following Childhood Cancer. *The Journal of Clinical Endocrinology and Metabolism*, 100(11), 4282-4290. <https://doi.org/10.1210/jc.2015-1690>

de Vathaire, F., Scwhartz, B., El-Fayech, C., Allodji, R. S., Escudier, B., Hawkins, M., Diallo, I., & Haddy, N. (2015). Risk of a Second Kidney Carcinoma Following Childhood Cancer: Role of Chemotherapy and Radiation Dose to Kidneys. *The Journal of Urology*, 194(5), 1390-1395. <https://doi.org/10.1016/j.juro.2015.06.092>

Dicks, E., Song, H., Ramus, S. J., Oudenhove, E. V., Tyrer, J. P., Intermaggio, M. P., Kar, S., Harrington, P., Bowtell, D. D., Group, A. S., Cicek, M. S., Cunningham, J. M., Fridley, B. L., Alsop, J., Jimenez-Linan, M., Piskorz, A., Goranova, T., Kent, E., Siddiqui, N., ... Pharoah, D. P. P. (2017). Germline whole exome sequencing and large-scale replication identifies FANCM as a likely high grade serous ovarian cancer susceptibility gene. *Oncotarget*, 8(31), 50930-50940. <https://doi.org/10.18632/oncotarget.15871>

Dietz, A. C., Chen, Y., Yasui, Y., Ness, K. K., Hagood, J. S., Chow, E. J., Stovall, M., Neglia, J. P., Oeffinger, K. C., Mertens, A. C., Robison, L. L., Armstrong, G. T., & Mulrooney, D. A. (2016). Risk and impact of pulmonary complications in survivors of childhood cancer: A report from the Childhood Cancer Survivor Study. *Cancer*,

122(23), 3687-3696. <https://doi.org/10.1002/cncr.30200>

Dincer, Y., Yüksel, S., Batar, B., Güven, M., Onaran, I., & Celkan, T. (2015). *DNA Repair Gene Polymorphisms and Their Relation with DNA Damage, DNA Repair, and Total Antioxidant Capacity in Childhood Acute Lymphoblastic Leukemia Survivors*. <https://doi.org/10.1097/MPH.0000000000000133>

Dixon, S. B., Bjornard, K. L., Alberts, N. M., Armstrong, G. T., Brinkman, T. M., Chemaitilly, W., Ehrhardt, M. J., Fernandez-Pineda, I., Force, L. M., Gibson, T. M., Green, D. M., Howell, C. R., Kaste, S. C., Kirchhoff, A. C., Klosky, J. L., Krull, K. R., Lucas, Jr., J. T., Mulrooney, D. A., Ness, K. K., ... Hudson, M. M. (2018). *Factors influencing risk-based care of the childhood cancer survivor in the 21st century*. <https://doi.org/10.3322/caac.21445>

Dumas, A., Berger, C., Auquier, P., Michel, G., Fresneau, B., Allodji, R. S., Haddy, N., Rubino, C., Vassal, G., Valteau-Couanet, D., Thouvenin-Doulet, S., Casagrande, L., Pacquement, H., El-Fayech, C., Oberlin, O., Guibout, C., & de Vathaire, F. (2016). Educational and occupational outcomes of childhood cancer survivors 30 years after diagnosis: A French cohort study. *British Journal of Cancer*, 114(9), 1060-1068. <https://doi.org/10.1038/bjc.2016.62>

Ehrhardt, M. J., Howell, C. R., Hale, K., Baassiri, M. J., Rodriguez, C., Wilson, C. L., Joshi, S. S., Lemond, T. C., Shope, S., Howell, R. M., Wang, Z., Srivastava, D., Mulrooney, D. A., Zhang, J., Robison, L. L., Ness, K. K., & Hudson, M. M. (2019). Subsequent Breast Cancer in Female Childhood Cancer Survivors in the St Jude Lifetime Cohort Study (SJLIFE). *Journal of Clinical Oncology: Official Journal of the*

- American Society of Clinical Oncology*, 37(19), 1647-1656.
<https://doi.org/10.1200/JCO.18.01099>
- Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews. Genetics*, 11(6), 446-450.
<https://doi.org/10.1038/nrg2809>
- El-Fayech, C., Haddy, N., Allodji, R. S., Veres, C., Diop, F., Kahlouche, A., Llanas, D., Jackson, A., Rubino, C., Guibout, C., Pacquement, H., Oberlin, O., Thomas-Teinturier, C., Scarabin, P.-Y., Chavaudra, J., Lefkopoulos, D., Giroud, M., Bejot, Y., Bernier, V., ... de Vathaire, F. (2017). Cerebrovascular Diseases in Childhood Cancer Survivors: Role of the Radiation Dose to Willis Circle Arteries. *International Journal of Radiation Oncology, Biology, Physics*, 97(2), 278-286.
<https://doi.org/10.1016/j.ijrobp.2016.10.015>
- Elsir, T., Smits, A., Lindström, M. S., & Nistér, M. (2012). Transcription factor PROX1: Its role in development and cancer. *Cancer Metastasis Reviews*, 31(3-4), 793-805.
<https://doi.org/10.1007/s10555-012-9390-8>
- Erčulj, N., Faganel Kotnik, B., Debeljak, M., Jazbec, J., & Dolžan, V. (2012). DNA repair polymorphisms influence the risk of second neoplasm after treatment of childhood acute lymphoblastic leukemia. *Journal of Cancer Research and Clinical Oncology*, 138(11), 1919-1930. <https://doi.org/10.1007/s00432-012-1265-4>
- Fang, X., Huang, X., Lu, J., & Su, D. (2024). Causal role of immune cells in thyroid cancer: A bidirectional Mendelian randomization study. *Frontiers in Immunology*, 15.

<https://doi.org/10.3389/fimmu.2024.1425873>

Felix, C. A., Walker, A. H., Lange, B. J., Williams, T. M., Winick, N. J., Cheung, N.-K. V., Lovett, B. D., Nowell, P. C., Blair, I. A., & Rebbeck, T. R. (1998). *Association of CYP3A4 genotype with treatment-related leukemia.*

<https://doi.org/10.1073/pnas.95.22.13176>

Fidler, M. M., Reulen, R. C., Winter, D. L., Kelly, J., Jenkinson, H. C., Skinner, R., Frobisher, C., & Hawkins, M. M. (2016). Long term cause specific mortality among 34 489 five year survivors of childhood cancer in Great Britain: Population based cohort study. *BMJ*, *354*, i4351. <https://doi.org/10.1136/bmj.i4351>

Figlioli, G., Bogliolo, M., Catucci, I., Caleca, L., Lasheras, S. V., Pujol, R., Kiiski, J. I., Muranen, T. A., Barnes, D. R., Dennis, J., Michailidou, K., Bolla, M. K., Leslie, G., Aalfs, C. M., ABCTB Investigators, Adank, M. A., Adlard, J., Agata, S., Cadoo, K., ... Peterlongo, P. (2019). The FANCM:p.Arg658* truncating variant is associated with risk of triple-negative breast cancer. *NPJ Breast Cancer*, *5*, 38. <https://doi.org/10.1038/s41523-019-0127-5>

Friedman, D. L., Whitton, J., Leisenring, W., Mertens, A. C., Hammond, S., Stovall, M., Donaldson, S. S., Meadows, A. T., Robison, L. L., & Neglia, J. P. (2010). Subsequent neoplasms in 5-year survivors of childhood cancer: The Childhood Cancer Survivor Study. *Journal of the National Cancer Institute*, *102*(14), 1083-1095. <https://doi.org/10.1093/jnci/djq238>

Gatta, G., Botta, L., Rossi, S., Aareleid, T., Bielska-Lasota, M., Clavel, J., Dimitrova, N., Jakab, Z., Kaatsch, P., Lacour, B., Mallone, S., Marcos-Gragera, R., Minicozzi, P.,

- Sánchez-Pérez, M.-J., Sant, M., Santaquilani, M., Stiller, C., Tavilla, A., Trama, A., ... EUROCARE Working Group. (2014). Childhood cancer survival in Europe 1999-2007: Results of EUROCARE-5--a population-based study. *The Lancet. Oncology*, 15(1), 35-47. [https://doi.org/10.1016/S1470-2045\(13\)70548-5](https://doi.org/10.1016/S1470-2045(13)70548-5)
- Gene Ontology Consortium, Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., ... Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1), iyad031. <https://doi.org/10.1093/genetics/iyad031>
- Ghosh, T., Chen, Y., Dietz, A. C., Armstrong, G. T., Howell, R. M., Smith, S. A., Mulrooney, D. A., Turcotte, L. M., Yuan, Y., Yasui, Y., & Neglia, J. P. (2021). Lung Cancer as a Subsequent Malignant Neoplasm in Survivors of Childhood Cancer. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 30(12), 2235-2243. <https://doi.org/10.1158/1055-9965.EPI-21-0250>
- Goswami, C., Chattopadhyay, A., & Chuang, E. Y. (2021). Rare variants: Data types and analysis strategies. *Annals of Translational Medicine*, 9(12), 961. <https://doi.org/10.21037/atm-21-1635>
- Gramatges, M. M., & Bhatia, S. (2018). *Evidence for Genetic Risk Contributing to Long-Term Adverse Treatment Effects in Childhood Cancer Survivors.*

<https://doi.org/10.1146/annurev-med-041916-124328>

Gramatges, M. M., Liu, Q., Yasui, Y., Okcu, M. F., Neglia, J. P., Strong, L. C., Armstrong, G. T., Robison, L. L., & Bhatia, S. (2014). Telomere Content and Risk of Second Malignant Neoplasm in Survivors of Childhood Cancer: A Report from the Childhood Cancer Survivor Study. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 20(4), 904-911. <https://doi.org/10.1158/1078-0432.CCR-13-2076>

Gramatges, M. M., Morton, L. M., Yasui, Y., Arnold, M. A., Neglia, J. P., Leisenring, W. M., Machiela, M. J., Dagnall, C. L., Chanock, S. J., Armstrong, G. T., Robison, L. L., Bhatia, S., & Lupo, P. J. (2019). *Telomere length-associated genetic variants and the risk of thyroid cancer in survivors of childhood cancer: A report from the Childhood Cancer Survivor Study (CCSS)*. <https://doi.org/10.1158/1055-9965.EPI-18-0972>

Green, D. M., Kawashima, T., Stovall, M., Leisenring, W., Sklar, C. A., Mertens, A. C., Donaldson, S. S., Byrne, J., & Robison, L. L. (2009). Fertility of Female Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. *Journal of Clinical Oncology*, 27(16), 2677-2685. <https://doi.org/10.1200/JCO.2008.20.1541>

Guldner, L., Haddy, N., Pein, F., Diallo, I., Shamsaldin, A., Dahan, M., Lebidois, J., Merlet, P., Villain, E., Sidi, D., Sakiroglu, O., Hartmann, O., Leftakopoulos, D., & de Vathaire, F. (2006). Radiation dose and long term risk of cardiac pathology following radiotherapy and anthracyclin for a childhood cancer. *Radiotherapy*

and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology, 81(1), 47-56. <https://doi.org/10.1016/j.radonc.2006.08.020>

Hadden, J. W. (2003). Immunodeficiency and cancer: Prospects for correction. *International Immunopharmacology*, 3(8), 1061-1071. [https://doi.org/10.1016/S1567-5769\(03\)00060-2](https://doi.org/10.1016/S1567-5769(03)00060-2)

Haddy, N., Diallo, S., El-Fayech, C., Schwartz, B., Pein, F., Hawkins, M., Veres, C., Oberlin, O., Guibout, C., Pacquement, H., Munzer, M., N'Guyen, T. D., Bondiau, P.-Y., Berchery, D., Laprie, A., Scarabin, P.-Y., Jouven, X., Bridier, A., Koscielny, S., ... de Vathaire, F. (2016). Cardiac Diseases Following Childhood Cancer Treatment. *Circulation*, 133(1), 31-38. <https://doi.org/10.1161/CIRCULATIONAHA.115.016686>

Haddy, N., Dondon, M.-G., Paoletti, C., Rubino, C., Mousannif, A., Shamsaldin, A., Doyon, F., Labbé, M., Robert, C., Avril, M.-F., Demars, R., Molinie, F., Lefkopoulos, D., Diallo, I., & de Vathaire, F. (2010). Breast cancer following radiotherapy for a hemangioma during childhood. *Cancer Causes & Control: CCC*, 21(11), 1807-1816. <https://doi.org/10.1007/s10552-010-9607-5>

Haddy, N., Le Deley, M. C., Samand, A., Diallo, I., Guérin, S., Guibout, C., Oberlin, O., Hawkins, M., Zucker, J.-M., & de Vathaire, F. (2006). Role of radiotherapy and chemotherapy in the risk of secondary leukaemia after a solid tumour in childhood. *European Journal of Cancer (Oxford, England: 1990)*, 42(16), 2757-2764. <https://doi.org/10.1016/j.ejca.2006.05.034>

Haddy, N., Mousannif, A., Tukenova, M., Guibout, C., Grill, J., Dhermain, F., Pacquement,

- H., Oberlin, O., El-Fayech, C., Rubino, C., Thomas-Teinturier, C., Le-Deley, M.-C., Hawkins, M., Winter, D., Chavaudra, J., Diallo, I., & de Vathaire, F. (2011). Relationship between the brain radiation dose for the treatment of childhood cancer and the risk of long-term cerebrovascular mortality. *Brain: A Journal of Neurology*, 134(Pt 5), 1362-1372. <https://doi.org/10.1093/brain/awr071>
- Hawkins, M. M., Lancashire, E. R., Winter, D. L., Frobisher, C., Reulen, R. C., Taylor, A. J., Stevens, M. C. G., & Jenney, M. (2008). The British Childhood Cancer Survivor Study: Objectives, methods, population structure, response rates and initial descriptive information. *Pediatric Blood & Cancer*, 50(5), 1018-1025. <https://doi.org/10.1002/pbc.21335>
- Henderson, T. O., Liu, Q., Turcotte, L. M., Neglia, J. P., Leisenring, W., Hodgson, D., Diller, L., Kenney, L., Morton, L., Berrington de Gonzalez, A., Arnold, M., Bhatia, S., Howell, R. M., Smith, S. A., Robison, L. L., Armstrong, G. T., Oeffinger, K. C., Yasui, Y., & Moskowitz, C. S. (2022). Association of Changes in Cancer Therapy Over 3 Decades With Risk of Subsequent Breast Cancer Among Female Childhood Cancer Survivors: A Report From the Childhood Cancer Survivor Study (CCSS). *JAMA Oncology*, 8(12), 1765-1774. <https://doi.org/10.1001/jamaoncol.2022.4649>
- Henderson, T. O., Moskowitz, C. S., Chou, J. F., Bradbury, A. R., Neglia, J. P., Dang, C. T., Onel, K., Novetsky Friedman, D., Bhatia, S., Strong, L. C., Stovall, M., Kenney, L. B., Barnea, D., Lorenzi, E., Hammond, S., Leisenring, W. M., Robison, L. L., Armstrong, G. T., Diller, L. R., & Oeffinger, K. C. (2016). Breast Cancer Risk in Childhood

Cancer Survivors Without a History of Chest Radiotherapy: A Report From the Childhood Cancer Survivor Study. *Journal of Clinical Oncology*, 34(9), 910-918.

<https://doi.org/10.1200/JCO.2015.62.3314>

Henderson, T. O., Rajaraman, P., Stovall, M., Constine, L. S., Olive, A., Smith, S. A., Mertens, A., Meadows, A., Neglia, J. P., Hammond, S., Whitton, J., Inskip, P. D., Robison, L. L., & Diller, L. (2012). Risk factors associated with secondary sarcomas in childhood cancer survivors: A report from the childhood cancer survivor study. *International Journal of Radiation Oncology, Biology, Physics*, 84(1), 224-230. <https://doi.org/10.1016/j.ijrobp.2011.11.022>

Hoeijmakers, J. H. J. (2001). Genome maintenance mechanisms for preventing cancer. *Nature*, 411(6835), 366-374. <https://doi.org/10.1038/35077232>

Inskip, P. D., Sigurdson, A. J., Veiga, L., Bhatti, P., Ronckers, C., Rajaraman, P., Boukheris, H., Stovall, M., Smith, S., Hammond, S., Henderson, T. O., Watt, T. C., Mertens, A. C., Leisenring, W., Stratton, K., Whitton, J., Donaldson, S. S., Armstrong, G. T., Robison, L. L., & Neglia, J. P. (2016). Radiation-Related New Primary Solid Cancers in the Childhood Cancer Survivor Study: Comparative Radiation Dose Response and Modification of Treatment Effects. *International Journal of Radiation Oncology, Biology, Physics*, 94(4), 800-807. <https://doi.org/10.1016/j.ijrobp.2015.11.046>

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27-30.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins,

- R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., ... MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434-443. <https://doi.org/10.1038/s41586-020-2308-7>
- Kelly, K. M., Perentesis, J. P., & Group (COG), O. behalf of the C. O. (2002). *Polymorphisms of drug metabolizing enzymes and markers of genotoxicity to identify patients with Hodgkin's lymphoma at risk of treatment-related complications*. <https://doi.org/10.1093/annonc/13.S1.34>
- Kiiski, J. I., Pelttari, L. M., Khan, S., Freysteinsdottir, E. S., Reynisdottir, I., Hart, S. N., Shimelis, H., Vilske, S., Kallioniemi, A., Schleutker, J., Leminen, A., Bützow, R., Blomqvist, C., Barkardottir, R. B., Couch, F. J., Aittomäki, K., & Nevanlinna, H. (2014). Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15172-15177. <https://doi.org/10.1073/pnas.1407909111>
- Knapke, S., Nagarajan, R., Correll, J., Kent, D., & Burns, K. (2012). *Hereditary cancer risk assessment in a pediatric oncology follow-up clinic*. <https://doi.org/10.1002/pbc.23283>
- Kony, S. J., De Vathaire, F., Chompret, A., Shamsaldim, A., Grimaud, E., Raquin, M.-A., Oberlin, O., Brugières, L., Feunteun, J., Eschwège, F., Chavaudra, J., Lemerle, J., & Bonaïti-Pellié, C. (1997). *Radiation and genetic factors in the risk of second*

malignant neoplasms after a first cancer in childhood.

[https://doi.org/10.1016/S0140-6736\(97\)01116-1](https://doi.org/10.1016/S0140-6736(97)01116-1)

Kwak, J. C., Ongusaha, P. P., Ouchi, T., & Lee, S. W. (2003). IFI16 as a Negative Regulator in the Regulation of p53 and p21Waf1*. *Journal of Biological Chemistry*, 278(42), 40899-40904. <https://doi.org/10.1074/jbc.M308012200>

Lacour, B., Goujon, S., Guissou, S., Guyot-Goubin, A., Desmée, S., Désandes, E., & Clavel, J. (2014). Childhood cancer survival in France, 2000-2008. *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation (ECP)*, 23(5), 449-457. <https://doi.org/10.1097/CEJ.0000000000000006>

Lacour, B., Guyot-Goubin, A., Guissou, S., Bellec, S., Désandes, E., & Clavel, J. (2010). Incidence of childhood cancer in France: National Children Cancer Registries, 2000-2004. *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation (ECP)*, 19(3), 173-181. <https://doi.org/10.1097/cej.0b013e32833876c0>

Lavergne, V., Sabnis, A., Tupule, A., Davidson, P. R., Kline, C., Matthay, K., Nicolaidis, T., Goldsby, R., Braunstein, S., Fogh, S. E., Sneed, P. K., Menzel, P., Nakamura, A., Dubois, S. G., Haas-Kogan, D. A., & Nakamura, J. L. (2020). *Germline MUTYH Mutation in a Pediatric Cancer Survivor Developing a Secondary Malignancy.* <https://doi.org/10.1097/MPH.0000000000001668>

Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-Variant Association Analysis: Study Designs and Statistical Tests. *American Journal of Human Genetics*, 95(1),

5-23. <https://doi.org/10.1016/j.ajhg.2014.06.009>

Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., Christiani, D. C., Wurfel, M. M., & Lin, X. (2012). Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. *The American Journal of Human Genetics*, *91*(2), Article 2. <https://doi.org/10.1016/j.ajhg.2012.06.007>

Lee, S., Wu, M. C., & Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics (Oxford, England)*, *13*(4), 762-775. <https://doi.org/10.1093/biostatistics/kxs014>

Li, B., & Leal, S. M. (2008). Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *American Journal of Human Genetics*, *83*(3), 311-321. <https://doi.org/10.1016/j.ajhg.2008.06.024>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, *27*(12), 1739-1740. <https://doi.org/10.1093/bioinformatics/btr260>

Lin, X., Lee, S., Wu, M. C., Wang, C., Chen, H., Li, Z., & Lin, X. (2016). Test for Rare Variants by Environment Interactions in Sequencing Association Studies. *Biometrics*, *72*(1), 156-164. <https://doi.org/10.1111/biom.12368>

Lipshultz, S. E., Diamond, M. B., Franco, V. I., Aggarwal, S., Leger, K., Santos, M. V., Sallan,

- S. E., & Chow, E. J. (2014). Managing chemotherapy-related cardiotoxicity in survivors of childhood cancers. *Paediatric Drugs*, *16*(5), 373-389. <https://doi.org/10.1007/s40272-014-0085-1>
- Lopes, J. L., Chaudhry, S., Lopes, G. S., Levin, N. K., & Tainsky, M. A. (2019). FANCM, RAD1, CHEK1 and TP53I3 act as BRCA-like tumor suppressors and are mutated in hereditary ovarian cancer. *Cancer Genetics*, *235-236*, 57-64. <https://doi.org/10.1016/j.cancergen.2019.04.061>
- Ma, Y. P., van Leeuwen, F. E., Cooke, R., Broeks, A., Enciso-Mora, V., Olver, B., Lloyd, A., Broderick, P., Russell, N. S., Janus, C., Ashworth, A., Houlston, R. S., & Swerdlow, A. J. (2012). FGFR2 genotype and risk of radiation-associated breast cancer in Hodgkin lymphoma. *Blood*, *119*(4), 1029-1031. <https://doi.org/10.1182/blood-2011-10-383380>
- Madsen, B. E., & Browning, S. R. (2009). A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. *PLOS Genetics*, *5*(2), e1000384. <https://doi.org/10.1371/journal.pgen.1000384>
- Magnani, C., Pastore, G., Coebergh, J. W., Viscomi, S., Spix, C., & Steliarova-Foucher, E. (2006). Trends in survival after childhood cancer in Europe, 1978-1997: Report from the Automated Childhood Cancer Information System project (ACCIS). *European Journal of Cancer (Oxford, England: 1990)*, *42*(13), 1981-2005. <https://doi.org/10.1016/j.ejca.2006.05.006>
- Mandelbaum, J., Bhagat, G., Tang, H., Mo, T., Brahmachary, M., Shen, Q., Chadburn, A., Rajewsky, K., Tarakhovsky, A., Pasqualucci, L., & Dalla-Favera, R. (2010). BLIMP1

is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. *Cancer Cell*, 18(6), 568-579.
<https://doi.org/10.1016/j.ccr.2010.10.030>

Martínez-Bueno, M., & Alarcón-Riquelme, M. E. (2019). Exploring Impact of Rare Variation in Systemic Lupus Erythematosus by a Genome Wide Imputation Approach. *Frontiers in Immunology*, 10, 258.
<https://doi.org/10.3389/fimmu.2019.00258>

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1), 122. <https://doi.org/10.1186/s13059-016-0974-4>

Meadows, A. T. (2003). Pediatric cancer survivors: Past history and future challenges. *Current Problems in Cancer*, 27(3), 112-126. [https://doi.org/10.1016/s0147-0272\(03\)00025-4](https://doi.org/10.1016/s0147-0272(03)00025-4)

Mertens, A. C., Liu, Q., Neglia, J. P., Wasilewski, K., Leisenring, W., Armstrong, G. T., Robison, L. L., & Yasui, Y. (2008). Cause-Specific Late Mortality Among 5-Year Survivors of Childhood Cancer: The Childhood Cancer Survivor Study. *JNCI Journal of the National Cancer Institute*, 100(19), 1368-1379.
<https://doi.org/10.1093/jnci/djn310>

Mertens, A. C., Mitby, P. A., Radloff, G., Jones, I. M., Perentesis, J., Kiffmeyer, W. R., Neglia, J. P., Meadows, A., Potter, J. D., Friedman, D., Yasui, Y., Robison, L. L., & Davies, S. M. (2004). *XRCC1 and glutathione-S-transferase gene polymorphisms and susceptibility to radiotherapy-related malignancies in survivors of Hodgkin*

disease: A report from the Childhood Cancer Survivor Study.

<https://doi.org/10.1002/cncr.20520>

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). *Sustainable data analysis with Snakemake* (10:33). F1000Research. <https://doi.org/10.12688/f1000research.29032.2>

Morgenthaler, S., & Thilly, W. G. (2007). A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). *Mutation Research*, 615(1-2), 28-56. <https://doi.org/10.1016/j.mrfmmm.2006.09.003>

Morton, L. M., Dores, G. M., Curtis, R. E., Lynch, C. F., Stovall, M., Hall, P., Gilbert, E. S., Hodgson, D. C., Storm, H. H., Johannesen, T. B., Smith, S. A., Weathers, R. E., Andersson, M., Fossa, S. D., Hauptmann, M., Holowaty, E. J., Joensuu, H., Kaijser, M., Kleinerman, R. A., ... van Leeuwen, F. E. (2013). Stomach Cancer Risk After Treatment for Hodgkin Lymphoma. *Journal of Clinical Oncology*, 31(27), 3369-3377. <https://doi.org/10.1200/JCO.2013.50.6832>

Morton, L. M., Karyadi, D. M., Hartley, S. W., Frone, M. N., Sampson, J. N., Howell, R. M., Neglia, J. P., Arnold, M. A., Hicks, B. D., Jones, K., Zhu, B., Dagnall, C. L., Karlins, E., Yeager, M. S., Leisenring, W. M., Yasui, Y., Turcotte, L. M., Smith, S. A., Weathers, R. E., ... Chanock, S. J. (2020). Subsequent Neoplasm Risk Associated With Rare Variants in DNA Damage Response and Clinical Radiation Sensitivity Syndrome Genes in the Childhood Cancer Survivor Study. *JCO Precision Oncology*, 4,

PO.20.00141. <https://doi.org/10.1200/PO.20.00141>

Morton, L. M., Sampson, J. N., Armstrong, G. T., Chen, T.-H., Hudson, M. M., Karlins, E., Dagnall, C. L., Li, S. A., Wilson, C. L., Srivastava, D. K., Liu, W., Kang, G., Oeffinger, K. C., Henderson, T. O., Moskowitz, C. S., Gibson, T. M., Merino, D. M., Wong, J. R., Hammond, S., ... Robison, L. L. (2017). Genome-Wide Association Study to Identify Susceptibility Loci That Modify Radiation-Related Risk for Breast Cancer After Childhood Cancer. *JNCI: Journal of the National Cancer Institute*, *109*(dix058). <https://doi.org/10.1093/jnci/dix058>

Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S. M., Roeder, K., & Daly, M. J. (2011). Testing for an Unusual Distribution of Rare Variants. *PLOS Genetics*, *7*(3), e1001322. <https://doi.org/10.1371/journal.pgen.1001322>

Neidhardt, G., Hauke, J., Ramser, J., Groß, E., Gehrig, A., Müller, C. R., Kahlert, A.-K., Hackmann, K., Honisch, E., Niederacher, D., Heilmann-Heimbach, S., Franke, A., Lieb, W., Thiele, H., Altmüller, J., Nürnberg, P., Klaschik, K., Ernst, C., Ditsch, N., ... Hahnen, E. (2017). Association Between Loss-of-Function Mutations Within the FANCM Gene and Early-Onset Familial Breast Cancer. *JAMA Oncology*, *3*(9), 1245-1248. <https://doi.org/10.1001/jamaoncol.2016.5592>

Nichols, K. E., Heath, J. A., Friedman, D., Biegel, J. A., Ganguly, A., Mauch, P., & Diller, L. (2003). *TP53, BRCA1, and BRCA2 tumor suppressor genes are not commonly mutated in survivors of Hodgkin's disease with second primary neoplasms.* <https://doi.org/10.1200/JCO.2003.12.042>

Nottage, K., Lanctot, J., Li, Z., Neglia, J. P., Bhatia, S., Hammond, S., Leisenring, W., Meadows, A., Srivastava, D., Robison, L. L., & Armstrong, G. T. (2011). Long-term risk for subsequent leukemia after treatment for childhood cancer: A report from the Childhood Cancer Survivor Study. *Blood*, *117*(23), 6315-6318. <https://doi.org/10.1182/blood-2011-02-335158>

Oeffinger Kevin C., Mertens Ann C., Sklar Charles A., Kawashima Toana, Hudson Melissa M., Meadows Anna T., Friedman Debra L., Marina Neyssa, Hobbie Wendy, Kadan-Lottick Nina S., Schwartz Cindy L., Leisenring Wendy, & Robison Leslie L. (2006). Chronic Health Conditions in Adult Survivors of Childhood Cancer. *New England Journal of Medicine*, *355*(15), 1572-1582. <https://doi.org/10.1056/NEJMsa060185>

Opstal-van Winden, A. W. J., de Haan, H. G., Hauptmann, M., Schmidt, M. K., Broeks, A., Russell, N. S., Janus, C. P. M., Krol, A. D. G., van der Baan, F. H., De Bruin, M. L., van Eggermond, A. M., Dennis, J., Anton-Culver, H., Haiman, C. A., Sawyer, E. J., Cox, A., Devilee, P., Hoening, M. J., Peto, J., ... van Leeuwen, F. E. (2019). Genetic susceptibility to radiation-induced breast cancer after Hodgkin lymphoma. *Blood*, *133*(10), 1130-1139. <https://doi.org/10.1182/blood-2018-07-862607>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical Research Ed.)*, *372*, n71.

<https://doi.org/10.1136/bmj.n71>

Persyn, E., Redon, R., Bellanger, L., & Dina, C. (2018). The impact of a fine-scale population stratification on rare variant association test results. *PLoS ONE*, 13(12). <https://doi.org/10.1371/journal.pone.0207677>

Peterlongo, P., Catucci, I., Colombo, M., Caleca, L., Mucaki, E., Bogliolo, M., Marin, M., Damiola, F., Bernard, L., Pensotti, V., Volorio, S., Dall'Olio, V., Meindl, A., Bartram, C., Sutter, C., Surowy, H., Sornin, V., Dondon, M.-G., Eon-Marchais, S., ... Radice, P. (2015). FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Human Molecular Genetics*, 24(18), 5345-5355. <https://doi.org/10.1093/hmg/ddv251>

Purdy, J. A. (2008). Dose to normal tissues outside the radiation therapy patient's treated volume: A review of different radiation therapy techniques. *Health Physics*, 95(5), 666-676. <https://doi.org/10.1097/01.HP.0000326342.47348.06>

Qin, N., Wang, Z., Liu, Q., Song, N., Wilson, C. L., Ehrhardt, M. J., Shelton, K., Easton, J., Mulder, H., Kennetz, D., Edmonson, M. N., Rusch, M. C., Downing, J. R., Hudson, M. M., Nichols, K. E., Zhang, J., Robison, L. L., & Yasui, Y. (2020). Pathogenic Germline Mutations in DNA Repair Genes in Combination With Cancer Treatment Exposures and Risk of Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer. *Journal of Clinical Oncology*, 38(24), 2728-2740. <https://doi.org/10.1200/JCO.19.02760>

Rao, D., Kimler, B. F., Nothnick, W. B., Davis, M. K., Fan, F., & Tawfik, O. (2015).

TRANSGELIN: A POTENTIALLY USEFUL DIAGNOSTIC MARKER DIFFERENTIALLY EXPRESSED IN TRIPLE-NEGATIVE AND NON-TRIPLE NEGATIVE BREAST CANCERS. *Human pathology*, 46(6), 876-883.

<https://doi.org/10.1016/j.humpath.2015.02.015>

Reulen, R. C., Frobisher, C., Winter, D. L., Kelly, J., Lancashire, E. R., Stiller, C. A., Pritchard-Jones, K., Jenkinson, H. C., Hawkins, M. M., & British Childhood Cancer Survivor Study Steering Group, for the. (2011). Long-term Risks of Subsequent Primary Neoplasms Among Survivors of Childhood Cancer. *JAMA*, 305(22), 2311-2319.

<https://doi.org/10.1001/jama.2011.747>

Richard, M. A., Lupo, P. J., Morton, L. M., Yasui, Y. A., Sapkota, Y. A., Arnold, M. A., Aubert, G., Neglia, J. P., Turcotte, L. M., Leisenring, W. M., Sampson, J. N., Chanock, S. J., Hudson, M. M., Armstrong, G. T., Robison, L. L., Bhatia, S., & Gramatges, M. M. (2020). Genetic variation in POT1 and risk of thyroid subsequent malignant neoplasm: A report from the Childhood Cancer Survivor Study. *PLOS ONE*, 15(2), e0228887.

<https://doi.org/10.1371/journal.pone.0228887>

Robison, L. L., & Hudson, M. M. (2014). Survivors of Childhood/Adolescent Cancer: Life-long Risks and Responsibilities. *Nature reviews. Cancer*, 14(1), 61-70.

<https://doi.org/10.1038/nrc3634>

Rökman, A., Ikonen, T., Seppälä, E. H., Nupponen, N., Autio, V., Mononen, N., Bailey-Wilson, J., Trent, J., Carpten, J., Matikainen, M. P., Koivisto, P. A., Tammela, T. L. J., Kallioniemi, O.-P., & Schleutker, J. (2002). Germline alterations of the RNASEL gene, a candidate HPC1 gene at 1q25, in patients and families with prostate

- cancer. *American Journal of Human Genetics*, 70(5), 1299-1304.
<https://doi.org/10.1086/340450>
- Saint Pierre, A., & Génin, E. (2014). How important are rare variants in common disease? *Briefings in Functional Genomics*, 13(5), 353-361.
<https://doi.org/10.1093/bfgp/elu025>
- Sapkota, Y., Turcotte, L. M., Ehrhardt, M. J., Howell, R. M., Arnold, M. A., Wilson, C. L., Leisenring, W., Wang, Z., Sampson, J., Dagnall, C. L., Karlins, E., Li, S. A., Hicks, B. D., Weathers, R., Smith, S. A., Shelton, K., Liu, Q., Tucker, M. A., Chanock, S. J., ... Yasui, Y. (2019). Genome-Wide Association Study in Irradiated Childhood Cancer Survivors Identifies HTR2A for Subsequent Basal Cell Carcinoma. *The Journal of Investigative Dermatology*, 139(9), 2042-2045.e8.
<https://doi.org/10.1016/j.jid.2019.02.029>
- Schaapveld, M., Aleman, B. M. P., van Eggermond, A. M., Janus, C. P. M., Krol, A. D. G., van der Maazen, R. W. M., Roesink, J., Raemaekers, J. M. M., de Boer, J. P., Zijlstra, J. M., van Imhoff, G. W., Petersen, E. J., Poortmans, P. M. P., Beijert, M., Lybeert, M. L., Mulder, I., Visser, O., Louwman, M. W. J., Krul, I. M., ... van Leeuwen, F. E. (2015). Second Cancer Risk Up to 40 Years after Treatment for Hodgkin's Lymphoma. *The New England Journal of Medicine*, 373(26), 2499-2511.
<https://doi.org/10.1056/NEJMoa1505949>
- Schork, N. J., Murray, S. S., Frazer, K. A., & Topol, E. J. (2009). Common vs. Rare Allele Hypotheses for Complex Diseases. *Current opinion in genetics & development*, 19(3), 212-219. <https://doi.org/10.1016/j.gde.2009.04.010>

Schubert, S., van Luttikhuizen, J. L., Auber, B., Schmidt, G., Hofmann, W., Penkert, J., Davenport, C. F., Hille-Betz, U., Wendeburg, L., Bublitz, J., Tauscher, M., Hackmann, K., Schröck, E., Scholz, C., Wallaschek, H., Schlegelberger, B., Illig, T., & Steinemann, D. (2019). The identification of pathogenic variants in BRCA1/2 negative, high risk, hereditary breast and/or ovarian cancer patients: High frequency of FANCM pathogenic variants. *International Journal of Cancer*, *144*(11), 2683-2694. <https://doi.org/10.1002/ijc.31992>

Servitzoglou, M., De Vathaire, F., Oberlin, O., Patte, C., & Thomas-Teinturier, C. (2015). Dose-Effect Relationship of Alkylating Agents on Testicular Function in Male Survivors of Childhood Lymphoma. *Pediatric Hematology and Oncology*, *32*(8), 613-623. <https://doi.org/10.3109/08880018.2015.1085933>

Shen, C. J., & Terezakis, S. A. (2021). The Evolving Role of Radiotherapy for Pediatric Cancers With Advancements in Molecular Tumor Characterization and Targeted Therapies. *Frontiers in Oncology*, *11*. <https://doi.org/10.3389/fonc.2021.679701>

Sherborne, A. L., Davidson, P. R., Yu, K., Nakamura, A. O., Rashid, M., & Nakamura, J. L. (2015). *Mutational Analysis of Ionizing Radiation Induced Neoplasms*. <https://doi.org/10.1016/j.celrep.2015.08.015>

Sohani, Z. N., Meyre, D., de Souza, R. J., Joseph, P. G., Gandhi, M., Dennis, B. B., Norman, G., & Anand, S. S. (2015). Assessing the quality of published genetic association studies in meta-analyses: The quality of genetic studies (Q-Genie) tool. *BMC Genetics*, *16*, 50. <https://doi.org/10.1186/s12863-015-0211-2>

Stanulla, M., Schaeffeler, E., Möricke, A., Coulthard, S. A., Cario, G., Schrauder, A.,

- Kaatsch, P., Dördelmann, M., Welte, K., Zimmermann, M., Reiter, A., Eichelbaum, M., Riehm, H., Schrappe, M., & Schwab, M. (2009). Thiopurine methyltransferase genetics is not a major risk factor for secondary malignant neoplasms after treatment of childhood acute lymphoblastic leukemia on Berlin-Frankfurt-Münster protocols. *Blood*, *114*(7), 1314-1318. <https://doi.org/10.1182/blood-2008-12-193250>
- Steliarova-Foucher, E., Colombet, M., Ries, L. A. G., Moreno, F., Dolya, A., Bray, F., Hesselning, P., Shin, H. Y., & Stiller, C. A. (2017). International incidence of childhood cancer, 2001–10: A population-based registry study. *The Lancet. Oncology*, *18*(6), 719-731. [https://doi.org/10.1016/S1470-2045\(17\)30186-9](https://doi.org/10.1016/S1470-2045(17)30186-9)
- Stiller, C. A., Allen, M. B., & Eatock, E. M. (1995). Childhood cancer in Britain: The National Registry of Childhood Tumours and incidence rates 1978-1987. *European Journal of Cancer (Oxford, England: 1990)*, *31A*(12), 2028-2034. [https://doi.org/10.1016/0959-8049\(95\)00428-9](https://doi.org/10.1016/0959-8049(95)00428-9)
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545-15550. <https://doi.org/10.1073/pnas.0506580102>
- Taylor, A. J., Frobisher, C., Ellison, D. W., Reulen, R. C., Winter, D. L., Taylor, R. E., Stiller, C. A., Lancashire, E. R., Tudor, E. C. G., Baggott, C., May, S., & Hawkins, M. M. (2009). *Survival after second primary neoplasms of the brain or spinal cord in*

survivors of childhood cancer: Results from the British childhood cancer survivor study. <https://doi.org/10.1200/JCO.2009.22.4386>

Teepen, J. C., van Leeuwen, F. E., Tissing, W. J., van Dulmen-den Broeder, E., van den Heuvel-Eibrink, M. M., van der Pal, H. J., Loonen, J. J., Bresters, D., Versluys, B., Neggers, S. J. C. M. M., Jaspers, M. W. M., Hauptmann, M., van der Heiden-van der Loo, M., Visser, O., Kremer, L. C. M., Ronckers, C. M., & DCOG LATER Study Group. (2017). Long-Term Risk of Subsequent Malignant Neoplasms After Treatment of Childhood Cancer in the DCOG LATER Study Cohort: Role of Chemotherapy. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 35(20), 2288-2298. <https://doi.org/10.1200/JCO.2016.71.6902>

Thirman, M. J., & Larson, R. A. (1996). Therapy-related myeloid leukemia. *Hematology/Oncology Clinics of North America*, 10(2), 293-320. [https://doi.org/10.1016/s0889-8588\(05\)70340-3](https://doi.org/10.1016/s0889-8588(05)70340-3)

Travis, L. B. (2006). *The epidemiology of second primary cancers.* <https://doi.org/10.1158/1055-9965.EPI-06-0414>

Travis, L. B., Gospodarowicz, M., Curtis, R. E., Clarke, E. A., Andersson, M., Glimelius, B., Joensuu, T., Lynch, C. F., van Leeuwen, F. E., Holowaty, E., Storm, H., Glimelius, I., Pukkala, E., Stovall, M., Fraumeni, J. F., Boice, J. D., & Gilbert, E. (2002). Lung cancer following chemotherapy and radiotherapy for Hodgkin's disease. *Journal of the National Cancer Institute*, 94(3), 182-192. <https://doi.org/10.1093/jnci/94.3.182>

- Tukenova, M., Diallo, I., Anderson, H., Hawkins, M., Garwicz, S., Sankila, R., El Fayeche, C., Winter, D., Rubino, C., Adjadj, E., Haddy, N., Oberlin, O., Moller, T., Langmark, F., Tryggvadottir, L., Pacquement, H., Svahn-Tapper, G., & de Vathaire, F. (2012). Second malignant neoplasms in digestive organs after childhood cancer: A cohort-nested case-control study. *International Journal of Radiation Oncology, Biology, Physics*, 82(3), e383-390. <https://doi.org/10.1016/j.ijrobp.2011.05.069>
- Turcotte, L. M., Liu, Q., Yasui, Y., Arnold, M. A., Hammond, S., Howell, R. M., Smith, S. A., Weathers, R. E., Henderson, T. O., Gibson, T. M., Leisenring, W., Armstrong, G. T., Robison, L. L., & Neglia, J. P. (2017). Temporal Trends in Treatment and Subsequent Neoplasm Risk Among 5-Year Survivors of Childhood Cancer, 1970-2015. *JAMA*, 317(8), 814-824. <https://doi.org/10.1001/jama.2017.0693>
- Turcotte, L. M., Liu, Q., Yasui, Y., Henderson, T. O., Gibson, T. M., Leisenring, W., Arnold, M. A., Howell, R. M., Green, D. M., Armstrong, G. T., Robison, L. L., & Neglia, J. P. (2019). Chemotherapy and Risk of Subsequent Malignant Neoplasms in the Childhood Cancer Survivor Study Cohort. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 37(34), 3310-3319. <https://doi.org/10.1200/JCO.19.00129>
- Turcotte, L. M., Whitton, J. A., Friedman, D. L., Hammond, S., Armstrong, G. T., Leisenring, W., Robison, L. L., & Neglia, J. P. (2015). Risk of Subsequent Neoplasms During the Fifth and Sixth Decades of Life in the Childhood Cancer Survivor Study Cohort. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 33(31), 3568-3575. <https://doi.org/10.1200/JCO.2015.60.9487>

- Uckunkaya, O., Nöllke, P., Hallmen, E., Becker, C., Stegmaier, S., Sparber-Sauer, M., Vokuhl, C., Koscielniak, E., & Hettmer, S. (2020). *Patterns of Prior and Subsequent Neoplasms in Children and Adolescents with Soft Tissue Sarcomas*. <https://doi.org/10.1097/MPH.0000000000001837>
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 1-21. <https://doi.org/10.1038/s43586-021-00056-9>
- Varan, A., & Kebudi, R. (2011). *Secondary malignant neoplasms after childhood cancer*. <https://doi.org/10.3109/08880018.2011.553879>
- Vassal, G., Schrappe, M., Pritchard-Jones, K., Arnold, F., Basset, L., Biondi, A., Bode, G., Eggert, A., Hjorth, L., Kamerić, L., Kamerić, N., Karner, S., Kearns, P., Kienesberger, A., Kowalczyk, J., Lack, P., Perilongo, G., Sullivan, R., Tsirou, A., ... Ladenstein, R. (2016). The SIOPE strategic plan: A European cancer plan for children and adolescents. *Journal of Cancer Policy*, 8, 17-32. <https://doi.org/10.1016/j.jcpo.2016.03.007>
- Veiga, L. H. S., Bhatti, P., Ronckers, C. M., Sigurdson, A. J., Stovall, M., Smith, S. A., Weathers, R., Leisenring, W., Mertens, A. C., Hammond, S., Neglia, J. P., Meadows, A. T., Donaldson, S. S., Sklar, C. A., Friedman, D. L., Robison, L. L., & Inskip, P. D. (2012). Chemotherapy and thyroid cancer risk: A report from the Childhood Cancer Survivor Study. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the*

American Society of Preventive Oncology, 21(1), 92-101.

<https://doi.org/10.1158/1055-9965.EPI-11-0576>

Veiga, L. H. S., Holmberg, E., Anderson, H., Pottern, L., Sadetzki, S., Adams, M. J., Sakata, R., Schneider, A. B., Inskip, P., Bhatti, P., Johansson, R., Neta, G., Shore, R., de Vathaire, F., Damber, L., Kleinerman, R., Hawkins, M. M., Tucker, M., Lundell, M., & Lubin, J. H. (2016). Thyroid Cancer after Childhood Exposure to External Radiation: An Updated Pooled Analysis of 12 Studies. *Radiation Research*, 185(5), 473-484. <https://doi.org/10.1667/RR14213.1>

Veres, C., Allodji, R. S., Llanas, D., Vu Bezin, J., Chavaudra, J., Mège, J. P., Lefkopoulos, D., Quiniou, E., Deutsh, E., de Vathaire, F., & Diallo, I. (2014). Retrospective reconstructions of active bone marrow dose-volume histograms. *International Journal of Radiation Oncology, Biology, Physics*, 90(5), 1216-1224. <https://doi.org/10.1016/j.ijrobp.2014.08.335>

Vodusek, A. L., Goricar, K., Gazic, B., Dolzan, V., & Jazbec, J. (2016a). Antioxidant defence-related genetic variants are not associated with higher risk of secondary thyroid cancer after treatment of malignancy in childhood or adolescence. *Radiology and Oncology*, 50(1), 80-86. <https://doi.org/10.1515/raon-2015-0026>

Vodusek, A. L., Goricar, K., Gazic, B., Dolzan, V., & Jazbec, J. (2016b). Antioxidant defence-related genetic variants are not associated with higher risk of secondary thyroid cancer after treatment of malignancy in childhood or adolescence. *Radiology and Oncology*, 50(1), 80. <https://doi.org/10.1515/RAON-2015-0026>

Vogiatzoglou, A. P., Moretto, F., Makkou, M., Papamatheakis, J., & Kretsovali, A. (2021).

Promyelocytic leukemia protein (PML) and stem cells: From cancer to pluripotency. *The International Journal of Developmental Biology*, 66(1-2-3), Article 1-2-3. <https://doi.org/10.1387/ijdb.210154av>

Wang, X., Sun, C.-L., Hageman, L., Smith, K., Singh, P., Desai, S., Hawkins, D. S., Hudson, M. M., Mascarenhas, L., Neglia, J. P., Oeffinger, K. C., Ritchey, A. K., Robison, L. L., Villaluna, D., Landier, W., & Bhatia, S. (2017). Clinical and Genetic Risk Prediction of Subsequent CNS Tumors in Survivors of Childhood Cancer: A Report From the COG ALTE03N1 Study. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 35(32), 3688-3696. <https://doi.org/10.1200/JCO.2017.74.7444>

Wang, Z., Liu, Q., Wilson, C. L., Easton, J., Mulder, H., Chang, T.-C., Rusch, M. C., Edmonson, M. N., Rice, S. V., Ehrhardt, M. J., Howell, R. M., Kesserwan, C. A., Wu, G., Nichols, K. E., Downing, J. R., Hudson, M. M., Zhang, J., Yasui, Y., & Robison, L. L. (2018). Polygenic Determinants for Subsequent Breast Cancer Risk in Survivors of Childhood Cancer: The St Jude Lifetime Cohort Study (SJLIFE). *Clinical cancer research: an official journal of the American Association for Cancer Research*, 24(24), 6230-6235. <https://doi.org/10.1158/1078-0432.CCR-18-1775>

Wang, Z., Wilson, C. L., Easton, J., Thrasher, A., Mulder, H., Liu, Q., Hedges, D. J., Wang, S., Rusch, M. C., Edmonson, M. N., Levy, S., Lanctot, J. Q., Caron, E., Shelton, K., Currie, K., Lear, M., Patel, A., Rosencrance, C., Shao, Y., ... Zhang, J. (2018). Genetic Risk for Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer. *Journal of Clinical Oncology: Official Journal of the American Society of*

- Clinical Oncology*, 36(20), 2078-2087. <https://doi.org/10.1200/JCO.2018.77.8589>
- Wells, G., Wells, G., Shea, B., Shea, B., O'Connell, D., Peterson, J., Welch, Losos, M., Tugwell, P., Ga, S. W., Zello, G., & Petersen, J. (2014). *The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses*. [https://www.semanticscholar.org/paper/The-Newcastle-Ottawa-Scale-\(NOS\)-for-Assessing-the-Wells-Wells/c293fb316b6176154c3fdbb8340a107d9c8c82bf](https://www.semanticscholar.org/paper/The-Newcastle-Ottawa-Scale-(NOS)-for-Assessing-the-Wells-Wells/c293fb316b6176154c3fdbb8340a107d9c8c82bf)
- Wingett, S. W., & Andrews, S. (2018). FastQ Screen : A tool for multi-genome mapping and quality control. *F1000Research*, 7, 1338. <https://doi.org/10.12688/f1000research.15931.2>
- Woo, M. H., Shuster, J. J., Chen, C., Bash, R. O., Behm, F. G., Camitta, B., Felix, C. A., Kamen, B. A., Pui, C.-H., Raimondi, S. C., Winick, N. J., Amylon, M. D., & Relling, M. V. (2000). Glutathione S-transferase genotypes in children who develop treatment-related acute myeloid malignancies. *Leukemia*, 14(2), 232-237. <https://doi.org/10.1038/sj.leu.2401660>
- Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *American Journal of Human Genetics*, 89(1), 82-93. <https://doi.org/10.1016/j.ajhg.2011.05.029>
- Xuan, D., Li, G., Cai, Q., Deming-Halverson, S., Shrubsole, M. J., Shu, X.-O., Kelley, M. C., Zheng, W., & Long, J. (2013). APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis*, 34(10), 2240-2243. <https://doi.org/10.1093/carcin/bgt185>

- Ying, S., Zhang, X., Sarkis, P. T. N., Xu, R., & Yu, X. (2007). Cell-specific regulation of APOBEC3F by interferons. *Acta Biochimica Et Biophysica Sinica*, 39(4), 297-304. <https://doi.org/10.1111/j.1745-7270.2007.00275.x>
- Zahnreich, S., & Schmidberger, H. (2021). Childhood Cancer: Occurrence, Treatment and Risk of Second Primary Malignancies. *Cancers*, 13(11), 2607. <https://doi.org/10.3390/cancers13112607>
- Zhuo, C., & al. (2015). *Screening for Novel Germline Rare Mutations Associated with Aggressive Prostate Cancer*. Wake Forest University. <https://apps.dtic.mil/sti/citations/AD1002994>
- Zuk, O., Schaffner, S. F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M. J., Neale, B. M., Sunyaev, S. R., & Lander, E. S. (2014). Searching for missing heritability: Designing rare variant association studies. *Proceedings of the National Academy of Sciences*, 111(4), E455-E464. <https://doi.org/10.1073/pnas.1322563111>

ANNEXES

Annexe A – 1. Stratégie de recherche PICOS- Pubmed

Population	Survivors	"survivorship"[MeSH Terms] OR "young ages"[All Fields] OR "survivorship"[All Fields] OR "survivors"[MeSH Terms] OR "cancer survivors"[MeSH Terms] OR "survival"[All Fields] OR "survival"[MeSH Terms] OR "surviving"[All Fields])
	AND	
	Childhood	"young adult"[All Fields] OR "young ages"[All Fields] OR "childhood"[All Fields] OR "child"[MeSH Terms] OR "child"[All Fields] OR "children"[All Fields] OR "childrens"[All Fields] OR "adolescent"[MeSH Terms] OR "adolescent"[All Fields] OR "infant"[MeSH Terms] OR "infant"[All Fields]
	Cancer	"rhabdomyosarcoma"[MeSH Terms] OR "brain neoplasms"[MeSH Terms] OR "central nervous system neoplasms"[MeSH Terms] OR "wilms tumor"[MeSH Terms] OR "neuroblastoma"[MeSH Terms] OR "retinoblastoma"[MeSH Terms] OR "osteosarcoma"[MeSH Terms] OR "sarcoma, ewing"[MeSH Terms] OR "bone neoplasms"[MeSH Terms] OR "lymphoma, non hodgkin"[MeSH Terms] OR "hodgkin disease"[MeSH Terms] OR "lymphoma"[MeSH Terms] OR "leukemia"[MeSH Terms] OR "neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "malignancies"[All Fields] OR "malignancy"[All Fields] OR "tumor"[All Fields] OR "carcinoma"[MeSH Terms] OR "cancer"[All Fields] OR "sarcoma"[MeSH Terms]
AND		
Intervention	Chemotherapy	"drug therapy"[MeSH Terms] OR "chemotherapy"[All Fields] OR "drug therapy"[MeSH Terms] OR "antineoplastic agents"[Pharmacological Action] OR "antineoplastic agents"[MeSH Terms] OR "antineoplastic agents"[All Fields] OR "alkylating agents"[Pharmacological Action] OR "alkylating agents"[MeSH Terms] OR "antimetabolites"[Pharmacological Action] OR "antimetabolites"[MeSH Terms] OR "anthracyclines"[MeSH Terms] OR "anthracycline"[All Fields] OR "vinca alkaloids"[MeSH Terms] OR "anticarcinogenic agents"[Pharmacological Action] OR "anticarcinogenic agents"[MeSH Terms] OR "vincristine"[All Fields] OR "anticancer drug"[All Fields] OR "anticancer drugs"[All Fields])
	OR	
	Radiotherapy	"radiotherapy"[MeSH Terms] OR "radiotherapy"[All Fields] OR "radiation therapy"[All Fields] OR "radiation treatment"[All Fields] OR "x-ray therapy"[MeSH Terms] OR "dose fractionation, radiation"[MeSH Terms] OR "dose response relationship, radiation"[MeSH Terms] OR "radiation dosage"[MeSH Terms] OR "radiation oncology"[MeSH Terms] OR "radiation injuries"[MeSH Terms] OR "radiation effects"[MeSH Terms] OR "radiation exposure"[MeSH Terms]
	OR	
Cancer treatment	"chemoradiotherapy"[MeSH Terms] OR "radiochemotherapy"[All Fields] OR "medical oncology"[MeSH Terms] OR "cancer treatment"[All Fields]	
AND		

Outcome	Second cancer	"neoplasms, second primary"[MeSH Terms] OR "second primary neoplasms"[All Fields] OR "second primary cancer"[All Fields] OR "second primary cancers"[All Fields] OR "Second Primary Malignancies"[All Fields] OR "Second Primary Tumor"[All Fields] OR "Subsequent Malignant Neoplasm"[All Fields] OR "Second Cancer"[All Fields] OR "Second Neoplasm"[All Fields] OR "Second Neoplasms"[All Fields] OR "Second Tumor"[All Fields] OR "Second Malignancy"[All Fields] OR "Second Malignancies"[All Fields] OR "neoplasms, radiation induced"[MeSH Terms] OR "radiation-induced neoplasms"[All Fields] OR "Radiation Induced Cancer"[All Fields] OR "Radiation Induced Tumor"[All Fields] OR "Radiation Induced Malignancy"[All Fields] OR "radiation-associated"[All Fields] OR "Radiation Induced Malignancies"[All Fields] OR "therapy associated neoplasm"[All Fields] OR "therapy related neoplasm"[All Fields] OR "treatment related neoplasm"[All Fields] OR "subsequent primary neoplasms"[All Fields] OR "Subsequent neoplasm"[All Fields] OR "Subsequent neoplasms"[All Fields] OR "Subsequent malignant neoplasms"[All Fields] OR "subsequent breast cancer*"[All Fields] OR "secondary breast cancer*"[All Fields] OR "second breast cancer*"[All Fields] OR "subsequent thyroid cancer*"[All Fields] OR "secondary thyroid cancer*"[All Fields] OR "second thyroid cancer*"[All Fields] OR "SMN"[All Fields]
	AND	
	Genetics variants	"variant"[All Fields] OR "variants"[All Fields] OR "genetic"[All Fields] OR "genetics"[All Fields] OR "genetics"[MeSH Terms] OR "sequence analysis, dna"[MeSH Terms] OR "whole exome sequencing"[MeSH Terms] OR "whole genome sequencing"[MeSH Terms] OR "genome-wide association study"[MeSH Terms] OR "genome-wide association study"[All Fields] OR "candidate gene association study"[All Fields] OR "genetic association studies"[MeSH Terms] OR "polymorphism, single nucleotide"[MeSH Terms] OR "single nucleotide polymorphism"[All Fields] OR "Structural Variation"[All Fields] OR "dna copy number variations"[MeSH Terms] OR "germ-line mutation"[MeSH Terms] OR "Genomic"[All Fields] OR "genomics"[MeSH Terms] OR "mutation"[MeSH Terms] OR "mutation"[All Fields] OR "mutations"[All Fields] OR "genetic predisposition to disease"[MeSH Terms] OR "genotype"[MeSH Terms] OR "genotype"[All Fields] OR "polymorphism, genetic"[MeSH Terms] OR "genes"[MeSH Terms] OR "genes"[All Fields] OR "gene"[All Fields] OR "sequencing"[All Fields] OR "whole exome sequencing/methods"[MeSH Major Topic] OR "biomarkers"[MeSH Terms] OR "biomarkers"[All Fields] OR "biomarker"[All Fields]
AND		
Study design	"cohort studies"[MeSH Terms] OR "cohort study"[All Fields] OR "retrospective studies"[MeSH Terms] OR "retrospective study"[All Fields] OR "prospective studies"[MeSH Terms] OR "prospective study"[All Fields] OR "follow up studies"[MeSH Terms] OR "follow up study"[All Fields] OR "longitudinal studies"[MeSH Terms] OR "longitudinal study"[All Fields] OR "meta analysis"[Publication Type] OR "meta analysis as topic"[MeSH Terms] OR "review"[Publication Type] OR "review literature as topic"[MeSH Terms] OR "review"[All Fields] OR "systematic review"[Publication Type] OR "systematic reviews as topic"[MeSH Terms] OR "systematic review"[All Fields] OR "case control studies"[MeSH Terms] OR "case control study"[All Fields] OR "cohort"[All Fields] OR "observational study"[Publication Type] OR "observational studies as topic"[MeSH Terms] OR "clinical trial"[Publication Type] OR "clinical trials as topic"[MeSH Terms] OR "genetic association studies"[MeSH Terms]	
NOT		
"nuclear weapons"[MeSH Terms] OR "nuclear weapons"[Title/Abstract] OR "atomic bomb"[Title/Abstract]		

Annexe A – 2. Stratégie de recherche PICOS- SCOPUS

Population	Childhood Survivors	TITLE-ABS-KEY (("first cancer in childhood" OR survivor OR survivors OR survivorship) AND ("young adult" OR "young ages" OR childhood OR child OR adolescent OR infant OR children))
	AND	
	Cancer	TITLE-ABS-KEY (cancer OR neoplasms OR malignant OR malignancy OR tumor OR carcinoma OR leukemia OR retinoblastoma OR sarcoma OR osteosarcoma OR "wilms tumor" OR nephroblastoma OR neuroblastoma OR lymphoma OR "hodgkin disease" OR astrocytoma OR leukaemia)
AND		
Intervention	Cancer Treatment	ALL (radiotherapy OR radiations OR radiation OR " x-ray therapy" OR chemotherapy OR alkylating OR anthracyclines OR anthracycline OR antineoplastic OR "anticancer drugs" OR "vinca alkaloid" OR antimetabolites OR Topoisomerase OR "Tyrosine kinase" OR epipodophyllotoxin OR antibiotics OR chemoradiotherapy)
AND		
Outcome	Second cancer	TITLE-ABS-KEY("second neoplasm" OR "second cancer" OR "second malignant neoplasm" OR "second tumor" OR "second malignancy" OR "second malignant tumor" OR "secondary cancer" OR "secondary neoplasm" OR "secondary cancers" OR "secondary tumor" OR "secondary malignancy" OR "second breast cancer" OR "subsequent breast cancer" OR "subsequent malignant neoplasm" OR "subsequent cancer" OR "subsequent primary neoplasms" OR "subsequent neoplasm" OR "subsequent malignancy" OR "second primary cancer" OR "second primary neoplasm" OR "second primary malignancy" OR "second primary tumor" OR "second primary sarcoma" OR "radiation induced neoplasm")
	AND	
	Genetics variants	TITLE-ABS-KEY(biomarkers OR genetic OR genetics OR exome OR genome OR transcriptome OR sequencing OR genomic OR genotype OR polymorphism OR snps OR mutations OR genes OR GWAS OR haplotype OR "copy number variation" OR "pathway analysis" OR " germline predisposition" OR variant OR genomic)
AND		
Study design	ALL (cohort OR "cohort study" OR "observational study" OR "case-control study" OR "prospective study" OR review OR meta-analysis OR "Retrospective Study" OR "cohort analysis" OR "longitudinal study")	
AND		
LIMIT-TO (DOCTYPE,"ar")OR LIMIT-TO (DOCTYPE,"re") OR LIMIT-TO (DOCTYPE,"ch") OR LIMIT-TO (DOCTYPE,"cp")		
AND		
LIMIT-TO (LANGUAGE,"English")		
AND		
EXCLUDE (EXACTKEYWORD,"A-bomb Survivors") OR EXCLUDE (EXACTKEYWORD,"Nuclear Accident") OR EXCLUDE (EXACTKEYWORD,"Nuclear Power Plant") OR EXCLUDE (EXACTKEYWORD,"Nuclear Power Plants") OR EXCLUDE (EXACTKEYWORD,"Nuclear Warfare") OR EXCLUDE (EXACTKEYWORD,"Nuclear Weapons") OR EXCLUDE (EXACTKEYWORD,"Nuclear Worker")		

Annexe A – 3. Stratégie de recherche PICOS- WOS

Population	Childhood Survivors	ALL = ((surviv*) AND ("young adult" OR "young ages" OR child* OR adolescent OR infant))
	Cancer	ALL = (cancer * OR neoplasm* OR malignant OR malignanc* OR tumo\$r* OR carcinoma OR leuk\$emia OR retinoblastoma OR sarcoma OR osteosarcoma OR "wilms tumor" OR nephroblastoma OR neuroblastoma OR lymphoma OR hodgkin OR menigioma)
AND		
Intervention	Cancer Treatment	ALL = (radiotherapy OR radiation* OR " x-ray therapy" OR chemotherapy OR alkylating OR anthracycline* OR antineoplastic OR "anticancer drug*" OR "vinca alkaloid" OR antimetabolites OR topoisomerase OR "Tyrosine kinase" OR epipodophyllotoxin OR antibiotic* OR chemoradiotherapy OR radiochemotherapy)
AND		
Outcome	Second cancer	ALL = ("second primary cancer*" OR "second primary neoplasm*" OR "second primary tumo\$r*" OR "second primary malignanc*" OR "secondary primary cancer*" OR "secondary primary neoplasm*" OR "secondary primary tumo\$r*" OR "secondary primary malignanc*" OR "subsequent malignant neoplasm*" OR "subsequent malignant tumo\$r*" OR "subsequent malignant cancer*" OR "subsequent neoplasm*" OR "subsequent tumo\$r*" OR "subsequent cancer*" OR "second neoplasm*" OR "second cancer*" OR "second tumo\$r*" OR "second malignanc*" OR "secondary neoplasm*" OR "secondary cancer*" OR "secondary tumo\$r*" OR "secondary malignanc*" OR "therapy associated neoplasm*" OR "treatment associated neoplasm*" OR "radiation induced neoplasm*" OR "therapy associated cancer*" OR "treatment associated cancer*" OR "radiation induced cancer*" OR "subsequent breast cancer*" OR "secondary breast cancer*" OR "second breast cancer*" OR "subsequent thyroid cancer*" OR "secondary thyroid cancer*" OR "second thyroid cancer*")
	Genetics variants	ALL = (biomarker* OR genetic* OR exome* OR genom* OR transcriptome OR sequencing OR genotype* OR polymorphism* OR snp* OR mutation* OR GWAS OR haplotype* OR "copy number variation*" OR "pathway analysis" OR " germline predisposition*" OR variant* OR rna-seq OR genotyping)
AND		
Study design	ALL = (cohort* OR "cohort stud*" OR "observational stud*" OR "case-control stud*" OR "prospective stud*" OR review OR meta-analysis OR "Retrospective Stud*" OR "longitudinal stud*")	
NOT		
TS = ("bomb" OR "nuclear power plant")		

ANNEXE B – Patients inclus dans l'étude et porteurs d'un syndrome génétique

Syndrome in FCCSS database	Cases	Controls
Neurofibromatosis Type 1 / von Recklinghausen Disease – <i>NF1</i>	3	1
Turcot Syndrome	1	0
Ataxia Telangiectasa – <i>ATM</i>	0	0
Gorlin Syndrome (Nevoid Basal Cell Carcinoma Syndrome / NBCCS / Basal Cell Nevus Syndrome) – <i>PTCH1</i>	0	1
Familial Adenomatous Polyposis (FAP) or Gardner Syndrome – <i>APC</i>	0	1
Li-Fraumeni Syndrome – <i>TP53</i>	5	1
Silver-Russel Syndrome	1	0
WAGR Syndrome (Wilms tumor, Aniridia, Genitourinary anomalies, and mental Retardation) – <i>WT1</i>	0	0

In total, 14 patients included in this study are carriers of a syndrome; and 220 patients are not carriers of any syndrome. However, it is important to note that this information is missing for 161 included patients.

Publications

Genetic Risk of Second Malignant Neoplasm after Childhood Cancer Treatment: A Systematic Review

Claire Ducos¹, Naïla Aba¹, Filippo Rosselli², Brice Fresneau^{1,3}, Baraah Al Ahmad Nachar², Monia Zidane¹, Florent de Vathaire¹, Simone Benhamou⁴, and Nadia Haddy¹



ABSTRACT

Second malignant neoplasm (SMN) is one of the most severe long-term risks for childhood cancer survivors (CCS), significantly impacting long-term patient survival. While radiotherapy and chemotherapy are known risk factors, the observed inter-individual variability suggests a genetic component contributing to the risk of SMN. This article aims to conduct a systematic review of genetic factors implicated in the SMN risk among CCS. Searches were performed in PubMed, Scopus, and Web of Sciences. Eighteen studies were included (eleven candidate gene studies, three genome-wide association studies, and four whole

exome/genome sequencing studies). The included studies were based on different types of first cancers, investigated any or specific types of SMN, and focused mainly on genes involved in drug metabolism and DNA repair pathways. These differences in study design and methods used to characterize genetic variants limit the scope of the results and highlight the need for further extensive and standardized investigations. However, this review provides a valuable compilation of SMN risk-associated variants and genes, facilitating efficient replication and advancing our understanding of the genetic basis for this major risk for CCS.

Introduction

Thanks to advances in treatment and improved diagnosis, 5-year childhood cancer survival increased over time, and has now reached 80% in most European countries (1). However, despite this achievement, two-thirds of survivors experience later complications (2). A serious long-term consequence of treatment of childhood cancer is the increased risk of developing second malignant neoplasms (SMN), which is three to five times greater than that of the general population (3).

Several studies have demonstrated a relationship between SMN development and treatment received for the first cancer, especially radiotherapy and chemotherapy (4–8). A dose-dependent relationship between SMN risk and the dose of ionizing radiation received has been established for several cancer types, as well as an increased risk of SMN at the radiation site (7). Similarly, most chemotherapy molecules, such as alkylating agents and anthracyclines, have been associated with the risk of several types of SMN (5, 8). While treatment is a recognized risk factor, it does not sufficiently account for the risk of SMN. Evidence of inter-individual variability suggests a genetic component (9–11).

Some studies investigated the associations between germline genetic variants and SMN risk. Initially, candidate gene and genome-wide association studies (GWAS) were used to identify common variants associated with this risk. With the introduction of advanced genotyping technologies, such as whole exome and genome sequencing (WES/WGS), attention has turned to the exploration of rare variants. This shift in focus brings forth valuable new information, contributing to a deeper understanding of the risk of SMN in childhood cancer survivors (CCS). Furthermore, these genetic association studies uncover novel genetic risk factors and offer an opportunity to investigate the interaction between genetics and previous cancer treatments to better personalize patient cancer treatment and improve CCS follow-up.

This review aims to provide an overview of genetic variants, genes, and potential gene-treatment interactions related to SMN risk.

Materials and Methods

The protocol for this systematic review has been previously registered on PROSPERO under the access number CRD42023388965. Finding reports follow the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines (12).

Information sources and search strategy

A systematic search of original articles published before June 2021 was performed, using PubMed, Scopus, and Web of Sciences. Additionally, a manual check was performed in early 2024 to identify any recent publications for potential inclusion. The search terms used for each database are presented according to the Population Intervention Comparator Outcome Study (PICOS) framework in Supplementary Tables S1–S3: Population—patients treated for childhood cancer; Intervention—patients treated with chemotherapy and/or radiotherapy; Comparator—not applicable for this review; Outcomes—studies that included genetic data involved in SMN risk in patients treated for childhood cancer; Study—case-control and cohort studies.

¹Radiation Epidemiology Team, Center for Research in Epidemiology and Population Health, INSERM Unit 1018, University Paris Saclay, Villejuif, France.

²CNRS UMR9019, Gustave Roussy Cancer Campus, Université Paris-Saclay, Equipe Labellisée Ligue Nationale Contre le Cancer Villejuif, France. ³Department of Children and Adolescents Oncology, Gustave Roussy Cancer Campus, Université Paris-Saclay, Villejuif, France. ⁴Oncostat Team, Center for Research in Epidemiology and Population Health, INSERM Unit 1018, University Paris Saclay, Villejuif, France.

S. Benhamou and N. Haddy contributed equally to this article.

Corresponding Author: Claire Ducos, Radiation Epidemiology Team in CESP, INSERM B2M, 32 rue Camille Desmoulins, Villejuif 94800, France. E-mail: Claire.ducos@gustaveroussy.fr

Cancer Epidemiol Biomarkers Prev 2024;33:999–1011

doi: 10.1158/1055-9965.EPI-24-0010

©2024 American Association for Cancer Research

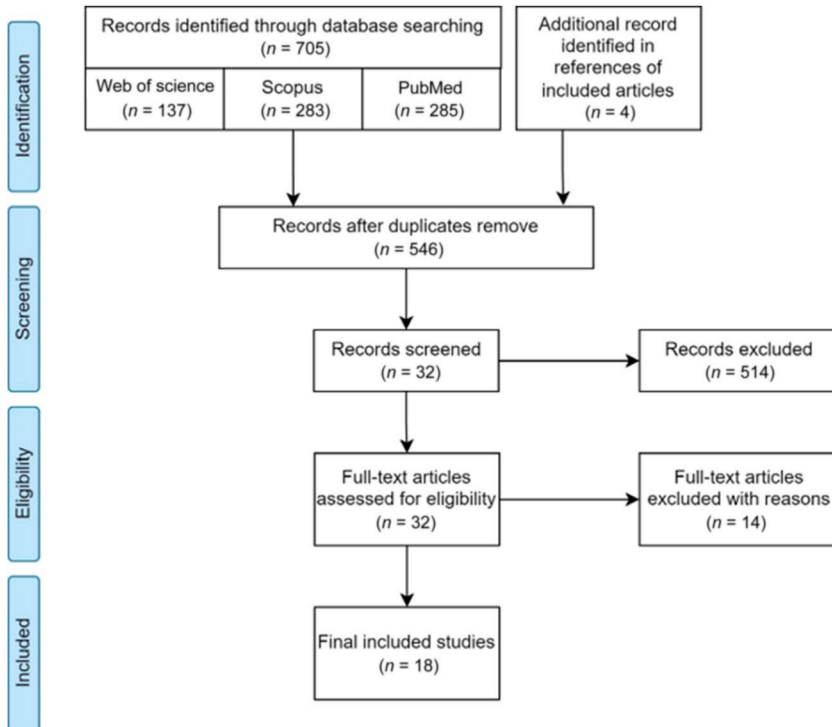


Figure 1. PRISMA flow chart of study inclusion. This figure shows the study selection process following PRISMA guidelines. It outlines the number of studies identified, screened for eligibility, assessed for inclusion, and included in this systematic review.

Manuscripts were excluded if they were not in English and were not about humans. We did not include articles dealing exclusively with cancer predisposition genes, such as *RBI*, *NFI*, and *TP53*. Reference lists of included manuscripts and previous relevant reviews were also searched. The titles and abstracts were reviewed in parallel by C. Ducos and N. Aba, using a priori exclusion and inclusion criteria. Conflicts were resolved between the authors or with the involvement of N. Haddy and S. Benhamou. The full texts of the remaining articles were screened to exclude those that did not meet the established inclusion criteria.

Data extraction

Data were recorded using a structured data extraction form to retrieve relevant information from the main text and supplementary material. This information includes details regarding the study and included patient characteristics and the main results of the association.

Quality assessment

We used the Q-GENIE scale (13) to assess the quality of the articles included in this review. This scale allows the evaluation of the classic biases accounted for in the Newcastle-Ottawa scale and the consideration of biases specific to genetic association studies. This scale judges articles based on 11 criteria, assigning a score from 1 (poor) to 7 (excellent). The overall score then classifies the article into three categories: poor, moderate, and high quality. Quality assessment is summarized in Supplementary Table S4.

Data availability

All data relevant to this study are included in the published article, its supplementary materials, and can also be found in the original articles included in this systematic review.

Results

Study selection

The search strategy identified 709 articles. After the removal of duplicates ($n = 163$), the remaining 546 articles were independently screened at the title/abstract level and 32 eligible studies were retained for full-text screening. Fourteen full-text articles were excluded as they did not meet the criteria outlined in the PICOS framework (Supplementary Table S5). A total of 18 studies was included in this systematic review and underwent data extraction and analysis. Selection phases can be found in the PRISMA flow chart (Fig. 1).

Study characteristics

Characteristics of studies are summarized in Table 1. The results in this review are categorized based on the genetic approach used in the studies: (i) eleven candidate gene studies (14–24), (ii) three GWAS (25–27), one of which (27) used a custom Illumina iSelect Array (iCOGS) containing 211,155 single-nucleotide polymorphisms (SNP), and (iii) four WES or WGS studies (28–31). All the studies were published between 2000 (24) and 2020 (14, 28, 29). Sample size of patients diagnosed with SMN after treatment for a first cancer during childhood, varied from 20 (19) to 207 (26) for candidate gene studies and GWAS, and from 47 (30) to 1,108 (28) for WES or WGS studies. Most of the published articles were based on the Childhood Cancer Survivors Study (CCSS; 31 participating centers in the United States and Canada) and/or SJLIFE cohorts (St. Jude Children’s Research Hospital, Memphis; refs. 14, 15, 22, 25–31).

Four articles specifically focused on survivors of Hodgkin lymphoma (20, 22, 25, 27), four on survivors of acute lymphoblastic leukemia (ALL; refs. 19, 21, 23, 24), and one on survivors of

Table 1. Characteristics of the 18 included studies.

Study	Sequencing method	DNA sample	Study design cohort	Country ethnicity	Follow-up	Childhood cancer	Year of treatment	Age at first cancer diagnosis	SMN
Richard and colleagues (14)	Candidate Gene	Blood, saliva, buccal cells	Case-Control CCSS SMN: 110 No SMN: 4956	US European and African ancestry	NA	Any childhood cancer	1970 to 1986	Mean (SD) Cases: 9 (5.5) Controls: 7.9 (5.9) Year: n (%) SMN <15: 96 (81.4) ≥15: 22 (18.6) No SMN	Thyroid
Gramatges and colleagues (15)	Candidate Gene	NA	Cohort CCSS SMN: 118 No SMN: 5206	US European ancestry	NA	Any childhood cancer	1970 to 1986	<15: 4251 (81.7) ≥5: 955 (18.3) Median (range) cases: 3.5 (0-21) controls: 5 (0-20)	Thyroid
Wang and colleagues (16)	Candidate Gene	Blood, saliva	Matched Case-Control Study COG SMN: 82 No SMN: 228 Validation: 25/54	US Matching on ancestry	Median Cases: 13.2 Controls: 13.2	Any childhood cancer	NA		CNS tumor
Applebaum and colleagues (17)	Candidate Gene	NA	Matched Case-Control Study COG SMN: 31 No SMN: 309	US European and African ancestry	Median time to SMN (range) Cases: 41 months (1-223)	Neuroblastoma ^a	1990 to 2006	Median Cases: 27.5 months Controls: 18 months	Any SMN
Vodusek and colleagues (18)	Candidate Gene	Biopsy - normal tissue	Matched Case-Control Study Patients followed up at Ljubljana SMN: 24 No SMN: 24	Slovenia	Median (range) Cases: 19.6 (9.0-23.6) Controls: 18.8 (12.80-27.6)	Any childhood cancer	1960 to 2006	Median (range) Cases: 12.7 (7.1-6.7) Controls: 12.9 (5-15.4)	Thyroid
Erčulj and colleagues (19)	Candidate Gene	Bone Marrow	Nested Matched Case-Control Patient followed up at Ljubljana SMN: 20 No SMN: 39	Slovenia	Median time at risk of SMN (range) Cases: 9.5 (2-34) Controls: 28 (10-37)	ALL	1971 to 2001	Median (range) Cases: 5 (1.3-12.3)	Any SMN
Ma and colleague (20) ^a	Candidate Gene	Blood	Case-Control National cohort study of breast cancer SMN: 232 (<20 years: 101) No SMN: 461 (<20 years: 140)	UK (Validation Netherlands) European ancestry	NA	Hodgkin lymphoma	1963 to 2003	Year: n (%) Cases Discovery: 0-19: 68 (48.6) Validation: 0-19: 33 (35.9) Controls Discovery: 0-19: 102 (33) Validation: 0-19: 38 (25) Year: n (%) <1-6: 72 (68.6) ≥6: 33 (31.4)	Breast
Stanulla and colleagues (21)	Candidate Gene	Peripheral blood or bone marrow	Cohort ALL-BFM cohort SMN: 105 No SMN: 814	Germany	Median time to SMN Year: n (%) <3: 33 (31.4) ≥3 - <10: 56 (53.3) ≥10: 16 (15.2)	ALL	>1979		Any SMN

(Continued on the following page)

Table 1. Characteristics of the 18 included studies. (Cont'd)

Study	Sequencing method	DNA sample	Study design cohort	Country ethnicity	Follow-up	Childhood cancer	Year of treatment	Age at first cancer diagnosis	SMN
Mertens and colleagues (22)	Candidate Gene	Buccal cell	Cohort CCSS SMN: 133 No SMN: 523	US	Years of Follow-up: n (%) 5-10: 3 (1) 11-15: 73 (11) 6-20: 197 (30) 21-25: 195 (30) <25: 182 (28) Controls are followed up for at least 2.5 years following the diagnosis of ALL	Hodgkin lymphoma (treated with RT)	1970 to 1986	Year: n (%) 0-5: 23 (3) 6-10: 84 (13) 11-15: 291 (45) 16-20: 252 (39)	Any SMN (in or out RT field)
Woo and colleagues (24)	Candidate Gene	Bone marrow or peripheral blood	Case-Control SJCRH, DFW, CHOP, POG SMN: 57 No SMN: 245	US	Controls are followed up for at least 2.5 years following the diagnosis of ALL	ALL	NA	Median (quartiles) Cases: 7.1 (3.5-11.1) Controls: 5.5 (3.4-10.2)	t-AML and t-MDS
Opstal-van Winden and colleagues (27)	GWAS ^c	Blood	Case-Control CCSS, British Hodgkin lymphoma cohort, Dutch Hodgkin lymphoma cohort SMN: 327 (250 ≤ 20 y) No SMN: 491 (316)	US, Netherlands, United Kingdom Adjusted on ancestry and ethnicity	Median (range) Cases: 24 (9-46) Controls: 30 (9-49)	Hodgkin lymphoma	1964 to 1999	Median (range) Cases: 19 (10-40) Controls: 22 (6-40)	Breast
Morton and colleagues (26)	GWAS	CCSS: blood, saliva or buccal cells SJLIFE: blood	Cohort CCSS and SJLIFE SMN: 207 No SMN: 2,774	US European ancestry	Median age at breast cancer (IQR) 39.2 (35.1-43.2) Median age at last follow-up (IQR) 32.9 (27.7-39.3)	Any childhood cancer	CSSS: 1970 to 1986 SJLIFE: 1962 to 2005	Median (IQR) SMN: 15.6 (13.5-18.0) No SMN: 6.1 (2.9-12.7)	Breast
Best and colleagues (25)	GWAS	Whole blood, saliva	Case-Control CCSS SMN: 158 (96 + 62) No SMN: 153 (82 + 71)	US European ancestry	Mean (SD) latency ^b Cases: 20 (5.8) Median (range) ^b Controls: 32 (27-38)	Hodgkin lymphoma	1970 to 1986	Mean (SD) ^b Cases: 15.9 (3.0) Controls: 15.6 (2.8)	Any SMN
Morton and colleagues (28)	WES	Blood, saliva, buccal cells	Matched Case-Control Study CCSS SMN: 1108 No SMN: 3883	US European ancestry	Mean SMN: 24.7 year No SMN: 32 year	Any childhood cancer	1970 to 1986	Year: n (%) Cases: <5: 274 (24.7)/5-9: 205 (18.5) 10-14: 299 (27)/≥15: 330 (29.8) Controls: <5: 1,678 (43.2)/5-9: 879 (22.6) 10-14: 738 (19)/≥15: 588 (15.1)	RT-SMN
Qin and colleagues (29)	WGS + WES	Blood, saliva, buccal cells	Cohort SJLIFE SMN: 495 No SMN: 3,907	US Adjusted on ancestry	Median (range) 22.3 (14.2-32.1) Median time from childhood cancer diagnosis to the first SMN 26.5 (19.1-33.9)	Any childhood cancer	>1962	Median (range) 6.3 (2.8-12.5)	Any SMN

(Continued on the following page)

Table 1. Characteristics of the 18 included studies. (Cont'd)

Study	Sequencing method	DNA sample	Study design cohort	Country ethnicity	Follow-up	Childhood cancer	Year of treatment	Age at first cancer diagnosis	SMN
Wang and colleagues (31)	WGS + WES	Blood	Cohort SJLIFE SMN: 492 No SMN: 2,514	US	Median (range) Cohort: 28 (6.1–54.7) SMN: 25.6 (18.6–32.7)	Any childhood cancer	>1962	Median (range) 7.1 (0.1–23.6)	Any SMN
Wang and colleagues (30)	Sequencing Study	Blood	Case–Controls SJLIFE, only female survivors SMN: 47 No SMN: 1,086	US European ancestry	Median (range) Cases: 31.2 (13.4–51.2) Controls: 28.1 (5.6–53.5)	Any childhood cancer	>1962	Median (range) SMN: 15.2 (3.1–21.2) No SMN: 6.1 (0.01–22.7)	Breast

Abbreviations: BFM, Berlin-Frankfurt-Münster; CHOP, Children's Hospital of Philadelphia; COG, Children's Oncology Group; DFW, Children's Medical Center/Southwestern University; IQR, Interquartile range; POG, Paediatric Oncology Group; SJCRH, St Jude Children's Research Hospital; t-MDS, therapy-related myelodysplastic syndrome.

^aOnly results for patients under 20 years of age are presented in this table.

^bCharacteristics for the discovery set.

^cGWAS performed using a custom Illumina iSelect Array, which includes 211,155 SNPs, specifically developed for the European Collaborative Oncological Gene-Environment Study project (iCOGS array).

neuroblastoma (17). The remaining articles did not target a specific type of first cancer. Included studies investigated the association between genetic variations and the risk of SMN (17, 19, 21, 22, 25, 29, 31), as well as specific SMNs, including breast cancer (20, 26, 27, 30), thyroid cancer (14, 15, 18), central nervous system (CNS) cancer (16), therapy-related acute myeloid leukemia (t-AML; refs. 23, 24), and radiation-related SMN (28).

Although this review focuses on CCS, two studies (20, 27) investigating the risk of breast cancer after Hodgkin lymphoma diagnosed before the ages of 36 and 40 were included, as results by age at diagnosis (≤ 20 , > 20 years) were reported.

Population stratification has the potential to introduce false positive or negative associations between genetic variants and SMN risk. To account for this confounding factor, eight focused, or conducted subtype analysis, on European descent (14, 15, 17, 20, 25, 26, 28, 30), and three studies adjusted for ancestry (16, 27, 29).

Quality assessment

The GWAS and WES articles all had a score greater than 45, implying high-quality studies. Overall, nine studies (14, 16, 25–31) were rated high quality, while nine were rated moderate quality (15, 17–24). The results of the quality control are comprehensively presented in Supplementary Table S4.

Results of syntheses

Main characteristics of the 18 included studies are summarized in **Tables 2–4**. A summary of the evidence is provided by study type: candidate gene studies (**Table 2**); GWAS (**Table 3**); and WES/WGS studies (**Table 4**).

Candidate gene studies

Among candidate gene studies, four considered any SMN and seven specific SMNs (**Table 2**).

Any SMN

In a study focusing on patients treated for Hodgkin lymphoma (22), the *GSTT1* null genotype was associated with an increase in risks of any SMN [relative risk (RR) = 1.5, 95% confidence interval (CI): 1.0–2.3] and SMN within the radiation field (RR = 1.4 95% CI, 0.9–2.1). Another study among neuroblastoma survivors (17) investigated 66 SNPs within 47 genes previously associated with SMN development. An increased SMN risk associated with rs861539 variant in the *XRCC3* gene [Odds Ratio (OR) = 2.04, 95% CI, 1.19–3.46] and a decreased SMN risk associated with rs17036651 variant in the *MSH2* gene (OR = 0.26, 95% CI, 0.09–0.81) were found. Significant associations were also found between SMN and rs709816 variant in the *NBN* gene (RR = 4.36; 95% CI, 1.2–16.0), and rs1799794 variant in the *XRCC3* gene (RR = 0.20; 95% CI, 0.04–0.99) among ALL survivors (19). Conversely, no association was observed in the fourth study focusing on *TPMT* gene (21).

Central nervous system

A significant association between CNS SMN and variant rs15869 in the *BRCA2* gene (OR = 2.5; 95% CI, 1.5–4.2) has been found (16). The variant rs1805389 in the *LIG4* gene was associated with any CNS and glioma risk, while rs25489 in the *XRCC1* gene and rs1673041 in the *POLD1* gene were associated with the risk of any CNS tumors and meningioma.

Table 2. Summary of findings from candidate gene association studies.

Study	Genes of interest	Outcome	Subtype analysis	Gene	Variants	Effect allele	OR 95% CI	P-value	Correction for multiple tests	Matching/adjustment covariables
Richard and colleagues (14)	60 SNP in/or near 14 telomere candidate genes	Thyroid	Total sample	POT1	rs58722976	G	HR = 6.1 (2.4–15.5)	0.0001	Bonferroni correction	Adjustment: sex, birth decade, age at diagnosis, primary cancer diagnosis, treatment exposures, thyroid nodules
Gramatges and colleagues (15)	9 SNP previously identified as associated with LTL	Thyroid	Adjusted on ancestry (European and African) Only with European ancestry individuals	POT1	rs58722976	G	HR = 8.0 (2.3–27.2)	0.0009		
Wang and colleagues (16)	129 SNPs in 46 genes implicated in the etiology of adult-onset primary CNS tumor	CNS	—	BRCA2	rs15869	C	2.5 (1.5–4.2)	<0.001	False Discovery Rate	Adjustment: birth year, primary cancer diagnosis, age at diagnosis, sex, treatment exposures Matching: primary cancer diagnosis, year of diagnosis, ethnicity, duration of follow-up Adjustment: age at diagnosis, sex, CRT
Wang and colleagues (16)	129 SNP in 46 genes implicated in the etiology of adult-onset primary CNS tumor	Glioma	—	LIG4 TP53 PHLDB1 CCDC26 XRCC1 POLD1 ERCC1 XRCC5 TP53	rs1805389 rs8079544 rs498872 rs4295627 rs25489 rs1673041 rs11615 rs828699 rs2909430	A G G A G A A A G	3.7 (1.4–9.9) 4.0 (1.1–15.2) 1.8 (1.0–3.3) 1.7 (0.9–3.2) 3.6 (1.1–11.5) 1.9 (1.1–3.1) 1.6 (1.0–2.5) 1.5 (1.0–2.2) 7.5 (1.5–36.9)	0.008 0.04 0.06 0.06 0.03 0.01 0.04 — 0.01	Matching: primary cancer diagnosis, year of diagnosis, ethnicity, duration of follow-up Adjustment: age at diagnosis, sex, CRT	
Applebaum and colleagues (17)	66 SNP in 47 genes previously reported to be associated with development of SMN Loci in antioxidant defense-related genes	Any SMN	—	LIG4 BRCA2 TPMT BRCA2 XRCC5 XRCC1 CDKN1A POLD1 XRCC3 MSH2	rs1805389 rs15869 rs2518471 rs15869 rs828699 rs25489 rs1801270 rs1673041 rs861539 rs17036651	A C G C A G C A T G	19.7 (2.5–156.4) 3.7 (1.3–10.2) 3.9 (0.9–16.7) 2.4 (1.2–4.9) 1.7 (0.9–3.1) 6.9 (1.1–43.6) 9.6 (1.1–85.4) 2.0 (1.0–3.8) 2.04 (1.19–3.46) 0.26 (0.09–0.81)	0.005 0.01 0.07 0.01 0.07 0.04 0.04 0.04 0.006 0.008	Matching: primary cancer diagnosis, year of diagnosis, ethnicity, duration of follow-up Adjustment: age at diagnosis, sex, CRT	
Vodusek and colleagues (18)	LocI in antioxidant defense-related genes	Thyroid	—	No association	No association			NA	NA	Matching: age at diagnosis, sex, primary cancer diagnosis, treatment (especially RT)

(Continued on the following page)

Table 2. Summary of findings from candidate gene association studies. (Cont'd)

Study	Genes of interest	Outcome	Subtype analysis	Gene	Variants	Effect allele	OR 95% CI	P-value	Correction for multiple tests	Matching/adjustment covariables
Ercüj and colleagues (19)	T1 DNA repair polymorphisms	Any SMN	—	<i>NBN</i>	rs709816	G	RR = 4.36 (1.2–16.0)	0.026	NA	Matching: CRT administration of epipodophylotoxins, cumulative doses of anthracyclines, sex and age at diagnosis
Ma and colleagues (20)	14 breast cancer risk SNP	Breast	<20 years old	<i>XRCC3</i> <i>FGFR2</i>	rs1799794 rs1219648	G	RR = 0.20 (0.04–0.99) 1.70 (1.16–2.50)	0.049	—	Matching: age and date at diagnosis, region of the country, duration of the follow-up
Stanulla and colleagues (21)	<i>TPMT</i>	Any SMN	—	No association	NA	NA	—	—	—	—
Mertens and colleagues (22)	<i>GSTM1</i> , <i>GSTT1</i> , <i>XRCC1</i>	Any SMN	SMN within the radiation field	<i>GSTM1/GSTT1</i>	<i>GSTM1</i> –/ <i>GSTT1</i> +	—	RR = 1.4 (0.91–2.12)	—	NA	Adjustment: age at diagnosis and smoking status
Bianco and colleagues (23)	<i>CYP3A4</i> , <i>CYP3A5</i> , <i>NGO1</i>	t-AML/t-MDS	—	No association	<i>GSTM1/GSTT1</i>	—	RR = 1.5 (1.03–2.29)	—	NA	NA
Woo and colleagues (24)	<i>GSTM1</i> , <i>GSTT1</i>	t-AML/t-MDS	only male	<i>GSTM1/GSTT1</i>	double null genotype	—	—	0.036	NA	Adjustment: ethnicity

Abbreviations: CRT, cranial radiation therapy; LTL, leukocyte telomere length; t-MDS, therapy-related myelodysplastic syndrome.

Breast cancer

One study (20) focusing on 14 SNPs previously associated with breast cancer risk in the general population reported a significant association between the variant rs1219648 in the *FGFR2* gene and breast SMN among patients younger than 20 years old (OR = 1.7; 95% CI, 1.16–2.5).

Leukemia

Two studies (23, 24) investigated the risk of leukemia as SMN but neither found any association with the studied SNPs (Table 2).

Thyroid cancer

Two studies investigated genes that control telomere length and the risk of thyroid SMN. The first one (15) did not find any association while the other (14) reported a significant association between the variant rs58722976 in the *POT1* gene and thyroid SMN [Hazard Ratio (HR) = 6.1; 95% CI, 2.4–15.5].

In a study focusing on six antioxidant genes (18), no significant association was found with secondary thyroid cancer after childhood or adolescent cancer treatment.

Genome-wide association studies

The results of the three selected GWAS (25–27) are summarized in Table 3. The first study (25) identified three SNPs significantly associated with the risk of developing a SMN in Hodgkin lymphoma survivors who received radiotherapy. Two SNPs, rs4946728 and rs1040411, in linkage disequilibrium (LD) involved in control of the *PRDM1* gene were associated with a high SMN risk: OR = 3.32 (95% CI, 2.25–74.90) and OR = 2.39 (95% CI, 1.73–3.30). These two SNPs were successfully replicated in a validation cohort. A third SNP (rs8083533) located within the intronic region of the *TAF4B* gene was significantly associated with SMN risk (OR = 1.86; 95% CI, 1.32–2.62) but was not replicated.

In the second study performed in the CCS cohort (26), three variants were associated with the risk of breast SMN in CCS who received radiotherapy to the breast. The variant rs4342822 located close to the *PROX1* gene and the variant rs74949440 in the *TAGLN* gene showed a high risk among female survivors receiving more than 10 Gy of radiation dose (HR = 1.92; 95% CI, 1.49–2.44 and HR = 2.59; 95% CI, 1.62–4.16, respectively). The rs17020562 variant located in the *RPS6K1* gene was associated with breast cancer risk in survivors who received less than 10 Gy to the breast (HR = 44.52; 95% CI, 15.06–131.62).

A polygenic risk score (PRS) composed of nine SNPs interacting with radiotherapy in the occurrence of breast cancer was used in the third GWAS (27). Although the PRS was associated with a higher risk of breast cancer following chest radiation therapy for Hodgkin lymphoma in the overall study, this association was not observed in the subgroup of patients treated for Hodgkin lymphoma before the age of 20.

Whole exome/genome studies

The results of the four studies using genomic and/or exomic sequencing data (28–31) are presented in Table 4.

Any SMN

The first sequencing study (31) examined the risk of SMN in CCS carrying pathogenic or likely pathogenic (P/LP) variants in a set of 60 cancer predisposition genes. Carriers of variants in these genes were more susceptible to developing any SMN (RR = 1.8; 95% CI, 1.2–2.6) or multiple cancers (RR = 3.1; 95% CI, 1.5–6.3). In addition, after stratification on radiation exposure, survivors who did not receive radiotherapy and carried P/LP variants were more likely

Table 3. Summary of findings from GWAS.

Study	Outcome	Subtype Analysis	Gene	Variants	Effect allele genotype	HR/OR 95% CI	P-value	Correction for multiple tests	Adjustment covariables
Opstal-van Winden and colleagues ^a (27)	Breast	Age at Hodgkin lymphoma treatment ≤20 years	RT-interaction-PRS: 9 SNP that showed a statistically significant interaction with RT-induced breast cancer	rs17020562	Lowest tertile	REF	P trend= 0.027	NA	BC-PRS, age at Hodgkin lymphoma diagnosis, country and ethnicity
Morton and colleagues (26)	Breast	<10 Gy to the breast	<i>RPS6KC1</i> (rare variant)	rs4342822	Middle tertile	HR = 44.52 (15.06–131.62)	5.00 × 10 ⁻⁹	Permutation method	Cohort, alkylating agent or platinum-containing chemotherapy, associated principal component
		≥10 Gy to the breast	<i>PROX1</i>	rs74949440	Highest tertile	HR = 1.92 (1.49–2.44)	7.09 × 10 ⁻⁹		
Best and colleagues (25)	Any SN	—	<i>TAGLN</i> (rare variant)	rs4946728	C	OR = 3.32 (2.25–4.90)	5.99 × 10 ⁻¹⁰	Permutation method	Age at Hodgkin lymphoma diagnosis, sex, year of Hodgkin lymphoma diagnosis, gonadal radiation in females and alkylating chemotherapy
			<i>PRDM1</i>	rs1040411	T	OR = 2.39 (1.73–3.30)	1.18 × 10 ⁻⁷		
			<i>TAF4B</i>	rs808533	T	OR = 1.86 (1.32–2.62)	4.00 × 10 ⁻⁴		

^aGWAS performed using a custom Illumina iSelect Array, which includes 211,155 SNPs, specifically developed for the European Collaborative Oncological Gene-Environment Study project (iCOGS array).

to develop any SMN (RR = 4.7; 95% CI, 2.4–9.3), non-malignant skin cancers (RR = 11.0; 95% CI, 2.9–41.4), or multiple SMNs (RR = 18.6; 95% CI, 3.5–99.3).

In the CCSS cohort, WES was used to investigate the effect of rare variants on the risk of radiotherapy-related second neoplasms (RT-SN; ref. 28). An association was found between variants in the homologous recombination repair (HRR) pathway and the risk of developing any RT-SN out of the field of radiotherapy (OR = 2.6; 95% CI, 1.7–3.9). Among genes belonging to this pathway, there was a suggestive association between the *FANCM* gene and the risk of RT-SN out-of-field (OR = 11.2; 95% CI, 4.1–30.8).

The risk of developing SMNs in carriers of pathogenic variants in genes belonging to DNA repair pathways was also investigated (29). An increased risk of any SMNs was found in carriers of variants in the HRR (RR = 1.5; 95% CI, 1.1–2.2), mismatch repair (MMR; RR = 2.1; 95% CI, 1.2–3.7), or nucleotide excision repair (NER; RR = 1.6; 95% CI, 1.1–2.5) pathways.

Breast

One study (31) showed that females who carried P/LP variants were more likely to develop breast cancer (RR = 9.4; 95% CI, 4.8–18.4). Patients who did not receive radiotherapy were still at risk of developing breast cancer (RR = 7.7; 95% CI, 2.4–44.4), and survivors who received irradiation were at higher risk (RR = 13.9; 95% CI, 6.0–32.2).

In a study (30) using a PRS derived from 170 breast cancer-associated loci, an increased risk was observed among survivors in the highest compared to the lowest quintile of the PRS (RR = 2.7; 95% CI, 1.0–7.3). The observed association remained consistent among patients who received radiotherapy.

When focusing on six pathways associated with breast SMN, one study (29) reported that pathogenic germline variants in HRR genes were significantly associated with an increased risk of breast cancer (RR = 3.7; 95% CI, 1.8–7.7), especially among survivors who received chest radiotherapy ≥ 20 Gy (RR = 4.4; 95% CI, 1.6–12.4), or with a cumulative dose of anthracyclines ≥ 98 mg/m² (RR = 4.4; 95% CI, 1.7–11.4). Similar results were found with pathogenic germline variants in Fanconi Anemia genes.

Sarcoma

Carriers of variants in the HRR pathway had a significant risk of developing sarcoma after receiving 10.45 mg/m² of alkylating agents (RR = 14.9; 95% CI, 4.0–38.0; ref. 29).

Discussion

This article provides a comprehensive review of the current understanding of the associations between genetic variants, genes, and pathways and the risk of SMN among CCS, through genetic association studies reported so far (14–31). Prior to the advent of high-throughput genotyping and sequencing technologies, the candidate gene approach was predominant. These studies focused on genes involved in drug metabolism (18, 21–24), in DNA repair pathways (19, 22), or genes known to have oncogenic effects (16, 17, 20). The candidate gene approach offers the benefit of concentrating on relevant genes, making them effective in identifying potential associations. However, these studies might overlook other genetic factors contributing to SMN risk.

Since the 2000s, GWAS have allowed the study of millions of variants, without *a priori* assumptions, providing a broader scope for uncovering genetic associations and potentially discovering novel common variants. Two GWAS (25, 26) included in this review

have primarily uncovered two SNPs associated with SMN risk. One SNP is located within the *PRDMI* gene (25), known as a tumor suppressor in B cells (32), and another within the *PROX1* gene (26), previously associated with breast cancer and various other types of cancer (33). The latter study (26) was unable to replicate the results from the first GWAS (25). This disparity may be attributed to differences in the selection criteria for individuals in both studies and the relatively modest sample sizes.

GWAS, limited by chip-based sequencing, may not effectively identify rare variants that could have a crucial impact on the risk of SMN. The missing heritability hypothesis (34) suggests that common variants only explain a fraction of heritability. New technologies like WES and WGS allow the exploration of rare genetic variants contributing to SMN occurrence. This review includes four articles using WES or WGS based on US CCS cohorts.

An increased risk of breast SMN was observed among CCS carrying the identified variants and receiving radiation therapy, highlighting the significance of genetic factors as modifiers in the association between treatment and SMN risk (30).

Two other studies investigated the role of genes involved in the DNA repair pathways or clinical radiation sensitivity in the SMN occurrence (28, 29). Both studies showed that carriers of variants in genes of the HRR pathway had an increased risk of developing RT-SN, any SMN, or breast cancer. Two key genes associated with SMN risk were also identified in the first study (28). The *FANCM* gene, previously associated with chemotherapy-related toxicity and cancer risk (35), was linked to an increased risk of out-of-field RT-SN. Additionally, the *EXO1* gene, involved in various DNA repair pathways has been associated with an increased susceptibility to radiation toxicity and cancer development (36, 37). Another sequencing study (31) also demonstrated a significant link between mutations in the SJCPC60 genes list and increased SMN risk, observed in both irradiated and non-irradiated groups.

Limitations

The literature on genetic associations for SMN after childhood cancer treatment is weak and limited by the variability in study design, methods, location of primary cancers, and outcomes. Some studies focused on radiation-induced SMNs, such as breast and thyroid cancer, while others examined any secondary neoplasms.

Many of the reviewed articles did not contain replication of the results, highlighting the need for future investigations to confirm the findings to ensure their credibility. Genetic association studies on the risk of developing a SMN in CCS have produced various and mostly unrelated results. Consequently, due to a lack of replication efforts, these genetic findings cannot yet be directly applied in clinical practice to adapt the treatment and the follow-up of children with cancer.

Population stratification represents another limitation of genetic studies leading to potential false positive associations. As most of the included articles focused on individuals of European ancestry, the results are not generalizable to non-European populations.

Additionally, correcting for multiple testing is an important issue in genetic association studies by applying correction methods, such as the Bonferroni correction, permutation, or controlling false discovery rate (FDR). In this review, the majority of the GWAS and WES/WGS studies effectively addressed this concern, while most gene candidate studies did not adequately consider it.

The WES/WGS studies often lack in statistical power to detect associations between rare genetic variants, their interaction with treatment and the risk of SMN. Due to this insufficient statistical power, the study of interaction between predisposition genes and the

Table 4. Summary of findings from whole exome and genome sequencing studies.

Study	Genes of interest	Subtype analysis	Outcome	Gene/pathway	RR (95%CI)	P-value	Correction for multiple tests	Matching/adjustment covariables
Morton and colleagues (28)	476 genes involved in DNA damage response or clinical radiation sensitivity syndromes	Out-of-fields RT-SMN excl. BCC	RT-SMN	HRR pathway	OR = 2.6 (1.7-3.9)	4.79×10^{-5}	Bonferroni correction	Matching: Sex, childhood cancer type, age at diagnosis, survival, radiation dose to the SMN location
	High-quality rare variants classified as potentially protein damaging using SnpEff and ClinVar	RT-SMN	RT-SMN	FANCM EXO1	OR = 11.2 (4.1-30.8) OR = 5.9 (3.6-9.8)	9.64×10^{-5} 3.31×10^{-5}		
		RT-SMN excl. BCC no RT	RT-SMN	NEIL3 TP53	OR = 0.1 (0.0-1.0) ∞	6.14×10^{-4} 9.99×10^{-8}		
Qin and colleagues (29)	127 genes in six major DNA repair pathway	Association between variant status and SMN	Any SMN	PGM in HRR ^a	1.5 (1.1-2.2)	0.051		
	Pathogenic/Likely Pathogenic mutations	Chest RT > 20 Gy	Breast	PGM in MMR ^c PGM in HRR ^a PGM in FA ^d PGM in HRR ^a	2.1 (1.2-3.7) 3.7 (1.8-7.7) 3.7 (1.3-10.6) 4.4 (1.6-12.4)	0.051 0.0016 0.034 0.02	FDR (Benjamini-Hochberg's method)	Adjustment: attained age, sex, and ancestry determined by PCA and clinical variables
		Anthracyclines doses ≥ 98 mg/m ²	Breast	PGM in FA ^d PGM in HRR ^a	5.5 (1.5-14.2) 4.4 (1.7-11.4)	0.03 0.01		
		Alkylating agents $\geq 10,451$ mg/m ²	Sarcoma	PGM in FA ^d PGM in HRR ^a	4.2 (1.4-12.32) 14.9 (4.0-38.0)	0.03 0.001		

(Continued on the following page)

Table 4. Summary of findings from whole exome and genome sequencing studies. (Cont'd)

Study	Genes of interest	Subtype analysis	Outcome	Gene/ pathway	RR (95%CI)	P-value	Correction for multiple tests	Matching/adjustment covariables
Wang and colleagues (31)	SJCPG60 genes ⁶ : established associations with monogenic cancer risk inherited in an autosomal dominant fashion at moderate to high penetrance	Association between variant status and SMN	Any SMN	Variant Carriers in SJCPG60	1.8 (1.2–2.6)	0.003	Holm's method	Adjustment: sex, age at diagnosis of the primary cancer, and treatment
			Breast		9.4 (4.8–18.2)	<0.001		
			Sarcoma		10.9 (4.7–25.2)	<0.001		
			Multiple SMN		3.1 (1.5–6.3)	0.002		
		Irradiated survivors	Breast	Variant Carriers in SJCPG60	13.9 (6.0–32.2)	–		
			Sarcoma		10.6 (4.3–26.3)	–		
	Pathogenic/Likely Pathogenic mutations	Non-irradiated survivors	Any SMN	Variant Carriers in SJCPG60	4.7 (2.4–9.3)	–		
			Breast		7.7 (2.4–44.4)	–		
			NMSC		11.0 (2.9–41.4)	–		
			Multiple SMN		18.6 (3.5–99.3)	–		
Wang and colleagues (30)	PRS 170 breast-cancer associated loci	All individuals	Breast	Variant Carriers PRS Q5	21.8 (7.1–66.8)	<0.001		
	Variants in SJCPG60 ⁶ (Wang and colleagues)	Chest RT	Breast	Variant Carriers	2.7 (1.0–7.3)	0.05	NA	Adjustment: age at primary diagnosis, RT, cumulative doses of alkylating agents and anthracyclines, genotype eigenvalues and attained age
	Pathogenic/Likely Pathogenic mutations	No chest RT			10.3 (3.6–29.8)	<0.001		
				PRS Q5 Variant Carriers	3 (1.1–8.1)	0.03		
					42.8 (9.9–183.8)	<0.001		

Abbreviation: BCC, basal cell carcinoma; FA, Fanconi Anemia; NMSC, non-melanoma skin cancer; PGM, pathogenic germline mutation.

⁶PGM in HRR pathway ($n = 42$): *BABAM1, BARD1, BLM, BRCC3, EME1, EME2, EXO1, FAM175A, GEN1, H2AFX, MDC1, MRE11A, MUS81, NBN, POLQ, RAD50, RAD51AP1, RAD51B, RAD51D, RAD52, RAD54B, RAD54L, RBBP8, RECOL, REV1, RTEL1, SHFM1, SLX1A, SLX1B, TP53BP1, UIM1, XRC2, XRC3, BRCA1, BRCA2, BRIP1, PALB2, RAD51, RAD51C, SLX4, RPA1, RPA2.*

⁶PGM in NER pathway ($n = 33$): *CCNH, CDK7, CETN2, DDB1, DDB2, ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, ERCC6, ERCC8, GTF2H1, GTF2H2, GTF2H3, GTF2H4, GTF2H5, LIG1, MMS19, MNAT1, POLK, RAD23A, RAD23B, RPA3, RPA4, UVSSA, XAB2, XPA, XPC, POLD1, POLE, RPA1, RPA2.*

⁶PGM in FA pathway ($n = 20$): *C19orf40, C1orf86, FAN1, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, BRCA1, BRCA2, BRIP1, PALB2, RAD51, RAD51C, SLX4.*

⁶PGM in MMR pathway ($n = 12$): *MLH1, MLH3, MSH2, MSH3, MSH4, MSH5, MSH6, PMS1, PMS2, PMS2L3, POLD1, POLE.*

⁶SJCPG60: *ALK, APC, BAP1, BMPRIA, BRAF, BRCA1, BRCA2, CBL, CDCT3, CDH1, CDK4, CDKN1C, CDKN2A, CEBPA, DICER1, EPCAM, FH, GATA2, HRAS, KRAS, MAP2K1, MAP2K2, MAX, MEN1, MLH1, MSH2, MSH6, NFI, NF2, NRAS, PALB2, PAX5, PHOX2B, PMS2, PRKAR1, PTCH1, PTEN, PTPN11, RAF1, RBI, RET, RUNX1, SDHA, SDHAF2, SDHB, SDHC, SDHD, SHOC2, SMAD4, SMARCA4, SMARCB1, SOST, STK11, SUFU, TMEM127, TP53, TSC1, TSC2, VHL, WT1.*

received treatment as a modifier of the genetic risk, is limited. Of the 18 included studies, only one investigated interactions (26), while the others conducted stratified analyses, based on the dose of chemotherapy or radiation therapy, often focusing on the site exposed to ionizing radiation (e.g., chest dose received for breast cancer). This raises concerns regarding the ability of these studies to fully capture the complex interplay between genetic predispositions and treatment response, which may have important implications for the development of personalized treatment strategies for pediatric cancer survivors.

Considering the relatively young age of the study population, a point frequently highlighted as a limitation in various studies, it becomes increasingly important to recognize that the influence of rare and common risk variants on second cancer risk could likely shift with advancing age and extended follow-up. In addition, a significant portion of the included studies are based on DNA samples collected from survivors of childhood cancer who were still alive at the time of the study, introducing a potential survival bias.

Future directions

In the future, international collaborations will be essential for studying the genetic risk of SMN following cancer treatment, enabling sufficient population sizes to carry out WES or WGS studies with adequate statistical power. For example, the PanCareSurFup cohort—a large European cohort which now provides strong evidence on late-effect epidemiological childhood cancer survival—should establish biobanks for extensive genetic studies.

Furthermore, international collaborations would allow for the replication of association results on independent validation cohorts. With the increasing number of large-scale genetic studies, investigating interactions between genetic effects and radiation therapy or chemotherapy doses should be an important objective to achieve.

References

- Botta L, Gatta G, Capocaccia R, Stiller C, Cañete A, Dal Maso L, et al. Long-term survival and cure fraction estimates for childhood cancer in Europe (EUROCare-6): results from a population-based study. *Lancet Oncol* 2022; 23:1525–36.
- Vassal G, Schrappe M, Pritchard-Jones K, Arnold F, Basset L, Biondi A, et al. The SIOPE strategic plan: a European cancer plan for children and adolescents. *J Cancer Pol* 2016;8:17–32.
- Mertens AC, Liu Q, Neglia JP, Wasilewski K, Leisenring W, Armstrong GT, et al. Cause-specific late mortality among 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study. *J Natl Cancer Inst* 2008;100:1368–79.
- Allodji RS, Tucker MA, Hawkins MM, Le Deley MC, Veres C, Weathers R, et al. Role of radiotherapy and chemotherapy in the risk of leukemia after childhood cancer: an international pooled analysis. *Int J Cancer* 2021;148:2079–89.
- Turcotte LM, Neglia JP, Reulen RC, Ronckers CM, Van Leeuwen FE, Morton LM, et al. Risk, risk factors, and surveillance of subsequent malignant neoplasms in survivors of childhood cancer: a review. *J Clin Oncol* 2018;36:2145–52.
- Turcotte LM, Whitton JA, Friedman DL, Hammond S, Armstrong GT, Leisenring W, et al. Risk of subsequent neoplasms during the fifth and sixth decades of life in the Childhood Cancer Survivor Study cohort. *J Clin Oncol* 2015;33:3568–75.
- 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 Biol Phys* 2016;94:800–7.
- Hawkins M, Bhatia S, Henderson TO, Nathan PC, Yan A, Teepen JC, et al. Subsequent primary neoplasms: risks, risk factors, surveillance, and future research. *Pediatr Clin North Am* 2020;67:1135–54.
- Armenian SH, Bhatia S. Chronic health conditions in childhood cancer survivors: is it all treatment-related—or do genetics play a role? *J Gen Intern Med* 2009;24:395–400.
- Gramatges MM, Bhatia S. Evidence for genetic risk contributing to long-term adverse treatment effects in childhood cancer survivors. *Annu Rev Med* 2018; 69:247–62.
- Bhatia S. Genetic variation as a modifier of association between therapeutic exposure and subsequent malignant neoplasms in cancer survivors. *Cancer* 2015;121:648–63.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
- Sohani ZN, Meyre D, de Souza RJ, Joseph PG, Gandhi M, Dennis BB, et al. Assessing the quality of published genetic association studies in meta-analyses: the quality of genetic studies (Q-Genie) tool. *BMC Genet* 2015;16:50.
- Richard MA, Lupo PJ, Morton LM, Yasui YA, Sapkota YA, Arnold MA, et al. Genetic variation in POT1 and risk of thyroid subsequent malignant neoplasm: a report from the Childhood Cancer Survivor Study. *PLoS One* 2020;15:e0228887.
- Gramatges MM, Morton LM, Yasui Y, Arnold MA, Neglia JP, Leisenring WM, et al. Telomere length-associated genetic variants and the risk of thyroid cancer in survivors of childhood cancer: a report from the Childhood Cancer Survivor Study (CCSS). *Cancer Epidemiol Biomarkers Prev* 2019;28:417–9.
- Wang X, Sun C-L, Hageman L, Smith K, Singh P, Desai S, et al. Clinical and genetic risk prediction of subsequent CNS tumors in survivors of childhood cancer: a report from the COG ALTE03N1 study. *J Clin Oncol* 2017;35: 3688–96.
- Applebaum MA, Vaksman Z, Lee SM, Hungate EA, Henderson TO, London WB, et al. Neuroblastoma survivors are at increased risk for second malignancies: a report from the International Neuroblastoma Risk Group Project. *Eur J Cancer* 2017;72:177–85.
- Vodusek AL, Goricar K, Gazic B, Dolzan V, Jazbec J. Antioxidant defence-related genetic variants are not associated with higher risk of secondary thyroid cancer after treatment of malignancy in childhood or adolescence. *Radiol Oncol* 2016;50:80–6.

A comprehension of germline predisposition to SMNs can significantly influence personalized clinical practices. It will allow for a more precise delineation of the risks and benefits associated with different treatment strategies. Additionally, the identification of genetic variants that can modify the risk of SMNs induced by chemotherapy and radiation therapy suggests integrating these genetic factors into the risk score for SMNs. This will help for the stratification of screening procedures for a more proactive approach to patient follow-up.

In conclusion, this systematic review represents a valuable tool for researchers, as it compiles all known genetic variants and genes associated with SMN risk in CCS. This will facilitate their efficient replication, thus contributing to the advancement of our understanding of the genetic basis of this important risk for survivors of childhood cancer.

Authors' Disclosures

No disclosures were reported.

Acknowledgments

C. Ducos received grants from the “Fondation ARC pour la Recherche sur le Cancer” and the Doctoral School of Public Health, both providing doctoral fellowships for C. Ducos. F. Rosselli received grants from “SIRIC EpiCure: INCA-DGOS-Inserm-ITMO Cancer_18002.”

Note

Supplementary data for this article are available at *Cancer Epidemiology, Biomarkers & Prevention Online* (<http://cebp.aacrjournals.org/>).

Received January 3, 2024; revised March 7, 2024; accepted May 22, 2024; published first May 27, 2024.

19. Erčulj N, Faganel Kotnik B, Debeljak M, Jazbec J, Dolžan V. DNA repair polymorphisms influence the risk of second neoplasm after treatment of childhood acute lymphoblastic leukemia. *J Cancer Res Clin Oncol* 2012;138:1919–30.
20. Ma YP, Van Leeuwen FE, Cooke R, Broeks A, Enciso-Mora V, Olver B, et al. FGFR2 genotype and risk of radiation-associated breast cancer in Hodgkin lymphoma. *Blood* 2012;119:1029–31.
21. Stanulla M, Schaeffeler E, Möricke A, Coultard SA, Cario G, Schrauder A, et al. Thiopurine methyltransferase genetics is not a major risk factor for secondary malignant neoplasms after treatment of childhood acute lymphoblastic leukemia on Berlin-Frankfurt-Münster protocols. *Blood* 2009;114:1314–8.
22. Mertens AC, Mitby PA, Radloff G, Jones IM, Perentesis J, Kiffmeyer WR, et al. XRCC1 and glutathione-S-transferase gene polymorphisms and susceptibility to radiotherapy-related malignancies in survivors of Hodgkin disease. *Cancer* 2004;101:1463–72.
23. Blanco JG, Edick MJ, Hancock ML, Winick NJ, Dervieux T, Amylon MD, et al. Genetic polymorphisms in CYP3A5, CYP3A4 and NQO1 in children who developed therapy-related myeloid malignancies. *Pharmacogenetics* 2002;12:605–11.
24. Woo M, Shuster J, Chen C, Bash R, Behm F, Camitta B, et al. Glutathione S-transferase genotypes in children who develop treatment-related acute myeloid malignancies. *Leukemia* 2000;14:232–7.
25. Best T, Li D, Skol AD, Kirchhoff T, Jackson SA, Yasui Y, et al. Variants at 6q21 implicate PRDM1 in the etiology of therapy-induced second malignancies after Hodgkin's lymphoma. *Nat Med* 2011;17:941–3.
26. Morton LM, Sampson JN, Armstrong GT, Chen T-H, Hudson MM, Karlins E, et al. Genome-wide association study to identify susceptibility loci that modify radiation-related risk for breast cancer after childhood cancer. *J Natl Cancer Inst* 2017;109:djx058.
27. Opstal-van Winden AWJ, de Haan HG, Hauptmann M, Schmidt MK, Broeks A, Russell NS, et al. Genetic susceptibility to radiation-induced breast cancer after Hodgkin lymphoma. *Blood* 2019;133:1130–9.
28. Morton LM, Karyadi DM, Hartley SW, Frone MN, Sampson JN, Howell RM, et al. Subsequent neoplasm risk associated with rare variants in DNA damage response and clinical radiation sensitivity syndrome genes in the Childhood Cancer Survivor Study. *JCO Precis Oncol* 2020;4:PO.20.00141.
29. Qin N, Wang Z, Liu Q, Song N, Wilson CL, Ehrhardt MJ, et al. Pathogenic germline mutations in DNA repair genes in combination with cancer treatment exposures and risk of subsequent neoplasms among long-term survivors of childhood cancer. *J Clin Oncol* 2020;38:2728–40.
30. Wang Z, Liu Q, Wilson CL, Easton J, Mulder H, Chang T-C, et al. Polygenic determinants for subsequent breast cancer risk in survivors of childhood cancer: the St Jude Lifetime Cohort Study (SJLIFE). *Clin Cancer Res* 2018;24:6230–5.
31. Wang Z, Wilson CL, Easton J, Thrasher A, Mulder H, Liu Q, et al. Genetic risk for subsequent neoplasms among long-term survivors of childhood cancer. *J Clin Oncol* 2018;36:2078–87.
32. Mandelbaum J, Bhagat G, Tang H, Mo T, Brahmachary M, Shen Q, et al. BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell like diffuse large B cell lymphoma. *Cancer Cell* 2010;18:568–79.
33. Elsir T, Smits A, Lindström MS, Nistér M. Transcription factor PROX1: its role in development and cancer. *Cancer Metastasis Rev* 2012;31:793–805.
34. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;11:446–50.
35. Catucci I, Osorio A, Arver B, Neidhardt G, Bogliolo M, Zanardi F, et al. Individuals with FANCM biallelic mutations do not develop Fanconi anemia, but show risk for breast cancer, chemotherapy toxicity and may display chromosome fragility. *Genet Med* 2018;20:452–7.
36. Keijzers G, Bakula D, Petr MA, Madsen NGK, Teklu A, Mkrtyan G, et al. Human exonuclease 1 (EXO1) regulatory functions in DNA replication with putative roles in cancer. *Int J Mol Sci* 2018;20:74.
37. Goellner EM, Putnam CD, Kolodner RD. Exonuclease 1-dependent and independent mismatch repair. *DNA Repair* 2015;32:24–32.

1 Rare Variants Associated with Pediatric Cancer Treatment-Related 2 Second Malignant Neoplasm Risk.

3 Claire Ducos^{1,2,3}, Brice Fresneau^{1,2,3,4}, Filippo Rosselli⁵, Giau Vu-Bezin^{1,2,3}, Boris
4 Schwartz^{1,2,3}, Rodrigue S. Allodji^{1,2,3}, Gaëlle Marenne⁶, Thomas E. Ludwig^{6,7}, Anne
5 Boland⁸, Jean-François Deleuze⁹, Helene Blanché⁹, Robert Olasso⁹, Julie Nys^{1,2,3},
6 Sarah Winter¹⁰, Franck Bourdeau¹⁰, Chiraz El-Fayech⁴, Carole Rubino^{1,2,3}, Ibrahima
7 Diallo¹¹, Florent de Vathaire^{1,2,3}, Simone Benhamou^{*2,3,12}, Nadia Haddy^{*1,2,3}

- 8 1. Radiation Epidemiology Team, Center for Research in Epidemiology and Population Health,
9 INSERM Unit 1018, Villejuif, France.
- 10 2. Research Department, Gustave Roussy, Villejuif, France.
- 11 3. University Paris Saclay, Villejuif, France.
- 12 4. Department of Children and Adolescents Oncology, Gustave Roussy Cancer Campus,
13 Université Paris-Saclay, Villejuif, France.
- 14 5. CNRS UMR9019, Gustave Roussy Cancer Campus, Université Paris-Saclay, Equipe
15 Labellisée Ligue Nationale Contre le Cancer, 94801 Villejuif, France.
- 16 6. Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France.
- 17 7. CHRU Brest, F 29200 Brest, France
- 18 8. Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine
19 (CNRGH), 91057, Evry, France
- 20 9. Centre d'Etude du Polymorphisme Humain, Fondation Jean Dausset, 75010 Paris, France.
- 21 10. SIREDO Paediatric Cancer Center, Institut Curie, Paris, France; INSERM U830, Laboratory
22 of Translational Research in Paediatric Oncology, Institut Curie, Paris, France; Paris
23 Sciences Lettres Research University, Paris, France.
- 24 11. Molecular Radiation Therapy and Therapeutic Innovation, INSERM U1030, Villejuif, France.
- 25 12. Oncostat Team, Center for Research in Epidemiology and Population Health, INSERM Unit
26 1018, University Paris Saclay, Gustave Roussy Cancer Campus, Villejuif, France.

27 * Equal Contribution

28 **Corresponding author:**

29 Claire Ducos

30 Radiation Epidemiology team in CESP, INSERM

31 B2M, 32 rue Camille Desmoulins, Villejuif 94800, France

32 Claire.ducos@gustaveroussy.fr

33 Phone: (+33) 6 89 31 68 68

34 **Running head:** Genetic risk of second cancer in childhood cancer survivors.

35 **Conflict of interest:** The authors declare no conflict of interest.

36 **Fundings:** C. Ducos received grants from the “Fondation ARC pour la Recherche sur
37 le Cancer” and the Doctoral School of Public Health (EDSP), both providing doctoral
38 fellowships for Claire Ducos. F. Rosselli received grants from “SIRIC EpiCure: INCa-
39 DGOS-Inserm-ITMO Cancer_18002”.

40 **Word counts:** abstract 267/275, main text 2982/3000

41

42 **Abstract**

43 **Purpose.** Survivors of childhood cancers are at risk of subsequent malignant
44 neoplasms (SMN). Exposure to radiation therapy and different chemotherapy agents
45 are known risk factors for SMN development. However, there remains interindividual
46 variability that could be attributed to genetic variations. Our study aims to identify rare
47 genetic variants associated with SMN risk and assess the influence of treatment on
48 the associated risk of these variants.

49 **Patients and Methods.** We conducted a whole-exome sequencing in a nested case-
50 control study within the French Childhood Cancer Survivors Study for 450 survivors
51 including 163 cases and 287 controls. Rare variants in genes within replication,
52 recombination and repair pathways were selected. Gene-based association tests were
53 conducted and a logistic regression analysis was used to estimate the effect of
54 selected genes on SMN risk, adjusting for relevant clinical variables. Specific analysis
55 restricted to breast and thyroid SMN were also performed. Stratified analyses based
56 on the radiation doses received by the bone marrow and specific organs were also
57 conducted.

58 **Results.** We have identified several genetic associations: the *RNASEL* and
59 *APOBEC3F* gene (Odds Ratio [OR]=5.43, 95% confidence interval [CI]: 1.76-20.44,
60 $p=0.0055$ and $OR=4.8$, 95% CI: 1.28-22.95, $p=0.027$, respectively) are significantly
61 associated with the risk of SMN, while the *FANCM* gene ($OR=4.65$, 95%CI: 1.06-
62 21.56, $p=0.041$) is associated with the risk of developing breast SMN.

63 **Conclusion.** This study provides new evidence regarding the involvement of genetics
64 in the development of SMN. Additional research is needed to confirm these results and
65 uncover the underlying mechanisms behind these associations. If replicated, these

66 findings will help identify survivors at higher risk of developing SMN, enabling tailored
67 treatment and follow-up strategies.

68

69 **Keyword.** Childhood Cancer survivors, Second Malignant Neoplasm, Genetic, Whole
70 Exome Sequencing, rare variants

71 **Introduction**

72 Thanks to advances in treatment and improved diagnosis, 5-year childhood cancer
73 survival increased over time, and now reaches 80% in most European countries (1).

74 Accordingly, the number of childhood cancer survivors (CCS) has risen, reaching an
75 estimated half a million in 2020 in Europe (2). However, two-thirds of these survivors
76 experience late complications associated to their primary cancer and its treatment (3).
77 Among these, the risk of developing subsequent malignant neoplasms (SMN) is three
78 to five times greater than that of the general population (4).

79 Several studies have shown a relationship between SMN development and treatment
80 received for the pediatric cancer, especially radiotherapy and chemotherapy (5–9). A
81 dose-response relationship between SMN risk and radiation dose received has been
82 established for several SMN types such as breast, or meningioma and sarcoma (9).
83 Most chemotherapy agents have also been associated with the risk of several types of
84 SMN. For instance, exposure to doxorubicin has been associated with subsequent
85 solid cancers and breast SMN, while exposure to cyclophosphamide has been
86 associated with subsequent sarcomas (8,10,11). While treatment is a recognized risk
87 factor of SMN, there is evidence of inter-individual variability in treatment response,
88 suggesting a genetic component (12).

89 DNA repair systems are crucial for preventing cancer by maintaining genetic integrity.
90 Ionizing radiation induces severe DNA damage, which, if not repaired efficiently, can
91 lead to variations and chromosomal abnormalities, resulting in cancer development
92 (13). Genetic variations within DNA repair genes, combined with chemotherapy or
93 radiotherapy, could modify the risk of SMN, which represent a significant challenge in
94 understanding SMN risk among CCS.

95 Several genome-wide association studies (GWAS) (14,15) and candidate gene studies
96 (16–26) have identified common variants primarily located in DNA repair or drug
97 metabolism pathways associated with the SMN risk following childhood cancer
98 treatment.

99 Rare and highly penetrant variants may also contribute to genetic susceptibility (27).

100 Three recent US studies (28–31) focused on specific pathways related to treatment-
101 associated cancers, particularly DNA repair pathways and cancer predisposition
102 genes. Stratified analyses based on the treatments received (15,28–30) suggested a
103 link between genetics and the treatment received for the first cancer, particularly
104 radiotherapy. However, limited numbers of studies exist on SMN risk in CCS, with few
105 using NGS to explore especially rare variants.

106 We conducted a discovery nested case-control study within the French Childhood
107 Cancer Survivors Study (FCCSS) cohort to identify rare genetic variants within the
108 replication, recombination, and repair pathways associated with SMN risk. Additionally,
109 we explored the impact of the received treatment on the risk associated with these
110 variants.

111 **Materials and Methods**

112 **Study participants**

113 **French Childhood Cancer Survivors Study (FCCSS) Cohort**

114 The FCCSS (<https://fccss.fr/>) aims to investigate any aspect of long-term childhood
115 cancer survivorship. The cohort has been previously described [31]. Briefly, it includes
116 7,670 patients treated for any malignancy except leukemia before the age of 21

117 between 1945 and 2001 in five French cancer centers and who survived at least 5
118 years after diagnosis.

119 Clinical and treatment data for the primary cancer, along with information on 20 genetic
120 syndromes (Table S1), were obtained from patient medical records.

121 SMN were identified through comprehensive follow-up, including medical reports, long-
122 term care, self-reported questionnaires, and linkage with the National Health Data
123 System (SNDS), followed by validation by a trained nurse. (33,34)

124 Radiation dose distributions were retrospectively reconstructed using patient-specific
125 voxel phantoms (35) and detailed treatment information, with an average dose
126 calculated for each organ [5]. Cumulative chemotherapy doses were determined in
127 grams per square meter of total body surface area and categorized into five major
128 classes according to their specific mechanisms of action (36).

129 A nested biobank within FCCSS of blood samples was established for 1,571 living
130 patients for whom we have valid contact address and who give a signed consent.
131 Lymphocytes were isolated from the blood samples and lymphoblastoid cell lines were
132 generated through Epstein-Barr virus infection either at GENETHON (Evry, France) or
133 at the CEPH-Biobank (Paris, France).

134 The FCCSS study has been approved by the French regulatory authorities and ethics
135 committee. Patients or their legal representatives have provided written informed
136 consent.

137 **Nested case-control study population**

138 A nested case-control study was conducted within the FCCSS cohort, focusing on
139 patients who provided a blood sample as of the date in 2019. Cases were defined as

140 individuals who developed SMN during the long-term follow-up. One to three controls
141 who did not develop an SMN were matched to each case on sex, age at first cancer,
142 year of diagnosis, type of first cancer, and length of follow-up, ensuring it was at least
143 as long as necessary to develop an SMN in the corresponding case. A total of 163
144 cases and 287 controls were considered for the study.

145

146 **Sequencing and bioinformatic analysis**

147 **Genetic data**

148 Exome sequencing was performed at the CEA-CNRGH using DNA extracted from
149 lymphocytes at the CEPH-Biobank (Fondation Jean Dausset-CEPH, Paris) using a salt
150 precipitation method automated on the Autopure (Qiagen). Library preparation was
151 performed using the SureSelect "All Exon V6" capture technology (Agilent
152 Technologies, CA, USA). These libraries were sequenced using the HiSeq 4000
153 sequencing kit (Illumina, CA, USA) to achieve an average coverage of at least 70X per
154 sample.

155 **Selection of 3 Replication, Repair or Recombination (3R) pathways genes**

156 Genes belonging to DNA replication, recombination, and/or repair (3R) pathways were
157 selected using Gene Ontology (GO), Molecular Signature Database (MSigDb) (37) and
158 Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Intersection of these
159 databases led to a list of 1,643 genes. Each gene was then reviewed by a biologist
160 specializing in 3R pathways (FP). The scoring system was established as follows: each
161 time a gene was identified in one of the three pathways by any of the three databases,

162 a score of one was assigned. Genes with a score of 2 or more were retained for
163 analyses, i.e. 969 genes (Figure 1, Table S2).

164 **Bioinformatic analysis**

165 All the bioinformatic steps are summarized in the supplementary material workflow
166 (Figure S2). High-quality paired-end reads were aligned to the GRCh38 human
167 genome reference using BWA-MEM (v1.9) (38). Variant calling was performed using
168 HaplotypeCaller from the Genome Analysis ToolKit (GATK, v4) (39) to generate gVCF
169 files followed by the generation of VCF files using GATK tools. Annotation was
170 performed with the VEP tool (40) and filtered using VQSR tools from GATK and the
171 RAVAQ package (41). Bioinformatic analysis was performed using R v4.1.2 and
172 Python v3.8.3 with Snakemake (42) pipelines.

173 Quality control procedures were applied to both variants and individuals using the
174 RAVAQ R package. Non-synonymous variants with insufficient sequencing depth (<10)
175 or low call rate (<95%) were excluded. Individuals with high heterozygosity,
176 relatedness >0.185, inbreeding rates >4% and low call rate (>90%) were
177 excluded.(41). Additionally, only rare variants, with a minor allele frequency (MAF)
178 below 1% in the gnomAD database (v3.1.2) (43) with predicted "high" or "moderate"
179 impact on gene function were retained.

180 **Statistical analysis**

181 Analysis were restricted to individuals of European descent by using a principal
182 component analysis (PCA) where common variants were matched to the 1000
183 Genomes (1000G) data (44). Therefore, we retained only the 395 individuals, including
184 141 cases and 254 controls, who successfully passed the quality control criteria. We
185 performed subgroup analyses for SMN breast and SMN thyroid cancers, as they are

186 the most prevalent both in the general population and in our study, with 48 and 69
187 cases, respectively. We also conducted sensitivity analysis including only patients
188 without any known genetic predisposition syndrome (381 patients).

189 All rare variants were assigned to genes within the predefined 3R gene list. Three
190 gene-based association tests were conducted (45), including SKAT (46), a burden test
191 (47), and SKAT-O (48). To ensure robust findings, only genes with a p-value < 0.05 in
192 at least two of these tests were selected for further analysis.

193 Due to the patient exclusions we had to forego the matching. Subsequently, the
194 statistical analyses were adjusted to account for the matching variables. The effect of
195 the selected genes on SMN risk was estimated using logistic regression adjusting for
196 sex, age at diagnosis of the first cancer, year of diagnosis, type of first cancer, length
197 of follow-up, and the first three principal components of PCA.

198 Stratified analyses were conducted according to the radiation dose received by the
199 bone marrow for any SMN, by breast for breast SMN and by the thyroid for thyroid
200 SMN analysis. For each analysis, two classes were defined using the mean radiation
201 dose received at each organ by controls as cut-off.

202 **Results**

203 Characteristics of the 141 SMN cases and 254 controls are described in Table 1.
204 Among the 141 cases, the SMN was diagnosed in thyroid (41%), breast (31%), bone
205 (11%), kidney and bladder (10%), and colon/rectum (7%).

206 The median age at diagnosis of the first cancer was 8 years (IQR: 4-12), and the
207 median follow-up duration was 25 years (IQR: 17-34) for both cases and controls
208 (Table 1).

209 **Variants identified through whole exome sequencing.**

210 A total of 1,215,288 variants in the 969 genes selected were identified. After exclusion
211 of common variants (defined as those with a MAF>1%) 67,706 rare variants were
212 detected. When focusing on genes in the 3R pathways, 3,314 variants with a high or
213 moderate effect across 583 genes were identified (Figure 1).

214 **Gene-based tests**

215 Gene-based results for genes identified by at least two rare variant tests methods are
216 presented in Table S4. Twenty genes were retained for further analyses and
217 individually included in a multivariate logistic regression. For breast and thyroid SMN,
218 eighteen and twenty-two genes, respectively, were considered (Figure 2B and 2C).

219 **Comprehensive model**

220 All statistically significant findings from the multivariate logistic regression are listed in
221 Table 2, with comprehensive results provided in Table S5.

222 Any SMN

223 We identified seven genes associated with an increased risk of any SMN in multivariate
224 models (Table 2): *APOBEC3F* (OR=4.8, 95% CI: 1.28-22.95, p=0.027), *DBF4*
225 (OR=4.10, 95% CI: 1.31-14.04, p=0.016), *ATR* (OR=3.13, 95% CI: 1.17-8.95,
226 p=0.026), *EGFR* (OR=10.67, 95%CI: 1.6-209.6, p=0.034), *RNASEL* (OR=5.43, 95%
227 CI: 1.76-20.44, p=0.0055), *HINFP* (OR=3.41, 95% CI: 1.06-12.03, p=0.043) and *XPC*
228 (OR=4.82, 95% CI: 1.49-18.57, p=0.012). The number of rare variants identified in
229 each gene and the number of cases and controls carrying at least one variant is shown
230 in Table 2.

231 Breast SMN

232 Eight genes were significantly associated with the risk of breast cancer in CCS: *BLM*
233 (OR=5.21, 95%CI: 1.5-19.11, p=0.0096) and *FANCM* (OR=4.65, 95%CI: 1.06-21.56,
234 p=0.041), *FAN1* (OR=4.27, 95%CI: 1.26-14.26, p=0.018), *PARP1* (OR=7.79, 95%CI:
235 1.15-22.04, p=0.034), *CTBP2* (OR=9.13, 95%CI :1.73-51.47, p=0.009), *ORC5*
236 (OR=18.41, 95%CI: 1.9-418.21, p=0.02), *ANXA6* (OR=16.63, 95%CI: 2.35-177.44,
237 p=0.008) and *PML* (OR=5.02, 95%CI:1.01-27.82, p=0.048). All these results are
238 presented in Table 2.

239 Thyroid SMN

240 The *HINFP* gene was also associated with the risk of thyroid cancer in CCS (OR=5.43,
241 95% CI: 1.26-23.5, p=0.022). In addition, there are five other genes with an
242 interpretable OR: *HIRA* (OR=8.46, 95% CI: 1.06-89.81, p=0.047), *IFI16* (OR=3.49,
243 95% CI: 1.28-9.32, p=0.013), *MCM5* gene (OR=5.7, 95% CI: 1.36-23.87, p=0.015) and
244 *RANBP2* (OR=8.68, 95% CI: 1.98-40.2, p=0.004). All these results are presented in
245 Table 2.

246 **Stratified analysis by radiation doses**

247 Stratified analyses by radiation doses are presented in Table S6. Despite the small,
248 some associations were suggested. For any SMN, patients who received low doses of
249 radiation (5Gy) to the bone marrow had an increased risk associated with the *XPC*
250 gene (OR=9.26, 95%CI: 1.62-77.92). In stratified analysis for breast SMN, among
251 individuals who received doses of radiation to the breast less than 4.7 Gy, the risk of
252 breast SMN was associated with *CTPBP2* (OR=32.23, 95% CI: 3.11-470.09), *BLM*
253 (OR=6.39, 95% CI: 1.46-29.07), *FAN1* (OR=7.41, 95% CI: 1.49-40.75), *PARP1*
254 (OR=8.9, 95% CI: 1.38-108.86) and *PML* (OR=10.63, 95% CI: 1.38-108.86) (Table 2).
255 Three genes showed potential associations with increased risk of thyroid SMN in

256 individuals exposed to lower thyroid doses (6.5 Gy): *HINFP* (OR=11.83, 95%CI: 1.47-
257 131.78), *IFI16* (OR=9.27, 95%CI: 2.7-33.77) and *RANBP2* (OR=32.33, 95%CI: 3.84-
258 348.64).

259 The findings of this study remain consistent following a sensitivity analysis that
260 excluded patients with a known genetic predisposition syndrome (Table S3).

261 **Discussion**

262 In this case-control study nested within the FCCSS, we investigated rare variants in
263 969 genes involved in DNA repair, replication, and recombination pathways and found
264 significant association associated with SMN risk. Notably, *APOBEC3F* and *RNASEL*
265 genes associated to overall SMN risk, *FANCM* with breast SMN, and *HINFP* and *IFI16*
266 genes with thyroid SMN, consistent with prior research.

267 *RNASEL* was associated with a 5-fold increased risk of SMN in our study, aligning with
268 its established role in carcinogenesis and its association with the prostate, breast, and
269 pancreatic cancers (49–52).

270 The identified rare variants rs151296858 and rs74315364, were exclusively found in
271 three cases and are in complete linkage disequilibrium, as confirmed by Rökman and
272 colleagues (51). The variant rs74315364 corresponds to a stop-gained variant
273 associated with prostate (50,52), pancreatic (49) and breast cancers.

274 All these studies emphasize the importance of this gene and its rare variations in
275 cancer development, highlighting the need for further investigation to confirm their
276 involvement in the risk of SMN among CCS.

277 *APOBEC3F* has been associated with breast cancer risk (53). In our study, only two
278 variants were found in seven cases and three controls. The variant rs146924719 was

279 present in all the cases, highlighting its important in the search for rare variants
280 associated with SMN risk among CCS. While our subgroup analysis was limited by
281 small sample sizes, it suggests a potential association between *APOBEC3F* and
282 breast SMN. These promising results have not been observed in previously published
283 NGS studies in CCS (28,29,31), necessitating further analysis.

284 These two genes represent novel findings, as they have not been identified in other
285 genetic association studies on the risk of SMN among CCS. This could be attributed
286 to the use of rare variant tests in our study (SKAT, SKATO, Burden) or methodological
287 differences like sequencing types (WES/WGS) and gene selection.

288 Our study also supports the association between the *FANCM* gene and breast SMN,
289 consistent with Morton et al. (29), who identified *FANCM* as strongly associated with
290 the risk of out-of-field radiotherapy-related SMN (RT-SN), including breast SMN. This
291 suggests that *FANCM* loss could result in reduced homologous recombination repair
292 efficiency.

293 This gene has been associated with hereditary breast cancer (54–56), triple-negative
294 breast cancer (57,58), and ovarian cancer (59,60). We also found an increased breast
295 SMN risk in patients exposed to high breast radiation doses while no association was
296 found at low doses.

297 Due to the limited sample size of our study, exploring gene-chemotherapy interactions
298 was unattainable as it was for the study of Morton and colleagues (29). However,
299 investigating the impact of chemotherapy on the gene association with SMN in CCS,
300 is important for the future research. Catucci et al. (61), support this by showing that
301 individuals with homozygous loss of function variations have a heightened
302 predisposition to cancer and increased chemotherapy toxicity.

303 We also found in our study on breast SMN several well-known cancer predisposition
304 genes, including *BLM*, associated with Bloom syndrome, and *PML*, genes involved in
305 numerous cellular processes (62).

306 Our study showed interesting associations between some genes and thyroid SMN.
307 *HINFP* was associated with an increased risk of both any SMN as thyroid specific SMN.
308 Furthermore, a rare variant in this gene, rs61753080, is particularly interesting as it
309 was present in five cases, including four cases of thyroid SMN. This variant has also
310 been previously associated with the risk of aggressive prostate cancer (63).

311 Additionally, *IFI16* was associated with an increased risk of thyroid SMN when
312 considering only patients who received high doses of radiation. This gene may have a
313 significant impact due to its direct interactions with *TP53*, well-known cancer
314 predisposition gene (64).

315 In contrast to the findings reported by Morton et al. (29), our study did not find a
316 significant association between the *EXO1* or *NEIL3* gene and SMN risk. These
317 differences could be attributed to variations in study populations and analysis
318 methodologies.

319 These observations highlight the importance of conducting context-specific genetic
320 investigations to fully comprehend the implications of genetic variations on SMN risks.
321 Further research is warranted to explore and validate the role of these previously
322 identified genes in SMN risk and to identify potential factors contributing to their lack of
323 association in our study.

324 While our study offers novel insights into rare variants associated with SMN in CCS,
325 certain limitations must be considered. The small sample size may be underpowered
326 to detect associations at exome-wide significance. Additionally, the focus on European

327 descent individuals, due to population stratification, limits the generalizability of our
328 findings and further reducing our size sample. Although our sample size was limited,
329 our results suggest a potential relationship between received radiation therapy and
330 specific genetic variants. However, the limited sample size restricted further analysis.
331 Conducting external validation would have been beneficial to strengthen our findings,
332 however, external datasets were not available for this analysis. Nevertheless, our
333 methodology was robust, performing three rare variant tests (SKAT, SKAT-O, and
334 Burden) and considering their intersection to reduce the rate of false positives.

335 **Conclusion and futures perspectives**

336 This study provides new insights into the risk of any SMN, as well as breast and thyroid
337 risks, in CSS, identifying rare genetic variants for genes involved in the 3R pathways.
338 It also confirmed the role of *FANCM* in the occurrence of breast cancer. Underling the
339 importance of considering both the genetic background and the treatment history of
340 CCS when assessing their risk for developing SMNs.

341 Future research requires validation studies in independent cohorts to confirm these
342 results, as well as exploring the biological impact of identified genetic variants through
343 functional analysis. International collaborations will be necessary, enabling sufficient
344 sample size to achieve exome-wide significance and to conduct interaction studies with
345 radiotherapy/chemotherapy as modifiers of genetic risk. With robust replication, the
346 findings of this study hold significant potential for clinical translation. This could lead to
347 better identification of patients at high risk of SMN providing an opportunity to
348 personalized treatment by considering the balance of benefits and risks. Additionally,
349 it could facilitate the implementation of personalized cancer surveillance programs to
350 reduce the long-term morbidity and mortality for CCS.

351 **References**

- 352 1. Botta L, Gatta G, Capocaccia R, et al.: Long-term survival and cure fraction estimates for
353 childhood cancer in Europe (EUROCARE-6): results from a population-based study.
354 *Lancet Oncol.* 23:1525–36., 2022
- 355 2. Vassal G, Schrappe M, Pritchard-Jones K, et al.: The SIOPE strategic plan: A European
356 cancer plan for children and adolescents. *J Cancer Policy* 8:17–32, 2016
- 357 3. Grabow D, Kaiser M, Hjorth L, et al.: The PanCareSurFup cohort of 83,333 five-year
358 survivors of childhood cancer: a cohort from 12 European countries. *Eur J Epidemiol*
359 33:335–49, 2018
- 360 4. Mertens AC, Liu Q, Neglia JP, et al.: Cause-Specific Late Mortality Among 5-Year
361 Survivors of Childhood Cancer: The Childhood Cancer Survivor Study. *J Natl Cancer Inst*
362 100:1368–79, 2008
- 363 5. Allodji RS, Tucker MA, Hawkins MM, et al.: Role of radiotherapy and chemotherapy in
364 the risk of leukemia after childhood cancer: An international pooled analysis. *Int J Cancer*
365 148:2079–89, 2021
- 366 6. Allodji RS, Haddy N, Vu-Bezin G, et al.: Risk of subsequent colorectal cancers after a
367 solid tumor in childhood: Effects of radiation therapy and chemotherapy. *Pediatr Blood*
368 *Cancer* 66:e27495, 2019
- 369 7. Turcotte LM, Neglia JP, Reulen RC, et al.: Risk, Risk Factors, and Surveillance of
370 Subsequent Malignant Neoplasms in Survivors of Childhood Cancer: A Review. *J Clin*
371 *Oncol.* 36:2145–52, 2018
- 372 8. Turcotte LM, Whitton JA, Friedman DL, et al.: Risk of Subsequent Neoplasms During the
373 Fifth and Sixth Decades of Life in the Childhood Cancer Survivor Study Cohort. *J Clin*
374 *Oncol* 33:3568–75, 2015
- 375 9. Inskip PD, Sigurdson AJ, Veiga L, et al.: Radiation-Related New Primary Solid Cancers
376 in the Childhood Cancer Survivor Study: Comparative Radiation Dose Response and
377 Modification of Treatment Effects. *Int J Radiat Oncol Biol Phys* 94:800–7, 2016
- 378 10. Teepen JC, van Leeuwen FE, Tissing WJ, et al.: Long-Term Risk of Subsequent
379 Malignant Neoplasms After Treatment of Childhood Cancer in the DCOG LATER Study
380 Cohort: Role of Chemotherapy. *J Clin Oncol Off* 5:2288–98, 2017
- 381 11. Henderson TO, Moskowitz CS, Chou JF, et al.: Breast Cancer Risk in Childhood Cancer
382 Survivors Without a History of Chest Radiotherapy: A Report From the Childhood Cancer
383 Survivor Study. *J Clin Oncol* 34:910–8, 2016
- 384 12. Armenian SH, Bhatia S.: Chronic Health Conditions in Childhood Cancer Survivors: Is it
385 All Treatment-Related—or Do Genetics Play a Role? *J Gen Intern Med* 24:395, 2009
- 386 13. Hoeijmakers JHJ.: Genome maintenance mechanisms for preventing cancer. *Nature*
387 411:366–74, 2001

- 388 14. Best T, Li D, Skol AD, et al.: Variants at 6q21 implicate PRDM1 in the etiology of
389 therapy-induced second malignancies after Hodgkin's lymphoma. *Nat Med* 17:941–
390 3,2017
- 391 15. Morton LM, Sampson JN, Armstrong GT, et al.: Genome-Wide Association Study to
392 Identify Susceptibility Loci That Modify Radiation-Related Risk for Breast Cancer After
393 Childhood Cancer. *J Natl Cancer Inst* 109(11):djj058, 2017
- 394 16. Richard MA, Lupo PJ, Morton LM, et al.: Genetic variation in POT1 and risk of thyroid
395 subsequent malignant neoplasm: A report from the Childhood Cancer Survivor Study.
396 *PLOS ONE* 15:e0228887, 2020
- 397 17. Gramatges MM, Morton LM, Yasui Y, et al.: Telomere length-associated genetic variants
398 and the risk of thyroid cancer in survivors of childhood cancer: A report from the
399 Childhood Cancer Survivor Study (CCSS). *Cancer Epidemiol Biomarkers Prev*
400 28(2):417-419, 2019
- 401 18. Wang X, Sun C-L, Hageman L, Smith K, et al.: Clinical and Genetic Risk Prediction of
402 Subsequent CNS Tumors in Survivors of Childhood Cancer: A Report From the COG
403 ALTE03N1 Study. *J Clin Oncol* 35:3688–96, 2017
- 404 19. Vodusek AL, Goricar K, Gazic B, et al.: Antioxidant defence-related genetic variants are
405 not associated with higher risk of secondary thyroid cancer after treatment of malignancy
406 in childhood or adolescence. *Radio Oncol* 50(1):80-86, 2016
- 407 20. Applebaum MA, Vaksman Z, Lee SM, et al.: Neuroblastoma survivors are at increased
408 risk for second malignancies: A report from the International Neuroblastoma Risk Group
409 Project. *Eur J Cancer* 72:177-185, 2017
- 410 21. Ma YP, van Leeuwen FE, Cooke R, et al.: FGFR2 genotype and risk of radiation-
411 associated breast cancer in Hodgkin lymphoma. *Blood* 119:1029–31, 2012
- 412 22. Erčulj N, Faganel Kotnik B, Debeljak M, et al.: DNA repair polymorphisms influence the
413 risk of second neoplasm after treatment of childhood acute lymphoblastic leukemia. *J*
414 *Cancer Res Clin Oncol* 138:1919–30, 2012
- 415 23. Stanulla M, Schaeffeler E, Möricke A, et al.: Thiopurine methyltransferase genetics is not
416 a major risk factor for secondary malignant neoplasms after treatment of childhood acute
417 lymphoblastic leukemia on Berlin-Frankfurt-Münster protocols. *Blood* 114:1314–8, 2009
- 418 24. Mertens AC, Mitby PA, Radloff G, et al.: XRCC1 and glutathione-S-transferase gene
419 polymorphisms and susceptibility to radiotherapy-related malignancies in survivors of
420 Hodgkin disease: A report from the Childhood Cancer Survivor Study. *Cancer*
421 101(6):1463-1472, 2004
- 422 25. Blanco JG, Edick MJ, Hancock ML, et al.: Genetic polymorphisms in CYP3A5, CYP3A4
423 and NQO1 in children who developed therapy-related myeloid malignancies.
424 *Pharmacogenet Genomics* 12:605, 2002
- 425 26. Woo MH, Shuster JJ, Chen C, et al.: Glutathione S-transferase genotypes in children who
426 develop treatment-related acute myeloid malignancies. *Leukemia* [Internet]. *Leukemia*
427 14(2):232-237, 2000

- 428 27. Eichler EE, Flint J, Gibson G, et al.: Missing heritability and strategies for finding the
429 underlying causes of complex disease. *Nat Rev Genet* 11(6):446-450, 2010
- 430 28. Qin N, Wang Z, Liu Q, et al.: Pathogenic Germline Mutations in DNA Repair Genes in
431 Combination With Cancer Treatment Exposures and Risk of Subsequent Neoplasms
432 Among Long-Term Survivors of Childhood Cancer. *J Clin Oncol* 38:2728–40,2020
- 433 29. Morton LM, Karyadi DM, Hartley SW, et al.: Subsequent neoplasm risk associated with
434 rare variants in DNA damage response and clinical radiation sensitivity syndrome genes
435 in the childhood cancer survivor study. 4:PO.20.00141, 2020
- 436 30. Wang Z, Liu Q, Wilson CL, et al.: Polygenic Determinants for Subsequent Breast Cancer
437 Risk in Survivors of Childhood Cancer: the St Jude Lifetime Cohort Study (SJLIFE). *Clin*
438 *Cancer Res* 24:6230–5, 2018
- 439 31. Wang Z, Wilson CL, Easton J, et al.: Genetic Risk for Subsequent Neoplasms Among
440 Long-Term Survivors of Childhood Cancer. *J Clin Oncol* 36:2078–87, 2018
- 441 32. Mansouri I, Allodji RS, Hill C, et al.: The role of irradiated heart and left ventricular
442 volumes in heart failure occurrence after childhood cancer. *Eur J Heart Fail* 21:509–18,
443 2019
- 444 33. Haddy N, Tartier L, Koscielny S, et al.: Repair of ionizing radiation-induced DNA damage
445 and risk of second cancer in childhood cancer survivors. *Carcinogenesis* 35:1745–9,2014
- 446 34. de Vathaire F, Haddy N, Allodji RS, et al.: Thyroid Radiation Dose and Other Risk Factors
447 of Thyroid Carcinoma Following Childhood Cancer. *J Clin Endocrinol Metab* 100:4282–
448 90, 2015
- 449 35. Veres C, Allodji RS, Llanas D, et al.: Retrospective reconstructions of active bone marrow
450 dose-volume histograms. *Int J Radiat Oncol Biol Phys* 90:1216–24, 2014
- 451 36. Charrier T, Haddy N, Schwartz B, et al.: Increased Cardiac Risk After a Second Malignant
452 Neoplasm Among Childhood Cancer Survivors: A FCCSS Study. *JACC CardioOncology*
453 5:792–803, 2023
- 454 37. Subramanian A, Tamayo P, Mootha VK, et al.: Gene set enrichment analysis: A
455 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl*
456 *Acad Sci* 102:15545–50, 2005
- 457 38. Li H, Durbin R.: Fast and accurate short read alignment with Burrows–Wheeler transform.
458 *Bioinformatics* 25:1754–60, 2009
- 459 39. Poplin R, Ruano-Rubio V, DePristo MA, et al.: Scaling accurate genetic variant discovery
460 to tens of thousands of samples. *BioRxiv* 201178, 2018
- 461 40. McLaren W, Gil L, Hunt SE, et al.: The Ensembl Variant Effect Predictor. *Genome Biol*
462 17:122, 2016
- 463 41. Marenne G, Ludwig TE, Bocher O, et al.: RAVAQ: An integrative pipeline from quality
464 control to region-based rare variant association analysis. *Genet Epidemiol* 46:256–65,
465 2022

- 466 42. Mölder F, Jablonski KP, Letcher B, et al.: Sustainable data analysis with Snakemake.
467 F1000Res 10:33, 2021
- 468 43. Karczewski KJ, Francioli LC, Tiao G, et al.: The mutational constraint spectrum
469 quantified from variation in 141,456 humans. *Nature* 581:434–43, 2020
- 470 44. Auton A, Abecasis GR, Altshuler DM, et al.: A global reference for human genetic
471 variation. *Nature* 526:68–74, 2015
- 472 45. Lee S, Wu MC, Lin X.: Optimal tests for rare variant effects in sequencing association
473 studies. *Biostat Oxf Engl* 13:762–75, 2012
- 474 46. Wu MC, Lee S, Cai T, et al.: Rare-Variant Association Testing for Sequencing Data with
475 the Sequence Kernel Association Test. *Am J Hum Genet* 89:82–93, 2011
- 476 47. Li B, Leal SM.: Methods for Detecting Associations with Rare Variants for Common
477 Diseases: Application to Analysis of Sequence Data. *Am J Hum Genet* 83:311–21, 2008
- 478 48. Lee S, Emond MJ, Bamshad MJ, et al.: Optimal Unified Approach for Rare-Variant
479 Association Testing with Application to Small-Sample Case-Control Whole-Exome
480 Sequencing Studies. *Am J Hum Genet* 91:224–37, 2012
- 481 49. Bartsch DK, Fendrich V, Slater EP, et al.: RNASEL germline variants are associated with
482 pancreatic cancer. *Int J Cancer* 117:718–22, 2005
- 483 50. Datta A, Mazumder MHH, Chowdhury AS, et al.: Functional and Structural
484 Consequences of Damaging Single Nucleotide Polymorphisms in Human Prostate Cancer
485 Predisposition Gene RNASEL. *BioMed Res Int* 2015:271458, 2015
- 486 51. Rökman A, Ikonen T, Seppälä EH, et al.: Germline alterations of the RNASEL gene, a
487 candidate HPC1 gene at 1q25, in patients and families with prostate cancer. *Am J Hum*
488 *Genet* 70:1299–304, 2002
- 489 52. Carpten J, Nupponen N, Isaacs S, et al.: Germline mutations in the ribonuclease L gene in
490 families showing linkage with HPC1. *Nat Genet* 30:181–4, 2002
- 491 53. Xuan D, Li G, Cai Q, et al.: APOBEC3 deletion polymorphism is associated with breast
492 cancer risk among women of European ancestry. *Carcinogenesis* 34:2240–3, 2013
- 493 54. Neidhardt G, Hauke J, Ramser J, et al.: Association Between Loss-of-Function Mutations
494 Within the FANCM Gene and Early-Onset Familial Breast Cancer. *JAMA Oncol* 3:1245–
495 8, 2017
- 496 55. Schubert S, van Luttikhuizen JL, Auber B, et al.: The identification of pathogenic variants
497 in BRCA1/2 negative, high risk, hereditary breast and/or ovarian cancer patients: High
498 frequency of FANCM pathogenic variants. *Int J Cancer* 144:2683–94, 2019
- 499 56. Peterlongo P, Catucci I, Colombo M, et al.: FANCM c.5791C>T nonsense mutation
500 (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast
501 cancer risk factor. *Hum Mol Genet* 24:5345–55, 2015

- 502 57. Kiiski JI, Pelttari LM, Khan S, et al.: Exome sequencing identifies FANCM as a
503 susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci U S A* 111:15172–
504 7, 2014
- 505 58. Figlioli G, Bogliolo M, Catucci I, et al.: The FANCM:p.Arg658* truncating variant is
506 associated with risk of triple-negative breast cancer. *NPJ Breast Cancer* 5:38, 2019
- 507 59. Dicks E, Song H, Ramus SJ, et al.: Germline whole exome sequencing and large-scale
508 replication identifies FANCM as a likely high grade serous ovarian cancer susceptibility
509 gene. *Oncotarget* 8:50930–40, 2017
- 510 60. Lopes JL, Chaudhry S, Lopes GS, et al.: FANCM, RAD1, CHEK1 and TP53I3 act as
511 BRCA-like tumor suppressors and are mutated in hereditary ovarian cancer. *Cancer Genet*
512 235–236:57–64, 2019
- 513 61. Catucci I, Osorio A, Arver B, et al.: Individuals with FANCM biallelic mutations do not
514 develop Fanconi anemia, but show risk for breast cancer, chemotherapy toxicity and may
515 display chromosome fragility. *Genet Med* 20:42–7, 2018
- 516 62. Vogiatzoglou AP, Moretto F, Makkou M, et al.: Promyelocytic leukemia protein (PML)
517 and stem cells: from cancer to pluripotency. *Int J Dev Biol* 66:85–95, 2021
- 518 63. Zhuo C: Screening for Novel Germline Rare Mutations Associated with Aggressive
519 Prostate Cancer. *Cancers* 14(15):3623, 2022
- 520 64. Kwak JC, Ongusaha PP, Ouchi T, et al.: IFI16 as a Negative Regulator in the Regulation
521 of p53 and p21Waf1*. *J Biol Chem.* 278:40899–904, 2003
- 522

523 **Abbreviations**

524

525 3R: DNA repair, replication and recombination pathway

526 CCS: Childhood Cancer Survivors

527 CI (95%CI): 95% Confidence Interval

528 FCCS: French Childhood Cancer Survivors Study

529 GO: Gene Ontology

530 GWAS: Genome-Wide Association Study

531 KEGG: Kyoto Encyclopedia of Genes and Genomes

532 MAF: Minor Allele Frequency

533 MSigDb: Molecular Signature Database

534 NGS: Next Generation Sequencing

535 OR: Odds Ratio

536 PCA: Principal Component Analysis

537 SKAT: Sequence Kernel Association Test

538 SKAT-O: Sequence Kernel Association Test Optimized

539 SMN: Subsequent Malignant Neoplasm

540 SNDS: National Health Data System

541

542 **Gene names**

543 **ADAR**: Adenosine Deaminase RNA Specific

544 **ADARB1**: Adenosine Deaminase RNA Specific B1

545 **ANKRD17**: Ankyrin Repeat Domain 17

546 **ANXA6**: Annexin A6

547 **APOBEC3F**: Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 3F

548 **ATR**: ATR Serine/Threonine Kinase

549 **BCL6**: B-Cell Lymphoma 6 Protein

550 **BLM**: Bloom Syndrome RecQ Like Helicase

551 **CCNE1**: Cyclin E1

552 **CHECK2**: Checkpoint Kinase 2

553 **CTBP2**: C-Terminal Binding Protein 2

554 **CYREN**: Cell Cycle Regulator Of NHEJ

555 **DBF4**: DBF4-CDC7 Kinase Regulatory Subunit

556 **EGFR**: Epidermal Growth Factor Receptor

557 **ENPP7**: Ectonucleotide Pyrophosphatase/Phosphodiesterase 7

558 **EXD2**: Exonuclease 3'-5' Domain Containing 2

559 **FAN1**: FANCD2 And FANCI Associated Nuclease 1

560 **FANCM**: Fanconi Anemia Complementation Group M

561 **GLI2**: Glioma-Associated Oncogene Family Zinc Finger 2

562 **HINFP**: Histone H4 Transcription Factor

563 **HIRA**: Histone Cell Cycle Regulator

564 **IFI16**: Interferon Gamma Inducible Protein 16

565 **KAT2B**: Lysine Acetyltransferase 2B

566 **KPNA2**: Karyopherin Subunit Alpha 2

567 **MAS1**: Mas-Related G Protein-Coupled Receptor A

568 **MCM5**: Minichromosome Maintenance Complex Component 5

569 **MCM6**: Minichromosome Maintenance Complex Component 6

570 **MLH1**: MutL Homolog 1

571 **ORC5**: Origin Recognition Complex Subunit 5

572 **PARP1**: Poly(ADP-Ribose) Polymerase 1

573 **PLRG1**: Pleiotropic Regulator 1

574 **PML**: Promyelocytic Leukemia Protein

575 **PRKN**: Parkin RBR E3 Ubiquitin Protein Ligase

576 **PURA**: Purine Rich Element Binding Protein A

577 **RAD9B**: RAD9 Checkpoint Clamp Component B

578 **RANBP2**: RAN Binding Protein 2

579 **RNASEL**: Ribonuclease L

580 **SETDB1**: SET Domain Bifurcated Histone Lysine Methyltransferase 1

581 **SPIRE1**: Spire Type Actin Nucleation Factor 1

- 582 **SWSAP1**: SWIM-Type Zinc Finger 7 Associated Protein 1
- 583 **TPR**: Translocated Promoter Region
- 584 **TRIM6**: Tripartite Motif Containing 6
- 585 **TRRAP**: Transformation/Transcription Domain Associated Protein
- 586 **XPC**: Xeroderma Pigmentosum, Complementation Group C
- 587 **ZBTB38**: Zinc Finger And BTB Domain Containing 38
- 588 **ZC3H12A**: Zinc Finger CCCH-Type Containing 12A
- 589

Table 1 – Clinical characteristics of included patients

Characteristics	Cases n (%) N = 141	Controls n (%) N = 254	All included patients n (%) N = 395
Age at childhood cancer, years			
<5	35 (24.8)	62 (24.4)	97 (24.6)
[5-9]	44 (31.2)	76 (29.9)	120 (30.4)
[10-14]	40 (28.4)	77 (30.3)	117 (29.6)
≥15	22 (15.6)	39 (15.4)	61 (15.4)
Sex			
Male	46 (32.6)	88 (34.6)	134 (33.9)
Female	95 (67.4)	166 (65.4)	261 (66.1)
Childhood cancer type			
Lymphoma	41 (29.1)	75 (29.5)	116 (29.4)
Renal	22 (15.6)	44 (17.3)	66 (16.7)
Bone	22 (15.6)	37 (14.6)	59 (14.9)
STS*	16 (11.3)	32 (12.6)	48 (12.2)
CNS*	13 (9.2)	25 (9.8)	38 (9.6)
SNS*	9 (6.4)	18 (7.1)	27 (6.8)
Others	18 (12.8)	23 (9.1)	41 (10.4)
Year of diagnosis of childhood cancer			
<1970	28 (19.9)	52 (20.5)	80 (20.3)
[1970-1979]	45 (31.9)	77 (30.3)	122 (30.9)
[1980-1989]	42 (29.8)	86 (33.9)	128 (32.4)
1990	26 (18.4)	39 (15.4)	65 (16.5)
Median age at diagnosis (IQR)			8 (4-12)
Time from childhood cancer diagnosis to first SMN/last follow-up, years			
<15	21 (14.9)	46 (18.1)	67 (17)
[15-24]	48 (34)	81 (31.9)	129 (32.7)
[25-34]	39 (27.7)	67 (26.4)	106 (26.8)
≥35	33 (23.4)	60 (23.6)	93 (23.5)
Median Follow-up (IQR)			25 (17-34)
SMN type			
Thyroid	58 (41.1)	-	58 (41.1)
Breast	44 (31.2)	-	44 (31.2)
Bone	15 (10.6)	-	15 (10.6)
Kidney and Bladder	14 (9.9)	-	14 (9.9)
Colorectal	10 (7.1)	-	10 (7.1)
Radiotherapy			
Yes	130 (92.9)	175 (68.9)	305 (77.2)
No	11 (7.8)	77 (30.3)	88 (22.3)
Mean radiation dose, Gy (CI)			
Bone Marrow	5 (4.4-5.6)	3.9 (3.4-4.3)	4.3 (3.9-4.7)
Breast	12.4 (9.4-15.4)	4.7 (3.3-6)	8.11 (6.5-9.8)
Thyroid	8 (6.3-9.7)	6.5 (5-7.9)	7.1 (6-8.2)
Chemotherapy			
Yes	129 (91.5)	208 (81.9)	337 (85.3)
No	12 (8.5)	44 (17.3)	56 (14.2)
Anthracycline			
Yes	52 (37.9)	108 (42.5)	160 (40.5)
No	89 (63.1)	146 (57.5)	235 (59.5)
Median, mg/m ² (IQR)	285.3 (160.5-403.6)	294.6 (199.2-379.1)	293.3 (179.8-382)
Alkylating agents			
Yes	79 (56)	143 (56)	222 (56)
No	62 (44)	111 (44)	173 (44)
Median, mg/m ² (IQR)	6771.8 (401.5-13240.4)	7524.3 (3595-13358.6)	8304.6 (5033.3-13772.6)

* Abbreviations: CNS, Central Nervous System; SNS, Sympathetic Nervous System; STS, Soft Tissue Sarcoma; SMN, Second Malignant Neoplasm

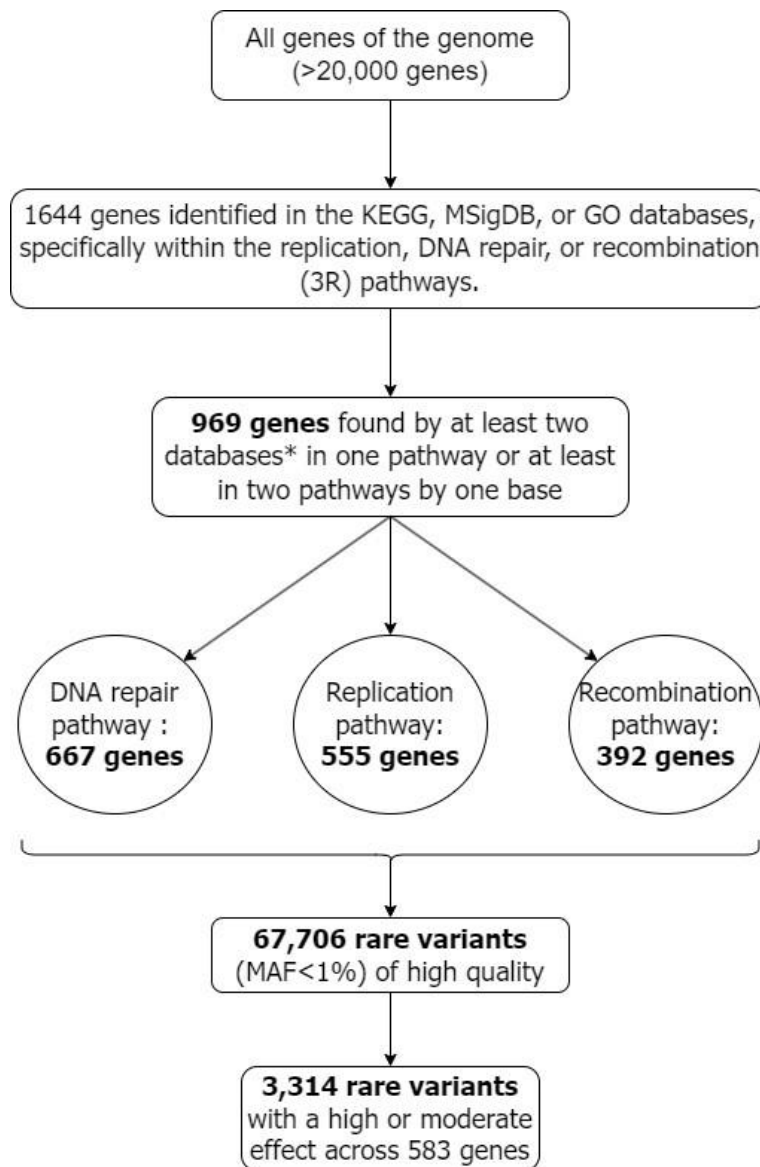
Table 2 – Multivariable association tests – Significant associations only

Gene	SMN type	OR	95%CI	P value	N variants (cases/controls)
APOBEC3F	Any SMN	4.8	(1.28-22.95)	0.027	2 (7/3)
	Breast	22.6	(2.06-781.04)	0.026	1 (3/1)
ATR	Any SMN	3.1	(1.17-8.95)	0.026	12 (11/7)
BCL6	Thyroid	45.3	(3.07-1479.93)	0.0086	4 (2/1)
BLM	Breast	5.2	(1.50-19.11)	0.0096	8 (7/6)
CTBP2	Breast	9.1	(1.73-51.47)	0.0089	6 (4/4)
DBF4	Any SMN	4.1	(1.35-14.04)	0.016	5 (10/5)
EGFR	Any SMN	10.7	(1.60-209.6)	0.034	4 (5/1)
	Thyroid	32.5	(3.62-806.66)	0.0066	4 (4/1)
EXD2	Any SMN	--	--	--	3 (5/0)
FAN1	Breast	4.2	(1.26-14.26)	0.018	7 (7/9)
FANCM	Breast	4.7	(1.06-21.56)	0.041	9 (5/6)
HINFP	Any SMN	3.4	(1.06-12.03)	0.043	5 (8/5)
	Thyroid	5.3	(1.26-23.50)	0.022	5 (5/5)
HIRA	Thyroid	8.5	(1.07-89.81)	0.047	3 (3/2)
IFI16	Thyroid	3.5	(1.28-9.32)	0.013	6 (9/15)
KPNA2	Any SMN	--	--	--	2 (3/0)
MAS1	Thyroid	24.0	(1.72-638.95)	0.022	3 (2/1)
MCM5	Thyroid	5.7	(1.36-23.87)	0.015	7 (5/6)
ORC5	Breast	18.4	(1.9-418.21)	0.020	2 (3/1)
PARP1	Breast	4.8	(1.15-22.04)	0.034	5 (6/4)
PML	Breast	5.1	(1.01-27.82)	0.048	7 (4/4)
RANBP2	Thyroid	8.7	(1.98-40.20)	0.004	8 (3/5)
RNASEL	Any SMN	5.4	(1.76-20.44)	0.0055	10 (8/4)
XPC	Any SMN	4.8	(1.49-18.57)	0.012	11 (9/4)
ZBTB38	Thyroid	132.1	(9.62-4654.71)	0.00088	3 (3/1)

Abbreviations: SMN, Second Malignant Neoplasm; OR, Odds Ratio; CI, 95% Confidence Interval

Significant association according to the logistic regression adjusted on clinical variable: sex, age at diagnosis of first cancer, type of first cancer and year of diagnosis of first cancer, follow-up and first two principal components of PCA to account for population stratification.

Figure 1 – Genes and variants selection workflow



Abbreviations: GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MSigDB, Molecular Signature Database; MAF, Minor Allele Frequency.

MAF (Minor Allele Frequency) represents the frequency of the less common allele in the population, providing insight into genetic variation. Rare variants are defined as those with a MAF below 1%.

Figure 2 – Gene-based tests – Gene Selection

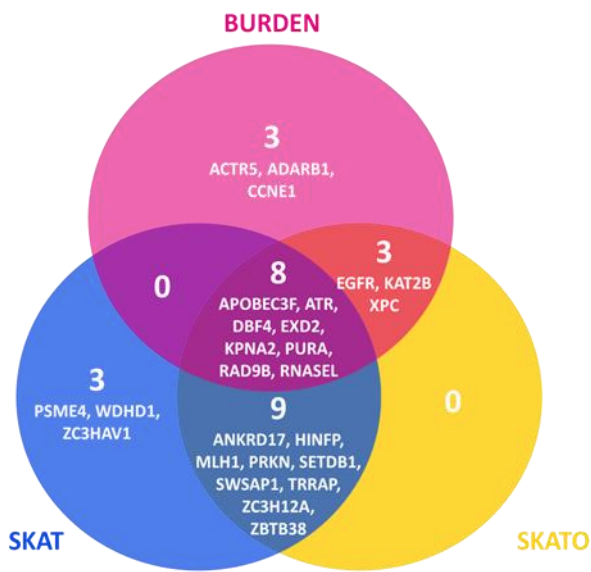


Figure 3A – Any SMN

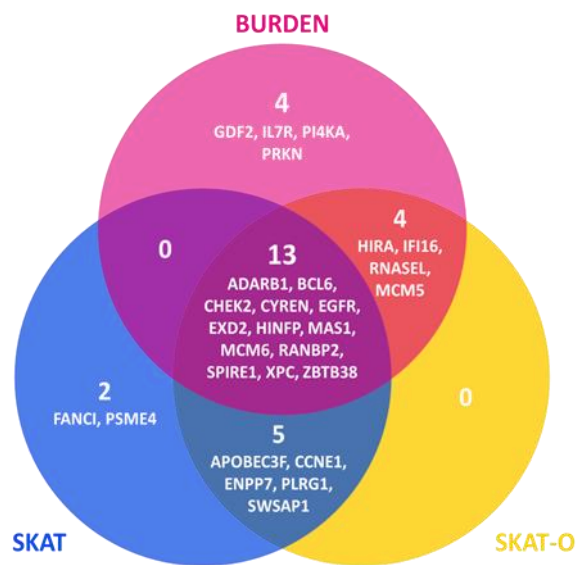


Figure 3B – Thyroid SMN

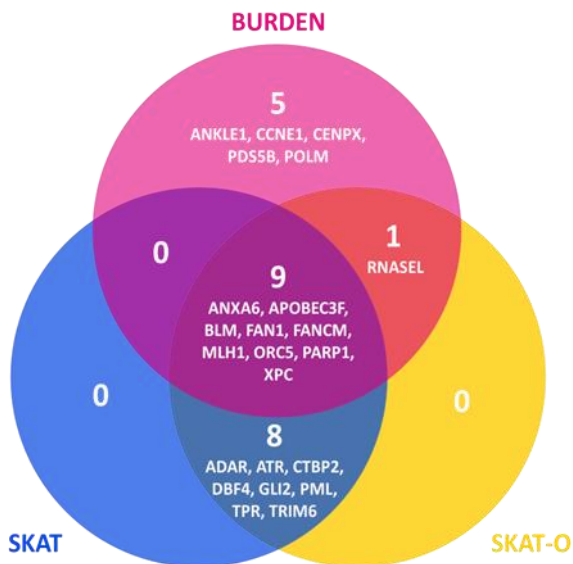


Figure 3C – Breast SMN