

Université de Montréal

**ÉTIOLOGIE DU BIAIS DE L'INACTIVATION DU CHROMOSOME  
X (ICX) DANS LES CELLULES SANGUINES DES FEMMES  
VIEILLISSANTES : SÉLECTION HÉMIZYGOTE ET ACQUISITION  
DE MUTATIONS SOMATIQUES**

par

Sami Ayachi

Département de biochimie et médecine moléculaire  
Faculté de médecine

Thèse présentée à la Faculté de médecine  
en vue de l'obtention du grade de doctorat en sciences  
Option médecine moléculaire

Avril 2022

© Sami Ayachi, 2022

Université de Montréal  
Faculté de médecine

*Cette thèse intitulée*

**ÉTIOLOGIE DU BIAIS DE L'INACTIVATION DU CHROMOSOME  
X (ICX) DANS LES CELLULES SANGUINES DES FEMMES  
VIEILLISSANTES : SÉLECTION HÉMIZYGOTE ET ACQUISITION  
DE MUTATIONS SOMATIQUES**

Présentée par  
Sami Ayachi

A été évaluée par un jury composé des personnes suivantes

Président-rapporteur : Sylvie Mader, Ph.D.  
Directeur de recherche : Lambert Busque, MD, FRCPC  
Membre du jury : Trang Hoang, Ph.D.  
Examineur externe : Denis Cournoyer, MD, FRCPC  
Représentant du doyen : Louis Gaboury, MD, FRCPC

## RÉSUMÉ

Les cellules souches hématopoïétiques (CSH) assurent une production constante des cellules sanguines tout au long de la vie, mais sont vulnérables à l'acquisition de mutations pouvant mener à une transformation maligne. Les mutations qui confèrent un avantage de croissance entraîneront une prolifération clonale. L'étude de la clonalité est centrale à la compréhension de ces phénomènes. Historiquement, l'analyse de la clonalité a été possible grâce au principe de l'inactivation du chromosome X (ICX) chez les femmes qui entraîne la création de deux populations cellulaires, celle avec le X-paternel actif et celle avec le X-maternel actif. Une déviation (biais) de la proportion théorique de 1 : 1 entre ces deux populations peut supposer une dominance clonale.

Nous avons démontré un biais significatif de l'ICX chez les femmes avec l'âge. Ce phénomène peut être expliqué par plusieurs causes dont la sélection hémizygote (un des deux X possède des allèles plus forts que l'autre) et l'acquisition de mutations dans une CSH. Nous posons l'hypothèse que ces deux phénomènes coexistent et peuvent être distingués par une approche génomique.

Nous avons recruté une cohorte de 2996 femmes canadiennes-françaises âgées entre 37 et 101 ans composée de 2172 individus issus de 321 familles et de 824 individus non apparentés. Deux tissus biologiques ont été recueillis : le sang périphérique (PMN, monocytes, lymphocytes T, lymphocytes B) et des cellules buccales. Le ratio de l'ICX a été déterminé par la méthode HUMARA, l'analyse de gènes associés à l'hématopoïèse clonale (19 gènes) a été faite par la méthode de séquençage NGS, et la cohorte a été génotypée à 700 625 loci polymorphiques de l'ADN (SNP). Des analyses bioinformatiques ont été

appliquées pour étudier la contribution génétique au biais de l'ICX. Nous démontrons que :

- (i) le biais de l'ICX est plus prévalent dans les cellules sanguines par rapport aux cellules épithéliales et maximal dans les cellules myéloïdes;
- (ii) le biais augmente avec l'âge seulement dans les cellules sanguines et que cette influence est plus marquée pour les neutrophiles;
- (iii) la concordance du biais est très importante pour les différents types cellulaires sanguins, suggérant un mécanisme opérant au niveau de la CSH ;
- (iv) il y a une composante héréditaire liée au biais de l'ICX;
- (v) la présence de mutations acquises (*TET2*, *DNMT3A*, etc.) explique seulement une partie du biais ;
- (vi) à l'aide d'analyses par liaison génétique la présence d'une région sur le chromosome X à Xq21 (LOD score 4.9) qui est associée au biais des lymphocytes T et une autre sur le chromosome 1 à 1q21 (LOD score 6) qui est associée au biais des neutrophiles.

Nous avons départagé la contribution liée à l'acquisition de mutations somatiques et identifié pour la première fois des régions liées à une prédisposition génétique. Nos travaux se poursuivront d'une part par l'analyse de gènes candidats dans les régions identifiées, et d'autre part nous tenterons d'identifier les cibles génétiques qui confèrent un potentiel de transformation maligne en utilisant une approche basée sur l'analyse du méthylome, de l'hydroxyméthylome et du transcriptome que nous venons de valider.

Notre étude démontre la complexité de l'adaptation de l'hématopoïèse au vieillissement et ouvre des portes sur l'identification de facteurs prédisposant aux cancers hématologiques.

**Mots clés:**

Vieillesse, Hématopoïèse, Clonalité, Inactivation du chromosome X, Sélection hémizygote, HUMARA, Linkage, GWAS, *TET2*, 5mC, 5hmC, NGS, RNA-Seq.



## **ABSTRACT**

Hematopoietic stem cells (HSC) ensure a constant lifelong production of blood cells, but are vulnerable to acquisition of mutations, which may lead to malignant transformation. Mutations that confer a growth advantage will lead to clonal derivation of cells. The study of clonality is central to the understanding of hematopoiesis adaptation to aging. Historically, the first clonality assays were based on the principle of X-chromosome inactivation (XCI) in women. Women are mosaics with half the cells with the paternal X active and the other half with the maternal one. A skewing from the theoretical 1:1 ratio between these two populations of cells could infer clonal derivation of cells.

More than 20 years ago, our team demonstrated, through analysis of (XCI) in women, that skewing increases with age. This intriguing phenomenon can be explained by several etiology including hemizygous selection (one of the 2 Xs has stronger alleles) or the acquisition of mutations giving a growth advantage. The first etiology is genetically predetermined and the second, acquired in somatic cells of bone marrow. We hypothesize that these two phenomena coexist and can be distinguished with a genomic approach.

To test our hypothesis, we investigated skewing in a cohort of 2996 French-Canadian women aged 37 to 101 comprised of 2172 related individuals from 321 families and 824 unrelated individuals. We analyzed XCI ratios at the HUMARA locus in epithelial cells, neutrophils, T-cells, monocytes, B-lymphocytes. We genotyped the cohort for clonal hematopoiesis and looked for germline heritable components by genome wide association studies and linkage analyses. We document that skewing was more prevalent in blood cells than in epithelial cells, and maximal in myeloid cells. Skewing increases with age only in blood cells. Intra-

individual correlation of skewing blood cell types was strongly correlated, suggesting selection influences operating at the HSC. Sibship analyses demonstrated heritability which was strongest when parental origin of skewing was taken into account. Clonal hematopoiesis accounted only for a small proportion of the skewing trait but its importance increased in the very old. Linkage analysis identified a region at Xq21 for skewing occurring in T-cells (LOD score 4.9) suggesting a hemizygous cell selection influence. We also identified a region at 1q21 for skewing in neutrophils (LOD score 6) suggesting a gene-gene interaction with X-linked genes.

We have demonstrated that age-associated skewing is a complex trait caused in part by acquired mutations and genetic predisposition variants. We will pursue our investigation using a candidate gene approach in the two identified regions and will try to identify genetic targets of oncogenic potential by a method based on analysis of the methylome, hydroxymethylome and transcriptome that was have validated in this cohort.

This thesis demonstrates the complexity of the adaptation mechanisms of hematopoiesis to aging and set the stage to identification of factors predisposing to hematological cancers.

**Key words:**

Aging, Hematopoiesis, Clonality, X chromosome inactivation, Hemizygous selection, HUMARA, Linkage, GWAS, TET2, 5mC, 5hmC, NGS, RNA-Seq.

## TABLE DES MATIÈRES

RÉSUMÉ .....	iii
ABSTRACT.....	v
TABLE DES MATIÈRES .....	vii
LISTE DES FIGURES .....	xi
LISTE DES TABLEAUX.....	xiii
LISTE DES ABRÉVIATIONS.....	xiv
REMERCIEMENTS.....	xviii
CHAPITRE 1 .....	1
1. INTRODUCTION .....	2
1.1 Introduction à la clonalité et à l'hématopoïèse clonale.....	2
1.2 L'inactivation du chromosome X .....	3
1.2.1 Historique sur la découverte de l'ICX .....	3
1.2.2 Mécanisme de l'ICX .....	5
1.2.2.1 Durant l'embryogenèse .....	5
1.2.3 Méthode permettant de mesurer l'ICX chez la femme .....	6
1.2.3.1 Glucose-6-phosphate-dehydrogenase : le paradigme des méthodes d'analyse de l'ICX .....	7
1.2.3.2 Polymorphe hypoxanthine-phosphoribosyltransférase: première méthode basée sur la méthylation différentielle entre Xa et Xi.....	8
1.2.3.3 Phosphoglycérate kinase: première méthode utilisant la réaction de polymérase en chaîne (PCR).....	9
1.2.3.4 Locus DXS255 et sonde M27 $\beta$ ; première méthode utilisant un locus multi-allélique .....	9
1.2.3.5 Méthodes basées sur l'analyse de transcrits polymorphiques de gène du chromosome X.....	10
1.2.3.6 La méthode de <i>Human Androgen-Receptor Gene (HUMARA)</i> .....	10
1.2.3.7 La méthode de XISTase.....	12
1.2.4 L'analyse de l'ICX dans la population saine.....	12
1.2.5 L'étiologie du biais de l'ICX .....	14

1.2.5.1	Biais de l'ICX non aléatoire primaire.....	14
1.2.5.2	Le biais de l'ICX secondaire.....	15
1.2.5.2.1	Études chez les femmes saines.....	15
1.2.5.2.1.1	La spécificité tissulaire du degré d'ICX .....	15
1.2.5.2.1.2	Effet de l'âge sur le biais de l'ICX des cellules sanguines.....	16
1.2.5.2.2	La sélection hémizygote.....	16
1.2.5.2.3	La sélection aléatoire dans le contexte de déplétion en cellules souches .	20
1.2.5.2.4	Les mutations acquises donnant un avantage de croissance.....	22
1.2.6	Les mécanismes des mutations les plus fréquentes DAT .....	26
1.2.6.1	<i>TET2</i> and <i>DNMT3</i> .....	26
1.2.6.2	<i>ASXL1</i> .....	28
1.3	L'étiologie de ARCH/CHIP.....	30
1.3.1	L'impact de la prédisposition génétique sur la progression des mutations dans les CSH.....	30
1.3.2	Le rôle de l'inflammation dans l'HC .....	31
1.3.3	L'impact de l'exposition écotoxique.....	33
1.3.4	L'impact de l'attaque du système immunitaire.....	33
1.3.5	Sommaire des causes de biais de l'ICX.....	35
1.3.6	Sommaire des données scientifiques.....	35
1.4	Projets de recherches .....	36
1.4.1	Mise en contexte .....	39
1.4.2	Hypothèse .....	40
1.4.3	Objectifs.....	40
	CHAPITRE 2.....	41
2.	<i>ETIOLOGY OF SKEWING OF X-CHROMOSOME INACTIVATION RATIOS OCCURRING IN BLOOD CELLS OF AGING WOMEN</i> .....	42
2.1	Abstract.....	43
2.2	Introduction.....	44
2.3	Materials and Methods.....	46
2.4	Results.....	50

2.5 Discussion.....	55
2.6 Tables and Figures.....	62
CHAPITRE 3.....	70
3. DÉTERMINANTS DE LA PÉNÉTRANCE ONCOGÉNIQUE DE LA MUTATION DE <i>TET2</i> CHEZ LES INDIVIDUS ÂGÉS .....	71
3.1 Résumé.....	71
3.2 Introduction.....	72
3.3 Matériel et méthodes :.....	74
3.3.1 Les sujets humains .....	74
3.3.2 Méthylome, hydroxyméthylome sur NGS Ion Torrent.....	75
3.3.3 Analyses bioinformatiques.....	79
3.3.4 RNA-Seq.....	80
3.3.4.1 Analyses bioinformatiques.....	81
3.4 Résultats.....	81
3.4.1 Contrôle de qualité des échantillons séquencés .....	82
3.4.1.1 Contrôle <i>spike-in</i> de l'enrichissement 5hmC .....	82
3.4.1.2 5mC : Contrôle par un gène endogène méthylé.....	83
3.4.1.3 Quantification de la bibliothèque 5mC et 5hmC .....	84
3.4.2 Contrôle de qualité des fichiers BAM.....	85
3.4.3 Couverture estimée de l'OFFSET .....	89
3.4.4 Résumé des résultats de CQ.....	91
3.4.5 Résultats des échantillons choisis .....	94
3.4.6 Analyse des régions méthylées différemment.....	94
3.4.7 L'expression génique dans les TET2 vs WT .....	96
CHAPITRE 4.....	98
4. DISCUSSION.....	99
4.1.1 Résumé des travaux de recherche .....	99
4.1.2 La prévalence du biais de l'ICX varie en fonction des tissus et des types de cellules .....	101
4.1.3 Le biais de l'ICX dans les cellules sanguines varie selon l'âge.....	103

4.1.4	Le biais de l'ICX est associé avec le VAF des mutations CHIP .....	104
4.1.5	L'héritabilité du biais de l'ICX est partiellement génétique .....	106
4.1.6	Des SNP sur le chromosome 23 et 1 sont associés avec le biais de l'ICX	107
4.1.7	Étude d'associations génomiques (GWAS) .....	112
4.2	Pénétrance oncogénique de la mutation de TET2 chez les individus âgés.....	113
4.2.1	Résumé des travaux de recherche .....	113
4.2.2	Processus de sélection de la méthode .....	114
4.2.3	Les méthodes d'analyses bioinformatiques .....	118
4.2.4	RNA-Seq et l'expression génique dans TET2 vs WT .....	121
4.2.5	Limitation de notre cohorte et effet faible de TET2 .....	122
CHAPITRE 5 .....		124
5.	CONCLUSION.....	125
5.1	Directions futures.....	127
RÉFÉRENCES .....		128
ANNEXES.....		143
ANNEXE 1 .....		144
ANNEXE 2 .....		155
ANNEXE 3 .....		176

## LISTE DES FIGURES

Figure 1: Schéma du choix aléatoire de l'inactivation du chromosome X (adapté de (30)) .....	6
Figure 2: La méthode <i>HUMARA</i> pour déterminer l'origine clonale d'une tumeur (adaptée de (49)).....	11
Figure 3: Modèle de biais de l'ICX (adaptée de (30)) .....	13
Figure 4: La sélection hémizygote.....	19
Figure 5: Modèle de la sélection aléatoire dans le contexte de déplétion en cellules souches (adaptée de (76)).....	21
Figure 6: Régulateurs épigénétiques de l'hématopoïèse clonale (adaptée de (3)) .....	29
Figure 7: Sommaire des causes de biais de l'ICX .....	35
Figure 8: Modèle explicatif de l'initiation du biais de l'ICX selon l'âge .....	38
Figure 9: Prevalence of XCI skewing varies according to tissue and blood cells types. .	64
Figure 10: Intra-individual correlation of XCI ratios between different tissues and blood cells types.....	65
Figure 11: Prevalence of XCI skewing is age-dependent in blood cells and maximal in PMN.....	66
Figure 12: Prevalence of XCI skewing in PMN and CHIP at different VAF in function of age.....	67
Figure 13: Regions associated with XCI skewing.....	69
Figure 14: Représentation schématique des méthodes du méthylome et hydroxyméthylome .....	75
Figure 15: Résultats de bio-analyseur .....	77
Figure 16: Représentation schématique des étapes de la méthode du méthylome.....	78

Figure 17: Spécificité de l'anticorps anti-5hmC par rapport au Spike-in .....	83
Figure 18: Spécificité de l'anticorps anti-5mC par rapport au TsH2B .....	84
Figure 19: Quantification par qPCR de la bibliothèque de 5mC et 5hmC .....	85
Figure 20: Enrichissement CpG incluant l'INPUT TET2-5mC .....	86
Figure 21: Enrichissement CpG incluant l'INPUT WT-5mC.....	87
Figure 22: Enrichissement CpG incluant l'INPUT TET2-5hmC.....	88
Figure 23: Enrichissement CpG incluant l'INPUT WT-5hmC.....	89
Figure 24: Expression génique dans les TET2 vs WT .....	97
Figure 25: Age distribution of the 2996 healthy women of the cohort .....	145
Figure 26: Prevalence of somatic mutations in the 347 mutated individuals of the cohort .....	146
Figure 27: Age-dependent XCI skewing in blood cells .....	149
Figure 28: MERLIN QTL Linkage analysis in epithelial cells .....	151
Figure 29: MERLIN QTL Linkage analysis in T cells.....	152
Figure 30: MERLIN VC Linkage analysis in monocytes .....	153
Figure 31: MERLIN VC Linkage analysis in B cells.....	154



## LISTE DES TABLEAUX

Table 1: Association between degree of skewing (DS) scores and blood cell counts ....	62
Table 2: Association between skewing degree of skewing (DS) scores and common medical conditions .....	62
Table 3: Intra-individual correlation of Psup between different cell types.....	63
Table 4: Heritability of skewing .....	63
Tableau 5: Liste des échantillons séquencés.....	81
Tableau 6: Couverture estimée de l'OFFSET pour les échantillons de 5mC par rapport à leur INPUT.....	90
Tableau 7: Couverture estimée de l'OFFSET pour les échantillons de 5hmC par rapport à leur INPUT.....	90
Tableau 8: Résumé des résultats de CQ pour les échantillons de 5mC .....	91
Tableau 9: Résumé des résultats de CQ pour les échantillons de 5hmC .....	93
Tableau 10: Échantillons finaux utilisés pour l'ensemble de données 5mC .....	94
Tableau 11: Échantillons finaux utilisés pour l'ensemble de données 5hmC .....	94
Tableau 12: Les DMR dans TET2 vs WT 5mC .....	96
Tableau 13: Résultats d'analyse par liaison génétique en 2-points avec Merlin sur le trait du biais de l'ICX.....	111
Tableau 14: Prevalence of XCI skewing is age-dependent in blood cells and maximal in PMN .....	147
Tableau 15: Prevalence of XCI skewing in PMN and CHIP at different VAF in function of age groups .....	148
Tableau 16: Correlation between XCI and VAF of CHIP mutations in PMN .....	150

## LISTE DES ABRÉVIATIONS

2-HG	2-hydroxyglutarate
2-OG	2-oxoglutarate
5-caC	5-carboxylcytosine
5-fC	5-formylcytosine
5hmC	5-hydroxyméthylcytosine
5mC	5-méthylcytosine
AAS	<i>Age-associated skewing</i>
ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
AR	<i>Androgen receptor</i>
ARCH	Hématopoïèse clonale liée à l'âge
ARN	Acide ribonucléique
ARNm	ARN messenger
<i>ASXL1</i>	Additional sex combs like 1
BC	Bucca epithelial cell
BS-seq	Séquençage au bisulfite
CHIP	Hématopoïèse clonale à potentiel indéterminé
CIHR	<i>Canadian Institute of Health Research</i>
CNV	Variation du nombre de copies
CQ	Contrôle de qualité
CSH	Cellules souches hématopoïétiques
CSPH	Cellules souches et de progéniteurs hématopoïétiques
CRP	Protéine C-réactive

CpG	5'-C-phosphate-G-3'
<i>DAT</i>	<i>DNMT3- ASXL1-TET2</i>
DIP	Immunoprécipitation de l'ADN
DHMR	Régions hydroxyméthylées différenciellement
DMR	Régions méthylées différenciellement
<i>DNMT</i>	ADN méthyl transférase
DS	<i>Degree of skewing</i>
DUP	Disomie uni-parentale féminine
DZ	<i>Dyzogotic</i>
FACS	<i>Fluorescence-activated cell sorting</i>
<i>G6PD</i>	Glucose-6-phosphate déshydrogénase
GLM	Modèles linéaires généralisés
GO	<i>Gene Ontology</i>
GRM	<i>Genetic relationship matrix</i>
GWAS	<i>Genome wide association studies</i>
H3	Histone 3
H3K4	Histone 3 lysine 4
HC	Hématopoïèse clonale
HPRT	Hypoxanthine-phosphoribosyltransférase
HSC	<i>Hematopoietic stem cells</i>
<i>HUMARA</i>	<i>Human Androgene-Receptor Gene</i>
ICX	Inactivation du chromosome X
IgG	Immunoglobuline G
IL	Interleukine
<i>IL2RG</i>	<i>Interleukine-2 receptor gamma chain gene</i>

<i>IRAQ1</i>	<i>Interleukine 1 receptor-associated Kinase 1</i>
kb	Kilobases
LMA	Leucémie myéloïde aiguë
LMC	Leucémie myéloïde chronique
MAF	<i>Minor allele frequency</i>
Mb	Mégabases
MPC	Mosaïcisme placentaire confiné
MZ	Monozygotic
NGS	Séquençage de nouvelle génération
NMP	Néoplasmes myéloprolifératifs
pb	Paire de bases
PCR	Réaction en chaîne par polymérase
<i>PGK</i>	Phosphoglycérate kinase
PMN	Polynucléaires neutrophiles
<i>PRC</i>	Complexe répressif <i>polycomb</i>
qPCR	PCR quantitatif
qRT-ASS-PCR	<i>Quantitative reverse-transcription allele-specific suppressive PCR</i>
QTL	<i>Quantitative Trait Loci</i>
RFLP	<i>Restriction fragment length polymorphism</i>
RNA-Seq	ARN sequencing
RT-PCR	PCR par transcriptase inverse
SMD	Syndrome myélodysplasique
SNP	Polymorphismes mononucléotidiques
STR	<i>Short tandem repeat</i>
<i>TERT</i>	Transcriptase inverse de la télomérase

<i>TET</i>	Translocation à dix-onze méthylcytosine dioxygénase
t-LMA	LMA liée au traitement
<i>TNF</i>	Tumor necrosis factor
VAF	Fréquence allélique variable
VC	<i>Variant Component</i>
VNTR	Variabiles de répétitions en tandem
WAS	Syndrome de Wiscott–Aldrich
Xa	Chromosome X activé
Xce	Élément de contrôle X
Xi	Chromosome X inactivé
Xic	Centre d'inactivation de X
<i>Xist</i>	Transcrit spécifique du X inactif
Xm	Chromosome X maternel
Xp	Chromosome X paternel

## **REMERCIEMENTS**

Il m'est agréable, au terme de ce travail, de remercier chaleureusement toutes les personnes qui ont, de près ou de loin, participé à l'accomplissement de mon projet.

C'est avec beaucoup d'émotions et de sincérité que je tiens à remercier mon directeur de recherche, Dr Lambert Busque, pour m'avoir accueilli dans son laboratoire et pour m'avoir donné l'occasion de travailler sur des projets stimulants et innovants dans un environnement sain et bienveillant. Plus qu'un directeur de recherche, vous avez été pour moi un mentor tant sur le plan professionnel que sur le plan humain. Je ne vous remercierai jamais assez pour votre disponibilité, votre écoute, votre soutien, votre ouverture d'esprit, votre confiance, votre bienveillance et surtout pour votre patience que j'ai à plusieurs reprises mise à rude épreuve. Tant de qualités qui m'ont permis d'évoluer et de m'épanouir pour devenir le scientifique accompli que je suis aujourd'hui, et qui m'inspireront pour toujours un profond respect et une grande admiration.

Je n'aurais jamais pu atteindre mes objectifs sans l'aide précieuse et le soutien de Manuel Buscarlet et de Sylvie Provost qui ont consacré de nombreuses heures à travailler fort, et parfois dans la précipitation, pour le bon avancement de mes projets. Je vous remercie infiniment pour toute l'aide que vous m'avez apportée. Un grand remerciement à Dre Marie-Pierre Dubé pour son temps précieux, son aide et ses bons conseils.

Merci à toutes les personnes qui ont siégé sur mes comités et qui m'ont donné de précieux conseils pour l'avancement de mes projets : Dre Julie Lessard, Dre Luigina Mollica, Dre Josée Hébert, Dr Benoît Coulombe.

Merci à tous les membres de mon jury qui ont accepté d'examiner mon travail : Dre Sylvie Mader, le président rapporteur, Dr Denis Cournoyer, examinateur externe, Dr Trang Hoang, membre du jury, et Dr Louis Gaboury, représentant du doyen.

Je n'aurais pas pu arriver au bout sans le soutien et l'amour de mes amis, qui loin de mon pays, sont devenus pour moi une deuxième famille. Merci à ma meilleur amie Amina Dahmani de m'avoir donné un coup de main dans la révision de ma thèse et à mon petit frère Clément Takola de m'avoir aidé dans le dessin de mes belles figures. Merci à mon grand frère Arturo David Faria pour ses conseils précieux et à mon *Bestie* Alex Sioui pour les moments de joie et de folie qu'on a passé ensemble, ainsi que tous les autres pour votre écoute et votre soutien. Vous avez fait partie de cette aventure et avez partagé avec moi les bons et les moins bons moments. Puisse notre amitié durer pour toujours.

Un remerciement spécial à Michel-Olivier Gratton qui est toujours venu à ma rescousse et qui a su m'épauler à chaque fois que j'en avais besoin. Un merci tout spécial à Lise Gratton, à toute la famille Gratton-Demers, Mélanie, Sébastien, Andréanne, Justin.

Merci à mon grand frère, Riad, à ma grande sœur et deuxième maman Lynda pour leur soutien infini. Merci à ma sœur et mon amie Salima pour tous les moments de joie et de folie. Merci à mon petit frère Mohamed-Amine que je considère comme mon fils et que je veille sur lui comme un adolescent même s'il ne l'est pas. Merci à ma belle-sœur Wafa et à mes adorables nièces et neveux, Pissa, Yakine, Rawnak, Moutayakin, Tej, Mountaha, Jad et Miraloo qui m'ont apporté l'équilibre, la douceur et l'amour dont j'avais besoin pour me ressourcer.

Enfin, à mes chers et précieux parents sans qui rien n'aurait été possible. Je vous remercie d'avoir cru en moi, de m'avoir soutenu dans mes choix et mes décisions, aux dépens de votre confort émotionnel et matériel. Merci de m'avoir donné l'amour, l'éducation et le confort nécessaire à mon épanouissement. Vous m'avez aussi transmis le goût du travail bien fait en héritage et c'est grâce à cela que je suis arrivée jusqu'à là aujourd'hui. Puissiez-vous être fiers de mes accomplissements.



**CHAPITRE 1**  
**INTRODUCTION ET REVUE DE LA LITTÉRATURE**

# **1. INTRODUCTION**

## **1.1 Introduction à la clonalité et à l'hématopoïèse clonale**

Les études sur la clonalité du développement cellulaire remontent à il y a près de 60 ans (1). Ces études ont permis de définir les principes de base de l'oncologie moderne en prouvant que les cancers résultent de la prolifération clonale de cellules (2). La prolifération cellulaire clonale implique qu'une population entière de cellules émane d'une seule cellule souche fondatrice qui a un avantage de croissance par rapport aux autres cellules.

La prolifération clonale de cellules peut être causée par plusieurs mécanismes cellulaires, génétiques ou épigénétiques. Au cours du développement d'un organisme, les cellules possèdent le même matériel génétique. Dans cet organisme, la régulation de l'expression des gènes permet le développement de différents types de tissus qui auront différentes fonctions (3). En vieillissant, les cellules peuvent acquérir des mutations somatiques (4-6). Des mutagènes d'origine interne par exemple le stress oxydatif, la formaldéhyde et l'acétaldéhyde, peuvent causer des erreurs dans la réplication de l'acide désoxyribonucléique (ADN) (7). Les mutagènes peuvent également être d'origine externe comme la présence des substances mutagènes liées au tabac, à la lumière ultraviolet (8), à la chimiothérapie (9) ou à la radiothérapie (10). Dans la plupart des cas, ces mutations n'auront aucun impact car elles vont toucher des régions non codantes du génome ou non régulatrices. Toutefois, si ces mutations affectent des gènes essentiels à la prolifération

cellulaire, la cellule mutée pourrait avoir une altération lui donnant un avantage de croissance. Ultimement, les mutations pourraient provoquer une croissance rapide des cellules qui pourrait se manifester par un phénotype de malignité et un diagnostic clinique de cancer (11, 12).

Nous nous intéresserons dans cette thèse à mieux comprendre ce paradigme dans le contexte de l'hématopoïèse. Les cellules souches hématopoïétiques ont la tâche titanesque de produire  $> 10^{11}$  cellules sanguines quotidiennement (13). Les cellules sanguines sont sujettes à l'acquisition de mutations somatiques pouvant conduire à l'hématopoïèse clonale (HC) et au développement de cancers hématologiques (3).

## **1.2 L'inactivation du chromosome X**

### **1.2.1 Historique sur la découverte de l'ICX**

Ohno et Hauschka ont démontré que chez les souris femelles, les chromosomes X des cellules somatiques sont hétéropycnotiques, c'est à dire qu'un des deux chromosomes X se comporte exactement de la même manière que les autosomes, où il reste dans un état étendu pendant l'interphase et la prophase. L'autre chromosome est dans un état extrêmement condensé, formant le Corpuscule de Barr. (14). Or, l'origine de cette forme hétéropycnotique était restée inconnue. Un an plus tard, Mary Lyon s'est basée sur le mosaïcisme de l'expression phénotypique des gènes récessifs chez les souris femelles hétérozygotes qui affectent la couleur de leur fourrure pour suggérer que la forme hétéropycnotique du chromosome X peut être d'origine maternel ou paternel dans

différentes cellules chez le même animal et qu'un des chromosome X est inactivé au cours du développement (15). Lyon a aussi expliqué que l'une des copies du chromosome X dans les cellules est fortement condensée et que les souris qui possèdent seulement une copie de chromosome X se développent en femelles fertiles. Lyon postula que l'un des deux chromosomes X est inactivé de façon aléatoire dans chaque cellule, que la descendance de chaque cellule maintiendra le même patron d'ICX résultant en un mosaïcisme avec 50% des cellules avec le chromosome paternel actif, et 50% avec le chromosome X maternel actif (15).

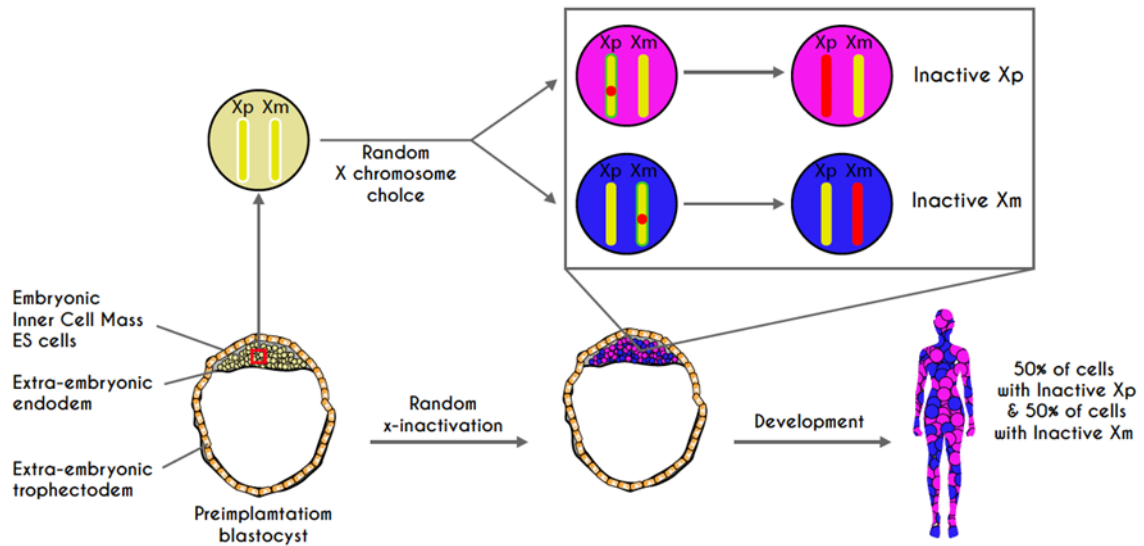
Dans la même période et indépendamment, Ernest Beutler a suggéré la même hypothèse mais chez l'humain. Il propose qu'un chromosome X est inactivé pour maintenir la parité de dose avec le seul chromosome X qui provient du mâle (16). C'est ce qu'on appelle la compensation de dose des chromosomes sexuels (17). Il a prouvé son hypothèse en utilisant les allèles du gène codant pour l'enzyme glucose-6-phosphate déshydrogénase (G6PD) situés sur le chromosome X (16). Il a observé que chez les femmes hétérozygotes pour le gène G6PD, il coexiste deux populations de cellules érythrocytaires: des cellules complètement déficientes et des cellules normales. Il conclut que dans les précurseurs des globules rouges, un seul chromosome X est actif et que l'inactivation affecte l'un ou l'autre des chromosomes dépendamment des cellules (16).

## **1.2.2 Mécanisme de l'ICX**

### **1.2.2.1 Durant l'embryogenèse**

Avant l'initiation de l'ICX dans l'embryon, les ovocytes de mammifères possèdent deux chromosomes X actifs. Le chromosome X inactif est réactivé avant l'entrée dans la prophase méiotique (18). Durant la spermatogenèse, le chromosome X actif est progressivement inactivé pendant la première prophase méiotique (19). Dans le processus de l'ICX, l'inactivation est réalisée par l'hétérochromatinisation du chromosome X, une forme de condensation de la chromatine (20). Les gènes du chromosome X inactif arrêtent d'être exprimés par la cellule ce qui permet la compensation de dose (17, 21). Dans l'embryon de la souris aux stades de deux ou quatre cellules, le Xp inactif est réactivé dans les cellules de la masse cellulaire interne au stade blastocyste. Par la suite, les cellules de l'épiblaste (stade blastocyste) inactivent au hasard et indépendamment une copie de chromosome X (22-24). Chez l'humain, le processus de l'ICX commence pendant les phases préimplantatoires du développement embryonnaire, vers le stade des huit cellules (25). Il existe deux formes différentes de l'ICX : l'ICX imprimée et l'ICX aléatoire. L'ICX imprimée entraîne une inactivation préférentielle du chromosome Xp et se produit dans les marsupiaux femelles et les tissus placentaires de souris (26, 27). Par contre, l'ICX aléatoire du chromosome X maternel ou paternel a lieu dans toutes les lignées de cellules somatiques de mammifères. L'ICX aléatoire n'a pas de préférence spécifique pour l'inactivation de l'un des chromosomes X parentaux (20, 28). Le chromosome X inactivé est mitotique. Il va être hérité par les futures cellules filles somatiques. Cette inactivation au hasard donne comme résultat des cellules mosaïques dans un individu où globalement on s'attend à ce qu'il y

aura une balance (50:50) d'expression des allèles des gènes liés aux deux chromosomes X parentaux (Figure 1, page 6) (29).



**Figure 1:** Schéma du choix aléatoire de l'inactivation du chromosome X (adapté de (30))

### 1.2.3 Méthode permettant de mesurer l'ICX chez la femme

Les prérequis des méthodes pour déterminer la clonalité basées sur l'ICX sont les suivants :

(i) être en mesure de distinguer les deux chromosomes X parentaux (polymorphisme lié au chromosome X), (ii) être capable de détecter leur état d'activation (expression, transcription ou méthylation différentielle). Les techniques présentées ci-après de façon chronologique rencontrent ces prérequis.

### **1.2.3.1 Glucose-6-phosphate-dehydrogenase : le paradigme des méthodes d'analyse de l'ICX**

Tel que mentionné précédemment, Ernest Beutler a utilisé les allèles du gène codant pour l'enzyme G6PD pour analyser l'ICX (16). Par la suite, l'analyse de l'ICX a été utilisée pour identifier l'origine et le développement des cellules tumorales (31). La première étude a été réalisée par Linder et Gartler, qui ont étudié les variantes A et B du gène G6PD chez 5 femmes hétérozygotes (32). Ils ont démontré que les femmes possèdent deux populations de cellules, et chaque population exprime une des variantes du gène G6PD (A ou B). En se basant sur l'électrophorèse des variantes A et B de ce gène, ils ont étudié 27 léiomyomes, des tumeurs bénignes formées de cellules musculaires lisses, provenant de 5 femmes hétérozygotes et 86 myomètres, la couche musculuse interne de la paroi utérine (33). Leur but était de déterminer si ce type de tumeur est déclenché par une seule cellule ou par plusieurs cellules en sachant que si la tumeur vient d'une seule cellule ils devraient observer qu'un seul phénotype de G6PD. Par contre si plusieurs cellules sont à l'origine de la tumeur ils devraient observer plusieurs phénotypes. Effectivement, ils ont découvert que les 27 léiomyomes ont une seule bande qui représente un seul allèle du gène G6PD ce qui indique que l'origine d'un léiomyome est une seule cellule versus les 85 des 86 myomètres qui ont 2 allèles (32). L'application du système de G6PD pour analyser l'inactivation de chromosome X a ouvert un grand champ de recherche en oncologie puisque cet outil pouvait être utilisé pour comprendre la base de la transformation maligne. Or, l'application de ce système était malheureusement très limitée par le niveau faible de l'hétérozygoté dans la population générale. L'évolution des variants a été favorisée dans les régions géographiques où la malaria est plus importante. L'hétérozygoté est de 35% chez les

femmes de descendance africaine et très faible dans la population caucasienne (34) . Cela a empêché l'analyse des maladies rares, l'analyse de grandes cohortes de femmes et l'analyse familiale (35). Cette limitation a poussé les chercheurs à trouver d'autres alternatives pour analyser la clonalité à partir de tests basés sur l'ICX.

### **1.2.3.2 Polymorphisme hypoxanthine-phosphoribosyltransférase: première méthode basée sur la méthylation différentielle entre Xa et Xi**

La nécessité d'identifier plus de sujets hétérozygotes pour des gènes sur le chromosome X a amené Vogelstein à proposer une stratégie inédite basée sur les avancés de la biologie moléculaire de l'époque (36). Son analyse est basée surtout sur 3 principes : (1) Un seul chromosome X est actif dans chaque cellule, (2) l'activation des gènes lié à l'X est accompagnée par un changement de méthylation des résidus de cytosine faisant en sorte qu'ils peuvent être reconnus par des enzymes de restriction endonucléaires sensibles à la méthylation de l'ADN, (3) les copies du chromosome X paternel et du chromosome X maternel peuvent être distinguées par le polymorphisme de longueur des fragments de restriction (ou RFLP, de l'anglais *restriction fragment length polymorphism*) (37). L'hétérozygotie des RFLP était accrue de façon substantielle par rapport au gène G6PD mais limitée à 25% des tumeurs. De plus, ce système d'analyse de l'ICX a également des limitations : (i) certaines tumeurs sont résistantes aux enzymes de restriction, (ii) cette analyse ne peut pas faire la distinction entre une tumeur monoclonale et une tumeur oligoclonale, (iii) la présence de cellules saines peut contaminer les échantillons et faire en sorte que les résultats indiquent que des tumeurs sont polyclonales même si celle-ci est d'origine monoclonale, ce qui peut être le cas pour plusieurs analyses (36).



### **1.2.3.3 Phosphoglycérate kinase: première méthode utilisant la réaction de polymérase en chaîne (PCR)**

Une autre limitation des analyses de clonalité était la nécessité d'avoir une importante quantité de cellules comme matériel de départ. La découverte de la réaction de polymérisation en chaîne (PCR) par Kary Mullis (38) a permis à Gilliland de développer un test d'une grande sensibilité analytique au locus de gène Phosphoglycerate kinase (PGK)(39). La méthode permet l'amplification de la partie polymorphique de PGK après une digestion sélective du chromosome X activé en utilisant des enzymes de restriction sensible à la méthylation comme le Hpa II et Bst XI. Cette méthode est plus sensible la rendant parfaitement adaptée pour étudier des populations cellulaires de quantité limitée. Le niveau d'hétérozygotie demeure toutefois faible à 25%.

### **1.2.3.4 Locus DXS255 et sonde M27 $\beta$ T; première méthode utilisant un locus multi-allélique**

Le locus DXS255 est hautement polymorphique en raison de la présence d'un nombre variable de répétitions en tandem (VNTR) (40, 41). Ces polymorphismes sont mis en évidence par la sonde M27 $\beta$ T (42, 43). La distinction entre Xa et Xi est basée sur la méthylation différentielle à ce site. Le taux élevé d'hétérozygotie est d'environ 90% (40) et présentait un grand avancement pour les études de clonalité (44). Toutefois, la méthode basée sur la technologie de Southern Blot était complexe et d'interprétation laborieuse. Cette technique est tombée en désuétude rapidement avec l'arrivée des technologies décrites ci-après.

### 1.2.3.5 Méthodes basées sur l'analyse de transcrits polymorphiques de gène du chromosome X.

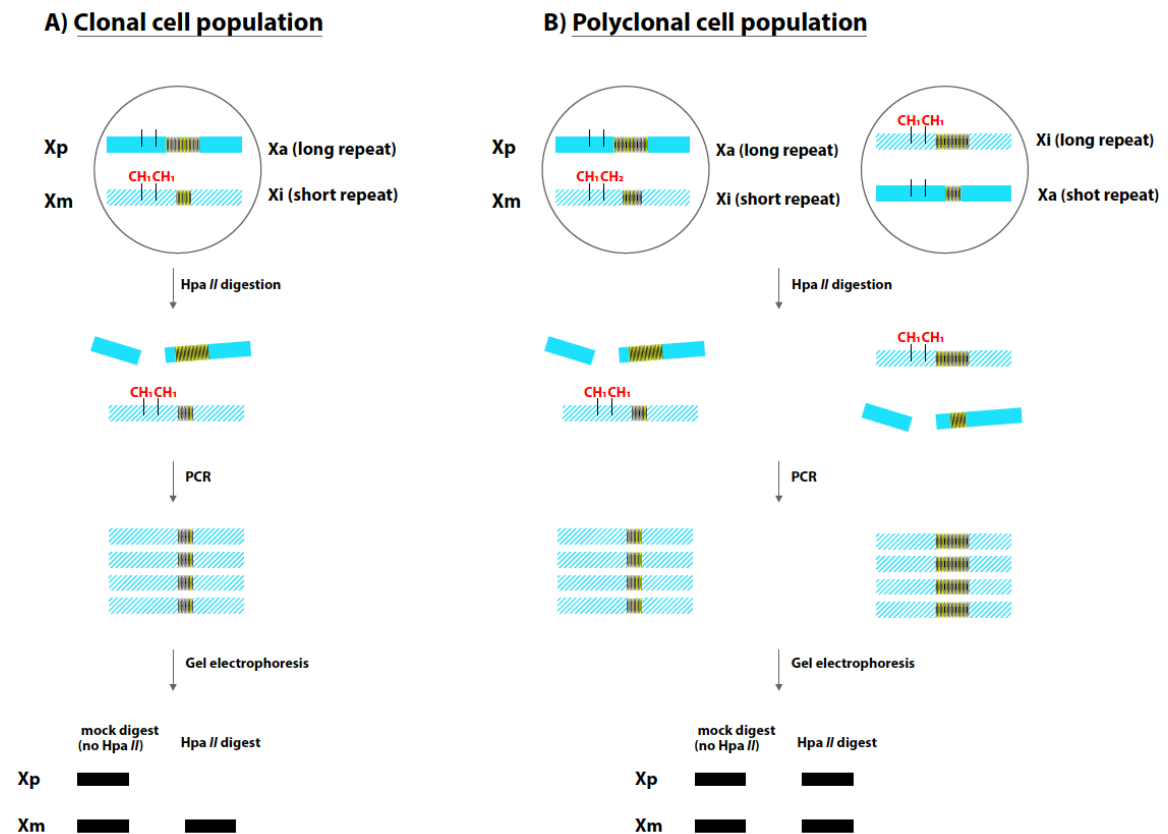
Plusieurs méthodes ont été développées par le groupe du Dr J. Prchal avec l'objectif de bénéficier de la sensibilité de la PCR, d'utiliser un polymorphisme fréquent et de ne pas recourir au principe de la méthylation différentielle. L'ARN doit être utilisé comme matériel de départ. La *quantitative reverse-transcription allele-specific suppressive PCR* (qRT-ASS-PCR) est un exemple de cette approche (45).

### 1.2.3.6 La méthode de *Human Androgen-Receptor Gene (HUMARA)*

En 1992, Allen *et al.* décrivent un test qui utilise une séquence hautement polymorphique dans le gène codant pour le récepteur d'androgène humain situé sur le chromosome X (AR ou *HUMARA* [GeneBank]) (46). En effet, le gène *HUMARA* est porteur d'une séquence CAG répétée en tandem (*short tandem repeat* (STR)) qui est hautement polymorphique dans la population et qui permet de distinguer entre les allèles paternels et les allèles maternels chez 90% des femmes (47). Les répétitions du CAG sont liées à quatre sites de méthylation qui sont méthylés sur l'Xi et non-méthylés sur le Xa (46). Les amorces PCR qui permettent l'amplification du locus sont en 5' des sites de méthylation différentielle entre Xa et Xi et en 3' de la séquences CAG, si bien que la digestion préalable de l'ADN avec l'enzyme de restriction HpaII (qui coupera Xa non-méthylé) permettra uniquement l'amplification de l'Xi. Les produits de PCR sont ensuite analysés par électrophorèse pour déterminer le nombre d'allèles et quantifier leur intensité (**Figure 2A**, page 11) (46, 47). Après avoir fait migrer le produit de PCR sur un gel les résultats sont les suivants : (i) Si

les deux bandes sont apparentes, le tissu étudié est d'origine polyclonale car la moitié des cellules ont le Xm inactif et l'autre moitié l'Xp inactif (**Figure 2B**, page 11) (ii) si une seule bande est observée, le tissu est d'origine monoclonale (**Figure 2A**, page 11) à moins que les deux allèles aient exactement le même nombre de répétitions CAG (47).

Busque *et al.* ont ensuite développé en 1994 un test basé sur l'ARN en utilisant la méthode de PCR par transcriptase inverse (RT-PCR) afin de démontrer la corrélation de ce dernier avec le test de *HUMARA* et la fiabilité des différents sites de méthylation sur ce locus du gène *HUMARA* (48).



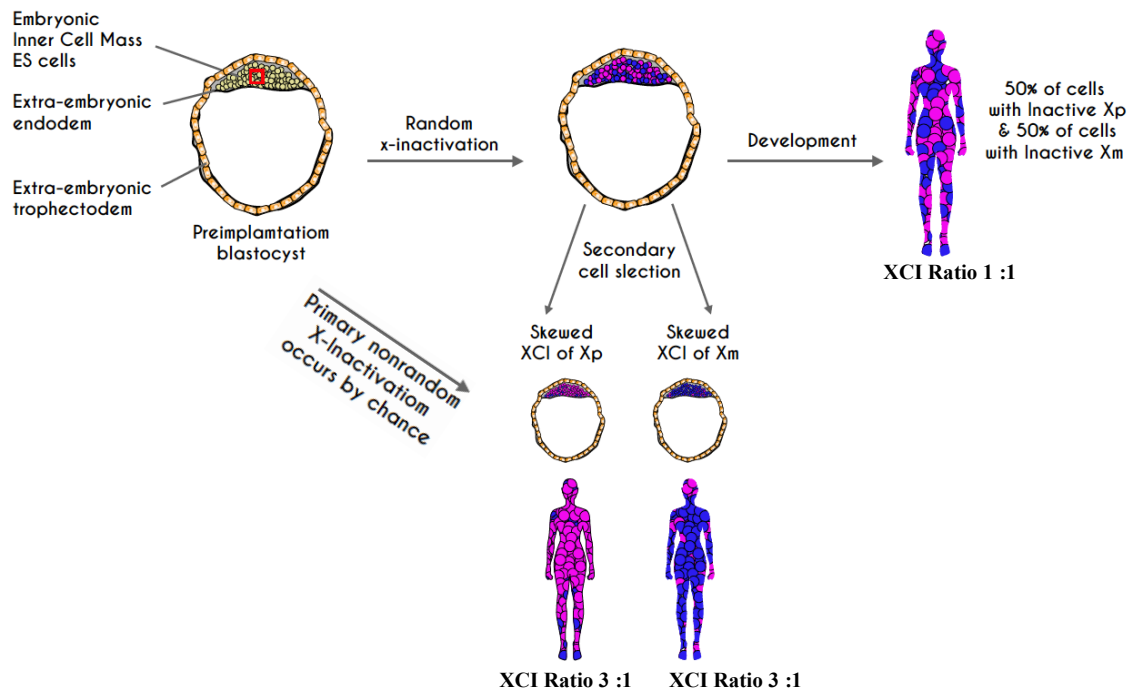
**Figure 2:** La méthode *HUMARA* pour déterminer l'origine clonale d'une tumeur (adaptée de (49))

### 1.2.3.7 La méthode de XISTase

Récemment, l'équipe de Zito *et al.* ont analysé l'ICX en utilisant une nouvelle technique basée sur une mesure métrique de l'expression de l'allèle spécifique au gène Xist provenant de données de RNA-Seq et DNA\_seq (29). L'utilisation du gène Xist est tirée d'une découverte en 1995 par l'équipe de Rupert *et al.* qui a identifié un polymorphisme exprimé dans le gène Xist et qui l'ont utilisé pour générer 2 amorces de l'allèle spécifique pour être utiliser pour le PCR (50). Ce qui est particulièrement plus important dans cette étude est l'utilisation du test *HUMARA* pour valider les résultats obtenus par XISTase. Ainsi, ils ont démontré que la corrélation de Spearman entre les quantifications était de 0,8 ( $P = 7 \times 10^{-5}$ ), révélant un degré élevé de reproductibilité entre XISTase et HUMARA dans les cellules lymphoblastes et les cellules du sang.

### 1.2.4 L'analyse de l'ICX dans la population saine.

Le chromosome X représente environ 5% du génome humain haploïde. Il est composé de 155 Mb et il contient plus de 800 gènes qui codent pour plusieurs protéines (51, 52). Chez les femmes saines, l'inactivation au hasard du chromosome X entraîne la création d'une mosaïque avec un ratio de 1:1 entre les cellules qui ont le Xp actif et celles qui ont le Xm actif (53). Une déviation significative par rapport au ratio de 1:1 est connu sous le nom de biais de l'ICX (ou *ICX skewing*, *nonrandom X inactivation* ou lyonization excessive) (Figure 3, page 13) (29). Plusieurs études ont retenu le critère d'un ratio  $\geq 3:1$  ce qui correspond à 75% des cellules avec le même X parental actif pour qualitativement catégoriser les sujets avec biais (54-56). Le biais de l'ICX peut être primaire ou secondaire (acquis).



**Figure 3:** Modèle de biais de l'ICX (adaptée de (30))

Dans l'ICX non aléatoire primaire, un facteur ou une modification modifie ce caractère aléatoire, de sorte que le Xp ou le Xm est de préférence inactivé. Dans la sélection cellulaire secondaire, l'ICX aléatoire se produit normalement mais la mutation d'un gène sur l'un ou l'autre X, conférant un avantage de croissance ou, au contraire, induisant une létalité cellulaire, favorise la sélection de cellules portant soit un Xp inactif, soit un Xm inactif.

## 1.2.5 L'étiologie du biais de l'ICX

### 1.2.5.1 Biais de l'ICX non aléatoire primaire

Les causes du biais de l'ICX non aléatoire primaire sont les événements qui se produisent durant l'embryogenèse (Figure 7: Sommaire des causes de biais de l'ICX page 35). Plusieurs hypothèses ont été proposées dans le but d'expliquer le biais de l'ICX et les divers facteurs qui peuvent contribuer à son initiation :

- (i) Nombre limité de cellules d'initiation ICX : Un des facteurs les plus importants est la présence d'un nombre limité de cellules au moment de l'initiation de l'ICX pendant l'embryogenèse. Ainsi, même si les chances que chaque cellule inactive un ou l'autre chromosome X est de 50%, la majorité des cellules pourraient choisir de désactiver par hasard le même chromosome X, par exemple le chromosome X paternel, quand le pool de cellules de départ est très petit. (53) (57) (53).
- (ii) Facteurs génétiques : Le biais de l'ICX est probablement déterminé génétiquement (58). Chez les souris, des distorsions de l'ICX aléatoires sont observés lorsque plusieurs souches de souris consanguines sont croisées. Comme ces écarts ne sont pas causés par la sélection des cellules, ce mécanisme a été appelé : ICX non aléatoire primaire. (59) (60). Alternativement, des SNP ont été trouvés dans le promoteur XIST humain, en corrélation avec le biais de l'ICX (61) (62). Aussi, plusieurs familles ont été identifiées dans lesquelles le biais de l'ICX est hérité, mais ce biais extrême pourrait également être expliqué par la présence d'allèles d'une maladie liée au chromosome X (63).

- (iii) Sélection cellulaire : Par conséquent, le biais de l'ICX pourrait également être causé par des mutations liées au chromosome X affectant la viabilité cellulaire, suivi d'une sélection négative ou même d'une sélection positive (53).
- (iv) Sauvetage trisomique : Une autre forme d'ICX non aléatoire primaire se retrouve dans ce qu'on appelle le sauvetage de zygotes trisomiques et pourrait également être une cause du développement de biais de l'ICX. (64).

### **1.2.5.2 Le biais de l'ICX secondaire**

Le biais de l'ICX secondaire se produit après l'initiation de l'inactivation du chromosome X aléatoire (Figure 3, page 13, Figure 7 page 35).

#### **1.2.5.2.1 Études chez les femmes saines**

##### **1.2.5.2.1.1 La spécificité tissulaire du degré d'ICX**

Le groupe de Fey *et al.* ont démontré que l'ICX est spécifique à un tissu précis en comparant des tissus normaux obtenus de la muqueuse gastro-intestinale et de la thyroïde. Ces tissus présentent un faible niveau de biais de l'ICX tandis que les leucocytes présents dans le sang présentent un niveau élevé de biais de l'ICX (65). Gale *et al.* ont comparé les profils de biais de l'ICX dans le sang périphérique (granulocytes, cellules B et cellules T), de la peau et des muscles de 20 femmes qui ont une formule sanguine normale ainsi que de la muqueuse du côlon de neuf individus et ils ont constaté que le biais de l'ICX est très élevé dans le sang périphérique (> 75% d'expression d'un allèle) mais pas dans les autres tissus analysés (66). En résumé, ces deux études ont démontré que le biais de l'ICX est

plus grand dans le sang périphérique par rapport à d'autres tissus ce qui suggère que le profil de biais de l'ICX est spécifique au tissu analysé.

#### **1.2.5.2.1.2 Effet de l'âge sur le biais de l'ICX des cellules sanguines.**

Fey *et al.* ont été les premiers à suggérer que le biais de l'ICX est corrélé avec l'âge. En utilisant le M27 $\beta$  ils ont démontré que des profils du biais de l'ICX ont été observés non seulement dans les leucémies mais également dans 21/105 échantillons (20%) de leucocytes normaux (67). De plus, le biais de l'ICX est significativement plus fréquent chez les femmes âgées (75 ans à 96 ans) par rapport aux enfants en bonne santé (âgés de 2 à 8 ans) et aux femmes plus jeunes (âgées de 20 à 58 ans) ( $P = .00125$ ; chi 2 test) (67). Par la suite, Busque *et al.* ont utilisé le test de *HUMARA* sur 295 femmes saines partagées sur 3 groupes : (1) les nouveau-nés, (2) les femmes âgées entre 28 et 32 ans et (3) les femmes âgées de  $\geq$  à 60 ans. L'incidence de biais de l'ICX (ratio d'allèles  $\geq$  3:1) était de 8,6% (14 sur 162) chez les nouveau-nés ( $P = 0,104$  versus les femmes entre 28 et 32 ans); 16,4% (11/67) de 28 à 32 ans ( $P = .0064$  versus le groupe  $\geq$  à 60 ans) et **37,9%** (25 sur 66) chez les femmes  $\geq$  60 ans (54). Ainsi, ces données ont démontré que le biais de l'ICX augmente avec l'âge et ceci a été confirmé par la suite par d'autres études (68, 69).

*Quels sont les facteurs et les mécanismes qui causent le biais secondaire de l'ICX ?*

#### **1.2.5.2.2 La sélection hémizygote**

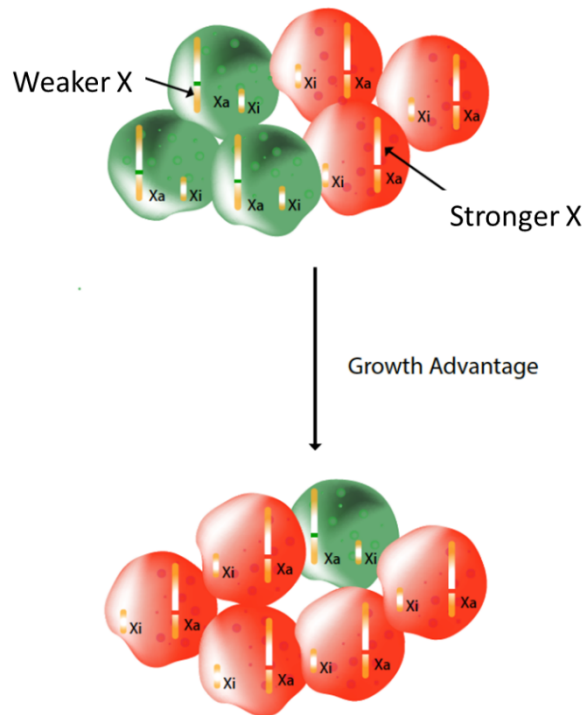
La sélection hémizygote s'explique par l'avantage concurrentiel pour toutes les cellules qui expriment un phénotype parental (Figure 4, page 19). Le processus de la sélection



hémizygote peut être une cause du biais de l'ICX. Cette théorie a été exposée par Abkowitz *et al* qui ont analysé le phénotype du G6PD des cellules du sang et des cellules progénitrices de la moelle des chattes âgées de la race Safari. Cette race est un premier croisement hybride entre les chats de Geoffroy (G) et des chats domestiques (d). Puisqu'il n'y a pas de recombinaisons entre les chromosomes X-G et X-d, les chattes safari ont la moitié de leurs cellules avec l'X-G actif et l'autre avec l'X-d actif. À la naissance, l'expression des allèles G et d sont équivalentes (ratio 50 :50) démontrant une inactivation aléatoire du chromosome X. Toutefois, Au fil du temps, ils ont observé un biais du ratio d'ICX chez près de 70% des chattes vieillissantes. Ils ont fait l'observation cruciale que le biais était toujours dans la direction de l'allèle G. Cette observation élimine les causes aléatoires qui auraient pu se produire autant dans les cellules G que d. Cela indique donc que les cellules qui ont le chromosome X-G dans l'état actif ont un avantage de croissance sur les cellules qui ont l'X-d actif. Puisque cela prend plusieurs années avant que la domination cellulaire se produise les auteurs déduisent qu'il y a des variants de gènes impliqués dans l'hématopoïèse qui se situe sur le chromosome X et que le variant G est plus fort que le variant d (70).

Plusieurs données soutiennent également une contribution génétique au trait. Christensen *et al.* ont analysé les cellules sanguines périphériques de 71 paires de jumeaux monozygotiques âgées de 73 à 93 ans et ils ont constaté qu'avec l'âge, une des populations cellulaires devient prédominante pour la plupart des femmes. Ils ont également observé une forte tendance pour la même lignée cellulaire à devenir prédominante chez 2 co-jumeaux. Cela suggère que les facteurs génétiques liés au chromosome X influencent la

cinétique des cellules souches hématopoïétiques humaines (71). Vickers *et al.* ont analysé le profil de l'ICX dans les cellules granulocytes de 33 monozygotes et 22 dizygotes des jumeaux âgés et ils ont observé une importante composante génétique dans la détermination de la méthylation asymétrique des allèles du chromosome X dans les granulocytes de jumeaux âgés. Ils ont aussi observé une corrélation du ratio de l'ICX entre les jumeaux monozygotes en comparaison aux jumeaux dizygotes. (72). L'héritabilité du biais de l'ICX dans les cellules polymorphonucléaires dans ces 2 études est à environ 68% (72) (71) ce qui suggère que le biais de l'ICX observé dans les cellules hématopoïétiques est, au moins en partie, génétiquement déterminé. En plus de la contribution génétique dans l'initiation de biais de l'ICX, y a-t-il d'autres mécanismes qui peuvent avoir un effet sur le biais de l'ICX dans les cellules hématopoïétiques des femmes vieillissantes ?



**Figure 4:** La sélection hémizygote.

Les femmes sont des mosaïques avec deux populations cellulaires distinguées seulement par les gènes exprimés par le chromosome X paternel ou maternel. S'il y a des allèles de gènes de différentes forces entre les deux populations, il y aura une sélection avec le temps.

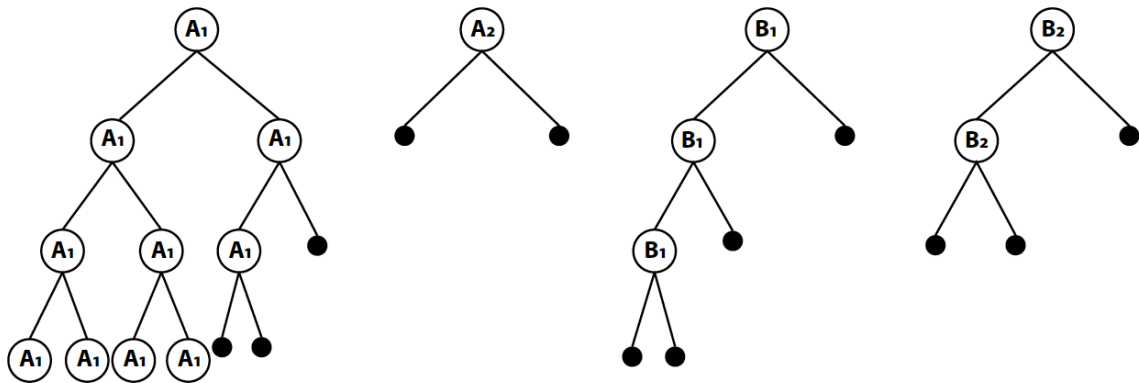
Le plus important est qu'ils soupçonnent qu'un mécanisme comparable est opérationnel chez les femmes. D'autant plus qu'il y a plusieurs exemples d'hémizygotie dans la différenciation des cellules du sang. Chez les femmes hétérozygotes pour deux troubles liés au chromosome X, le syndrome de Wiscott–Aldrich (WAS) et le G6PD. Chez les femmes atteintes du WAS toutes les plaquettes et cellules T expriment un seul phénotype parental G6PD (A ou B) alors que les deux isoenzymes A et B sont présentes parmi les globules rouges et les granulocytes ce qui implique des plaquettes défectueuses et les lymphocytes T échouent à devenir mature chez les femmes hétérozygotes. Cette observation suggère que la sélection contre les cellules exprimant le défaut de Wiskott-Aldrich a lieu dans le système hématopoïétique de la femme hétérozygote (73). Aussi, dans une étude sur deux

femmes hétérozygotes atteintes de l'agammaglobulinémie liée au chromosome X aux allèles A et B de G6PD, les cellules T et les neutrophiles avaient des niveaux similaires d'actif enzymatique A et B mais les cellules B avaient uniquement une activité liée à la forme A de G6PD. Le défaut génétique responsable de l'agammaglobulinémie liée au chromosome X est intrinsèque à la lignée des lymphocytes B. Ainsi, ces résultats démontrent qu'il y a un avantage de croissance sélective pour les cellules normales B (mais pas les cellules T ou les granulocytes) chez les femmes atteintes de l'agammaglobulinémie (74). En somme, la domination d'un phénotype de chromosome X parental unique parmi les cellules sanguines est un phénomène fréquent avec le vieillissement et démontre que ceci résulte d'une sélection hémizygote.

#### **1.2.5.2.3 La sélection aléatoire dans le contexte de déplétion en cellules souches**

Toutes les cellules hématopoïétiques sont dérivées de cellules souches pluripotentes qui ont engagé, suite à une série de voies de différenciation, leur progéniture vers les différentes lignées de cellules sanguines (75). Le système a un énorme potentiel de réserve et l'hématopoïèse échoue rarement même chez les patients âgés qui ont été exposés à une chimiothérapie (76). Malheureusement la régulation de cellules souches est peu connue. La capacité de génération de cellules souches d'une lignée est cependant limitée. Après un certain nombre de générations c'est-à-dire qu'une fois qu'elles ont «vieilli», une cellule souche quitte le pool des cellules souches pour former deux précurseurs de lignées cellulaires. Sa part dans la formation du sang est reprise par la cellule souche suivante la plus ancienne (77). Il existe cependant des preuves que de nombreuses décisions d'engagement dans le système hématopoïétique ont un élément aléatoire ou stochastique

(78). Le taux global d'auto-renouvellement à chaque division dans le pool de cellules souches doit être de 0,5 pour maintenir l'état d'équilibre. Au niveau de la cellule unique, toute cellule souche en division pourrait avoir un taux d'auto-renouvellement de 1,0 (les deux cellules filles sont des cellules souches), 0,5 (une cellule fille est une cellule souche et une entre dans le pool de cellules de différenciation) ou 0 (les deux cellules filles subissent une différenciation terminale) (Figure 5, page 21). Ce dernier modèle prédit qu'avec le temps, le pool de cellules souches deviendrait composé de descendants qui émanent de moins en moins du pool original de cellules souches, mais sans baisse du nombre total de cellules (76). Gale *et al.* a soulevé l'hypothèse de la déplétion en cellules souches avec le vieillissement pour expliquer l'émergence du biais de l'ICX (76). Toutefois, en absence de capacité de mesurer le nombre de cellules souches et leur capacité de division, et de faire une corrélation avec le biais de l'ICX, la déplétion en cellules souches comme cause de biais de l'ICX demeure une hypothèse impossible à vérifier.



**Figure 5:** Modèle de la sélection aléatoire dans le contexte de déplétion en cellules souches (adaptée de (76)).

L'ICX est déterminé par le modèle relatif d'expression des cellules souches (A) ou (B) au moment de l'ICX, par exemple avec une expression 50:50 des deux allèles comme indiqué ici. Au moment de la division des cellules souches, certaines cellules s'engageront dans la différenciation terminale (•). Avec le temps, bien que le nombre absolu de cellules souches reste le même, certaines des cellules souches seront perdues par différenciation terminale des deux cellules filles. L'hématopoïèse sera donc dérivée des descendants de quelques-uns des membres originaux de la population de cellules souches et cela se reflètera dans l'expression relative des allèles A et B

#### **1.2.5.2.4 Les mutations acquises donnant un avantage de croissance**

Les cellules souches et les progéniteurs hématopoïétiques accumulent des mutations somatiques tout au long de la vie d'un individu, dont la plupart sont des mutations passagères non pathogéniques sans conséquence fonctionnelle ou qui ont le potentiel de contribuer à une expansion clonale qui conduit à l'hématopoïèse clonale [118] [119]. Ceci dit, est ce que les mutations somatiques qui donne un avantage de croissance ont un lien avec le biais de l'ICX ?

Notre groupe a joué un rôle très important dans l'investigation des causes génétiques et non génétiques du biais de l'ICX. Busque *et al.* ont été intrigués par le fait que le biais de l'ICX était plus fréquent chez les cellules myéloïdes que chez les cellules T ou les cellules buccales et par le fait que le biais de l'ICX est dépendant de l'âge (54) . Busque *et al.* ont émis l'hypothèse que le biais de l'ICX était probablement multifactoriel et que certains individus pourraient acquérir une mutation qui induit l'hématopoïèse clonale. Pour tester

cette hypothèse, Busque *et al.* ont recruté 2292 femmes de 37 à 101 ans d'origine canadienne-française sans aucun trouble hématologique connu. La cohorte comprenait 1734 individus appartenant à 311 familles et 558 personnes non apparentées. Ils ont d'abord décidé d'analyser à l'aide du séquençage d'exome l'ADN de 3 femmes âgées présentant un biais de l'ICX connu dans leurs cellules myéloïdes et leurs cellules T polyclonales. Ils ont identifié dans une de ces 3 femmes des mutations somatiques dans *TET2*, *DNMT3A* et *SLC39A12* (4). L'extension de l'analyse aux différents groupes d'âge a conduit à l'identification des mutations somatiques de faux-sens, de non-sens et du décalage de cadre de lecture de *TET2* dans l'ADN des polynucléaires neutrophiles (PMN). Ces mutations n'étaient pas présentes dans les cellules lymphocytaires ou épithéliales dans 10 des 179 (5,6%) sujets âgés avec un biais de l'ICX, aucun sur 105 sujets âgés sans biais de l'ICX, et aucun des 96 sujets plus jeunes avec un biais de l'ICX. Les paramètres hématologiques des individus qui ont une mutation sur *TET2* ne différaient pas de leurs homologues du même âge. Cette étude a été la première à signaler qu'une mutation acquise dans un gène associé au cancer myéloïde, *TET2*, dépend de l'âge et est compatible avec une hématopoïèse normale (4). Busque *et al.* ont également suggéré qu'une mutation acquise dans un gène conducteur est un petit contributeur à l'acquisition du biais de l'ICX dans la population générale. C'est justement ce point important qui sera adressé dans cette thèse après avoir observé le lien possible entre le biais de l'ICX et les mutations somatiques acquises.

Une autre étude de Laurie *et al.* a rapporté un mosaïcisme clonale dépendant de l'âge dans les grandes anomalies chromosomiques en utilisant des données de bio-puces de SNP chez

plus de 50 000 personnes recrutées pour une études d'association pangénomique (GWAS). Ils ont démontré que le mosaïcisme clonal a été détecté chez 2% à 3% des personnes âgées de plus de 50 ans versus 0,5% chez des personnes âgées de moins de 50 ans. Des caractéristiques de nombreuses anomalies de la mosaïque clonale sont également retrouvées dans les cancers hématologiques comme la suppression de régions communes de gènes associés à ces cancers (79). Une étude similaire de Jacobs *et al.*, où 31 717 cas de cancer et 26 136 individus sans cancer provenant de 13 études de GWAS ont été analysés, a observé de grandes anomalies chromosomiques dans un sous-ensemble de clones dans l'ADN obtenu à partir d'échantillons de sang ou des cellules buccales (80). Par ailleurs, une autre étude par Forsberg *et al.* ont démontré une accumulation des différentes variantes dans les génomes nucléaires *in vivo* dans des cohortes stratifiées selon l'âge chez 318 jumeaux monozygotiques et chez 296 individus célibataires. Ils ont aussi observé des régions récurrentes et des mutations sur des gènes spécifiques où la plupart de ces mutations sont des délétions (81).

En 2014, trois groupes ont rapporté un ensemble d'analyse de données d'exome d'ADN provenant de grandes cohortes où ils ont observé la présence des mutations dépendantes de l'âge dans des gènes clés (82) (83) (84). Bien que ces trois cohortes aient des critères d'inclusion différents (cancer non hématologique, troubles psychiatriques ou maladies cardiovasculaires), ils ont toutefois identifié un ensemble similaire de gènes, suggérant un phénomène universel lié à l'âge. Ainsi, plus de 70 gènes différents ont été identifiés. Ceux qui étaient les plus fréquemment mutés étaient les modificateurs épigénétiques *DNMT3A*, *TET2* et *ASXL1*, surnommés les mutations *DAT*. Les données prospectives disponibles pour



deux de ces études ont révélé que les individus atteints d'hématopoïèse clonale avaient un risque accru d'environ 10 fois de développer un cancer hématologique (84) (6). Le risque relatif de mortalité était modeste à 1.4. Cependant, l'hématopoïèse clonale était associée à un risque accru d'évènements cardiovasculaires (85). Ceci suggère qu'il pourrait y avoir un lien fort entre les mutations somatiques acquises et la possibilité de développer des maladies cardiovasculaires. Pour bien clarifier les différentes nomenclatures utilisées, l'hématopoïèse clonale liée à l'âge (ARCH) fait référence à l'expansion sans rapport avec le type de mutation (86) tandis que le CHIP est causé par une mutation chez un gène conducteur à un VAF > 2% (87). Notre équipe a aussi rapporté l'analyse d'une grande cohorte de 2530 individus normaux âgés de 55 à 101 ans en utilisant une approche plus sensible qui cible les gènes CHIP (88). Buscarlet *et al.* ont documenté une forte prévalence de mutations et une prépondérance encore plus élevée pour les mutations dans DNMT3A ou *TET2* qui sont des gènes essentiels à la méthylation et la déméthylation de l'ADN. La taille du clone, c'est-à-dire le VAF, était significative dans une grande proportion d'individus affectés. Le VAF moyen était de 14,3% ce qui correspond à 28,6% des cellules provenant de cellules souches mutées. Près de la moitié de la cohorte mutée (48,7%) avait un VAF  $\geq$  10%. Or, il n'y avait pas de différences significatives dans les paramètres des cellules sanguines, par exemple le nombre des cellules, entre les mutants et les contrôles. Les modèles de restriction de la lignée DNMT3A sont compatibles avec des cellules souches pluripotentes alors que les mutations *TET2* se produisent principalement dans les cellules myéloïdes et parfois dans les cellules B (89). Plus inquiétant pour notre compréhension actuelle des lésions pré-malignes, Young *et al.* ont rapporté des mutations

dans *TET2* ou *DNMT3A* à de très basses fréquences chez 95% des individus âgés de 50 à 60 ans, suggérant que ces mutations sont presque omniprésentes après 50 ans (90).

## **1.2.6 Les mécanismes des mutations les plus fréquentes DAT**

### **1.2.6.1 *TET2* and *DNMT3***

La méthylation et la déméthylation de l'ADN jouent un rôle essentiel dans la régulation des gènes et dans la formation de l'identité cellulaire. La méthylation, en particulier au niveau des promoteurs, est généralement considérée comme une marque répressive. Par contre la déméthylation est généralement considérée comme une marque d'activation de gène. Au cœur de ce processus de régulation se trouvent les enzymes responsables de l'établissement des marques de méthylation et de la déméthylation, l'ADN méthyl transférase (*DNMT*) et la famille des dix-onze translocases (*TET*), qui sont capables d'éliminer ces marques. En général, le *DNMT1* est considéré comme ayant un rôle dans le maintien des marques de méthylation établies, tandis que le *DNMT3A / B* est responsable de l'établissement des marques de méthylation *de novo* (Figure 6, page 29) (91). Les trois membres de la famille *TET*, *TET1*, *TET2* et *TET3* sont des méthylcytosine dioxygénases principalement responsables de la catalyse de la conversion de la 5mC en 5-hydroxyméthylcytosine (5hmC) et à des événements d'hydroxylation ultérieurs (92). La perte de *TET2* est associée à l'hyperméthylation de l'ADN, tandis que les mutations de perte de fonction dans l'ADN méthyltransférase 3A (*DNMT3A*) sont associées à l'hypométhylation. Cette association persiste dans la transformation maligne (93). Alors que d'autres membres de la famille *TET* et *DNMT* sont capables de catalyser ces réactions, *TET2* et *DNMT3A* restent les

principaux membres de la famille qui sont mutés dans les cancers hématologiques (94) , et les gènes les plus fréquemment mutés dans l'HC (95).

Les conséquences de la perte de fonction de DNMT3A ont été reportées par des études mécanistiques chez la souris. Le *knock-out* complet de *Dnmt3a* et *Dnmt3b* chez la souris, et donc l'élimination de la méthylation de l'ADN *de novo*, a conduit à une réduction de l'auto-renouvellement de CSH, mais sans biais dans l'engagement de lignage (96). Des études ont révélé que, avec le temps, la perte de *Dnmt3a* seule était suffisante pour entraîner l'expansion des cellules souches, conduisant inévitablement au développement de leucémies lymphoïdes et myéloïdes. Ces leucémies présentent une hypométhylation globale avec des altérations des lignées cellulaires spécifiques (97). Pour comprendre le rôle de la méthylation dans l'HC, Spencer *et al.* ont effectué un séquençage au bisulfite du génome entier du sang d'un patient de 9 ans qui possédait une mutation de la lignée germinale DNMT3A-R882H (98). Ils ont trouvé une hypométhylation similaire à celle des patients atteints de LMA par rapport à un frère ou une sœur en bonne santé. Malgré les mises en garde concernant la nature germinale de la mutation et l'âge du patient, ces données suggèrent que le modèle d'hypométhylation, trouvé dans la LMA où *DNMT3A* est muté, précède le développement leucémique et peut jouer un rôle dans la progression de l'HC (98).

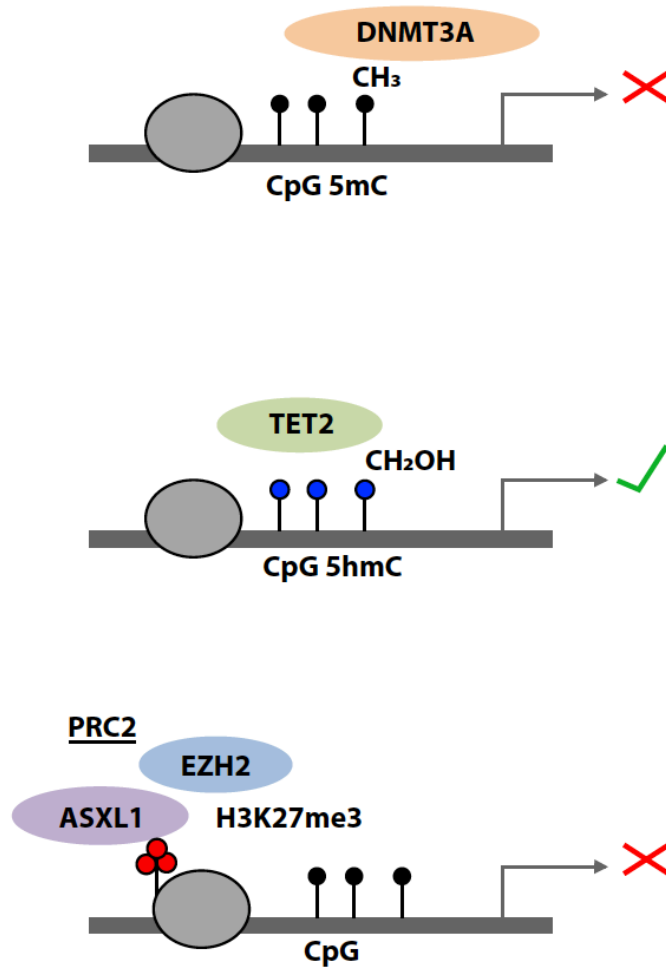
Contrairement à la perte de *DNMT3A*, les mutations de perte de fonction *TET2* sont associées à une hyperméthylation. Des études chez la souris utilisant un mutant de perte de fonction *TET2* ont démontré une expansion des cellules Lin<sup>-</sup> Sca1<sup>+</sup> cKit<sup>+</sup> (LSK) avec une

diminution des niveaux de 5hmC (99). Notre équipe a démontré que chez les individus sains qui sont HC, le rapport de 5hmC/5mC correspondait avec *TET2* VAF, de sorte que le VAF supérieur conduit à une baisse de 5hmC. Ceci soutient la prévalence de ces processus biochimiques chez l'humain (88). Par contre, les mécanismes exacts par lesquels la perte de fonction de *TET2* entraîne une augmentation de l'expansion des cellules souches restent inconnus. Une étude suggère que la perte de 5mC et les augmentations de 5hmC sont associées à la différenciation, ce qui indiquerait que *TET2* joue un rôle essentiel dans le développement des érythrocytes en régulant les gènes spécifiques de la lignée via la déméthylation oxydative de l'ADN par *TET2* (3, 100).

#### **1.2.6.2 ASXL1**

En plus de la régulation de la méthylation de l'ADN décrite ci-dessus, la régulation épigénétique des histones par des modifications post-traductionnelles est un autre régulateur majeur de l'expression génique et du fonctionnement des cellules (101). Le complexe répressif *polycomb* (PRC) présente un intérêt particulier compte tenu de son rôle dans la biologie des cellules souches (102). *ASXL1* est l'un des trois homologues mammifères du gène de la drosophile *Asx*, membre du PRC, qui présente ses effets à travers deux PRC différents. Le complexe PRC2 catalyse la méthylation de la lysine 27 de l'histone H3 (H3K27me3), une réaction qui est exécutée par les méthyltransférases EZH1 et EZH2. Pendant ce temps, le complexe PRC1 se lie à H3K27me3 par la protéine CBX et catalyse l'ubiquitination de la lysine 119 sur l'histone H2A (H2AK119ub) (103). Ces deux marques conduisent à une répression de la transcription et à la formation d'un état

d'hétérochromatine à plusieurs locus à travers le génome, y compris les gènes *Hox* (104).  
 (104) (3).



**Figure 6:** Régulateurs épigénétiques de l'hématopoïèse clonale (adaptée de (3))  
 Schéma de DNMT3A catalysant la méthylation de CpG (•). *TET2* générant 5hmC (•).  
 ASXL1 générant une triméthylation H3K27 (•) en recrutant le membre du complexe  
 PRC2 EZH2.

### 1.3 L'étiologie de ARCH/CHIP

#### 1.3.1 L'impact de la prédisposition génétique sur la progression des mutations dans les CSH

Les facteurs, autres que le vieillissement, qui prédisposent à l'acquisition et à la progression de ces mutations sont jusqu'à présent moins connus (Annexe 2, Figure 2, page 164). La prédisposition génétique a été rapportée dans plusieurs cancers, y compris les NMP qui partagent plusieurs mutations en commun avec l'HC. Dans les NMP, il existe une prédisposition génétique à acquérir les mutations JAK2 V617F conférées par un haplotype germinale (105) (106). Des études de GWAS dans des groupes familiaux de NMP ont identifié un certain nombre de variantes génétiques associées à un risque accru de développer un NMP clonal. L'association la plus forte découverte à ce jour est la présence de l'haplotype *JAK2* 46/1. Aussi, de nombreuses études ont trouvé des variantes supplémentaires dans d'autres gènes, notamment dans le gène de la transcriptase inverse de la télomérase (*TERT*) (107). Pour ce qui concerne l'HC, Zink *et al.* ont trouvé une délétion germinale dans l'intron 3 du gène *TERT* qui prédispose à l'HC. Notre équipe Buscarlet *et al.* ont aussi démontré pour la première fois, l'héritabilité des mutations à l'aide d'une cohorte de fratrie. Ils ont identifié un risque familial significatif de 2,7 pour les mutations *TET2* mais pas pour *DNMT3A*. La mutation *TET2* montre une agrégation familiale suggérant un locus prédisposant (95). L'étude importante de Thompson *et al.* a documenté une contribution génétique significative à la perte de cellules sanguines du chromosome Y associée à l'âge chez les hommes, une situation qui partage plusieurs similitudes avec l'HC puisque la perte du chromosome Y est une forme de mosaïcisme clonal qui a été associée

au développement du cancer (108) et des évènements cardiovasculaires (109). En plus des facteurs génétiques, le microenvironnement des CSH joue un rôle important dans la clonalité des cellules malignes.

### **1.3.2 Le rôle de l'inflammation dans l'HC**

Le microenvironnement des CSH est un autre facteur qui a suscité un grand intérêt dans la pathogenèse de l'ARCH. En fait, Vas *et al.* ont rapporté que le microenvironnement vieillissant influence la clonalité des cellules malignes de manière indépendante du CSH et peut donc également contribuer à l'ARCH (110). L'association de CHIP avec des maladies cardiovasculaires (84) (111) et pulmonaires chroniques (112) (95) soulève la possibilité que l'inflammation chronique associée à l'âge (113) peut être un dénominateur clé commun entre ces conditions médicales et l'émergence de l'HC. Il existe des preuves suggérant qu'un milieu pro-inflammatoire favorise les cellules souches mutées. Par exemple, les cellules souches de la leucémie myéloïde chronique (LMC) ont une expression accrue du récepteur de l'interleukine (IL) -1 et une sensibilité à l'IL-1 par rapport aux CSH normales et elles sont donc favorisées dans un environnement inflammatoire (114). Abegunde *et al.* ont découvert qu'un environnement pro-inflammatoire soutenu par le TNF-alpha favorise l'expansion des clones mutants Tet2 chez la souris (115). On ne sait actuellement pas si l'inflammation favorise l'hématopoïèse clonale indépendamment d'autres conditions médicales associées à l'âge ou si elles sont liées d'une manière ou d'une autre. Plus intéressant, une autre étude a démontré que les cellules myéloïdes mutantes Tet2 ont une capacité déficiente à résoudre l'inflammation causée par une production accrue d'IL-6 (116) et les souris déficientes en Tet2 sont prédisposées à l'athérosclérose

(117). Récemment, Meisel *et al.* ont constaté que la myéloprolifération pré-leucémique n'était observée que chez les souris déficientes en Tet2 (Tet2 - / -) avec un dysfonctionnement de leur petite barrière intestinale associée à une translocation bactérienne et à une production accrue d'IL-6. La myéloprolifération pré-leucémique n'a pas été développée chez les souris Tet2 - / - *germe-free* traitées aux antibiotiques et la neutralisation de la myéloprolifération pré-leucémique par IL-6 a inversé la tendance anti-inflammatoire, ce qui suggère qu'une intervention anti-inflammatoire peut empêcher l'expansion clonale (118). De plus, les cytokines pro-inflammatoires sont impliquées dans la pathogenèse des SMD et sont en cours d'évaluation comme cibles thérapeutiques potentielles (119).

Chez l'homme, Jaiswal *et al.* ont constaté que chez les patients atteints de CHIP et de maladies cardiovasculaires, les niveaux d'IL-8 étaient plus élevés dans un sous-ensemble de patients porteurs de mutations *TET2* (85). Cook *et al.* ont rapporté des niveaux accrus d'IL-6 et des comorbidités plus grandes chez les patients atteints de CHIP (120). Plus récemment, notre équipe Busque *et al.* a découvert dans une cohorte de 1887 personnes de plus de 70 ans que les porteurs de CHIP avaient des niveaux 21% plus élevés de CRP à haute sensibilité que leurs homologues non porteurs ( $eb = 1,21$ , intervalle de confiance à 95% [IC]: 1,08 - 1,36,  $p = 0,001$ ), confirmant davantage l'association entre l'inflammation et CHIP (ANNEXE 3, page 176) (121). En conclusion, ces études soutiennent l'hypothèse que l'inflammation favorise l'expansion clonale initiale des CSH mutées et accélère la transformation maligne et le développement de leucémie.



Les causes potentielles d'ARCH discutées jusqu'à présent s'appliquent à la population vieillissante normale. Cependant, dans certaines conditions iatrogènes, une sélection et une expansion clonale peuvent se produire, c'est-à-dire dans le cas d'exposition génotoxique.

### **1.3.3 L'impact de l'exposition écotoxique**

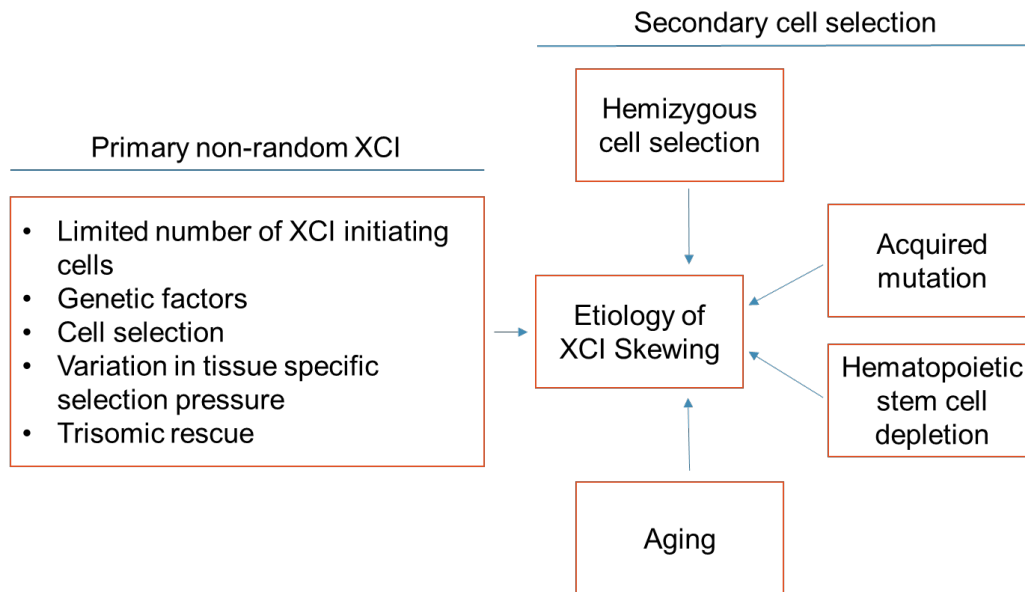
Les analyses clonales du sang de patients qui ont développé une LMA liée au traitement (t-LMA) ont contribué à élucider la physiopathologie de ces complications. Des mutations spécifiques tels que TP53 ou PPM1D peuvent conférer au CSH une résistance au stress génotoxique tel que la chimiothérapie ou la radiothérapie et favoriser l'expansion. Ces mutations peuvent conduire au développement de la t-LMA (122) (123) (124). Récemment, une étude génomique sur une grande cohorte de patients a révélé que les mutations TP53 et PPM1D étaient enrichies dans les SMD induits par le traitement par rapport aux SMD primaires, offrant des preuves à l'appui que les clones hébergeant ces mutations se développent sous pression sélective induite par le traitement. Ces études suggèrent que le stress génotoxique exogène sélectionne des clones naturels qui hébergent des mutations conférant une résistance à ces agents toxiques. Cependant, la valeur prédictive des mutations pour cette complication rare peut être spécifique de certains gènes, comme *TP53* et *PPM1D*, et ne pas être la même pour d'autres gènes, comme *TET2* par exemple (122).

### **1.3.4 L'impact de l'attaque du système immunitaire**

L'attaque immunitaire de CSH favorise la sélection de cellules avec un profil mutationnel différent de celui du vieillissement ou du stress génotoxique. Chez les patients plus jeunes, cela inclut des mutations dans les délétions *PIGA*, *BCOR / BCORL1* et *HLA*. Chez les

individus plus âgés, des mutations dans *DNMT3A* et *ASXL1* sont également documentées, par contre les mutations *TET2* sont sous-représentées (123). Des mutations liées à l'ARCH ont également été identifiées dans le contexte des syndromes d'insuffisance médullaire (125). Le schéma des mutations identifiées dans l'anémie aplasique présente certaines différences avec le schéma observé dans la population vieillissante. L'anémie aplasique est une maladie qui touche la moelle osseuse, où la formation des cellules du sang (globules rouges, globules blancs et plaquettes) est très diminuée en raison d'une destruction des CSH dans la moelle osseuse. Les patients atteints d'anémie aplasique présentent une prévalence élevée de mutations (45%) qui peuvent être sous catégorisées comme des mutations plus spécifiques de l'anémie aplasique (*PIGA*, *BCOR* et *BCORLI*) et d'autres mutations classiquement documentées dans ARCH (*DNMT3A* et *ASXL1*). En effet, l'incidence des mutations *DNMT3A* et *ASXL1* augmente avec l'âge tandis que les mutations *PIGA* et *BCOR/BCORLI* sont présentes dans tous les groupes d'âge (126). Les mutations *PIGA* et *BCOR/BCORLI* ont été associées à une réponse accrue à l'intervention immunothérapeutique tandis que les mutations ARCH classiques ont été associées à une moindre sensibilité. Il semble que, selon la mutation spécifique, certaines peuvent favoriser l'évasion de l'attaque immunitaire tandis que d'autres peuvent favoriser l'auto-renouvellement du CSH (122).

### 1.3.5 Sommaire des causes de biais de l'ICX



**Figure 7:** Sommaire des causes de biais de l'ICX

### 1.3.6 Sommaire des données scientifiques

Les points principaux à retenir de cette revue de littérature est : (i) l'importance des études de clonalité sur l'évolution de l'oncologie moderne, (ii) la découverte du ICX, son rôle dans les études de la clonalité, le mécanisme de son inactivation et les méthodes permettant de le mesurer, (iii) l'analyse de l'ICX dans la population normale et la découverte du biais de ICX, (iv) l'étiologie du biais de l'ICX, (v) le rôle des mutations somatiques ARCH/CHIP dans la population normale et son rôle avec le biais de ICX, et (vi) l'étiologie de ARCH/CHIP.

## 1.4 Projets de recherches

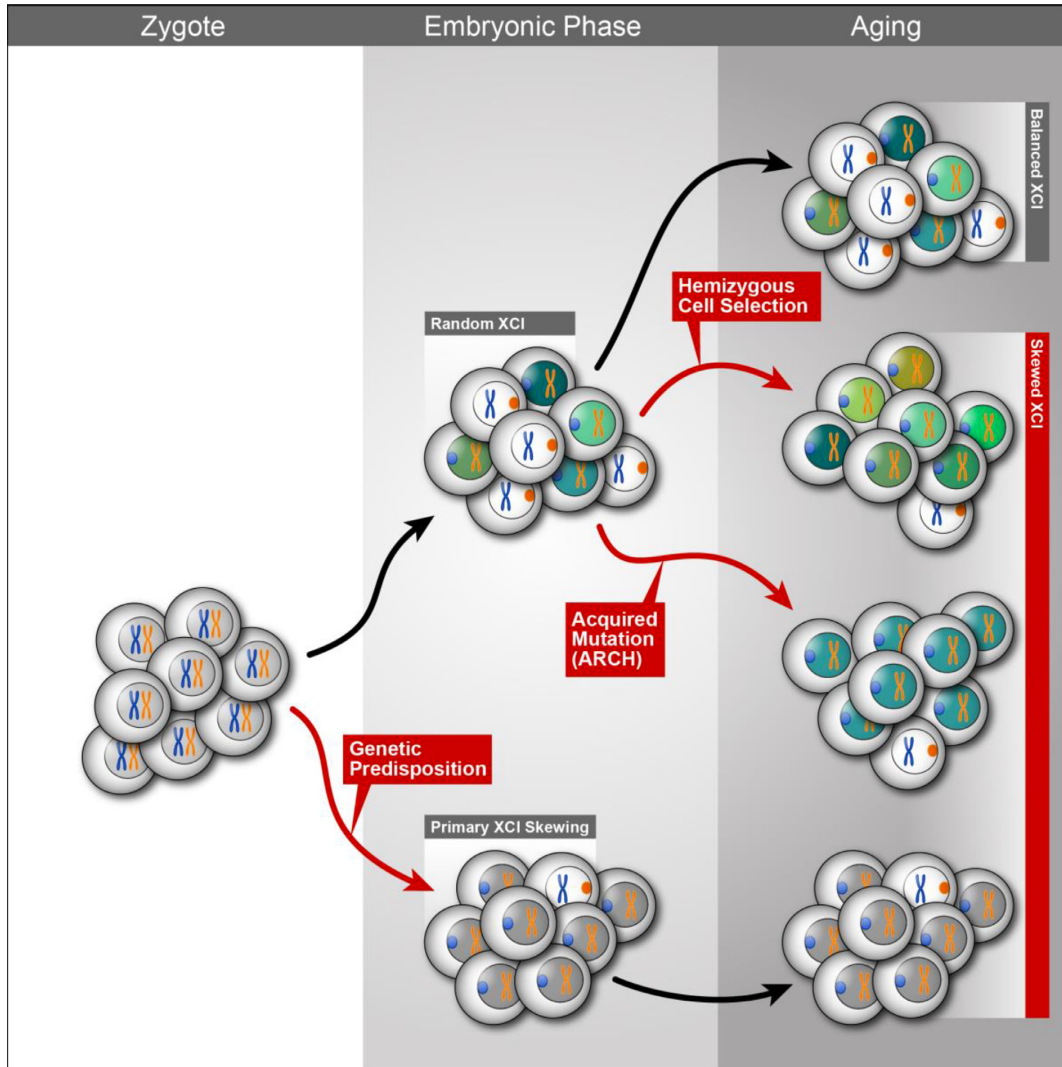
Le développement d'une maladie implique souvent des processus dynamiques qui commencent des années ou des décennies avant l'apparition clinique. Dans de nombreux cas, cependant, le processus de pathogenèse n'est détecté qu'après que le patient ne présente des symptômes qui conduisent à une maladie cliniquement apparente (127). Le cancer se développe en raison des effets combinés de multiples mutations somatiques, qui sont acquises à des moments différents de la vie (128). Dans certains modèles de développement du cancer, des mutations précoces conduisent à des expansions clonales par des cellules souches ou d'autres cellules progénitrices (129). De telles expansions clonales augmentent considérablement la probabilité que des mutations coopérantes ultérieures surviennent dans des cellules qui contiennent déjà les mutations initiatrices précédentes (84). Le fait que les cellules soient clonales signifie qu'elles sont dérivées d'un seul précurseur qui a subi une mutation somatique. La clonalité est une description de la lignée ainsi que l'homogénéité du tissu testé. Ces aspects de clonalité peuvent être observés dans l'acquisition d'anomalies cytogénétiques sur la transformation leucémique dans la myélofibrose avec métaplasie myéloïde (130) Pour comprendre la pathogenèse des maladies prolifératives, il est important de savoir dans quelle mesure les expansions clonales se produisent et précèdent le cancer (127). La dynamique des populations des cellules souches hématopoïétiques peut précéder de nombreux cancers hématologiques, y compris les NMP (131), les SMD (132), et la LMA (133, 134). Par exemple, chez certains patients, les cellules souches portant un ensemble des mutations présentes dans les cellules cancéreuses sont capables de survivre à la chimiothérapie ; par la suite, ces cellules

acquièrent de nouvelles mutations, déclenchant une rechute (135-137). Cela suggère que des expansions clonales de cellules souches peuvent avoir existé avant le développement du cancer (84).

La découverte du principe de base selon lequel le cancer est clonal est étroitement reliée à l'évolution de méthodes suffisamment sophistiquées pour déterminer de façon précise l'expansion clonale des cellules (138, 139). Une des premières méthodes qui a prouvé la clonalité est l'ICX chez les femmes saines (15, 16, 32). Le développement des études de clonalité a joué un rôle déterminant en médecine et en hématologie. Ces études ont été utilisées pour établir l'origine unique des tumeurs et pour différencier les états non malins des états malins (140). Plus important encore, elles ont été utilisées pour élucider la hiérarchie hématopoïétique (141, 142).

Tel que décrit précédemment, l'ICX est un processus qui se déroule lors du développement embryonnaire chez la femme où l'un des chromosomes X est inactivé dans chaque cellule. La descendance ultérieure de chaque cellule conserve le même chromosome X inactivé, ce qui résulte en un organisme qui est une mosaïque de cellules exprimant les gènes du chromosome X<sub>p</sub> ou X<sub>m</sub> (Figure 1, page 6). L'ICX est un processus stochastique puisque le fait qu'une cellule germinale et sa progéniture expriment X<sub>p</sub> ou X<sub>m</sub> dépend du hasard (84). L'hypothèse retenue est que le ratio de l'ICX entre les allèles des gènes liés au chromosome X entre le X<sub>p</sub> et le X<sub>m</sub> est de 1:1. Si une déviation significative du ratio de l'ICX se présente, ceci signifie qu'il y a un biais de l'ICX où la plus grande part de

l'expression des gènes provient d'un seul allèle Xp ou Xm (Figure 8, page 38, Modèle explicatif de l'initiation du biais de l'ICX selon l'âge) (54-56).



**Figure 8:** Modèle explicatif de l'initiation du biais de l'ICX selon l'âge

Dans cette thèse nous avons élaboré deux projets différents mais qui permettent d'approfondir nos connaissances sur l'impact du vieillissement sur l'HC chez l'humain dans le but de mieux comprendre le rôle des composants génétiques et des mutations somatiques. Ces projets sont :

- **Chapitre 2 :** *Etiology of skewing of X-chromosome inactivation ratios occurring in blood cells of aging women.* Sami Ayachi, Sylvie Provost, Marie-France Gagnon, Robert Mio, Manuel Buscarlet, Rene Fouodjio, Luigina Mollica, Marie-Pierre Dubé, and Lambert Busque (article en soumission à Leukemia)
- **Chapitre 3 :** Déterminants de la pénétrance oncogénique de la mutation de *TET2* chez les individus âgés.

#### 1.4.1 Mise en contexte

Notre laboratoire s'intéresse au vieillissement de l'hématopoïèse chez l'humain. Des études menées il y a 25 ans ont démontré que la clonalité, tel que démontré par l'ICX chez les femmes s'accroît avec l'âge (54). Phénomène intrigant pouvant être expliqué par plusieurs causes dont (i) la sélection hémizygote (les cellules avec un des deux X parentaux dans l'état actif dominant l'autre population) et (ii) l'acquisition de mutation donnant un avantage de croissance (*TET2*, *DNMT3A*, *ASXL1*, etc.). Ces deux causes sont complètement différentes. La première étiologie est génétiquement prédéterminée et la seconde est acquise dans les cellules somatiques de moelle osseuse (Figure 8, page 38, ANNEXE 2, page 155) (4, 70).

### **1.4.2 Hypothèse**

L'hypothèse formulée pour ce projet de recherche est que le biais de l'ICX dans les cellules sanguines des femmes vieillissantes est causé par deux phénomènes, la sélection hémizygote et les mutations somatiques acquises.

### **1.4.3 Objectifs**

Les objectifs de cette thèse consistent à :

- 1- Identifier la corrélation entre le biais de l'ICX et le vieillissement dans des tissus spécifiques.
- 2- Identifier la corrélation entre le biais de l'ICX et les mutations somatiques acquises et évaluer l'impact de CHIP sur l'ICX.
- 3- Identifier une cohorte d'individus sans mutation somatique pour tester l'héritabilité du trait de biais.
- 4- Effectuer une analyse GWAS et par liaison génétique pour identifier le SNP associé au trait de biais de l'ICX.



## CHAPITRE 2

Article

### ETIOLOGY OF SKEWING OF X-CHROMOSOME INACTIVATION RATIOS OCCURRING IN BLOOD CELLS OF AGING WOMEN

Sami Ayachi<sup>1,3</sup>, Sylvie Provost<sup>4</sup>, Marie-France Gagnon<sup>1,2,3</sup>, Robert Mio<sup>1</sup>, Manuel  
Buscarlet<sup>1</sup>, Rene Fouodjio<sup>4</sup>, Luigina Mollica<sup>1,2,3</sup>, Marie-Pierre Dubé<sup>3,4</sup>, and Lambert  
Busque<sup>1,2,3\*</sup>

<sup>1</sup>Research Center, <sup>2</sup>Hematology Division, Hôpital Maisonneuve-Rosemont, <sup>3</sup>Université  
de Montréal, Montréal, Canada, <sup>4</sup>Beaulieu-Saucier Pharmacogenomics Center, Montreal  
Heart Institute, Montréal, Canada.

¶Contributed equally to this work; \*Corresponding author.

Article en soumission à Leukemia

## **2. *ETIOLOGY OF SKEWING OF X-CHROMOSOME INACTIVATION RATIOS OCCURRING IN BLOOD CELLS OF AGING WOMEN***

Cet article met en lumière la contribution des mutations somatiques acquises et la sélection hémizygote dans l'initiation du biais de l'ICX chez les femmes vieillissantes.

### Contributions détaillées de chacun des auteurs :

J'ai effectué la majorité des expériences dans cet article, soit environ 80%. Robert Mio a effectué les premiers tests de HUMARA (environ 20%), ce qui a permis de démarrer le projet. J'ai effectué le plan des analyses statistiques, les plans pour les expériences du GWAS et de liaison génétique et j'ai participé à la réalisation de l'expérience de GWAS et de liaison génétique. Sylvie Provost, René Fouodjio et Marie Pierre Dubé ont réalisé les analyses statistiques, les analyses bioinformatiques du GWAS et de liaison génétique. J'ai également participé à l'interprétation des résultats. J'ai participé à la rédaction de l'article, plus précisément à la partie "matériel et méthodes", aux légendes des figures, et j'ai fait les figures. Manuel Buscarlet a participé dans la collecte des données. Marie-France Gagnon et Luigina Mollica ont participé à l'interprétation et la rédaction de l'article. Dr Lambert Busque a effectué la conception et le design des expériences, l'analyse et l'interprétation des résultats, la rédaction de l'article et a supervisé toutes les étapes du projet.

## 2.1 Abstract

Women are mosaics of two populations of cells differentiated by parental expression of alleles of the X-chromosomes (15). Skewing towards one or the other population occurs in aging women (54, 143). Germline variants causing hemizygous cells selection and/or clonal hematopoiesis (CH) are hypothesized as etiology of skewing. We investigated skewing in a cohort of 2172 women from 321 families and 824 unrelated individuals. We analyzed X-chromosome inactivation (XCI) ratios at the HUMARA locus in epithelial cells, neutrophils, monocytes, T cells, and B cells. We genotyped the cohort to look for germline variants and sequenced CH's associated genes. Skewing was more prevalent in myeloid cells (41%), intermediate in T cells (22.5%) and lowest in epithelial cells (14.3%). Sibship analyses demonstrated heritability of skewing ( $H^2$ : 0.20-0.31,  $p < 10^{-6}$ ). Linkage studies identified a region at Xq12 for skewing occurring in T cells (LOD score 4.8), suggesting a hemizygous cell selection process. We identified a region at 1q21 for skewing in neutrophils (LOD score 6), suggesting epistasis with X-linked genes. CH was also documented but accounted only for a small proportion of the skewing trait except in the very old. We conclude that age-associated skewing (AAS) of XCI ratios is predominantly caused by genetically-driven lifelong selection processes.

**Key words:** X-chromosome inactivation; skewing; aging; hematopoiesis; heritability; hemizygous cell selection; CHIP; GWAS; linkage.

## 2.2 Introduction

In 1961, Mary Lyon proposed that each X-chromosome in excess of one is randomly inactivated in cells of the developing female embryo (15). Therefore, following X-chromosome inactivation (XCI), females are mosaics with two cell populations expressing either alleles of maternal ( $X_m$ ) or paternal ( $X_p$ ) origin. Analysis of XCI ratios between  $X_p$  and  $X_m$  and its deviation (skewing) toward one parental allele has allowed to gain valuable insight into clonal disorders or carrier states of X-linked genetic disorders (reviewed in (1)).

Interestingly, an increased frequency of XCI skewing has also been documented in healthy aging women (54, 143), reaching ~40% in women aged 60 or more (54), and 67% in centenarians (144). A number of mechanisms have been put forward to explain age-associated skewing (AAS), including mutation-driven clonal hematopoiesis, stochastic clonal dominance, and genetic predisposition (144-147). The latter has been supported by several observations, including a strong correlation in the degree and direction of skewing between monozygotic (MZ) twins (148), a greater concordance in the direction of skewing in MZ than dizygotic (DZ) twin pairs (72). Kristiansen *et al.* (144) recently confirmed these observations in cohorts of MZ and DZ twins pairs of different age groups supporting a genetically dependent selection of blood cells with aging. Zito *et al.* reported significant heritability of skewing in blood cells of older but not in younger twins, and not in other tissues indicating that the selection process may be blood cells specific and takes several decades to manifest (149).

Abkowitz *et al.* (150) provided further evidence for an X-linked genetic component to AAS of blood cells in aging cats. They analyzed the blood cells of female Safari cats (first-cross hybrids between Geoffroy and domestic cats) and showed that XCI was random in young cats, but skewed in 67% of older cats with consistent bias towards the Geoffroy allele (150) indicating a hemizygous cell selection process. However, no groups have identified genomic region(s) or candidate genes responsible for the selection process.

Importantly, hemizygous cell selection is a polyclonal process as all cells with the same parental X in the active state will be affected. In contrast, AAS can also be caused by acquisition of a somatic mutation conferring a clonal growth advantage, as we have demonstrated in a small subset of elderly women with skewing that harbored mutation in the *TET2* gene (4). CH is a phenocopy of hemizygous cell selection in regard to XCI measurement. Fortunately, it is now possible to identify subjects with CH to better phenotype subjects for analysis.

Deciphering the different etiology of AAS is a daunting task as it is multifactorial, age-dependent, and limited to women. In this study, we used a genomic approach in a cohort of nearly 3000 women including 321 families and unrelated subjects of French ancestry from Quebec, and identified for the first time evidences for germline and acquired genetic contribution to the trait.

## 2.3 Materials and Methods

### Subjects

The study population included 3002 related and unrelated women of French ancestry recruited from the general community in the province of Quebec, Canada. Six women were excluded (active cancer, three X chromosomes), leaving 2996 active subjects. The *family cohort* comprised 2172 women from 321 families, aged between 30 and 101 years (mean  $63.5 \pm 10.5$  years). Families had 3-12 female siblings (mean 5.3, median 5). In 68 families, the mother was also recruited. The *unrelated cohort* comprised 824 women with an average age of  $73.5 \pm 9.0$  years (range 50 to 100 years). All women answered a medical questionnaire and gave an informed consent. The study was approved by the Maisonneuve-Rosemont Hospital's Ethics Committee.

### Specimens

Blood cells were obtained by venipuncture. Buccal epithelial cells (BC) were obtained using a cotton and paper swab (Whatman Bioscience, NJ, USA). Complete blood counts were obtained from a GenS automated cell counter (Beckman Coulter, CA, USA). Blood cells were separated into polymorphonuclear (PMN) and mononuclear fractions by standard density gradient centrifugation (Ficoll-Paque, GE Healthcare, NJ, USA). Fluorescence-activated cell sorting (FACS) was used to isolate T cells (CD3+), B cells (CD19+), and monocytes (CD14+) from the mononuclear cell fraction (FACStar-plus, Becton Dickinson, CA, USA). As interim analysis of all blood cell types from 1144 subjects from the family cohort showed that PMNs and T cells presented the most

discordant XCI ratios, only these two blood cell types were analyzed for the remainder of subjects. DNA and RNA were isolated from cells as previously described (151).

## **XCI ratio determination**

### HUMARA assay

HUMARA assay was performed as initially described (46) with minor modifications (151).

### XCI scores.

1. *Degree of skewing (DS) score.* This score expresses the deviation from a theoretical 50% XCI ratio between  $X_p$  and  $X_m$ , and ranges from 0 (no skewing) to 0.5 (complete skewing). Individuals with a DS score  $\geq 0.25$  are qualitatively categorized as skewed. A DS of 0.25 corresponds to a 3:1 ratio between  $X_p$  and  $X_m$ , and this threshold has been widely used in the literature to define skewing. (44, 54-56, 152, 153) The DS does not provide insight into the direction of skewing ( $X_p$  vs  $X_m$ ).

2. *Proportion of cells presenting the superior HUMARA allele (greater number of CAG repeats) on the active X chromosome (Psup) score.* Psup specifies the direction of skewing as females have the same superior allele in all cells. Psup can be used to establish intra-individual skewing correlations. Psup is also a surrogate and partial indicator for parental origin of X chromosome when comparing sibs from the same family for heritability, genome wide association studies (GWAS) or linkage analyses. However, Psup cannot ascertain the parental origin of X chromosomes in sisters that have inherited different maternal X chromosomes.

The formulas used to calculate skewing scores are available in Supplemental Material 1.

### **Myeloid gene sequencing to identify clonal hematopoiesis of indeterminate potential (CHIP)**

PMNs were sequenced at high coverage on an Ion Proton sequencer (Thermo Fisher Scientific) using the Ampliseq AML panel, as previously described (95). The AML panel covers the following 19 genes which are recurrently mutated in myeloid cancers: *DNMT3A*, *TET2*, *ASXL1*, *TP53*, *JAK2*, *BRAF*, *CBL*, *CEBPA*, *FLT3*, *GATA2*, *IDH1*, *IDH2*, *KIT*, *KRAS*, *NPM1*, *NRAS*, *PTPN11*, *RUNX1*, and *WT1* (95). Each sample was sequenced at a minimum of 4000x mean coverage (corresponding to 95% >500x). CHIP was considered present if the Variant Allele Fraction (VAF) was  $\geq 2\%$ .

### **Genotyping for GWAS and Linkage analyses**

Genome-wide genotyping was performed on 2996 individuals using 200ng of their genomic DNA. The Illumina Infinium Global Screening Array v3-MD (Illumina, San Diego, CA) including 700,625 genomic markers was used and processed according to the manufacturer's specifications. Imputation was performed using IMPUTE2 (v2.3.2) and phasing was performed using the SHAPEIT2 algorithm (v.2r790) using the 1000 Genomes reference data released on June 16 2014 including all population samples (distributed through the IMPUTE2 web site).(154-156) A total of 11,345,359 genetic variants with imputation probability of  $\geq 0.80$  and completion rate of  $\geq 98\%$  were obtained. For GWAS analysis, a total of 4,499,752 genetic variants with minor allele frequency (MAF)  $>5\%$



were included. The reference genome GRCh37 (hg19) was used for SNP positions. Details of data cleanup and imputation are presented in Supplemental Material 1.

### **Statistical analyses for descriptive variables**

Descriptive statistics were performed using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). We excluded outliers as defined by values falling outside of mean  $\pm 4 \times \text{STD}$ . Normality of each continuous endpoint (without outlier) was evaluated prior to the statistical analysis using the Kolmogorov-Smirnov test at a 5% significance level. As continuous traits did not follow a normal distribution, inverse normal transformation was performed. Generalized linear regression models were adjusted for age and family when appropriate. Bonferroni correction was applied for multiple testing.

### **Heritability**

*Heritability in nuclear families was evaluated using SOLAR Eclipse version 8.4.2 (General). Mega2 version 5.0.0 was used to produce files in SOLAR format.* The heritability of skewing was evaluated in 1734 women belonging to 311 families from the family cohort.

### **Genome Wide Association Studies (GWAS)**

GWAS were performed with SAIGE\_0.43.3 by using genotyped SNPs to construct a genetic relationship matrix (GRM) in order to perform genetic variant association tests (genotyped and imputed) while accounting for family relatedness (157). A threshold of  $5E-08$  was used to assess for genome-wide significance. Markers from pseudo-autosomal regions were analyzed as autosomal variants.

## **Linkage**

We applied linkage analysis in the familial cohort. The Québec population has expanded rapidly from 8500 founders over 400 years (reviewed in (158)), enriching the population for possible founder mutations and large haplotype segments. Linkage analyses were performed on nuclear families using MERLIN 1.1.2 for two-point analysis and MINX to analyze X chromosome markers (159). Mega2 v5.0.0 was used to produce files in QTDT format from PLINK file format. Allele frequencies were computed from genotyped pedigree founders, randomly selected samples from each pedigree, and singleton individuals. No Mendelian error were detected using PEDCHECK 1.2.). Only genotyped SNPs were used. Analyses are performed using Variant Component (VC) and Quantitative Trait Loci (QTL) analyses. The skewing phenotype based on DS and Psup scores were analyzed separately for PMN and T cells, as these were the two phenotypes with the highest number of pedigrees available for analysis (approximately 300). Sensitivity linkage analyses were conducted in CHIP-free individuals only. To account for the multiple phenotypes (8 phenotypes), the significance of LOD score was set at 3.9.

## **2.4 Results**

### **XCI ratios**

XCI ratios were determined in 2996 women. Of these, 306 (10.2%) were homozygous at the HUMARA locus and were therefore excluded from the skewing analyses. For the remaining 2690 participants, HUMARA scores were not available in all tissues. The available data was as follows: PMN (DS n=2536, pasup n=2364), monocytes (DS n=897,

pasup n=892), T cells (DS n=2244, pasup n=2245, B cells (DS n=936, pasup n=932) and BC (DS=2341, pasup n=2339). HUMARA reproducibility of results was evaluated by the “coefficient of repeatability” described by Bland et Altman (160) with a variance of less than 3.3% for 1027 duplicate PMN samples.

### **Prevalence of skewing in different tissues and cell types**

The prevalence of skewing, defined by a DS score  $\geq 0.25$ , was evaluated in epithelial cells and in various blood cell types. Only women for whom samples from all cell types were available (n=777) were included in this analysis to control for age (mean age  $63.3 \pm 9.7$  years). Of these women, 14.3% had skewing in epithelial cells (BC), 41.4% in PMN, 41.7% in monocytes, 30.1% in B cells, and 22.5% in T cells (Figure 9, page 64).

### **Association between skewing and biological parameters**

No association was evidenced between skewing (DS scores) and peripheral blood counts (Table 1, page 62). Skewing was also independent of male/female child ratio, parity, number of miscarriages, and smoking status (data not presented). Table 2 shows the most common medical conditions reported by study subjects, none of which was associated with skewing.

### **Intra-individual correlations of XCI ratios between different blood cells and epithelial cells**

We observed a strong correlation of P<sub>sup</sub> scores amongst different blood cell types (Figure 10, page 65, Table 3, page 63). P<sub>sup</sub> scores of PMN were best correlated with those of

monocytes ( $r=0.9065$ ), followed by those of B cells ( $r=0.8404$ ), and of T cells ( $r=0.7473$ ). Monocytes had P<sub>sup</sub> scores which correlated better with those of B cells ( $r=0.8148$ ) than of T cells ( $r=0.6734$ ). The correlation between T and B cells ( $r=0.7228$ ) was comparable to that between T cells and either PMN ( $r=0.7473$ ) or monocytes ( $r=0.6734$ ). However, the XCI ratios of epithelial cells and blood cells were correlated to a lesser degree ( $r=0.4784$  to  $0.6105$ ), suggesting common and tissue-specific etiology of skewing.

### **Age effect on skewing**

The wide age range of our cohort allowed us to evaluate the influence of age on skewing prevalence. In epithelial cells, DS was stable across ages (BETA=-0.0009,  $P=0.62$ ), implying that age does not affect skewing prevalence in these cells. In contrast, the prevalence of skewing increased with age in hematopoietic cells. The age effect was maximal in PMN (BETA =0.0142,  $P<0.000001$ ) and monocytes (BETA =0.0121,  $P=0.00055$ ), lower in T cells (BETA=0.0071,  $P=0.00037$ ), and not statistically significant in B cells (BETA =0.0068,  $P=0.049$ ) for which less samples were available for analysis. The steady increase in skewing prevalence with age in PMN in contrast to T cells is depicted in Figure 11, and raises the hypothesis of specific additional age-dependent etiology for this lineage, such as CHIP acquisition.

### **CHIP identification and contribution to skewing**

CHIP data was available for 2530 of the 2690 individuals informative for XCI ratios (94%). CHIP was present in 347/2530 (13.71%) individuals and CHIP VAF correlates with XCI skewing as previously reported (95). This is expected, as the increase in clonal cells will

influence XCI ratios. However, to evaluate the contribution of CHIP to PMN skewing, the relative proportion of CHIP vs non-CHIP individuals and the VAF of CHIP must be taken into account. A VAF of 25% (50% clonal cells) is needed to correspond to a  $DS \geq 0.25$  in the absence of other causes of skewing. Figure 11 shows that CHIP is age-dependent, however, the relative proportion of individuals with CHIP at a VAF of 10% or 25% is relatively small until 80 years of age. In very old individuals, CHIP may account for a significant proportion of PMN skewing.

### **Heritability of skewing**

The heritability of skewing was evaluated in 1734 women from 311 families using a sib-pair approach. Heritability of DS and P<sub>sup</sub> scores is presented in Table 4. As DS scores do not take into account the direction of skewing, sib-pairs skewed in the direction of the same or the opposite parental allele were considered concordant for the skewing trait. In contrast, the P<sub>sup</sub> score favors the concordance between parental allele among sibs of the same family. Therefore, heritability of P<sub>sup</sub> scores support an X-linked mechanism of skewing. The heritability of P<sub>sup</sub> scores was demonstrated in both epithelial cells and blood cells, whereas that of DS scores was demonstrated only for blood cells. The heritability was greater for T cells than for PMN. These results justified a genome wide approach to attempt to identify genetic loci associated with skewing.

### **GWAS**

GWAS analysis using SAIGE program was performed in related and unrelated subjects for the total available cohort (T) and after removing individuals with CHIP (C-), with the

hypothesis that they could represent phenocopies of the trait and decrease power. We also used both XCI scores (DS and P<sub>sup</sub>) although the parental direction of skewing cannot be evaluated in the unrelated part of the cohort. Phenotypes analyzed and number of genotyped individuals included in the analyses were: BC DS (T=2252, C-=1646); BC P<sub>sup</sub> (T=2250, C-=1644); PMN DS (T=2449, C-=1797); PMN P<sub>sup</sub> (T=2280, C-=1659); T cell DS (T=2165, C-=1578); T cell P<sub>sup</sub> (T=2162, C-=1577); monocyte DS (T=866, C-=622); monocyte P<sub>sup</sub> (T=861, C-=619); B cell DS (T=899, C-=650); B cell P<sub>sup</sub> (T=896, C-=648). No loci reached the predetermined threshold of 5E-08 for any of these phenotypes.

### **Linkage analysis**

No significant linkage was found when using the DS score in PMN or T cells by QTL or VC linkage analysis, in the total population or in the sensitivity analysis with CHIP negative individuals only. This may be attributable to the decreased power (increased noise) of the DS phenotype which does not account for the direction of skewing. However, when using the P<sub>sup</sub> score we found significant linkage signals. In T cells, QTL analysis identified a linkage region at rs2497911 with a LOD score = 4.8 on chromosome X at position 66,627,440 (hg19) (Figure 13A, page 69). The linkage region (defined as LOD > 3.9) spanned 2 Mb (Figure 13B, page 69). The LOD score signal was slightly reduced to 3.5 in sensitivity analyses with CHIP-negative individuals. As the HUMARA locus is in this region, we looked for association between the HUMARA alleles P<sub>sup</sub> scores and found no association (P=0.86), ruling out this gene as a candidate for skewing. For skewing assessed using the P<sub>sup</sub> score in PMN by VC linkage, we found a LOD = 6.01 at rs6427005 located on chromosome 1q21.3 at position 153,678,097 (Figure 13C, page 69). The linkage

region extended over a total of 41 Mb (Figure 13D, page 69). The LOD score was slightly reduced in the CHIP-negative only analysis. QTL analysis provided concordant a signal at the region (LOD 3.1) but did not reach our predetermined level of significance.

## **2.5 Discussion**

Among several distinguishing biological features, females in contrast to males are mosaics of two cell populations that differ only by the parental origin of the expressed X-linked genes. One of the most intriguing phenomenon is AAS of XCI skewing that occurs in the normal population. This phenomenon is identifiable in the neonate and its prevalence increases progressively thereafter (54). Although AAS allowed for the identification of mutation-driven clonal hematopoiesis in aging individual (4), little is known on the germline contribution to skewing. Identifying gene(s) involved in lifelong selection of blood cells may provide novel insight in hematopoiesis adaptation to aging. To that end, we extended the biological characterization of AAS of XCI conducting a comprehensive study that took into account both germline and acquired events leading to skewing in a cohort specifically designed and phenotyped to maximize our chance to decipher this complex trait.

XCI ratio can be determined by several methodologies (reviewed in (1)). Here we used the HUMARA assay which has the enormous advantage of being informative in close to 90% of women (46). We have previously demonstrated the validity of this assay by showing concordance with various transcriptional assays in subjects from this cohort and a previous cohort (151, 161). The validity of HUMARA has also been confirmed by other

investigators using different technologies (149, 162). Although doubts were raised in the past (163, 164), we are convinced that this approach, applied to a large population such as that studied herein, is reliable and robust.

In accordance with the literature, we found the prevalence of skewing to be higher in blood cells (22.5% - 41.7%) than in epithelial cells (14.3%).(165). Most previous studies analyzed only PMN and T cells, but we also evaluated B cells and monocytes. We showed that the prevalence of skewing is highest and concordant in PMN and monocytes (41.4%-41.7%), intermediate in B cells (30.1%), and lowest in T cells (22.5%). This may be related to the different turnover rates of these blood cells, closeness in hematopoietic development, or for T cell the fact that peripheral homeostatic expansion is not dependent on hematopoietic stem progenitor cells (166).

The concordance of X-chromosome parental origin of skewing between tissues of the same individual can inform us on the tissue specificity of skewing. Psup skewing scores obtained in different blood cell types of a given individual were strongly correlated (Figure 10, page 65). These findings are in line with previous reports (167, 168), and suggest that skewing occurs at the hematopoietic stem cells (HSC) level, prior to lineage specification. Interestingly, less concordance was seen between blood cells and epithelial cells, suggesting that in addition to shared common mechanism, further hematopoietic-specific mechanisms contribute to skewing.



The prevalence of skewing increased with age in all blood cell types. Although the association was not statistically significant for B cells, this was probably due to a lower number of women from which B cells were available for analysis. The age-effect was maximal in PMN, for which skewing prevalence increased relentlessly. In T cells, the prevalence of skewing rather stabilized in later life (Figure 11, page 66). One potential explanation for this would be the additive effect of acquired mutations (CHIP) in myeloid cells. To evaluate this possibility, we studied the prevalence of CHIP (measured in PMN) in the same age groups in function of VAF to compare with the prevalence of skewing (Figure 12, page 6). In this study, the qualitative definition of skewing was a DS score of  $\geq 0.25$ , which corresponds to the presence of 50% clonal cells, or a VAF of 25%. We demonstrated that the proportion of individuals with CHIP at significant VAF to influence XCI skewing ratios (VAF of 10% or 25%) is small until the age of 80. In the very old, CHIP accounts for a significant proportion of skewing. However, we could not take into account clonal hematopoiesis due to non-driver mutations (112) or acquired clonal mosaicism (169).

Skewing was not associated with significant differences in blood cell counts (Table 3, page 63), indicating that blood cell populations in which one of the two X chromosomes predominates over the other still obey to the homeostatic regulatory mechanisms of hematopoiesis. There was no association between skewing and common medical conditions (Table 4, page 63). Moreover, no link could be detected between skewing and smoking status or parity. This contrasts with a number of previous studies that reported a higher incidence of spontaneous abortion in women with skewed XCI ratios in their blood

cells (170-172). If skewed women were carriers of X-linked mutant alleles, and skewing was caused by negative selection against such alleles, one would expect a lower incidence of male birth in skewed women. In our study, the ratio of male:female birth was identical in skewed and non-skewed women, repudiating the mutant allele hypothesis.

Twin studies (72, 144, 148, 149) and Abkowitz's cat study (150) have provided evidences for a genetic component to acquired skewing. However, neither model allows for genetic locus identification. Furthermore, traditional multi-generation based approaches are inappropriate to study an age-dependent trait such as skewing. Herein, we evaluated heritability of the skewing trait among elderly sisters using a sib-pair approach. Fortunately, we had access to large families of French ancestry from Quebec with an average of 5.6 female siblings per family. This population exhibits a significant founder effect (158), reducing the number of potential contributing allelic variants. We measured heritability with DS score, which do not take into account the direction of skewing, and with P<sub>sup</sub> scores which partially (as discussed in Methods) take into account the parental origin of the X-linked dominant allele. The heritability of skewing was demonstrated by analyzing P<sub>sup</sub> scores in epithelial cells, PMN and T cells. Heritability of skewing was also demonstrated by analyzing DS scores in PMN and T cells, but appeared of less magnitude than that observed by analyzing P<sub>sup</sub> scores. This argues in favor of an X-linked genetic component to skewing, and supports the hemizygous cell selection hypothesis.

The demonstration of a heritable component to the skewing traits motivated us to conduct GWAS and linkage analyses looking for germline predisposing loci. We conducted

sensitivity analyses in the subgroup of the population without any CHIP in order to reduce potential phenocopies of germline variants. In the GWAS approach including related and unrelated subjects, no loci reached the pre-determined level of significance in any of the skewing scores or tissues. This may in part be explained by the multifactorial and polygenic nature of the skewing trait. It is also possible that our cohort was underpowered for association analysis. Alternatively, linkage analysis can be a powerful approach in founder populations, and we obtained interesting results. First, when using the DS score, which is insensitive to the parental origin of skewing and possibly influenced by multiple causes including genetic, environmental and stochastic, we did not find linkage signals. However, when using the Psup score that partially takes into account the parental origin of skewing, we found linkage signals.

First, we identified a region at Xq12-13 for the T cells (CD3+) skewing trait with a LOD score of 4.8 (Figure 13A, page 69). The region of interest has 2 Mb and includes several genes, including the androgen receptor (AR) (Figure 13B, page 69). To rule out that the AR gene was the source of the signal, we performed an association study using this multiallelic loci that was negative. Although speculative, this region also includes the interleukin-2 receptor gamma chain gene (*IL2RG*, OMIM:308380) which would constitute a candidate gene of prime importance. *IL2RG* encodes an indispensable component of the IL2 receptor (173), and mutations in this gene have been associated with severe combined immunodeficiency (174). We speculate that variants in this gene may favor T cells harboring the stronger allele on the active X-chromosome causing the phenotype of skewing in this tissue. Interestingly, the *IL2RG* product also heterodimerizes with other

cytokine receptors and could influence the development of other blood lineages supporting relative concordance of skewing in different blood cell types (174). Second, using the PMN P<sub>sup</sub> score, we identified a strong signal at chromosome 1q21.3 with a LOD score of 6 (Figure 13C, page 69). The region of interest is large, spanning 41 Mb and including several genes (Figure 13D, page 69). This result is surprising as we expected to map genes on the X-chromosome. Although speculative, the 1q21.3 region code for several genes including members of the S100 protein family which has been found to have a signaling feedback loop with the Interleukine 1 receptor-associated Kinase 1 (*IRAQ1*) (175) which resides at Xq28, supporting the hypothesis that autosomal genes could be in epistasis with X loci and support hemizygous cell selection.

In conclusion, we identified for the first-time chromosomal regions associated with the AAS of XCI trait, confirming the hemizygous cell selection process in blood cells of normal aging women. AAS is a complex trait caused in part by germline genetic variants and acquired mutations. Fine mapping and sequencing of candidate genes in the identified genetic regions associated with skewing are necessary to identify the causative variants.

## **Acknowledgements**

This research was funded in part by Canadian Institute of Health Research (CIHR). We thank study participants, research nurses and research assistants.

## **Author's contribution**

SA, SP: Collection and data assembly, data analysis and interpretation, manuscript writing.

MFG and LM interpretation and manuscript writing. RM, MB: collection of data. MPD,

SP and RF: statistical analyses. MPD: data analysis, interpretation, and manuscript writing.

LB: Conception and design, data analysis and interpretation, manuscript writing, financial support (grants), provision of study material.

## **Disclosure**

No conflicts of interest are declared.

## 2.6 Tables and Figures

**Table 1:** Association between degree of skewing (DS) scores and blood cell counts

Counts	Epithelial <i>P</i>	PMN <i>P</i>	T cell <i>P</i>
Total WBC	0.2839	0.4165	0.8016
PMN	0.7442	0.4567	0.8011
Lymphocyte	0.0200	0.9945	0.8392
Monocyte	0.1446	0.0472	0.6977
Platelet	0.0936	0.1239	0.4479
Hemoglobin	0.6185	0.6844	0.7585

Analysis was performed on family members and unrelated individuals, controlling for age and family. WBC: white blood cells; PMN: polymorphonuclear cells. Significance of *P* value was set at  $P < 0.003$  after Bonferonni correction.

**Table 2:** Association between skewing degree of skewing (DS) scores and common medical conditions

Medical conditions	Epithelial <i>P</i>	PMN <i>P</i>	T cell <i>P</i>
Cancer	0.7425	0.3475	0.3774
Arthritis	0.6686	0.4066	0.9296
Asthma	0.0959	0.1047	0.7089
Diabetes	0.9357	0.2745	0.5101
Eczema	0.6403	0.7743	0.1672
Thyroid problems	0.8088	0.9335	0.5098
Allergy	0.7403	0.1696	0.7987

Analysis was performed on family members and unrelated individuals, controlling for age and family. PMN: polymorphonuclear cells.

**Table 3:** Intra-individual correlation of Psup between different cell types

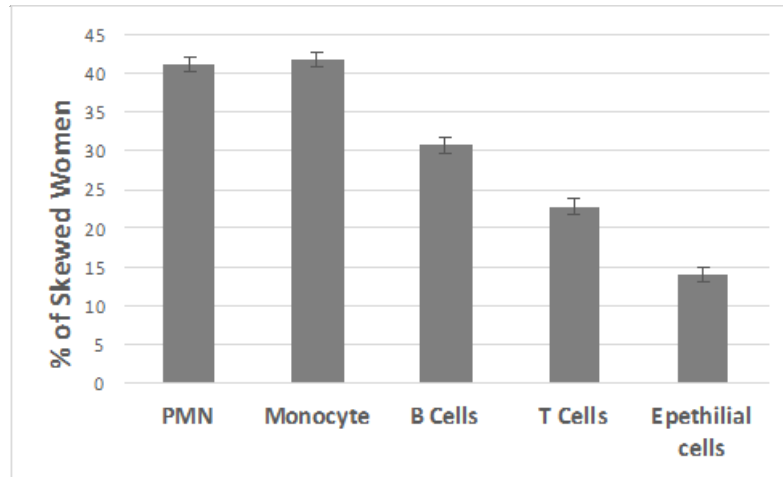
Cells	<i>r</i>	<i>P</i>
PMN vs Monocyte	0.9065	<10 <sup>-4</sup>
PMN vs B cell	0.8404	<10 <sup>-4</sup>
PMN vs T cell	0.7473	<10 <sup>-4</sup>
Monocyte vs B cell	0.8148	<10 <sup>-4</sup>
Monocyte vs T cell	0.6734	<10 <sup>-4</sup>
T cell vs B cell	0.7228	<10 <sup>-4</sup>
Epithelial cell vs PMN	0.5570	<10 <sup>-4</sup>
Epithelial cell vs Monocyte	0.4784	<10 <sup>-4</sup>
Epithelial cell vs B cell	0.5574	<10 <sup>-4</sup>
Epithelial cell vs T cell	0.6105	<10 <sup>-4</sup>

XCI ratios were evaluated using the Psup value which indicates the proportion of cells presenting the superior HUMARA allele on the active X chromosome. Spearman correlation coefficient (*r*) obtained in PMN (n = 2364 samples), monocytes (n = 933 samples), B cells (n = 893 samples), T cells (n = 2246 samples) and epithelial cells (n = 2340 samples). PMN: polymorphonuclear cells.

**Table 4:** Heritability of skewing

	Solar		
	H <sup>2</sup>	SD	<i>P</i>
<u>DS scores</u>			
Epithelial cell	0.045	0.039	0.114
PMN	0.097	0.042	0.005
T cell	0.255	0.050	<10 <sup>-9</sup>
<u>Psup scores</u>			
Epithelial cell	0.310	0.053	<10 <sup>-13</sup>
PMN	0.205	0.048	<10 <sup>-6</sup>
T cell	0.303	0.053	<10 <sup>-12</sup>

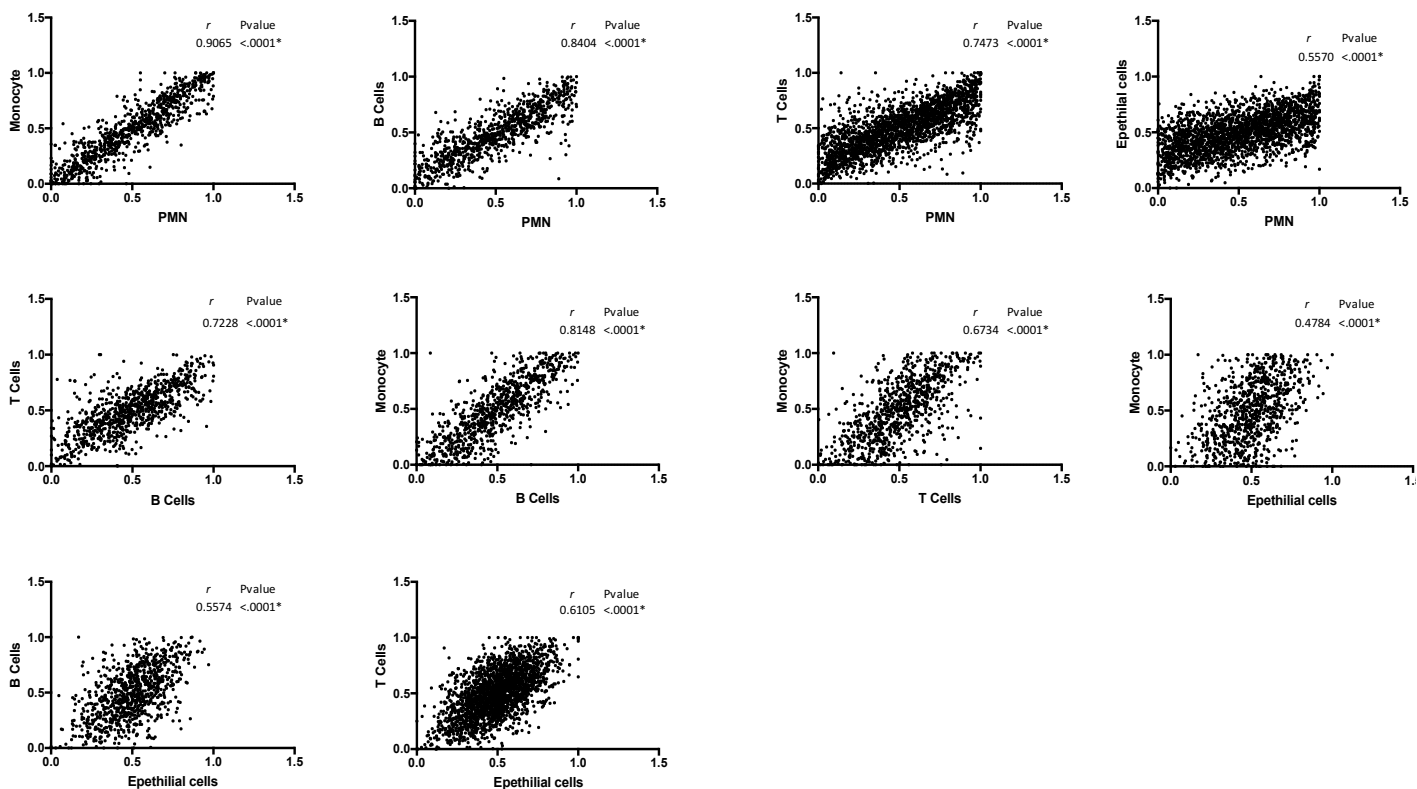
Heritability (H<sup>2</sup>) of skewing scores was calculated using Solar (see Methods). The family cohort used for this analysis comprised 1734 women from 311 families. PMN: polymorphonuclear cells; SD: standard deviation



**Figure 9:** Prevalence of XCI skewing varies according to tissue and blood cells types.

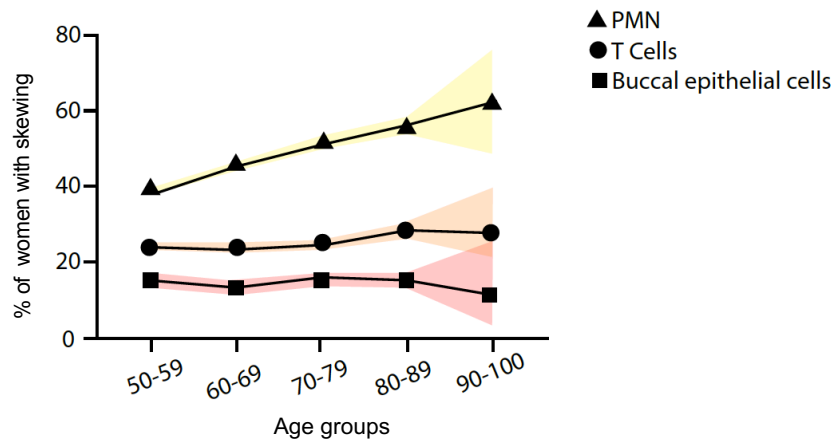
Skewing was defined as a DS score of  $\geq 0.25$ , which corresponds to a ratio  $\geq 3:1$  between the two parental Xs. The prevalence of skewing was highest and similar in PMN and monocytes (41.4% and 41.7% respectively), intermediate in B lymphocytes (30.1%) and lower in T lymphocytes (22.5%). The prevalence of skewing was lowest in buccal epithelial cells (14.3%). In order to control for age, only women for whom samples from all cell types were available were included in this analysis ( $n=777$ , mean age  $63.3 \pm 9.7$  years).





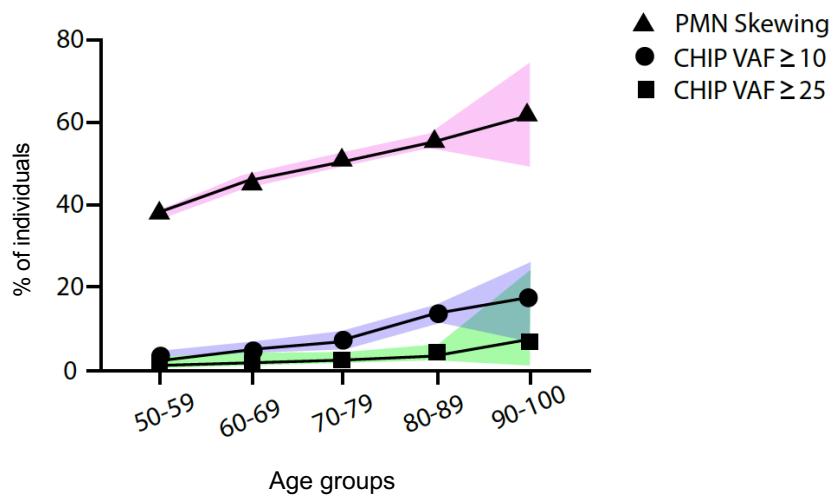
**Figure 10:** Intra-individual correlation of XCI ratios between different tissues and blood cells types

XCI ratios were calculated using the Psup score which indicates the proportion of cells with the superior HUMARA allele in the active state. Psup scores vary from 0% (in cases where there is complete skewing toward the inferior allele) and 100% (in cases where there is complete skewing toward the superior allele). A ratio of 50% represents a balanced XCI ratio between both alleles. Psup scores allow to evaluate the concordance of the parental origin of skewing. Each dot represents an individual. Intra-individual correlations ( $r$ ) between blood cell types and epithelial cells were statistically significant for every combination. However, the strength of the correlation varied considerably between combinations. Correlation was strongest between PMN and monocytes but weaker when any blood cell type was compared to epithelial cells (see Table 3, page 63).



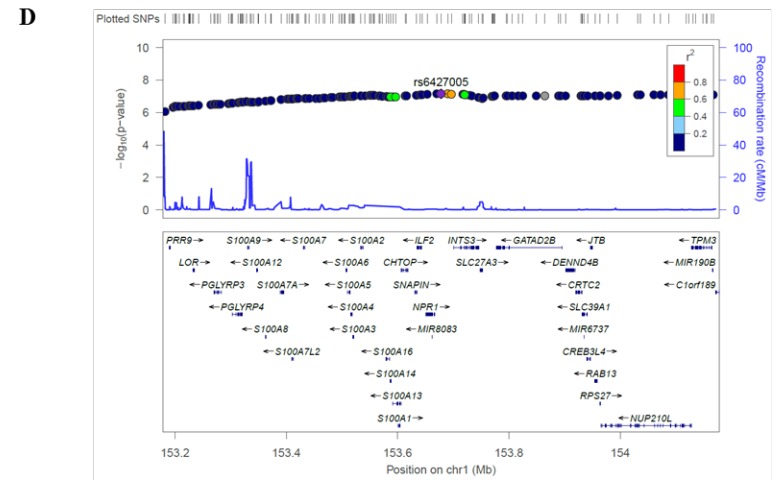
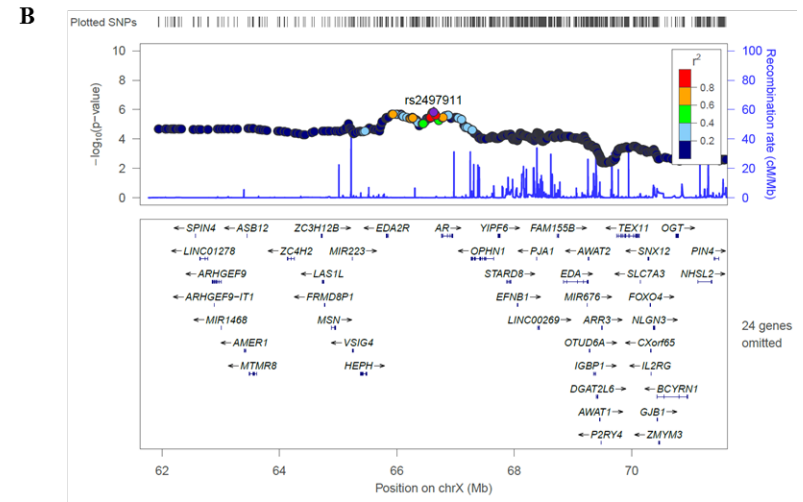
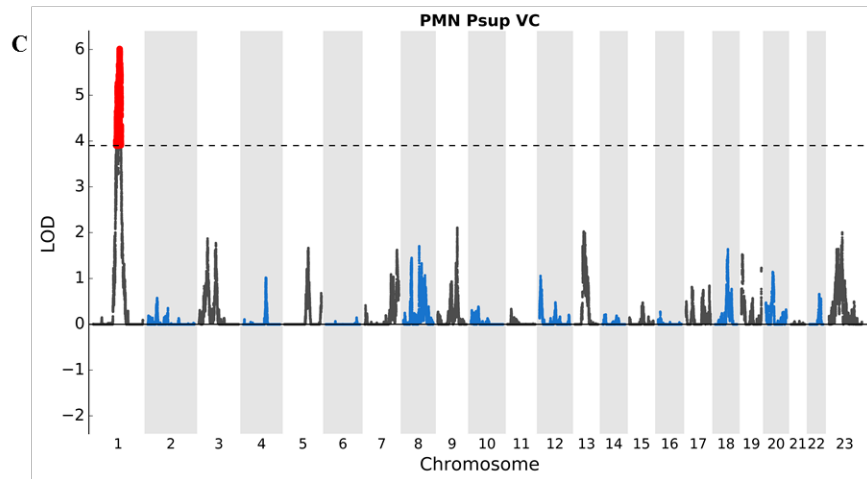
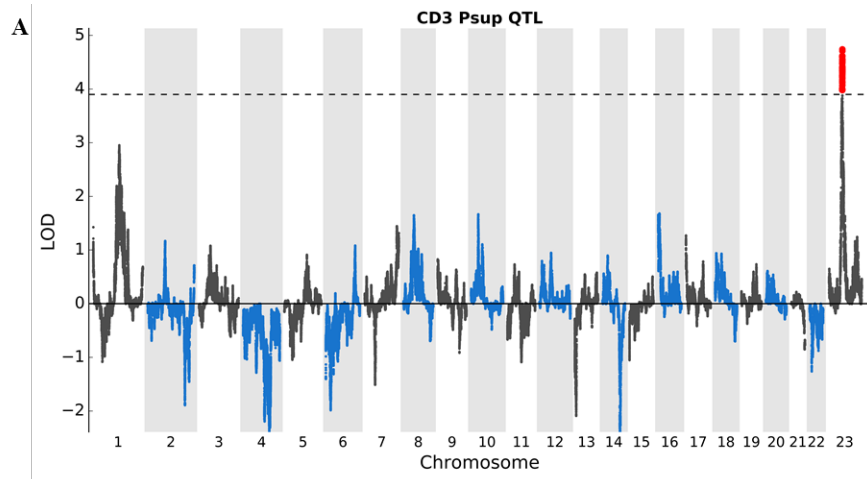
**Figure 11:** Prevalence of XCI skewing is age-dependent in blood cells and maximal in PMN.

The prevalence of XCI skewing correlates strongly with age in PMN (BETA =0.0142,  $P < 0.00000$ ), less in T-cells (BETA=0.0071,  $P = 0.00037$ ), and not at all in epithelial cells (BETA=-0.0009,  $P = 0.62623$ ). This figure shows the prevalence of skewing ( $DS \geq 0.25$ ) in different age groups. The shaded areas represent the confidence intervals (95%).



**Figure 12:** Prevalence of XCI skewing in PMN and CHIP at different VAF in function of age.

In order to evaluate if CHIP is a significant contributor to PMN XCI skewing, we calculated the prevalence of individuals with CHIP at a sufficient VAF to influence XCI ratio. A VAF of 10% corresponds to 20% clonal cells and an increase in DS of 0.1. A VAF of 25% corresponds to 50% clonal cells and an increase in DS of 0.25. The prevalence curves presented above indicate that CHIP is not a predominant cause of PMN skewing in the cohort. However, as individuals age, the CHIP contribution increases and may be responsible, in part, for the age-effect on PMN skewing which is more important than in other blood cell types. PMN (n = 2372), CHIP VAF  $\geq$  10 (n = 167), CHIP VAF  $\geq$  25 (n = 58) distributed by age group. The shaded areas represent the confidence intervals (95%).



**Figure 13:** Regions associated with XCI skewing.

(A) Manhattan plot of twopoint QTL linkage analysis for CD3 P<sub>sup</sub> controlling for age. LOD score threshold was set to 3.9 for 476,199 variants. (B) Regional plot (LocusZoom)a(176) of single-nucleotide polymorphisms (SNPs) located on chromosome X for T cells (CD3) P<sub>sup</sub> using twopoint QTL linkage analysis. The first y-axis shows the negative log<sub>10</sub> of P values for genotyped genetic variants, the second y-axis shows the recombination rate from HapMap reference samples (blue line). (C) Manhattan plot of twopoint VC linkage analysis for PMN P<sub>sup</sub> controlling for age. LOD score threshold was set to 3.9 for 476,199 variants. (D) Regional plot (LocusZoom) of single-nucleotide polymorphisms (SNPs) located on chromosome 1 for PMN P<sub>sup</sub> using two point VC linkage analysis, controlling for age. The first y-axis shows the negative log<sub>10</sub> of P values for genotyped genetic variants, the second y-axis shows the recombination rate from HapMap reference samples (blue line). Base pair positions are given according to hg19. The degree of linkage disequilibrium ( $r^2$ ) of each genetic variant with the best SNP estimated from the study population is displayed.

## **CHAPITRE 3**

### **DÉTERMINANTS DE LA PÉNÉTRANCE ONCOGÉNIQUE DE LA MUTATION DE *TET2* CHEZ LES INDIVIDUS ÂGÉS**

### **3. DÉTERMINANTS DE LA PÉNÉTRANCE ONCOGÉNIQUE DE LA MUTATION DE *TET2* CHEZ LES INDIVIDUS ÂGÉS**

#### **3.1 Résumé**

La relation entre l'hématopoïèse vieillissante et le développement des cancers myéloïdes est inconnue. Notre groupe a documenté que des mutations acquises se produisent dans les cellules sanguines d'individus vieillissants (4, 6, 177). Il a démontré que > 10% des personnes de plus de 65 ans ont une mutation de perte de fonction dans un gène appelé *TET2*. *TET2* appartient à la famille des enzymes de TET (*TET1*, 2 et 3) (178) (179) (180). *TET2* subit très souvent une mutation somatique dans les cellules cancéreuses hématopoïétiques, spécifiquement dans les tumeurs malignes myéloïdes, et son inactivation est en corrélation avec un phénotype d'hyperméthylation (4, 122). L'objectif global de notre recherche est de comprendre comment les mutations *TET2* mènent à la transformation. Pour identifier les marques épigénétiques spécifiques au *TET2* nous avons utilisé des méthodes comme le méthylome, l'hydroxyméthylome, le NGS et l'expression génique (RNA-Seq) qui nous a permis de déterminer les régions géniques affectées par la mutation *TET2* ou les régions qui pourraient contribuer, avec la mutation *TET2*, au déclenchement d'un cancer.

### 3.2 Introduction

Le vieillissement est l'un des facteurs de risque les plus importants pour le développement de tumeurs malignes myéloïdes telles que le SMD, la LMA et les NMP (3). Cependant, la relation entre l'hématopoïèse vieillissante et le développement de ces cancers est inconnue. L'identification des facteurs génétiques ou épigénétiques de la transformation hématologique associée à l'âge est cruciale pour prévenir et traiter plus efficacement ces cancers. Récemment, notre groupe et d'autres ont documenté que des mutations acquises se produisent dans les cellules sanguines des individus vieillissants (4, 6, 177). Notre groupe a montré que plus de 10% des personnes de plus de 65 ans ont une mutation de perte de fonction dans un gène appelé *TET2* (4). *TET2* appartient à la famille des enzymes TET (*TET1*, 2 et 3) qui fait partie de la superfamille des di-oxygénases dépendantes du Fe<sup>2+</sup> et du 2-oxoglutarate (2-OG) (179). Les enzymes TET2 modifient la méthylation en catalysant l'oxydation du groupe méthyle de 5mC pour former des 5-hydroxyméthylcytosines (5hmC) qui, elles même, peuvent continuer à être oxydées par les protéines TET en 5-formylcytosine (5-fC) et 5-carboxylcytosine (5-caC) (181). Ce processus permet, entre autres, la déméthylation des cytosines (181). Les bases modifiées par les protéines TET sont fortement enrichies en amplificateurs, promoteurs et corps géniques et ont un effet important sur l'expression des gènes (181). Par conséquent, *TET2* est considéré comme un gène suppresseur de tumeur et un gardien de la méthylation inappropriée de l'ADN se produisant dans notre génome (182). Les mutations que nous avons documentées affectent la fonction *TET2* comme le prouve le niveau inférieur de 5hmC chez ces individus (183). De manière très importante, tous les sujets avaient une



hématopoïèse clonale prouvant l'avantage de croissance conféré par la mutation à la cellule souche hématopoïétique tel que documentée par l'inactivation du chromosome X chez les sujets féminins. Cependant, tous les sujets mutés avaient une formule sanguine normale et aucun phénotype malin. (95, 183). Puisque *TET2* est le gène le plus fréquemment muté dans le cancer myéloïde, nous postulons que les mutations survenant chez les individus vieillissants normaux sont l'une des étapes initiales de la transformation maligne.

L'objectif global de notre recherche est de comprendre comment les mutations *TET2* conduisent à la transformation. Nous allons caractériser les différences de méthylation de l'ADN, définie comme régions méthylées différentiellement (DMR), et l'hydroxyméthylation, définie comme régions hydroxyméthylées différentiellement (DHMR), à l'échelle du génome dans les *TET2* VS WT, au niveau des corps géniques, des *transcription start sites* (tss, des exons, des introns et des îlots CpG et leur effet sur les marques épigénétiques et les voies de signalisation. La méthylation des dinucléotides CpG est un mécanisme épigénétique important pour contrôler la régulation des gènes(184), et les altérations de la méthylation de l'ADN jouent un rôle crucial dans de nombreuses maladies, en particulier le cancer (185). Les objectifs spécifiques de notre projet sont de valider la méthode de méthylome, d'hydroxyméthylome et du transcriptome ainsi que d'identifier quels gènes ou qu'ensemble de gènes sont affectés par le dysfonctionnement de *TET2* et lesquels sont importants dans la transformation maligne. Par conséquent, notre premier objectif est d'identifier l'impact de la mutation *TET2* sur l'épigénome (méthylome, hydroxyméthylome) et notre deuxième objectif est d'identifier lesquelles de ces marques épigénétiques ont un potentiel oncogène.

### **3.3 Matériel et méthodes :**

#### **3.3.1 Les sujets humains**

Note laboratoire a recruté prospectivement de 1998 à 2006 une cohorte de plus de 4000 femmes (également  $\approx$  500 hommes) sans aucun trouble hématologique connu. Notre équipe a développé, optimisé et validé un protocole de séquençage de nouvelle génération (NGS) robuste pour détecter avec sensibilité et précision l'incidence des mutations somatiques TET2 chez les sujets de notre cohorte vieillissante. TET2 a été séquençé à une couverture élevée ( $95\% > 500x$ ) chez 1771 individus âgés de 50 à 101 ans. 116 individus mutés sur le gène *TET2* ont été identifiés, dont 10 individus porteurs de 2 mutations. Les 126 mutations *TET2* (50 faux-sens, 32 non-sens, 40 décalages de cadre et 4 sites d'épissage) identifiées sont réparties dans tout le gène avec des mutations faux-sens regroupées dans des régions connues pour être importantes pour les activités catalytiques. L'impact fonctionnel des mutations identifiées a été évalué à l'aide d'outils bioinformatiques pour prédire les effets fonctionnels et en quantifiant directement le phénotype épigénétique associé à la perte de fonction TET2 (niveaux réduits de 5hmC) par spectrométrie de masse. Notre équipe a démontré que ces mutations affectent directement le niveau de 5hmC proportionnellement à la VAF suggérant qu'elles ont un impact négatif sur le maintien épigénétique de l'hématopoïèse vieillissante (95, 183).

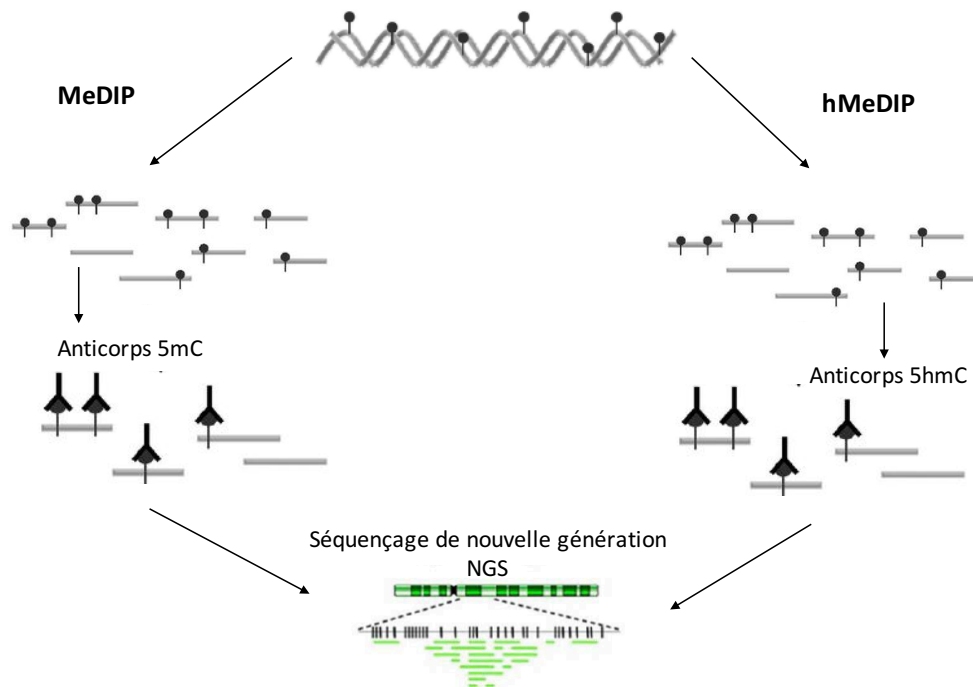
En utilisant cette cohorte, nous avons conçu 3 groupes de sujets: (i) un groupe de 7 individus vieillissants normaux sans aucune mutation dans TET2 (*TET2wt*) qui agira

comme groupe témoin, (ii) un groupe de 10 sujets vieillissants avec des mutations TET2 (TET2mut), et (iii) un groupe de patients atteints d'un cancer myéloïde et présentant une mutation dans TET2 (TET2cancer).

### 3.3.2 Méthylome, hydroxyméthylome sur NGS Ion Torrent

Nous avons choisi la méthode d'immunoprécipitation de l'ADN (DIP) qui consiste à utiliser un anticorps ciblant 5hmc ou 5mc pour immunoprécipiter un fragment d'ADN méthylé ou hydroxyméthylé pour ensuite les séquencer (Figure 14, page 75).

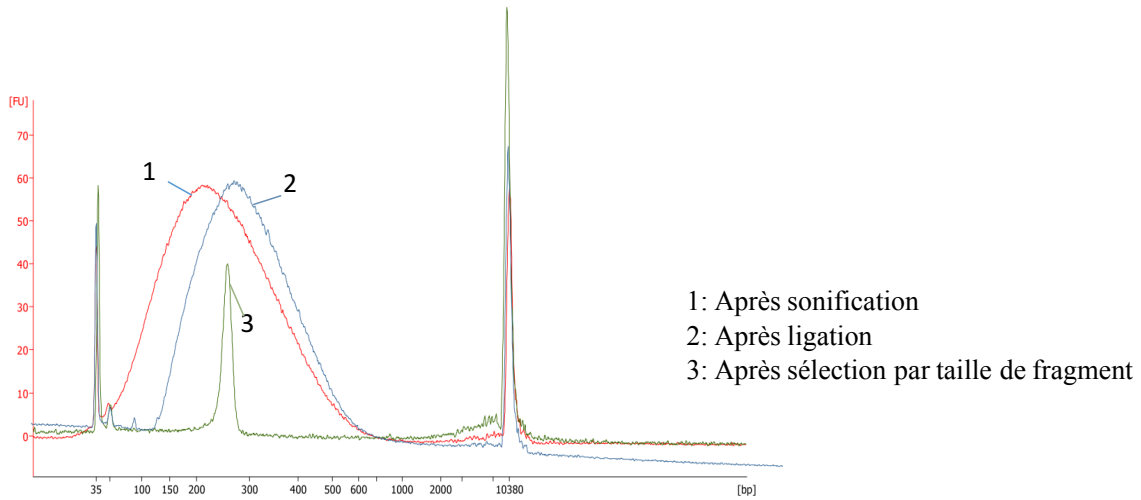
#### Immunoprécipitation de l'ADN méthylé /hydroxyméthylé (MeDIP, hMeDIP)



**Figure 14:** Représentation schématique des méthodes du méthylome et hydroxyméthylome

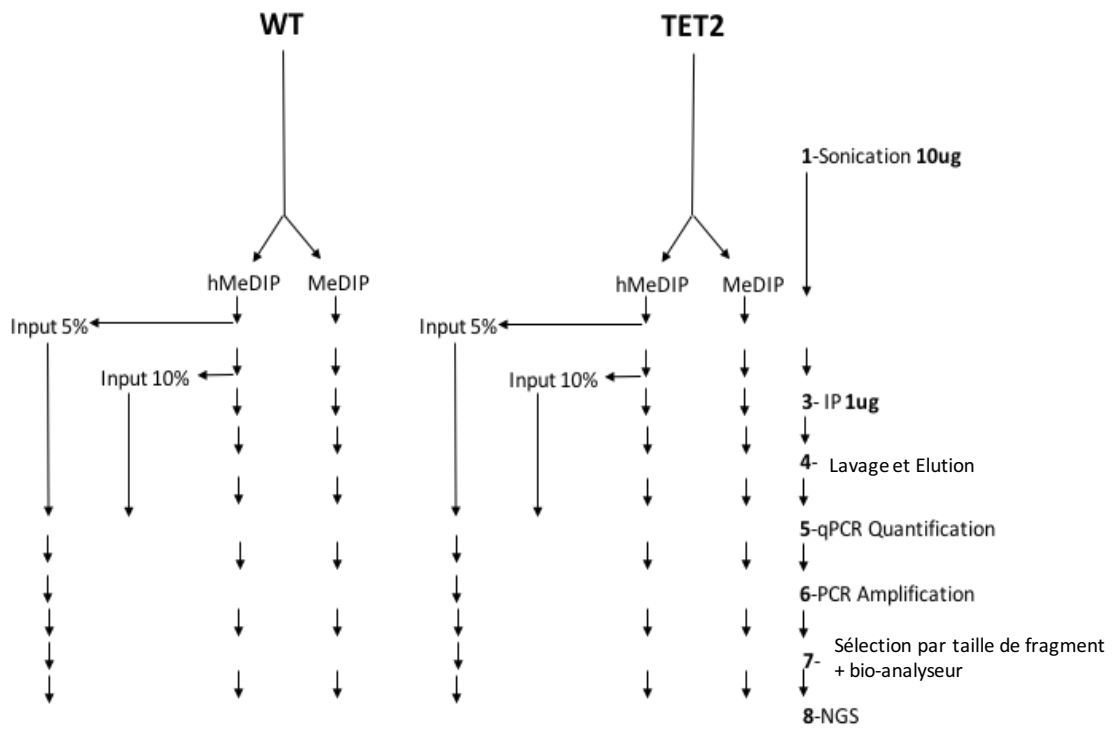
La première étape est la sonication qui consiste à fragmenter l'ADN en fragment de 50 à 300 pb de taille (Figure 14, page 75). Par la suite, une librairie est construite. Tout d'abord, les fragments sont réparés et ligés par des adaptateurs spécifiques en utilisant le kit de *NEBNext Fast DNA Library Prep Set for Ion Torrent* qui est compatible avec l'Ion Torrent. Ces fragments ligés sont ensuite immunoprécipités par des anticorps spécifiques au 5mC et 5hmC qui proviennent de la compagnie Cell Signaling *SimpleDIP Methylated DNA IP* (#76853). Après une série de lavages et d'élutions, les fragments sont quantifiés par la méthode de PCR quantitatif (qPCR) en utilisant des amorces spécifiques aux adaptateurs et gènes contrôles. Cette étape est cruciale car la quantification des fragments immunoprécipités sera comparée à l'INPUT et aux contrôles négatifs IgG. Les valeurs obtenues permettent de s'assurer de la qualité initiale des librairies et de calculer l'enrichissement de l'immunoprécipitation. Après avoir quantifié les fragments, ces derniers sont amplifiés en utilisant le kit *NEBNext Fast DNA Library Prep Set for Ion Torrent*, purifiés et quantifiés par la même méthode de qPCR pour quantifier le produit final. Les fragments amplifiés et purifiés sont sélectionnés par rapport à leur taille à l'aide d'un *E\_Gel® SizeSelect™* (Figure 15, page 77). Les fragments sélectionnés sont prêts à être chargés sur une puce par le chef pour le séquençage sur l'Ion Proton. La plateforme Ion Torrent (Proton) a été utilisée pour le séquençage (fragments de 100-200 pb à une seule extrémité, +/- 30M lectures par échantillon). Pour chaque échantillon (individu), nous avons X\_5mC, X\_5hmC et X\_IN (IN pour INPUT). Nos données sont présentées dans un fichier BAM où les gènes sont annotés et alignés avec leur position.

### Profil des fragments immunoprécipités après amplification



**Figure 15:** Résultats de bio-analyseur

Le pic 1 représente la taille des fragments après la sonication. Le pic 2 représente la taille des fragments après la ligation par les adaptateurs. Le pic 3 représente la taille des fragments après le E Gel® SizeSelect™.



**Figure 16:** Représentation schématique des étapes de la méthode du méthylome

### 3.3.3 Analyses bioinformatiques

Les données ont été analysées à l'aide d'une combinaison d'outils et de langages de programmation, principalement Python, *bash scripting* et R. Les échantillons ont tout d'abord subi une série de pré-traitements et de tests de contrôle de qualité (CQ). Les données ont été annotées avec des corps géniques, des tss, des exons, des introns et des îlots CpG à l'aide de ressources de la base de données UCSC hg19 *golden path*. Le CQ a été réalisé à l'aide du package *qsea* R (186) sur une base par échantillon (vs INPUT). Une analyse de corrélation a été effectuée à l'aide du MEDIPS *package* pour mieux visualiser les données dans leur ensemble (187). Les PEAK CALLING ont été identifiés sur tous les échantillons en utilisant *macs2* pour identifier les régions enrichies (188). Les pics obtenus ont été utilisés pour créer des images instantanées à l'aide du navigateur de génome IGV pour estimer la qualité des données.

Le nombre total des pics déterminés pour chaque échantillon était également évalué. Le modèle QSEA peut prendre en compte la variation du nombre de copies (CNV) et tenir compte de leur influence sur la densité de lecture MeDIP. La CNV peut être calculée en externe et importée dans QSEA ou estimée directement dans QSEA. QSEA utilise en interne le *package* bioconducteur HMMcopy pour estimer la CNV à partir des données de séquençage de la bibliothèque INPUT (186).

Pour ce qui est de la couverture estimée de l'OFFSET, la densité CpG moyenne est déterminée par fragment et les régions sans CpG sont utilisées pour estimer le bruit de fond. Il s'agit d'une estimation du bruit expérimental et des valeurs supérieures à 0,9

indiquent un manque d'efficacité de l'enrichissement (186). L'enrichissement relatif estimé a été étalonné en utilisant la méthode d'étalonnage à l'aveugle dans QSEA car aucune donnée expérimentale de validation n'était disponible. Par exemple, dans les échantillons de tissus adultes, nous pouvons supposer une méthylation élevée dans des régions à faible densité de CpG et des niveaux de méthylation moyens linéairement décroissants pour une densité de CpG plus élevée.

### **3.3.4 RNA-Seq**

Pour le séquençage de nouvelle génération, tous les réactifs et appareils provenaient de Thermo Fisher Scientific. La méthode de RNA-Seq utilisée commence par l'isolation de l'ARN messager (ARNm) des échantillons d'ARN totale de 200 à 1000 ng à l'aide du kit Dynabeads mRNA DIRECT™ Micro Kit par la suite le kit Ion Total RNA-Seq Kit v2 est utilisé pour produire des bibliothèques compatibles avec la technologie d'Ion Torrent. Suite à l'enrichissement de l'ARNm à l'aide de colonnes d'oligo-dT, celui-ci est fragmenté à l'aide d'une ARNase et purifié. Les fragments d'ARNm sont hybridés avec des adaptateurs universel pour ensuite être convertis en ADNc par une transcriptase inverse. Finalement des adaptateurs et code-barres sont ajoutés avant d'être amplifié, purifié et quantifié. Les bibliothèques sont ensuite chargées sur l'instrument Ion Chef (kit Ion PI Hi-Q Chef) pour la préparation du modèle et chargée sur les puces Ion P1 v3.



### 3.3.4.1 Analyses bioinformatiques

Les données ont été analysées à l'aide d'une combinaison d'outils bioinformatiques, principalement R, DESeq2 et EdgeR afin d'obtenir les gènes qui sont statistiquement différentiellement exprimés entre *TET2* et *WT*.

## 3.4 Résultats

Dans un projet précédent dans notre laboratoire, un total de 2542 individus ont été séquencés dans le but d'identifier des mutations somatiques tel que *TET2* et *DNMT3* (126). Nous avons identifié 402 mutations somatiques chez 347 individus. Pour *TET2*, 134 mutations somatiques ont été identifiées. Dans le cadre de mon projet de thèse, on a sélectionné les 10 individus *TET2* mutés avec le VAF le plus élevé dans le but d'avoir un effet plus prononcé de la mutation *TET2* sur la méthylation et l'hydroxyméthylation. Les échantillons sélectionnés sont présentés dans le Tableau 5 (page 81) ci-dessous.

**Tableau 5:** Liste des échantillons séquencés

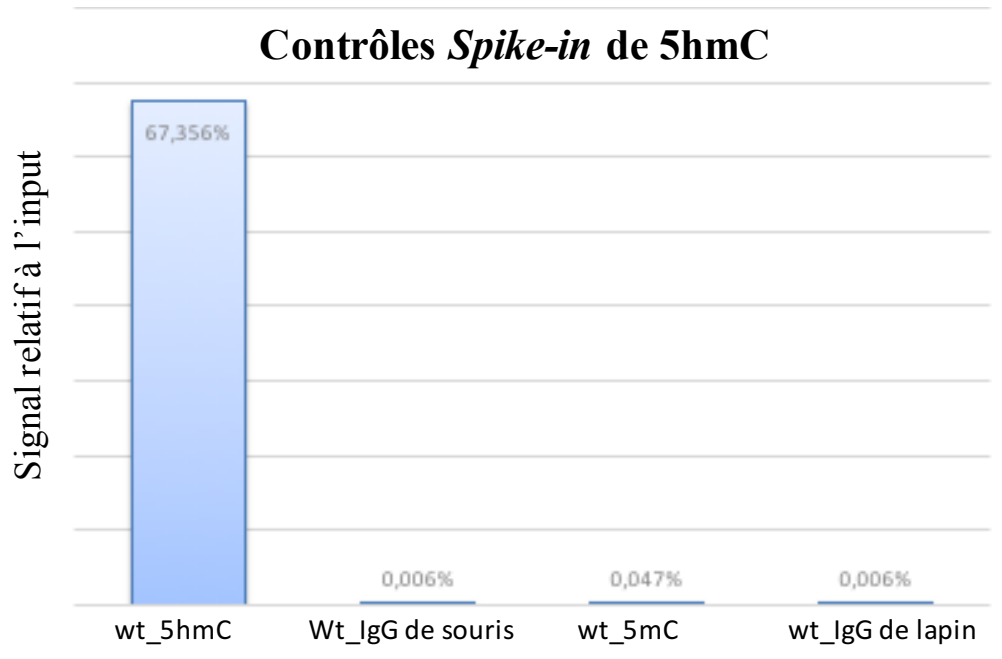
<i>TET2</i> (n=10)		<i>WT</i> (n=7)	
5hmC	5mC	5hmC	5mC
128.F	128.F	21.A	21.A
224.B	224.B	74.D	74.D
229.A	229.A	380.C	380.C
385.B	385.B	515.A	515.A
447.E	447.E	009.D	009.D
Z.189	Z.189	298.XA	298.XA
Z.286	Z.286	Z.0234	Z.0234
Z.304	Z.304		
Z.863	Z.863		
246.E	246.E		

### 3.4.1 Contrôle de qualité des échantillons séquencés

#### 3.4.1.1 Contrôle *spike-in* de l'enrichissement 5hmC

Avant de procéder à l'étape de l'immunoprécipitation, nous avons d'abord testé la spécificité des anticorps et le niveau de l'enrichissement des fragments méthylés et hydroxyméthylés. Pour ce faire, nous avons commencé par tester la spécificité de l'anticorps anti-5hmC et son contrôle négatif IgG de souris en utilisant un Spike-in, une forme de fragment hydroxyméthylé, ajouté dans nos échantillons. Les fragments du Spike-in sont ensuite immunoprécipités grâce à l'anticorps anti-5hmC pour être quantifiés par qPCR. L'anticorps anti-5mC et l'IgG de lapin, son contrôle négatif, ont aussi été utilisés avec le Spike-in pour démontrer sa spécificité. La

Figure 17 (page 83) démontre que l'anticorps anti-5hmC a immunoprécipité 67% du Spike-in par rapport à l'IgG de souris. Nous avons aussi observé que le Spike-in n'a pas été enrichi par l'anticorps anti-5mC. Ces résultats indiquent que l'anticorps anti-5hmC est spécifique aux fragments 5hmC.

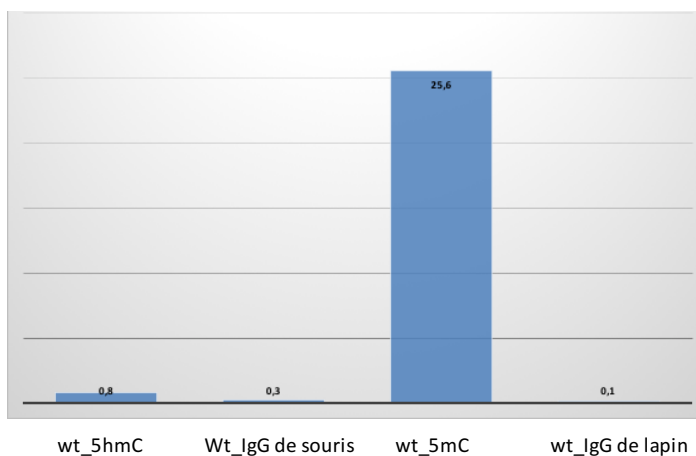


**Figure 17:** Spécificité de l'anticorps anti-5hmC par rapport au *Spike-in*

### 3.4.1.2 5mC : Contrôle par un gène endogène méthylé

Nous avons aussi testé la spécificité de l'anticorps anti-5mC et son IgG de lapin. Nous avons utilisé une amorce d'un gène appelé *TsH2B*, qui est connu pour être méthylé dans les cellules adultes. Ces amorces ont été fournies dans le kit de signalisation cellulaire. En utilisant 1µg d'ADN total (100%), et après incubation avec 4 anticorps contre 5hmC, IgG de souris, 5mC et IgG de lapin, nous avons enrichi, comme démontré dans la Figure 18 (page 84), 25,6% de promoteur méthylé *TsH2B*. 0,8% de 5hmC sur ce promoteur a été détecté par un anticorps 5hmC. Cela suggère que ce promoteur pourrait également être hydroxyméthylé.

## 5mC: qPCR sur la méthylation du gène promoteur endogène TsH2B



**Figure 18:** Spécificité de l'anticorps anti-5mC par rapport au TsH2B

### 3.4.1.3 Quantification de la bibliothèque 5mC et 5hmC

Après avoir testé la spécificité des deux anticorps, nous avons procédé à l'étape de l'immunoprécipitation. Les fragments immunoprécipités sont liés aux extrémités 3' et 5' par des codes-barres pour pouvoir les quantifier par qPCR et les séquencer par la suite. Par exemple, tel que représenté dans la Figure 19 (page 85), en utilisant 1 $\mu$ g d'ADN total (100%), nous avons enrichi par cette méthode 3,81% de 5mC et 0,22% de 5hmC.

Ensuite, les fragments enrichis par l'anticorps 5mC et 5hmC sont amplifiés pour obtenir la quantité suffisante nécessaire pour le séquençage. Après amplification, nous avons sélectionné sur gel les fragments de 275 à 300 pb pour obtenir un séquençage optimal (Figure 15, page 75). Une fois que les fragments sélectionnés sont chargés sur une puce par le chef, le séquençage sur le proton peut avoir lieu. Après avoir séquencé nos fragments, nous avons obtenu des fichiers appelés BAM pour chaque échantillons WT-5mC, WT-

5hmC, TET2-5mC et TET2-5hmC. Ces fichiers BAM passent ensuite par une procédure de CQ afin de sélectionner ceux qui répondent aux critères de séquençage avant de procéder aux analyses bioinformatiques.

### 5hmC and 5mC: Quantification de la Librairie Quantification des fragments ligués

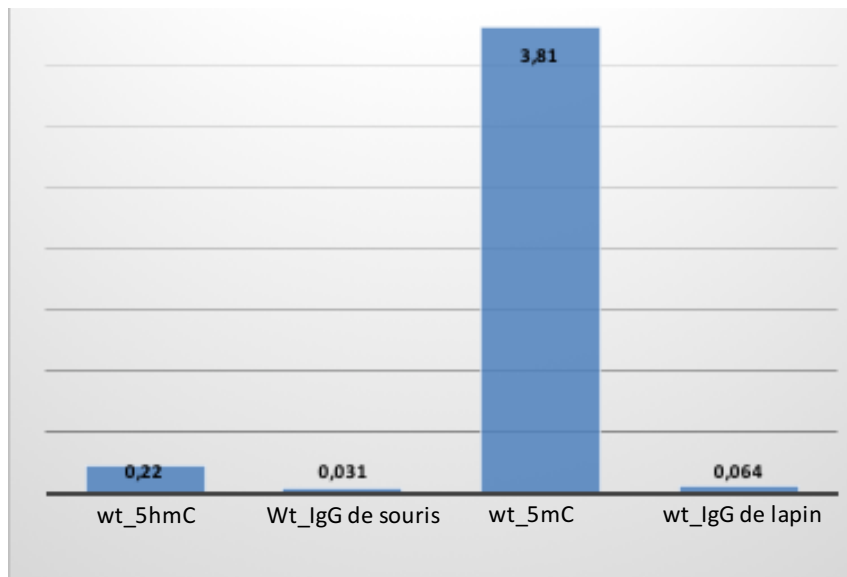
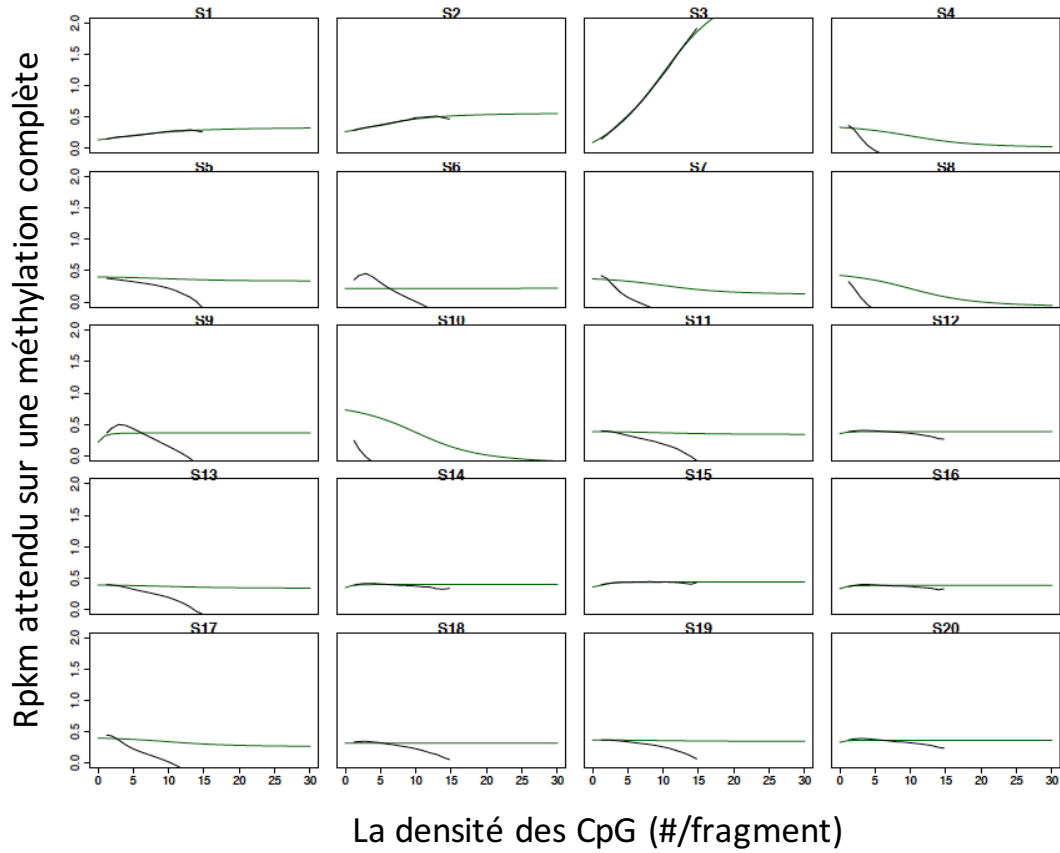


Figure 19: Quantification par qPCR de la bibliothèque de 5mC et 5hmC

#### 3.4.2 Contrôle de qualité des fichiers BAM

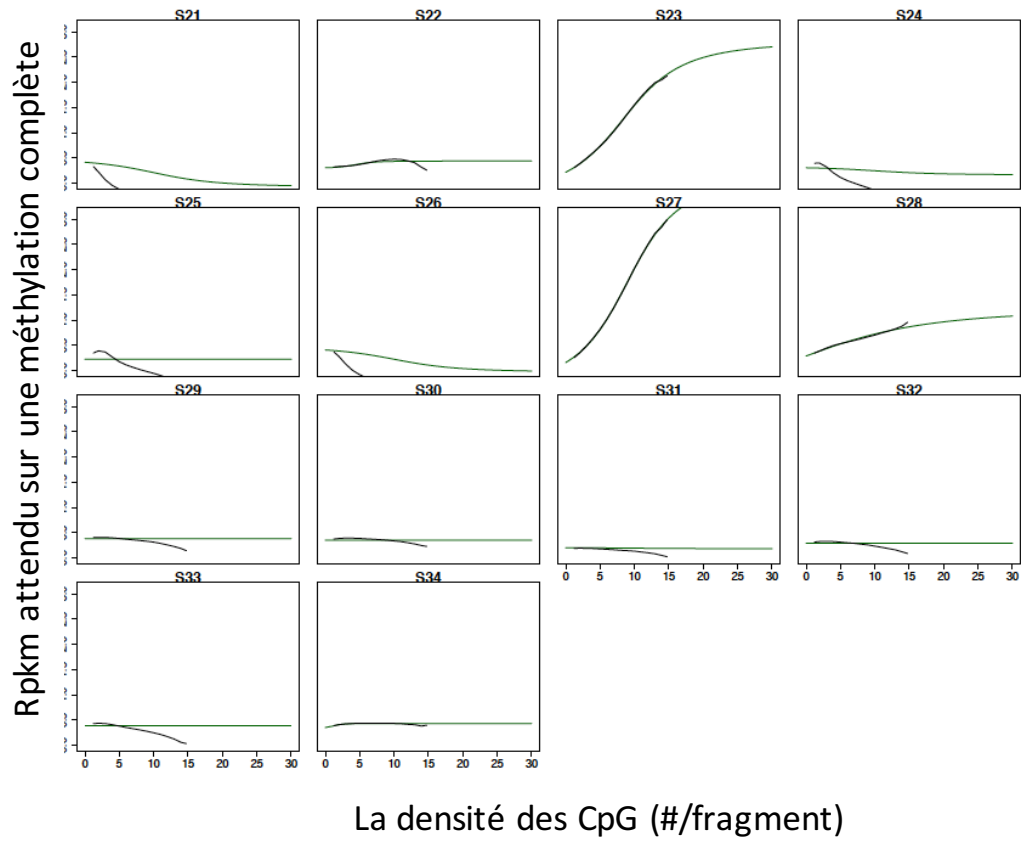
Afin d'effectuer le CQ du séquençage, des méthodes dans le langage R sont utilisées pour trier les échantillons qui ont passé le CQ de ceux qui ont échoués. Nous avons calculé l'enrichissement de CpG de chaque individu, l'enrichissement de CpG par rapport à son INPUT (Figure 20, page 86), les PEAK CALLING dans 5mC et 5hmC (Tableau 8, page 91, et Tableau 9, page 93) et estimé la couverture de l'OFFSET (Tableau 6, page 90, et Tableau 7, page 90).

### Analyse de l'enrichissement des CpG



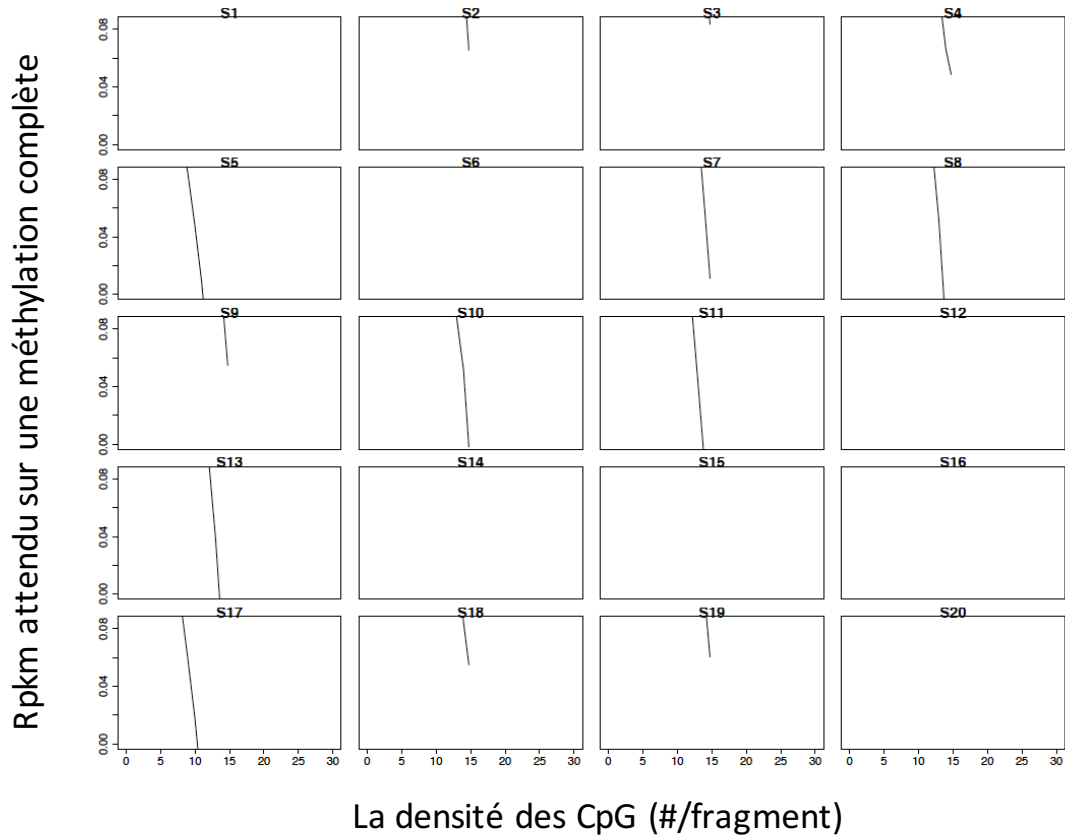
**Figure 20:** Enrichissement CpG incluant l'INPUT TET2-5mC

## Analyse de l'enrichissement des CpG



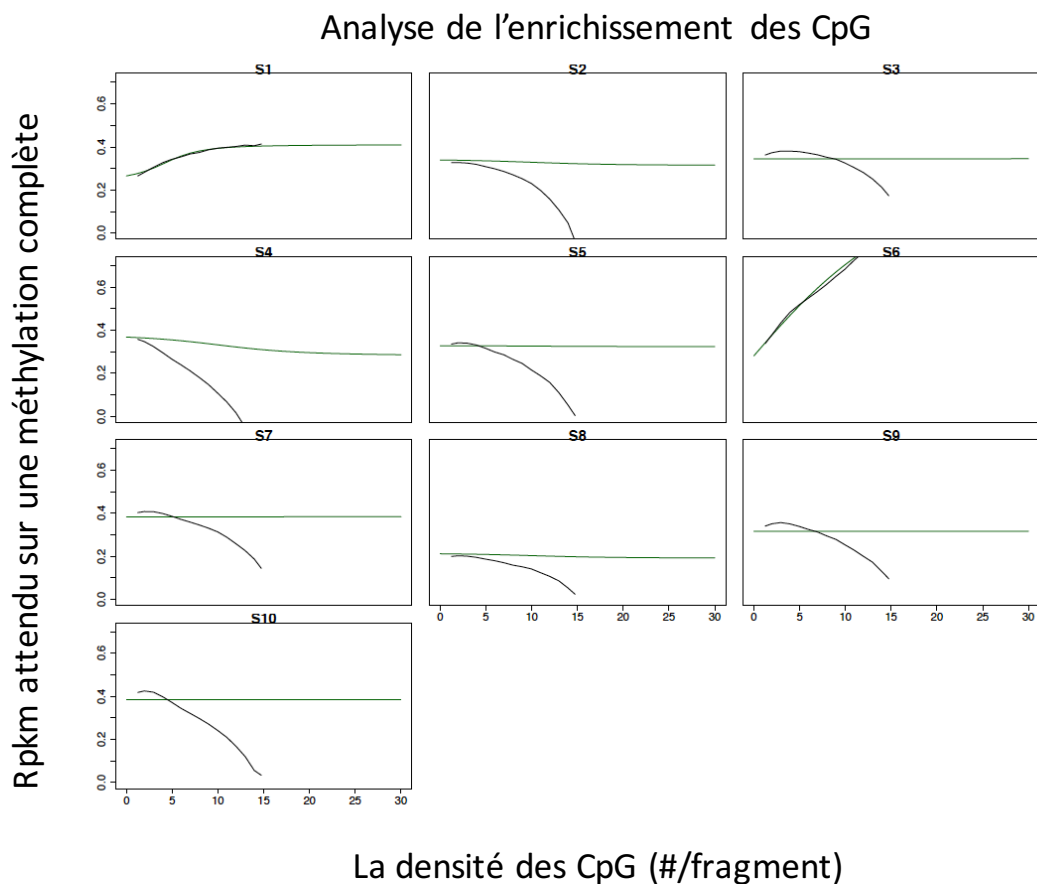
**Figure 21:** Enrichissement CpG incluant l'INPUT WT-5mC

### Analyse de l'enrichissement des CpG



**Figure 22:** Enrichissement CpG incluant l'INPUT TET2-5hmC





**Figure 23:** Enrichissement CpG incluant l'INPUT WT-5hmC

### 3.4.3 Couverture estimée de l'OFFSET

Le décalage de couverture estimé a été déterminé pour chaque échantillon par rapport à son INPUT. Il s'agit d'une estimation du bruit expérimental et des valeurs supérieures à 0,9 indiquent un manque d'efficacité d'enrichissement (186). Après avoir estimé la couverture de l'OFFSET sur les échantillons de 5mC et 5hmC, nous avons obtenu cette liste qui démontre la valeur de chaque échantillon.

**Tableau 6:** Couverture estimée de l'OFFSET pour les échantillons de 5mC par rapport à leur INPUT

<b>Échantillons</b>	<b>OFFSET</b>
Tet2_128	0.42
Tet2_224.B	0.83
Tet2_229.A	0.29
F Tet2_246	1.04
Tet2_385.B	1.24
Tet2_447.E	0.67
Tet2_Z.189	1.15
Tet2_Z.286	1.32
Tet2_Z.304	0.67
Tet2_Z.863	2.26
WT_009.D	1.31
WT_21.A	0.97
WT_298.XA	0.69
WT_380.C	0.97
WT_515.A	E 0.72
WT_74.D	1.29
WT_Z.0234	0.49

**Tableau 7:** Couverture estimée de l'OFFSET pour les échantillons de 5hmC par rapport à leur INPUT

<b>Échantillons</b>	<b>OFFSET</b>
Tet2_128.F	1.00
Tet2_224.B	0.98
Tet2_229.A	1.08
Tet2_246.E	0.80
Tet2_385.B	1.07
Tet2_447.E	1.04
Tet2_Z.189	1.03
Tet2_Z.286	1.10
Tet2_Z.304	1.05
Tet2_Z.863	0.94
WT_009.D	0.82
WT_21.A	1.05
WT_380.C	1.07
WT_515.A	1.14
WT_74.D	1.01

### 3.4.4 Résumé des résultats de CQ

Sur la base de ces résultats, un résumé du CQ a été compilé mettant en évidence les échantillons individuels recommandés pour une analyse continue.

**Tableau 8:** Résumé des résultats de CQ pour les échantillons de 5mC

Sample	Sample Number for CpGEnrichment graphs	Sample Number for INPUT CpGEnrichment graphs	Number of Peaks 5mC	Number of Reads 5mC	Gains vs Losses	Saturation	Offset	Pass/Fail
Tet2_128.F	S1	S11	2053	18.04	Gains All (3% max)	0.99	0.4212821 92570642	Pass
Tet2_224.B	S2	S12	3866	43.8	Both (slightly higher gains 0.6% max)	0.99	0.8276573 04973697	Pass
Tet2_229.A	S3	S13	23454	16.07	High losses CpG (1.5%)	0.99	0.2870282 91646671	?
Tet2_246.E	S4	S14	377706	35.3	More losses than gains (12% loss in exons and 4% in CpG)	0.99	1.0382076 5450982	Fail
Tet2_385.B	S5	S15	1256	40.4	More loss CpG and TSS (0.04%) gains exon / gene/ introns	0.99	1.2373942 572898	Fail
Tet2_447.E	S6	S16	20614	36	More gains (4%) except CpG islands (3% losses)	1	0.6690712 23357035	Pass
Tet2_Z.189	S7	S17	81472	41.5	More gains in all (0.6%) little gain in CpG	0.99	1.1485554 0751164	Fail

Sample	Sample Number for CpGEnrichment graphs	Sample Number for INPUT CpGEnrichment graphs	Number of Peaks 5mC	Number of Reads 5mC	Gains vs Losses	Saturation	Offset	Pass/Fail
Tet2_Z.286	S8	S18	76337	35.9	More gains on average (1%) more losses in CpG (0.1%)	0.98	1.3186574 1297365	Fail
Tet2_Z.304	S9	S19	16239	46	More gains on average (8%) more losses in CpG (1%)	1	0.6671617 86059888	Pass
Tet2_Z.863	S10	S20	357190	30.2	More losses in everything (10%)	0.97	2.2633072 426716	Fail
WT_009.D	S21	S28	455420	42.1	Gains and losses but more losses (15%) esp CpG	0.87	1.3134762 4353447	Fail
WT_21.A	S22	S29	11013	27.8	Gains in all (0.6%)	0.99	0.9690632 58238819	Fail
WT_298.X A	S23	S30	17660	37	Losses in all, highest in CpG (3%)	0.99	0.6862022 03141241	?
WT_380.C	S24	S31	255	53.5	More Gains in all except CpG and TSS (about even gains / losses) 0.06%	1	0.9695262 59854654	Fail
WT_515.A	S25	S32	4212	22.06	Gains in all, close to even in CpG (1.5%)	0.99	0.7151830 57674551	Pass
WT_74.D	S26	S33	564909	48.8	More losses in everything (30%)	0.99	1.2922761 271085	Fail
WT_Z. 0234	S27	S34	161748	34.7	Losses in all, highest in CpG (12%)	1	0.4911157 65447921	?

**Tableau 9:** Résumé des résultats de CQ pour les échantillons de 5hmC

Sample	Sample Number for CpGEnrichment graphs	Sample Number for INPUT CpGEnrichment graphs	Number of Peaks 5HmC	Number of Reads 5HmC	Gains vs Losses	Saturation	Offset	Pass/Fail
Tet2_128.F	S1	S11	41	39.9	Gains and losses (even CpG) gains only TSS (0.003%)	0.98	0.9964053 81716814	Fail
Tet2_224.B	S2	S12	981	35.5	High losses in CpG and TSS (0.025%)	0.99	0.9812839 58171037	Fail
Tet2_229.A	S3	S13	30	67.2	Mainly losses except TSS (0.025%)	0.99	1.0832624 1438667	Fail
Tet2_246.E	S4	S14	1330	33.7	Gains and losses more gains in all esp CpG (2.5%)	0.98	0.7977750 00064857	Pass
Tet2_385.B	S5	S15	13603	35	Gains in all except CpG (0.3%)	0.99	1.0664958 1460783	Fail
Tet2_447.E	S6	S16	91	43.5	More gains in TSS more losses in everything else (0.008%)	0.99	1.0419697 4225106	Fail
Tet2_Z.189	S7	S17	665	35.9	Gains and losses but more gains esp CpG and TSS (0.25%)	0.99	1.0347984 7628357	Fail
Tet2_Z.286	S8	S18	6	46.7	No difference in CpG or TSS (0.0035)	0.99	1.1007290 1820659	Fail

### 3.4.5 Résultats des échantillons choisis

Après avoir retiré de l'analyse les échantillons de faible qualité, au total 7 échantillons étaient analysables ; 2 échantillons de 5hmC et 5 échantillons de 5mC. Le contrôle de qualité a révélé des problèmes similaires, mais légèrement moins graves, dans la qualité de l'échantillon pour l'ensemble 5mC, tandis que l'ensemble 5hmC, qui se compose d'un seul échantillon WT et d'un échantillon TET2, n'est pas statistiquement fiable.

**Tableau 10:** Échantillons finaux utilisés pour l'ensemble de données 5mC

Échantillons	Résultats de CQ
Tet2_128.F	PASS
Tet_224.B	PASS
Tet2_447.E	PASS
Tet2_Z.304	PASS
WT_515.A	PASS

**Tableau 11:** Échantillons finaux utilisés pour l'ensemble de données 5hmC

Échantillons	Résultats de CQ
Tet2_246.E	PASS
WT_009.D	PASS

### 3.4.6 Analyse des régions méthylées différentiellement

L'analyse de méthylation différentielle dans QSEA est basée sur des modèles linéaires généralisés (GLM) et un test de rapport de vraisemblance, similaire aux tests effectués pour détecter les gènes exprimés de manière différentielle, mis en œuvre dans DESeq2 (189) ou EdgeR (190) En bref, les nombres de lectures sont modélisés avec une distribution binomiale négative avec une moyenne et un paramètre de dispersion. Pour chaque fenêtre,

le GLM est ajusté avec une fonction de lien logarithmique. Pour les tests de signifiante, un modèle réduit est ajusté et le rapport de vraisemblance (LR) des modèles est comparé à une distribution de 'Chi-squared'.

Les résultats de l'analyse DMR et DHMR ont été filtrés pour les régions avec une annotation et une valeur p ajustée à  $< 0,01$  et filtrés séparément pour les régions avec une valeur p ajustée à  $< 0,1$ . La valeur p ajustée de filtrage  $< 0,01$  a abouti à 18 535 régions pour l'ensemble 5hmC et 2 régions pour l'ensemble 5mC. Le filtrage  $p_{adj} < 0,1$  a identifié 596 333 régions pour 5hmC et 30 régions pour 5mC.

**Tableau 12:** Les DMR dans TET2 vs WT 5mC

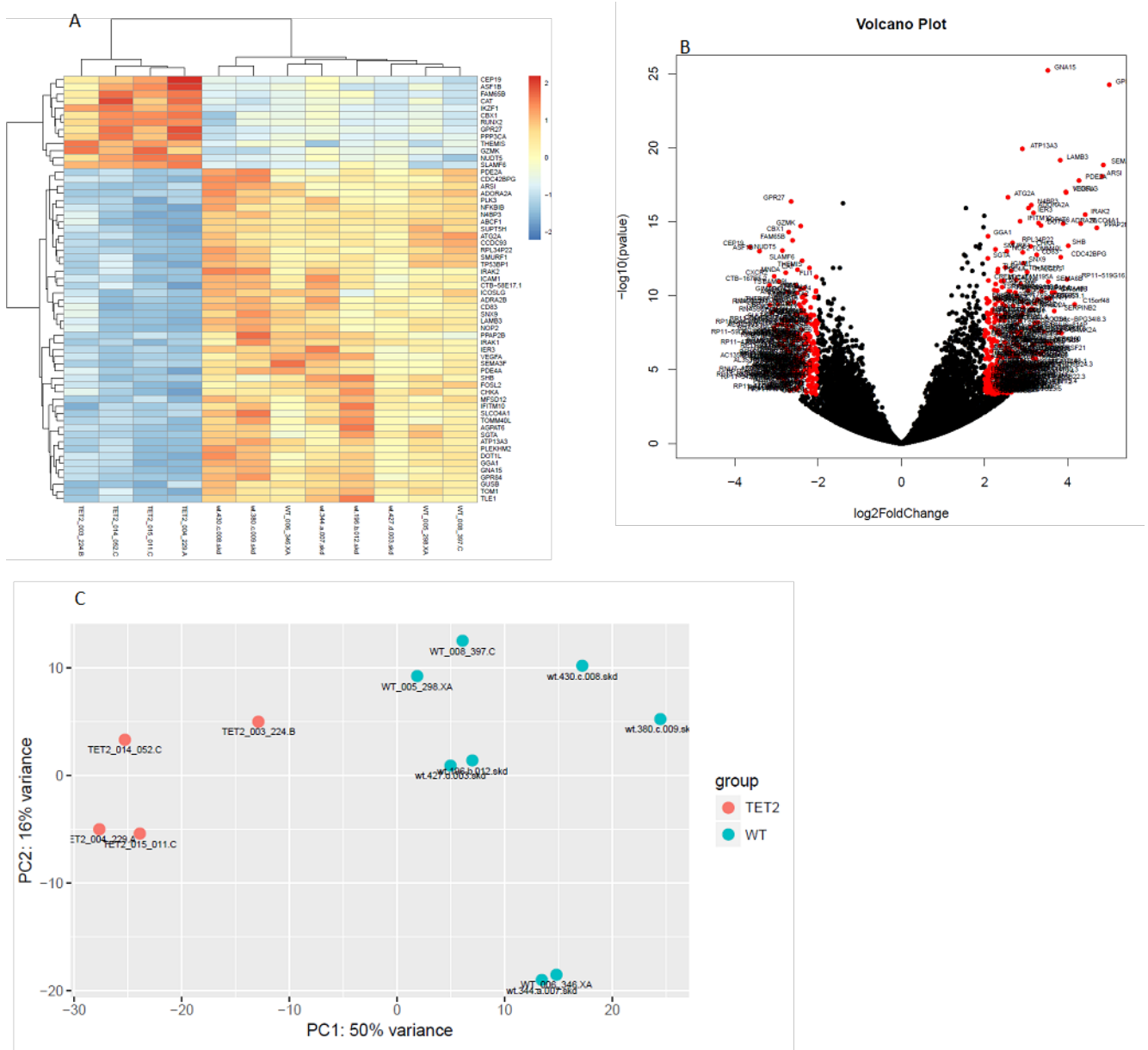
chr	window_start	window_end	CpG_density	CpG_density.1	Gene bodies
chr1	106633501	106634000	1.71334100266756	1.71334100266756	
chr1	121426501	121427000	5.94131164464807	5.94131164464807	EMBP1
chr2	91821501	91822000	2.67752185562371	2.67752185562371	
chr2	91974001	91974500	2.01911711124135	2.01911711124135	
chr3	66537001	66537500	3.98890999545766	3.98890999545766	LRIG1
chr3	96870501	96871000	2.03451698942099	2.03451698942099	EPHA6
chr5	46354001	46354500	1.52615291344495	1.52615291344495	HCN1
chr7	17404501	17405000	2.0734486695675	2.0734486695675	MEOX2
chr7	61099001	61099500	4.30402808183102	4.30402808183102	
chr7	129722501	129723000	0.98822923475088	0.98822923475088	KLHDC10
chr8	15414501	15415000	1.90838540216302	1.90838540216302	TUSC3
chr8	43805501	43806000	3.20142325985684	3.20142325985684	FNTA
chr8	43832001	43832500	4.35676095534684	4.35676095534684	HGSNAT
chr8	43836001	43836500	5.67201478185152	5.67201478185152	
chr8	43837501	43838000	4.0924113581565	4.0924113581565	
chr8	46843001	46843500	5.65672779928732	5.65672779928732	
chr9	46812001	46812500	4.93898387490912	4.93898387490912	KGFLP1
chr9	46828001	46828500	2.27741372114796	2.27741372114796	
chr9	66808501	66809000	2.62791956057473	2.62791956057473	
chr10	42534501	42535000	5.88729259013734	5.88729259013734	
chr11	48856501	48857000	2.64283770027575	2.64283770027575	
chr11	48898501	48899000	2.58787770901264	2.58787770901264	
chr12	130877001	130877500	15.530252271221	15.530252271221	
chr14	19033501	19034000	1.66704026566557	1.66704026566557	
chr14	22619001	22619500	3.05306517611968	3.05306517611968	TCRA
chr17	71144501	71145000	6.88507122630349	6.88507122630349	SDK2
chr19	44580001	44580500	1.46898229431959	1.46898229431959	ZNF284
chr20	26313001	26313500	2.51875849871985	2.51875849871985	
chrX	14122001	14122500	2.98010327425215	2.98010327425215	TRAPPC2
chrX	61716001	61716500	7.16259503062319	7.16259503062319	

### 3.4.7 L'expression génique dans les TET2 vs WT

Afin d'identifier la modulation de l'expression génique dans les TET2 vs WT, un RNA-Seq a été effectué sur les individus qui possèdent une mutation sur TET2 vs les individus WT. Notre cohorte possède des échantillons d'ADN et aussi d'ARN qui ont été extraits au moment du recrutement et du prélèvement du sang des individus de notre cohorte. Le RNA-Seq a été effectué sur 4 individus TET2 et 5 individus WT. Les résultats sont présentés



dans la Figure 24 (page 97). Une légère modulation de l'expression de certains gènes est observée mais cette différence n'est pas statistiquement significative.



**Figure 24:** Expression génique dans les TET2 vs WT

(A) Heat-map qui représente l'expression des gènes dans les TET2 VS WT, (B) Volcano Plot qui démontre les gènes qui sont surexprimés VS les gènes qui sont sous-exprimés dans les TET2 VS WT, (C) PCA regroupement les TET2 et les WT.

**CHAPITRE 4**  
**DISCUSSION**

## 4. DISCUSSION

### 4.1 Étiologie du ratio du biais de l'inactivation du chromosome X dans les cellules sanguines de femmes vieillissantes

#### 4.1.1 Résumé des travaux de recherche

Le vieillissement de l'hématopoïèse chez l'humain est le champ d'intérêt du laboratoire de Dr Busque depuis une vingtaine d'années (52). De nombreuses études, dont la nôtre, ont démontré une augmentation du ratio de biais de l'ICX dans les cellules sanguines après l'âge de 50 à 60 ans (54) (71). Plusieurs causes peuvent expliquer ce phénomène (122). Dans le cadre des travaux de recherche de cette thèse, nous nous sommes concentrés sur (i) la sélection hémizygote (les cellules avec un des deux X parentaux dans l'état actif dominant l'autre population) qui est considérée comme génétiquement prédéterminée et (ii) l'acquisition de mutations acquises dans les cellules somatiques de moelle osseuse et donnant un avantage de croissance dont, par exemple, *TET2*, *DNMT3A* et *ASXL1*.

La première partie de nos travaux porte sur la caractérisation moléculaire de ces deux phénomènes et particulièrement sur l'identification de variants génétiques responsables de la sélection hémizygote. Pour ce faire, nous avons utilisé les cellules sanguines de 3002 individus âgés entre 37 et 100 ans dont certains avaient des liens familiaux (apparentés) et d'autres sans lien (non apparentés). Ceux qui sont liés forment 311 familles. Les cellules sanguines de ces individus ont été séparées en fractions PMN et mononucléaires qui se présentent en cellules T (CD3 +), cellules B (CD19 +) et monocytes (CD14 +). Ces fractions de cellules sanguines ont été comparées avec les cellules épithéliales de chacun

de ces individus. Par la suite nous avons mesuré le ratio de l'ICX en utilisant la méthode de HUMARA permettant ainsi d'obtenir deux mesures différentes, le DS et le Psup. Ces mesures ont été utilisées dans nos calculs statistiques dans le but de répondre aux objectifs suivants :

- 1- Identifier la corrélation entre le biais de l'ICX et le vieillissement dans des tissus spécifiques,
- 2- Identifier la corrélation entre le biais de l'ICX et les mutations somatiques acquises et évaluer l'impact de CHIP sur l'ICX,
- 3- Identifier une cohorte d'individus sans mutation somatique pour tester l'héritabilité du trait de biais,
- 4- Effectuer une analyse GWAS et par liaison génétique pour identifier le SNP associé au trait de biais de l'ICX.

Les résultats générés par ces études ont révélé plusieurs découvertes pertinentes dans la compréhension de l'étiologie du biais de l'ICX en permettant de faire la distinction entre le phénomène de sélection hémizygote et l'acquisition de mutations donnant un avantage de croissance ainsi que la découverte pour la première fois des régions sur l'ADN liées au biais de l'ICX. Ces régions, qui sont représentées par des SNP sur le chromosome 1 et le chromosome X, démontrent la présence d'une composante génétique contribue au biais de l'ICX et la possibilité qu'une interaction entre ces gènes puisse exister et probablement l'interruption de cette interaction peut être la cause de la sélection hémizygote. Ces résultats sont en préparation pour soumission dans le journal Blood. Dans cette section nous mettrons en contexte la pertinence de ces résultats dans la compréhension de la

relation entre le biais de l'ICX et les types de tissus et de cellules sanguines, l'âge, le VAF des mutations CHIP, et l'héritabilité. De plus, nous allons mettre en évidence l'importance de la découverte, grâce aux études de génotypage, des régions sur l'ADN qui sont associées au biais de l'ICX.

#### **4.1.2 La prévalence du biais de l'ICX varie en fonction des tissus et des types de cellules**

Nous avons démontré que la prévalence du biais de l'ICX est plus élevée dans les cellules sanguines que dans les cellules épithéliales (

**Figure 9**, page 64). Nos données démontrent que les tissus dérivés du sang expriment la prévalence la plus élevée de biais de l'ICX, ce qui est cohérent avec les travaux antérieurs de Zito *et al.* qui ont également démontré une augmentation de la prévalence du biais de l'ICX dans les cellules sanguines mais pas dans les cellules adipeuses et cutanées (29). Le fait que le biais de l'ICX est inférieur dans les cellules T par rapport aux PMN pourrait être expliqué par la probabilité que ce phénomène ne se produise que dans les cellules souches myéloïdes. Il est possible également que le biais de l'ICX se produise également dans le pool de cellules souches lymphoïdes, mais ceci serait masqué dans la population de cellules T car une proportion importante de cellules T présentes dans le sang se forment plus tôt dans la vie, avant l'initiation du biais de l'ICX, tel que démontré par Gale *et al.* (145). De plus, un renouvellement cellulaire plus fréquent pourrait augmenter le risque de biais, comme dans les cellules myéloïdes et, par conséquent, la longévité plus longue des cellules T entraînerait moins de biais de l'ICX.

Afin d'examiner la corrélation du biais de l'ICX entre les tissus, nous avons comparé les niveaux de biais en utilisant la mesure P<sub>sup</sub> dans tous les échantillons de tous les individus qu'ils soient CHIP ou non. Nous avons observé une forte corrélation des scores P<sub>sup</sub> chez différents types de cellules sanguines (Table 1, page 62; Figure 10, page 65). Ces résultats concordent avec l'étude de Zito *et al.* qui ont démontré dans une cohorte de jumeaux que le biais de l'ICX est similaire entre les cellules immunitaires mais diffère avec les cellules adipeuses et les cellules cutanées (29). La corrélation la plus forte de P<sub>sup</sub> est entre les PMN et les monocytes ce qui indique que les tissus myéloïdes peuvent partager des niveaux de biais de l'ICX très similaires, suivis par la corrélation entre les cellules PMN et les cellules B puis celle avec les cellules T. Les monocytes présentent une corrélation plus forte avec les cellules B qu'avec les cellules T. La corrélation entre les cellules sanguines et les cellules épithéliales est la plus faible. L'analyse des cellules de moelle osseuse de trois femmes en bonne santé dont leurs cellules sanguines périphériques présentent un biais de l'ICX a révélé une bonne corrélation entre les cellules progénitrices hématopoïétiques immatures (cellules CD34<sup>pos</sup>CD38<sup>neg</sup>) et tous les types de cellules sanguines (données non présentées). Ceci suggère que le biais de l'ICX se produit au niveau des cellules hématopoïétiques souches avant la différenciation des lignées cellulaires. Ainsi, nous avons documenté une concordance du biais de l'ICX entre les cellules précurseurs hématopoïétiques CD34<sup>pos</sup>CD38<sup>neg</sup> et toutes les lignées de cellules sanguines.

De plus, nos données démontrent que le biais de l'ICX est spécifique aux tissus et indiquent que les motifs de l'ICX pourraient être partiellement contrôlés par des mécanismes de

régulation spécifiques aux tissus, ce qui est en concordance avec les travaux précédents de Zito *et al.* (29).

#### **4.1.3 Le biais de l'ICX dans les cellules sanguines varie selon l'âge.**

Nous savons que les niveaux de biais de l'ICX dans les cellules sanguines augmentent avec l'âge, comme le démontrent plusieurs études incluant la nôtre (152, 191, 192) (54). Cependant, les connaissances sur la relation entre le biais de l'ICX et le vieillissement des cellules sanguines sont très limitées. Afin d'explorer ceci, nous avons étudié l'association entre l'âge et le ratio de biais de l'ICX dans les PMN, les monocytes, les cellules B, les cellules T et les cellules épithéliales (Figure 27, page 149).

La prévalence de biais de l'ICX augmente avec l'âge dans toutes les cellules sanguines, mais l'association n'était pas statistiquement significative dans les cellules B ni dans les cellules épithéliales. Nous pensons fortement que le faible nombre de cellules B est la cause pour laquelle la corrélation entre le biais de l'ICX et l'âge n'est pas statistiquement significative dans les cellules B. La corrélation entre l'âge et le biais de l'ICX est forte dans les PMN et les cellules CD14 mais faible dans les cellules T, ce qui pourrait également être expliqué par la longévité des cellules myéloïdes par rapport aux cellules T tel qu'expliqué auparavant.

L'prévalence de biais de l'ICX augmente progressivement avec l'âge, de sorte que près de 60% des femmes de plus de 80 ans présentent un biais de l'ICX dans leurs PMN et leurs monocytes. Cela suggère que la plupart des femmes pourraient éventuellement présenter

un biais de l'ICX dans leurs cellules myéloïdes si elles disposent de suffisamment de temps et que le biais de l'ICX serait de ce fait un phénotype très répandu dans les populations vieillissantes. Nous avons donc démontré que le biais de l'ICX augmente avec l'âge et que ce phénomène est spécifique et distinctif aux tissus sanguins, puisque l'on n'observe pas de biais dans les cellules épithéliales, ni de biais de l'ICX lié à l'âge dans les cellules épithéliales.

#### **4.1.4 Le biais de l'ICX est associé avec le VAF des mutations CHIP**

Les mutations CHIP sont associées à un risque accru de cancer et de mortalité et leurs prévalences augmentent avec l'âge, tout comme le biais de l'ICX (87, 95). Cependant, il n'y a aucune preuve d'une corrélation entre le biais de l'ICX et les mutations CHIP. Afin d'explorer ceci, nous avons étudié l'association entre le biais de l'ICX et les mutations CHIP en utilisant la mesure de DS et les VAF des mutations CHIP que nous avons obtenus par séquençage de l'ADN des individus de la même cohorte en utilisant la technologie de NGS (95). Ces travaux nous ont donc permis de démontrer que le DS des cellules PMN est positivement associé avec le VAF des mutations CHIP, avec le VAF des individus porteurs de mutations sur *TET2* et avec le VAF des individus porteurs de mutations sur *DNMT3* (Tableau 16, page 150). Ces résultats ont été abordés sommairement dans une de nos études publiées dans Blood en 2017 (95).

Une étude précédente a suggéré qu'une sélection positive de cellules porteuses d'une mutation somatique conduit à une HC et à un biais de l'ICX car les cellules sélectionnées portent le même chromosome X parental inactivé (29). Cette HC induite par une mutation



somatique (ou CHIP) est connue pour se produire chez des personnes âgées en bonne santé (4, 112). Les résultats de cette étude concordent avec nos résultats, ce qui suggère fortement que le biais de l'ICX corrèle avec les mutations CHIP.

Une autre étude antérieure réalisée sur des patients atteints de polyarthrite rhumatoïde (193) rapporte que l'HC module l'inflammation chronique chez ces patients. Ceci a été effectué en comparant les résultats de mutations avec des paramètres cliniques dans le but d'étudier les associations entre l'HC et le phénotype clinique. Ils ont découvert que les mutations *TET2* et *DNMT3* sont les mutations les plus courantes chez ces patients et ont suggéré que ces mutations pourraient donner un avantage de survie aux clones affectés. Ils ont également démontré que la prévalence de l'HC augmente avec l'âge chez les patients atteints de polyarthrite rhumatoïde. Cela dit, la hausse du biais de l'ICX chez les personnes atteintes de cette maladie par rapport à ceux qui ne sont pas atteints pourrait s'expliquer par l'augmentation de la prévalence de l'HC avec l'âge chez les patients atteints (193). Par ailleurs, l'étude de Zito *et al.* a démontré que le biais de l'ICX augmente chez les fumeurs plus âgés (29) ce qui est cohérent avec l'augmentation du l'HC observée chez les fumeurs (88, 112). Dans l'ensemble, ces données, y compris les nôtres, suggèrent fortement un lien possible entre le biais de l'ICX et l'HC.

Les résultats de nos travaux démontrent donc que le phénomène de biais de l'ICX peut être causé par l'acquisition de mutations donnant un avantage de croissance dans les cellules somatiques de moelle osseuse, ce qui est complètement différent des causes génétiquement prédéterminées.

#### 4.1.5 L'héritabilité du biais de l'ICX est partiellement génétique

Les études chez les jumeaux sont un modèle solide pour étudier l'héritabilité des caractères complexes. Des études antérieures réalisées sur des jumeaux ont démontré que le biais de l'ICX dans les échantillons de sang est héréditaire :  $h^2 = 0,68$  dans les granulocytes de jumeaux vieillissants,  $0,58$  dans les cellules sanguines périphériques et  $0,34$  dans les lignées cellulaires lymphoblastoïdes (29, 71, 72, 194). Cependant, ces études n'ont pas examiné l'héritabilité dans les PMN, les cellules T et les cellules épithéliales en utilisant les mesures de DS et de P<sub>sup</sub>. Pour estimer l'influence des effets génétiques (héritabilité) et des facteurs environnementaux sur le biais de l'ICX dans ces trois tissus, nous avons évalué les tissus prélevés chez 1734 femmes, dont certaines sont des CHIP et d'autres non CHIP, issues de 311 familles franco-canadiennes en utilisant une approche par paires fraternelles. En utilisant le DS, nous avons été en mesure de démontrer que le biais de l'ICX est héritable dans les PMN ( $h^2 = 0,097$ ), dans les cellules T ( $h^2 = 0,255$ ) mais pas dans les cellules épithéliales ( $h^2 = 0,045$ ) (Table 2, page 62). Par contre, en utilisant le P<sub>sup</sub>, nous avons trouvé que le biais de l'ICX est héritable dans les PMN ( $h^2 = 0,205$ ), dans les cellules T ( $h^2 = 0,303$ ) mais également dans les cellules épithéliales ( $h^2 = 0,310$ ) (Table 2, page 62). Ainsi, nous avons constaté que l'héritabilité de biais de l'ICX dans le DS est présente dans les tissus sanguins mais pas dans les cellules épithéliales. La présence de l'héritabilité dans les cellules sanguines des femmes âgées uniquement confirme notre hypothèse selon laquelle il existe une association entre le biais de l'ICX et l'HC et ceci est cohérent avec les travaux de Zito *et al.* où ils ont démontré que le biais de l'ICX est héritable dans le sang des femmes de plus de 55 ans mais pas chez les individus plus jeunes

ou dans d'autres tissus (29). Nous avons également constaté qu'en retirant les individus CHIP de l'analyse (données non-présentées), l'effet est peu important sur l'héritabilité. Ceci est possiblement dû à un petit nombre d'individus ou parce que les mutations CHIP pourraient être des phénocopies, causant ainsi un impact sur l'héritabilité.

Des études antérieures sur des jumeaux ont affirmé que le biais de l'ICX dans le sang avait une composante génétique (29, 71, 72, 194). En outre, l'héritabilité du biais de l'ICX dans les cellules sanguines chez les femmes vieillissantes confirme le rôle des variants génétiques en tant que régulateurs du biais de l'ICX qui varient selon l'âge. Dans l'ensemble, nos données convergent pour suggérer un locus sur un gène lié au chromosome X, prédisposant le biais de l'ICX. Cette hypothèse est soutenue par des études qui ont indiqué que (i) la sélection hémizygote pourrait également favoriser des motifs de biais de l'ICX dans les cellules saines, (ii) une sélection positive de cellules portant une mutation liée au chromosome X conduit au biais de l'ICX (29, 70).

Cependant, le seul moyen qui peut nous aider à trouver un variant possiblement associé au biais de l'ICX chez les femmes vieillissantes est d'utiliser l'analyse par liaison génétique, une méthode d'analyse statistique développée pour cartographier les gènes par rapport à leurs emplacements chromosomiques en fonction d'un trait héréditaire.

#### **4.1.6 Des SNP sur le chromosome 23 et 1 sont associés avec le biais de l'ICX**

Afin d'étudier l'implication possible d'un variant génétique associé au biais de l'ICX dans les cellules sanguines de femmes apparentées, nous avons réalisé une analyse par liaison

génétiq ue sur notre cohorte qui contient 311 familles. Compte tenu de nos résultats d'héritabilité qui ont démontré que le biais de l'ICX est héréditaire dans les cellules sanguines, nous nous attendions à ce que l'étude de liaison puisse identifier une région associée avec le trait de biais de l'ICX. En effet, des régions intéressantes ont été identifiées sur les chromosomes 1 et X qui pourraient être associées à un trait de biais de l'ICX.

L'analyse QTL réalisée sur tous les échantillons de cellules T en Psup a révélé une région associée sur le chromosome X qui dépasse le seuil de 3,9 LOD (Figure 13A, page 69). En effet, cette région a un score LOD de 5 ce qui est statistiquement significatif. Cette très grande région contient de nombreux gènes intéressants tels que l'*IL2RG* (Figure 13B, page 69), qui est un gène qui code pour la chaîne gamma commune du récepteur, également connue sous le nom de sous-unité de récepteur d'interleukine-2 gamma. Ce récepteur est commun pour plusieurs cytokines et joue un rôle important dans la régulation du système immunitaire notamment les cellules T (195). Cela dit, plusieurs études nous laissent penser que nous pourrions trouver des régions d'intérêt associées au biais de l'ICX sur le chromosome X (70, 74). En effet, une de ces études a démontré chez les chats femelles Safari qu'une forme excessive de biais de l'ICX sans conséquence pathologique se développe lors du vieillissement et est le résultat de la sélection hémizygote. Ils ont aussi démontré un avantage compétitif pour tous les CSH avec un phénotype chromosomique X spécifique et, qu'un gène du chromosome X (ou des gènes) régule la réplication, la différenciation et / ou la survie des CSH (70). L'autre étude portant sur deux femmes hétérozygotes atteinte d'agammaglobulinémie liée au chromosome X a démontré que le défaut génétique responsable de cette maladie affecte les cellules T ainsi que les cellules

B. Les auteurs de cette étude considèrent que des études de cartographie permettront l'identification et la fonction du gène responsable de l'agammaglobulinémie liée au chromosome X (74). Dans nos résultats, la présence d'un SNP sur le chromosome X dans les cellules T avec un LOD fort suggèrent fortement que cette région pourrait être liée au trait de biais de l'ICX. En somme, la présence de gènes importants dans la régulation du système immunitaire, notamment les cellules T, tel que l'*IL2RG*, renforce notre hypothèse que probablement un de ces gènes est responsable de la sélection hémizygote.

Lors de nos analyses par liaison génétique, une autre région sur le chromosome 1 avec un score LOD très élevé de 6 a été retrouvée grâce à une analyse VC sur tous les échantillons de PMN Psup (Figure 13C, page 69). Cette région contient de nombreux gènes intéressants tels que l'*IL6R* (Figure 13D, page 69). Le gène de l'*IL6R* code pour la glycoprotéine 130 (gp130) qui est un récepteur commun à plusieurs cytokines de la famille de l'IL6. gp130 joue un rôle central dans la régulation de plusieurs processus biologiques tels que la régulation de la réponse immunitaire, le métabolisme et le système nerveux (196). Par ailleurs, l'analyse QTL effectuée dans les échantillons de PMN et de cellules B a identifié une région associée sur le chromosome 1 avec un score LOD de 3,2 et 3,4 respectivement. Ce score LOD est inférieur au seuil de 3.9 LOD. Fait intéressant, toutes ces régions sont situées sur les mêmes locus du chromosome 1.

La détection d'une autre région possiblement associée au biais de l'ICX sur le chromosome 1 nous a conduit à concevoir plusieurs hypothèses : (i) cette région pourrait être associée à un gène ou un sous-ensemble de gènes qui pourraient jouer un rôle dans la régulation des

gènes sur le chromosome X, (ii) cette région pourrait être associée à des gènes avec un polymorphisme associé à un avantage de croissance cellulaire. En effet, cette région possède aussi un nombre de gènes importants tel que l'*IL6R* et l'interaction entre certains gènes pourrait jouer un rôle dans la régulation de la croissance des cellules. L'interruption de cette interaction causée par un défaut sur un gène lié au chromosome X pourrait être la cause de l'avantage de croissance qui conduit au biais de l'ICX.

Lors de nos analyses par liaison génétique, nous avons également détecté d'autres régions associées au biais de l'ICX sur différents chromosomes (Figure 28 à Figure 31, pages 151 à 154). Ces régions sont représentées dans le Tableau 13 (page 111). Elles pourraient être importantes puisqu'il est possible que d'autres gènes situés sur divers chromosomes jouent aussi un rôle sur l'initiation du biais de l'ICX. Ceci dit, nous nous sommes concentrés sur le chromosome X, puisque nous pensons fortement que la sélection hémizygote est liée à un gène sur ce chromosome, et sur le chromosome 1 puisqu'une région de celui-ci sort dans plusieurs analyses avec un LOD très fort.

**Tableau 13:** Résultats d'analyse par liaison génétique en 2-points avec Merlin sur le trait du biais de l'ICX

Mesure	Analyse	Individus	Tissus	Chromosome	SNP
DS	QTL	ALL	Monocytes	-	-
			Cellules B	-	-
			Cellules T	-	-
			PMN	-	-
			Cellules épithéliales	-	-
		CHIP Neg	Monocytes	-	-
			Cellules B	-	-
			Cellules T	-	-
			PMN	-	-
			Cellules épithéliales	-	-
	VC	ALL	Monocytes	-	-
			Cellules B	-	-
			Cellules T	-	-
			PMN	-	-
			Cellules épithéliales	-	-
		CHIP Neg	Monocytes	Chr 6	rs116343052
Cellules B	Chr 2	rs74685148			
Cellules T	-	-			
PMN	-	-			
Cellules épithéliales	-	-			
Pasup	QTL	ALL	Monocytes	-	-
			Cellules B	Chr1	rs12116550
			Cellules T	Chr23	rs4827384
			PMN	Chr1	rs6671166
			Cellules épithéliales	-	-
		CHIP Neg	Monocytes	-	-
			Cellules B	-	-
			Cellules T	Chr23	rs2497911
			PMN	Chr1	rs74636473
			Cellules épithéliales	Chr18	rs77943753
	VC	ALL	Monocytes	-	-
			Cellules B	Chr1	rs9426935
			Cellules T	-	-
			PMN	Chr1	rs6427005
			Cellules épithéliales	-	-
		CHIP Neg	Monocytes	-	-
Cellules B	-	-			
Cellules T	Chr8	rs6468592			
PMN	Chr1	rs151277197			
Cellules épithéliales	-	-			

#### 4.1.7 Étude d'associations génomiques (GWAS)

Selon nos résultats d'héritabilité dans le biais de l'ICX, une composante génétique pourrait jouer un rôle dans l'initiation du biais de l'ICX. Nous avons émis l'hypothèse qu'un variant génétique pourrait jouer un rôle de régulateur de biais de l'ICX dans les cellules sanguines qui varie selon l'âge. Afin d'explorer cette hypothèse, nous avons mené une étude d'association génomique (GWAS) en utilisant une cohorte de 3000 femmes apparentées et non apparentées. Ces femmes en bonne santé ont été recrutées sans *a priori*. Les femmes âgées de 55 ans et plus ont été séquencées pour identifier les gènes mutés de manière récurrente dans les tumeurs malignes myéloïdes (*DNMT3A*, *TET2*, *JAK2*, *ASXL1*, *TP53*, *CBL*, *KRAS*, *RUNX1*, *IDH2*, *GATA2*, *NRAS* et *CEBPA*). Ces gènes sont impliqués dans le CHIP. Bien que l'héritabilité ait démontré que le biais de l'ICX est héréditaire dans les cellules sanguines, le GWAS n'a détecté aucun locus de gènes lié au trait de biais de l'ICX. L'explication possible de ces résultats est que le biais de l'ICX est un trait complexe et qu'il peut y avoir un grand nombre de gènes impliqués. La forte prévalence du biais de l'ICX dans les cellules sanguines chez les femmes en bonne santé pourrait également compliquer la découverte d'un gène spécifique associé au trait sur le chromosome X. Dans les travaux antérieurs de Zink *et al.* (112) sur la base de l'analyse de 11 262 Islandais, une délétion de la lignée germinale dans l'intron 3 du gène de la *TERT* qui prédispose à l'HC a été trouvée. Dans une étude récente basée sur l'analyse de 97 691 participants de diverses ascendances, 3 loci associés au statut CHIP incluant *TERT* et *TET2* ont également été trouvés (197). Ainsi, notre cohorte peut être trop petite pour trouver un gène spécifique responsable du biais de l'ICX. Une autre possibilité pour expliquer pourquoi notre cohorte ne nous a pas



donné les mêmes résultats peut être due aux origines différentes des individus de ces cohortes.

## **4.2 Pénétrance oncogénique de la mutation de TET2 chez les individus âgés**

### **4.2.1 Résumé des travaux de recherche**

Le vieillissement est l'un des facteurs de risque les plus importants dans le développement de tumeurs myéloïdes malignes (122). Cependant, la relation entre l'hématopoïèse vieillissante et le développement de ces cancers est inconnue. Récemment, notre groupe et d'autres ont documenté que des mutations acquises se produisent dans les cellules sanguines d'individus vieillissants (4, 6, 177). Notre groupe a démontré que > 10% des personnes de plus de 65 ans ont une mutation de perte de fonction dans un gène appelé *TET2*. *TET2* appartient à la famille des enzymes de TET (*TET1*, 2 et 3) (178). Les enzymes *TET2* modifient la méthylation en catalysant l'oxydation du groupe méthyle de la 5mC pour former des 5hmC (179). Le 5mC et le 5hmC ont des rôles régulateurs spécifiques dans l'expression génique et ils sont considérés comme une modification épigénétique indépendante (180). *TET2* subit très souvent une mutation somatique dans les cellules cancéreuses hématopoïétiques, spécifiquement dans les tumeurs malignes myéloïdes, et son inactivation est en corrélation avec un phénotype d'hyperméthylation (4, 122).

L'objectif global de notre recherche est de comprendre comment les mutations *TET2* mènent à la transformation. Notre objectif est de mesurer l'impact de la mutation *TET2* sur l'épigénome (méthylome, hydroxyméthylome) puis d'identifier lesquelles de ces marques

épigénétiques ont un potentiel oncogène. Nous avons conçu 3 groupes de sujets : (i) un groupe d'individus vieillissants normaux sans aucune mutation dans *TET2* ( $TET2_{wt}=7$ ), (ii) un groupe de sujets vieillissants avec des mutations *TET2* ( $TET2_{mut}=10$ ) et (iii) un groupe de patients atteints d'un cancer myéloïde et présentant une mutation dans *TET2* ( $TET2_{cancer}=0$ ).

Nous avons utilisé des méthodes de méthylome et d'hydroxyméthylome suivies d'un séquençage à l'aide du NGS et de l'expression génique (RNA-Seq) pour identifier les marques épigénétiques spécifiques des 3 groupes. Cela permettra de déterminer si *TET2mut* affecte 5hmC ou 5mC dans des locus spécifiques du génome par rapport à *TET2wt*. Une fois les marques épigénétiques spécifiques de *TET2mut* identifiées, nous les comparerons à celles documentées dans le groupe *TET2cancer*. Nous pourrions identifier ceux qui sont enrichis en cancer, donc potentiellement oncogènes. Nous estimons que la comparaison des 3 groupes permettra de déterminer les régions géniques affectées par la mutation *TET2* ou les régions qui pourraient contribuer avec la mutation *TET2* afin de déclencher le cancer.

#### **4.2.2 Processus de sélection de la méthode**

Il existe de nombreuses méthodes pour cibler la modification de la cytosine; (i) les méthodes de résolution à un seul nucléotide qui utilise des dérivés d'oxydation de 5mC impliquant l'utilisation d'un séquençage au bisulfite (BS-seq) ou d'un séquençage de troisième génération, et (ii) la méthode d'enrichissement d'affinité. Les méthodes de résolution à un seul nucléotide sont très intéressantes puisqu'elles sont très spécifiques pour

cibler les modifications 5mC et 5hmC mais nécessitent des appareils de séquençage très puissants qui ont la capacité de séquencer des fragments d'ADN de plus grande taille. Étant donné que les appareils de séquençage de notre labo sont ceux d'Ion torrent, nous n'étions pas en mesure d'utiliser les méthodes de résolution à un seul nucléotide puisque les appareils d'Ion torrent disponibles peuvent seulement séquencer des fragments de 250 pb maximum. Par conséquent, la méthode d'enrichissement d'affinité a été sélectionnée. L'enrichissement par affinité est obtenu en utilisant des anticorps qui reconnaissent spécifiquement les modifications ou les dérivés de l'oxydation 5mC ou 5hmC (198).

La méthode DIP a été choisie pour le 5mC puisqu'il est abondant sur le génome humain, presque 5% des cytosines sont méthylées, et ceci facilite son immunoprécipitation (199). À l'inverse, le 5hmC n'est pas très abondant sur le génome où 0.05% des cytosines sont hydroxyméthylées (200). Ceci signifie que son immunoprécipitation peut être difficile et que nous voulions recourir à la méthode appelé hMe-Seal qui est plus fiable pour obtenir une bonne quantité de 5hmC. Cette méthode est basée sur un étiquetage chimique sélectif au 5hmC avec une biotine. En exploitant le processus de glucosylation naturelle du 5hmC par la  $\beta$ -glucosyltransferase, plusieurs méthodes ont été développées pour marquer chimiquement et enzymatiquement le 5hmC avec la biotine (198). Malheureusement, la trousse de cette méthode n'est plus sur le marché. Nous avons donc été dans l'obligation d'utiliser la même méthode que le 5mC en se basant sur le fait que plusieurs études l'ont utilisé antérieurement comme méthode d'enrichissement pour le 5hmC (201).

La méthode DIP consiste à utiliser un anticorps ciblant le 5mc ou le 5hmc pour immunoprécipiter les fragments d'ADN méthylé ou hydroxyméthylé pour ensuite les séquencer (Figure 14, page 75). Cette méthode nous donne la possibilité d'observer les cytosines modifiées dans tout le génome avec un biais minimum. En revanche, l'utilisation des anticorps peut causer des problèmes de bruits de fond puisque les anticorps peuvent immunoprécipiter des fragments non-spécifiques qui peuvent influencer les résultats de l'enrichissement. Ce bruit de fond de fragments non-spécifiques est d'autant plus critique dans le cas du 5hmC qui est en faible quantité sur le génome. Cela dit, la forte affinité des anticorps contre le 5mc et le 5hmC a été démontrée grâce aux contrôles positifs et négatifs fournis avec la trousse de la compagnie Cell-Signalling qui a été utilisée pour effectuer le DIP. Ainsi, le gène endogène fourni comme contrôle pour le 5mC et le Spike fourni comme contrôle pour le 5hmC ont permis de démontrer que l'enrichissement obtenu par cette méthode est spécifique (

Figure 17, page 83, et Figure 18, page 84).

Après l'immunoprécipitation les fragments enrichis ont été quantifiés par qPCR. Cette quantification a démontré que l'enrichissement par immunoprécipitation du 5mC était efficace et spécifique par rapport à l'IgG de souris, le contrôle négatif (Figure 19, page 85). Pour ce qui est de l'enrichissement par immunoprécipitation du 5hmC, celle-ci a donné les résultats attendus puisque le 5hmC est 20 fois moins abondant dans la cellule que dans le 5mC (Figure 19, page 85). Par contre, la présence d'un enrichissement de l'IgG semble indiquer que l'anticorps anti-5hmC est moins spécifique. En se basant sur les contrôles et les valeurs de quantification par qPCR, nous avons validé la méthode de méthylome et

d'hydroxyméthylome et cette méthode pourra être utilisée pour identifier les cibles génétiques qui confèrent un potentiel de transformation maligne.

L'instrument Ion Chef<sup>TM</sup> prépare 2 puces à la fois et chaque puce contient un seul individu. L'ion Proton<sup>TM</sup> séquence une puce à la fois. La lecture de chacune de nos puces a obtenu entre 30 et 40 millions de reads par individu. Certains papiers démontrent que 40 millions de reads sont suffisants pour chaque individu, tandis que d'autres montrent qu'il faut obtenir 80 à 100 millions de reads par individu (202). Il est donc possible que le nombre de reads obtenu par individu a nui aux tests de CQ et qu'on doit probablement fournir 80 à 100 millions de reads par individu pour couvrir complètement les fragments à séquencer afin d'obtenir un séquençage de plus grande qualité.

L'autre problème rencontré avec cette méthode est qu'il n'y a pas de trousse commercialement disponible commercialement pour faire le meDIP ou hmeDIP par NGS, encore moins pour les instruments d'Ion Torrent. De ce fait, un protocole a dû être développé puis validé en sélectionnant les anticorps et les réactifs de préparation de la bibliothèque et en configurant les conditions appropriées (quantité, taille, temps) pour le NGS et les instruments de séquençage d'Ion Torrent. Les fragments obtenus ont dû être analysés sur un bio-analyseur avant la fragmentation par sonication, puis une sélection des fragments de 250pb a été effectuée. Deux méthodes de qPCR ont été réalisées ; une après l'immunoprécipitation pour quantifier les fragments immunoprécipités et une autre après l'amplification des fragments immunoprécipités. Toutes ces étapes ont comme objectif de favoriser un enrichissement et un séquençage spécifique. Or, une de ces étapes est peut-

être responsable du rejet des échantillons par l'analyse de CQ. Cette étape a lieu après la quantification des fragments immunoprécipités. Ces fragments doivent être amplifiés par PCR selon leur quantité, c'est-à-dire que le nombre de cycle doit être augmenté afin de compenser pour une faible quantité de fragments. Par exemple, si la quantité du 5mC est entre 5-10 ng, les fragments doivent être amplifiés de 10 à 12 cycles. Dans le cas où la quantité de 5mC est faible, 2 cycles de PCR sont ajoutés pour obtenir la quantité requise pour le séquençage. En revanche la quantité de fragments 5hmC est toujours faible donc le nombre de cycle devaient être augmenté pour atteindre de 14 à 16 cycles. En rétrospective, nous pensons que l'augmentation des cycles de PCR ont également amplifié les fragments non-spécifique qui peuvent avoir influencé le séquençage. Or, il n'est pas possible de déterminer si cette étape a endommagé notre bibliothèque jusqu'à ce que les fragments soient séquencés et les analyses de CQ soient réalisées. Compte tenu du fait que la quantité des fragments de 5mC et 5hmC immunoprécipités est un élément crucial pour l'efficacité du séquençage, la méthode développée pourrait être améliorée en vérifiant préalablement l'efficacité de l'affinité des anticorps anti-5mC et anti-5hmC, en utilisant les anticorps d'un autre fournisseur ou en utilisant une trousse de meilleure qualité.

#### **4.2.3 Les méthodes d'analyses bioinformatiques**

L'analyse bioinformatique a été réalisée en utilisant des programmes développés spécifiquement pour analyser le méthylome et l'hydroxyméthylome. L'analyse de CQ est constituée de plusieurs étapes : le calcul de l'enrichissement de CpG de chaque individu, le calcul de l'enrichissement de CpG par rapport à son INPUT, le calcul des PEAK CALLING dans les 5mC et 5hmC, et l'estimation de la couverture de l'OFFSET. Ces

étapes permettent de trier les échantillons afin de conserver seulement ceux qui répondent aux critères de CQ. Dans le nombre d'échantillons qui ont été rejetés plusieurs correspondent aux échantillons de 5hmC. Le rejet de ces échantillons démontre qu'il y a eu des problématiques soit au niveau du processus d'enrichissement qui a résulté en un faible enrichissement du 5mC ou du 5hmC, soit au niveau de l'amplification des fragments non-spécifiques qui a augmenté le bruit de fond et qui a nui à la qualité du séquençage.

D'autre part, une étape de CQ a été ajoutée à notre procédure dans le but de déterminer la spécificité de l'enrichissement en comparant l'immunoprécipitation à son INPUT. Cette étape n'est pas recommandée par les études qui ont utilisé la même approche (201) puisqu'ils pensent que l'INPUT n'est pas nécessaire pour le meDIP et le hmeDIP car, contrairement au ChIP-seq, l'INPUT pour le meDIP est requis seulement dans le cas du séquençage du génome entier. Le Génome Science Center a ainsi indiqué que l'INPUT n'est pas utile aux analyses des méthodes meDIP et hmeDIP (203). Ainsi, l'utilisation de l'INPUT peut aussi nuire aux tests de CQ.

L'analyse différentielle de la méthylation a été réalisée sur les échantillons qui ont répondu aux critères de CQ. Dans les échantillons de 5mC séquencés, ceci représente 4 TET2 et un seul WT. L'analyse des DMR effectuée sur ces échantillons a identifié 30 DMR dans TET2 vs WT. Ce nombre est très faible par rapport aux études qui ont fait le même type d'analyse chez des souris *tet2 knock out* (204). Dans les échantillons de 5hmC séquencés, un seul TET2 et un seul WT répondaient aux critères de CQ. L'analyse de DHMR a été réalisée sur ceux-ci malgré le faible nombre d'échantillons. Bien que le nombre de DHMR

identifiés est très grand, aucun n'est statistiquement significatif. De plus, l'analyse *Gene Ontology* (GO) des gènes qui contiennent ces DMR et ces DHMR ne donne aucune voix de signalisation intéressante. Ainsi, nous sommes arrivés à la conclusion que les résultats de DMR et DHMR ne sont pas fiables pour deux principales raisons : (1) la faible qualité des échantillons, relevée par les tests de CQ, est probablement la cause qui a nui à l'identification d'un nombre élevé de DMR et de DHMR, (2) la fiabilité des tests statistiques a été grandement réduite par le faible nombre d'échantillons analysés. Il est également probable que l'effet de TET2 soit faible vu que le TET2 de l'autre allèle toujours fonctionnel puisse compenser la fonction du TET2 muté ce qui aura comme résultat de réduire son impact ou de le camoufler.

Dans le but d'améliorer l'approche utilisée, il faudrait (i) augmenter le nombre d'échantillons TET2 et WT via le recrutement d'un plus grand nombre d'individus, (ii) améliorer le processus d'enrichissement du 5mC ou du 5hmC en utilisant des anticorps spécifiques au 5mC ou au 5hmC de qualité, (iii) s'assurer que les étapes d'immunoprécipitation soient spécifiques par rapport à l'IgG utilisée comme contrôle, (iv) augmenter la quantité totale de fragments 5mC ou 5hmC immunoprécipités afin de réduire le bruit de fond causée par une amplification non-spécifique, et (v) augmenter le nombre de reads pour couvrir toutes les régions méthylées et hydroxylées immunoprécipitées. Pour réaliser ceci, il faut séquencer un seul individu par série de séquençage ce qui résulterait en une augmentation du nombre de reads pour assurer la couverture de toutes les régions immunoprécipités. Une autre approche serait d'utiliser des sujets qui ont une transformation maligne avancée puisque ceux-ci ont un phénotype qui est similaire à TET



*knock out* et permettrait ainsi d'identifier des DMR et des DHMR statistiquement significatifs. Finalement, une approche complètement différente pourrait être utilisée comme du *single cell* puisque c'est une méthode très pertinente où l'effet de TET2 est très faible.

#### **4.2.4 RNA-Seq et l'expression génique dans TET2 vs WT**

Le RNA-Seq a été utilisé pour séquencer des échantillons TET2 et WT dans le but d'identifier les gènes qui sont surexprimés ou sous-exprimés dans les individus TET2 versus WT. Le protocole utilisé pour le RNA-Seq est éprouvé par notre laboratoire et des analyses de CQ sur nos échantillons ont été réalisées. L'efficacité du séquençage est basée sur le nombre de reads. Ainsi, le seuil de 10 millions de reads a été établi et les quelques échantillons avec un nombre de reads inférieur à ce seuil ont été rejetés. Le rejet de ces échantillons était causé pas un ARN de départ de mauvaise qualité.

L'analyse effectuée sur les TET2 vs WT a démontré une légère modulation de l'expression des gènes mais aucun changement dans l'expression de gène n'a été considéré comme statistiquement significatif (Figure 24, page 97). Une raison qui pourrait expliquer ce résultat est la méthode de RNA-Seq utilisée mais celle-ci est fiable. De plus, le séquençage des échantillons s'est bien déroulé tel que confirmé par les tests de CQ. L'absence de changement significatif dans l'expression des gènes dans les TET2 vs WT pourrait également être expliquée par le faible nombre d'échantillons utilisés, c'est-à-dire seulement 4 TET2 et 5 WT, ce qui aurait un impact sur les tests statistiques appliqués. Par ailleurs, il est probable que l'effet de TET2 ne puisse pas être détectable dans le contexte

où la présence de l'allèle TET2 toujours fonctionnel vient compenser la fonction du TET2 muté et l'impact du TET2 muté sur l'expression génique serait donc négligeable. Dans le but de contrer ces problématiques, il faudrait (i) augmenter le nombre de nos échantillons TET2 et WT, et/ou (ii) utiliser des échantillons prélevés chez des sujets qui ont une transformation maligne avancée.

#### **4.2.5 Limitation de notre cohorte et effet faible de TET2**

Malgré le fait que nous avons validé la méthode de méthylome et de hydroxyméthylome, nous avons quand même rencontré des problèmes qui pourraient expliquer la raison pour laquelle nous n'avons pas trouvé des DMR et des DHMR statistiquement significatif. Bien que notre méthode de RNA-Seq soit fiable, nous n'avons pas pu identifier une modulation du gène dans TET2 vs WT. Ces résultats nous laissent à penser que c'est possible que notre cohorte n'a pas suffisamment de puissance car on n'a pas beaucoup d'échantillons de TET2 pour pouvoir identifier des gènes qui peuvent être surexprimés ou sous-exprimés dans les TET2 vs WT ou pour identifier des DMR et des DHMR. Dans ce cas nous devons recruter plus des individus âgés, extraire l'ADN et l'ARN de leurs cellules sanguines, séquencer l'ADN pour identifier ceux qui ont une mutation sur TET2 et séquencer l'ARN des individus qui ont une mutation sur TET2 pour augmenter le nombre TET2 muté et refaire l'analyse avec un plus grand nombre d'échantillons.

Un autre point très important est que les individus de notre cohorte qui possèdent une mutation sur TET2 ont une fréquence de VAF entre 20% et 40%, ce qui signifie que 40 à 80% de leurs cellules possèdent une mutation sur TET2 et que les autres cellules sont WT.

En plus, la présence chez ces individus d'un allèle muté signifie conséquemment que l'autre allèle porte le gène TET2 WT qui est toujours fonctionnel et qui peut ainsi compenser pour l'absence de TET2. En somme, la présence de cellules WT et d'un allèle WT dans les cellules mutées ont comme répercussion probable de compenser l'effet de la mutation de TET2 sur l'expression des gènes. Ceci explique probablement pourquoi il n'a pas été possible d'observer une différence statistiquement significative entre des gènes exprimés chez les individus de la cohorte TET2 versus ceux de la cohorte WT. De plus, même un individu avec 80% de cellules possédant un TET2 muté (VAF à 40%) n'est pas équivalent à un individu TET2 *knock out* et l'effet de TET2 dans ces deux types d'individus n'est pas le même. Ainsi, l'effet de la mutation TET2 dans les TET2 mutés est faible et peut expliquer la légère modulation de l'expression génique non significative observée grâce au RNA-Seq (Figure 24, page 97). En fait, il est probable que l'impact de la mutation TET2 puisse seulement être observable, avec les approches que nous avons utilisées, qu'en travaillant avec des individus qui ont eu déjà une transformation maligne ou un stade avancé de cancer myéloïde. Des changements importants de méthylation et d'hydroxyméthylation qui ont un impact sur l'expression génique ne peuvent se produire qu'avec la transformation maligne. En reprenant les travaux réalisés avec une cohorte de ce type d'individus, il serait donc envisageable d'observer une différence statistiquement significative entre des gènes exprimés puisque l'effet de TET2 muté serait plus important et permettrait d'identifier des DMR et des DHMR ainsi que des gènes dont leur expression est modulée par TET2.

## **CHAPITRE 5**

### **CONCLUSION ET DIRECTIONS FUTURES**

## 5. CONCLUSION

La variation du ratio du biais de l'ICX selon l'âge est un phénomène biologique observé dans les cellules sanguines d'une proportion significative de femmes vieillissantes en bonne santé. Une compréhension approfondie des principes de base de ce phénomène permet de mieux comprendre la biologie de l'hématopoïèse vieillissante étant donné que l'âge est un facteur de risque majeur dans le développement de cancers hématologiques. L'accès à une cohorte exceptionnelle qui contient plus que 3000 individus apparentés et non apparentés permet de réaliser les travaux requis pour des études sur le biais de l'ICX. Cette cohorte a déjà été validée puisqu'elle a permis à notre laboratoire de réaliser plusieurs projets importants publiés dans des journaux scientifiques dans le but de comprendre l'impact du vieillissement chez les cellules sanguines mais aussi d'élucider le lien entre l'inflammation, les mutations somatiques et les maladies cardiovasculaires (ANNEXE 3, page 176) (121).

Les travaux présentés dans cette thèse démontrent que le biais de l'ICX est spécifique et distinctif aux tissus sanguins ce qui suggère que les motifs de l'ICX pourraient être partiellement contrôlés par des mécanismes de régulation spécifiques à ces tissus. Nos travaux ont aussi démontré que le biais de l'ICX est moins important dans les cellules T par rapport aux PMN ce qui suggère que ce phénomène pourrait également être spécifique aux cellules myéloïdes. Ces résultats sont d'autant plus intéressants que le biais de l'ICX augmente selon l'âge des individus.

Nos travaux mettent aussi en évidence la présence d'une corrélation entre le biais de l'ICX et les mutations CHIP ce qui suggère fortement un lien possible entre le biais de l'ICX et l'HC. Ces résultats renforcent l'idée et valide notre hypothèse que le phénomène de biais de l'ICX peut être causé par l'acquisition de mutations donnant un avantage de croissance dans les cellules somatiques de la moelle osseuse, ce qui est un mécanisme distinct des causes génétiquement prédéterminées que l'on retrouve dans le processus de la sélection hémizygote. Ainsi, nos travaux permettent de faire la distinction entre deux causes, la sélection hémizygote et l'acquisition de mutations qui donnent le même phénotype de l'ICX (Figure 8, page 38).

Par ailleurs, nos données démontrent l'héritabilité du biais de l'ICX dans les cellules sanguines chez les femmes vieillissantes ce qui confirme le rôle des variants génétiques en tant que régulateurs du biais de l'ICX qui varient selon l'âge. Aussi, ces résultats suggèrent fortement l'existence d'un locus liée au chromosome X prédisposant le biais de l'ICX. En effet, les travaux présentés dans cette thèse révèlent pour la première fois qu'il y a des régions sur l'ADN, précisément sur le chromosome 1 et le chromosome X, qui pourraient être associées au trait du biais de l'ICX. Ces régions pourraient être associées à un gène ou un sous-ensemble de gènes qui pourrait jouer un rôle dans l'initiation du biais du chromosome X et être associés à des gènes avec un polymorphisme associé à la sélection hémizygote.

## 5.1 Directions futures

La prochaine étape consisterait donc à examiner les gènes associés à ces régions dans chaque chromosome dans le but d'identifier spécifiquement ces gènes d'intérêt. Ceci pourrait être réalisé en effectuant du *fine mapping* sur les gènes *IL2RG* and *IL6R* afin d'identifier précisément le(s) gène(s) qui est/sont muté(s). Par la suite, l'identification de certaines régions sur le chromosome X et le chromosome 1, qui sont associées à des gènes jouant un rôle dans l'inflammation, permet de formuler l'hypothèse que l'inflammation puisse être impliquée dans le déclenchement du biais de l'ICX. Cette hypothèse avait déjà été soulevée par de nombreuses études qui soutiennent que l'inflammation favorise l'expansion clonale initiale des CSH, accélère la transformation maligne et le développement de la leucémie (121) (142). Ainsi, le rôle de l'inflammation sur le biais de l'ICX pourrait être une nouvelle étiologie du biais de l'ICX.

D'autre part, nos travaux se poursuivront dans le but d'identifier les cibles génétiques qui confèrent un potentiel de transformation maligne en utilisant une approche basée sur l'analyse du méthylome, de l'hydroxyméthylome et du transcriptome que nous venons de valider.

En conclusion, nos travaux démontrent la complexité de l'adaptation de l'hématopoïèse au vieillissement et ouvre des portes sur l'identification de facteurs prédisposant aux cancers hématologiques et permettent d'apporter une petite pierre à l'édifice colossal qu'il nous reste à bâtir dans les prochaines années pour approfondir nos connaissances sur l'hématopoïèse clonale.

## RÉFÉRENCES

1. Ayachi S, Buscarlet M, Busque L. 60 Years of clonal hematopoiesis research: From X-chromosome inactivation studies to the identification of driver mutations. *Experimental hematology*. 2020.
2. Fialkow PJ. The origin and development of human tumors studied with cell markers. *The New England journal of medicine*. 1974;291(1):26-35.
3. Bowman RL, Busque L, Levine RL. Clonal Hematopoiesis and Evolution to Hematopoietic Malignancies. *Cell Stem Cell*. 2018;22(2):157-70.
4. Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nature genetics*. 2012;44(11):1179-81.
5. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine*. 2014;371(26):2477-87.
6. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine*. 2014;371(26):2488-98.
7. Swenberg JA, Lu K, Moeller BC, Gao L, Upton PB, Nakamura J, et al. Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol Sci*. 2011;120 Suppl 1(Suppl 1):S130-S45.
8. Pfeifer GP. Environmental exposures and mutational patterns of cancer genomes. *Genome medicine*. 2010;2(8):54.
9. McNerney ME, Godley LA, Le Beau MM. Therapy-related myeloid neoplasms: when genetics and environment collide. *Nature reviews Cancer*. 2017;17(9):513-27.
10. Tucker MA, Coleman CN, Cox RS, Varghese A, Rosenberg SA. Risk of second cancers after treatment for Hodgkin's disease. *The New England journal of medicine*. 1988;318(2):76-81.
11. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458(7239):719-24.
12. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-21.
13. Catlin SN, Busque L, Gale RE, Guttorp P, Abkowitz JL. The replication rate of human hematopoietic stem cells in vivo. *Blood*. 2011;117(17):4460-6.
14. Ohno S, Hauschka TS. Allocycly of the X-Chromosome in Tumors and Normal Tissues. *Cancer research*. 1960;20(4):541-5.
15. Lyon MF. Sex chromatin and gene action in the mammalian X-chromosome. *American journal of human genetics*. 1962;14:135-48.



16. Beutler E, Yeh M, Fairbanks VF. The normal human female as a mosaic of X-chromosome activity: studies using the gene for C-6-PD-deficiency as a marker. *Proceedings of the National Academy of Sciences of the United States of America*. 1962;48(1):9-16.
17. Brockdorff N, Turner BM. Dosage compensation in mammals. *Cold Spring Harbor perspectives in biology*. 2015;7(3):a019406.
18. Kratzer PG, Chapman VM. X chromosome reactivation in oocytes of *Mus caroli*. *Proceedings of the National Academy of Sciences*. 1981;78(5):3093.
19. Lifschytz E, Lindsley DI. Sex chromosome activation during spermatogenesis. *Genetics*. 1974;78(1):323-31.
20. Starmer J, Magnuson T. A new model for random X chromosome inactivation. *Development (Cambridge, England)*. 2009;136(1):1.
21. Ercan S. Mechanisms of x chromosome dosage compensation. *Journal of genomics*. 2015;3:1-19.
22. Takagi N, Sasaki M. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*. 1975;256(5519):640-2.
23. Cheng MK, Distèche CM. Silence of the fathers: early X inactivation. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2004;26(8):821-4.
24. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science (New York, NY)*. 2004;303(5658):644-9.
25. van den Berg IM, Laven JS, Stevens M, Jonkers I, Galjaard RJ, Gribnau J, et al. X chromosome inactivation is initiated in human preimplantation embryos. *American journal of human genetics*. 2009;84(6):771-9.
26. Graves JAM. MAMMALS THAT BREAK THE RULES: Genetics of Marsupials and Monotremes. *Annual review of genetics*. 1996;30(1):233-60.
27. Takagi N, Sasaki M. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*. 1975;256(5519):640-2.
28. Wang M, Lin F, Xing K, Liu L. Random X-chromosome inactivation dynamics in vivo by single-cell RNA sequencing. *BMC Genomics*. 2017;18(1):90.
29. Zito A, Davies MN, Tsai P-C, Roberts S, Andres-Ejarque R, Nardone S, et al. Heritability of skewed X-inactivation in female twins is tissue-specific and associated with age. *Nature communications*. 2019;10(1):5339.
30. Morey C, Avner P. Genetics and epigenetics of the X chromosome. *Annals of the New York Academy of Sciences*. 2010;1214:E18-33.
31. Fialkow PJ, Martin GM, Klein G, Clifford P, Singh S. Evidence for a clonal origin of head and neck tumors. *International journal of cancer*. 1972;9(1):133-42.

32. Linder D, Gartler SM. Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. *Science*. 1965;150(3692):67-9.
33. Boyer SH, Porter IH, Weilbacher RG. Electrophoretic heterogeneity of glucose-6-phosphate dehydrogenase and its relationship to enzyme deficiency in man. *Proceedings of the National Academy of Sciences of the United States of America*. 1962;48(10):1868-76.
34. Chen GL, Prchal JT. X-linked clonality testing: interpretation and limitations. *Blood*. 2007;110(5):1411-9.
35. Busque L, Gilliland DG. X-inactivation analysis in the 1990s: promise and potential problems. *Leukemia*. 1998;12(2):128-35.
36. Vogelstein B, Fearon ER, Hamilton SR, Feinberg AP. Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. *Science*. 1985;227(4687):642-5.
37. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*. 1980;32(3):314-31.
38. Mullis KB, Faloona FA. [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*. 155: Academic Press; 1987. p. 335-50.
39. Gilliland DG, Blanchard KL, Levy J, Perrin S, Bunn HF. Clonality in myeloproliferative disorders: analysis by means of the polymerase chain reaction. *Proceedings of the National Academy of Sciences*. 1991;88(15):6848-52.
40. Fraser NJ, Boyd Y, Brownlee GG, Craig IW. Multi-allelic RFLP for M27 beta, an anonymous single copy genomic clone at Xp11.3-Xcen [HGM9 provisional no. DXS255]. *Nucleic acids research*. 1987;15(22):9616-.
41. Fraser NJ, Boyd Y, Craig I. Isolation and characterization of a human variable copy number tandem repeat at Xcen-p11.22. *Genomics*. 1989;5(1):144-8.
42. Abrahamson G, Fraser NJ, Boyd J, Craig I, Wainscoat JS. A highly informative X-chromosome probe, M27 beta, can be used for the determination of tumour clonality. *British journal of haematology*. 1990;74(3):371-2.
43. Hodges E, Howell WM, Boyd Y, Smith JL. Variable X-chromosome DNA methylation patterns detected with probe M27 beta in a series of lymphoid and myeloid malignancies. *British journal of haematology*. 1991;77(3):315-22.
44. Vogelstein B, Fearon ER, Hamilton SR, Preisinger AC, Willard HF, Michelson AM, et al. Clonal analysis using recombinant DNA probes from the X-chromosome. *Cancer research*. 1987;47(18):4806-13.
45. Swierczek SI, Agarwal N, Nussenzveig RH, Rothstein G, Wilson A, Artz A, et al. Hematopoiesis is not clonal in healthy elderly women. *Blood*. 2008;112(8):3186-93.
46. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-

receptor gene correlates with X chromosome inactivation. *American journal of human genetics*. 1992;51(6):1229-39.

47. Comertpay S, Pastorino S, Tanji M, Mezzapelle R, Strianese O, Napolitano A, et al. Evaluation of clonal origin of malignant mesothelioma. *J Transl Med*. 2014;12:301.

48. Busque L, Zhu J, DeHart D, Griffith B, Willman C, Carroll R, et al. An expression based clonality assay at the human androgen receptor locus (HUMARA) on chromosome X. *Nucleic acids research*. 1994;22(4):697-8.

49. Mio R. Genetic-epidemiologic analysis of X-inactivation skewing in human females : suggestive evidence for two distinct traits. Montréal, Canada: Université de Montréal; 2003.

50. Rupert JL, Brown CJ, Willard HF. Direct Detection of Non-Random X Chromosome Inactivation by Use of a Transcribed Polymorphism in the XIST Gene. *European Journal of Human Genetics*. 1995;3(6):333-43.

51. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, et al. The DNA sequence of the human X chromosome. *Nature*. 2005;434(7031):325-37.

52. Helena Mangs A, Morris BJ. The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Curr Genomics*. 2007;8(2):129-36.

53. Gribnau J, Stefan Barakat T. X-chromosome inactivation and its implications for human disease. *bioRxiv*. 2017:076950.

54. Busque L, Mio R, Mattioli J, Brais E, Blais N, Lalonde Y, et al. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood*. 1996;88(1):59-65.

55. Allen RC, Nachtman RG, Rosenblatt HM, Belmont JW. Application of carrier testing to genetic counseling for X-linked agammaglobulinemia. *American journal of human genetics*. 1994;54(1):25-35.

56. Gale RE, Wheadon H, Linch DC. X-chromosome inactivation patterns using HPRT and PGK polymorphisms in haematologically normal and post-chemotherapy females. *British journal of haematology*. 1991;79(2):193-7.

57. Teklenburg G, Weimar CH, Fauser BC, Macklon N, Geijsen N, Heijnen CJ, et al. Cell lineage specific distribution of H3K27 trimethylation accumulation in an in vitro model for human implantation. *PloS one*. 2012;7(3):e32701.

58. Renault NK, Pritchett SM, Howell RE, Greer WL, Sapienza C, Ørstavik KH, et al. Human X-chromosome inactivation pattern distributions fit a model of genetically influenced choice better than models of completely random choice. *European journal of human genetics : EJHG*. 2013;21(12):1396-402.

59. Cattanaach BM, Isaacson JH. Controlling elements in the mouse X chromosome. *Genetics*. 1967;57(2):331-46.

60. Naumova A, Olien L, Bird L, Smith M, Verner A, Leppert M, et al. Genetic mapping of X-linked loci involved in skewing of X chromosome inactivation in the human. *European journal of human genetics : EJHG*. 1998;6:552-62.

61. Plenge RM, Hendrich BD, Schwartz C, Arena JF, Naumova A, Sapienza C, et al. A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. *Nature genetics*. 1997;17(3):353-6.
62. Wang Z, Yan A, Lin Y, Xie H, Zhou C, Lan F. Familial skewed x chromosome inactivation in adrenoleukodystrophy manifesting heterozygotes from a Chinese pedigree. *PloS one*. 2013;8(3):e57977.
63. Ruttum MS, Lewandowski MF, Bateman JB. Affected females in X-linked congenital stationary night blindness. *Ophthalmology*. 1992;99(5):747-52.
64. Lau AW, Brown CJ, Peñaherrera M, Langlois S, Kalousek DK, Robinson WP. Skewed X-chromosome inactivation is common in fetuses or newborns associated with confined placental mosaicism. *American journal of human genetics*. 1997;61(6):1353-61.
65. Fey MF, Peter HJ, Hinds HL, Zimmermann A, Liechti-Gallati S, Gerber H, et al. Clonal analysis of human tumors with M27 beta, a highly informative polymorphic X chromosomal probe. *The Journal of clinical investigation*. 1992;89(5):1438-44.
66. Gale R, Wheadon H, Boulos P, Linch D. Tissue specificity of X-chromosome inactivation patterns. *Blood*. 1994;83(10):2899-905.
67. Fey MF, Liechti-Gallati S, von Rohr A, Borisch B, Theilkäs L, Schneider V, et al. Clonality and X-inactivation patterns in hematopoietic cell populations detected by the highly informative M27 beta DNA probe. *Blood*. 1994;83(4):931-8.
68. Mossner M, Nolte F, Hütter G, Reins J, Klaumünzer M, Nowak V, et al. Skewed X-inactivation patterns in ageing healthy and myelodysplastic haematopoiesis determined by a pyrosequencing based transcriptional clonality assay. *Journal of Medical Genetics*. 2013;50(2):108-17.
69. Sharp A, Robinson D, Jacobs P. Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. *Human Genetics*. 2000;107(4):343-9.
70. Abkowitz JL, Taboada M, Shelton GH, Catlin SN, Gutterop P, Kiklevich JV. An X chromosome gene regulates hematopoietic stem cell kinetics. *Proceedings of the National Academy of Sciences*. 1998;95(7):3862-6.
71. Christensen K, Kringen M, Hagen-Larsen H, Skytthe A, Bathum L, Jeune B, et al. X-linked genetic factors regulate hematopoietic stem-cell kinetics in females. *Blood*. 2000;95:2449-51.
72. Vickers MA, McLeod E, Spector TD, Wilson IJ. Assessment of mechanism of acquired skewed X inactivation by analysis of twins. *Blood*. 2001;97(5):1274-81.
73. Prchal J, Carroll A, Prchal J, Crist W, Skalka H, Gealy W, et al. Wiskott-Aldrich syndrome: cellular impairments and their implication for carrier detection. *Blood*. 1980;56(6):1048-54.
74. Conley ME, Brown P, Pickard AR, Buckley RH, Miller DS, Raskind WH, et al. Expression of the Gene Defect in X-Linked Agammaglobulinemia. *New England Journal of Medicine*. 1986;315(9):564-7.

75. Till JE, McCulloch EA. Hemopoietic stem cell differentiation. *Biochimica et biophysica acta*. 1980;605(4):431-59.
76. Gale RE, Fielding AK, Harrison CN, Linch DC. Acquired skewing of X-chromosome inactivation patterns in myeloid cells of the elderly suggests stochastic clonal loss with age. *British journal of haematology*. 1997;98(3):512-9.
77. Rosendaal M, Hodgson GS, Bradley TR. Organization of haemopoietic stem cells: the generation-age hypothesis. *Cell and tissue kinetics*. 1979;12(1):17-29.
78. Till JE, McCulloch EA, Siminovitch L. A STOCHASTIC MODEL OF STEM CELL PROLIFERATION, BASED ON THE GROWTH OF SPLEEN COLONY-FORMING CELLS. *Proceedings of the National Academy of Sciences of the United States of America*. 1964;51(1):29-36.
79. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature genetics*. 2012;44(6):642-50.
80. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nature genetics*. 2012;44(6):651-8.
81. Forsberg Lars A, Rasi C, Razzaghian Hamid R, Pakalapati G, Waite L, Thilbeault Krista S, et al. Age-Related Somatic Structural Changes in the Nuclear Genome of Human Blood Cells. *The American Journal of Human Genetics*. 2012;90(2):217-28.
82. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature medicine*. 2014;20(12):1472-8.
83. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine*. 2014;371(26):2488-98.
84. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoun SF, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine*. 2014;371(26):2477-87.
85. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *The New England journal of medicine*. 2017;377(2):111-21.
86. Shlush LI. Age-related clonal hematopoiesis. *Blood*. 2018;131(5):496-504.
87. Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9-16.
88. Buscarlet M, Provost S, Zada YF, Barhdadi A, Bourgoin V, Lépine G, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood*. 2017;130(6):753-62.

89. Buscarlet M, Provost S, Zada YF, Bourgoïn V, Mollica L, Dubé M-P, et al. Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. *Blood*. 2018;132(3):277-80.
90. Young AL, Challen GA, Birman BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nature communications*. 2016;7(1):12484.
91. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nature Reviews Genetics*. 2013;14(3):204-20.
92. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science (New York, NY)*. 2009;324(5929):930-5.
93. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell*. 2010;18(6):553-67.
94. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *New England Journal of Medicine*. 2016;374(23):2209-21.
95. Buscarlet M, Provost S, Zada YF, Barhdadi A, Bourgoïn V, Lepine G, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood*. 2017;130(6):753-62.
96. Tadokoro Y, Ema H, Okano M, Li E, Nakauchi H. De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *Journal of Experimental Medicine*. 2007;204(4):715-22.
97. Challen Grant A, Sun D, Mayle A, Jeong M, Luo M, Rodriguez B, et al. Dnmt3a and Dnmt3b Have Overlapping and Distinct Functions in Hematopoietic Stem Cells. *Cell stem cell*. 2014;15(3):350-64.
98. Spencer DH, Russler-Germain DA, Ketkar S, Helton NM, Lamprecht TL, Fulton RS, et al. CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell*. 2017;168(5):801-16.e13.
99. Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell*. 2011;20(1):11-24.
100. Ge L, Zhang R-p, Wan F, Guo D-y, Wang P, Xiang L-x, et al. TET2 plays an essential role in erythropoiesis by regulating lineage-specific genes via DNA oxidative demethylation in a zebrafish model. *Molecular and cellular biology*. 2014;34(6):989-1002.
101. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*. 2016;17(8):487-500.
102. Avgustinova A, Benitah SA. Epigenetic control of adult stem cell function. *Nature Reviews Molecular Cell Biology*. 2016;17(10):643-58.

103. Fisher CL, Berger J, Randazzo F, Brock HW. A human homolog of Additional sex combs, ADDITIONAL SEX COMBS-LIKE 1, maps to chromosome 20q11. *Gene*. 2003;306:115-26.
104. Schwartz YB, Pirrotta V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nature Reviews Genetics*. 2013;14(12):853-64.
105. Jones AV, Chase A, Silver RT, Oscier D, Zoi K, Wang YL, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nature genetics*. 2009;41(4):446-9.
106. Kilpivaara O, Mukherjee S, Schram AM, Wadleigh M, Mullally A, Ebert BL, et al. A germline JAK2 SNP is associated with predisposition to the development of JAK2V617F-positive myeloproliferative neoplasms. *Nature genetics*. 2009;41(4):455-9.
107. Tashi T, Swierczek S, Prchal JT. Familial MPN Predisposition. *Current Hematologic Malignancy Reports*. 2017;12(5):442-7.
108. Forsberg LA, Rasi C, Malmqvist N, Davies H, Pasupulati S, Pakalapati G, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nature genetics*. 2014;46(6):624-8.
109. Haitjema S, Kofink D, van Setten J, van der Laan SW, Schoneveld AH, Eales J, et al. Loss of Y Chromosome in Blood Is Associated With Major Cardiovascular Events During Follow-Up in Men After Carotid Endarterectomy. *Circulation Cardiovascular genetics*. 2017;10(4):e001544.
110. Vas V, Senger K, Dörr K, Niebel A, Geiger H. Aging of the microenvironment influences clonality in hematopoiesis. *PloS one*. 2012;7(8):e42080.
111. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *New England Journal of Medicine*. 2017;377(2):111-21.
112. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*. 2017;130(6):742-52.
113. Franceschi C, Garagnani P, Vitale G, Capri M, Salvioli S. Inflammaging and 'Garb-aging'. *Trends in Endocrinology & Metabolism*. 2017;28(3):199-212.
114. Zhang B, Chu S, Agarwal P, Campbell VL, Hopcroft L, Jørgensen HG, et al. Inhibition of interleukin-1 signaling enhances elimination of tyrosine kinase inhibitor-treated CML stem cells. *Blood*. 2016;128(23):2671-82.
115. Abegunde SO, Buckstein R, Wells RA, Rauh MJ. An inflammatory environment containing TNF $\alpha$  favors Tet2-mutant clonal hematopoiesis. *Experimental hematology*. 2018;59:60-5.
116. Zhang Q, Zhao K, Shen Q, Han Y, Gu Y, Li X, et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature*. 2015;525(7569):389-93.

117. Fuster JJ, MacLauchlan S, Zuriaga MA, Polackal MN, Ostriker AC, Chakraborty R, et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science (New York, NY)*. 2017;355(6327):842.
118. Meisel M, Hinterleitner R, Pacis A, Chen L, Earley ZM, Mayassi T, et al. Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. *Nature*. 2018;557(7706):580-4.
119. Sallman DA, List A. The central role of inflammatory signaling in the pathogenesis of myelodysplastic syndromes. *Blood*. 2019;133(10):1039-48.
120. Cook EK, Izukawa T, Young S, Rosen G, Jamali M, Zhang L, et al. Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. *Blood advances*. 2019;3(16):2482-6.
121. Busque L, Sun M, Buscarlet M, Ayachi S, Feroz Zada Y, Provost S, et al. High-sensitivity C-reactive protein is associated with clonal hematopoiesis of indeterminate potential. *Blood advances*. 2020;4(11):2430-8.
122. Ayachi S, Buscarlet M, Busque L. 60 Years of clonal hematopoiesis research: From X-chromosome inactivation studies to the identification of driver mutations. *Experimental hematology*. 2020;83:2-11.
123. Busque L, Buscarlet M, Mollica L, Levine RL. Concise Review: Age-Related Clonal Hematopoiesis: Stem Cells Tempting the Devil. *Stem Cells*. 2018;36(9):1287-94.
124. Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS, et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature*. 2015;518(7540):552-5.
125. Cooper JN, Young NS. Clonality in context: hematopoietic clones in their marrow environment. *Blood*. 2017;130(22):2363-72.
126. Yoshizato T, Dumitriu B, Hosokawa K, Makishima H, Yoshida K, Townsley D, et al. Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. *New England Journal of Medicine*. 2015;373(1):35-47.
127. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine*. 2014;371(26):2477-87.
128. Nowell PC. The clonal evolution of tumor cell populations. *Science (New York, NY)*. 1976;194(4260):23-8.
129. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature*. 2001;414(6859):105-11.
130. Mesa RA, Li CY, Ketterling RP, Schroeder GS, Knudson RA, Tefferi A. Leukemic transformation in myelofibrosis with myeloid metaplasia: a single-institution experience with 91 cases. *Blood*. 2005;105(3):973-7.
131. Jamieson CH, Gotlib J, Durocher JA, Chao MP, Mariappan MR, Lay M, et al. The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and



- predisposes toward erythroid differentiation. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(16):6224-9.
132. Jaiswal S, Ebert BL. MDS is a stem cell disorder after all. *Cancer Cell*. 2014;25(6):713-4.
133. Potter NE, Greaves M. Cancer: Persistence of leukaemic ancestors. *Nature*. 2014;506(7488):300-1.
134. Vasanthakumar A, Godley LA. On the origin of leukemic species. *Cell stem cell*. 2014;14(4):421-2.
135. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012;481(7382):506-10.
136. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*. 2014;506(7488):328-33.
137. Corces-Zimmerman MR, Hong WJ, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(7):2548-53.
138. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458(7239):719-24.
139. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306-13.
140. Adolphson CC, Prchal JT, Carroll AJ. Colonic Villous Adenoma: Monoclonal Origin. *JAMA*. 1982;247(6):829-30.
141. Adamson JW, Fialkow PJ, Murphy S, Prchal JF, Steinmann L. Polycythemia vera: stem-cell and probable clonal origin of the disease. *The New England journal of medicine*. 1976;295(17):913-6.
142. Prchal JT, Throckmorton DW, Carroll AJ, Fuson EW, Gams RA, Prchal JF. A common progenitor for human myeloid and lymphoid cells. *Nature*. 1978;274(5671):590-1.
143. Fey MF, Liechti-Gallati S, von Rohr A, Borisch B, Theilkas L, Schneider V, et al. Clonality and X-inactivation patterns in hematopoietic cell populations detected by the highly informative M27 beta DNA probe. *Blood*. 1994;83(4):931-8.
144. Kristiansen M, Knudsen GP, Bathum L, Naumova AK, Sorensen TI, Brix TH, et al. Twin study of genetic and aging effects on X chromosome inactivation. *Eur J Hum Genet*. 2005;13(5):599-606.
145. Gale RE, Fielding AK, Harrison CN, Linch DC. Acquired skewing of X-chromosome inactivation patterns in myeloid cells of the elderly suggests stochastic clonal loss with age. *British journal of haematology*. 1997;98(3):512-9.

146. Sandovici I, Naumova AK, Leppert M, Linares Y, Sapienza C. A longitudinal study of X-inactivation ratio in human females. *Hum Genet.* 2004;115(5):387-92.
147. Abkowitz JL, Catlin SN, Gutterop P. Evidence that hematopoiesis may be a stochastic process in vivo. *Nature medicine.* 1996;2(2):190-7.
148. Christensen K, Kristiansen M, Hagen-Larsen H, Skytthe A, Bathum L, Jeune B, et al. X-linked genetic factors regulate hematopoietic stem-cell kinetics in females. *Blood.* 2000;95(7):2449-51.
149. Zito A, Davies MN, Tsai PC, Roberts S, Andres-Ejarque R, Nardone S, et al. Heritability of skewed X-inactivation in female twins is tissue-specific and associated with age. *Nature communications.* 2019;10(1):5339.
150. Abkowitz JL, Taboada M, Shelton GH, Catlin SN, Gutterop P, Kiklevich JV. An X chromosome gene regulates hematopoietic stem cell kinetics. *Proceedings of the National Academy of Sciences of the United States of America.* 1998;95(7):3862-6.
151. Bolduc V, Chagnon P, Provost S, Dubé MP, Belisle C, Gingras M, et al. No evidence that skewing of X chromosome inactivation patterns is transmitted to offspring in humans. *The Journal of clinical investigation.* 2008;118(1):333-41.
152. Tonon L, Bergamaschi G, Dellavecchia C, Rosti V, Lucotti C, Malabarba L, et al. Unbalanced X-chromosome inactivation in haemopoietic cells from normal women. *British journal of haematology.* 1998;102(4):996-1003.
153. Sharp A, Robinson D, Jacobs P. Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. *Human genetics.* 2000;107(4):343-9.
154. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.
155. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda).* 2011;1(6):457-70.
156. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics.* 2012;44(8):955-9.
157. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature genetics.* 2018;50(9):1335-41.
158. Scriver CR. Human genetics: lessons from Quebec populations. *Annual review of genomics and human genetics.* 2001;2:69-101.
159. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics.* 2002;30(1):97-101.
160. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-10.

161. Busque L, Paquette Y, Provost S, Roy DC, Levine RL, Mollica L, et al. Skewing of X-inactivation ratios in blood cells of aging women is confirmed by independent methodologies. *Blood*. 2009;113(15):3472-4.
162. Mossner M, Nolte F, Hütter G, Reins J, Klaumünzer M, Nowak V, et al. Skewed X-inactivation patterns in ageing healthy and myelodysplastic haematopoiesis determined by a pyrosequencing based transcriptional clonality assay. *J Med Genet*. 2013;50(2):108-17.
163. Agarwal N, Nussenzveig RH, Swierczek SI, Parker CJ, Prchal JT. Does HUMARA assay for assessment of clonal hematopoiesis have shortcomings? *Blood*. 2009;114(11):2357-8; author reply 8-9.
164. Swierczek SI, Piterkova L, Jelinek J, Agarwal N, Hammoud S, Wilson A, et al. Methylation of AR locus does not always reflect X chromosome inactivation state. *Blood*. 2012;119(13):e100-9.
165. Gale RE, Wheadon H, Boulos P, Linch DC. Tissue specificity of X-chromosome inactivation patterns. *Blood*. 1994;83(10):2899-905.
166. Johnson PL, Yates AJ, Goronzy JJ, Antia R. Peripheral selection rather than thymic involution explains sudden contraction in naive CD4 T-cell diversity with age. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(52):21432-7.
167. Prchal JT, Guan YL, Prchal JF, Barany F. Transcriptional analysis of the active X-chromosome in normal and clonal hematopoiesis. *Blood*. 1993;81(1):269-71.
168. Prchal JT, Prchal JF, Belickova M, Chen S, Guan Y, Gartland GL, et al. Clonal stability of blood cell lineages indicated by X-chromosomal transcriptional polymorphism. *J Exp Med*. 1996;183(2):561-7.
169. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature genetics*. 2012;44(6):642-50.
170. Sangha KK, Stephenson MD, Brown CJ, Robinson WP. Extremely skewed X-chromosome inactivation is increased in women with recurrent spontaneous abortion. *American journal of human genetics*. 1999;65(3):913-7.
171. Lanasa MC, Hogge WA, Kubik CJ, Ness RB, Harger J, Nagel T, et al. A novel X chromosome-linked genetic cause of recurrent spontaneous abortion. *Am J Obstet Gynecol*. 2001;185(3):563-8.
172. Uehara S, Hashiyada M, Sato K, Sato Y, Fujimori K, Okamura K. Preferential X-chromosome inactivation in women with idiopathic recurrent pregnancy loss. *Fertil Steril*. 2001;76(5):908-14.
173. Asao H, Okuyama C, Kumaki S, Ishii N, Tsuchiya S, Foster D, et al. Cutting edge: the common gamma-chain is an indispensable subunit of the IL-21 receptor complex. *J Immunol*. 2001;167(1):1-5.

174. Noguchi M, Nakamura Y, Russell SM, Ziegler SF, Tsang M, Cao X, et al. Interleukin-2 receptor gamma chain: a functional component of the interleukin-7 receptor. *Science (New York, NY)*. 1993;262(5141):1877-80.
175. Goh JY, Feng M, Wang W, Oguz G, Yatim S, Lee PL, et al. Chromosome 1q21.3 amplification is a trackable biomarker and actionable target for breast cancer recurrence. *Nature medicine*. 2017;23(11):1319-30.
176. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-7.
177. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature medicine*. 2014;20(12):1472-8.
178. Ko M, An J, Pastor WA, Koralov SB, Rajewsky K, Rao A. TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunological reviews*. 2015;263(1):6-21.
179. Solary E, Bernard OA, Tefferi A, Fuks F, Vainchenker W. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia*. 2014;28(3):485-96.
180. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Molecular cell*. 2011;42(4):451-64.
181. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes & development*. 2016;30(7):733-50.
182. Xu Y-P, Lv L, Liu Y, Smith MD, Li W-C, Tan X-M, et al. Tumor suppressor TET2 promotes cancer immunity and immunotherapy efficacy. *The Journal of clinical investigation*. 2019;129(10):4316-31.
183. Buscarlet M, Tessier A, Provost S, Mollica L, Busque L. Human blood cell levels of 5-hydroxymethylcytosine (5hmC) decline with age, partly related to acquired mutations in TET2. *Experimental hematology*. 2016;44(11):1072-84.
184. Bird A. DNA methylation patterns and epigenetic memory. *Genes & development*. 2002;16(1):6-21.
185. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-74.
186. Lienhard M, Grasse S, Rolff J, Frese S, Schirmer U, Becker M, et al. QSEA—modelling of genome-wide DNA methylation from sequencing enrichment experiments. *Nucleic acids research*. 2016;45(6):e44-e.
187. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics*. 2013;30(2):284-6.

188. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*. 2008;9(9):R137.
189. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550.
190. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research*. 2012;40(10):4288-97.
191. Hatakeyama C, Anderson CL, Beaver CL, Peñaherrera MS, Brown CJ, Robinson WP. The dynamics of X-inactivation skewing as women age. *Clinical genetics*. 2004;66(4):327-32.
192. Knudsen GP, Pedersen J, Klingenberg O, Lygren I, Ørstavik KH. Increased skewing of X chromosome inactivation with age in both blood and buccal cells. *Cytogenetic and genome research*. 2007;116(1-2):24-8.
193. Savola P, Lundgren S, Keränen MAI, Almusa H, Ellonen P, Leirisalo-Repo M, et al. Clonal hematopoiesis in patients with rheumatoid arthritis. *Blood Cancer J*. 2018;8(8):69.
194. Kristiansen M, Knudsen GP, Bathum L, Naumova AK, Sørensen TI, Brix TH, et al. Twin study of genetic and aging effects on X chromosome inactivation. *European journal of human genetics : EJHG*. 2005;13(5):599-606.
195. Dwyer CJ, Knochelmann HM, Smith AS, Wyatt MM, Rangel Rivera GO, Arhontoulis DC, et al. Fueling Cancer Immunotherapy With Common Gamma Chain Cytokines. *Front Immunol*. 2019;10:263-.
196. West NR. Coordination of Immune-Stroma Crosstalk by IL-6 Family Cytokines. *Front Immunol*. 2019;10(1093).
197. Bick AG, Weinstock JS, Nandakumar SK, Fulco CP, Bao EL, Zekavat SM, et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature*. 2020;586(7831):763-8.
198. Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*. 2014;156(1-2):45-68.
199. Matsuda S, Yasukawa T, Sakaguchi Y, Ichiyanagi K, Unoki M, Gotoh K, et al. Accurate estimation of 5-methylcytosine in mammalian mitochondrial DNA. *Scientific reports*. 2018;8(1):5801.
200. Skvortsova K, Zotenko E, Luu P-L, Gould CM, Nair SS, Clark SJ, et al. Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA. *Epigenetics & chromatin*. 2017;10:16-.
201. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes & development*. 2016;30(7):733-50.
202. Corley MJ, Zhang W, Zheng X, Lum-Jones A, Maunakea AK. Semiconductor-based sequencing of genome-wide DNA methylation states. *Epigenetics*. 2015;10(2):153-66.

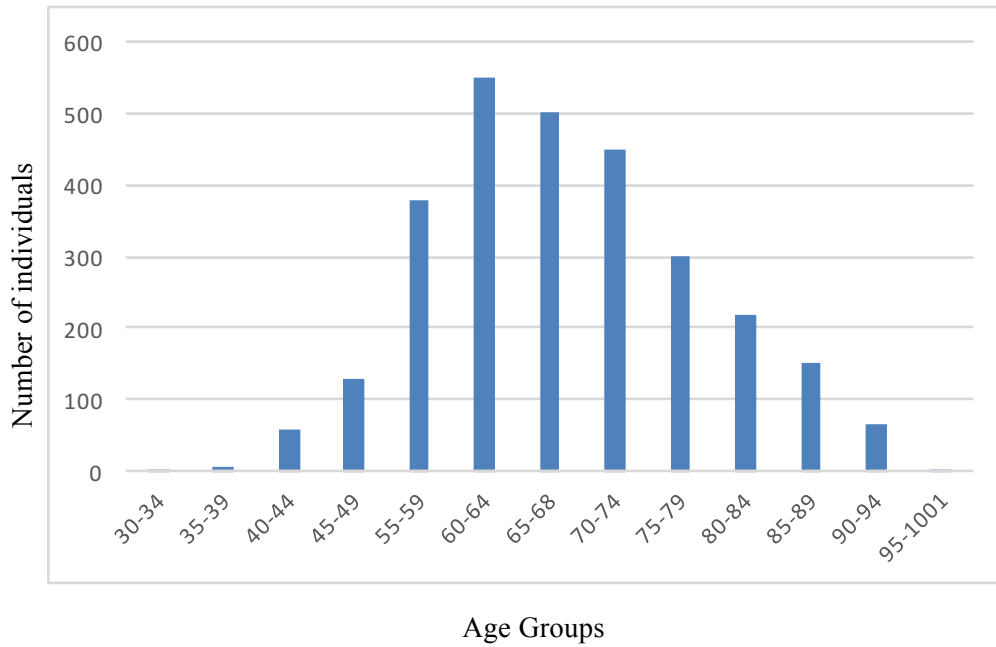
203. Blaschke K, Ebata KT, Karimi MM, Zepeda-Martínez JA, Goyal P, Mahapatra S, et al. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*. 2013;500(7461):222-6.
204. Zhang X, Su J, Jeong M, Ko M, Huang Y, Park HJ, et al. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nature genetics*. 2016;48(9):1014-23.

## **ANNEXES**

**ANNEXE 1**

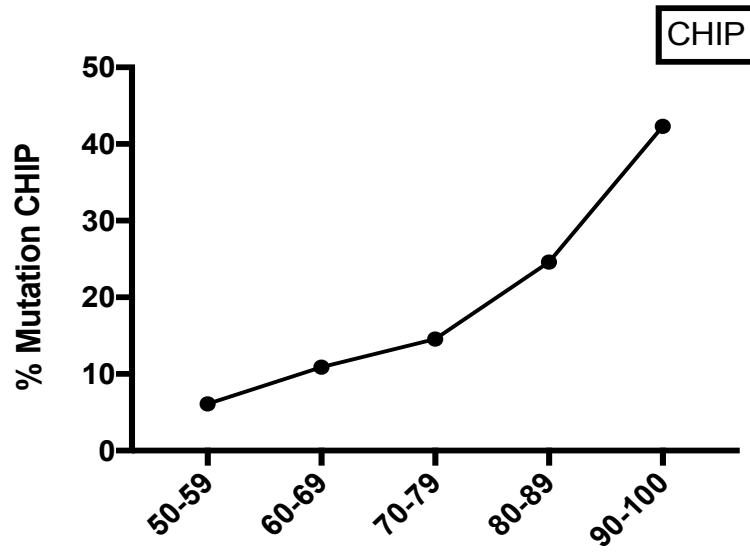
**FIGURES ET TABLEAUX SUPPLÉMENTAIRES**





Cohort	
<b>Individuals</b>	2996
<b>Age Range</b>	30-101
<b>Related</b>	2172
<b>Unrelated</b>	824

**Figure 25:** Age distribution of the 2996 healthy women of the cohort



Mutations	
Mutated Individuals	347 (13.7%)
Somatic Mutations	402
Mean Age	73.2

**Figure 26:** Prevalence of somatic mutations in the 347 mutated individuals of the cohort

**Tableau 14:** Prevalence of XCI skewing is age-dependent in blood cells and maximal in PMN

**PMN**

Age groups	Count of pmn_DS	Average of pmn_DS2	StdDev of pmn_DS2	IC95	Mean-	Mean+
50-59	505	0,212578769	0,134562578	0,011736178	0,200842591	0,224314947
60-69	912	0,237463339	0,137283833	0,008909843	0,228553496	0,246373182
70-79	630	0,25349115	0,148885239	0,011625977	0,241865173	0,265117127
80-89	283	0,270405659	0,150181587	0,017497321	0,252908338	0,28790298
>90	42	0,297475683	0,126630669	0,038296788	0,259178895	0,33577247

**T cells**

Age groups	Count of CD3_DS	Average of CD3_DS2	StdDev of CD3_DS3	IC95	Mean-	Mean+
50-59	470	0,16059917	0,115322604	0,010425901	0,150173269	0,171025071
60-69	807	0,169847429	0,11809005	0,008147505	0,161699924	0,177994934
70-79	549	0,17916636	0,121774198	0,010186317	0,168980043	0,189352677
80-90	239	0,184348529	0,118071102	0,014968999	0,16937953	0,199317528
>90	25	0,183828403	0,109421509	0,042892444	0,140935959	0,226720846

**Epithelial cells**

Age groups	Count of fb_DS	Average of fb_DS2	StdDev of fb_DS3	IC95	Mean-	Mean+
50-59	478	0,145441985	0,101159126	0,009068578	0,136373407	0,154510563
60-69	847	0,134208197	0,100127425	0,006743097	0,127465099	0,140951294
70-79	576	0,138300115	0,106543723	0,008700911	0,129599204	0,147001026
80-90	256	0,139463353	0,100674871	0,012332445	0,127130908	0,151795798
>90	28	0,116659688	0,093791675	0,034740284	0,081919404	0,151399972

**Tableau 15:** Prevalence of XCI skewing in PMN and CHIP at different VAF in function of age groups

**PMN skewing**

Age groups	Count of age	Count of #Mutation	%	StdDev	Interval 95%	Mean-	Mean+
50-59	377	23	6,100795756	0,01140	2,23	2,23	2,23
60-69	1064	115	10,80827068	0,00729	1,43	1,43	1,43
70-79	742	107	14,42048518	0,01088	2,13	2,13	2,13
80-90	347	87	25,07204611	0,02167	4,25	4,25	4,25
>90	36	15	41,66666667	0,08275	16,22	16,22	16,22
Grand Total	2566	347					

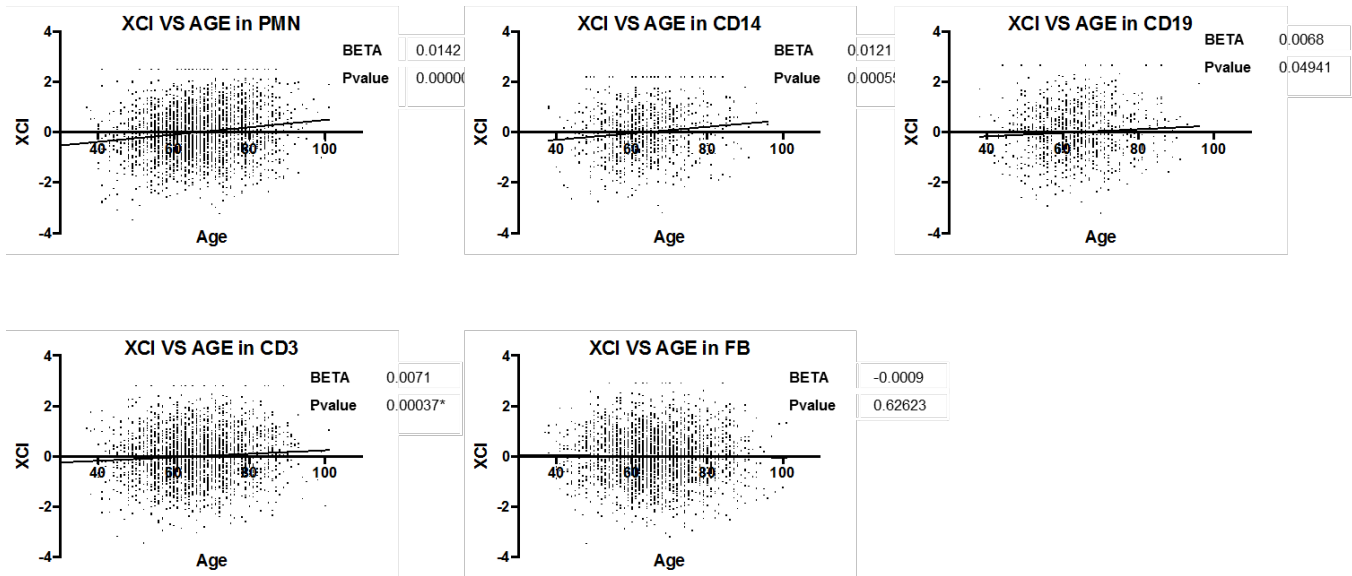
**CHIP VAF >10**

Age groups	Count of age	Count of #Mutation	%	StdDev	Interval 95%	Mean-	Mean+
50-59	377	7	1,856763926	0,00643	1,26	1,26	1,26
60-69	1064	52	4,887218045	0,00507	0,99	0,99	0,99
70-79	742	54	7,277628032	0,00806	1,58	1,58	1,58
80-90	347	45	12,96829971	0,01680	3,29	3,29	3,29
>90	52	9	17,30769231	0,05244	10,28	10,28	10,28
Grand Total	2582	167					

**CHIP VAF >25**

Age groups	Count of age	Count of #Mutation	%	StdDev	Interval 95%	Mean-	Mean+
50-59	377	2	0,530503979	0,00346	0,68	0,68	0,68
60-69	1064	16	1,503759398	0,00286	0,56	0,56	0,56
70-79	742	22	2,964959569	0,00526	1,03	1,03	1,03
80-90	347	14	4,034582133	0,00984	1,93	1,93	1,93
>90	52	4	7,692307692	0,03694	7,24	7,24	7,24
Grand Total	2582	58					

### Age Effect on XCI Skewing in All

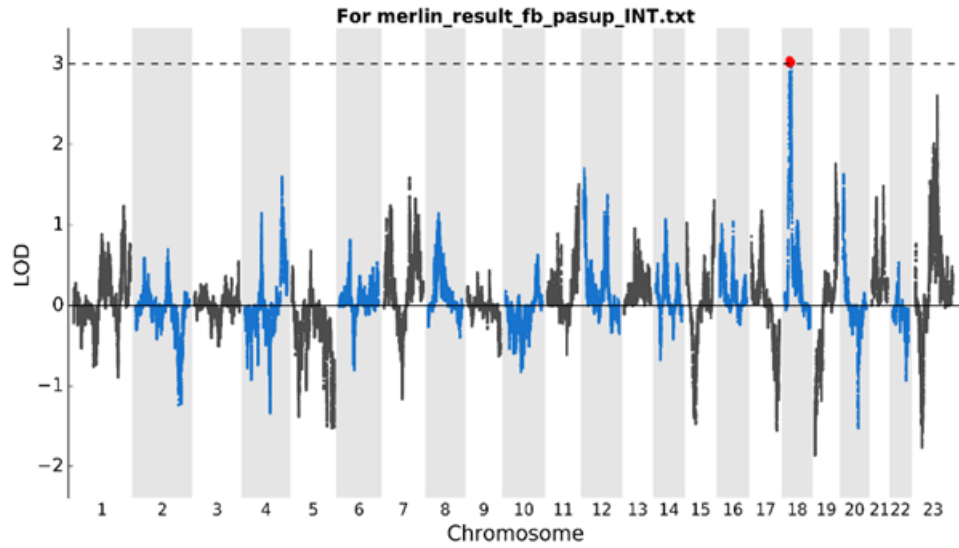


**Figure 27:** Age-dependent XCI skewing in blood cells

XCI skewing increases with age in blood cells but not in epithelial cells. The incidence of skewing increases with age in all blood cells but the association was not statistically significant in B cells nor in epithelial cells. Each plot shows the association of XCI ratio with age in every tissues cells type. Each dot represent an individual. Beta and Pvalue are indicated. PMN (n = 2536 samples), monocyte (n = 885 samples), B cells (n = 923 samples), T cells (n = 2222 samples) and epithelial cells (n = 2310 samples)

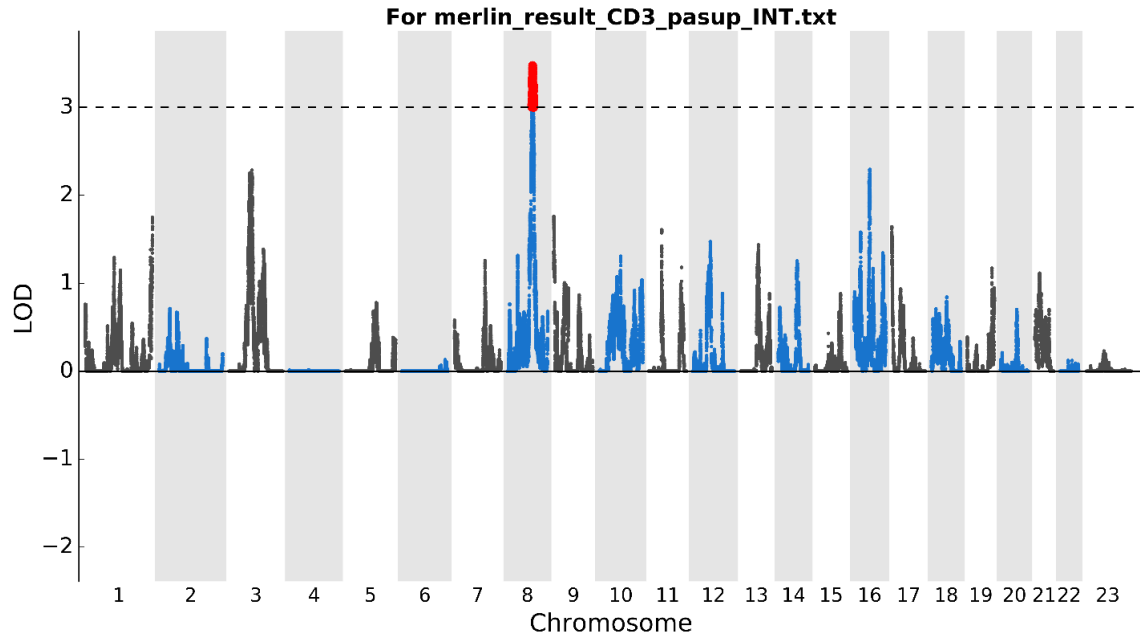
**Tableau 16:** Correlation between XCI and VAF of CHIP mutations in PMN

<b>SUB GROUP</b>	<b>ENDPOINT</b>	<b>PREDICTOR</b>	<b>N</b>	<b>BETA</b>	<b>Pvalue</b>
ALL (mutations)	PMN	VAF_ALL	305	0,0176	0,0003
		VAF_TET2	108	0,0199	0,0070
		VAF_DNMT3A	185	0,0239	0,0029
<sup>1</sup> pmn ds analysis was adjusted for categorie_pmn_ds (0/1)					



**Figure 28:** MERLIN QTL Linkage analysis in epithelial cells

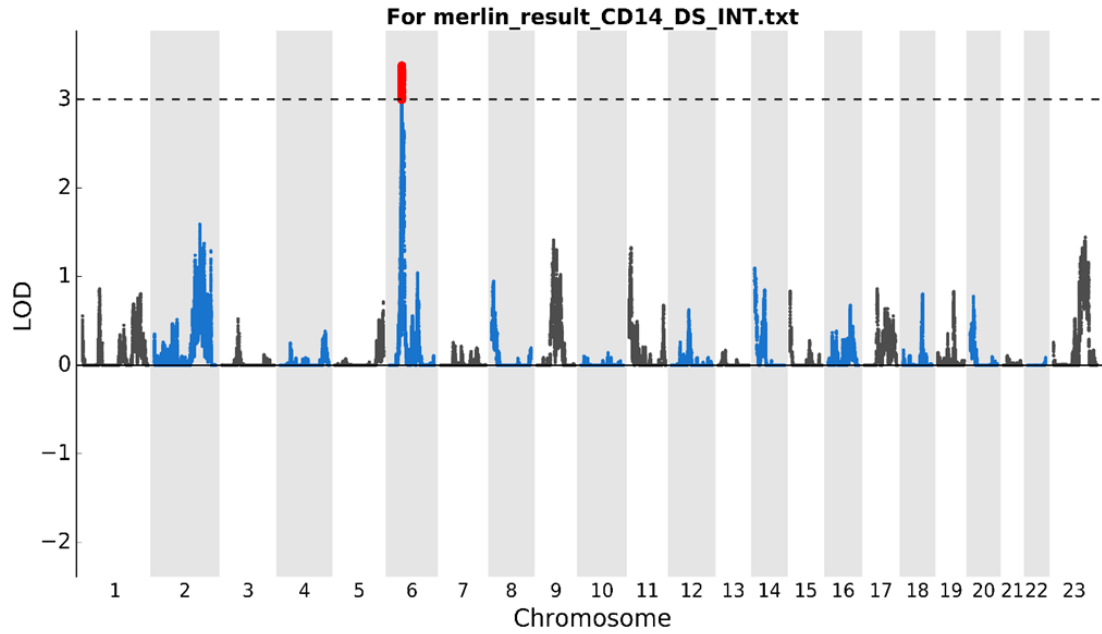
Manhattan plot of twopoint QTL linkage analysis for epithelial cell Psup controlling for age. CHIP mutations removed from the cohort. LOD score threshold was set to 3.9 for 476,199 variants.



**Figure 29:** MERLIN QTL Linkage analysis in T cells

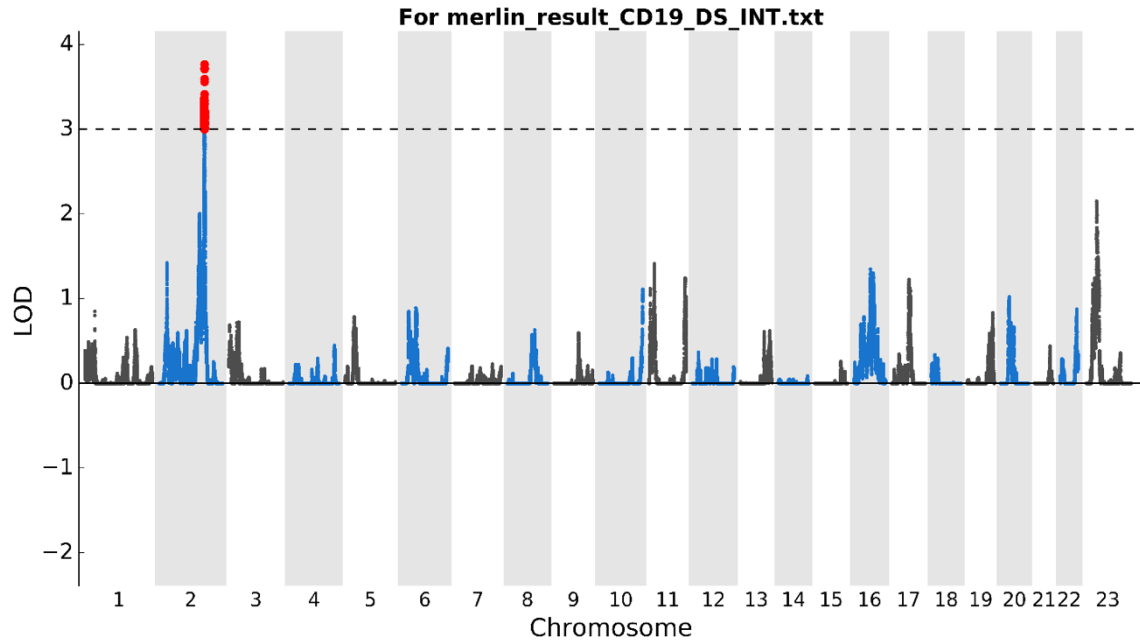
Manhattan plot of twopoint QTL linkage analysis for T cell Psup controlling for age. CHIP mutations removed from the cohort. LOD score threshold was set to 3.9 for 476,199 variants.





**Figure 30:** MERLIN VC Linkage analysis in monocytes

Manhattan plot of twopoint VC linkage analysis for monocytes Psup controlling for age. CHIP mutations removed from the cohort. LOD score threshold was set to 3.9 for 476,199 variants.



**Figure 31:** MERLIN VC Linkage analysis in B cells

Manhattan plot of twopoint VC linkage analysis for B cells Psup controlling for age. CHIP mutations removed from the cohort. LOD score threshold was set to 3.9 for 476,199 variants.

**ANNEXE 2**

**60 YEARS OF CLONAL HEMATOPOIESIS RESEARCH: FROM X-  
CHROMOSOME INACTIVATION STUDIES TO THE IDENTIFICATION OF  
DRIVER MUTATIONS**

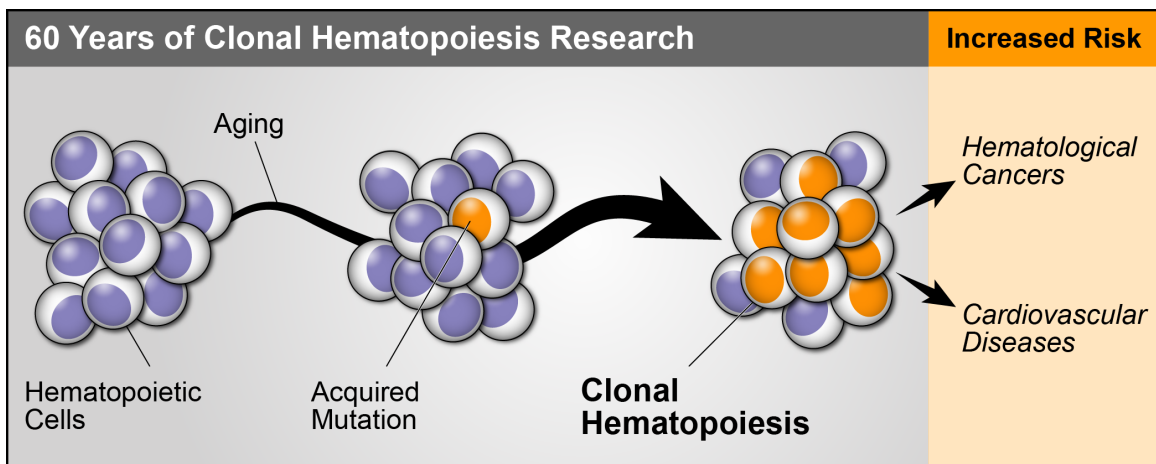
## 60 years of clonal hematopoiesis research: from X-Chromosome Inactivation studies to the identification of driver mutations.

**Sami Ayachi**<sup>1,3</sup>, Manuel Buscarlet<sup>1</sup> and \*Lambert Busque<sup>1,2,3</sup>

<sup>1</sup>Research Center, <sup>2</sup>Hematology Division, Hôpital Maisonneuve-Rosemont; <sup>3</sup>Université de Montréal, Montreal, Québec, Canada.

\*Corresponding author: Lambert Busque MD FRCPC, Hôpital Maisonneuve-Rosemont, 5415 boul. De l'Assomption, Montréal, Québec, Canada H1T 2M4. 514-252-3400 Ext 3741. [lbusque.hmr@ssss.gouv.qc.ca](mailto:lbusque.hmr@ssss.gouv.qc.ca)

### Visual abstract



## ABSTRACT

The history of clonal hematopoiesis (CH) research was punctuated by several seminal discoveries that have forged our understanding of cancer development. The clever application of the principle of random X-chromosome inactivation (XCI) in females lead to the development of the first test to identify clonal derivation of cells. Initially limited by a low level of informativeness, the applicability of these assays expanded with differential methylation-based assays at highly polymorphic genes such as the human androgen receptor (*HUMARA*). Twenty years ago, the observation that skewing of XCI ratios increases as women age, was the first clue that lead to the identification of mutations in the *TET2* gene in hematologically normal aging individuals. In 2014, large-scale genomic approaches of 3 cohorts allowed to define CH, which was demonstrated to increase the risk of developing hematological cancers and cardiovascular diseases. These observations created a fertile field of investigation aimed at investigating the etiologies and consequences of CH. The most frequently mutated genes in CH are *DNMT3A*, *TET2* and *ASXL1*, which have a role in hematopoietic stem cell (HSC) development and self-renewal. These mutations confer a competitive advantage to the CH clones. However, the penetrance of CH is age-dependent but incomplete, suggesting the influence of extrinsic factors. Recent data attribute a modest role to genetic predisposition, but several observations point to the impact of a proinflammatory milieu that advantages the mutated clones. CH may be a barometer of non-healthy aging and interventions devised at curbing its initiation or progression should be a research priority.

«The origin and development of human tumors studied with cell markers» is the title of a 1974 review paper by the late Philip J. Fialkow (1). It articulates both the necessity for and fascination with the study of clonality. The human body has to produce, over its lifespan, trillions of quality cells originating from a single fertilized egg. When this proliferation system derails, we are at risk of cancer. In this short review we will provide a historical perspective of the study of clonal hematopoiesis and outline future opportunities and challenges.

### **X-CHROMOSOME INACTIVATION (XCI) ANALYSIS.**

In 1961, Mary Lyon proposed that each X-chromosome in excess of one is randomly inactivated in cells of the developing female embryo (2). Therefore, following XCI, females present a mosaic pattern of two cell populations, one expressing maternal (X<sub>m</sub>), the other paternal (X<sub>p</sub>) X-linked genes. Almost concurrently, Ernest Beutler, proved this theory in human females heterozygous for alleles of the Glucose-6-Phosphate Dehydrogenase (G6PD) gene (3). These observations paved the way to the first clonality studies. Linder and Gartler studied G6PD variants (A and B) by electrophoresis in 5 female heterozygotes and documented that 27 samples of leiomyomas had only one band in contrast to 85 of 86 myometrium samples that had both alleles present, indicating that these tumors had a single cell origin (4). This study opened a fertile field of investigation that has allowed investigators to unravel some of the basic tenets of modern oncology (5).

The applicability of XCI analysis with the G6PD assay was drastically limited by the low level of heterozygosity in the general population. This precluded analysis of rare disorders, analysis of large cohorts of females, and familial analysis. In the mid 1980's, Vogelstein *et al.* made a seminal contribution to the field by identifying DNA polymorphisms linked to residues differentially methylated between the active (X<sub>a</sub>) and inactive X-chromosome (X<sub>i</sub>) (6). Differential methylation was validated as a surrogate for protein expression. Gilliland *et al.* developed a PCR-based method allowing, for the first time, to study small populations of cells (7). Another clever way of improving the informative value of XCI assays was to look for transcriptional polymorphisms of X-linked genes. Prchal *et al.* have been instrumental in developing such assays (8, 9). These assays have the theoretical advantage of not relying on methylation patterns to assess the activity of the X-chromosome, since they directly rely on the exclusive transcriptional activity of X<sub>a</sub>.

In 1992, XCI studies expanded with the introduction of a novel assay of the human androgen-receptor gene (AR or HUMARA [GeneBank]) by Allen *et al.* (10). This assay benefitted from a multiallelic CAG short tandem repeat (STR) in the first exon of the gene that is closely linked to differentially methylated cytosine residues between X<sub>a</sub>. The AR CAG repeat is polymorphic in 90% of females of all racial groups (11) rendering XCI analysis feasible in the majority of females with a single PCR-based assay. Interestingly, an RNA-based assay was developed and correlated with the HUMARA assay, supporting reliability of the differential methylation patterns at this locus (12). In fact, the concordance of transcriptional assays with HUMARA was challenged only once (13) and ultimately further validated (14, 15). The HUMARA assay is currently the most frequently used method of XCI analysis (PubMed, December 2019: 324 publication using HUMARA key word).

### **XCI ANALYSIS OF THE NORMAL POPULATION.**

Significant skewing of XCI ratio from the normal, theoretical 1:1 ratio has been artificially determined to be  $\geq 3:1$  (which corresponds to the expression of 75% of one allele)(16-21). Determination of the prevalence of skewing should be straightforward since it can be simply estimated by using XCI assays in healthy females. However, initial estimation was contradictory. Vogelstein *et al.* found only 3 of 81 (3.7%) normal females to have skewing using the phosphoglycerate kinase (PGK) and hypoxanthine phosphoribosyl transferase (HPRT) X-inactivation probes (20). Gale *et al.* found significant skewing in blood-derived cells in 23% of normal females using PGK and HPRT probes (18) and 22% of females using M27 $\beta$  (21), which has been confirmed in other studies (20, 22). Discordance in the incidence of skewing was initially explained by the diversity of assays used, the different criteria for skewing, and the small population sizes. However, a more plausible explanation came from studies that analyzed XCI patterns of different tissues in the same female. For example, Fey *et al.* reported that the incidence of skewing was low in gastrointestinal mucosa and thyroid tissue, but was significantly higher in blood cells (23). Gale *et al.* reported that 45% of females analyzed had different patterns of XCI when blood derived cells were compared to muscle and skin samples (19). Interestingly, blood cells in both studies were more frequently skewed than other tissues. These studies suggested that XCI patterns were tissue specific. The second, and perhaps most important clue to the discrepancy in incidence of skewing came with the analysis of XCI patterns in females of different age groups. An age difference in skewing was first suggested by Fey *et al.*, who reported a higher incidence of skewing in older females compared to children using the M27 $\beta$  probe (24). We characterized the effect of age on skewing in a large cross sectional study using the more robust and validated HUMARA assay in 295 normal females from three age groups: (i) neonates, (ii) 28-32 years old, and (iii) age greater than 60 years old. We documented an incidence of skewing (ratio  $\geq 3:1$ ) of 8.6% in neonates, 16.4% in the 28-32 year olds and 37.9% in women greater than 60 years old ( $P < 0.0001$  versus neonates,  $P = 0.064$  versus 28 to 32 years) (17). These results have been confirmed by other investigators (15, 25-31).

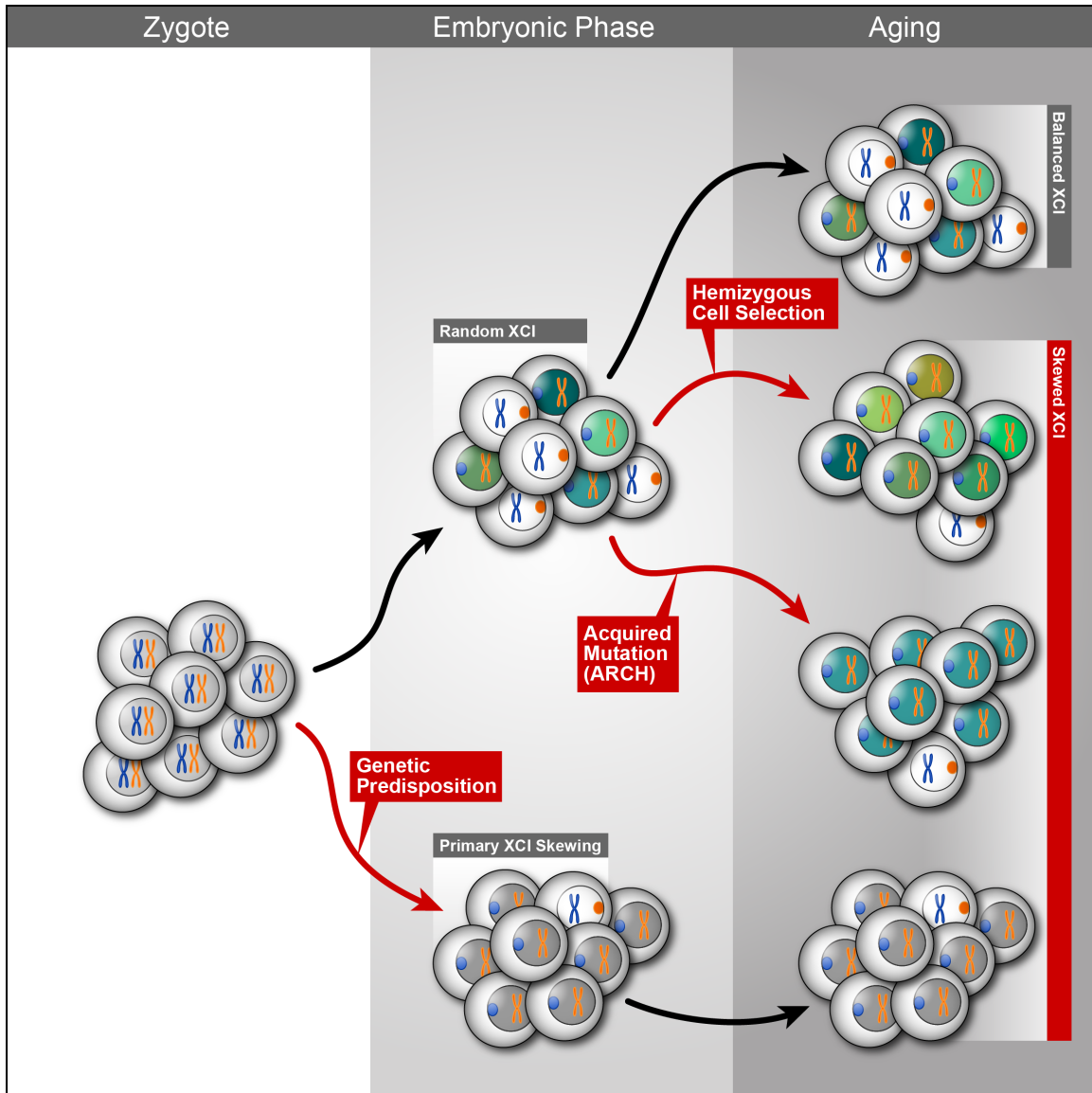
### **INVESTIGATING THE ETIOLOGY OF XCI ACQUIRED SKEWING**

A number of mechanisms have been put forward to explain the increased prevalence of skewing observed in hematopoietic cells including (i) acquired clonal hematopoiesis (mutation driven), (ii) stochastic clonal dominance caused by hematopoietic stem cell (HSC) depletion, and (iii) genetic predisposition (26, 32-34). Several lines of evidence have supported a genetic contribution to the trait. Christensen *et al.* analyzed peripheral blood cells from 71 elderly monozygotic (MZ) twin pairs, and observed a strong correlation in the degree and direction of skewing between co-twins (35). Vickers *et al.* studied the peripheral blood polymorphonuclear cells (PMNs) of 29 MZ and 18 dizygotic (DZ) twin pairs, and observed a better intraclass correlation of XCI ratios between MZ twins than between DZ twins (36). They estimated heritability to account for 68% of the skewing observed in PMNs. Kristiansen *et al.* reported a comparable heritability estimate in a larger cohort of elderly MZ ( $n=82$ ) and DZ ( $n=112$ ) twin pairs (33). These studies collectively suggest that the skewing of XCI observed in hematopoietic cells is, at least in part, genetically determined. Evidence for an X-linked genetic component to skewing was

provided by Abkowitz *et al.* (37), who analyzed the hematopoietic cells of ageing female Safari cats (first-cross hybrids between Geoffroy and domestic cats), and showed that XCI was random in young cats, but skewed in 67% of older cats. As recombination between the two parental X-chromosomes cannot occur in this animal model, the fact that skewing always favored the Geoffroy X-chromosome (37) suggested a hemizygous cell selection process and was not compatible with random processes.

To investigate the genetic and non-genetic components of acquired skewing of XCI, we recruited subjects comprised of 2530 women of French-Canadian ancestry without any known hematological disorders, ranging from 55 to 101 years old. The cohort comprised 1727 individuals belonging to 435 families and 803 unrelated individuals. We were intrigued by the observation that skewing was more prevalent in myeloid cells than in T-cells or buccal cells and more age-dependent (data not shown). We hypothesized that skewing was probably multifactorial and that some subjects could have acquired true, mutation-driven clonal hematopoiesis. We first analyzed DNA from three elderly women with known skewing in their myeloid and polyclonal T-cells using exome sequencing. We identified somatic mutations in TET2, DNMT3A, and SLC39A12 in one of them (38). Extension of the analysis to various age groups led to the identification of missense, nonsense, and frameshift somatic TET2 mutations in the DNA from PMN, but not from lymphocyte or epithelial cells, in 10/179 (5.6%) elderly subjects with XCI skewing, 0/105 elderly subjects without XCI skewing, and 0/96 younger subjects with XCI skewing. The hematologic parameters of TET2 mutant individuals did not differ from those of age-matched counterparts. This study was the first to demonstrate that an acquired mutation in a myeloid cancer associated gene (TET2) is age-dependent and compatible with normal hematopoiesis (38). It also suggested that an acquired mutation in a driver gene is a small contributor to acquired XCI skewing in the general population (Figure 1). Simultaneously, Laurie *et al.* showed an age-dependent clonal mosaicism for large chromosomal anomalies using single nucleotide polymorphism (SNP) microarray data from over 50,000 individuals recruited for genome-wide association studies (39). Clonal mosaicism was detected in 2-3% of elderly individuals in contrast to 0.5% of individuals younger than 50 years. Furthermore, the presence of clonal mosaicism was associated with a 10-fold increased risk of developing a hematological cancer. Jacobs *et al.* (40) and Forsberg *et al.* (41) reported similar findings.





**Figure 1.** One of the two X-chromosomes is inactivated during the early embryonic phase with a theoretical ratio of 1:1 between paternal X and maternal X-chromosome. The inactive X is represented as the Barr body. Deviation from the 1:1 ratio may occur stochastically due to the low number of cells present at the timing of X-Chromosome Inactivation (XCI), or because of the presence of a lethal allele on one of the two X-chromosome (reviewed in (42)). The majority of individuals will have balance XCI ratios after birth which will be maintained lifelong. After the age 30, a significant proportion of women will have skewing of their XCI ratio in favor of one of the two parental X-chromosomes (17). Several different etiologies may explain this acquired skewing of XCI. According to the Abkowitz's model (37) this may be due to a selective growth advantage conferred by germline alleles of genes residing on the X-chromosome. For example, if an allele of a specific gene, residing on the paternal-X, is competitively stronger than the allele of the same gene on the maternal X, all cells with the paternal X active will have a growth advantage to the cells with the maternal X active, there will be hemizygous cell selection

with time. Importantly, all progenitors with paternal X-chromosome active will contribute to hematopoiesis. This proliferation is polyclonal but with a skewed XCI profile. In contrast, a stem cell can have an acquired mutation in a driver or non-driver gene conferring a growth advantage which will lead to true clonal derivation of cells. Because XCI patterns are faithfully transmitted to daughter cells, all cell will have the same X in the active state, leading to a skewed pattern of XCI. Only identification of specific mutation (most efficiently with NGS) has allowed to distinguish these two different phenomenon with the same XCI phenotype (38).

## **MUTATION DRIVEN CLONAL HEMATOPOIESIS IN THE NORMAL POPULATION**

In 2014, three groups reported analysis of DNA exome datasets from large cohorts of subjects and documented age-dependent mutations in driver genes (43-45). These three cohorts had different inclusion criteria (non-hematological cancer, psychiatric disorder or cardiovascular disease), yet they identified a similar set of genes, suggesting a universal age-associated phenomenon. Although more than 70 different genes were identified, the most frequently mutated ones were the epigenetic modifiers DNMT3A, TET2 and ASXL1, nicknamed the «DAT» mutations. Importantly, the prospective data available for two of these studies revealed that subjects with clonal hematopoiesis had a roughly 10-fold heightened risk of developing a hematological cancer (44, 45). The relative risk of mortality was modest at 1.4. However, clonal hematopoiesis was associated with an increased risk of cardiovascular events, which was further confirmed (46). This new entity has several designations, but it is generally accepted that Age-Related Clonal Hematopoiesis (ARCH) refers to clonal expansion irrelevant of the type of mutation (47), and Clonal Hematopoiesis of Indeterminate Potential (CHIP), to clonal hematopoiesis caused by a mutation in a driver gene at a Variant Allele Fraction (VAF) of >2% (48).

Zink *et al.*, using a whole-genome sequencing (WGS) approach on a large cohort from Iceland, confirmed the high prevalence of mutations in driver gene, but an even higher rate of mutations in non-driver candidates (49). We recently reported the analysis of a large cohort of 2530 normal individuals, aged 55 to 100 years old, using a more sensitive (relative to a genome-based approach) gene-targeted approach (50). We documented a high prevalence of mutations and even higher preponderance for mutations in DNMT3A or TET2 - genes central to DNA methylation and demethylation. Clone size (VAF) was significant in a large proportion of affected individuals. The average VAF was 14.3%, which corresponds to 28.6% of cells originating from mutated stem cells. Almost half of the mutated cohort (48.7%) had a VAF  $\geq$ 10%. There were no significant differences in blood cell parameters (cell numbers and indices) between mutant and aged-matched controls, except for a tendency toward reduced PMN in TET2 mutants. Interestingly, we have noted familial aggregation for TET2 mutations (*vide infra*) as well as mutations in DNMT3A and TET2 originating at different levels of the hematopoietic hierarchy. DNMT3A lineage restriction patterns are compatible with a pluripotent stem cell origin, whereas TET2 mutations occur mainly in myeloid cells and sometimes in B-cells (51). More disturbing to our current understanding of pre-malignant lesions, Young *et al.* demonstrated mutations in TET2 or DNMT3A at very low frequencies in 95% of

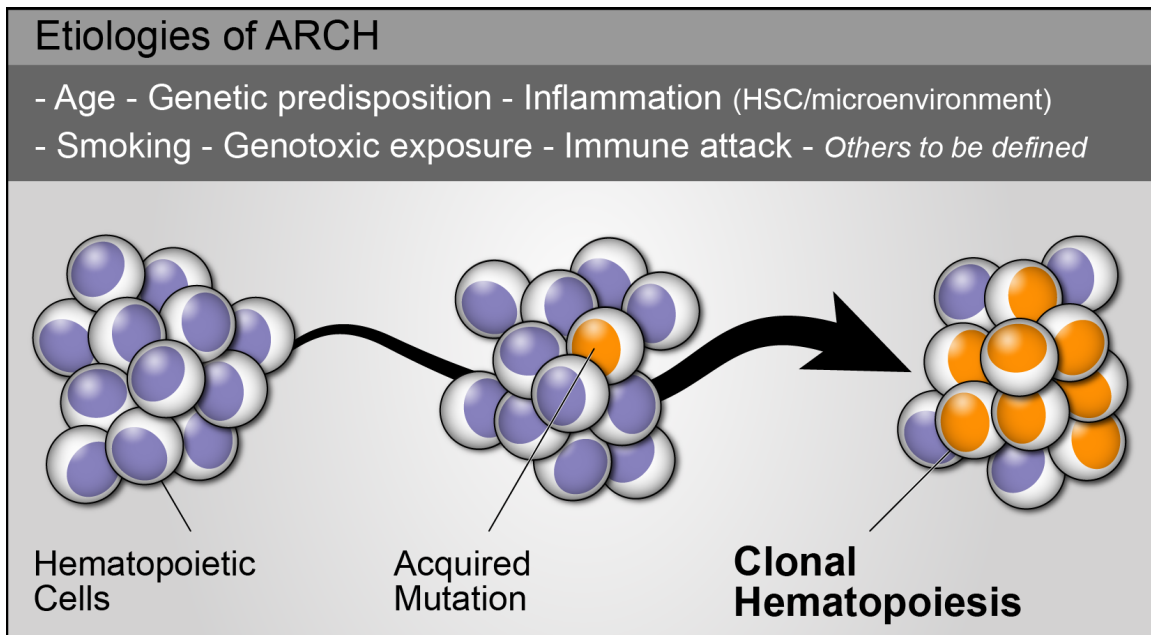
individuals aged 50 to 60 years old, suggesting that these mutations are almost ubiquitous after the age of 50 (52).

### **ETIOLOGY OF ARCH/CHIP**

Aging HSC face the potential of exhaustion. Several parameters differentiate old versus young HSC (reviewed in (53)), such as cumulating random DNA damage (54), reduced telomere length (55, 56), increased polarity (57), reduced autophagy (58), and epigenetic reprogramming (59). Some of these changes are presumably signs of imminent demise, but some others are possibly adaptation mechanisms for sustainability. One thing is certain: throughout the aging process, fewer and fewer HSC contribute to hematopoietic output, although this does not impact overall production. Currently, the main recognized candidate driver genes are DNMT3A and TET2 (50). Studies in mice have demonstrated that loss of function of either of these genes leads to increased HSC self-renewal capacity. Loss of *Dnmt3a* leads to expansion of HSC number (60) with an almost infinite serial transplantation capacity (61). Conditional *Tet2* loss-of-function induces Lineage<sup>-</sup> Sca1<sup>+</sup>cKit<sup>+</sup> (LSK) cell expansion concomitant to decreased 5-hydroxymethyl-cytosine (5hmC) levels and increased self-renewal capacity of HSC (62-64). The specific TET2 or DNMT3A mutations that have been documented in the blood cells of aging individuals are mainly loss-of-function mutations, and they include indels, frameshift, nonsense, splice site and missense mutations which abrogate catalytic function. It is therefore reasonable to speculate that HSC mutated in either of these two genes have increased self-renewal capacity, and hence are able to outcompete unmutated HSC. However, despite the almost ubiquitous presence of these mutations over the age of fifty (52), only a fraction will develop an expanded clone (>2% VAF) over time, suggesting inter-individual variation in the competitiveness of unmutated versus mutated HSC.

Knowledge about the factors (other than aging) that predispose to acquisition/progression of these mutations is so far incomplete and should be investigated (Figure 2). Genetic predisposition has been demonstrated in several cancer including Myeloproliferative Neoplasms (MPNs) which share several driver mutations with CH. In MPNs, there is a genetic predisposition to acquire JAK2 V617F mutations conferred by a germline haplotype (65, 66). Furthermore, germline variations at the TERT locus have been associated with familial clustering of MPN (reviewed in (67)). For CH, Zink *et al.* have shown an association between a small germline deletion in intron 3 of the telomerase reverse transcription (TERT) gene (49) and ARCH. Studying a cohort sip-ships, we identified a significant family risk of 2.7 for TET2 mutations but not for DNMT3A (50). Interestingly, Thompson *et al.* documented a significant genetic contribution to age-associated Y-chromosome loss in blood cells of men (68), a situation that shares several similarities with CH: loss of Y-chromosome is a form of clonal mosaicism that has been associated with cancer development (69) and cardiovascular events (70). Studying large cohorts of several hundreds of thousands of men, they identified recurrent predisposition loci and estimated the heritability of the trait as 31.7%. However, recently two groups looked at the concordance of ARCH in monozygotic and dizygotic twin pairs (71, 72) and the results refuted a genetic contribution to ARCH. Hansen *et al.* conducted a population-based study of 594 twins from 299 pairs aged 73-94 years with a follow-up of 20 years. Mutations were identified in 36% of twins, but no different concordance was observed

between monozygotic and dizygotic twin pairs. In contrast to our observation (50), no difference was noted between DNMT3A vs TET2. Furthermore, the size of the clone did not influence results (71). Similar results were obtained by Fabre *et al.* in a smaller cohort of twins (72). These results are in sharp contrast with those obtained when looking at XCI skewing in twins discussed above, suggesting that ARCH is mainly sporadic and that XCI skewing is partly genetic. Clearly, more studies are needed to conclude on the quantitative contribution of genetic factors leading to ARCH.



**Figure 2.** Etiologies of ARCH/CHIP are numerous and not completely defined at this time. Aging is prerequisite but intrinsic factors such as the survival or stem cell renewal advantage conferred by the acquired mutation is central to its capacity to outgrow normal stem cells. The genetic predisposition contribution is controversial at this time, but extrinsic factors such as a proinflammatory milieu of the microenvironment seems to play a central role in clonal progression.

Another factor that has gained interest in the pathogenesis of ARCH is the HSC microenvironment. In fact, Vas *et al.* have shown that the aging microenvironment influences clonality of malignant cells in a HSC-independent manner and may therefore also be a contributor to ARCH (73). The association of CHIP with cardiovascular (44-46) and chronic pulmonary (49, 50) diseases raises the possibility that age-associated chronic inflammation, often called « inflammaging » (74), may be a key common denominator between these medical conditions and the emergence of CH. There is evidence suggesting that a proinflammatory milieu favors mutated stem cells. For example, chronic myelogenous leukemia (CML) stem cells have increased expression of IL-1 receptor and sensitivity to IL-1 in comparison to normal HSC, and are therefore favored in an inflammatory environment (75). Abegunde *et al.* demonstrated that a pro-inflammatory environment supported by TNF- $\alpha$  promotes the expansion of Tet2 mutant clones in mice (76). It is currently unclear if inflammaging promotes clonal hematopoiesis independently

from other age-associated medical conditions or if they are somehow linked. Interestingly, Tet2 mutant myeloid cells have a deficient capacity to resolve inflammation caused by increased production of IL-6 (77), and Tet2 deficient mice are predisposed to atherosclerosis (78). Recently, Meisel *et al.* demonstrated that preleukemic myeloproliferation (PMP) was only seen in Tet2 deficient mice (Tet2<sup>-/-</sup>) with a dysfunction of their small-intestinal barrier associated with bacterial translocation and increased IL-6 production (79). Preleukemic myeloproliferation failed to develop in germ-free and antibiotic treated Tet2<sup>-/-</sup> mice, and neutralization of IL-6 reversed preleukemic myeloproliferation, suggesting that anti-inflammatory intervention may prevent clonal expansion. Furthermore, proinflammatory cytokines have been implicated in the pathogenesis of myelodysplastic syndromes and are evaluated as potential therapeutic targets (reviewed in (80)). In humans, Jaiswal *et al.* demonstrated that patients with CHIP and cardiovascular disease had higher level of IL-8 in a subset of patients with TET2 mutations (46). Cook *et al.* showed increased levels of IL-6 and increased co-morbidities in patients with CHIP (81). Furthermore, Myers SanMiguel *et al.* demonstrated that transplantation of Dnmt3aR8878H<sup>+</sup> bone marrow cells into young versus old congenic recipient mice, lead to an accelerated expansion of the Dnmt3aR8878H<sup>+</sup> cells in the older mice (82). TNF- $\alpha$  and M-CSF were identified by RNA-seq to have higher expression levels in older mice. Taken together these studies support the hypothesis that low grade inflammation promotes the initial clonal expansion of mutated HSC, and that subsequently, as the clone size increases, it directly further impedes the control of inflammation which may ultimately affect organs such as the heart, lung, pancreas, brain, etc.

The potential etiologies of ARCH discussed so far applied to the normal aging population. However, under certain iatrogenic conditions clonal selection and expansion may occur. This is the case for genotoxic exposure. Clonal analyses of blood from patients who have developed treatment-related AML (t-AML) have been instrumental in elucidating the physiopathology of these complications. Wong *et al.* documented that 4 out of 22 patients with t-AML had, prior to exposure to chemotherapy, the exact TP53 mutation documented in their AML specimen, but at a very low VAF, supporting expansion of this clone under the influence of cytotoxic agents (83). Gillis *et al.* also identified pre-existing TP53 mutations as a risk factor for t-AML (84). Interestingly, they showed that TET2 mutations were prevalent pre-chemotherapy but not associated with treatment-related hematological cancer. Recently a genomic study on a large cohort of patients found TP53 and PPM1D mutations to be enriched in therapy-induced MDS compared to primary MDS (85), offering supporting evidence that clones harboring these mutations thrive under therapy-induced selective pressure. This was demonstrated in patients undergoing autologous transplantation for lymphoma where pre-existing CHIP was associated with a fourfold increased risk of t-MDS or t-AML [86].

Likewise, mutations in the TP53 modulator PPM1D have been detected in the blood of patients with ovarian [87] and lung [88] cancer after therapy [89–91]. Studying a large cohort of patients treated for nonhematologic cancers, Coombs *et al.* [92] recently documented that CHIP was prevalent in 25% of them, associated with age, prior radiation, tobacco smoking and adverse outcomes. The strongest association was found between TP53 and PPM1D mutations and prior treatment with chemotherapy and radiation.

Furthermore, cell lines bearing PPM1D mutations have been reported to outcompete normal cells when exposed to cytotoxic DNA-damaging agents [93]. Interestingly, Comen *et al.* [94] have found that tumor-infiltrating leukocytes (TILeuks) present in the biopsy of four of seven breast cancer patients that later developed t-AML harbored the same mutations as the one documented in the in the AML clone. Taken together, these data suggest that exogenous genotoxic stress selects naturally occurring clones that harbor mutations conferring resistance to these toxic agents. It may be possible to identify patients at risk of developing treatment related cancer before treatment. However, the predictive value of mutations for this rare complication may be specific for certain genes (such as TP53 and PPM1D) and not be the same for other genes (such as TET2 for example). Screening for mutations in individuals submitted to intensive chemotherapy/radiotherapy may allow a better evaluation of the risk–benefit ratio of therapy.

ARCH-related mutations have also been identified in the context of bone marrow failure syndromes (reviewed in [95]). The pattern of mutations identified in aplastic anemia bears certain differences with that found in the general aging population. Aplastic anemia patients have a high prevalence of mutations (45%), which can be subcategorized as mutations more specific to aplastic anemia (PIGA, BCOR, and BCORL1) and other mutations classically documented in ARCH (DNMT3A and ASXL1) [96]. The incidence of DNMT3A and ASXL1 mutations increases with age, whereas PIGA and BCOR/BCORL1 mutations are equally represented across all age groups [96]. PIGA and BCOR/BCORL1 mutations have been associated with increased responsiveness to immunotherapeutic intervention, while classic ARCH mutations have been associated with less sensitivity. Aplastic anemia is also subject to loss of human leukocyte antigen (HLA) class I allele, which may help escape immune attack [97,98]. Therefore, it seems as if, depending on the specific mutation, some may favor evasion of immune attack while others may favor HSC self-renewal.

### **MUTATIONS IN OTHER TISSUES THAN BLOOD**

Age-associated somatic mutation accumulation is not limited to hematopoietic stem cells [99]. The most frequently mutated normal tissues are skin, esophagus, and lung, suggesting that environmental exposure during aging promotes somatic mosaicism [100]. Middle-aged individuals with sun-exposed skin have high burdens of cancer-associated somatic mutations [101]. Interestingly, study of esophagus biopsies reveals a high frequency of NOTCH1 mutation and TP53. The NOTCH1 mutations were significantly age dependent and underrepresented in esophageal cancer, indicating a selective growth advantage that would be minimally oncogenic. The study of all tissues, including blood, will help to elucidate common mechanisms of adaptation to aging and early cancer development.

### **FUTURE DIRECTIONS**

Clonal hematopoiesis offers a new opportunity to understand the biology and adaptation mechanisms of aging hematopoiesis, and provides insight into the mechanisms underlying malignant transformation. Furthermore, it might shed new light on common denominators of age-associated medical conditions such as heart disease. CHIP may be a biomarker of global non-heathy aging. As such, its presence may command changes in life style (102) and potential medical intervention. If the link between CHIP and inflammation is solidly

confirmed, it will be important to see if medical interventions aimed at controlling inflammation, such as with the anti-IL1 canakinumab antibody (103), or the recently published COLCOT study (104) that used low-dose colchicine to prevent recurrent cardiovascular events, will curb and prevent expansion of ARCH-related clones and the risk of associated hematological cancers.

#### **ACKNOWLEDGEMENTS**

Part of the original work that supported this review was funded for LB by the Canadian Institutes of Health Research (CIHR) and Leukemia Lymphoma Society of Canada (LLSC). We are thankful to Drs. Natasha Szuber and Luigina Mollica for insightful comments.

## REFERENCES

1. Fialkow PJ. The origin and development of human tumors studied with cell markers. *The New England journal of medicine*. 1974;291(1):26-35.
2. Lyon MF. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet*. 1962;14:135-48.
3. Beutler E, Yeh M, Fairbanks VF. The normal human female as a mosaic of X-chromosome activity: studies using the gene for C-6-PD-deficiency as a marker. *Proc Natl Acad Sci U S A*. 1962;48:9-16.
4. Linder D, Gartler SM. Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. *Science*. 1965;150(3692):67-9.
5. Busque L, Gilliland DG. X-inactivation analysis in the 1990s: promise and potential problems. *Leukemia*. 1998;12(2):128-35.
6. Vogelstein B, Fearon ER, Hamilton SR, Feinberg AP. Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. *Science*. 1985;227(4687):642-5.
7. Gilliland DG, Blanchard KL, Levy J, Perrin S, Bunn HF. Clonality in myeloproliferative disorders: analysis by means of the polymerase chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*. 1991;88(15):6848-52.
8. Prchal JT, Guan YL. A novel clonality assay based on transcriptional analysis of the active X chromosome. *Stem Cells*. 1993;11 Suppl 1:62-5.
9. Luhovy M, Liu Y, Belickova M, Prchal JF, Prchal JT. A novel clonality assay based on transcriptional polymorphism of X chromosome gene p55. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation*. 1995;1(2):81-7.
10. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet*. 1992;51(6):1229-39.
11. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*. 1992;12(2):241-53.
12. Busque L, Zhu J, DeHart D, Griffith B, Willman C, Carroll R, et al. An expression based clonality assay at the human androgen receptor locus (HUMARA) on chromosome X. *Nucleic Acids Res*. 1994;22(4):697-8.
13. Swierczek SI, Agarwal N, Nussenzveig RH, Rothstein G, Wilson A, Artz A, et al. Hematopoiesis is not clonal in healthy elderly women. *Blood*. 2008;112(8):3186-93.
14. Busque L, Paquette Y, Provost S, Roy DC, Levine RL, Mollica L, et al. Skewing of X-inactivation ratios in blood cells of aging women is confirmed by independent methodologies. *Blood*. 2009;113(15):3472-4.



15. Mossner M, Nolte F, Hutter G, Reins J, Klaumunzer M, Nowak V, et al. Skewed X-inactivation patterns in ageing healthy and myelodysplastic haematopoiesis determined by a pyrosequencing based transcriptional clonality assay. *J Med Genet.* 2013;50(2):108-17.
16. Allen RC, Nachtman RG, Rosenblatt HM, Belmont JW. Application of carrier testing to genetic counseling for X-linked agammaglobulinemia. *Am J Hum Genet.* 1994;54(1):25-35.
17. Busque L, Mio R, Mattioli J, Brais E, Blais N, Lalonde Y, et al. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood.* 1996;88(1):59-65.
18. Gale RE, Wheadon H, Linch DC. X-chromosome inactivation patterns using HPRT and PGK polymorphisms in haematologically normal and post-chemotherapy females. *Br J Haematol.* 1991;79(2):193-7.
19. Gale RE, Wheadon H, Boulos P, Linch DC. Tissue specificity of X-chromosome inactivation patterns. *Blood.* 1994;83(10):2899-905.
20. Vogelstein B, Fearon ER, Hamilton SR, Preisinger AC, Willard HF, Michelson AM, et al. Clonal analysis using recombinant DNA probes from the X-chromosome. *Cancer Res.* 1987;47(18):4806-13.
21. Gale RE, Wheadon H, Linch DC. Assessment of X-chromosome inactivation patterns using the hypervariable probe M27 beta in normal hemopoietic cells and acute myeloid leukemic blasts. *Leukemia.* 1992;6(7):649-55.
22. Puck JM, Stewart CC, Nussbaum RL. Maximum-likelihood analysis of human T-cell X chromosome inactivation patterns: normal women versus carriers of X-linked severe combined immunodeficiency. *American journal of human genetics.* 1992;50(4):742-8.
23. Fey MF, Peter HJ, Hinds HL, Zimmermann A, Liechti-Gallati S, Gerber H, et al. Clonal analysis of human tumors with M27 beta, a highly informative polymorphic X chromosomal probe. *The Journal of clinical investigation.* 1992;89(5):1438-44.
24. Fey MF, Liechti-Gallati S, von Rohr A, Borisch B, Theilkas L, Schneider V, et al. Clonality and X-inactivation patterns in hematopoietic cell populations detected by the highly informative M27 beta DNA probe. *Blood.* 1994;83(4):931-8.
25. Champion KM, Gilbert JG, Asimakopoulos FA, Hinshelwood S, Green AR. Clonal haemopoiesis in normal elderly women: implications for the myeloproliferative disorders and myelodysplastic syndromes. *Br J Haematol.* 1997;97(4):920-6.
26. Gale RE, Fielding AK, Harrison CN, Linch DC. Acquired skewing of X-chromosome inactivation patterns in myeloid cells of the elderly suggests stochastic clonal loss with age. *Br J Haematol.* 1997;98(3):512-9.
27. Tonon L, Bergamaschi G, Dellavecchia C, Rosti V, Lucotti C, Malabarba L, et al. Unbalanced X-chromosome inactivation in haemopoietic cells from normal women. *Br J Haematol.* 1998;102(4):996-1003.

28. Racchi O, Mangerini R, Rapezzi D, Rolfo M, Gaetani GF, Ferraris AM. X chromosome inactivation patterns in normal females. *Blood Cells Mol Dis.* 1998;24(4):439-47.
29. El Kassar N, Hetet G, Briere J, Grandchamp B. X-chromosome inactivation in healthy females: incidence of excessive lyonization with age and comparison of assays involving DNA methylation and transcript polymorphisms. *Clin Chem.* 1998;44(1):61-7.
30. Hatakeyama C, Anderson CL, Beever CL, Penaherrera MS, Brown CJ, Robinson WP. The dynamics of X-inactivation skewing as women age. *Clin Genet.* 2004;66(4):327-32.
31. Sharp A, Robinson D, Jacobs P. Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. *Hum Genet.* 2000;107(4):343-9.
32. Sandovici I, Naumova AK, Leppert M, Linares Y, Sapienza C. A longitudinal study of X-inactivation ratio in human females. *Hum Genet.* 2004;115(5):387-92.
33. Kristiansen M, Knudsen GP, Bathum L, Naumova AK, Sorensen TI, Brix TH, et al. Twin study of genetic and aging effects on X chromosome inactivation. *Eur J Hum Genet.* 2005;13(5):599-606.
34. Abkowitz JL, Catlin SN, Gutterop P. Evidence that hematopoiesis may be a stochastic process in vivo. *Nat Med.* 1996;2(2):190-7.
35. Christensen K, Kristiansen M, Hagen-Larsen H, Skytthe A, Bathum L, Jeune B, et al. X-linked genetic factors regulate hematopoietic stem-cell kinetics in females. *Blood.* 2000;95(7):2449-51.
36. Vickers MA, McLeod E, Spector TD, Wilson IJ. Assessment of mechanism of acquired skewed X inactivation by analysis of twins. *Blood.* 2001;97(5):1274-81.
37. Abkowitz JL, Taboada M, Shelton GH, Catlin SN, Gutterop P, Kiklevich JV. An X chromosome gene regulates hematopoietic stem cell kinetics. *Proc Natl Acad Sci U S A.* 1998;95(7):3862-6.
38. Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet.* 2012;44(11):1179-81.
39. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet.* 2012;44(6):642-50.
40. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet.* 2012;44(6):651-8.
41. Forsberg LA, Rasi C, Razzaghian HR, Pakalapati G, Waite L, Thilbeault KS, et al. Age-related somatic structural changes in the nuclear genome of human blood cells. *Am J Hum Genet.* 2012;90(2):217-28.
42. Busque L, Buscarlet M, Mollica L, Levine RL. Concise Review: Age-Related Clonal Hematopoiesis: Stem Cells Tempting the Devil. *Stem Cells.* 2018;36(9):1287-94.

43. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med.* 2014;20(12):1472-8.
44. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014;371(26):2477-87.
45. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371(26):2488-98.
46. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med.* 2017;377(2):111-21.
47. Shlush LI. Age-related clonal hematopoiesis. *Blood.* 2018;131(5):496-504.
48. Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood.* 2015;126(1):9-16.
49. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood.* 2017;130(6):742-52.
50. Buscarlet M, Provost S, Zada YF, Barhdadi A, Bourgoin V, Lepine G, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood.* 2017;130(6):753-62.
51. Buscarlet M, Provost S, Zada YF, Bourgoin V, Mollica L, Dube MP, et al. Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. *Blood.* 2018;132(3):277-80.
52. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat Commun.* 2016;7:12484.
53. de Haan G, Lazare SS. Aging of hematopoietic stem cells. *Blood.* 2018;131(5):479-87.
54. Beerman I. Accumulation of DNA damage in the aged hematopoietic stem cell compartment. *Semin Hematol.* 2017;54(1):12-8.
55. Rudolph KL, Chang S, Lee HW, Blasco M, Gottlieb GJ, Greider C, et al. Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell.* 1999;96(5):701-12.
56. Mollica L, Fleury I, Belisle C, Provost S, Roy DC, Busque L. No association between telomere length and blood cell counts in elderly individuals. *J Gerontol A Biol Sci Med Sci.* 2009;64(9):965-7.
57. Florian MC, Dorr K, Niebel A, Daria D, Schrezenmeier H, Rojewski M, et al. Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation. *Cell Stem Cell.* 2012;10(5):520-30.

58. Ho TT, Warr MR, Adelman ER, Lansinger OM, Flach J, Verovskaya EV, et al. Autophagy maintains the metabolism and function of young and old stem cells. *Nature*. 2017;543(7644):205-10.
59. Kramer A, Challen GA. The epigenetic basis of hematopoietic stem cell aging. *Semin Hematol*. 2017;54(1):19-24.
60. Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet*. 2011;44(1):23-31.
61. Jeong M, Park HJ, Celik H, Ostrander EL, Reyes JM, Guzman A, et al. Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *Cell Rep*. 2018;23(1):1-10.
62. Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*. 2011;20(1):11-24.
63. Quivoron C, Couronne L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell*. 2011;20(1):25-38.
64. Ho PA, Kopecky KJ, Alonzo TA, Gerbing RB, Miller KL, Kuhn J, et al. Prognostic implications of the IDH1 synonymous SNP rs11554137 in pediatric and adult AML: a report from the Children's Oncology Group and SWOG. *Blood*. 2011;118(17):4561-6.
65. Jones AV, Chase A, Silver RT, Oscier D, Zoi K, Wang YL, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet*. 2009;41(4):446-9.
66. Kilpivaara O, Mukherjee S, Schram AM, Wadleigh M, Mullally A, Ebert BL, et al. A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nat Genet*. 2009;41(4):455-9.
67. Tashi T, Swierczek S, Prchal JT. Familial MPN Predisposition. *Curr Hematol Malig Rep*. 2017;12(5):442-7.
68. Thompson DJ, Genovese G, Halvardson J, Ulirsch JC, Wright DJ, Terao C, et al. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature*. 2019;575(7784):652-7.
69. Forsberg LA, Rasi C, Malmqvist N, Davies H, Pasupulati S, Pakalapati G, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet*. 2014;46(6):624-8.
70. Haitjema S, Kofink D, van Setten J, van der Laan SW, Schoneveld AH, Eales J, et al. Loss of Y Chromosome in Blood Is Associated With Major Cardiovascular Events During Follow-Up in Men After Carotid Endarterectomy. *Circulation Cardiovascular genetics*. 2017;10(4):e001544.
71. Hansen JW, Pedersen DA, Larsen LA, Husby S, Clemmensen SB, Hjelmberg J, et al. Clonal hematopoiesis in elderly twins: concordance, discordance and mortality. *Blood*. 2019.

72. Fabre MA, McKerrell T, Zwiebel M, Vijayabaskar MS, Park NR, Wells PM, et al. Concordance for clonal hematopoiesis is limited in elderly twins. *Blood*. 2019.
73. Vas V, Senger K, Dorr K, Niebel A, Geiger H. Aging of the microenvironment influences clonality in hematopoiesis. *PLoS One*. 2012;7(8):e42080.
74. Franceschi C, Garagnani P, Vitale G, Capri M, Salvioli S. Inflammaging and 'Garb-aging'. *Trends Endocrinol Metab*. 2017;28(3):199-212.
75. Zhang B, Chu S, Agarwal P, Campbell VL, Hopcroft L, Jorgensen HG, et al. Inhibition of interleukin-1 signaling enhances elimination of tyrosine kinase inhibitor-treated CML stem cells. *Blood*. 2016;128(23):2671-82.
76. Abegunde SO, Buckstein R, Wells RA, Rauh MJ. An inflammatory environment containing TNFalpha favors Tet2-mutant clonal hematopoiesis. *Exp Hematol*. 2018;59:60-5.
77. Zhang Q, Zhao K, Shen Q, Han Y, Gu Y, Li X, et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature*. 2015;525(7569):389-93.
78. Fuster JJ, MacLauchlan S, Zuriaga MA, Polackal MN, Ostriker AC, Chakraborty R, et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science*. 2017;355(6327):842-7.
79. Meisel M, Hinterleitner R, Pacis A, Chen L, Earley ZM, Mayassi T, et al. Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. *Nature*. 2018;557(7706):580-4.
80. Sallman DA, List A. The central role of inflammatory signaling in the pathogenesis of myelodysplastic syndromes. *Blood*. 2019;133(10):1039-48.
81. Cook EK, Izukawa T, Young S, Rosen G, Jamali M, Zhang L, et al. Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. *Blood Adv*. 2019;3(16):2482-6.
82. SanMiguel JM, Loberg M, Heuer S, Stearns T, Young K, Trowbridge J. Cell-Extrinsic Stressors from the Aging Bone Marrow (BM) Microenvironment Promote Dnmt3a-Mutant Clonal Hematopoiesis. *Blood*. 2019;134(Supplement\_1):5.
83. Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS, et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature*. 2015;518(7540):552-5.
84. Gillis NK, Ball M, Zhang Q, Ma Z, Zhao Y, Yoder SJ, et al. Clonal haemopoiesis and therapy-related myeloid malignancies in elderly patients: a proof-of-concept, case-control study. *Lancet Oncol*. 2017;18(1):112-21.
85. Lindsley RC, Saber W, Mar BG, Redd R, Wang T, Haagenon MD, et al. Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. *N Engl J Med*. 2017;376(6):536-47.

86. Gibson CJ, Lindsley RC, Tchekmedyian V, et al. Clonal hematopoiesis associated with adverse outcomes after autologous stem-cell transplantation for lymphoma. *J Clin Oncol*. 2017;35: 1598–1605.
87. Swisher EM, Harrell MI, Norquist BM, et al. Somatic mosaic mutations in PPM1D and TP53 in the blood of women with ovarian carcinoma. *JAMA Oncol*. 2016;2:370–372.
88. Zajkowicz A, Butkiewicz D, Drosik A, Giglok M, Suwinski R, Rusin M. Truncating mutations of PPM1D are found in blood DNA samples of lung cancer patients. *Br J Cancer*. 2015;112: 1114–1120.
89. Ruark E, Snape K, Humburg P, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. *Nature*. 2013;493:406–410.
90. Akbari MR, Lepage P, Rosen B, et al. PPM1D mutations in circulating white blood cells and the risk for ovarian cancer. *J Natl Cancer Inst*. 2014;106:djt323.
91. Cardoso M, Paulo P, Maia S, Teixeira MR. Truncating and missense PPM1D mutations in early onset and/or familial/hereditary prostate cancer patients. *Genes Chromosomes Cancer*. 2016;55:954–961.
92. Coombs CC, Zehir A, Devlin SM, et al. Therapy-related clonal hematopoiesis in patients with non-hematologic cancers is common and associated with adverse clinical outcomes. *Cell Stem Cell*. 2017;21:374–382.e4.
93. Hsu JI, Dayaram T, Tovy A, et al. PPM1D mutations drive clonal hematopoiesis in response to cytotoxic chemotherapy. *Cell Stem Cell*. 2018;23:700–713.e6.
94. Comen EA, Bowman RL, Selenica P, et al. Evaluating clonal hematopoiesis in tumor-infiltrating leukocytes in breast cancer and secondary hematologic malignancies. *J Natl Cancer Inst*. 2020;112:107–110.
95. Cooper JN, Young NS. Clonality in context: hematopoietic clones in their marrow environment. *Blood*. 2017;130:2363–2372.
96. Yoshizato T, Dumitriu B, Hosokawa K, et al. Somatic mutations and clonal hematopoiesis in aplastic anemia. *N Engl J Med*. 2015;373:35–47.
97. Babushok DV, Duke JL, Xie HM, et al. Somatic HLA Mutations expose the role of class I-mediated autoimmunity in aplastic anemia and its clonal complications. *Blood Adv*. 2017;1:1900–1910.
98. Katagiri T, Sato-Otsubo A, Kashiwase K, et al. Frequent loss of HLA alleles associated with copy number-neutral 6pLOH in acquired aplastic anemia. *Blood*. 2011;118:6601–6609.
99. Blokzijl F, de Ligt J, Jager M, et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*. 2016;538:260–264.
100. Yizhak K, Aguet F, Kim J, et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science*. 2019;364:eaaw0726.
101. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. 2015;348:880–886.

102. Libby P, Ebert BL. CHIP (clonal hematopoiesis of indeterminate potential). *Circulation*. 2018;138:666–668.
103. Ridker PM, Everett BM, Thuren T, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med*. 2017;377:1119–1131.
104. Tardif JC, Kouz S, Waters DD, et al. Efficacy and safety of low-dose colchicine after myocardial infarction. *N Engl J Med*. 2019;381:2497–2505.

**ANNEXE 3**

**HIGH-SENSITIVITY C-REACTIVE PROTEIN IS ASSOCIATED WITH  
CLONAL HEMATOPOIESIS OF INDETERMINATE POTENTIAL**



# High-sensitivity C-Reactive Protein is associated with Clonal Hematopoiesis of Indeterminate Potential

Lambert Busque<sup>1,2,3,\*¶</sup>, Maxine Sun<sup>3,4</sup>, Manuel Buscarlet<sup>1</sup>, **Sami Ayachi**<sup>1,3</sup>, Yassamin Feroz Zada<sup>4</sup>, Sylvie Provost<sup>4</sup>, Vincent Bourgoïn<sup>1</sup>, Luigina Mollica<sup>1,2,3</sup>, Marlies Meisel<sup>5</sup>, Reinhard Hinterleitner<sup>5</sup>, Bana Jabri<sup>6</sup>, Marie-Pierre Dubé<sup>3,4\*</sup> and Jean-Claude Tardif<sup>3,4\*¶</sup>

<sup>1</sup>Research Center, <sup>2</sup>Hematology Division, Hôpital Maisonneuve-Rosemont, <sup>3</sup>Université de Montréal, Montréal, Canada

<sup>4</sup>Beaulieu-Saucier Pharmacogenomics Center, Montreal Heart Institute, Montréal, Canada.

<sup>5</sup>Department of Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA

<sup>6</sup>Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA University of Pittsburgh School of Medicine

\*Submission corresponding author: Lambert Busque MD FRCPC, Hôpital Maisonneuve-Rosemont, Montréal, QC, Canada H1T 2M4. 514-252-3400 Ext 3741; [lbusque.hmr@ssss.gouv.qc.ca](mailto:lbusque.hmr@ssss.gouv.qc.ca);

¶Address reprint requests to Dr. Busque, Hôpital Maisonneuve-Rosemont, 5415 boul. de l'Assomption, Montréal, QC, Canada, [lbusque.hmr@ssss.gouv.qc.ca](mailto:lbusque.hmr@ssss.gouv.qc.ca); or Dr. Marie-Pierre Dubé, Montreal Heart Institute, QC, Canada H1T1C8, [marie-pierre.dube@mhi-rc.org](mailto:marie-pierre.dube@mhi-rc.org); or Dr. Jean-Claude Tardif, the Montreal Heart Institute, 5000 Belanger St., Montreal, QC H1T 1C8, Canada, [jean-claude.tardif@icm-mhi.org](mailto:jean-claude.tardif@icm-mhi.org)

## Abstract

Clonal Hematopoiesis of Indeterminate Potential (CHIP) is predictive of hematological cancers and cardiovascular diseases, but the etiology of CHIP initiation and clonal expansion is unknown. Several lines of evidence suggest that proinflammatory cytokines may favor mutated Hematopoietic Stem Cell (HSC) expansion. To investigate the potential link between inflammation and CHIP, we performed targeted deep sequencing of 11 genes previously implicated in CHIP in 1887 subjects aged over 70 years-old from the Montreal Heart Institute (MHI) Biobank, of which 1359 had prior coronary artery disease (CAD), and 528 controls did not. We assessed association of CHIP with log (ln) transformed high-sensitivity C-reactive protein (hs-CRP), a validated biomarker of inflammation. CHIP was identified in 427 of the 1887 subjects (22.6%). CHIP mutations were more frequently identified in *DNMT3A* (11.6%) and *TET2* (6.1%), with a higher proportion of *TET2* mutations occurring in controls than in patients with CAD (9.0% vs 4.9%,  $P < 0.001$ ). CHIP carriers had 21% higher hs-CRP levels compared to their non-carrier counterparts ( $e\beta = 1.21$ , 95% confidence interval [CI]: 1.08-1.36,  $P = 0.001$ ). A similar effect was observed in the subgroup of patients with known CAD ( $e\beta = 1.22$ , 95% CI: 1.06-1.41,  $P = 0.005$ ). These findings confirm the association between inflammation and CHIP. This association may open investigational avenues aimed at documenting mechanisms linking inflammation to clonal progression, and ultimately supports prevention interventions to attenuate CHIP's impact on cardiovascular disease and cancer.

**Key point:** Subjects with CHIP have higher level of hs-CRP

## Introduction

Clonal Hematopoiesis (CH) occurring in normal aging subjects, initially suggested by X-chromosome inactivation (XCI) studies,<sup>1,3</sup> is caused by acquired mutations in genes recurrently mutated in hematological cancers,<sup>4,7</sup> and in non-driver candidates.<sup>6,8</sup> CH prevalence increases significantly in patients aged above 60 years-old, and confers an increased risk of progression to hematological cancers and cardiovascular diseases.<sup>5,6,8,9</sup> The precise risk associated with the presence of CH in healthy individuals is uncertain, hence the creation of a clinical entity named Clonal Hematopoiesis of Indeterminate Potential (CHIP).<sup>10</sup>

Little is known about the etiology of clone initiation and clonal expansion. Genetic predisposition is controversial. Zink *et al.* showed an association between a small germline deletion in intron 3 of the telomerase reverse transcription gene (*TERT*)<sup>8</sup> and CH. Studying a cohort of sib-ships, we identified a significant 2.7-fold increase in the familial risk for mutation in *TET2* but not in *DNMT3A*.<sup>11</sup> However, a strong genetic contribution to CH was recently refuted when the concordance of CH was studied in monozygotic and dizygotic twin pairs.<sup>12,13</sup>

The association between CHIP and both cardiovascular<sup>5,6,9</sup> and chronic pulmonary diseases,<sup>11</sup> raised the possibility that age-associated chronic inflammation<sup>14</sup> may be a key common denominator between these medical conditions.<sup>15</sup> Studies in mice have supported the role of inflammation in clonal expansion of mutated Hematopoietic Stem Cells (HSC). Abegunde *et al.* demonstrated that a proinflammatory environment supported by TNF- $\alpha$  promotes the expansion of *Tet2* mutant clones in mice,<sup>16</sup> and we found that inflammation was a key driver of preleukemic myeloproliferation in *Tet2* deficient mice (*Tet2*<sup>-/-</sup>).<sup>17</sup> In humans, Jaiswal *et al.* demonstrated higher IL-8 levels in patients with CHIP and cardiovascular disease in a subset of 12 individuals with *TET2* mutations.<sup>9</sup> A trend towards increased levels of IL-6 was also observed in patients with CHIP.<sup>18</sup>

We report here a statistically significant correlation between high-sensitivity C-Reactive Protein (hs-CRP), a validated and routinely available biomarker of inflammation,<sup>19</sup> and CHIP.

## Methods

### Study population

We selected all subjects aged 70 years-old or over (1940) from participants of the Montreal Heart Institute (MHI) biobank, an ongoing prospective cohort including 23,000 individuals for the purpose of clinical and genetic research,<sup>20</sup> that had hs-CRP level tested. Participants were recruited on a voluntary basis during any hospital visit, regardless of the presence or stage of heart disease. All subjects underwent a medical questionnaire by a research nurse and their electronic chart was reviewed. DNA, plasma and serum were collected at baseline.

For the purpose of the current study, patients with coronary artery disease (CAD) were defined as those with a prior history of myocardial infarction (MI), percutaneous coronary intervention (PCI), and/or coronary artery bypass graft (CABG) surgery. Subjects also needed to be aged 70 years-old or above.

The protocol was approved by MHI's ethics committee and performed in accordance to the Helsinki declaration.

### **hs-CRP**

hs-CRP concentration was measured by quantitative immuno-nephelometric analysis on a Dimension Vista 500 Intelligent Lab System (Siemens Healthineers).

### **CHIP determination by Next Generation Sequencing (NGS)**

Subject's DNA (n=1940) was sequenced at high coverage (95% >500x) on an Ion Proton sequencer using a custom Ampliseq "CHIP" panel (Thermo Fisher Scientific) designed to target the top 11 genes reported in CHIP (*ASXL1*, *CBL*, *DNMT3A*, *GNAS*, *GNB1*, *JAK2* (chr9:5073674- 5073808), *PPM1D*, *SF3B1* (exons 14-16), *SRSF2*, *TET2* and *TP53*)<sup>5-7,11</sup> with 202 amplicons covering 38.49 kb. The panel coverage, specificity and sensitivity were validated (see Supplemental Figure 1). Mutations were considered present if the Variant Allele Frequency (VAF) was  $\geq 2\%$  as defined by Steensma *et al.*<sup>10</sup> Base calling, alignment (hg19) and variant calling were performed in instrument by TorrentServer v5.8.0 (Thermo Fisher Scientific). Subsequently, mutations were annotated and filtered using IonReporter v5.8 (Thermo Fisher Scientific) and only exonic and splice site mutations with a minor allele frequency (MAF)  $\leq 0.001$  were kept for further annotation. Frameshift, nonsense, in-frame deletions or insertions, splice sites and predicted consequential missense mutations (based on ClinVar, FATHMM or PolyPhen) were considered significant. In absence of non-hematological tissue for germline status confirmation, mutations with a VAF of 50% or 100% ( $\pm 4\%$ ) were considered potentially germline and excluded (n=9) except if they co-occurred with other somatic mutations (n=9) (see Supplemental Table 1).

### **Statistical analyses**

Bivariate associations were evaluated using the Fisher's Exact test and the Kruskal-Wallis test for categorical and continuously coded variables, respectively. hs-CRP was modelled as  $\ln(\text{hs-CRP})$  due to its non-normal distribution. The geometric mean was calculated by taking the antilog of the mean of the log-transformed hs-CRP data, as previously described.<sup>21</sup> Normality of residuals from a generalized linear regression (GLM) model adjusted for age, sex, body mass index (BMI), and previous history of CAD with  $\ln(\text{CRP})$  was confirmed. GLMs tested the associations of CHIP-associated mutations and levels of  $\ln(\text{hs-CRP})$  after adjusting for age, sex, body mass index (BMI), and CAD status at baseline. Stratified analyses were performed according to CAD status. Back transformation of the regression beta term for  $\ln(\text{CRP})$  was derived as  $e^\beta$ , and the 95% confidence interval (CI) as  $e^{\beta \pm 1.96 \times \text{standard error}}$ . The percent difference in hs-CRP between CHIP+ and CHIP- was derived as  $(e^\beta - 1) \times 100$ . Additional analyses were performed according to the CHIP VAF, and categorized as VAF  $\geq 0.10$  vs. no mutation. Analyses were performed using SAS v.9.4 and R version 3.5.1. In sensitivity analyses, we adjusted for additional confounding factors, such as diabetes mellitus at baseline, as well as statin, aspirin, and beta-blocker use at baseline. In sub-group analyses, we focused on those without a history of cancer at study entry.

## Results

### Study population

The population was comprised of 1940 subjects, of which 9 were excluded because of potential germline mutations (see below) and another 44 were excluded because they had cardiovascular disease without CAD at baseline. The remaining 1887 participants included 1359 patients with at least one previous account of MI, PCI, or CABG, and 528 patients without a previous history of CAD.

### CHIP-associated mutations

We identified 427 CHIP carriers among the 1887 participants (22.6%, Figure 1). Of these, 331 (17.5%) had a single mutation, and 96 (5.1%) had more than one mutation. The prevalence and the relative proportion between the 11 candidate genes were similar to previous reports.<sup>5-8,11</sup> Mutations in *DNMT3A*, *TET2* and *ASXL1* accounted for the majority of mutations (82.9%). The mean VAF was 13.3%. The prevalence of CHIP carriers was slightly higher in the non-CAD than in the CAD cohort (25.3% vs 21.5%,  $P=0.075$ ). The relative prevalence between the different genes was similar between the two cohorts, except for *TET2* whose prevalence was higher in the non-CAD cohort (9.0% vs 4.9%,  $P<0.001$ ). All mutations are described in Supplemental Table 1.

### Population characteristics according to CHIP carrier status

Subjects were segregated according to their CHIP carrier status (CHIP- or CHIP+) in three groups: (i) all subjects, (ii) subjects with CAD, and (iii) subjects without CAD. Table 1 describes the baseline characteristics of the three groups. Univariable analyses showed that CHIP carriers were significantly older (mean: 75.0 vs. 74.4 years-old,  $P=0.002$ ), were less affected by dyslipidemia (85% vs. 88%,  $P=0.041$ ), and had fewer previous PCI (30% vs. 38%,  $P=0.006$ ) than CHIP non-carriers. hs-CRP was significantly higher in CHIP carriers compared to non-carriers (median: 1.60 vs. 1.41 mg/L,  $P=0.009$ ). Smoking, past or current, was not associated with CHIP.

### Primary analysis (entire cohort)

Our primary analysis focused on all patients, regardless of CAD status, and consisted in performing a multivariable generalized linear regression analysis for prediction of hs-CRP ( $n=1887$ ). CHIP carriers of any gene mutations had significantly higher hs-CRP than non-CHIP carriers in adjusted analyses ( $e^{\beta}=1.21$ , 95% CI: 1.08-1.36,  $P=0.001$ , Table 2). Focusing on mutations in individual genes, CHIP carriers of mutations in *DNMT3A* had higher hs-CRP than those without mutations in that gene ( $e^{\beta}=1.17$ , 95% CI: 1.01-1.36,  $P=0.04$ ). No other significant effect was observed for any of the remaining individual genes.

### Secondary analyses (CAD and non-CAD subgroups)

Among the patients with CAD at baseline ( $n=1359$ ), CHIP carriers were older (mean: 75.1 vs. 74.5 years), had fewer previous PCI (44% vs. 52%), more congestive heart failure (24% vs 18%), and higher hs-CRP (median: 1.60 vs. 1.39 mg/L) than non-CHIP carriers (all  $P<0.05$ , Table 1). In the non-CAD cohort ( $n=528$ ), CHIP carriers were also significantly older than non-carriers (74.9 vs. 74.1 years) but had lower rates of current smoking (0.8% vs. 4.6%), and hs-CRP levels were not significantly different between those two groups. In multivariable generalized linear regression analyses focusing on patients with a history of CAD at baseline,

CHIP carriers of any mutations had higher hs-CRP than non-CHIP carriers after adjusting for covariates ( $e^{\beta}=1.22$ , 95% CI: 1.06-1.41,  $P=0.005$ , Table 2). This association was not significant in the smaller subgroup of patients without a history of CAD at baseline ( $e^{\beta}=1.15$ , 95% CI: 0.94-1.40,  $P=0.178$ ).

### **Sensitivity, sub-group and additional analyses**

When limiting analyses to individuals with VAF  $\geq 10\%$ , as compared to those without CHIP, carriers of any gene mutation also had higher  $\ln(\text{CRP})$  ( $e^{\beta}=1.23$ , 95% CI: 1.05-1.45,  $P=0.038$ ) in adjusted analyses (data not shown). This was not different than the entire CHIP cohort ( $e^{\beta}=1.21$ ). However, we could not document a quantitative correlation between  $\ln(\text{CRP})$  and CHIP due to limited sample size in patients with VAF  $\geq 10\%$  ( $n=191$ ). In sensitivity analyses, we adjusted for additional confounders (statin, aspirin, beta-blocker use at baseline, as well as diabetes mellitus) that could have influenced the results of our analyses. The additional adjustment of these confounding factors in our multivariable analyses did not affect the magnitude of our findings. Finally, because patients with a cancer diagnosis at study entry are likely to have received chemotherapy or radiation therapy, which could result in clonal expansion, we also conducted a sub-group analysis in patients without cancer at baseline ( $n=1541$ ). The relationship between CHIP and  $\ln(\text{CRP})$  in this cohort remained unchanged.

## **Discussion**

Young *et al.* documented mutations in *TET2* or *DNMT3A* at very low frequencies in 95% of individuals aged 50 to 60 years, suggesting that these genes are almost ubiquitously mutated after the age of 50.<sup>22</sup> However, only a fraction of mutated individuals will have clonal expansion of a magnitude corresponding to CHIP,<sup>10</sup> and an associated increased risk of hematological cancers and cardiovascular disease. It is therefore of prime importance to identify the factors associated with clonal progression. Thus, we evaluated the hypothesis that inflammation is associated with clonal development of CHIP-related mutations in a single center cohort of 1887 individuals aged over 70, comprising individuals with (1359) and without (528) CAD at baseline. We showed that hs-CRP is significantly higher in subjects with CHIP than in those without CHIP, supporting the hypothesis that there is a link between inflammation and CHIP's pathogenesis in aging individuals (Figure 3).

We selected patients aged above 70 to maximize the prevalence of CHIP, which was indeed 22.6% in this cohort. Surprisingly, patients without CAD at baseline had slightly higher prevalence of CHIP than those with CAD at baseline (25.3% vs 21.5%,  $P<.075$ ). Since CHIP is associated with an increased risk of CAD, we expected a higher prevalence of CHIP in this subgroup. We speculate that selecting patients aged  $>70$  years-old may have systematically introduced a survival bias, where CHIP carriers with CAD may have been left out due to earlier death. It is also intriguing that the prevalence of *TET2* mutations was significantly lower in the CAD cohort than in the non-CAD subgroup. It is possible that specific CHIP-associated genes have a different survival impact. Expanding the cohort to include younger patients may ultimately allow us to answer this question.

The primary objective of this study was to document an association between inflammation using hs-CRP and CHIP carrier status. CRP was discovered in 1930 in the serum of patients with pneumococcal pneumonia.<sup>23</sup> It is a pentraxin produced by the liver and an acute phase

reactant,<sup>24</sup> IL-6, IL-1 and TNF $\alpha$  are the main inducers of CRP.<sup>25,26</sup> CRP measurement is commonly used to evaluate tissue injury, infection, and inflammatory diseases.<sup>27</sup> The development of high-sensitivity assays<sup>28</sup> has been invaluable for the investigation of low grade chronic inflammation in different human diseases.<sup>27</sup> The value of hs-CRP as a predictor of cardiovascular events was demonstrated 20 years ago.<sup>29</sup> For example, in a cohort of 28,263 normal women, hs-CRP measured only once at baseline was the strongest predictor of coronary heart disease death, nonfatal myocardial infarction, stroke, or the need for coronary revascularization procedures amongst other biomarkers including IL-6. We adopted a similar approach using baseline hs-CRP measurement in a cohort of patients with and without CAD.

The association between hs-CRP and CHIP was demonstrated in both univariable and multivariable analyses for the entire study population and for the cohort with CAD. This demonstration was made using an adequately large, single center, cohort of older individuals, and a commonly available biomarker of inflammation (Figure 2). The differences in hs-CRP levels between CHIP and non-CHIP carriers within non-CAD cohorts was slightly lower, and the difference did not reach statistical significance. This might be related to the fact that our non-CAD cohort was smaller than the CAD counterpart (528 vs 1359) or that the influence of inflammation as measured by hs-CRP has a greater influence on CHIP development in CAD patients. Interestingly, two recent studies did not identify a clear relationship between hs-CRP and CHIP. Bick *et al.* analyzed exome sequences from 35,416 individuals from the UK biobank without prevalent cardiovascular disease (similar to our non-CAD cohort).<sup>30</sup> They reported a modest and significant increase in hs-CRP (2.49 vs 2.85, P=0.03) in univariate analysis, but this association was lost in the covariate-adjusted model (P=0.08). Further, in a recent TOPMed communication<sup>31</sup>, Bick *et al.* identified in data from diverse genetic cohorts (n=32) that CHIP was associated with IL-6 (P=0.0035, n=11762), IL-1b (P= 2.4 X10<sup>-4</sup>, n= 598), but not with hs-CRP (P=0.10, n=22092). These discordant results are surprising as both IL-6 and IL-1 are inducers of CRP production by the liver.<sup>25</sup> It is possible that CRP is not a sensitive or specific inflammation biomarker of CHIP compared to IL-6 or IL-1, and/or that our single center and laboratory approach increased the capacity to identify a true relationship in contrast to large heterogeneous cohorts. Nevertheless, these studies confirm the association between biomarkers of inflammation and CHIP.

The demonstration of an association between inflammation and CHIP in human subjects does not address the causal relationship between the two entities. Does inflammation promote clonal expansion of CHIP or do CHIP clone(s) drive inflammatory response? Zhang *et al.* demonstrated that myeloid derived cells with loss of *TET2* maintain higher expression of IL-6, and that *TET2* is required to resolve inflammation and specifically repressed IL-6.<sup>32</sup> Fuster *et al.* studied clonal expansion of *tet2*-mutant cells in atherosclerosis-prone mice and demonstrated that *TET2*-deficient macrophages exhibited an increase in NLRP inflammasome mediated IL-1  $\beta$  secretion.<sup>33</sup> Interestingly, Bick *et al.* demonstrated that a relatively common coding mutation in the IL6-Receptor (pAsp358Ala) which significantly reduces signalization of IL-6, attenuated cardiovascular risk of CHIP carriers vs non-CHIP carriers,<sup>30</sup> supporting a causal link between CHIP, inflammation and CAD. Taken together this indicates that CHIP's mutation increase inflammation cytokines profile, at least at the cellular level. However, there are also several lines of evidence that suggest that increased inflammation promotes clonal expansion of mutated HSC (reviewed in<sup>34</sup>). In a key experiment, Cai *et al.* showed a rapid increase in the frequency and absolute number of Tet2-KO mature myeloid cells and HSCs in response to inflammatory stress, and demonstrated an enhanced production of inflammatory

cytokines, including interleukin-6 (IL-6) by these cells.<sup>35</sup> Taken together, these studies suggest that extrinsic inflammation mediators may selectively support clonal expansion of mutated HSCs which will eventually contribute to excessive cytokine release perpetuating a feedback loop.<sup>34</sup>

The difference of hs-CRP levels between CHIP carriers and non-CHIP carriers was relatively modest ( $\pm 20\%$ ). However, a small quantitative difference in inflammatory cytokines may have a lever effect if it is linked with increase receptor expression. Such a model has been demonstrated for chronic myelogenous leukemia (CML) in which stem cells have increased IL-1 receptor activity,<sup>36</sup> and increased proliferation and survival in a proinflammatory environment.<sup>37</sup> Furthermore, in a recent meta-analysis of inflammatory cytokine profile of 697 individuals with myelodysplastic syndromes versus controls, Shi *et al.* demonstrated a small but significant increase levels of TNF, IL-8 and IL-6.<sup>38</sup> In a cohort of patients with ulcerative colitis, Zhang *et al.* documented a small but significantly increased level of serum INF $\gamma$  but not TNF- $\alpha$  in subjects with *DNMT3A* mutation.<sup>39</sup> Therefore, it is possible that small but long standing difference in inflammation level may have a biological impact. Interestingly, treatment with the anti-IL1b antibody canakinumab led to significant reductions of not only cardiovascular disease,<sup>40</sup> but also incident lung cancer in patients with CAD.<sup>41</sup> Furthermore, the recent demonstration that utilization of a low dose of the anti-inflammatory drug colchicine reduces cardiovascular events after MI may further justify the prospective evaluation of such approaches in CHIP carriers.<sup>42</sup>

This study also allowed us to make additional observations. CHIP regroups alterations occurring in several different genes, but clinical outcomes have been usually estimated according to the CHIP-carrier status<sup>5,6</sup> rather than gene-specific estimation. Recently, some studies have demonstrated gene specificities in regards to risk of progression to acute myeloid leukemia,<sup>43</sup> genetic predisposition<sup>11</sup> or lineage restriction<sup>44</sup>. We wanted to study the relationship between inflammation and specific CHIP-associated genes. We were able to document a positive association between *DNMT3A* and hsCRP in the entire cohort. No other association was statistically significant, although quantitative differences were similar for *DNMT3A*, *TET2* and *ASXL1*. A significantly larger cohort would be necessary to address this question. We documented a significant association between heart failure and CHIP carrier status. This is in line with several previous observations in mice<sup>45</sup> and humans.<sup>46,47</sup> In contrast, we did not find an association with smoking in this cohort where prevalence was very low.

In conclusion, this study highlights the role of inflammation in CHIP. The etiology of CHIP is probably multifactorial and several other factors need to be identified. Clinical trials should test whether anti-inflammatory therapy can reduce CHIP progression and related diseases.

## Acknowledgements

This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) and by the Leukemia Lymphoma Society of Canada (LLSC).

## Authorship contributions

Contributions: LB, JCT, MPD conceived the study. MB performed and analyzed the sequencing data with the support of SA and VB. MB generated figures. MS performed statistical analyses with the support of YFZ and SP under the supervision of MPD. LM, MM, RH and JB revised the manuscript and provided insightful suggestions.

## Conflict-of-interest disclosure

The authors declare no competing financial interests.



## References.

1. Fey MF, Liechti-Gallati S, von Rohr A, et al. Clonality and X-inactivation patterns in hematopoietic cell populations detected by the highly informative M27 beta DNA probe. *Blood*. 1994;83(4):931-938.
2. Busque L, Mio R, Mattioli J, et al. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood*. 1996;88(1):59-65.
3. Ayachi S, Buscarlet M, Busque L. 60 Years of clonal hematopoiesis research: From X-chromosome inactivation studies to the identification of driver mutations. *Exp Hematol*. 2020.
4. Busque L, Patel JP, Figueroa ME, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet*. 2012;44(11):1179-1181.
5. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371(26):2488-2498.
6. Genovese G, Kahler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. 2014;371(26):2477-2487.
7. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*. 2014;20(12):1472-1478.
8. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*. 2017;130(6):742-752.
9. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med*. 2017;377(2):111-121.
10. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9-16.
11. Buscarlet M, Provost S, Zada YF, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood*. 2017;130(6):753-762.
12. Hansen JW, Pedersen DA, Larsen LA, et al. Clonal hematopoiesis in elderly twins: concordance, discordance and mortality. *Blood*. 2019.
13. Fabre MA, McKerrell T, Zwiebel M, et al. Concordance for clonal hematopoiesis is limited in elderly twins. *Blood*. 2019.
14. Franceschi C, Garagnani P, Vitale G, Capri M, Salvioli S. Inflammaging and 'Garb-aging'. *Trends Endocrinol Metab*. 2017;28(3):199-212.
15. Busque L, Buscarlet M, Mollica L, Levine RL. Concise Review: Age-Related Clonal Hematopoiesis: Stem Cells Tempting the Devil. *Stem Cells*. 2018.
16. Abegunde SO, Buckstein R, Wells RA, Rauh MJ. An inflammatory environment containing TNFalpha favors Tet2-mutant clonal hematopoiesis. *Exp Hematol*. 2018;59:60-65.

17. Meisel M, Hinterleitner R, Pacis A, et al. Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. *Nature*. 2018;557(7706):580-584.
18. Cook EK, Izukawa T, Young S, et al. Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. *Blood Adv*. 2019;3(16):2482-2486.
19. Ridker PM. From C-Reactive Protein to Interleukin-6 to Interleukin-1: Moving Upstream To Identify Novel Targets for Atheroprotection. *Circ Res*. 2016;118(1):145-156. Montreal Heart Institute Hospital Biobank. Vol. 2018.
21. Tardif JC, Rhainds D, Brodeur M, et al. Genotype-Dependent Effects of Dalcetrapib on Cholesterol Efflux and Inflammation: Concordance With Clinical Outcomes. *Circ Cardiovasc Genet*. 2016;9(4):340-348.
22. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat Commun*. 2016;7:12484.
23. Tillett WS, Francis T. Serological Reactions in Pneumonia with a Non-Protein Somatic Fraction of Pneumococcus. *J Exp Med*. 1930;52(4):561-571.
24. Oliveira EB, Gotschlich C, Liu TY. Primary structure of human C-reactive protein. *J Biol Chem*. 1979;254(2):489-502.
25. Moshage HJ, Roelofs HM, van Pelt JF, et al. The effect of interleukin-1, interleukin-6 and its interrelationship on the synthesis of serum amyloid A and C-reactive protein in primary cultures of adult human hepatocytes. *Biochem Biophys Res Commun*. 1988;155(1):112-117.
26. Yoshizaki K. Pathogenic role of IL-6 combined with TNF-alpha or IL-1 in the induction of acute phase proteins SAA and CRP in chronic inflammatory diseases. *Adv Exp Med Biol*. 2011;691:141-150.
27. Pepys MB, Berger A. The renaissance of C reactive protein. *BMJ*. 2001;322(7277):4-5.
28. Rifai N, Ridker PM. High-sensitivity C-reactive protein: a novel and promising marker of coronary heart disease. *Clin Chem*. 2001;47(3):403-411. 9
29. Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N Engl J Med*. 2000;342(12):836-843.
30. Bick AG, Pirruccello JP, Griffin GK, et al. Genetic IL-6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. *Circulation*. 2019.
31. Bick AG, Weinstock JS, Nandakumar SK, et al. Inherited Causes of Clonal Hematopoiesis of Indeterminate Potential in TOPMed Whole Genomes. *bioRxiv*. 2019:782748.
32. Zhang Q, Zhao K, Shen Q, et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature*. 2015;525(7569):389-393.

33. Fuster JJ, MacLauchlan S, Zuriaga MA, et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science*. 2017;355(6327):842-847.
34. King KY, Huang Y, Nakada D, Goodell MA. Environmental influences on clonal hematopoiesis. *Exp Hematol*. 2019.
35. Cai Z, Kotzin JJ, Ramdas B, et al. Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis. *Cell Stem Cell*. 2018;23(6):833-849 e835.
36. Zhang B, Li M, McDonald T, et al. Microenvironmental protection of CML stem and progenitor cells from tyrosine kinase inhibitors through N-cadherin and Wnt-beta-catenin signaling. *Blood*. 2013;121(10):1824-1838.
37. Zhang B, Chu S, Agarwal P, et al. Inhibition of interleukin-1 signaling enhances elimination of tyrosine kinase inhibitor-treated CML stem cells. *Blood*. 2016;128(23):2671-2682.
38. Shi X, Zheng Y, Xu L, Cao C, Dong B, Chen X. The inflammatory cytokine profile of myelodysplastic syndromes: A meta-analysis. *Medicine (Baltimore)*. 2019;98(22):e15844.
39. Zhang CRC, Nix D, Gregory M, et al. Inflammatory cytokines promote clonal hematopoiesis with specific mutations in ulcerative colitis patients. *Exp Hematol*. 2019;80:36-41 e33.
40. Ridker PM, Everett BM, Thuren T, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med*. 2017;377(12):1119-1131.
41. Ridker PM, MacFadyen JG, Thuren T, et al. Effect of interleukin-1beta inhibition with canakinumab on incident lung cancer in patients with atherosclerosis: exploratory results from a randomised, double-blind, placebo-controlled trial. *Lancet*. 2017;390(10105):1833-1842.
42. Tardif JC, Kouz S, Waters DD, et al. Efficacy and Safety of Low-Dose Colchicine after Myocardial Infarction. *N Engl J Med*. 2019.
43. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*. 2018;559(7714):400-404.
44. Buscarlet M, Provost S, Zada YF, et al. Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. *Blood*. 2018;132(3):277-280.
45. Sano S, Oshima K, Wang Y, et al. Tet2-Mediated Clonal Hematopoiesis Accelerates Heart Failure Through a Mechanism Involving the IL-1beta/NLRP3 Inflammasome. *J Am Coll Cardiol*. 2018;71(8):875-886.
46. Dorsheimer L, Assmus B, Rasper T, et al. Association of Mutations Contributing to Clonal Hematopoiesis With Prognosis in Chronic Ischemic Heart Failure. *JAMA Cardiol*. 2019;4(1):25-33.

47. Dorsheimer L, Assmus B, Rasper T, et al. Hematopoietic alterations in chronic heart failure patients by somatic mutations leading to clonal hematopoiesis. *Haematologica*. 2019. 10.

**Table 1.** Baseline characteristics of the study population, according to clonal hematopoiesis of indeterminate potential (CHIP) status (univariate analysis).

	All patients		Without baseline coronary artery disease		With baseline coronary artery disease	
	CHIP+	CHIP-	CHIP+	CHIP-	CHIP+	CHIP-
<i>N</i> (%)	427(22.6)	1460 (77.4)	134(25.3)	394(74.7)	293(21.5)	1066(78.5)
Age, yr Mean (SD)	75.0 (3.4)*	74.4 (3.3)	74.9 (3.2)*	74.1 (3.3)	75.0 (3.5)*	74.5(3.3)
Weight, kg Mean (SD)	77.5 (15.5)	78.2 (14.4)	77.0 (16.5)	75.0 (14.9)	77.8 (15.0)	79.4 (14.1)
Height, m Mean (SD)	1.65 (0.09)	1.66 (0.08)	1.64 (0.10)	1.63 (0.09)	1.66 (0.08)	1.67 (0.08)
BMI, Mean (SD)	28.197 (5.041)	28.245 (4.423)	28.390 (5.265)	27.915 (4.446)	28.108 (4.942)	28.367 (4.410)
Male, <i>n</i> (%)	292 (68.4%)	1044 (71.5%)	67 (50.0%)	183 (46.4%)	225 (76.8%)	861 (80.8%)
<b>Cardiovascular risk factors – <i>n</i> (%)</b>						
Dyslipidemia	361 (84.5%)*	1287 (88.3%)	84 (62.7%)	267 (67.9%)	277 (94.5%)	1020 (95.8%)
Hypertension	327 (76.8%)	1095 (75.3%)	82 (61.7%)	239 (60.7%)	245 (83.6%)	856 (80.7%)
Diabetes	120 (28.1%)	408 (28.0%)	26 (19.4%)	73 (18.5%)	94 (32.1%)	335 (31.5%)
Ever smoker	301 (70.7%)	988 (67.7%)	79 (59.4%)	218 (55.3%)	222 (75.8%)	770 (72.3%)
Current smoker	21 (4.9%)	71 (4.9%)	1 (0.8%)*	18 (4.6%)	20 (6.8%)	53 (5.0%)
<b>CVD history – <i>n</i> (%)</b>						
Coronary heart disease	293 (68.6%)	1066 (73.0%)	-	-	-	-
Previous MI	193 (45.3%)	649 (44.5%)	0 (0.0%)	0 (0.0%)	193 (66.1%)	649 (60.9%)
Previous PCI	130 (30.4%)*	550 (37.7%)	0 (0.0%)	0 (0.0%)	130 (44.4%)*	550 (51.6%)
Previous CABG	163 (38.2%)	601 (41.2%)	0 (0.0%)	0 (0.0%)	163 (55.6%)	601 (56.4%)
Stroke	42 (9.9%)	154 (10.6%)	0 (0.0%)	0 (0.0%)	42 (14.4%)	154 (14.5%)
Angina	250 (58.7%)	924 (63.5%)	0 (0.0%)	0 (0.0%)	250 (85.6%)	924 (87.1%)
PVD	85 (19.9%)	306 (21.0%)	0 (0.0%)	0 (0.0%)	85 (29.0%)	306 (28.8%)
Angiography	279 (65.3%)	1018 (69.7%)	15 (11.2%)	52 (13.2%)	264 (90.1%)	966 (90.6%)
CHF	72 (16.9%)	195 (13.4%)	1 (0.7%)	2 (0.5%)	71 (24.3%)*	193 (18.2%)
<b>Medication – <i>n</i> (%)</b>						
Aspirin	310 (72.6%)	1123 (76.9%)	66 (49.3%)	193 (49.0%)	244 (83.3%)	930 (87.2%)
Antiplatelet	329 (77.0%)	1164 (79.7%)	67 (50.0%)	199 (50.5%)	262 (89.4%)	965 (90.5%)
Statin	360 (84.3%)	1276 (87.4%)	83 (61.9%)	259 (65.7%)	277 (94.5%)	1017 (95.4%)
ACE	144 (33.7%)	538 (36.8%)	24 (17.9%)	77 (19.5%)	120 (41.0%)	461 (43.2%)
ARB	157 (36.8%)	496 (34.0%)	43 (32.1%)	130 (33.0%)	114 (38.9%)	366 (34.3%)
Beta-blockers	287 (67.2%)	970 (66.4%)	55 (41.0%)	143 (36.3%)	232 (79.2%)	827 (77.6%)
<b>Lipids, mmol/L – mean (SD)</b>						
Total cholesterol	3.89 (1.00)	3.86 (0.98)	4.44 (1.02)	4.59 (1.11)	3.63 (0.88)	3.59 (0.76)
Triglycerides	1.92 (1.03)	1.99 (1.02)	1.92 (1.08)	2.04 (1.03)	1.93 (1.00)	1.97 (1.01)
HDL	1.23 (0.37)	1.22 (0.36)	1.32 (0.36)	1.34 (0.42)	1.19 (0.36)	1.18 (0.33)
LDL	1.80 (0.84)	1.75 (0.83)	2.28 (0.89)	2.36 (0.99)	1.58 (0.71)	1.52 (0.62)
<b>hs-CRP mg/L – median (IQR), min-max</b>						
hs-CRP	1.60 (2.96)*,	1.41 (2.29),	1.55 (2.99),	1.45 (2.23),	1.60 (2.88)*,	1.39 (2.30),

	0.09-115.0	0.08-80.70	0.09-92.80	0.08-61.30	0.09-115.0	0.08-80.70
<b>Cancer history – n (%)</b>						
Any cancer	89 (20.8%)	257 (17.6%)	26 (19.4%)	67 (17.0%)	63 (21.5%)	190 (17.8%)
Hematological cancer	6 (1.4%)	17 (1.1%)	3 (2.2%)*	0 (0.0%)	3 (1.0%)	17 (1.5%)
<b>Incident/recurrent cancer – n (%)</b>						
Any cancer	34 (7.9%)	111 (7.6%)	11 (8.2%)	25 (6.3%)	23 (7.8%)	86 (8.0%)
Hematological cancer	7 (1.6)	21 (1.4%)	0 (0.0%)	2 (0.5%)	2 (0.6%)	4 (0.3%)

ACE: angiotensin-converting enzyme; ARB: angiotensin receptor blockers; BMI: body mass index; CABG: coronary artery bypass graft; CAD: coronary artery disease; CRP: C-reactive protein; CVD: cardiovascular disease; HDL: high density lipoprotein; HF: heart failure; IQR: interquartile range; LDL: low density lipoprotein; MI: myocardial infarction; PCI: percutaneous coronary revascularization intervention; STD: standard deviation.

\* When  $P < 0.05$  per Fisher's Exact test or Kruskal-Wallis two-sided significance test.

**Table 2.** Association between CHIP and hs-CRP at study entry (multivariate analysis).

	All patients (n = 1887)		Without baseline coronary artery disease (n = 528)		With baseline coronary artery disease (n = 1359)	
	CHIP+	CHIP-	CHIP+	CHIP-	CHIP+	CHIP-
<b>Any mutations</b>						
n	427	1460	134	394	293	1066
hs-CRP GM, mg/L (IQR)	1.85 (0.81-3.77)	1.5 (0.71-3.00)	1.86 (0.83-3.82)	1.51 (0.77-3.02)	1.84 (0.81-3.72)	1.49 (0.69-3.01)
Back-transformed $\beta$ (95% CI)*	1.21 (1.08-1.36)		1.15 (0.94-1.40)		1.22 (1.06-1.41)	
% difference	21%		15%		22%	
P value*	0.001		0.178		0.005	
<b>DNMT3A</b>						
n	218	1669	66	462	152	1207
hs-CRP GM, mg/L (IQR)	1.82 (0.80-3.79)	1.54 (0.73-1.43)	1.91 (0.81-4.72)	1.55 (0.77-3.10)	1.79 (0.78-3.66)	1.54 (0.71-3.13)
Back-transformed $\beta$ (95% CI)*	1.17 (1.01-1.36)		1.23 (0.95-1.59)		1.15 (0.95-1.38)	
% difference	17%		23%		15%	
P value*	0.04		0.125		0.148	
<b>TET2</b>						
n	115	1772	48	480	67	1292
hs-CRP GM, mg/L (IQR)	1.78 (0.81-3.46)	1.56 (0.73-3.19)	1.77 (0.82-2.77)	1.58 (0.77-3.27)	1.79 (0.76-4.01)	1.55 (0.71-3.17)
Back-transformed $\beta$ (95% CI)*	1.13 (0.92-1.38)		0.92 (0.68-1.24)		1.24 (0.94-1.62)	
% difference	13%		-8%		24%	
P value*	0.253		0.585		0.128	
<b>ASXL1</b>						
n	52	1835	14	514	38	1321
hs-CRP GM, mg/L (IQR)	1.93 (0.88-3.64)	1.56 (0.73-3.19)	1.77 (0.96-2.45)	1.59 (0.77-3.27)	1.99 (0.83-4.25)	1.55 (0.71-3.17)
Back-transformed $\beta$ (95% CI)*	1.26 (0.94-1.69)		1.09 (0.64-1.86)		1.31 (0.92-1.87)	
% difference	26%		9%		31%	
P value	0.128		0.751		0.138	
<b>PPM1D</b>						
n	38	1849	9	519	29	1330
hs-CRP GM, mg/L (IQR)	1.64 (0.67-3.43)	1.57 (0.74-3.19)	1.84 (0.60-5.09)	1.59 (0.78-3.23)	1.59 (0.74-3.28)	1.56 (0.71-3.18)
Back-transformed $\beta$ (95% CI)*	0.98 (0.69-1.39)		1.14 (0.59-2.20)		0.95 (0.63-1.43)	
% difference	-2%		14%		-5%	
P value	0.909		0.705		0.805	

CHIP: clonal hematopoiesis of indeterminate potential, CHIP+: carrier of mutation (s). CHIP-: Non-carrier of mutation, CAD: coronary artery disease, GM: geometric mean, IQR: interquartile range, Std. Error: Standard error  
N.B. The sample size related to carriers of mutations in *TP53*, *SF3B1*, *SRSF2*, *CBL*, and *JAK2* were insufficient to perform analyses in these sub-groups. The presented P values for mutations yes vs. no status differences in log-transformed changes in hs-CRP.

\* Generalized linear regression model-derived  $\ln(\text{hs-CRP})$  with adjustment for age, sex, and coronary artery disease status at baseline-adjusted where CHIP- was the referent category.

Figure 1. **Prevalence and distribution of somatic mutations.** (A) Description of the cohort (n=1887). (B) Prevalence of the 540 somatic mutations identified in 427 individuals of the cohort. (C) Mutation signature of all single nucleotide substitutions (see Supplemental Figure 2). (D) Variant allele fraction (VAF) distribution of all somatic mutations. (E) Co-occurrence of the all somatic mutations in the 427 mutated individuals. Darker shades represent double mutation in the same gene. (F) Contribution of individual genes to the total number of observed somatic mutations. (G) Average VAF of somatic mutation for each gene.



Figure 2. **High sensitivity C-reactive protein concentration (mg/L) in CHIP- and CHIP+ individuals.** Box-and-Whisker plot of the hs-CRP concentration (mg/L) in all CHIP- (n=1460) and CHIP+ (n=427) individuals of the cohort. The center line denotes the statistical median, while the box contains the first (Q1) to third (Q3) quartiles. Whiskers mark 3/2 times the interquartile range. Outliers are not represented.

\* When  $P < 0.05$  per Kruskal-Wallis two-sided significance test.

Figure 3. **Risk factors for CHIP.** CHIP is age dependent. Several factors may influence its development. These include genetic predisposition, exposure to genotoxic and immune attack (reviewed in<sup>15</sup>). This study further confirms the potential role of an inflammatory state in CHIP development.

Figure 1.

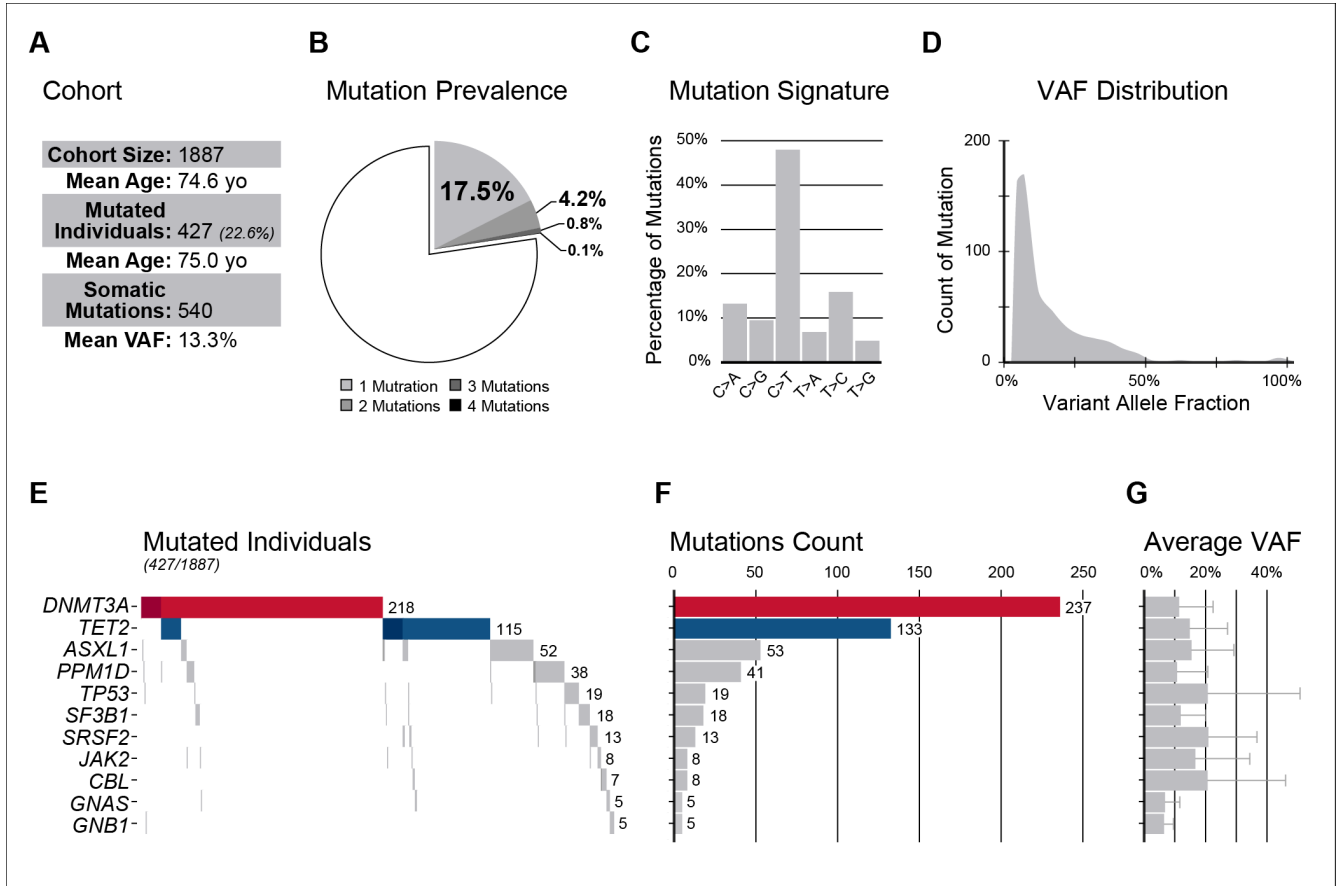


Figure 2.

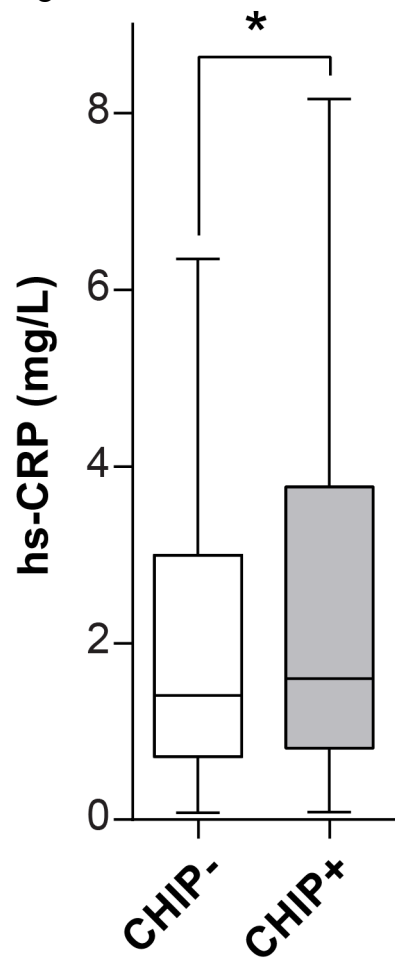


Figure 3.

