



GENHOTEL-EA3886



Thèse de doctorat

Ecole doctorale Génomes Aux Organismes

Présentée par

Monsieur Mohamed Sahbi Ben Kilani

Pour obtenir le grade de

Docteur de l'Université d'Evry Val d'Essonne

Domaine :

Génétique Humaine

Sujet de la thèse :

**Analyses des Variations de Nombre de Copies de gènes candidats dans
la Polyarthrite Rhumatoïde**

Thèse présentée et soutenue à l'Université d'Evry le 10/12/2014 devant le jury composé de :

Pr Henri-Jean GARCHON
Pr Jean-Michel DUPONT
Pr Laetitia MICHOU
Dr Elham HASSEN
Dr Sophie GARNIER
Pr Elisabeth PETIT-TEIXEIRA

Rapporteur
Rapporteur
Examinatrice
Examinatrice
Examinatrice
Directrice de thèse

Nom du Laboratoire: GenHotel-EA3886
2 rue Gaston Crémieux, 91057, Evry CEDEX

« La quête du savoir n'a plus comme finalité l'innovation, mais la survie ! »

Remerciements

Je tiens à adresser mes sincères remerciements :

Aux membres du jury qui ont accepté d'évaluer mon travail de thèse.

Au Professeur Jean-Michel Dupont et au Professeur Henri-Jean Garchon d'avoir accepté d'être les rapporteurs de ce manuscrit.

Au Pr Laetitia Michou, le Dr Elham Hassen et le Dr Sophie Garnier pour avoir accepté d'examiner mon mémoire et faire partie de mon jury de thèse.

Je remercie chaleureusement ma directrice de thèse, Pr Elisabeth Petit-Teixeira pour son accueil il y a déjà 5 ans, pour son soutien, sa gentillesse et sa bienveillance durant toutes ces années. Ses conseils m'ont été précieux sur le plan scientifique et sa confiance m'a permis d'élargir mes compétences et gagner en assurance. Merci infiniment pour ton écoute de tous les instants, ton amitié et ton dévouement que ce soit sur le plan scientifique ou humain.

Un grand merci au Dr Valérie Chaudru pour son soutien depuis son arrivée au laboratoire, pour sa gentillesse, sa patience et son aide précieuse dans la partie statistique de ce travail. Je remercie également Pr Javier Perea pour ses remarques et ses suggestions qui m'ont permis d'enrichir mes travaux et d'apporter des améliorations à la rédaction du mémoire de thèse.

Au Pr François Cornélis pour son accueil chaleureux et pour ses conseils avisés durant toutes ces années.

A tous les anciens membres de l'équipe de GenHotel qui sont partis en cours de route. Merci pour ces supers souvenirs, ces moments de complicité et cette ambiance à la fois sérieuse et conviviale.

Au Dr Robert Olasso, pour ses suggestions techniques et pour les discussions scientifiques mais aussi son soutien pour la suite de ma carrière.

Je voudrais remercier tous les stagiaires qui sont passés au laboratoire et particulièrement ceux que j'ai encadrés pour leur contribution à ce travail, pour leur dynamisme, et leur motivation.

Merci également à tous les professeurs du département biologie de l'université d'Evry qui m'ont enseigné pendant mon cursus. Je remercie aussi ceux avec qui j'ai fait les TP qui se sont bien déroulés pendant ces trois années d'enseignements.

Je tiens aussi à remercier chaleureusement Mme Florence Hervy, Secrétaire département biologie, qui m'a facilité les démarches administratives pour mener à bien mon travail de thèse et toujours avec le sourire et la bonne humeur.

Je dédie ce travail à ma femme Ines qui a toujours été là pour me soutenir surtout dans les moments de doutes et qui m'a toujours poussé vers l'avant pour réaliser mes rêves. Merci pour ta patience, ta joie et pour tout le bonheur que tu m'apportes jour après jour. Je dédie également cette thèse à mon ange Nahawend qui est arrivée alors que j'entamais ma deuxième année de thèse et a inondé ma vie de joie et de bonheur. Tu me donnes le courage et la force de continuer et d'aller de l'avant pour t'offrir, comme le disait Cabrel, "un ciel aussi joli que des milliers de roses".

J'offre également ce travail à mes chers parents qui ont fait beaucoup de sacrifices pour que je puisse être là où je suis aujourd'hui et également à ma sœur qui m'a toujours soutenu. Merci pour votre patience, votre confiance et tout l'amour que vous me donnez.

Je dédie également cette thèse à chaque membre de ma grande et immense famille, ceux qui m'ont soutenu et encouragé à chaque étape de ma vie. Je remercie également ma belle-famille qui m'a été d'un grand réconfort et d'une aide précieuse durant ces dernières années. A ceux

qui ne sont plus parmi nous mais qui ont toujours cru en moi et même loin de moi, c'est grâce à leur amour que je suis devenu ce que je suis.

Enfin, je remercie du fond du cœur tous mes amis d'ici et de Tunisie qui ont été là pour me remonter le moral durant ces derniers mois difficiles et pour me changer les idées quand j'étais submergé par le travail. Merci également à tous de m'avoir supporté et de m'avoir encouragé pendant toutes ces années.

Résumé

Les Variations de Nombre de Copies (CNVs) sont des séquences supérieures à 1kb ayant subi une délétion, une duplication ou une inversion de fragments d'ADNs et sont présents avec une fréquence de 12% dans le génome humain. Leur caractérisation ainsi que leur association dans les maladies complexes sont, à ce jour, en plein essor. Nous avons donc entrepris l'étude des CNVs de gènes candidats dans les familles de patients atteints de Polyarthrite Rhumatoïde (PR) en utilisant plusieurs outils de biologie moléculaire.

Les CNVs de quatre gènes candidats (*GSTM1*, *GSTT1*, *FCGR3B* et *CCL3L1*) ont été investigués par PCR classique, qPCR ou *Droplet Digital PCR* (ddPCR). Mais très vite, la PCR classique a montré qu'elle n'est pas applicable à tous les gènes et que la qPCR était peu sensible et peu reproductible. Par contre, grâce à la ddPCR, nous avons tout d'abord pu quantifier avec précision les CNVs des gènes *GSTM1*, *FCGR3B* et *CCL3L1*. Par la suite, nous avons pu identifier les génotypes en nombre de copies pour le gène *GSTM1* et mettre en évidence une duplication en tandem qui a été confirmée par une *Long Range PCR*. La transmission des CNVs au sein des familles étudiées nous a ensuite permis d'identifier tous les génotypes en nombre de copies du gène *FCGR3B* et certains génotypes du gène *CCL3L1*. Des événements de recombinaison *de novo* ont été mis en évidence pour le gène *FCGR3B* ainsi que des duplications en tandem transmises des parents aux patients pour les trois gènes investigués par ddPCR. Aucune association n'a été mise en évidence entre la PR et les CNVs des quatre gènes candidats. Néanmoins, une transmission préférentielle de la délétion du gène *GSTT1* a été mise en évidence, après stratification, dans le sous-groupe de patients séropositifs pour le Facteur Rhumatoïde. Des tendances non significatives ont été observées dans les sous-groupes de cas index avec un âge de début de la maladie inférieur à 40 ans, avec présence de nodules (pour le gène *GSTM1*), et avec la présence d'auto-anticorps (pour les gènes *FCGR3B* et *CCL3L1*).

En conclusion, le développement de l'utilisation d'une technologie de pointe et l'accès à des échantillons familiaux sont des atouts indispensables pour caractériser les variants génomiques de nombre de copies, facteurs de risque potentiels des maladies complexes

Mots-clés : Variation de Nombre de Copies, Polyarthrite Rhumatoïde, Gènes candidats, *Droplet Digital PCR*, Transmission familiale.

Abstract

Copy Number Variations (CNVs) are sequences up to 1kb resulting from deletion, duplication and inversion of DNA regions and are present with a frequency of 12% in the human genome. CNVs characterization and association to complex diseases are still subjects of controversies. We then enhanced the study of CNVs for candidate genes in Rheumatoid Arthritis (RA) families, with different methodologies.

CNVs of four candidate genes (*GSTM1*, *GSTT1*, *FCGR3B* et *CCL3LI*) were investigated with standard PCR, quantitative PCR (qPCR) or with Droplet Digital PCR (ddPCR). First, we found that standard PCR was not applicable for all genes and that the qPCR was not sensitive and reproducible for CNVs quantification. Second, the ddPCR methodology allowed us to quantify CNVs of *GSTM1*, *FCGR3B*, and *CCL3LI* with high resolution and to characterize copy number genotypes of *GSTM1* gene, leading to the identification of tandem duplicated copies. A Long Range PCR confirmed this result. CNVs transmission in families allowed the identification of all copy number genotypes for *FCGR3B* gene and some of *CCL3LI* genotypes. *De novo* recombination events were highlighted for *FCGR3B* gene and transmission of the duplication from one parent to the offspring was observed for all genes characterized by ddPCR. No association between RA and candidate genes CNVs was found, but after stratification, we observed a significant preferential transmission of *GSTT1* deletion in the subgroup of patients seropositive for Rheumatoid Factor. We also showed a non-significant tendency in the subgroups of patient with an age at onset inferior to 40 years, with presence of nodules (for *GSTM1* gene), and with presence of auto-antibodies (for *FCGR3B* and *CCL3LI* gene).

In conclusion, digital PCR is currently the most adequate methodology to accurately genotype CNVs. Analysis of familial sample leads to the identification of duplication events and to the characterization of genotypes, essential for CNVs association studies in complex diseases.

Key words: Copy Number Variation, Rheumatoid Arthritis, Candidate genes, Droplet Digital PCR, Familial transmission.

Table des matières

REMERCIEMENTS	I
RESUME	IV
ABSTRACT	V
ABREVIATIONS	XVI
CHAPITRE 1 : LA POLYARTHRITE RHUMATOÏDE.....	1
I. DEFINITION	1
A. ASPECTS CLINIQUES DE LA PR.....	1
B. EPIDEMIOLOGIE DE LA PR.....	3
1. <i>Prévalence</i>	3
2. <i>Incidence</i>	3
3. <i>Mortalité et survie dans la PR</i>	4
C. PHYSIOPATHOLOGIE DE LA PR.....	4
1. <i>Les acteurs cellulaires impliqués dans la PR</i>	4
2. <i>Les acteurs intercellulaires : les cytokines</i>	8
3. <i>Acteurs de la signalisation cellulaire</i>	9
4. <i>Production d'auto-anticorps</i>	10
D. TRAITEMENTS.....	13
II. ETIOLOGIE DE LA POLYARTHRITE RHUMATOÏDE	15
A. LA COMPOSANTE GENETIQUE	16
1. <i>L'héritabilité</i>	16
2. <i>Approches des études génétiques</i>	16
3. <i>Le gène HLA-DR</i>	19
4. <i>Le gène PTPN22</i>	20
5. <i>Les autres facteurs de risque génétique mineurs</i>	21
B. LA COMPOSANTE ENVIRONNEMENTALE.....	22
1. <i>L'exposition au tabac</i>	22
2. <i>Les autres facteurs</i>	23

CHAPITRE 2 : LES VARIATIONS DE NOMBRE DE COPIES	24
I. DEFINITION DES VARIATIONS DE NOMBRE DE COPIES (CNVS).....	24
A. FORMATION DES CNVS.....	25
1. <i>Caractéristiques des CNVs</i>	25
2. <i>Exemple de réarrangements récurrents</i>	26
3. <i>Exemples de réarrangements non-récurrents</i>	28
B. MODES DE TRANSMISSION DES CNVS	30
C. EFFETS DES CNVS SUR L'EXPRESSION	31
D. METHODES D'IDENTIFICATION.....	32
1. <i>Analyses du génome entier</i>	32
2. <i>Analyses ciblées sur des gènes candidats</i>	43
E. CNVS ET MALADIES.....	53
1. <i>Les CNVs dans les maladies rares</i>	53
2. <i>Les CNVs dans les maladies fréquentes</i>	54
CHAPITRE 3 : LES CNVS DANS LA PR	58
I. LES CNVS ASSOCIES A LA PR PAR ETUDE DU GENOME ENTIER.....	58
II. LES CNVS ASSOCIES A LA PR PAR ETUDE DE GENES CANDIDATS	60
1. <i>La famille des Glutathion S-Transférase</i>	61
2. <i>La famille de gènes du Récepteur Fc Gamma</i>	70
3. <i>La famille des « chemokine (C-C motif) ligand »</i>	73
4. <i>Les autres gènes avec des CNVs dans la PR</i>	76
CHAPITRE 4 : LES OBJECTIFS	78
CHAPITRE 5 : PATIENTS & METHODES	80
I. PATIENTS	80
II. CARACTERISATION DES CNVS DU GENE <i>GSTM1</i>	80
A. IDENTIFICATION DE LA PRESENCE/ABSENCE DU GENE.....	80
B. IDENTIFICATION DU GENOTYPE PAR PCR CLASSIQUE.....	82
C. QUANTIFICATION DES CNVS	83
1. <i>Quantification par PCR quantitative</i>	83
2. <i>Quantification par le système QuantStudio 12K flex</i>	84
D. MISE EN EVIDENCE DE LA DUPLICATION EN TANDEM.....	85
1. <i>Mise en évidence par la PCR Long Range</i>	85

2.	<i>Mise en évidence par la PCR QX200 Droplet Digital</i>	86
III.	CARACTERISATION DES CNVS DU GENE <i>GSTT1</i>	88
A.	IDENTIFICATION DE LA PRESENCE/ABSENCE DU GENE	88
B.	IDENTIFICATION DU GENOTYPE	89
IV.	CARACTERISATION DES CNVS DU GENE <i>FCGR3B</i>	89
A.	QUANTIFICATION PAR QPCR.....	89
B.	QUANTIFICATION PAR DDPCR.....	90
V.	CARACTERISATION DES CNVS DU GENE <i>CCL3L1</i>	90
A.	MISE AU POINT PAR QPCR	90
B.	MISE AU POINT DE LA DDPCR	91
1.	<i>Caractérisation des CNVs</i>	92
VI.	ETUDE D'ASSOCIATIONS DANS LES FAMILLES TRIO	92
A.	LE TEST DE DESEQUILIBRE DE TRANSMISSION (TDT)	92
B.	TEST DU RISQUE RELATIF DES GENOTYPES (GRR).....	93
C.	ASSOCIATION ENTRE LES FACTEURS ASSOCIES A LA PR ET LES CNVS	93
CHAPITRE 6 : RESULTATS	95
I.	CARACTERISATION DES CNVS DU GENE <i>GSTM1</i>	95
A.	IDENTIFICATION DE LA PRESENCE/ABSENCE DU GENE	95
B.	IDENTIFICATION DU GENOTYPE	95
C.	QUANTIFICATION DES CNVS	96
1.	<i>Quantification par PCR quantitative</i>	96
2.	<i>Quantification par le système QuantStudio 12K flex</i>	97
D.	MISE EN EVIDENCE DE LA DUPLICATION EN TANDEM.....	99
1.	<i>Mise en évidence par la PCR Long Range</i>	99
2.	<i>Mise en évidence par la PCR QX200 Droplet Digital</i>	99
3.	<i>Comparaison entre qPCR, dPCR et ddPCR</i>	103
E.	TRANSMISSION DES CNVS DANS LES FAMILLES	103
F.	ANALYSE FAMILIALE DE L'ASSOCIATION <i>GSTM1</i> ET PR	104
G.	RELATION ENTRE CNVS DU GENE ET FACTEURS DE LA PR	106
II.	CARACTERISATION DES CNVS DU GENE <i>GSTT1</i>	107
A.	IDENTIFICATION DE LA PRESENCE/ABSENCE.....	107
B.	IDENTIFICATION DU GENOTYPE	108
C.	ANALYSE FAMILIALE DE L'ASSOCIATION ENTRE <i>GSTT1</i> ET LA PR.....	109

D.	RELATION ENTRE LES CNVS DE <i>GSTT1</i> ET LES FACTEURS DE LA PR	111
III.	CARACTERISATION DES CNVS DU GENE <i>FCGR3B</i>	112
A.	QUANTIFICATION PAR QPCR.....	112
B.	QUANTIFICATION PAR DDPCR.....	112
1.	<i>Test d'interférence</i>	112
2.	<i>Test de reproductibilité</i>	113
3.	<i>Tests de digestion</i>	114
4.	<i>Caractérisation des CNVs</i>	115
C.	COMPARAISON QPCR ET DDPCR.....	115
D.	TRANSMISSION DES CNVS DANS LES FAMILLES	116
E.	ANALYSE FAMILIALE DE L'ASSOCIATION ENTRE <i>FCGR3B</i> ET LA PR	117
F.	RELATION ENTRE LES CNVS DE <i>FCGR3B</i> ET LES FACTEURS DE LA PR.....	118
IV.	CARACTERISATION DES CNVS DU GENE <i>CCL3L1</i>	120
A.	MISE AU POINT PAR QPCR	120
B.	CARACTERISATION PAR DDPCR	121
1.	<i>Mise au point par ddPCR</i>	121
2.	<i>Test d'interférence</i>	121
3.	<i>Test de reproductibilité</i>	122
4.	<i>Test de digestion</i>	123
5.	<i>Caractérisation des CNVs</i>	123
C.	TRANSMISSION DES CNVS DANS LES FAMILLES	124
D.	RELATION ENTRE LES CNVS DE <i>CCL3L1</i> ET LES FACTEURS DE LA PR	125
	CHAPITRE 7 : DISCUSSION.....	127
I.	CARACTERISATION DES CNVS DU GENE <i>GSTM1</i>	127
II.	CARACTERISATION DES CNVS DU GENE <i>GSTT1</i>	131
III.	CARACTERISATION DES CNVS DU GENE <i>FCGR3B</i>	133
IV.	CARACTERISATION DES CNVS DU GENE <i>CCL3L1</i>	138
	CHAPITRE 8 : CONCLUSION	144
	REFERENCES BIBLIOGRAPHIQUES.....	144
	ANNEXES.....	144

Table des Figures

Figure 1 : Evolution clinique de la Polyarthrite Rhumatoïde : a) chez un individu sain et b) chez un patient PR.....	1
Figure 2: Différentes phases de la pathophysiologie de la polyarthrite rhumatoïde.	5
Figure 3: Le réseau cytokinique dans la PR.	8
Figure 4: Cascades de signalisation cellulaire d'apoptose et de survie dans les fibroblastes de la synoviale rhumatoïde	10
Figure 5: La réaction de citrullination.....	12
Figure 6 : Modes d'actions des traitements ciblés disponibles	15
Figure 7: Principe de l'analyse des paires de germains	17
Figure 8: Sillon de fixation des antigènes et la substitution des acides aminés de HLA et leur influence sur la susceptibilité de la PR.....	20
Figure 9: Réseau d'interaction entre les différents variants associés à la PR).....	22
Figure 10: Exemples de réarrangements récurrents et non récurrents.	26
Figure 11: Réarrangements génomiques résultants d'une recombinaison entre deux LCRs....	27
Figure 12: <i>Non-homologous end-joining</i> (NHEJ) chez les vertébrés.	29
Figure 13: Mécanisme de réarrangements de type <i>Fork Stalling and Template Switching</i>	30
Figure 14: Localisation des variants génétiques associées à des variations de l'expression génique	32
Figure 15: Etapes de l'hybridation génomique comparative (CGH) conventionnelle et sur puces.....	33
Figure 16: Utilisation des génotypes de SNPs pour l'identification de délétion par analyse de ségrégation familiale.	35
Figure 17: Cinq classes de Variations de Nombre de Copie identifiés par l'aCGH et la plateforme SNP.	37
Figure 18: Cinq approches pour détecter des CNVs à partir de données de séquences NGS..	39

Figure 19: Etapes de la méthode <i>Paire-end Mapping</i>	40
Figure 20: Stratégie d'amplification de la région de délétion du gène <i>GSTM1</i>	44
Figure 21: Principe de la technologie TaqMan® pour la quantification des CNVs.	45
Figure 22: Méthode de calcul de nombre de copies en fonction des valeurs Ct.	46
Figure 23: Diagramme schématique de trois systèmes de PRT au niveau du locus du gène <i>CCL3L1</i>	48
Figure 24: Schéma des différentes étapes de la MLPA	49
Figure 25: Principe de la Microfluidique Digital PCR.....	51
Figure 26: Principe de la <i>Droplet Digital PCR</i>	52
Figure 27: La réaction de conjugaison de la phase II de la détoxification des xénobiotiques .	61
Figure 28: Mise en évidence de la duplication en tandem du gène <i>GSTM1</i>	63
Figure 29: Expression différentiel du gène <i>GSTM1</i> en fonction du nombre de copies.....	65
Figure 30: Localisation du cluster de gène de la famille <i>GST</i> theta sur le chromosome 22q11.2. La délétion de <i>GSTT1</i> se produit par recombinaison des régions répétées de part et d'autre du gène.....	66
Figure 31: Expression du gène <i>GSTT1</i> en fonction des trois génotypes	67
Figure 32: Structure des classes de FcγRs avec la localisation des SNPs fonctionnels caractérisés	70
Figure 33: Cluster des gènes FcγRs au niveau du chromosome 1.	71
Figure 34: Région dupliquée de FCGR3 avec les différents clusters de <i>breakpoints</i>	72
Figure 35: Interaction entre CCR5 et CCL3L1 lors de l'infection par le virus HIV	74
Figure 36: Région génomique de <i>CCL3L1</i> au niveau du chromosome 17	75
Figure 37: Position des amorces P1, P2 et P3 au niveau des gènes <i>GSTM1</i> et <i>GSTM4</i>	81
Figure 38 : Schéma de la distribution des échantillons sur les puces. Chaque carré contient 64 nano-puits de 33nl chacun.....	84
Figure 39 : Principe de l'amplification de la région dupliquée. Les deux traits rouges correspondent aux amorces XLP1 et XLP2.	85

Figure 40 : Digestion avec l'enzyme HaeIII dans la région du gène GSTM1 et l'impact sur l'estimation des CNVs par la ddPCR. Les traits verticaux bleus indiquent les sites de coupure de l'enzyme HaeIII	87
Figure 41 : Profil de migration de la PCR-présence/absence du gène GSTM1 sur gel d'agarose pour 11 échantillons.....	95
Figure 42 : Profil de migration des produits PCR issus de la séquence délétée (4,7kb) et de la séquence du gène (625pb) sur gel d'agarose pour 11 échantillons.....	96
Figure 43 : Quantification des CNVs du gène GSTM1 par qPCR pour 18 échantillons.....	97
Figure 44 : Exemples de résultats obtenus par la dPCR pour deux échantillons avec un nombre de copies différents.	98
Figure 45 : Distribution du nombre de copies obtenus par qPCR et dPCR.	98
Figure 46 : Profil électrophorétique de la longue PCR sur gel d'agarose 0,6% obtenu pour l'échantillon 14.....	99
Figure 47 : Tests d'interférence des sondes de <i>GSTM1/RNase P</i> et <i>GSTM1/RPP30</i>	100
Figure 48 : Test de reproductibilité intra et inter-plaques.....	100
Figure 49: Gradient de HaeIII permettant la mise en évidence de la duplication en tandem.	101
Figure 50: Résultats obtenus par ddPCR.....	102
Figure 51: Diagramme de Venn indiquant le nombre d'échantillons ayant le même nombre de copies retrouvé par les trois méthodes	103
Figure 52 : Profil de migration des produits de PCR-Multiplex du gène <i>GSTT1</i> sur gel d'agarose pour 12 échantillons.....	108
Figure 53: Profil de migration des amplifications de la région délétée et du gène <i>GSTT1</i> pour 6 échantillons.....	108
Figure 54 : Distribution des CNVs dans les 96 échantillons issus des familles trio.	112
Figure 55 : Tests d'interférence des sondes de <i>FCGR3B/RPP30</i>	113
Figure 56: Résultats du test de reproductibilité intra et inter-plaque pour <i>FCGR3B</i>	113
Figure 57: Tests de digestion avec un gradient d'unité enzymatique pour HaeIII et FokI réalisé sur un échantillon de 1,6 copies.	114

Figure 58 : Distribution des CNVs dans notre échantillon de familles.....	115
Figure 59 : Analyse de la corrélation entre les CNVs obtenus par qPCR et ddPCR.	116
Figure 60 : Courbes de dissociation des amorces pour <i>CCL3LI</i> à différentes températures (entre 66°C et 56°C) d'hybridation (a) et à 60°C (b).	120
Figure 61: Séparation des <i>plots</i> positifs et négatifs pour le test en EVA Green (a) et le test avec les sondes (b).....	121
Figure 62: Tests d'interférence des sondes de <i>CCL3LI/RPP30</i>	122
Figure 63: Résultats du test de reproductibilité intra et inter-plaque pour <i>CCL3LI</i>	122
Figure 64: Résultats des tests de digestion avec deux enzymes (MseI et AluI) à différentes unités et avec deux protocoles.....	123
Figure 65: Distribution des CNVs du gène <i>CCL3LI</i> dans les échantillons de 93 familles trio.	124
Figure 66: Schéma récapitulatif résumant les différentes étapes à suivre pour la quantification des CNVs par ddPCR en fonction de leurs caractéristiques.	143

Liste des tableaux

Tableau 1: Critères pour la classification de la Polyarthrite Rhumatoïde (American College of Rheumatology 1987 et 2010).....	2
Tableau 2 : Résumé des DMARDs biologiques disponibles pour le traitement de la PR	14
Tableau 3: Liste des plateformes de détection de CNVs.....	38
Tableau 4: Maladies autoimmunes associées à des CNVs dans le génome humain.....	56
Tableau 5: Résultats d'associations significatives de la plupart CNVs présents dans les régions indiquées. Les gènes contiennent, ou chevauchent des CNVs. Les positions sont issues de la base de données NCBI	59
Tableau 6: Pourcentage des individus suivant le nombre de copie de <i>FCGR3B</i> et en fonction des différentes méthodes de quantification utilisées	72
Tableau 7: Caractéristiques des cas index PR.....	80
Tableau 8 : Caractéristique des amorces d'amplification	81
Tableau 9 : Caractéristiques des amorces d'amplification du gène <i>GSTM1</i> et de la séquence de délétion	82
Tableau 10 : Caractéristiques des amorces d'amplification de la région dupliquée	86
Tableau 11 : Caractéristiques des amorces d'amplification du gène <i>GSTT1</i> et du gène de la β -globine.....	89
Tableau 12 : Caractéristiques des amorces d'amplification de la région délétée et du gène ...	89
Tableau 13 : Caractéristiques des amorces d'amplification et de la sonde spécifique du gène <i>CCL3L1</i>	91
Tableau 14 : Modèle de calcul de l'Odds Ratio.	93
Tableau 15: Modèles de transmission des CNVs dans 199 familles trio	104
Tableau 16: Résultats du Test de Déséquilibre de Transmission (TDT) et du Risque Relatif des Génotypes (GRR) dans 198 familles trio de patients PR.....	105
Tableau 17: Analyse de la relation entre les CNVs du gène <i>GSTM1</i> et les facteurs associés à la PR (200 cas index)	107

Tableau 18 : Résultats du Test de Déséquilibre de Transmission (TDT) et du Risque Relatif des Génotypes (GRR) 197 les familles trio de patients PR.....	110
Tableau 19 : Analyse de la relation entre les CNVs du gène <i>GSTT1</i> et les facteurs associés à la PR (199 cas-index).....	111
Tableau 20 : Transmission des CNVs du gène <i>FCGR3B</i> dans 188 familles trio.....	117
Tableau 21: Résultats du Test de Déséquilibre de Transmission (TDT) et du Risque Relatif des Génotypes (GRR) pour le gène <i>FCGR3B</i> dans 185 familles trio de patients PR	118
Tableau 22: Analyse de la relation entre les CNVs du gène <i>FCGR3B</i> et les facteurs associés à la PR chez 197 cas-index	119
Tableau 23: Transmission des CNVs du gène <i>CCL3L1</i> dans 93 familles trio.....	125
Tableau 24: Analyse de la relation entre les CNVs du gène <i>CCL3L1</i> et les facteurs associés à la PR chez 98 cas-index	126

Abréviations

AASV Systémique	Anticorps cytoplasmiques Anti-neutrophiles associés à une Vascularite
ACPA	Anticorps anti-peptide citrulliné
ACR	American College of Rheumatology
ADCC	Cytotoxicité cellulaire dépendante des anticorps
ADN	Acide Désoxyribonucléique
AKA	Anti-Keratin Antibodies
AKT	Protein Kinase B
ALOX5AP	Arachidonate 5-Lipoxygenase-Activating Protein
APF	Anti-Perinuclear Factor
APP	Amyloid Precursor Protein
AS	Assemblage de Séquence <i>de novo</i>
ASD	Autism Spectrum Disorder
ASTN2	Astrotactin 2
BAC	Chromosomes Artificiels Bactériens
B2M	Beta-2-Microglobulin
BIP	Immunoglobulin heavy gene Binding Protein
BMX	Bone Marrow Kinase
C4	Complément 4
CCL3	Chemiokine C-C motif Ligand 3
CCL3L1	Chemiokine C-C motif Ligand 3-Like 1
CCL3L3	Chemiokine C-C motif Ligand 3-Like 3

CCL3P1	Chemiokine C-C motif Ligand 3 Pseudogene 1
CCP	Cyclic Citrullinated Peptide
CD	Cluster de Différenciation
CGH	Comparative Genomic Hybridation
CMH	Complexe Majeur d’Histocompatibilité
CMT1A	Charcot-Marie-Tooth Type 1A
CN	Copy Number
CNP	Copy Number Polymorphism
CNV	Copy Number Variation
CNVR	Copy Number Variation Region
CPA	Cellules Présentatrices d’Antigène
CRHR1	Corticotropin Releasing Hormone Receptor 1
CTLA-4	Cytotoxic T Lymphocyte Associated antigen 4
CTNND2	Catenin (cadherin-associated protein), Delta 2
dPCR	Digital PCR
ddPCR	Droplet Digital PCR
DEFB	Beta-Défensine
DGV	Database of Genomic Variation
DL	Déséquilibre de Liaison
DMARDs	Disease-Modifying Antirheumatic Drugs
DMD	Dystrophin Muscular Dystrophy
dMIQE	Minimum Information for the publicaton of Digital PCR Experiment
DSB	Double-Strained DNA Break

ELISA	Enzyme-Linked Immuno-Sorbent Assay
FAS	Apoptosis Stimulating Fragment
Fc	Fragment cristallisable
FCGR	Low-affinity Fc Gamma Receptor
FCGR2C	Low-affinity Fc Gamma Receptor 2C
FCGR3A	Low-affinity Fc Gamma Receptor 3A
FCGR3B	Low-affinity Fc Gamma Receptor 3B
FISH	Fluorescent In Situ Hybridation
FoSTeS	Fork Stalling and Template Switching
FR	Facteur Rhumatoïde
FRET	Fluorescence Resonance Energy Transfer
GO	Gene Ontology
GRR	Risque Relatif des Génotypes
GS	Gold Standard
GST	Glutathion S-Transférerase
GSTM1	Glutathion S-Transférerase M1
GSTM2	Glutathion S-Transférerase M2
GSTM4	Glutathion S-Transférerase M4
GSTT1	Glutathion S-Transférerase T1
GWAS	Genome Wide Association Studies
HLA	Human Leukocyte Antigen
HLA-DR	Human Leukocyte Antigen Receptor
HLA-DRB1	Human Leukocyte Antigen Receptor Beta chain 1

HNPP	Hereditary Neuropathy with liability to Pressure Palsies
HR	Homologous Recombination
HSD17B10	Hydroxysteroid (17-beta) Dehydrogenase 10
HUWE1	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase
IFNγ	Interféron γ
Ig	Immunoglobuline
IKKg	I κ B Kinase-g
IL	Interleukine
IRF1	Interferon Regulatory Factor 1
ISV	Intermediate Size Variation
ITAM	Immunoreceptor Tyrosine-based Activation Motif
ITIM	Immunoreceptor Tyrosine-based Inhibitory Motif
ITP	Purpura Thrombopénique Idiopathique
JAK	Janus Kinase
JNK	Jun amino-terminal kinases
LBs	Lymphocytes B
LCE	Late Cornified Envelope
LCE3B	Late Cornified Envelope 3B
LCE3C	Late Cornified Envelope 3C
LCP2	Lymphocyte Cytosolic Protein 2
LCR	Low Copy Repeat
LCV	Large Copy Number

LINEs	Long Interspersed repeated Sequences
LITAF	Lipopolysaccharide-Induced TNF Factor
LTs	Lymphocytes T
LTRs	Long Terminal Repeat
MAF	Minor Allele Frequency
MAPK	Mitrogen Activated Protein Kinase
MAPT	Microtubule-Associated Protein Tau
MCC	Molecular Copy-number Counting
MECP2	Methyl CpG binding Protein 2
MIP	Macrophages Inflammatory Protein
MIR	Mammalian-wide Interspersed Repeat
MLPA	Multiplex-Ligation-Dependent Probe Amplification
mTOR	Mammalian Target of Rapamycin
mAb	Monoclonal Antibodies
MEPS	Minimum Efficient Processing Segment
MIF	Macrophage migration Inhibitory Factor
MTX	Methotrexate
MYT1L	Myelin Transcription factor 1-Like
NAHR	Non-Allelic Homologous Recombination
NCF1	Neutrophil Cytosolic Factor 1
NFκB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End-Joining

NK	Natural Killer
OR	Odds Ratio
p53	Cellular Tumor Antigen
PADI	Peptidylarginine-Désaminase
PCR	Polymerase Chain Reaction
PEM	Paired-End Mapping
PI-3	PhosphoInositide-3
PI-3K	PhosphoInositide-3 Kinase
PLP	Proteolipid Protein 1
PMP22	Peripheral Myelin Protein 22
PR	Polyarthrite Rhumatoïde
PRT	Paralogue Ratio Test
PTPN22	Protein Tyrosine Phosphatase Non-receptor22
qPCR	Quantitative PCR
RAI1	Retinoic Acid Induced 1
RD	Read Depth
RIBC1	RIB43A domain with Coiled-coils 1
ROS	Reactive Oxygen Species
RM	Retard Mental
RNP	Nuclear Ribonucleoprotein
SAPK	Stress-activated protein kinases
SD	Segmental Duplication
SE	Shared Epitope

SLE	Lupus Systémique Erythémateux
smPCR	Single Molecule PCR
SMS	Smith-Magenis Syndrom
SMC1A	Structural Maintenance of Chromosomes 1A
SNP	Single Nucleotide Polymorphism
SR	Split Read
STAT	Transducer and Activator of Transcription
SV	Structural Variation
Syk	Spleen Tyrosine Kinase
TDT	Test de Déséquilibre de Transmission
TE	Tobacco Exposure
TCR	T Cell Receptor
TBE	Tris-Borate-EDTA
Th	Lymphocytes T helper
TLR	Toll Like Receptor
TNF	Tumor Necrosis Factor
TNFAIP3	Tumor Necrosis Factor, Alpha-Induced Protein 3
TNIP1	TNFAIP3 Interacting Protein 1
Treg	Lymphocytes T régulateur
VPREB1	Pre-B lymphocyte 1
WGS	Whole Genome Sequencing
WES	Whole Exome Sequencing
WTCCC	Wellcom Trust Case Control Consortium

Chapitre 1 : La Polyarthrite Rhumatoïde

I. Définition

La polyarthrite rhumatoïde (PR) est une maladie auto-immune systémique caractérisée par une arthrite chronique, destructive et invalidante. C'est une maladie multifactorielle car plusieurs facteurs interviennent dans le déclenchement de la maladie, comme par exemple : les facteurs hormonaux, le terrain génétique prédisposé et des facteurs environnementaux (Sany J 1997). Lorsque tous ces facteurs sont réunis, ils activent une réponse immunitaire innée et acquise incontrôlée qui se traduit par une réaction inflammatoire exagérée, en particulier au niveau de la membrane synoviale des articulations (Smolen and Steiner 2003) (Figure 1).

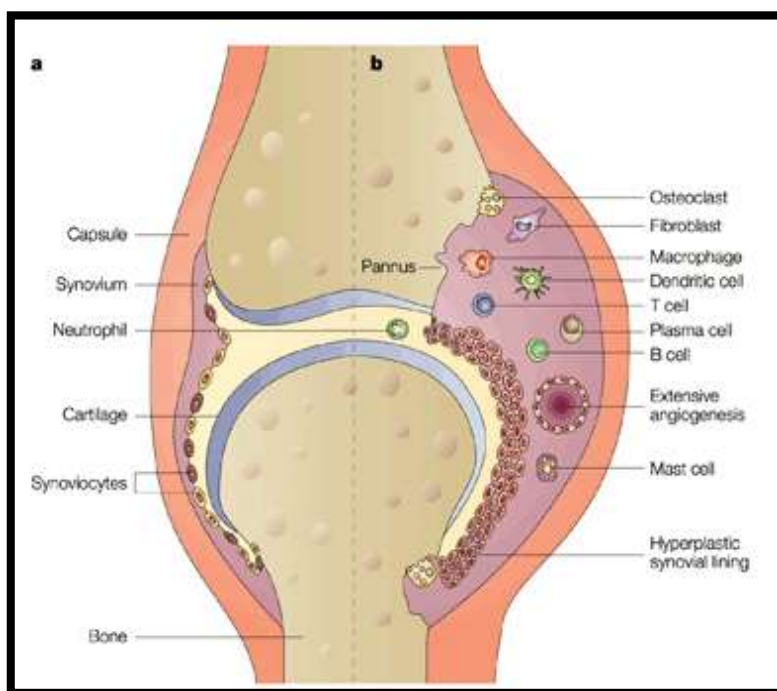


Figure 1 : Evolution clinique de la Polyarthrite Rhumatoïde : a) chez un individu sain et b) chez un patient PR (d'après Smolen and Steiner 2003).

A. Aspects cliniques de la PR

La PR est une polyarthrite inflammatoire chronique qui touche de multiples articulations avec une distribution bien caractéristique, induisant des douleurs, des déformations articulaires et réduction de la qualité de vie (Firestein 2003). Les patients se présentent généralement avec une douleur et une raideur articulaire dans plusieurs jonctions.

L'apparition des symptômes se fait sur plusieurs semaines voir des mois, débutant par une articulation et souvent accompagnée par un amaigrissement et une fatigue. La déclaration de la maladie survient plus rapidement chez 15% des patients. Dans 8 à 15% des cas, les symptômes débutent après un évènement initiateur spécifique comme une infection par exemple (Harris 2005).

La PR affecte préférentiellement les femmes avec un *sex ratio* de 2 hommes pour 4 femmes. La moyenne d'âge de déclenchement de la maladie est entre 45 et 50 ans et la prévalence est de 1% dans la population mondiale (Firestein 2003). La description de la maladie a évolué depuis 1859, année de la première définition de la PR par Garrod, et plusieurs critères ont été établis par l'*American College of Rheumatology* (ACR) dans les années 80 (Arnett, Edworthy et al. 1988). Ils ont été révisés en 2010 pour établir une nouvelle liste représentée dans le tableau 1 (Aletaha, Neogi et al. 2010).

Tableau 1: Critères pour la classification de la Polyarthrite Rhumatoïde (American College of Rheumatology 1987 et 2010). D'après Aletaha D, et al. 2010.

	Score
Target population (Who should be tested?): Patient who	
1) have at least 1 joint with definite clinical synovitis (swelling)	
2) with the synovitis not better explained by another disease	
Classification criteria for RA (score-based algorithm: add score categories A-D; a score of $\geq 6/10$ is needed for classification of a patient as having definite RA)	
A. Joint involvement	
1 large joints	0
2-10 large joints	1
1-3 small joints (with or without involvement of large joints)	2
4-10 small joints (with or without involvement of large joints)	3
> 10 joints (at least 1 small joint)	5
B. Serology (at least 1 test result is needed for classification)	
Negative RF <i>and</i> negative ACPA	0
Low-positive RF <i>or</i> low-positive ACPA	2
High-positive RF <i>or</i> high positive ACPA	3
C. Acute-phase reactants (at least 1 test result is needed for classification)	
Normal CRP <i>and</i> normal ESR	0
Abnormal CRP <i>or</i> normal ESR	1
D. Duration of symptoms	
< 6 weeks	0
≥ 6 weeks	1

RA : Rheumatoid Arthritis ; RF : Rheumatoid Factor ; ACPA : Antibodies to Citrullinated Protein Antigen

B. Epidémiologie de la PR

Plusieurs études ont estimé la prévalence et l'incidence de la PR au cours de ces dernières décennies, montrant une variation considérable de la fréquence de cette maladie dans les différentes populations (Gabriel, Crowson et al. 1999; Saraux, Guedes et al. 1999; Simonsson, Bergman et al. 1999; Symmons 2001). Cette différence peut être due aux méthodes d'analyses comme par exemple les différences dans les méthodes d'identification des patients et leur enregistrement. Néanmoins, ces différences n'expliquent pas cette grande variabilité entre les populations, d'autres facteurs peuvent influencer cette prévalence tels que l'accès aux soins, les facteurs environnementaux et les facteurs génétiques (MacGregor A 2003).

1. Prévalence

C'est le nombre de personnes atteintes d'une pathologie donnée par rapport à la population totale. La moyenne de la prévalence dans les pays de l'Europe du Sud est estimée à 3,3 cas pour 10^3 , dans les pays de l'Europe du Nord à 5 et dans les pays développés à 3,5. Une étude réalisée en Amérique du Nord a montré une prévalence estimée à 10,7 cas pour 10^3 (Alamanos, Voulgari et al. 2006). Peu d'études ont été réalisées dans plusieurs endroits du monde comme certains pays d'Afrique, la Russie et une grande partie de l'Asie, ce qui peut induire une sous-estimation de la prévalence mondiale. La population des indiens natifs d'Amérique a le record le plus élevé de développement de PR, avec une prévalence de 5,3% pour les indiens Pima (Del Puente, Knowler et al. 1989) et de 6,8% chez les indiens Chippewa (Harvey, Lotze et al. 1981). Par contre, il existe des groupes de populations qui ont une très faible prévalence comme en Afrique du Sud (Brighton, de la Harpe et al. 1988) et au Nigéria (Silman, Ollier et al. 1993) où aucun patient PR n'a été retrouvé sur respectivement 500 et 2000 adultes.

2. Incidence

C'est le nombre de nouveaux cas par an dans une population. L'incidence moyenne annuelle de la PR observée dans les pays de l'Europe du Sud est de 16,5 cas pour 10^5 . Pour les pays de l'Europe du Nord, l'incidence moyenne annuelle est de 29 et elle est de 38 en Amérique du Sud. Il existe une différence significative entre les pays de l'Europe du Nord et du Sud dans la population d'hommes uniquement. Les pays de l'Europe du Sud ont une incidence et une prévalence plus faible que les pays de l'Europe du Nord et l'Amérique du

Nord, mais cette différence reste non significative dans la population des femmes. Ce manque de significativité peut être lié au nombre limité d'études réalisées en Europe du Sud. En outre, plusieurs études portent sur une faible taille d'échantillon (Alamanos, Voulgari et al. 2006).

3. Mortalité et survie dans la PR

Le taux de mortalité est plus élevé pour les patients atteints de PR que dans la population générale. Le taux de mortalité varie largement entre les études réalisées : il est élevé dans les études réalisées en milieu hospitalier et relativement faible dans les études réalisées sur les populations. La survie des patients PR diminue de 3 à 10 ans en fonction de la sévérité de la maladie et l'âge du déclenchement de la PR. Les causes de mortalité ne diffèrent pas entre les patients PR et la population générale à laquelle ils appartiennent. La conclusion est que la majorité des individus atteints meurent des mêmes causes que la population générale, mais à un âge plus précoce (Gabriel, Crowson et al. 1999; Doran, Pond et al. 2002).

C. Physiopathologie de la PR

Beaucoup d'études ont montré que, au cours de la maladie, il y avait une dérégulation dans le fonctionnement du système immunitaire ce qui a permis de classer la PR comme maladie auto-immune. En effet, de nombreux signes d'auto-réactivité avec production d'auto-anticorps ont été décrits. Plusieurs acteurs cellulaires de l'immunité innée et adaptative, comme les macrophages, les cellules dendritiques, les lymphocytes B et T, entrent en jeu et interagissent entre eux et avec d'autres cellules de l'organisme, comme les synoviocytes et les ostéocytes, pour arriver à la destruction articulaire (Figure 2) (Burmester, Feist et al. 2014).

1. Les acteurs cellulaires impliqués dans la PR

a) Les cellules présentatrices d'antigène

Certaines cellules de l'immunité innée comme les macrophages et les cellules dendritiques sont capables de présenter un antigène aux lymphocytes T pour enclencher la réponse immunitaire adaptative, elles sont donc appelées Cellules Présentatrices d'Antigène (CPA). Ces cellules expriment à la surface de leur membrane des molécules nommées *Human Leukocyte Antigen* (HLA) (Sprenger, Schlagenhauser et al.) de classe II, qui sont indispensables dans l'activation des lymphocytes T (Pettit and Thomas 1999). Ces CPAs, ainsi que leurs protéines effectrices, ont été retrouvées dans le liquide synovial de patients PR,

ce qui indique qu'elles sont impliquées dans l'inflammation articulaire et la destruction du cartilage et de l'os (Figure 1 et 2).

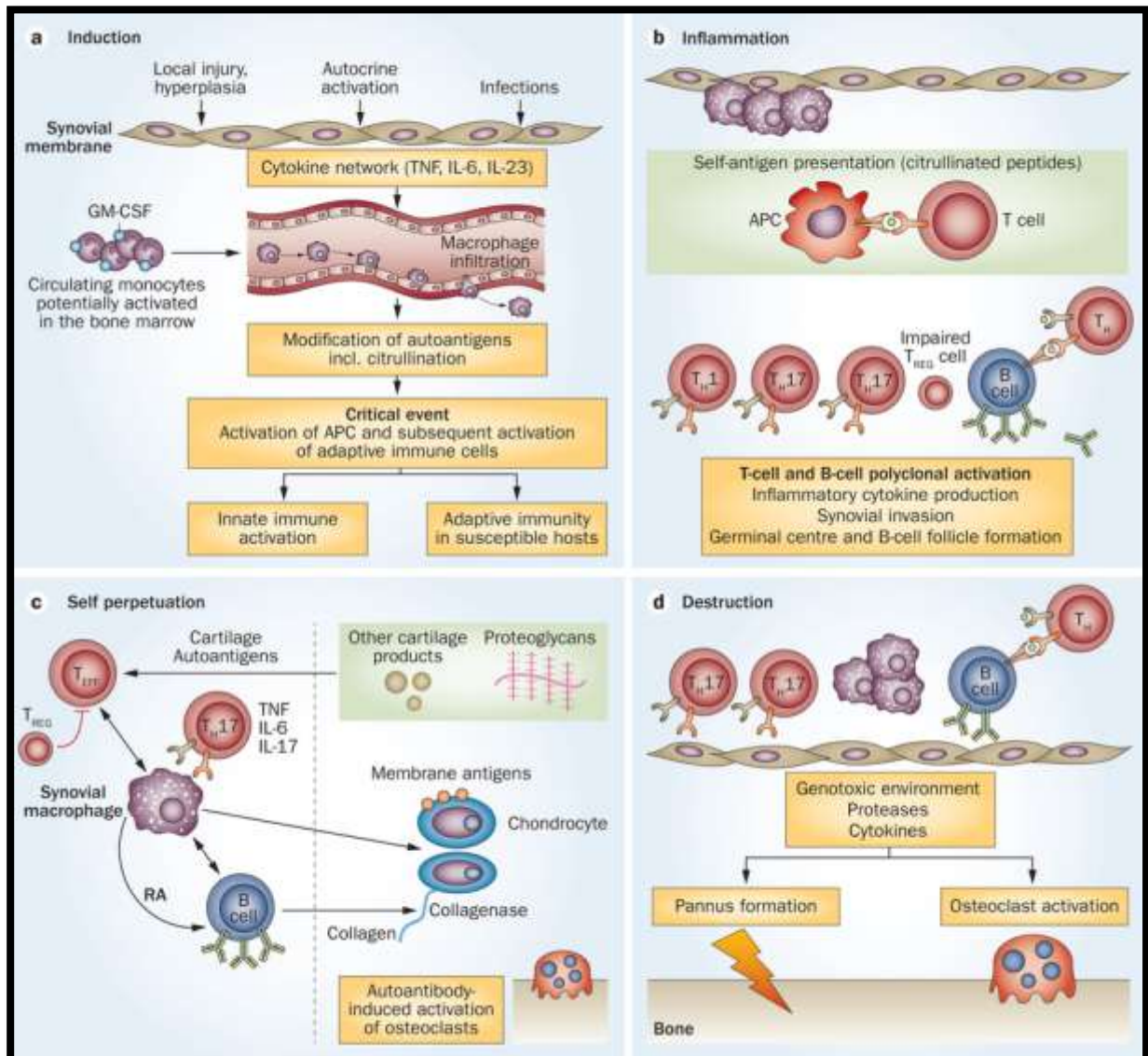


Figure 2: Différentes phases de la pathophysiologie de la polyarthrite rhumatoïde (d'après Burmester, Feist et al. 2014).

Lors de l'inflammation, les cellules du système immunitaire inné sont attirées dans la synoviale rhumatoïde par des chimiokines. Les cellules dendritiques sont plus particulièrement attirées par les chimiokines *Macrophages Inflammatory Protein* (MIP- α et β) qui sont retrouvées en abondance dans la synoviale. Il a été rapporté que les infections par certains microorganismes étaient un facteur déclenchant dans la PR par activation de la réponse innée, via l'activation des récepteurs *Toll Like* (TLR) (Klinman 2003; Morel, Miossec et al. 2004). Une fois activées, les cellules dendritiques induisent l'inflammation de la

synoviale et sont capables d'activer les T *helper* (Th) 1 ou la réponse Th2 en fonction des cytokines présentes dans l'environnement.

b) Les lymphocytes T

Les lymphocytes T (LTs) autoréactifs sont capables de réagir avec les molécules du soi. La proportion de ces LTs chez les patients PR est plus élevée que chez les individus sains et serait due à une dérégulation lors de la sélection thymique. En effet, lors de la maturation des LTs dans le thymus, ceux qui reconnaissent les auto-peptides associés aux molécules du Complexe Majeur d'Histocompatibilité (CMH) de classe II, présentées par les CPA, sont éliminés par apoptose. Chez les patients PR, il est supposé que cette sélection négative est déficiente, ce qui a pour conséquence la survie d'un grand nombre de LTs auto-réactifs (Weyand, McCarthy et al. 1995). Après reconnaissance d'un antigène, les LTs naïfs se différencient en LTs producteurs d'Interféron γ (IFN γ), d'Interleukine 2 (IL2) ou encore d'IL17. Cette réponse est dite de type Th1 par opposition à une réponse de type Th2 qui se traduit par une production d'IL4. Dans la synoviale rhumatoïde, les lymphocytes Th1 sont particulièrement abondants (Berek and Schroder 1997; Panayi, Corrigan et al. 2001). Deux autres types de cellules effectrices ont été récemment mis en évidence, les Th17 et les Treg, ce qui a permis d'apporter des nouveaux éléments quant à l'immunopathologie de la PR. Le Th17 produit l'IL23 qui joue un rôle important dans les lésions tissulaires induites par Th17. De plus, l'identification des cibles cellulaires et les effets biologiques de l'IL17 produite par les Th17, renforce l'hypothèse que ces cellules ont un rôle dans les lésions synoviales et articulaires chez les patients PR. Par contre les Treg jouent un rôle essentiel dans le maintien de la tolérance au soi. En effet, après avoir démontré l'importance des Treg dans la suppression des cellules CD4 et CD8 au niveau organe spécifique et l'auto-immunité systémique chez les rongeurs, il a été suggéré qu'une défaillance des Tregs contribuait à l'inflammation chez les patients PR (Steinman 2007).

c) Les lymphocytes B

Jusqu'à récemment, le rôle des lymphocytes B (LBs) a été minimisé. Cependant, l'efficacité des biothérapies avec un anticorps dirigé contre le marqueur CD20 des LBs (le rituximab), qui induit le blocage des LBs chez les patients atteints de PR, souligne leur importance dans la physiopathologie (De Vita, Zaja et al. 2002). Les LBs peuvent endosser le rôle de véritable CPA car ils sont capables de présenter des antigènes aux lymphocytes CD4+. L'expression des molécules HLA-DR, observées à la surface des LBs présents dans la

synoviale rhumatoïde, conforte l'argumentation de l'implication des LBs dans la présentation des antigènes (Eliaou, Andary et al. 1988; Tighe, Chen et al. 1993). Les lymphocytes B produisent certains auto-anticorps détectés chez les patients atteints de PR tels que le facteur rhumatoïde (FR, cf page 10) et les anticorps anti-collagène. Il existe d'autres auto-anticorps qui peuvent être impliqués dans la physiopathologie de la maladie sans être spécifiques à la PR, comme BIP (*immunoglobulin heavy gene binding protein*) et *human RNP-33* (RA33). En revanche, il existe des auto-anticorps qui sont plus spécifiques de la PR comme les anticorps dirigés contre les peptides citrullinés (ACPA) (van Boekel, Vossenaar et al. 2002)

d) Les synoviocytes

Les synoviocytes constituent le principal composant cellulaire de la couche bordante de la membrane synoviale et sont divisés en deux groupes : les macrophages et les synoviocytes fibroblastiques. Les premiers font partie de la lignée monocyttaire, tandis que les seconds ont plutôt une origine mésenchymateuse. Les macrophages se distinguent des synoviocytes fibroblastiques par l'expression de molécules HLA de classe II et de récepteurs Fc des immunoglobulines (Firestein, Alvaro-Gracia et al. 1990). Lorsqu'ils sont activés, les synoviocytes macrophagiques seraient les véritables moteurs de la réaction inflammatoire produisant deux médiateurs : des médiateurs primaires ne nécessitant pas de synthèse protéique (les prostaglandines, leucotriènes, les radicaux libres et les enzymes contenus dans les granules et qui participent de façon importante dans la destruction des tissus) et des médiateurs secondaires qui requiert une synthèse protéique comme les cytokines pro-inflammatoires IL1 et le *Tumor Necrosis Factor alpha* (TNF α) (Russo-Marie 1996). Ces derniers induisent la synthèse de médiateurs primaires par une action autocrine et paracrine sur les chondrocytes, les cellules endothéliales et les synoviocytes. Ces cellules vont produire à leur tour des facteurs de croissances et des cytokines pro-inflammatoires.

Les synoviocytes ont une capacité de prolifération qui ressemble dans certains aspects aux cellules cancéreuses. En effet, ils possèdent des caractéristiques morphologiques semblables, une perte d'inhibition de contact et l'activation de plusieurs oncogènes comme les cellules tumorales. La prolifération anormale des synoviocytes de PR pourrait s'expliquer, comme pour les tumeurs, par un défaut d'apoptose (Pap, Franz et al. 2000; Pope 2002). Les protéines p53 et *FAS ligand* ainsi que les voies de signalisation NF κ B et PI-3 kinase sont particulièrement impliquées dans le phénomène de résistance de synoviocytes à l'apoptose.

2. Les acteurs intercellulaires : les cytokines

Les interactions cellulaires entre les macrophages, les lymphocytes B et T, et les cellules non-hématopoïétiques (fibroblastes, tissus de connexion et les cellules de l'os) sont importantes dans la pathogénicité de la PR et sont facilitées par l'action des cytokines libérées par les cellules activées. Il existe quatre grandes familles de cytokines : les interleukines, les interférons, les facteurs de croissance et les chimiokines. Il est de plus en plus évident que la présence de certaines cytokines pro-inflammatoires dans le sérum et le liquide synovial reflète leur implication dans la pathogénicité et la progression de la PR. Ainsi, le TNF α avec plusieurs interleukines, notamment IL1, 6, 7, 12/23, 15, 17, 18, 32 et 35, et des protéines de la famille des interférons, ont un taux élevé dans le sérum de patients (Szekanecz and Koch 2007; Malemud 2013) ainsi que dans le liquide synovial (Figure 3) (Oppenheimer-Marks, Davis et al. 1991).

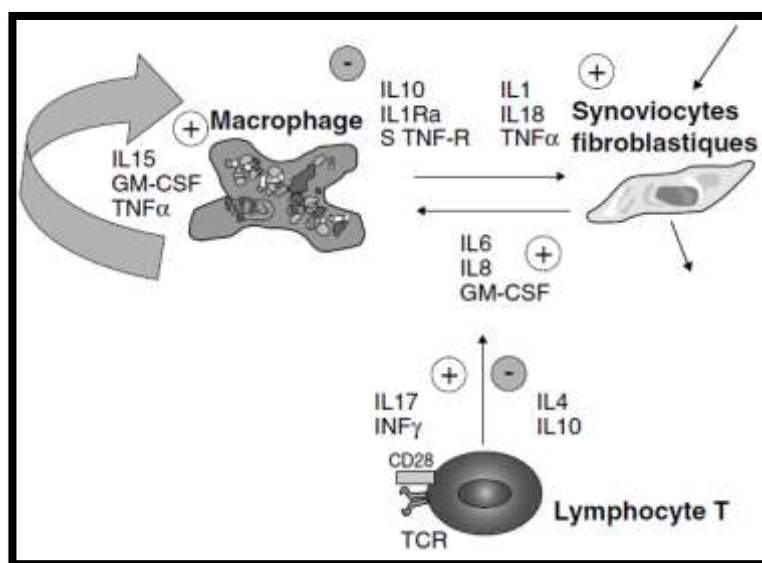


Figure 3: Le réseau cytokinique dans la PR (d'après Morel, Miossec et al. 2004).

Comme cité précédemment, les lymphocytes Th1 sont les plus abondants dans la synovite rhumatoïde mais, paradoxalement, la production de cytokines (IL2 et IFN γ) par ces cellules est faible dans les articulations des patients. Néanmoins, les cellules CD4+, Th1 et Th0 sécrètent plus d'IL17 chez les patients PR que chez les patients atteints d'arthrose (Miossec 2003). Les synoviocytes produisent également des cytokines telles que les TNF α , l'IL1, l'IL15, l'IL18 et l'IL6 mais aussi des facteurs de croissance et des chimiokines qui sont présents à une forte concentration dans le liquide synovial mais également dans le sérum de

patients atteints de PR (Firestein, Alvaro-Gracia et al. 1990; Alvaro-Gracia, Zvaifler et al. 1991). Il est donc clair que les cytokines permettent la communication entre les différents acteurs cellulaires pour aboutir aux phénotypes associés à la PR.

3. Acteurs de la signalisation cellulaire

Dans les cellules, la fixation d'un ligand sur son récepteur permet la transduction du message sous forme d'une cascade de signalisation cellulaire qui aboutit à l'activation des facteurs de transcription qui vont permettre à leur tour de transformer le message chimique en une action cellulaire (comme par exemple la sécrétion de protéines effectrices). L'activation de ces facteurs de transcription est induite par une cascade de phosphorylation grâce à des protéines kinases qui ont une activité phosphorylante.

De récentes études ont montré que plusieurs protéines kinases étaient dérégulées dans la PR et dans d'autres maladies autoimmunes. Dès lors, il a été observé que le dysfonctionnement du signal intracellulaire, induit par les cytokines pro-inflammatoires, était responsable d'une survie anormale des cellules immunitaires (Carames, Lopez-Armada et al. 2008; Mayer, Rehberg et al. 2011), de l'apoptose des chondrocytes au niveau des articulations (Carames, Lopez-Armada et al. 2008; Malemud, Sun et al. 2012) et/ou d'une résistance à l'apoptose des cellules dans les tissus synoviaux de patients PR (Liu and Pope 2004; Jonsson, Allen et al. 2005; Korb, Pavenstadt et al. 2009; Malemud 2011).

Des dérégulations dans certaines voies de signalisation comme JAK/STAT, SAPK/MAPK et PI-3K/AKT/mTOR (Asquith and McInnes 2007; Malemud and Reddy 2008; Malemud and Pearlman 2009; Wu and Mohan 2009; Malemud 2010), ainsi que des activités anormales de Syk (*Spleen Tyrosine Kinase*) (Coffey, DeGuzman et al. 2012), BMX (*Bone Marrow Kinase*) (Gottar-Guillier, Dodeller et al. 2011) ont toutes été retrouvées chez des patients avec une PR active.

Il a ensuite été observé que l'activité de BMX était nécessaire pour la phosphorylation de p38 kinase et JNK ainsi que pour l'activation de NFκB (Gottar-Guillier, Dodeller et al. 2011; Razani, Reichardt et al. 2011). Il semble donc que BMX soit responsable de l'activation de p38, JNK et NFκB, qui sont essentiels pour la survie et l'apoptose des cellules de la réponse inflammatoire (Malemud 2013). Les différentes voies de signalisation sont illustrées dans la Figure 4.

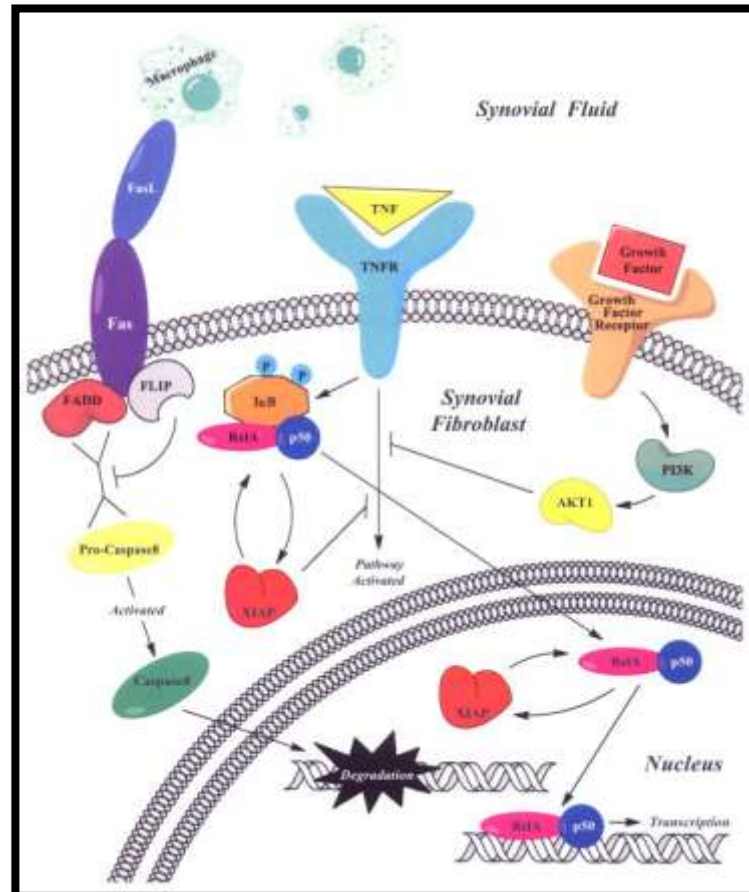


Figure 4: Cascades de signalisation cellulaire d'apoptose et de survie dans les fibroblastes de la synoviale rhumatoïde (d'après Malemud 2013).

4. Production d'auto-anticorps

a) Le facteur rhumatoïde

Le facteur rhumatoïde (FR) a été décrit pour la première fois il y a environ 90 ans (Meyer 1922), depuis un grand nombre d'études se sont intéressées à son incidence, sa nature et sa spécificité. L'anticorps FR est présent chez environ 75% des patients, mais sa spécificité reste limitée puisqu'il est également trouvé chez des patients avec d'autres maladies autoimmunes, des maladies infectieuses et, avec une faible fréquence, dans la population saine (3 à 5%). Malgré sa faible spécificité, le FR reste néanmoins utilisé partout dans le monde comme un marqueur de diagnostic dans la PR. En effet, il figure dans la liste des critères ACR pour la PR.

Les FRs sont des anticorps dirigés contre le fragment cristallisable de la molécule Immunoglobuline G (IgG). Ils sont retrouvés dans tous les sous-groupes d'immunoglobulines

(IgE, IgM, IgA et IgG). Le FR-IgG a une capacité d'auto-fixation, ce qui induit la formation d'un large complexe immunitaire. La question de savoir si le FR fait partie des symptômes de la PR reste encore non élucidée bien qu'il soit retrouvé de façon significative chez les cas ayant une inflammation agressive des articulations (van Boekel, Vossenaar et al. 2002).

Le test de mise en évidence est réalisé dans les laboratoires cliniques par la technique d'ELISA (*Enzyme-Linked Immuno-Sorbent Assay*). Certaines études suggèrent que le taux élevé d'IgA-FR est un paramètre de l'activité de la maladie (Houssien, Jonsson et al. 1997; Pai, Pai et al. 1998). La détermination en routine combinant IgM-FR, IgG-FR et IgA-FR est recommandée pour augmenter la sensibilité, la spécificité du diagnostic et la valeur prédictive (Swedler, Wallman et al. 1997; Jonsson, Steinsson et al. 1998).

b) Les anticorps anti-CCP

Les anticorps anti-peptide cyclique citrulliné ont initialement été identifiés dans les années 60 sous l'appellation d'*anti-perinuclear factor* (APF) comme étant des auto-anticorps spécifiques à la PR réagissant contre les granules de keratohyaline qui sont agrégés dans la région peri-nucléaire des cellules épithéliales buccale chez l'homme (Nienhuis and Mandema 1964). En 1979, une nouvelle étude propose les anticorps anti-keratine (*anti-keratin antibodies*, AKA) comme étant d'autres auto-anticorps spécifiques de la PR (Young, Mallya et al. 1979). Ces deux auto-anticorps très spécifiques de la PR, sont suspectés d'être le même et seul auto-anticorps puisqu'ils ont tendance à être détectés simultanément. En outre, malgré leurs spécificités, ces deux auto-anticorps ne sont pas cliniquement utilisés en routine car leur antigène était encore inconnu et les techniques mis en place pour leur détection étaient difficiles à mettre en œuvre (Mimori 2005).

C'est en 1993 que l'antigène d'AKA a été mis en évidence pour la première fois en étant reconnu chez 75% des patients PR. C'est une protéine de 40kDa isolée à partir des tissus de la peau humaine et appelée filagrine (Simon, Girbal et al. 1993). Il a été démontré par la suite qu'APF reconnaissait aussi cette molécule avec la même spécificité qu'AKA, puisque la molécule cible d'APF reconnaissait le précurseur de la filagrine, la pro-filagrine (Sebbag, Simon et al. 1995). Une désamination d'un résidu arginine en une citrulline a été mise en évidence dans les molécules de la filagrine, cette réaction étant réalisée par l'enzyme peptidylarginine-désaminase (PADI) (Figure 5). Cette citrulline s'est révélée être très

importante dans la reconnaissance de l'épitope par les auto-anticorps dans la PR (Schellekens, de Jong et al. 1998). Le peptide cyclique citrulliné (*cyclic citrullinated peptide*, CCP) est une molécule dans laquelle deux résidus Serine sont convertis en Cystéine, ce qui donne une forme cyclique à la molécule grâce à la création de ponts di-sulfures. Cette conformation est retrouvée dans la majorité des épitopes de la filagrine (Schellekens, Visser et al. 2000).

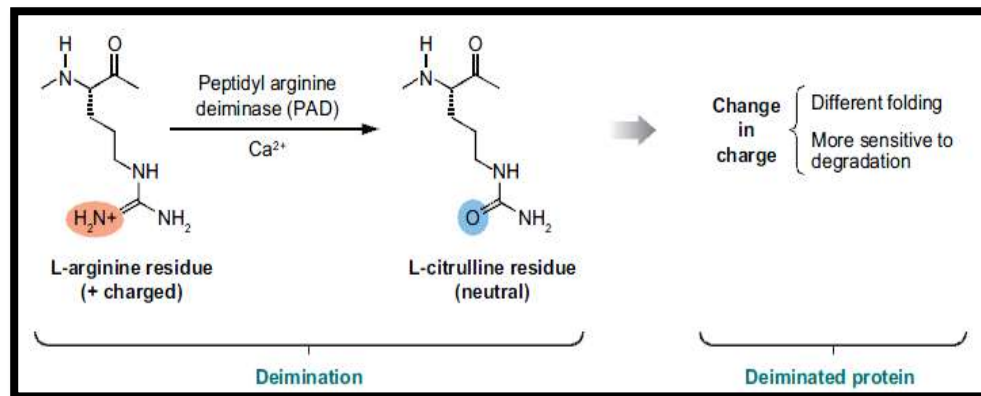


Figure 5: La réaction de citrullination (d'après Klareskog, Ronnelid et al. 2008).

La spécificité des anticorps anti-CCP (ACPA)/filagrine dans la PR est supérieure à 90% dans la plupart des études, avec une sensibilité qui varie entre 33% et 87,2%. Cette différence pourrait souligner une histoire génétique et raciale, ou tout simplement une différence dans les antigènes ou les techniques utilisées. Les anti-CCP ont été également décrits comme étant spécifiques également à la PR débutante puisque 50 à 70% des patients avec une PR débutante sont ACPA positifs (Avouac, Gossec et al. 2006; Nishimura, Sugiyama et al. 2007). Peu d'individus dans une population de contrôles sains sont positifs pour les anti-CCP (environ 2%) (Klareskog, Ronnelid et al. 2008).

c) Les autres auto-anticorps

En 1989, l'équipe de Hassfeld et collaborateurs a décrit un antigène de 33kDa reconnu par le sérum de patients PR. Cet antigène correspond à RA33 dont les auto-anticorps ont été détectés dans le sérum de 36% des patients contre seulement 1% dans le sérum normal (Hassfeld, Steiner et al. 1989).

La protéine calpaïne est retrouvée avec des taux élevés dans la synoviale lors de l'inflammation et semble être sécrétée par les cellules synoviales. Il a été suggéré que la calpaïne jouait un rôle dans dégradation du cartilage dans la PR (Yamamoto, Shimizu et al.

1992; Szomor, Shimizu et al. 1995). Des anticorps anti-calpastine, inhibiteurs de la calpaïne, ont été détectés chez 45% des patients PR, mais également dans d'autres maladies autoimmunes (Despres, Talbot et al. 1995; Mimori, Sukanuma et al. 1995). Ces auto-anticorps anti-calpastine pourraient augmenter l'activité de la calpaïne induisant la destruction du cartilage (Yamamoto, Shimizu et al. 1992).

Les patients PR produisent plusieurs auto-anticorps dirigés contre plusieurs auto-antigènes, mais la plupart d'entre eux sont non spécifiques à la PR. Ces anticorps non spécifiques peuvent être dirigés contre une large variété de protéines du cartilage comme par exemple le collagène et les fibronectines. Des anticorps anti-phospholipides et anti-neutrophiles peuvent également être trouvés dans le sérum de patients PR (van Boekel, Vossenaar et al. 2002).

D. Traitements

Durant cette dernière décennie, les stratégies thérapeutiques de la PR ont largement changé et sont orientées selon trois axes : la réduction de l'inflammation, le ciblage moléculaire spécifique à des mécanismes impliqués dans la pathogénicité et les traitements personnalisés. Plusieurs études cliniques ont montré que le traitement précoce et agressif avec les DMARDs (*Disease-modifying antirheumatic drugs*) conventionnels, comme le méthotrexate (MTX), les glucocorticoïdes et autre anti-inflammatoires, peut être efficace pour contrôler l'activité inflammatoire et le développement des érosions chez plusieurs patients (Mottonen, Hannonen et al. 1999; Svensson, Boonen et al. 2005).

Avec le développement des biothérapies, le premier traitement ciblé a vu le jour en 1998 et est utilisé pour le blocage des cytokines de la famille du TNF. L'Etanercept, premier inhibiteur du TNF α approuvé pour l'utilisation clinique, est une protéine de fusion recombinante qui agit comme un récepteur soluble en fixant le TNF soluble, empêchant ainsi sa fixation avec son récepteur. Il peut être utilisé en monothérapie (Genovese, Bathon et al. 2002) ou combiné avec le MTX (Emery, Breedveld et al. 2008), ce qui permet d'avoir un meilleur résultat qu'avec le MTX ou l'etanercept seuls (van der Heijde, Klareskog et al. 2006).

Le succès de ce nouveau traitement a permis le développement de nouvelles molécules qui ciblent plusieurs intervenants dans la réponse inflammatoire (Tableau 2 et Figure 6).

L'Adalimumab, l'Infliximab, le Golimumab et le Certolizumab ont également pour cible thérapeutique le TNF. L'Anakinra cible l'IL1 alors que le Tocilizumab est spécifique pour l'IL6. D'autres molécules ciblent des marqueurs cellulaires comme le CD20 des lymphocytes B pour bloquer l'activité de ces derniers (le Rituximab) ou les CD80/CD86 pour empêcher la co-stimulation des lymphocytes T (Abatacept) (Colmegna, Ohata et al. 2012; Vivar and Van Vollenhoven 2014).

L'administration de ces traitements a révélé une grande variabilité de réponses chez un pourcentage important de patients PR. Les causes biologiques de ces variations ne sont pas encore identifiées, mais plusieurs hypothèses existent, comme la différence d'expression des cytokines chez les patients (Ulfgren, Grondal et al. 2000). De récentes études ont montré que les patients ayant une forte expression de TNF α répondaient mieux aux anti-TNFs (Wijbrandts, Dijkgraaf et al. 2008), alors que les individus avec un taux élevé d'ACPA ou de FR et une prolifération de LBs synoviaux étaient plus réceptifs aux traitements ciblant les LBs.

Tableau 2 : Résumé des DMARDs biologiques disponibles pour le traitement de la PR (d'après Vivar and Van Vollenhoven 2014)

Name	Target	Format	Mechanism	Administration	Approximate half-life*
Etanercept	TNF	Recombinant human fusion protein of the TNF receptor and the Fc portion of IgG1.	Works as a decoy receptor. It binds to soluble TNF, blocking the binding to its receptor	sc. injection once (50 mg), or twice (25 mg) a week	3-6 days
Adalimumab	TNF	Fully human IgG1 MAb	Binding to TNF	sc. injection once every second week	13 days
Infliximab	TNF	Chimeric murine-human IgG1 MAb	Binding to soluble and mb bound TNF	iv. infusion every 4-8 weeks	9 days
Golimumab	TNF	Fully human IgG1 MAb	Binding to soluble and mb bound TNF	sc. injection once a month or iv. Infusion at 0 and 4 weeks, thereafter every 8 weeks	13 days
Certolizumab-pegol	TNF	Humanized pegylated anti-TNF Fab' fragment	Binding to TNF	sc. injection once every second week	14 days
Anakinra	IL-1	Recombinant human IL-1 receptor antagonist	Binding to IL-1 type-1 receptor	sc. injection once a day	4-6 hours
Tocilizumab	IL-6	Humanized recombinant IgG1 MAb	Binding to soluble and membrane bound IL-6 receptor	iv. infusion every 4 weeks	10-13 days
Rituximab	B cells	Chimeric murine-human IgG1 MAb	Binding to CD20 and depletion of CD20+ B cells	Two initial infusions 14 days apart. Courses may be repeated every 6 months or more frequently depending on disease activity	18 days (range: 5-76 days)
Abatacept	T cells	Recombinant human fusion protein of the extracellular domain of CTLA-4 and the Fc portion of IgG1	Binding to CD80/ CD86, blocking T-cell co-stimulation	iv. infusion every 4 weeks or sc. injection once a week	13 days (range: 8-25 days)

CTLA-4, cytotoxic T lymphocyte associated antigen 4; DMARDs, disease-modifying anti-rheumatic drugs; IgG, immunoglobuline G; IL, interleukin; iv, intraveineux; MAb, anticorps monoclonal; mb, membrane; sc, sous-cutané; TNF, tumor necrosis factor.

*Le temps de demi-vie correspond à celui nécessaire pour avoir un effet biologique indiqué par les fabricants.

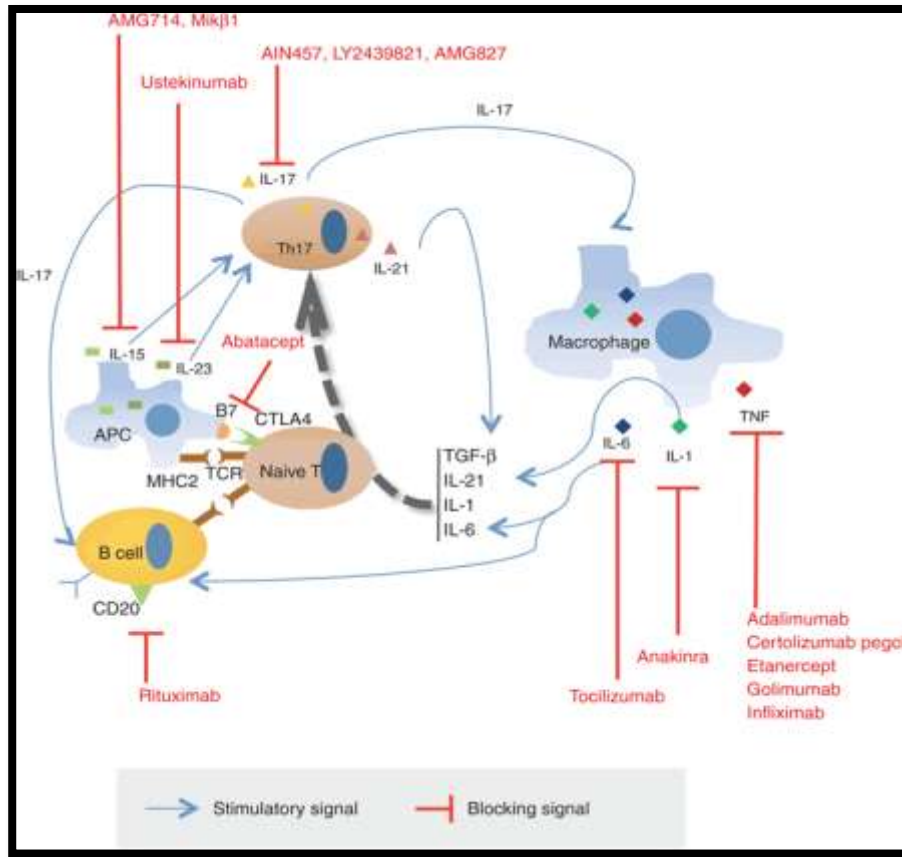


Figure 6 : Modes d’actions des traitements ciblés disponibles (d’après Colmegna, Ohata et al. 2012).

II. Etiologie de la Polyarthrite Rhumatoïde

La PR est connue pour être une maladie complexe et multifactorielle. Plusieurs facteurs peuvent intervenir dans la susceptibilité de développement de la PR. Les études d’agrégation familiale dans la PR montrent une prévalence qui augmente chez les apparentés de premier degré des patients (2 à 12%) (del Junco, Luthra et al. 1984; Silman, MacGregor et al. 1993). Des études faites en Finlande et en Angleterre sur des jumeaux montrent un taux de concordance plus élevé chez les jumeaux monozygotes que chez les dizygotes. Selon MacGregor et collaborateurs, les facteurs génétiques contribuent pour environ 60% du risque total dans la PR, ce qui suggère que la composante génétique a un impact significatif sur la susceptibilité à la maladie et qu’il existe cependant une part environnementale liée à la maladie (MacGregor, Snieder et al. 2000). D’autres facteurs stochastiques semblent intervenir également dans l’évolution de la maladie (Seldin, Amos et al. 1999).

A. La composante génétique

1. L'héritabilité

L'héritabilité est la proportion de la variance phénotypique qui peut être attribuée à la variance génétique plutôt qu'à la variance environnementale. L'ensemble de l'héritabilité de la PR avec ACPA+ et ACPA- est similaire (60-70%) (MacGregor, Snieder et al. 2000; van der Helm-van Mil, Wesoly et al. 2005). Il existe néanmoins une controverse quant à la contribution du locus HLA (explicité dans la page 18) dans la variabilité génétique. Précédemment, la contribution du locus HLA a été estimée à 37% (Gregersen, Silver et al. 1987; Deighton, Walker et al. 1989; van der Woude, Houwing-Duistermaat et al. 2009) alors que de récentes études l'estiment à seulement 11% (van der Helm-van Mil, Wesoly et al. 2005; van der Woude, Houwing-Duistermaat et al. 2009). La différence entre ces deux estimations peut être due au fait que la susceptibilité de la PR est aussi associée à des allèles protecteurs en plus des allèles partagés (*shared epitope*, SE). De plus, la contribution du locus HLA à la variabilité dans la PR ACPA+ est différente de celle dans la PR ACPA- (40% et 2% respectivement) (van der Woude, Houwing-Duistermaat et al. 2009).

2. Approches des études génétiques

a) Analyses de liaison

Elles ont pour but d'identifier les régions génomiques contenant des gènes de prédisposition aux maladies en ciblant les individus apparentés. L'approche la plus communément utilisée, lorsqu'il n'y a pas d'a priori sur le mode de transmission d'un ou de plusieurs facteurs génétiques, est celle de l'étude des paires de germains atteints.

Théoriquement, les apparentés atteints devraient avoir plus d'haplotypes identiques partagés par descendant dans la région du variant causal. Ainsi, les régions chromosomiques montrant le plus d'allèles partagés ont plus de chance de contenir des gènes de susceptibilité (Abel and Dessein 1998) (Figure 7). Dans une étude de liaison, le génome entier peut être criblé partiellement ou complètement (*Whole Genome Scan*, WGS) en utilisant des marqueurs microsatellites ou par SNP (*Single Nucleotide Polymorphism*) (Orozco, Rueda et al. 2006).

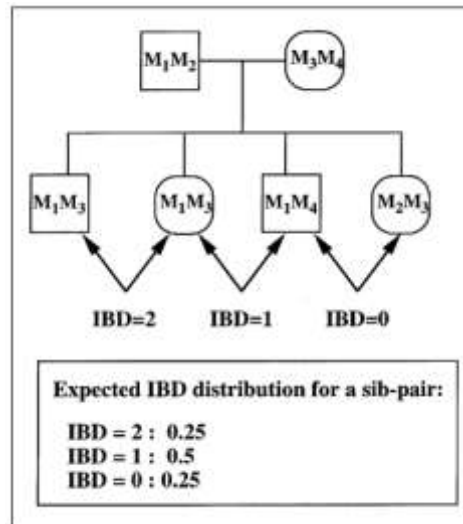


Figure 7: Principe de l'analyse des paires de germains (d'après Abel and Dessein 1998).

Plusieurs études de liaisons génétiques ont été menées pour la PR, dans différentes populations (Européenne, Américaine et Japonaise). Le résultat le plus probant est la liaison avec le chromosome 6q21 dans la région du locus HLA. Plusieurs autres liaisons ont été suggérées dans des régions en dehors de celle de HLA, mais peu d'entre elles ont montré une liaison significative ou ont été répliquées dans d'autres études (Merriman, Twells et al. 1997; Gaffney, Kearns et al. 1998; Shai, Quismorio et al. 1999).

Les régions identifiées étaient de grandes tailles, il a donc été nécessaire d'affiner la cartographie en utilisant une densité plus élevée de marqueurs suivie d'une approche "gènes candidats" afin de mettre en évidence le variant responsable de l'association avec ces régions. Grâce à cette approche, le gène *PTPN22* (*protein tyrosine phosphatase non-receptor22*) a été identifié dans la région 1p13 et est devenu le gène le mieux caractérisé dans la PR, après le gène *HLA-DRB1* (Begovich, Carlton et al. 2004).

b) L'approche gène candidat

Cette approche part du postulat qu'un gène est potentiellement impliqué dans la pathogénicité de la maladie. Le gène candidat choisi est généralement analysé dans une cohorte cas/contrôles. Les génotypes et les fréquences alléliques sont déterminés par génotypage ou séquençage direct puis comparés entre les cas et les contrôles (Marian 2012). Dans les études d'association, l'objectif est de détecter si un marqueur génétique est impliqué dans la susceptibilité de la maladie. Le *design* de l'étude est basé sur la sélection d'un set bien

défini d'individus atteints (les cas) et un groupe de contrôles sains, tous les deux issus de la même population (Lander and Schork 1994).

L'approche gène candidat est plus puissante que les analyses de liaisons génétiques dans la détection des effets mineurs contribuant à la prédisposition de la maladie. Néanmoins, pour avoir une étude avec des résultats probants, le *design* de l'étude doit être fait avec précaution pour ne pas avoir des associations faussement positives (Cordell and Clayton 2005). Le choix du gène candidat peut se faire selon son implication fonctionnelle dans les mécanismes pathogéniques (gène candidat fonctionnel) et, ou bien en considérant sa localisation dans une région génomique auparavant décrite dans les études de liaison (Dunbar 2006).

Un grand nombre de polymorphismes de gènes impliqués dans la pathogénicité de la PR a été investigué mais les résultats sont conflictuels, puisque peu d'entre eux ont été répliqués (John and Worthington 2001). Tout de même, plusieurs gènes candidats ont été associés à la PR comme les gènes *PTPN22*, *CTLA-4*, *PADI4* et *MIF* qui seront décrits par la suite.

c) Les analyses association génome entier (GWAS)

Les GWAS permettent d'investiguer les maladies fréquentes, en testant l'association de variations génomiques fréquentes sur de grandes cohortes de cas/contrôles. Ces analyses sont réalisées à l'aide de puces de génotypage. Plusieurs facteurs peuvent faire varier la puissance des GWAS mais le plus déterminant reste la taille de l'échantillon utilisé. Malgré la sensibilité des GWAS, il est également nécessaire de répliquer les résultats trouvés. Les variants associés avec le phénotype dans la population initiale et dans le set de réplification sont considérés comme des facteurs de risque génétique.

Plus de 1200 GWAS ont été réalisées avec succès sur plus de 200 traits complexes en identifiant des milliers de loci associés (Marian 2012). Dans la PR, les GWAS réalisées ont été mises en œuvre pour détecter des associations avec des variants fréquents dans la population avec une fréquence de l'allèle mineur (*minor allele frequency*, MAF) supérieure à 5%. La plupart des variants identifiés à ce jour par les GWAS dans la PR n'expliquent qu'une modeste partie de la composante génétique avec un Odds Ratio (OR) maximal de 1.5 ou moins (Raychaudhuri, Remmers et al. 2008; Stahl, Raychaudhuri et al. 2010). Actuellement, la centaine de variants génétiques associés à la PR n'explique qu'environ 16% de sa

composante génétique (Stahl, Raychaudhuri et al. 2010; Okada, Terao et al. 2012; Okada, Wu et al. 2014)

3. Le gène HLA-DR

C'est au cours des années 1970 (Stastny 1976) que le locus HLA a été associé à la PR pour la première fois. Les analyses génétiques qui ont suivi ont permis d'affirmer que le locus HLA est le facteur génétique à risque le plus prédominant, son implication ayant été estimée à environ 30% dans la composante génétique (Deighton, Walker et al. 1989) puis corrigé à 11% (van der Helm-van Mil, Wesoly et al. 2005; van der Woude, Houwing-Duistermaat et al. 2009). Les allèles du gène *HLA-DRB1* associés à la PR codent pour la séquence « *shared epitope* » (SE) dans la troisième région hypervariable de la chaîne DRB1 (Gregersen, Silver et al. 1987). Les motifs d'acides aminés du SE, QKRAA, QQRAA et KKRAA, sont connus pour être associés à la susceptibilité de la PR, alors que le motif DERAA est par contre responsable d'un effet protecteur.

Dans la population caucasienne, la PR est plus particulièrement associée aux allèles HLA-DRB1 04*01, DRB1 04*04 et DRB1 01*01. L'analyse d'une cohorte de 91 paires de jumeaux monozygotiques a montré une concordance accrue pour la PR observée dans les paires présentant les allèles SE. De plus, un risque 5 fois plus élevé pour la PR a été observé chez les jumeaux homozygotes pour les allèles SE comparés à ceux négatifs pour les allèles SE (Jawaheer, Thomson et al. 1994). Une exploration faite sur l'association des facteurs génétiques avec les auto-anticorps a montré que les allèles SE n'étaient pas associés au FR, la différence de pourcentage entre les patients FR-positifs et FR-négatifs n'étant pas significative (Michou, Teixeira et al. 2008). Cette même étude a démontré une association significative entre la présence des ACPA et les allèles SE, à l'état hétérozygote comme à l'état homozygote. Ceci est confirmé par un autre travail plus récent qui détaille la titration des ACPA en fonction du nombre de copies de l'allèle SE (Balsa, Cabezon et al. 2010).

En 2012, une étude a démontré, en reprenant les données de la GWAS de 2010 (Stahl, Raychaudhuri et al. 2010), que le risque de PR associé au gène *HLA-DRB1* était fortement corrélé avec le résidu acide aminé à la position 11 localisé en dessous du sillon de fixation de l'antigène DRβ1. Les acides aminés 71 et 74 de la chaîne latérale constituant la surface du sillon de fixation de l'antigène sont également associés avec la PR. De plus, cette étude a trouvé des allèles à risque indépendants dans les gènes *HLA-B* et *HLA-DPBI* et dans les deux

cas, les signaux de ces deux régions étaient expliqués par une variation d'un seul acide aminé situé en dessous du sillon de fixation de l'antigène respectif. Il existe donc des variants génétiques dans *HLA-B*, *HLA-DRB1* et *HLA-DPB1* affectant en total 5 positions d'acides aminés et qui expliquent presque la totalité de variance du risque associé à la PR causé par la région du CMH (Figure 8) (Viatte, Plant et al. 2013).

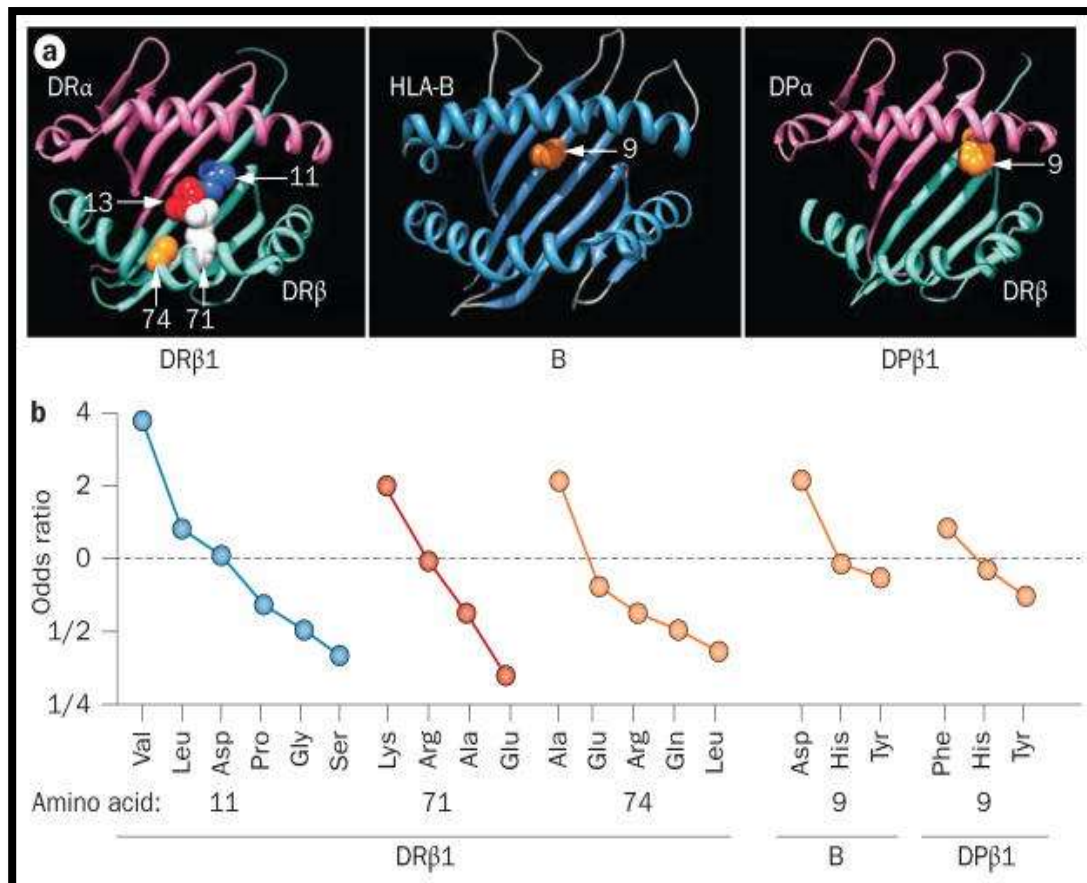


Figure 8: Sillon de fixation des antigènes et la substitution des acides aminés de HLA et leur influence sur la susceptibilité de la PR. a) Modèles en trois dimensions pour la molécule du CMH de classe I (*HLA-B*) et de classe II (*HLA-DRB1* et *HLA-DPB1*). b) odds ration des associations avec la PR en fonction des acides aminés à la position 11, 71, 74, et 9 pour les deux régions (DRβ1 et DPβ1) (d'après Viatte, Plant et al. 2013).

4. Le gène *PTPN22*

C'est en 2004 qu'une association significative entre le gène *PTPN22* (*protein tyrosine phosphatase non-receptor22*) et la PR a été démontrée (Begovich, Carlton et al. 2004). Ce gène code pour une phosphatase qui inhibe la voie de signalisation des TCRs en déphosphorylant les kinases de la famille Src, comme Lck ou Fyn, bien qu'il soit évident que l'allèle à risque affecte l'activité enzymatique de la phosphatase (Rhee and Veillette 2012). Le polymorphisme en question est de type SNP (rs24766601), et modifie la séquence protéique

Arg620 en Trp620. L'équipe de Begovich et collaborateurs en 2004, a décrit une association de la PR avec l'allèle mineur T OR=1,65 (IC [1,23-2,20]). De plus les génotypes (TT et TC), induits par le polymorphisme, confèrent un risque pour la maladie plus élevé que le génotype CC. Une relation entre les variants alléliques de *PTPN22* et les auto-anticorps a été décrite dans une étude menée par Orozco et son équipe en 2007, et qui a montré que l'allèle T, combiné à la présence des ACPA, était significativement associé au développement de la PR (OR de 10,06 ; IC [1,88 ; 53,73]). Ce résultat reflète la haute spécificité de l'allèle T, combiné aux ACPA, pour la polyarthrite rhumatoïde, suggérant que le variant allélique T pourrait influencer le devenir de la maladie en influençant la production des ACPA (Orozco, Pascual-Salcedo et al. 2008).

Par ailleurs, l'influence de ce polymorphisme sur la réponse immunitaire reste controversée puisque en 2005 un gain de fonction a été décrit (Vang, Congia et al. 2005), alors que les études fonctionnelles réalisées par la suite étaient en contradiction. En effet, en 2011, Zhang et collaborateurs ont observé que le l'allèle T du SNP rs24766601 était un allèle de perte de fonction qui induit la déstabilisation de la protéine PTPN22, le même résultat a été observé avec son homologue chez la souris (Zhang, Zahir et al. 2011). Le variant de la phosphatase devient alors la cible de dégradation par les protéases calpaines et par ubiquitination via le protéasome, ce qui a pour effet l'augmentation, l'activation et la sélection positive des LTs, mais également les cellules dendritiques et des LBs. En 2012, *PTPN22* a été associé au développement dans le thymus des Treg (Brownlie, Miosge et al. 2012; Maine, Hamilton-Williams et al. 2012) et ainsi qu'à d'autres mécanismes moléculaire comme le déséquilibre dans l'expression des variants transcrits de *PTPN22* (Ronninger, Guo et al. 2012); (Harrison, Southam et al. 2012).

5. Les autres facteurs de risque génétique mineurs

D'autres loci de susceptibilité dans la PR ont été identifiés, par exemple, des variants du gène *CTLA-4* (*cytotoxic T-lymphocyte associated antigen 4*), ou encore dans les séquences des chaînes alpha et beta du récepteur de l'interleukine 2 (IL2) (Consortium 2007). Une GWAS réalisée en 2010, regroupant les résultats de plusieurs études d'associations, a mis en évidence l'association de sept nouveaux loci avec la PR, conduisant à plus de 30 allèles à risque identifiés (Stahl, Raychaudhuri et al. 2010). La dernière GWAS en date est celle de 2014 qui a été réalisée avec de très grandes cohortes (29 880 cas-index et 73 758 contrôles) et environ 4 millions de SNPs. Cette étude a permis l'identification de 42 nouveaux loci de

risques associés à la PR sur le génome entier, ce qui amène le nombre total de loci à risque dans la PR à 101 dont 98 correspondent à des gènes candidats. Elle a démontré également que ces gènes sont des cibles d'approches thérapeutiques déjà approuvés (Okada, Wu et al. 2014). Plusieurs des protéines de ces facteurs mineurs peuvent interagir ensemble (Figure 9), ce qui augmente la susceptibilité de la maladie (Nakaoka, Cui et al. 2011; Rossin, Lage et al. 2011).

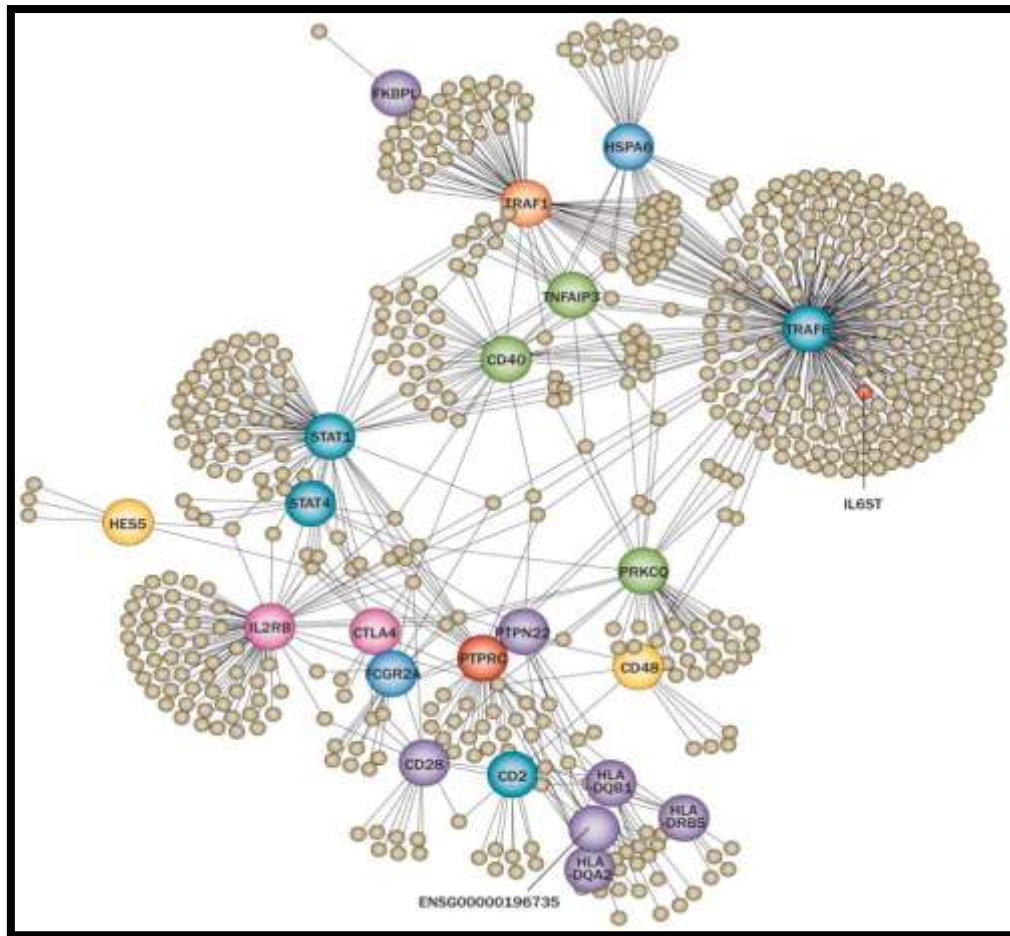


Figure 9: Réseau d'interaction entre les différents variants associés à la PR (d'après Viatte, Plant et al. 2013).

B. La composante environnementale

1. L'exposition au tabac

Il est maintenant établi que l'exposition au tabac est le facteur environnemental majeur dans l'accroissement du risque pour la PR. L'étude de cette voie d'investigation pourrait aider à mieux élucider la pathologie moléculaire de la PR (Stolt, Bengtsson et al. 2003).

Depuis la fin des années 80, la recherche de relations entre ce facteur et la PR n'a cessé de s'intensifier et surtout de se spécialiser. La première étude faite sur cette association a été réalisée en 1987 par une équipe anglaise et qui a démontré une forte relation entre l'habitude de fumer et la PR (Larsen 1995). D'autres recherches sont venues confirmer la première étude de 1987 avec de nouvelles informations : le risque lié au tabac est dépendant du nombre de paquets fumés par an et, également, dépendant de l'âge (Hutchinson and Moots 2001). De nombreux travaux ont continué sur le même chemin en essayant à chaque fois d'associer plus précisément ce facteur à la PR, en le reliant aux facteurs génétiques et aux auto-anticorps.

2. Les autres facteurs

Comme la PR touche plus de femmes que d'hommes, les études épidémiologiques se sont intéressées à l'influence de la grossesse, de l'allaitement et des facteurs hormonaux endogènes ou exogènes comme facteurs de risque dans la PR. Il a été démontré que au cours de la grossesse, le risque de développer la maladie était faible alors que dans l'année qui suit l'accouchement, le risque devient plus élevé (Silman, Kay et al. 1992). De plus, l'allaitement a été associé à un risque 5 fois plus élevé dans la PR (Brennan and Silman 1994). Les facteurs hormonaux exogènes comme la pilule ou les traitements de substitution ne modifient pas l'incidence de la PR mais semble retarder son début et sa sévérité (Cutolo, Seriola et al. 2002).

Plusieurs agents infectieux ont été rapportés comme étant des facteurs à risque dans la PR, comme le virus Epstein-Barr, la bactérie *E. Coli* et les mycobactéries. Ces associations entre la PR et les agents infectieux ont été suggérées par l'augmentation du taux d'anticorps ou de la charge microbienne chez les patients, comparés à d'autres individus. Mais cette association reste controversée et doit être complétée par d'autres études épidémiologiques et fonctionnelles (Klinman 2003).

Chapitre 2 : Les variations de Nombre de Copies

I. Définition des Variations de nombre de Copies (CNVs)

Les SNP étaient considérés comme les variations génomiques prédominantes et susceptibles d'avoir un effet important sur le phénotype (Freeman, Perry et al. 2006). Cependant, les études d'association sur génome entier réalisées au cours des dernières années n'ont expliqué qu'une faible part de la composante génétique (Stahl, Raychaudhuri et al. 2010). C'est pour cela que, depuis une dizaine d'année, les Variations de Nombre de Copies, sont considérées comme des variations génomiques indispensables à investiguer dans la composante génétique des maladies complexes (Almal and Padh 2012).

En 2004, deux équipes ont montré que les CNVs étaient largement représentés dans le génome d'individus sains (Iafrate, Feuk et al. 2004; Sebat, Lakshmi et al. 2004). Ce résultat a été depuis reproduit dans plusieurs populations (Tuzun, Sharp et al. 2005; Conrad, Andrews et al. 2006; McCarroll, Hadnott et al. 2006). Il apparaît alors à la communauté scientifique l'importance des CNVs et de l'étude de leurs implications dans la variation des phénotypes.

Une définition des CNVs a été proposée en 2006 par l'équipe de Feuk et ses collaborateurs. Un CNV est défini comme une variation de copies d'une région génomique dont la séquence est supérieure à 1kb et le nombre de copies diffère par rapport à un génome de référence. Il est considéré comme un polymorphisme si sa fréquence dans la population est supérieure ou égale à 1%. Après concertation, il a été décidé que la nomination "CNV" regroupait trois sous-groupes: les larges variants de nombre de copies (LCVs, (Iafrate, Feuk et al. 2004)), les polymorphisme de nombre de copies (CNP, (Sebat, Lakshmi et al. 2004)) et les variants de taille intermédiaire (ISV, (Tuzun, Sharp et al. 2005)). Les retrotransposons ainsi que les insertions/délétions des éléments transposables ont été exclu du classement lors d'un atelier "*The effect of genomic structural variation on gene expression and human disease*" organisé par le "*Wellcome Trust Sanger Institute*" en 2005 (Wellcome Trust Sanger Institute, 2005). Les CNVs peuvent donc présenter, au sein du génome, soit un gain de nombre de copies (insertion ou duplication) soit une perte (délétion ou des génotypes nuls) et ceci par rapport à une séquence de référence du génome (Scherer, Cheung et al. 2003).

Ces modification peuvent apparaître lors de la méiose comme à la mitose, tel que l'a mis en évidence une étude qui a montré que même de vrais jumeaux pouvaient avoir un

nombre de copies différent et que les CNV peuvent être différents entre les organes ou tissus au sein d'un même individu (Piotrowski, Bruder et al. 2008).

Les CNVs semblent également jouer un rôle prépondérant dans l'évolution et l'adaptation des espèces. En effet, il a été suggéré que lorsqu'une variation altère une fonction d'un gène en changeant son niveau d'expression, ce variant altère indirectement une capacité de l'organisme et agit indirectement sur sa sélection. Ainsi, les CNVs comprenant des gènes fonctionnels sont plus probablement dus à une adaptation, par rapport aux CNVs localisés dans les régions intergéniques. Ces adaptations peuvent être dues, pour la plus part, à une sélection purificatrice (ou négative) de CNVs contenant des gènes (Cooper, Nickerson et al. 2007) et, avec une moindre fréquence, à des CNVs soumis à une sélection positive (Kryukov, Pennacchio et al. 2007; Conrad, Pinto et al. 2010). La sélection purificatrice agit sur des variants qui réduisent certaines capacités afin de les éliminer de la population. De plus, les délétions des larges CNVs (>500kb) contiennent généralement des gènes fonctionnels, indiquant une forte sélection négative (Charlesworth, Morgan et al. 1993; Itsara, Wu et al. 2010). La sélection positive, quant à elle, agit sur les nouveaux variants ou les variants neutres qui deviennent avantageux en fonction de l'environnement. Mais, parmi les milliers de CNVs identifiés, seule une petite fraction montre une évidence empirique de sélection positive chez les êtres humains (Gonzalez, Kulkarni et al. 2005; Kidd, Newman et al. 2007; Xue, Sun et al. 2008; Hardwick, Machado et al. 2011). Parmi ces gènes qui ont subi une sélection positive, il y a les familles de gènes du système immunitaire, les gènes impliqués dans le développement neurologique et certains gènes intervenant dans les voies métaboliques (Johansson, Lundqvist et al. 1993; Hirayasu, Ohashi et al. 2008; McLean, Reno et al. 2011).

A. Formation des CNVs

1. Caractéristiques des CNVs

Le changement de nombre de copies requiert un changement dans la structure du chromosome en joignant deux séquences d'ADN préalablement séparées. De cette affirmation, deux types de réarrangement se distinguent : les réarrangements récurrents (ou fréquents) et les réarrangements non-récurrents.

Les premiers possèdent la même taille et des bordures fixes chez plusieurs individus. Ils surviennent dans des régions contenant des séquences appelées « *Low Copy Repeat* »

(LCRs) qui possèdent une forte homologie. Les CNVs issus de ces changements sont, en très grande partie, les résultats de recombinaison appelée Recombinaison Homologue Non-Allélique (*Non-Allelic Homologous Recombination*, NAHR) (Hastings, Lupski et al. 2009). Les réarrangements non-récurrents, quant à eux, sont de tailles variables et donc de bordures distinctes, et peuvent survenir à n'importe quel endroit du génome. Lorsqu'ils sont fréquents, ils partagent souvent une région génomique commune et se produisent entre des séquences très petites (2 à 15pb) sans recombinaison homologue via un mécanisme de *Non-Homologous End-Joining* (NHEJ) ou de *Fork Stalling and Template Switching* (FoSTeS) (Gu, Zhang et al. 2008) (Figure 10).

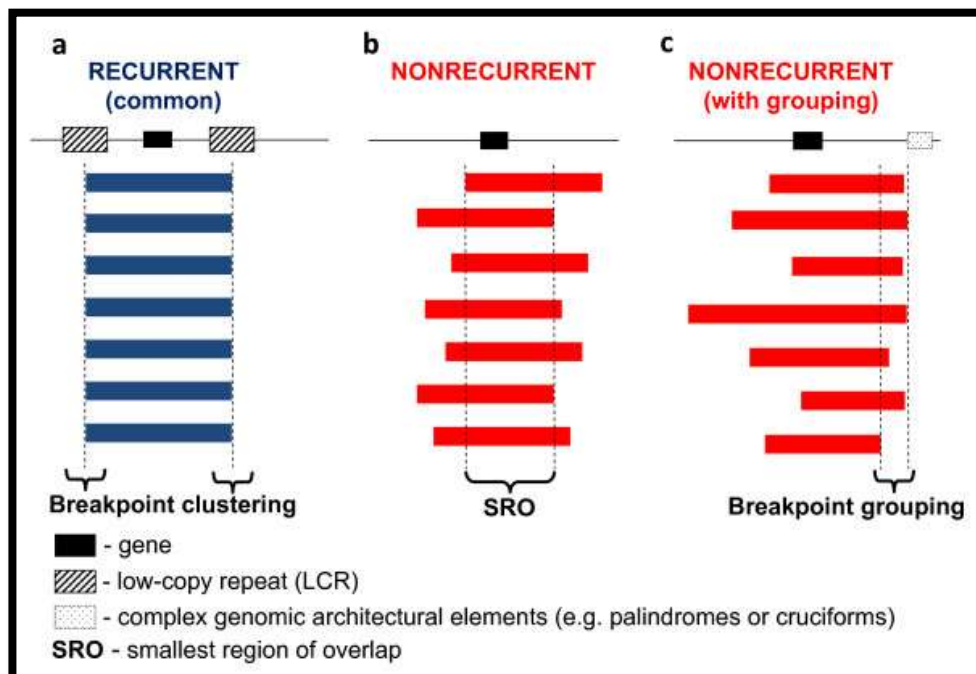


Figure 10: Exemples de réarrangements récurrents et non récurrents. a) réarrangements récurrents ayant les mêmes tailles et ayant des points de cassures (*Breakpoints*) regroupés. Ces réarrangements sont pour la plus part du à des recombinaisons homologues non alléliques (NAHR). Les deux blocs hachés représentent les *Low Copy Repeats* (LCRs). b) réarrangement non-récurrent avec des *breakpoints* dispersés dans le génome. Tous ces réarrangements partagent une région commune qui se chevauche (*Smallest Region of Overlap*, SRO). c) exemple de réarrangement non-récurrent avec des *breakpoints* groupés en une seule région (d'après Hastings, Lupski et al. 2009).

2. Exemple de réarrangements récurrents

a) *La Recombinaison Homologue Non-Allélique*

Les réarrangements récurrents sont dus, dans la majorité des cas, au NAHR qui se fait entre deux LCRs. Ces derniers, appelés aussi duplications segmentaires (*Segmental*

Duplication) (Stankiewicz and Lupski 2002; Shaw and Lupski 2004; Weiss, Shen et al.), sont des régions spécifiques de l'ADN avec une taille entre 10 à 300kb et qui ont entre 95% et 97% de similarité (Stankiewicz and Lupski 2002; Bailey and Eichler 2006). Lors des méioses ou mitoses, des LCR non alléliques peuvent s'aligner (faux-alignement). Un crossing-over entre ces deux régions induit un réarrangement chromosomique avec une duplication/délétion, lorsque les deux LCRs ont la même orientation sur un même chromosome (intrachromosomique), ou sur deux chromosomes différents (interchromosomique; Figure 11,a1). Cette recombinaison peut donner également des inversions lorsque les deux LCRs sont en opposition sur un même chromosome (Figure 11, a2) (Stankiewicz and Lupski 2002).

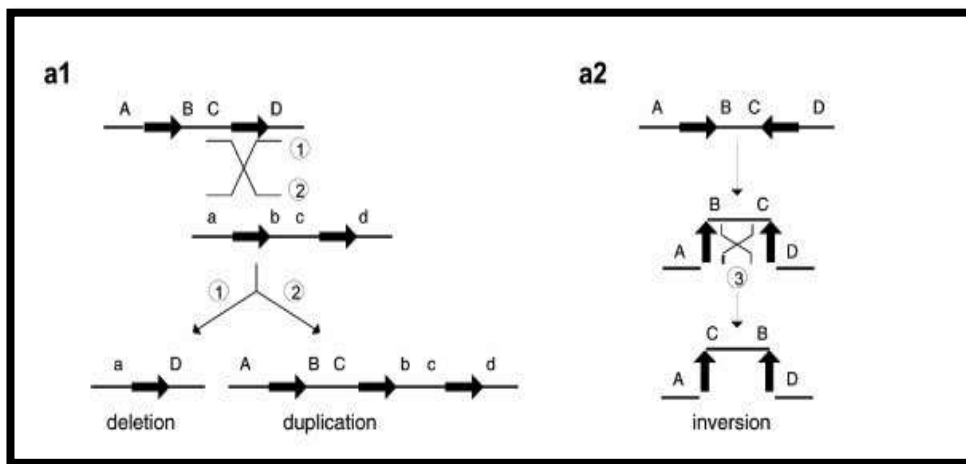


Figure 11: Réarrangements génomiques résultant d'une recombinaison entre deux LCRs. Les LCRs ont représentés par les flèches en gras qui indiquent l'orientation de ces fragments. a1) Recombinaison entre les deux régions répétées orientées dans le même sens induisant des délétions et/ou duplication. a2) Recombinaison entre deux LCRs inversés induisant une inversion. Les lettres en majuscule et en minuscule correspondent aux deux versions alléliques d'un gène. Les diagonales indiquent un évènement de recombinaison (d'après Hastings, Lupski et al. 2009).

L'équipe de Liu P et collaborateurs a montré que la fréquence des NAHR était fortement corrélée avec la taille des LCRs de part et d'autre de la région recombinée, mais aussi avec la distance entre ces deux répétitions (Liu, Lacaria et al. 2011). Ces segments doivent avoir une taille minimale partageant une forte similarité entre les LCRs et sont appelés *Minimum Efficient Processing Segment* (MEPS) (Gu, Zhang et al. 2008). Les MEPS chez l'être humain, et lors de la méiose, semblent nécessiter une longueur d'homologie entre 300 et 500pb, estimée lors de l'étude du réarrangement génomique dans certaines maladies (Reiter, Hastings et al. 1998). Lors des mitoses, les MEPS diffèrent de ceux des méioses, comme l'a démontré Steinmann et collaborateurs au niveau des délétions somatiques du gène

NF1, flanqué par des séquences ayant une longueur d'homologie de 114pb (Steinmann, Cooper et al. 2007). Mais ces conditions ne sont pas fixes puisque il y a aussi la distance entre les LCRs qui influe sur l'apparition des NAHR (Gu, Zhang et al. 2008).

De plus, ces NAHR peuvent être également le résultat de recombinaisons entre des séquences répétées telles que des séquences *Alu* ou LINEs (Lee, Carvalho et al. 2007).

3. Exemples de réarrangements non-récurents

a) Fusion des extrémités non-homologues

La fusion des extrémités non-homologues (*non-homologous end-joining*, NHEJ) est, avec la réparation homologue (*homologous repair*, HR), l'un des mécanismes majeur impliqué dans la réparation des cassures double-brins (*double-strand DNA break*, DSB) chez les eucaryotes et a été décrit chez tous les organismes, des bactéries jusqu'aux mammifères (Roth, Porter et al. 1985; Lieber, Ma et al. 2003; Weterings and van Gent 2004). Ce mécanisme est utilisé en routine dans les cellules humaines, que ce soit pour réparer les cassures (DSB) physiologiques, comme les recombinaisons des régions VDJ impliquées dans le système immunitaire, ou les DSB pathologiques induits par les radiations ionisantes ou les espèces réactives de l'oxygène (*Reactive Oxygen Species*, ROS). Les défaillances de NHEJ héritées causent environ 15% des immunodéficiences sévères combinées chez l'homme (Schwarz, Ma et al. 2003). Ce mécanisme est souvent considéré également comme le mécanisme majeur induisant des chromosomes transloqués dans le cancer (Lieber, Lu et al. 2008).

Les NHEJs se forment suivant quatre étapes: identification des cassures double-brins (DBS); création de ponts moléculaires entre les deux terminaisons cassées; modification des terminaisons pour les rendre compatibles et pour pouvoir faire la ligation; ligation des deux brins (Figure 12) (Weterings and van Gent 2004). Cette procédure souligne deux caractéristiques des NHEJs: la première est qu'ils ne nécessitent pas obligatoirement de LCRs ou de MEPS; la seconde est que ce mécanisme laisse des sites "cicatrices" qui informent sur la présence d'une ligation des terminaisons (Lieber 2008). En effet, il a été possible pour deux équipes, Nobile et collaborateurs ainsi que Toffolatti et collaborateurs, d'identifier des sites de cassure grâce aux données de séquençage obtenues à partir de 19 patients atteints de dystrophie musculaire ayant une délétion non-récurente au niveau des introns 47 et 48 du gène *DMD* (Nobile, Toffolatti et al. 2002; Toffolatti, Cardazzo et al. 2002). Ils ont trouvé que

cette délétion n'était pas flanquée de LCR mais qu'il existait des micro-homologies (2 à 4 nucléotides) dans 7 cas, des courtes insertions (1 à 5 nucléotides) dans trois cas et des courtes duplications (jusqu'à 25pb) dans 3 cas. Enfin, d'autres jonctions contenant des séquences courtes avec des origines inconnues et sans aucune micro-homologie sont observées.

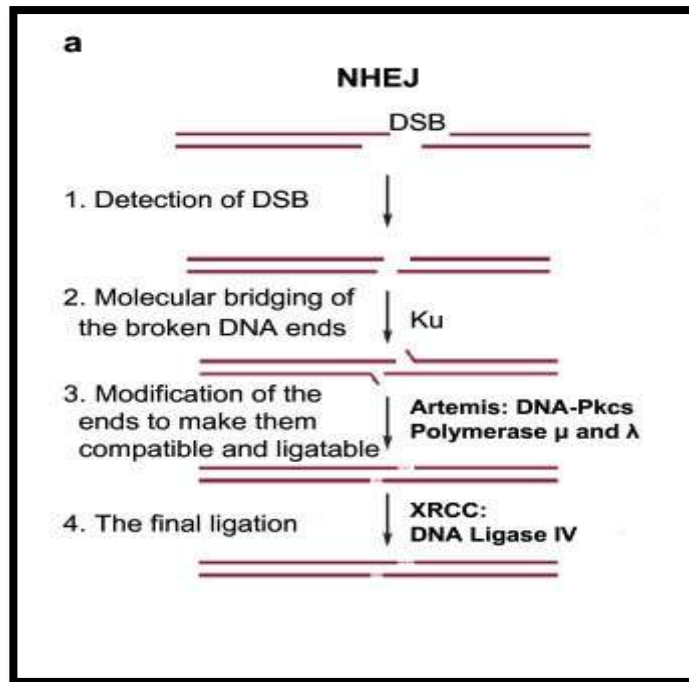


Figure 12: Non-homologous end-joining (NHEJ) chez les vertébrés. Une cassure sur les deux brins (*Double Strand Break*, DSB) se produit et est corrigée par le mécanisme NHEJ. Ku, Artemis et XRCC constituent la machinerie enzymatique catalysant la réaction de réparation (d'après Hastings, Lupski et al. 2009).

Ces évènements concordent avec les résultats du mécanisme NHEJ. Le plus intéressant est que 42% des cassures identifiées surviennent dans des éléments répétés de l'ADN tel que les LTRs, les LINES, les séquences Alu, les MIR et les MER2, mais aussi d'autres motifs connus comme induisant des DSB, comme le motif TTTAAA, présents à proximité de plusieurs jonctions (Nobile, Toffolatti et al. 2002; Toffolatti, Cardazzo et al. 2002).

b) Fork Stalling and Template Switching

En 2007, l'équipe de Lee et collaborateurs a mis en évidence un nouveau mécanisme de réarrangement en étudiant les régions entourant le gène *PLP1* (*Proteolipid Protein 1*) identifiant un réarrangement non-récurrent associé à l'apparition des délétions et duplications de ce gène. La cartographie des points de cassure (*Breakpoints*) dans cette région a révélé que ces recombinaisons n'induisaient pas de simples duplications ou délétion, mais des fragments

inversés ou transloqués dans d'autres régions. Des micro-homologies, de 2 à 5 nucléotides, ont été identifiées dans chaque jonction des points de cassure.

Ce modèle de réarrangement chromosomique suggère que lors de la réplication de l'ADN, quand la fourche de réplication progresse, l'extrémité 3' de l'amorce d'ADN peut changer de matrice pour un autre ADN simple brin situé sur une fourche de réplication voisine grâce aux micro-homologies qui y sont présentes, à condition que la fourche de réplication initiale soit bloquée (Figure 13) (Lee, Carvalho et al. 2007).

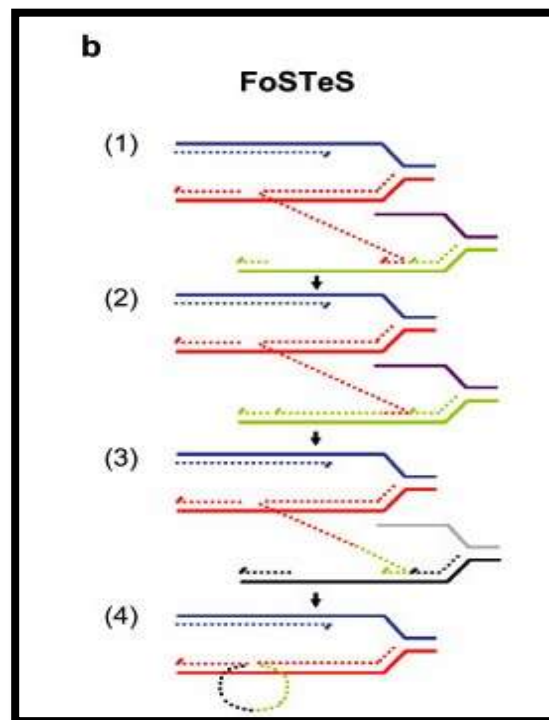


Figure 13: Mécanisme de réarrangements de type *Fork Stalling and Template Switching*. (1) Après décrochage de la fourche de réplication originale, le brin (en pointillés rouges) se désengage et s'hybride sur une deuxième fourche (les lignes en vert et violet). (2) il y a extension de l'amorce nouvellement formée et synthèse de l'ADN (pointillés verts). (Aletaha, Neogi et al.) il peut arriver que ce dernier se désengage une autre fois et s'hybride sur une troisième fourche de réplication (lignes noires et grises). (4) une nouvelle synthèse d'ADN se produit (d'après Hastings, Lupski et al. 2009).

B. Modes de transmission des CNVs

Depuis la description des CNVs au début des années 2000, l'étude de la transmission des CNVs dans le génome entier a été investiguée. Les premiers résultats ont montré que les CNVs peuvent être transmis des parents vers les descendants. Par exemple, si le premier parent a deux copies du gène sur un chromosome et aucune sur l'autre, et que le second parent possède une copie sur chaque chromosome, l'enfant peut hériter soit d'une seule copie ou de trois copies, ce qui peut avoir des conséquences potentielles dans les deux situations (White et al, 2006). Par ailleurs les CNVs peuvent survenir de façon *de novo*, mais ce cas est très peu

fréquent (1% des cas) (Redon, Ishikawa et al. 2006; McCarroll, Kuruvilla et al. 2008). Et, lorsqu'ils sont grandes tailles, ils sont souvent associés à des maladies génétiques (Strawzki 2010).

Grâce aux études de transmission des CNVs, il a été mis en évidence que certains CNVs étaient en Déséquilibres de Liaison (DL) avec certains SNPs à proximité. En effet, il a été démontré que les délétions fréquentes sont la trace de mutations ancestrales qui sont en DL avec des SNPs à proximité (Hinds, Kloek et al. 2006; McCarroll, Hadnott et al. 2006). Cependant, ces résultats ne sont pas correctement vérifiés dans le cas des régions riches en duplication (Locke, Sharp et al. 2006) car les techniques de génotypage de SNPs ne couvrent pas de manière optimale les régions ayant une importante variation structurale (McCarroll, Hadnott et al. 2006; Redon, Ishikawa et al. 2006)

C. Effets des CNVs sur l'expression

Les SNPs sont depuis longtemps considérés comme étant associés à des variations phénotypiques soit avec des effets causals directs sur les gènes, soit indirectement avec des effets sur des variants à proximité, impliqués dans des fonctions régulatrices (des SNP en déséquilibre de liaison par exemple). Plus récemment, l'effet des CNVs sur les variations phénotypiques attire de plus en plus l'attention des scientifiques. Il est devenu évident que les CNVs ont des conséquences sur la régulation de l'expression des gènes ou l'altération de la séquence codante (Stankiewicz and Lupski 2006).

C'est ce qu'a essayé de démontrer une équipe en 2009, en étudiant l'association entre 14 925 transcrits et des variations génétiques (SNPs et CNVs) présentes sur le génome entier. Ils ont trouvé que 83,6% des transcrits avec des variations d'expressions géniques, était associés à des SNPs, alors que 17,7% seulement sont associés avec des CNVs (Figure 14A et B). Uniquement 1,3% des transcrit sont associés à des SNPs et des CNVs simultanément.

Parmi les 17,7% de transcrits associés aux CNVs, plus de la moitié (53%) sont associés à des CNVs qui sont à proximités de régions codantes ou qui affectent des gènes régulateurs qui ont un impact sur l'expression génique d'autres gènes. Ces CNVs sont de diverses natures : les délétions, les duplications, les CNVs multialléliques, et les CNVs complexes (Stranger, Forrest et al. 2007).

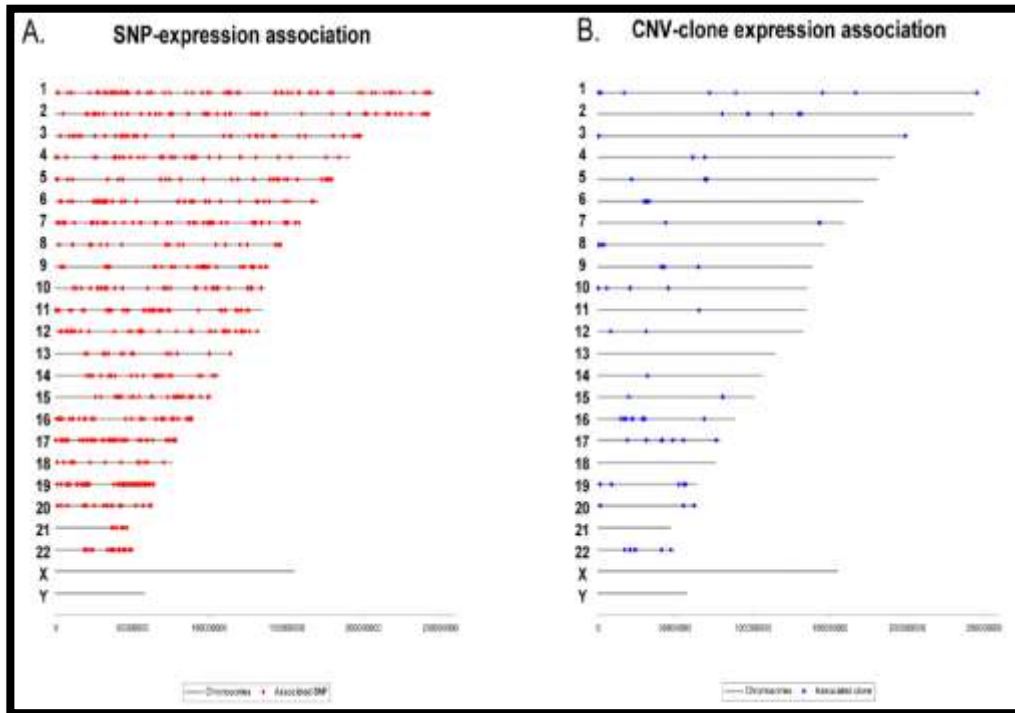


Figure 14: Localisation des variants génétiques associés à des variations de l'expression génique. A) Les SNPs associés et B) Les CNVs associés. Les points rouges et bleus correspondent à la localisation variants SNPs et CNVs (d'après Stranger, Forrest et al. 2007).

D. Méthodes d'identification

1. Analyses du génome entier

Avant la découverte des techniques à haut débit, les techniques de cytogénétique ont mis en évidence le rôle essentiel des anomalies chromosomiques. Mais l'analyse cytogénétique conventionnelle (avec un seuil de résolution de 400 bandes distinctes par génome haploïde) ne détecte pas les réarrangements de moins de 5Mb. En effet, les microremaniements (1 à 4Mb) ne peuvent être détectés par les techniques cytogénétiques moléculaires tel que la FISH (Fluorescence In Situ par hybridation) ou la peinture sur chromosome (Beri-Dexheimer, Bonnet et al. 2007). Néanmoins, elles ont permis de faire des cartographies physiques qui ont révélé la présence de séquences flanquantes et des remaniements, laissant donc envisager des évènements de recombinaison homologue qui seraient à l'origine de délétions et de duplications (Stankiewicz and Lupski 2002).

Ces données ont fait apparaître un besoin de nouvelles approches permettant de caractériser les déséquilibres chromosomiques de petite taille (1 à 4 Mb) à l'échelle pan-

génomique, c'est à dire sur un grand nombre de loci de façon simultanée (Beri-Dexheimer, Bonnet et al. 2007).

a) *L'Hybridation Génomique Comparative*

Au cours de la dernière décennie, l'Hybridation Génomique Comparative (CGH) a été largement utilisée pour cartographier les variations de nombre de copies à l'échelle du chromosome. Elle a été décrite pour la première fois en 1992 par Kallioniemi et collaborateurs, utilisant des pools d'ADN génomique marqués différemment, un fluorophore pour l'ADN « test » et un fluorophore différent pour l'ADN de référence (Kallioniemi, Kallioniemi et al. 1992). Ces deux ADN sont ensuite dénaturés et simultanément hybridés sur des chromosomes immobilisés et dénaturés, issus de la métaphase, en présence d'une concentration élevée d'ADN Cot-1 non marqué (Figure 15). Après hybridation, le ratio de fluorescence est mesuré pour chaque chromosome créant ainsi un profil génome entier des variants de nombre de copies dans le pool d'ADN (Snijders, Pinkel et al. 2003).

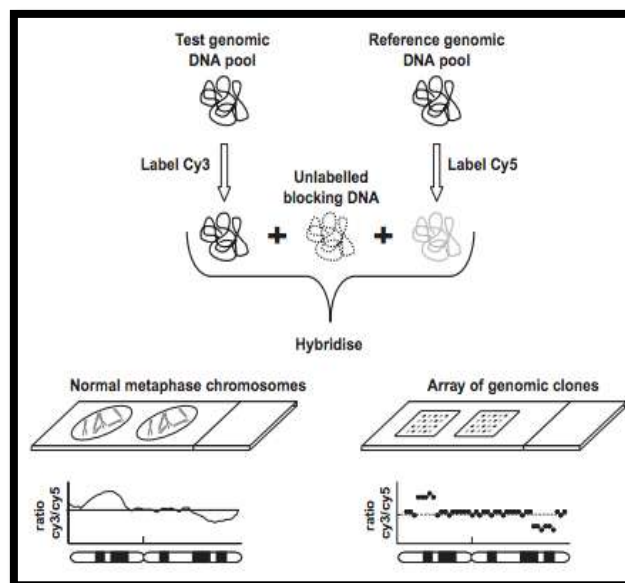


Figure 15: Etapes de l'hybridation génomique comparative (CGH) conventionnelle et sur puces. L'ADN test est marqué avec un fluorophore (exemple le Cy3) et l'ADN de référence est marqué par un autre fluorophore (exemple le Cy5). Les deux pools d'ADN sont mélangés avec un ADN non marqué qui sert à bloquer les régions répétitives du génome. Après dénaturation, le mélange d'ADN est hybridé à des chromosomes métaphasiques normaux, dans le cas de la CGH conventionnelle, ou à une puce de clones d'ADN génomique, dans le cas de l'aCGH. Un ratio de fluorescence des deux fluorophores est mesuré le long de chaque chromosome pour la CGH et sur chaque puce pour l'aCGH (d'après Snijders, Pinkel et al. 2003).

L'avantage majeur de la CGH était de pouvoir interroger le génome entier sur la présence de CNVs en une seule expérience. Néanmoins, elle présente plusieurs inconvénients,

notamment la limite de la résolution, qui est de 10Mb pour la détection de perte de copies, et qui peut être inférieure pour la détection des gains de copies (Snijders, Pinkel et al. 2003). De plus, cette technique demande une grande expertise pour l'obtention de chromosomes métaphasiques de qualité satisfaisante, ainsi que pour l'acquisition des images et leurs analyses (Redon and Carter 2009).

Les données du projet international de séquençage du génome humain, publiées depuis la moitié des années 90, issus de la construction et de la caractérisation des bibliothèques formées de larges inserts de clones comme les chromosomes artificiels bactériens (BAC) (Cheung, Nowak et al. 2001; McPherson, Marra et al. 2001) ont permis la modification de la méthode CGH en remplaçant les chromosomes métaphasiques par ces clones sur Puce ("Array") représentant des régions précises des chromosomes (Figure 15).

Cette nouvelle technique fut appelée "*matrix-CGH*" puis "*array-CGH*" (aCGH) qui est resté le nom communément utilisé jusqu'à aujourd'hui. Le développement de cette nouvelle méthode a permis d'augmenter de 10 fois la résolution par rapport à la CGH classique, ouvrant ainsi la possibilité d'identifier les micro-remaniements qui étaient, jusque-là, indétectables. De plus, pour la première fois, les points de cassure des délétions et duplications ont pu être localisés directement sur les séquences assemblées du génome humain (Redon and Carter 2009).

Après les premières études réalisées par l'équipe d'Jafrate et celle de Sebat en 2004 montrant que les CNVs étaient présents dans tout le génome humain, une première carte des CNVs a été établie en 2006, utilisant deux techniques dont la CGH sur puces (aCGH). Cette carte a montré que 12% du génome contenait des CNVs. Elle a permis de montrer également que, parmi les CNVs observés, 24% étaient associés à des duplications successives (LTRs) et que des NAHR étaient à l'origine de la plupart des CNVs (Redon, Ishikawa et al. 2006).

Une année plus tard, une autre étude utilisant des puces CGH a été réalisée en deux étapes : une première de cartographie préliminaire du génome, et une seconde, réalisée avec une forte résolution, afin d'identifier les régions avec CNVs (CNVRs) déjà décrites dans la base de donnée "*Database of Genomic Variations (DGV)*" mais aussi celles non encore identifiées. Un total de 2570 gènes avec des CNVs ont été identifiés, dont 1286 étaient déjà répertoriés dans la DGV et 1284 nouvellement décrits, tous ayant une fonction biologique ou cellulaire connue (de Smith, Tsalenko et al. 2007).

Depuis ces études, il y a eu plusieurs améliorations pour augmenter la résolution en utilisant plusieurs types de séquences pour construire les puces CGH, dont les clones de grandes ou petites tailles, des clones d'ADN complémentaire (ADNc), des produits PCR génomiques et des oligonucléotides (Carter 2007).

b) Analyses de SNPs

Grâce aux données SNP produites par le projet HapMap (Consortium 2005), il semblait possible d'identifier des CNVs à partir de ces données. En 2006, une équipe a pensé à utiliser ces données de génotypage de SNPs pour la caractérisation directe des CNVs et plus précisément des délétions (Figure 16) (McCarroll, Hadnott et al. 2006).

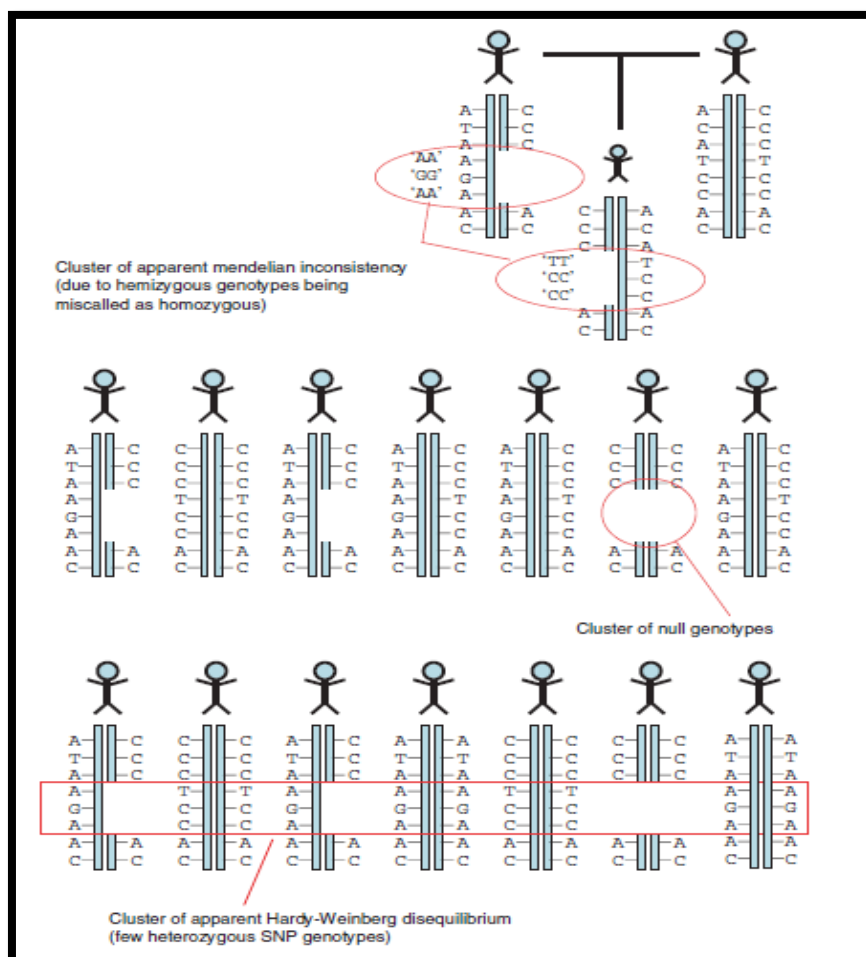


Figure 16: Utilisation des génotypes de SNPs pour l'identification de délétion par analyse de ségrégation familiale. La ségrégation de la délétion laisse une empreinte dans les données de génotype SNPs, induisant la perte physique d'un cluster de SNPs, mis en évidence par les erreurs mendéliennes, le déséquilibre de Hardy-Weinberg, ou la présence de génotypes nuls (d'après McCarroll, Hadnott et al. 2006).

Ils développent alors une approche permettant d'identifier des délétions, en supposant que la ségrégation de ces dernières laisserait une empreinte dans les génotypes des SNPs et seraient mises en évidence par l'écart à une transmission mendélienne, par le déséquilibre de Hardy-Weinberg, et en prenant en compte les génotypes nuls.

Généralement, les études génétiques excluent les SNPs présentant l'une des trois situations citées ci-dessus, les considérant comme étant due à des erreurs de génotypage ou à des artefacts. Pour considérer que cette déviation est due à une variation structurale, il faut regarder si le SNP en question se trouve dans un cluster d'autres SNPs se comportant de manière identique pour un même individu. Pour confirmer cette hypothèse, les chercheurs ont analysé 1,3 million de SNP de 269 individus issus du projet international HapMap phase I (Consortium 2005) et ils ont caractérisé une cartographie du génome avec 541 délétions potentielles de taille allant de 1 à 745kb. A partir de ces candidats, environ 278 loci (de 1 à 230kb) ont été identifiés dans plusieurs individus non affiliés suggérant que ces délétions ségrégent avec une fréquence élevée dans la population humaine, parmi lesquelles 120 sont des délétions homozygotes.

Le fait le plus intéressant, c'est que 94% de ces délétions (507/541 total) étaient nouvellement décrites. Ceci pourrait être dû au fait que ces données SNP permettaient d'avoir une résolution plus fine que les autres techniques précédemment utilisées, et que les délétions identifiées dans cette étude étaient pour la plupart de petite taille. En effet, 55% de ces variants identifiés ont une taille inférieure à 8kb et seulement 11 délétions étaient associées avec des CNVs de large taille (Tuzun, Sharp et al. 2005).

Dans l'étude de Redon et collaborateurs en 2006, l'utilisation d'une plateforme de génotypage SNP sur génome entier a permis non seulement d'identifier des délétions mais aussi des duplications, même si la résolution était moindre dans ce cas que l'aCGH (Redon, Ishikawa et al. 2006). Néanmoins, ils ont pu identifier des LCRs qui pouvaient générer des CNVs par NAHR dans des régions dupliquées du génome, et des CNVs avec des séquences fortement homologues comme les séquences *Alu*. Cette différence dans l'identification des régions répétées qui peuvent induire des CNVs est due au fait que peu de SNPs informatifs sont présents dans ces régions et donc les méthodes utilisant ces données ont plus de mal à identifier les CNVs (Figure 17).

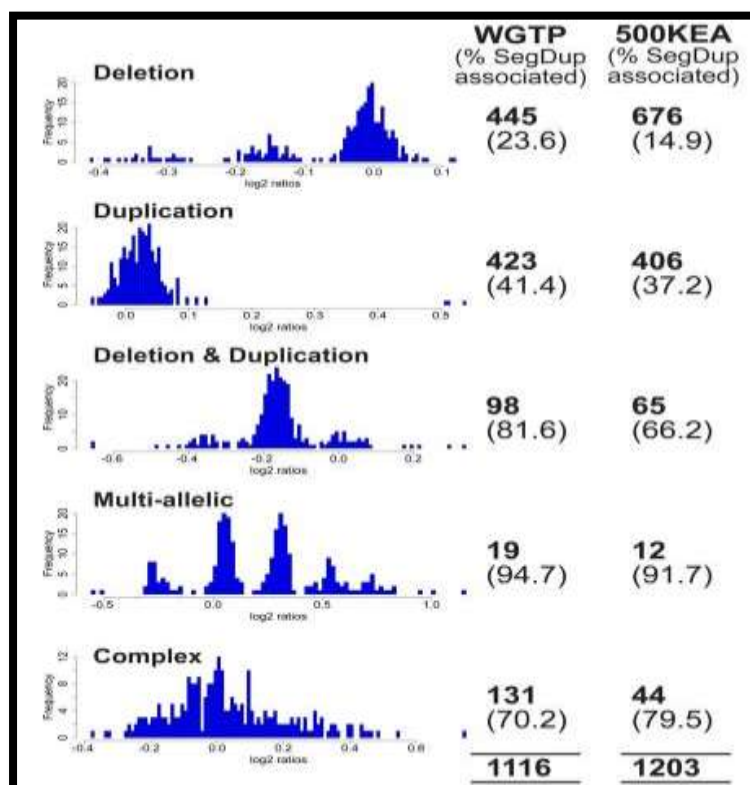


Figure 17: Cinq classes de Variations de Nombre de Copie identifiés par l'aCGH et la plateforme SNP. Le nombre de chaque classe de CNV identifié par les deux méthodes est indiqué ainsi que la proportion des CNVs qui chevauchent des duplications segmentaires (d'après Redon, Ishikawa et al. 2006).

Par contre, avec l'aCGH, les sondes utilisées peuvent être sélectionnées avec une plus grande distribution dans le génome, une plus grande densité dans les régions désirées et peuvent être dessinées spécialement pour l'identification de régions répétées où il y a un manque de SNPs informatifs. C'est pour cela que, depuis quelques années, les études de CNVs par SNPs sont complétées par des sondes aCGH, ce qui a augmenté la performance de la méthode (McCarroll, Kuruvilla et al. 2008).

a) *Comparaison entre plateformes CGH-array et SNP-array*

Une récente étude a comparé 12 plateformes différentes d'identification de CNVs se basant sur trois méthodologies : l'aCGH, la SNP-array et l'association SNP/CGH arrays. En effet, plusieurs leaders mondiaux ayant des plateformes d'identification de CNVs se partagent ce marché. Cette étude a comparé la performance de ces plateformes sur un génome de référence issu du projet HapMap (Consortium 2003) par rapport à deux listes de CNVs considérés comme références *Gold Standard* (GS) : l'une issue du "1000 Genome Consortium" (1000GC) (Abecasis, Altshuler et al. 2010; Mills, Walter et al. 2011) et l'autre "ultra high-resolution aCGH" (42MGS) (Conrad, Pinto et al. 2010). Ces données CNVs sont

considérées comme les plus précises et les plus complètes actuellement disponibles. Les plateformes comparées diffèrent par la méthode d'analyse (aCGH, SNPs ou les deux associés), par la cible étudiée (génomome entier ou des CNVs spécifiques) et par le nombre de sondes utilisées (de 292 milles à 4,2 millions) (Tableau 3).

Tableau 3: Liste des plateformes de détection de CNVs (d'après Haraksingh, Abyzov et al. 2011)

Plateformes	Cibles	Nombre de sondes	Espacement entre les sondes	Total de CNVs concordant avec 1000GC*	Total de CNVs concordant avec de 42MGS*
aCGH	-Génomome entier	4,2M	284bp	83,87	115,112
	-CNVs			408,386	310,275
	-Génomome entier	2,1M	1,1kb-1,2kb	73,78	98,95
	-CNVs			286,103	251,253
	-Génomome entier	720milles	2,5kb	24,26	29,29
	-CNVs			160,179	166,178
SNPs+aCGH	-Génomome entier	946milles (CGH)	2,2kb	46	67
		906milles (SNPs)			
	-Génomome entier	292milles (CGH)	7,2kb au total	4,5	15,16
		119milles (SNPs)			
SNPs	-Génomome entier	1,4M	1,2kb	62,68	122,122

M : Millions ; GS : Gold Standard

* Nombre de CNVs identifié par chaque méthode et qui sont concordant avec les GS

Au vu des résultats, il était clair que les plateformes les plus performantes sont celles désignées spécifiquement pour des CNVs ciblés. Pour un CNV donné, ces plateformes utilisent un nombre plus élevé de sondes (Tableau 3). De plus, même les plateformes utilisant moins de sondes mais désignées spécifiquement pour la détection de CNVs sont plus performante que celle interrogeant le génomome entier, ou celles utilisant des SNPs ou des SNPs/aCGH combinés (Tableau 3).

Malgré les avantages et les inconvénients de chaque méthode, aucune n'a pu détecter complètement les CNVs répertoriés dans les deux listes GS (~80% des CNVs pour 1000GC et ~40% des CNVs pour 42M GS). Ceci est dû au fait que dans 1000 GC il y a un large pourcentage de séquences répétées (22% Alus, 1% LINEs) qui sont difficiles à identifier par

aCGH et il y a peu de SNP informatifs. De plus, 75% des CNVs de 1000GC et 18% des CNVs du 42MGS sont de taille inférieure à 1kb alors que la majorité de ces plateformes ne peuvent détecter les CNVs inférieurs à 1kb car elles ne possèdent pas le nombre suffisant de sondes pour couvrir ces petites variations (Haraksingh, Abyzov et al. 2011).

Cette étude montre donc que la détection de CNVs par les méthodes d'identification à grande échelle a ses avantages et ses inconvénients. Mais avec l'apparition du séquençage de 2^{ème} et 3^{ème} génération sur génome entier, et les méthodes d'analyses permettant l'extraction des données CNVs, SNPs ou transposons, l'utilisation des arrays sera de plus en plus remplacée par les méthodes permettant une résolution à l'échelle de la paire de base (Abecasis, Altshuler et al. 2010). Néanmoins, les *arrays* restent une solution plus économique pour l'évaluation d'un génome entier lors des études biologiques ou de mise en place de diagnostics cliniques (Haraksingh, Abyzov et al. 2011).

b) Séquençage nouvelle génération sur génome entier

Grâce au développement du séquençage de 2^{ème} et 3^{ème} génération (*Next Generation Sequencing*, NGS) au cours de ces dernières années, l'analyse des CNVs à grande échelle sur le génome entier a vu émerger plusieurs approches très sensibles dans la détection de ces variations génomiques.

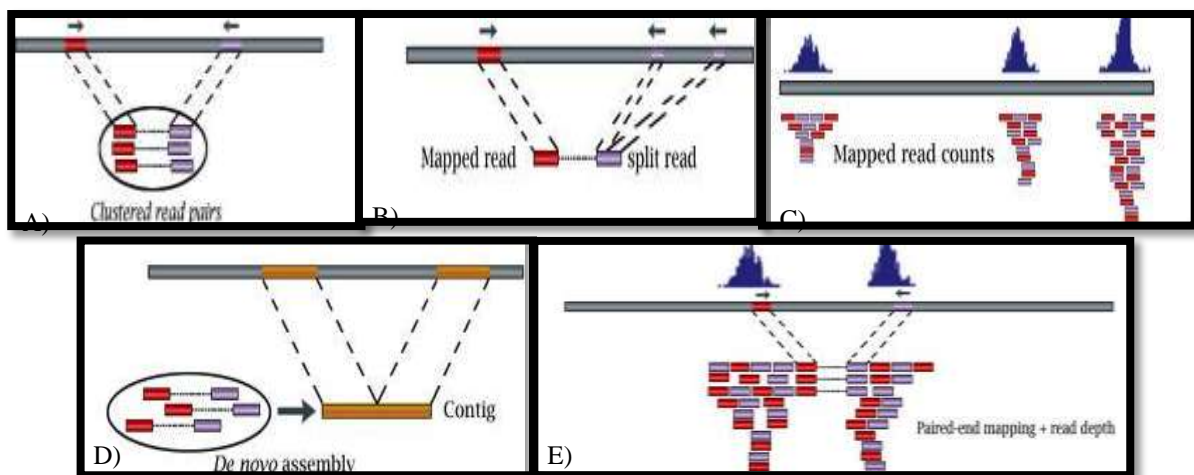


Figure 18: Cinq approches pour détecter des CNVs à partir de données de séquences NGS. A) *Paired-end mapping* (PEM) détecte les CNVs par alignement discordants des séquences. B) L'approche *Split read* utilise l'alignement incomplet de séquences pour identifier les petits CNVs. C) L'approche basée sur la *Red depth* détecte les CNVs en comptant le nombre de séquences chevauchantes à chaque région génomique. Dans la figure, les *reads* sont alignés à trois régions de l'Exome. D) La méthode d'assemblage *de novo* (AS) détecte les CNVs en alignant des contigs sur un génome de référence. E) l'approche combinée utilise simultanément les approches RD et PEM pour la détection de CNVs (d'après Zhao, Wang et al. 2013).

Ces approches peuvent être classées en 5 méthodes différentes: le séquençage bidirectionnel des extrémités de fragments issus du génome (*Paired-End Mapping*, PEM) - la lecture discontinue (*Split Read*, SR) - la profondeur de lecture (*Read Depth*, RD) - l'assemblage *de novo* du génome (AS) - la combinaison de ces méthodes citées précédemment (Figure 18) (Zhao, Wang et al. 2013).

(1) Paired-End Mapping

Le principe de la PEM se base sur la préparation de séquences d'ADN d'environ 3kb obtenues par fragmentation mécanique. Puis, après protection des sites de restriction EcoRI, les extrémités de ces fragments sont liguées avec des adaptateurs biotinylés qui permettent de circulariser les séquences d'ADN grâce au site EcoRI présents sur les deux adaptateurs. Ensuite, l'ADN circularisé est fragmenté aléatoirement par ultrason et les ADN ayant les *paired-ends* seront isolés par affinité via des adaptateurs biotinylés à la streptavidine. Il y a par la suite construction de librairie génomique et séquençage (Figure 18A et 19). Ces séquences sont enfin alignées puis comparées à un génome de référence.

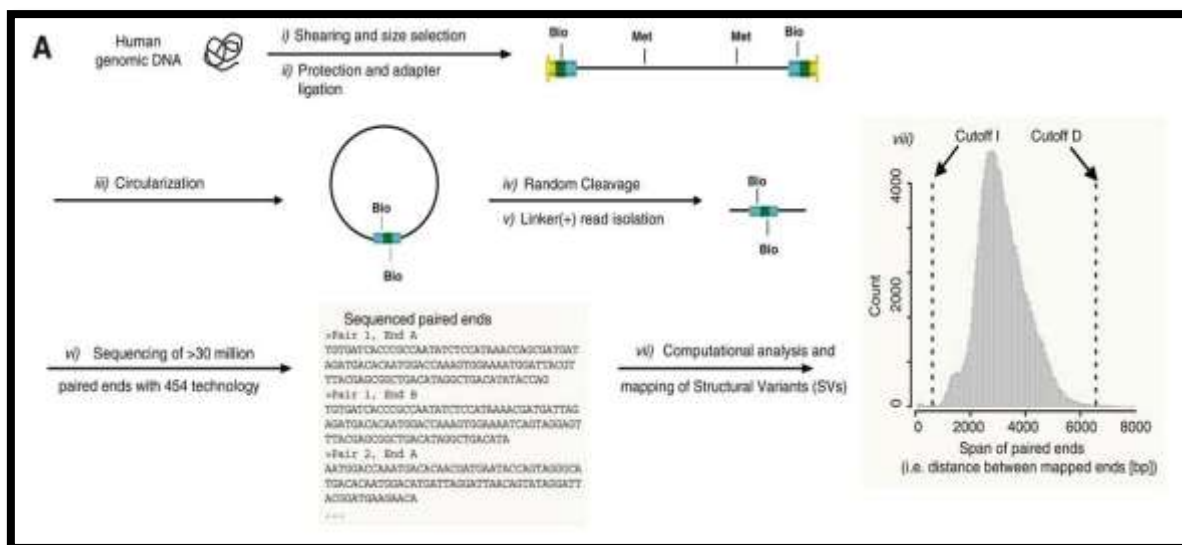


Figure 19: Etapes de la méthode *Paired-end Mapping*. i) L'ADN est clivé en fragments de 3kb ; ii) ligation des extrémités des fragments générés à des adaptateurs en forme d'épingle à cheveux biotinylés ; iii) Les fragments sont circularisés ; iv) puis clivés encore une fois aléatoirement ; v) les fragments positifs sont isolés ; vi) la librairie construite est séquençée. vii) les séquences *Paired-ends* sont analysées (viii) pour déterminer la distribution des séquences (d'après Korbel, Urban et al. 2007).

Les avantages de cette méthode se résument au fait que la PEM augmente la résolution des variations structurales (SVs) détectées jusqu'à un niveau qui permet la validation par PCR,

et qu'elle ne nécessite pas de préparation de librairie d'ADN par clonage, ce qui permet un gain de temps. Par contre, comme toutes les autres techniques d'analyses à grand-échelle, elle possède une limite de détection dans les régions avec des copies multiples ayant une forte similarité et dans l'identification des longues répétitions (>3kb). De plus, le choix d'utiliser des fragments de 3kb entrave la détection des insertions >3kb (Korbel, Urban et al. 2007).

(2) L'approche Split read

Cette méthode utilise des paires de séquences différentes issues des lectures de séquences NGS (*reads*). La première séquence de chaque paire est alignée à un génome de référence alors que la deuxième ne chevauche pas du tout ou que partiellement le génome de référence (Figure 18B). Ces *reads* non chevauchants ou partiellement chevauchants fournissent une identification précise, à une paire de base près, des points de cassures des SVs et des CNVs. Ensuite ces séquences non-complètement alignées sont divisées en plusieurs fragments. Le premier et le dernier fragment de chaque *read* clivé est par la suite aligné avec un génome de référence ce qui permet d'avoir la position exacte de début et de la fin de chaque événement d'insertion ou/et de délétion. Néanmoins, cette approche dépend fortement de la taille des *reads* et n'est applicable qu'à des régions du génome de référence (Zhao, Wang et al. 2013).

(3) L'approche Read Depth

Cette approche est devenue, depuis peu, la plus utilisée pour l'estimation des CNVs à partir de données NGS. Le principe de cette méthode se base sur le fait que la profondeur de couverture d'une région du génome est corrélée avec le nombre de copies de la région (Payami, Nutt et al. 2003) (Figure 18C). Par exemple, un gain de copies devrait avoir une intensité plus élevée que celle attendue (Teo, Pawitan et al. 2012). Contrairement aux deux précédentes approches, qui utilisent les informations de la position des CNVs, la RD peut quantifier de façon exacte le nombre de copies. De plus, cette dernière méthode peut détecter les larges insertions et les CNVs dans les régions complexes du génome, qui sont difficilement détectés par la PEM et la SR (Yoon, Xuan et al. 2009).

(4) Approche de l'assemblage de novo

Contrairement aux autres approches précédemment citées, qui alignent d'abord les *reads* de NGS avec un génome de référence, cette méthode reconstruit tout d'abord le

fragment d'ADN grâce à l'assemblage de contigs issus de *reads* de petite taille qui se chevauchent. Ces contigs assemblés sont ensuite comparés au génome de référence, ce qui permet d'identifier les régions du génome avec des CNVs (Figure 18D).

(5) L'approche combinée

Bien qu'il y ait eu beaucoup de progrès dans chacune des quatre approches citées précédemment (PEM, SR, RD, et AS) et que le nombre d'outils d'analyses n'ait cessé d'augmenter, aucune d'elle n'a été capable de détecter un spectre complet de CNVs avec de fortes sensibilités et spécificité. Avec les avantages et les faiblesses de chacune de ces méthodes, elles pourraient se compléter les une les autres. La méthode PEM identifie plusieurs types de SVs (délétion, insertion, inversion...) (Korbel, Urban et al. 2007), mais elle ne peut pas estimer la valeur de nombre de copies (CN) avec précision et identifier les larges insertions qui sont plus grandes que les bibliothèques construites (3kb). Par contre, la méthode RD peut prédire le CN d'une région avec précision, et a une bonne performance pour la détection des larges CNVs. Elle est également applicable pour le séquençage sur génome entier ou sur l'exome entier. Néanmoins, elle ne permet pas la détection précise des points de cassures, ou des copies neutres comme les inversions, la translocation ou les CNVs inférieurs à 1kb. Dans l'approche combinée, l'association entre la PEM et la RD peut être envisagée (Figure 18E). Pour pallier aux limitations de chacune des méthodes, des outils de bioinformatique ont été mis en place pour combiner les deux approches afin d'augmenter la performance de détection des CNVs et réduire le nombre de faux-positifs (Zhao, Wang et al. 2013).

c) Le séquençage entier de l'exome

Au cours ces dernières années, les scientifiques se sont intéressés de plus en plus au séquençage des régions codantes (*Whole Exome Sequencing*, WES), car en plus de l'intérêt biologique de cette méthode, elle reste tout de même moins chère que le séquençage du génome entier (*Whole Genome Sequencing*, WGS). La WES est largement utilisée, que ce soit pour les maladies mendéliennes ou pour les maladies complexes, et a le potentiel d'identifier rapidement les mutations d'un seul nucléotide ou les CNVs dans les régions codantes.

Puisque les régions ciblées sont de petites tailles (100 à 300pb), l'identification et le génotypage des CNVs à partir des données produites par WES est plus difficile qu'à partir de données WGS. Par ailleurs, vu que les exons ne sont pas contigus, les méthodes comme la PEM, la SR ou l'AS ne peuvent pas s'appliquer (Tan, Wang et al. 2014). La seule stratégie

d'identification des CNVs adaptée avec succès du WGS vers le WES est la RD. Suivant le même principe que pour le WGS, la méthode RD va d'abord calculer le nombre de *reads* alignés à chaque exon cible, puis la moyenne de chaque valeur de RD pour chaque exon va être considérée par la suite comme un signal brut de RD qui sera utilisé pour l'analyse.

La limitation inhérente du WES est qu'elle produit des signaux seulement pour 1 à 2% du génome humain, due aux actuels kits d'extraction de l'exome. Ceci ne permet donc pas de détecter le spectre complet de CNVs et les points de cassures. Néanmoins, le WES permet d'avoir un rapide aperçu des signatures CNVs pour certaines maladies ou phénotypes (Zhao, Wang et al. 2013).

2. Analyses ciblées sur des gènes candidats

Depuis les premières études qui ont démontré la présence inattendue d'un nombre élevé de CNVs dans le génome humain, il y a de cela dix ans (Iafrate, Feuk et al. 2004; Sebat, Lakshmi et al. 2004), des centaines de nouveaux CNVs ont été identifiés dans les laboratoires de diagnostic et de recherche partout dans le monde. La validation et la réplication de ces CNVs sont devenues une priorité. Il fallait donc voir si l'on réussissait à observer le même CNV pour plusieurs individus ou à confirmer un CNV d'un même individu avec plusieurs plateformes (Ceulemans, van der Ven et al. 2012). Par ailleurs, la détection des CNVs par puces explorant le génome entier était moins sensible pour l'analyse individuelle de gènes ou de séquences d'intérêt. Pour pallier à ce problème, il fallait mettre en place des techniques qui pouvaient explorer l'impact des CNVs dans les maladies avec une haute-résolution, à haut-débit et surtout à faibles coûts.

a) La PCR classique

La Réaction de Polymérisation en Chaîne (PCR) classique avec révélation des amplicons sur gel était la plus répandue pour la détection rapide d'un CNV dans un gène ou dans une région d'intérêt. Elle peut être utilisée de diverses façons en fonction de l'information qu'on souhaite avoir sur le CNV : détection de la présence ou de l'absence du CNV, la quantification relative du CNVs, ou l'identification du génotype. Pour les deux premiers cas, c'est généralement facile à réaliser. Il suffit de désigner un couple d'amorces à l'intérieur du gène et un autre couple pour amplifier une séquence d'un gène qui n'a pas de CNVs qui servira comme contrôle interne. L'amplification de ces deux séquences permettra de déterminer si le gène est présent (profil d'amplification avec deux bandes, une pour le gène

cible et une pour le contrôle interne) ou si il y a une délétion complète du gène (profil PCR avec seulement la bande de la séquence contrôle). L'inconvénient de cette méthode c'est qu'elle ne permet pas de savoir si le gène est présent à l'état homozygote ou hétérozygote. Grâce au développement des logiciels d'analyse des profils PCR, il a été possible de compléter l'interprétation des résultats pour différencier les génotypes homozygotes et hétérozygotes. En effet, en 1997, une équipe a réussi à mettre en place un protocole qui permettait la quantification relative d'un gène en comparant l'intensité de la bande du gène d'intérêt avec celle du gène contrôle, un ratio est alors calculé et le nombre de copie peut être estimé (McLellan, Oscarson et al. 1997). C'est sur ce principe que se basera la PCR quantitative en temps réel par la suite. Mais cette stratégie de quantification des CNVs ne permet pas d'identifier les réarrangements neutres (duplication sur un chromosome et la délétion sur l'autre) et reste peu fiable.

La troisième stratégie consiste à désigner deux paires d'amorces, une première qui amplifie une séquence du gène d'intérêt, et la deuxième s'hybride de part et d'autre du gène. Si le gène est présent sur les deux chromosomes (génotype homozygote), il y a amplification avec le premier couple d'amorces et pour le deuxième la distance sera trop grande pour avoir une amplification. Si le gène est présent à l'état hétérozygote (présence du gène sur un chromosome et sa délétion sur l'autre), il y aura une amplification du premier couple d'amorces mais aussi du deuxième puisqu'avec la délétion, la distance entre les amorces sera réduite. Ces dernières pourront amplifier la région cible. La séquence amplifiée par la deuxième paire d'amorces contiendra le point de cassure. Enfin, s'il y a une complète délétion du gène, le premier couple d'amorces ne donnera pas d'amplification, alors que le deuxième donnera une amplification de la région délétée (Buchard, Sanchez et al. 2007) (Figure 20).

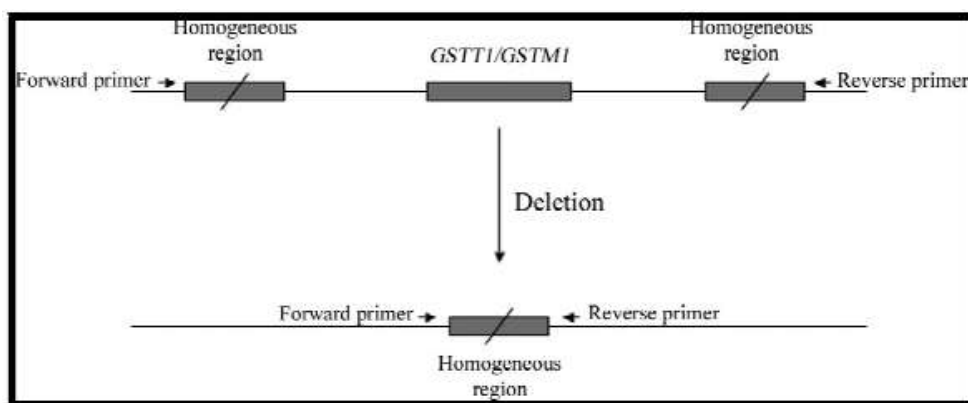


Figure 20: Stratégie d'amplification de la région de délétion du gène *GSTM1* (d'après Buchard, Sanchez et al. 2007).

Malgré le concept astucieux de cette PCR, elle semble, néanmoins, ne pas être applicable à tous les CNVs des autres gènes puisqu'elle nécessite une bonne cartographie de la région d'intérêt et surtout connaître les points de cassures entourant le gène cible.

b) La PCR quantitative

La PCR quantitative (qPCR) en temps réel est l'une des méthodes les plus utilisées pour la validation des résultats CNVs obtenus par les plateformes de détections de CNVs à grande-échelle (aCGH ou SNP-array). Son principe se base sur la quantification de l'intensité de fluorescence libérée lors de la réaction de PCR soit par des sondes hydrolysables (TaqMan®) (Wilke, Duman et al. 2000), des sondes Scorpions (Solinas, Brown et al. 2001), et des sondes FRET (Shengqi, Xiaohong et al. 2002) ou par des molécules fluorescentes qui s'intercalent entre les brins de l'ADN (SYBR green) (Simpson, Feeney et al. 2000).

Pour la quantification des CNVs, c'est la technologie TaqMan® qui est la plus répandue car elle permet une quantification plus précise et plus spécifique que le SYBR green, et elle est moins coûteuse que les autres types de sondes. Elle contient un fluorophore rapporteur à l'extrémité 5' et un marqueur d'atténuation du signal à l'extrémité 3' (*quencher*) (Figure 21).

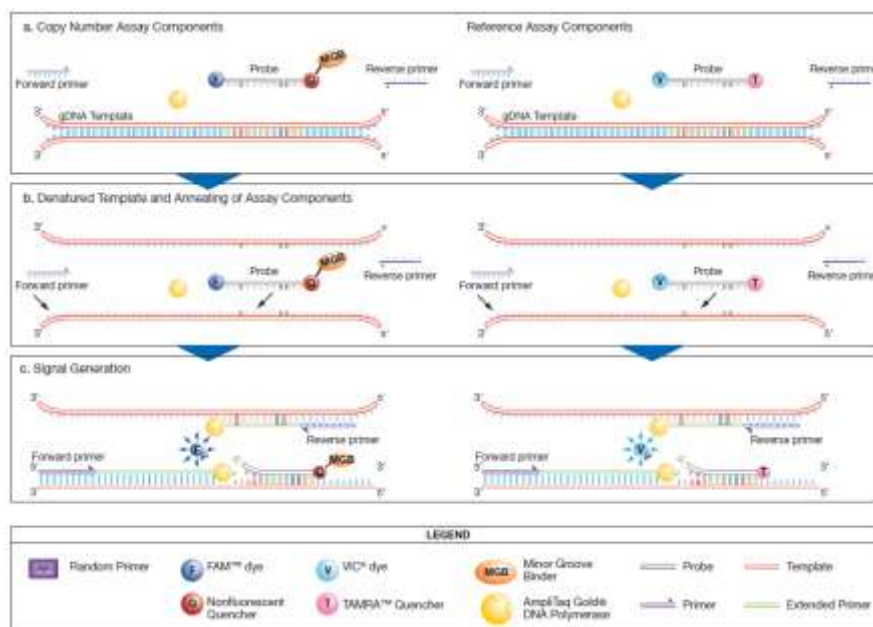


Figure 21: Principe de la technologie TaqMan® pour la quantification des CNVs. D'après Copy Caller AppliedBiosystems ».

Lorsque la sonde est intacte, il n'y a pas de fluorescence car elle est atténuée par le quencher, mais une fois la sonde hybridée et l'étape d'élongation déclenchée, la polymérase va hydrolyser la sonde du côté 5' grâce à son activité 5' exonucléasique. Le fluorophore va être ainsi libéré et émet une fluorescence qui va être détectée à chaque cycle. Pour quantifier les CNVs d'un gène cible, il faut co-amplifier avec lui un gène de référence (qui ne contient pas de CNV), permettant par la suite le calcul du nombre de copies.

Au cours de la réaction, plus il y a accumulation d'amplicons, plus l'intensité de fluorescence augmente par hydrolyse des sondes ou par intercalation d'agents fluorescents. La quantification se fait alors par détermination du nombre de cycle (*threshold cycle*, Ct) à partir duquel l'intensité devient supérieure au seuil fixé et qui dépend directement de la quantité de départ du gène cible. Deux CNs, dont un est élevé et l'autre faible, vont avoir respectivement une augmentation significative de fluorescence qui sera précoce avec un Ct faible et tardive avec un Ct élevé (Lee and Jeon 2008) (Figure 22).

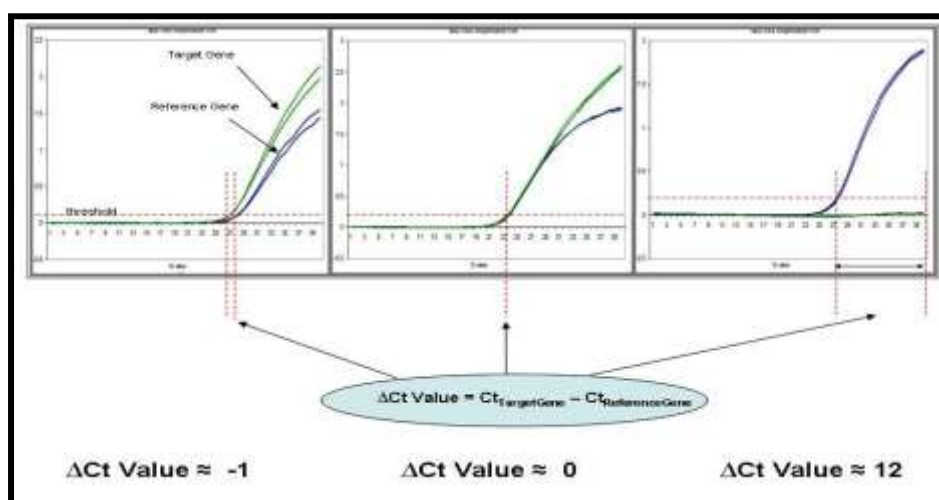


Figure 22: Méthode de calcul de nombre de copies en fonction des valeurs Ct (d'après Rose-Zerilli, Barton et al. 2009).

Il existe deux méthodes de calcul de CNVs par qPCR, soit par quantification relative, en calculant la différence entre le Ct du gène cible et celui d'un gène de référence, (méthode du ΔCt), soit par quantification avec une courbe standard, avec des individus contrôles de CN connu, qui permet de quantifier les CNs inconnus.

La qPCR offre une meilleure précision que la PCR classique, mais également plus de facilité pour la quantification en routine puisqu'il n'y a pas d'étapes de migration. Elle est devenue parmi les méthodes les plus utilisées pour la quantification de CNVs. Cependant, il y a tout de même des limitations importantes dans l'application de la qPCR pour la quantification de CNVs. La plus récurrente est la variabilité des CN identifiés en fonction du nombre de réplicats. Il existe tout de même plusieurs solutions récentes pour corriger cette limite (Weaver, Dube et al. 2010).

Enfin, cette méthode est idéale pour l'analyse des CNVs ayant un nombre de copies allant de 0 à 4 copies dans une petite population. Mais cette méthode devient très couteuse quand il s'agit d'étudier plusieurs CNVs sur de grandes cohortes, notamment lors des études d'association.

c) Test de Ratio entre paralogues

Pour pallier à certains défauts de la qPCR, une nouvelle variante de PCR multiplex a été décrite en 2007, reprenant certains principes de la qPCR comme l'utilisation de la fluorescence et d'un gène de référence pour la quantification des CNV. Mais, contrairement à la qPCR qui se base sur l'hybridation et l'hydrolyse des sondes fluorescentes différentes pour le gène cible et la référence, cette nouvelle méthode utilise une paire unique d'amorces pour le gène cible et le gène de référence, contenant un fluorophore. En effet, dans la qPCR, le gène et la référence sont indépendants et n'ont pas, généralement, les mêmes propriétés d'amplification ce qui pourrait induire un problème de reproductibilité entre les résultats. Il a donc été proposé que le gène de référence devrait avoir une similarité avec le gène cible. C'est sur ce principe que l'utilisation du paralogue du gène cible (qui est issu d'un évènement de duplication au cours de l'évolution) a été proposée.

Les amorces sont donc choisies de façon concise pour pouvoir amplifier les deux paralogues (gène cible et gène référence) mais également pour pouvoir les différencier avec une petite variation de taille entre les deux amplicons (Armour, Palla et al. 2007) (Figure 23). Ces derniers sont enfin séparés par plusieurs méthodes comme la restriction enzymatique, le pyroséquençage ou encore l'électrophorèse capillaire (Ceulemans, van der Ven et al. 2012). L'analyse se fait par calcul du ratio d'intensité de fluorescence entre les deux paralogues, d'où son nom *Paralogue Ratio test* (PRT) (Figure 23).

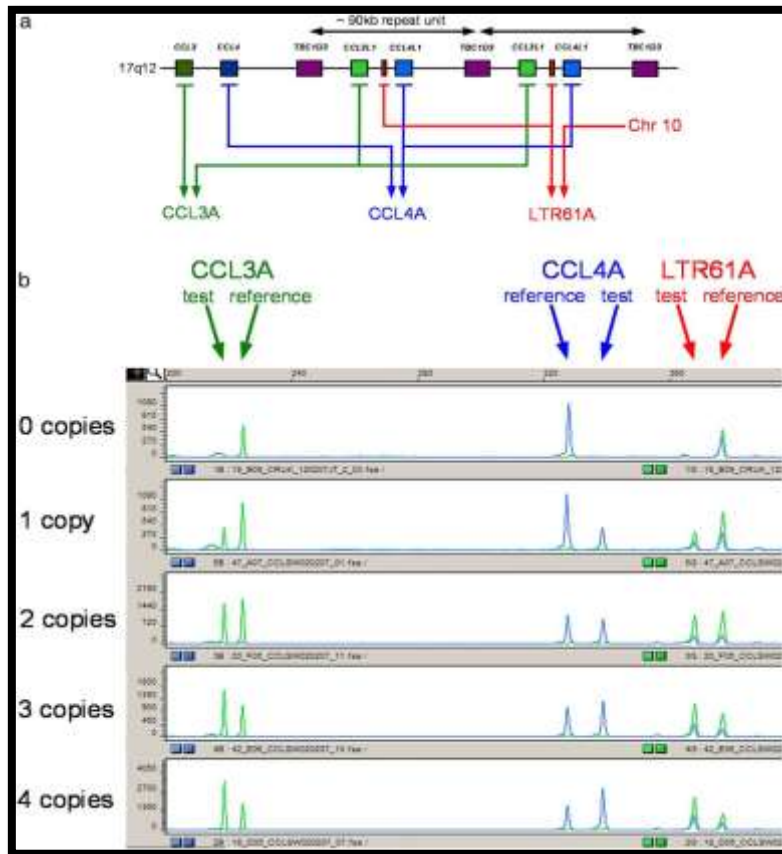


Figure 23: Diagramme schématique de trois systèmes de PRT au niveau du locus du gène *CCL3L1*. a) Le ratio "CCL3A" (en vert) compare la quantité de produits PCR du gène *CCL3L1* par rapport à ceux de *CCL3* qui est considéré comme gène de référence avec 1 copie par génome haploïde. De même, le système "CCL4A" (en bleu) compare les produits de *CCL4L1* et *CCL4*. Enfin le système "LTR61A" (en rouge) compare les produits d'une séquence LTR dans la région du CNV par rapport un gène de référence qui se trouve au niveau du chromosome 10. b) Profil électrophorétique montrant les résultats de PRT issus d'individus avec 1, 2, 3 et 4 copies de *CCL3L1* et *CCL4L* (d'après Armour, Palla et al. 2007).

Cette méthode permet de quantifier les CNVs rapidement, à coûts réduits et avec une faible quantité d'ADN génomique (10-20ng). De plus, contrairement à la qPCR, la PRT ne nécessite pas l'utilisation d'un grand nombre de réplicats, une seule réaction suffit. Mais l'inconvénient est que le taux d'erreur peut atteindre 20% pour les échantillons avec un nombre de copies élevé (>4). L'autre inconvénient de la PRT est qu'elle dépend de la présence de séquences paralogues au sein du génome. (Armour, Palla et al. 2007).

d) **Multiplex-Ligation-dependent Probe Amplification**

La *Multiplex-Ligation-dependent Probe Amplification* (MLPA) est de loin la méthode de PCR multiplex la plus utilisée pour l'analyse de CNVs. Son principe diffère des autres variantes de PCR quantitatives puisqu'il y a deux séquences (séquences X et Y) qui sont utilisées et qui sont spécifiques à un gène cible. La séquence X contient une sonde

fluorescente qui a une taille définie. Enfin, à l'extrémité de chaque séquence X et Y se trouve une paire d'amorces universelles. Ces séquences X et Y vont s'hybrider spécifiquement à un gène et, après ligation de ces deux séquences, une amplification par PCR est réalisée grâce aux amorces universelles (Figure 23).

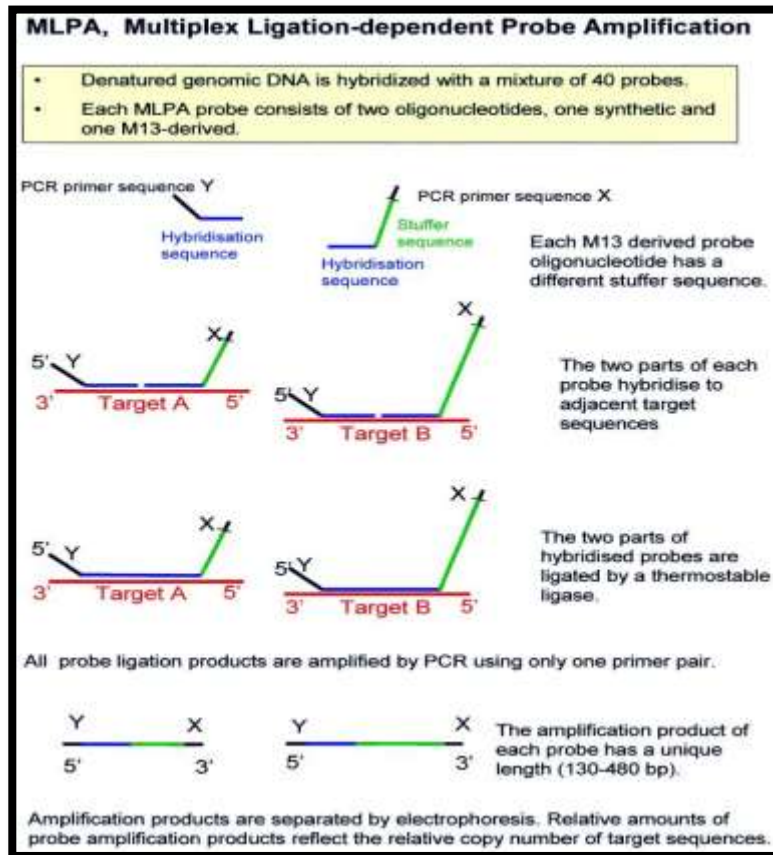


Figure 24: Schéma des différentes étapes de la MLPA (d'après Schouten, McElgunn et al. 2002).

Ainsi, plusieurs CNVs de gènes différents peuvent être ciblés dans une seule réaction en dessinant simplement les séquences d'hybridation spécifiques à chaque gène et avec une longueur de sondes différente. Il est donc possible d'identifier chaque CNVs après PCR puisque chaque amplicon a une taille différente de l'autre (de 130 à 480pb). Ces derniers sont enfin séparés par électrophorèse capillaire et un ratio de fluorescence est calculé entre celle du gène cible et celle du gène de référence. Cette méthode permet d'avoir une quantification avec une sensibilité élevée et de différencier deux homologues avec une paire de base de différence. Mais cette spécificité élevée peut rapidement devenir un inconvénient. En effet, la présence d'un SNP dans la séquence du gène où s'hybrident les sondes de MLPA (qui font environ 20-30 nucléotides) peut affaiblir l'efficacité de l'hybridation ou de la ligation ce qui

peut engendrer des résultats faux-positifs de délétions. C'est pourquoi il faut confirmer les délétions avec d'autres méthodologies (Schouten, McElgunn et al. 2002).

e) *La PCR digitale*

Récemment, une nouvelle méthode appelée *Single molecule PCR* (smPCR) a été mise en place pour l'analyse des CNVs en partant d'une seule molécule d'ADN de départ obtenue par dilution en série. Dès lors, chaque aliquot sera soit positif ou négatif pour la séquence cible, c'est de là que les termes "*Digital PCR*" ou "*Molecular Copy-number Counting* (MCC)" ont été attribués à cette nouvelle méthode (McCaughan and Dear 2010). Cette approche se base sur une dilution de l'ADN dans des aliquots contenant moins d'une molécule de génome haploïde par microlitre. Une première PCR multiplex avec un nombre de cycle limité est réalisée pour chaque aliquot afin d'amplifier la séquence cible. Les produits de cette PCR sont ensuite dispatchés dans une autre plaque pour faire une deuxième PCR nichée avec des amorces spécifiques pour le gène cible et le gène de référence. Les produits PCR sont par la suite séparés par électrophorèse en gel puis un comptage simultané des aliquots positifs ou négatifs pour le gène cible et la référence se fait pour estimer le MCC et donc calculer le nombre de copies du gène cible par rapport au gène de référence qui ne varie pas (Daser, Thangavelu et al. 2006; McCaughan 2009). Cette méthode, rapide et flexible, permet la caractérisation d'une large gamme d'aberrations chromosomiques avec la capacité d'analyser plusieurs cibles simultanément. Elle est également très sensible, même en présence d'ADN de mauvaise qualité ou en faible quantité (Ceulemans, van der Ven et al. 2012).

Cette approche a été améliorée en développant un système de nanofluide sur micro-Puces, la *Digital-array*, permettant l'analyse de CNVs en se basant sur le principe de MCC. Cette technique est capable de quantifier avec précision des échantillons d'ADN en se basant sur le fait que les molécules d'ADN individuelles sont réparties aléatoirement dans plus de 9000 nano-puits de réactions où aura lieu la PCR avec deux sondes fluorescentes spécifiques pour le gène cible et le gène de référence (Figure 25a).

La concentration en copies/ μ l est calculée pour chaque séquence en comptant le nombre d'événements positifs et négatifs pour le gène cible et le gène de référence (Figure 25b-d). Le nombre de copies relatif par génome haploïde est estimé par le ratio entre la concentration du gène cible et du gène de référence qui est dans ce cas une copie par génome haploïde. Le nombre de copies relatif doit être égal à 1 lorsque le gène possède 2 copies. S'il

est supérieur à 1, le gène est dupliqué et s'il est inférieur à 1, le gène est délété. Cette technique permet d'identifier des CNVs supérieurs à 7 copies (Qin, Jones et al. 2008).

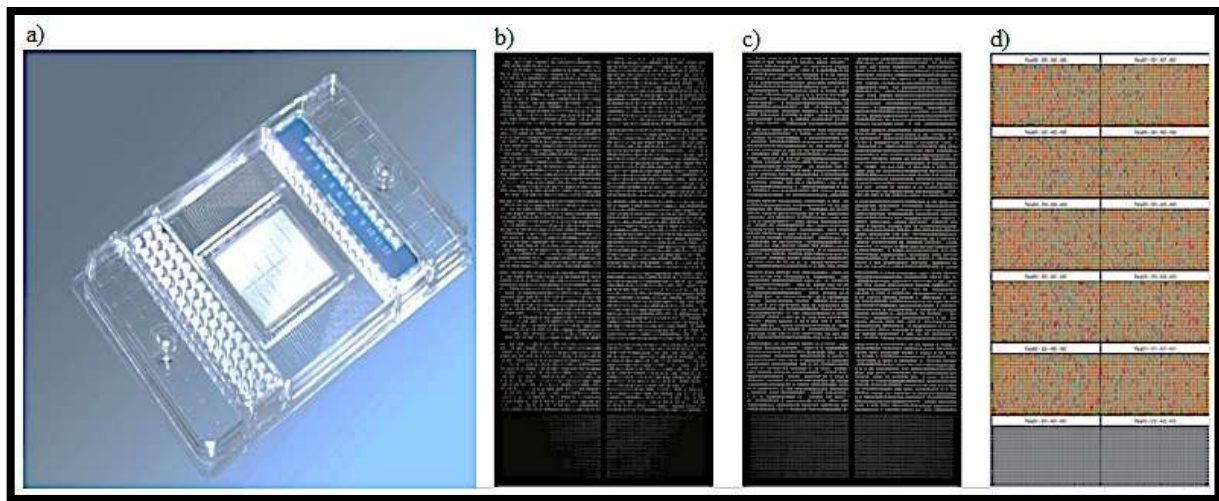


Figure 25: Principe de la Microfluidique Digital PCR. a) Cassette utilisée pour la digitale PCR. b) exemple de puce avec les évènements positifs pour le gène cible et pour le gène de référence (c). d) résultats combinés par le logiciel pour le gène cible (en jaune) et le gène référence (en rouge) (d'après Qin, Jones et al. 2008).

Une autre variante de digital PCR a vu le jour récemment remplaçant la distribution des molécules d'ADN dans des nano-puits par le partitionnement d'un volume final de réaction normale de qPCR en 20 000 gouttelettes (*Droplets*) grâce au principe d'émulsion d'où son appellation *Droplet Digital PCR* (ddPCR). Une huile pour l'émulsion et le volume réactionnel sont déposés dans deux puits différents d'une cassette (Figure 26a). Ensuite, les gouttelettes sont générées grâce au générateur de droplets, qui va appliquer une pression pour aspirer l'huile qui va stabiliser les droplets et le mélange réactionnel à travers des nano-circuits et mélanger ces deux solutions pour former des microgouttes (Figure 26b) qui seront collectées dans un troisième puits (Figure 26a). Elles sont ensuite transférées vers une plaque de PCR classique et amplifiées par un thermocycleur conventionnel. Enfin, un lecteur de droplets va mesurer le signal fluorescent dans chaque gouttelette et compter le nombre d'évènements positifs et négatifs. Un logiciel dédié va estimer la concentration en copie/ μ l pour le gène cible et le gène de référence et calculer le nombre de copies équivalent (Figure 26c-f) (Hindson, Ness et al. 2011).

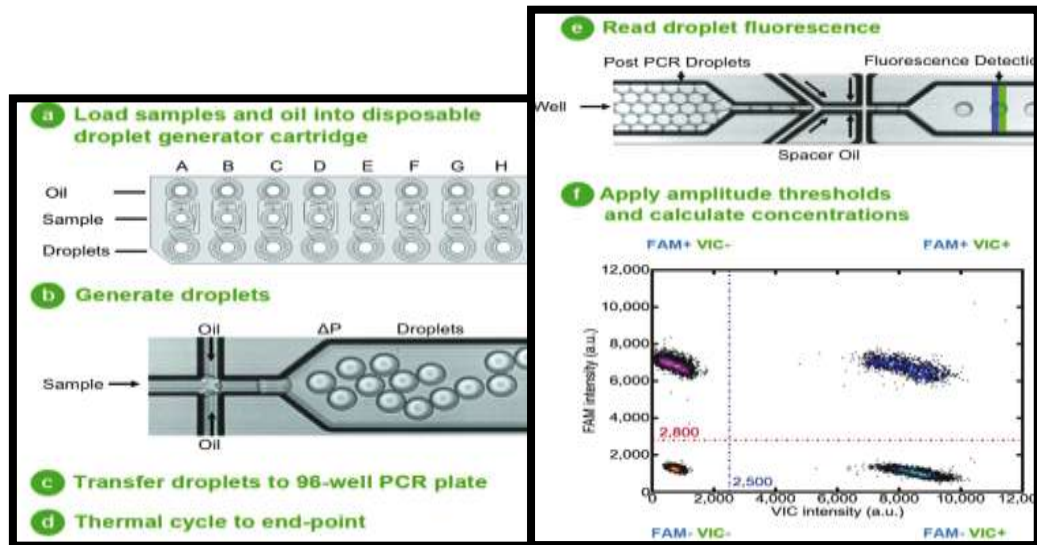


Figure 26: Principe de la Droplet Digital PCR. a) Les échantillons sont chargés dans une cassette avec huit puits b) Une aspiration appliquée dans les nano-canaux va permettre la formation d'environ 20000 droplets faisant un nano-litre en moins de 2min. c) La solution contenant de *droplets* est transférée vers une plaque de 96 puits. d) une amplification est réalisée dans un thermocycleur conventionnel. e) la plaque est chargée dans un lecteur de fluorescence qui va lire *droplet* par *droplet* pour détecter la fluorescence correspondante au gène cible et au gène de référence avec une fréquence d'environ 1000 droplets par seconde. Le nombre d'événements positifs et négatifs est alors calculé ce qui permet l'estimation de la concentration en copies/ μ l du gène et de la référence, et leur intervalle de confiance (95%) estimé par la loi de Poisson (d'après Hindson, Ness et al. 2011).

L'avantage de cette technologie par rapport à sa concurrente précédemment citée est qu'elle s'affranchit de l'étape de dilution limite, qui peut s'avérer une source d'erreurs si elle n'est pas faite correctement. En effet, à concentration élevée d'ADN ou lorsque ce dernier est de mauvaise qualité, la dilution limite n'est plus optimale et les puces vont être chargées d'évènements positifs et le calcul de CN ne peut plus se faire. De plus, la ddPCR utilise un principe plus simple pour la détection de duplication en tandem qui se base sur le choix d'une enzyme de restriction qui coupe de part et d'autre de la région amplifiée.

En effet, si les deux copies se retrouvent dans la même *droplet*, lors du comptage, elles sont considérées comme un seul évènement alors que lorsqu'elles sont séparées par une enzyme de restriction, les deux copies vont être comptées comme deux évènements distincts. Le nombre de copies dans le premier cas est donc sous-estimé et dans le deuxième cas le CN a la bonne valeur. Ces duplications peuvent être détectées également par la Digital-array mais nécessitent une première PCR pour amplifier les deux copies dupliquées en tandem, puis une PCR nichée se fait dans les nano-puits avec les sondes fluorescentes (Qin, Jones et al. 2008). La ddPCR est donc plus économique et plus simple puisqu'elle n'utilise pas de puces, consomme moins de réactifs, et comporte moins d'étapes (Hindson, Ness et al. 2011).

E. CNVs et maladies

1. Les CNVs dans les maladies rares

Les maladies rares sont dues à des troubles génomiques et sont pour la plupart associées à de grandes aberrations chromosomiques, mais il existe aussi plusieurs délétions ou duplications sub-microscopiques décrites depuis les années 80 (Schmickel 1986). La microdélétion du chromosome 15q11.2q12 associée au syndrome de Prader-Willi a été identifiée en 1981 (Ledbetter, Riccardi et al. 1981). Plusieurs délétions variables ont été identifiées dans la majorité des patients atteints d' α -thalassémie (Flint, Hill et al. 1986), et de dystrophie musculaire de Duchenne (Francke, Ochs et al. 1985).

La maladie de Charcot-Marie-Tooth de type 1A (CMT1A) et la neuropathie héréditaire avec hypersensibilité à la pression (HNPP) sont des maladies des troubles génomiques bien connues, causées, dans la vaste majorité des cas, par des CNVs du gène *Peripheral Myelin Protein 22 (PMP22)*. En effet, sa duplication, causant la CMT1A, et sa délétion réciproque, causant la HNPP, se produit au niveau d'un fragment de 1,4Mb sur le chromosome 17p12 (Lupski, de Oca-Luna et al. 1991; Hoogendijk, Hensels et al. 1992).

Le syndrome de Smith-Magenis (SMS) est une maladie neurocomportementale avec un retard mental (RM), une dérégulation du sommeil, et des caractéristiques craniofaciales spécifiques accompagnés d'anomalies congénitales. Cette maladie résulte d'une haplo-insuffisance du gène *RAI1 (Retinoic Acid Induced 1)* sur le chromosome 17p11.2 (Slager, Newton et al. 2003). Dans 70 à 80% des patients atteints de SMS, le gène *RAI1* est délété dans une même région d'environ 3,7Mb, flanquée par des LCRs (Park, Stankiewicz et al. 2002). La duplication réciproque de cette région de 3,7Mb a été décrite chez les patients atteints par le syndrome de Potocki-Lupski qui se distingue du SMS par une hypotonie infantile, un retard de croissance, un syndrome autiste spécifique, une apnée du sommeil et des anomalies des structures cardiaques (Potocki, Bi et al. 2007).

Un autre syndrome, lié à une micro-délétion de 500-650kb flanquée par des LCRs au niveau du chromosome 17q21.31, cause notamment un retard de développement, une hypotonie, des dimorphismes faciaux et une épilepsie. Il a été estimé que cette délétion était responsable jusqu'à 1% des cas de retards mentaux. Deux gènes candidats ont été proposés

comme étant à risque : *CRHR1* (*Corticotropin Releasing Hormone Receptor 1*) et *MAPT* (*Microtubule-Associated Protein Tau*).

2. Les CNVs dans les maladies fréquentes

Récemment, un grand nombre de CNVs a été identifié comme ayant un effet sur la susceptibilité de certaines maladies fréquentes, parmi lesquelles nous pouvons citer : les désordres neurologiques complexes, les maladies cardiovasculaires, les maladies du métabolisme, le cancer, les maladies autoimmunes et infectieuses, et d'autres maladies fréquentes (Almal and Padh 2012). En effet, en 2010, une étude d'association de CNVs fréquents sur génome entier a été menée par le *Wellcome Trust Case Control Consortium* (WTCCC) sur de larges cohortes de cas/contrôles pour 8 maladies fréquentes qui ont un impact majeur sur la santé publique (trouble bipolaire, cancer du sein, maladie coronarienne, maladie de Crohn, hypertension, PR, et diabète de type 1 et 2). Trois loci ont été observés avec une association significative avec 4 des maladies étudiées, qui étaient en DL avec des SNPs et donc avaient été identifiés auparavant lors des études GWAS (Craddock, Hurler et al. 2010).

a) Les CNVs et les troubles neurologiques

Une triplification du gène *SNCA* (*Synuclein Alpha*) au niveau du chromosome 4q21 a été associée à la forme autosomale dominante de la maladie de Parkinson. Ce gain de copies change l'expression de ce gène dont le taux de la protéine correspondante dans le sang et l'ARNm dans le tissu cérébrale est deux fois plus exprimé avec un large dépôt d'agrégats de cette protéine dans le cerveau (Singleton, Farrer et al. 2003).

Pour la maladie d'Alzheimer, le gain en nombre de copies du gène *APP* (*Amyloid Precursor Protein*), localisé au niveau du chromosome 21q21, a été proposé comme étant un facteur causal (Rovelet-Lecrux, Hannequin et al. 2006). Cette duplication a été associée au déclenchement précoce de la forme autosomale dominante de la maladie d'Alzheimer dans la population Allemande (Sennvik, Fastbom et al. 2000; Cabrejo, Guyant-Marechal et al. 2006; Rovelet-Lecrux, Hannequin et al. 2006).

Le retard mental et les troubles du développement sont peut-être parmi les phénotypes les plus complexes avec plusieurs CNVs associés (Almal and Padh 2012). Six duplications chevauchantes, au niveau du chromosome Xp11.22, ont été observées dans le RM lié à l'X

chez 6 hommes non apparentés. Cette duplication couvre une région de 370kb contenant 4 gènes (*SMC1A*, *RIBC1*, *HSD17B10* et *HUWE1*) dont trois candidats sont susceptibles de causer le phénotype de RM (Roeleveld, Zielhuis et al. 1997). Les CNVs du gène *MECP2* (*Methyl CpG binding Protein 2*) localisés au niveau Xq28 ont été également associés au retard du développement, et l'encéphalopathie infantile létale. (Van Esch, Bauters et al. 2005; del Gaudio, Fang et al. 2006; Bauters, Van Esch et al. 2008; Carvalho, Zhang et al. 2009).

L'Autisme est un trouble du comportement caractérisé par une large variété de symptômes. En effet, ces derniers varient d'une personne à l'autre d'où son appellation *Autism spectrum disorder* (Bailey, Phillips et al. 1996; Stefansson, Rujescu et al. 2008). Des délétions et des duplications ont été observées pour l'ASD comme la duplication de la région 15q11-13 et la délétion de la région 16p11.12. Cette dernière présente aussi une micro-duplication réciproque qui a été associée à l'autisme et présente dans environ 1% des cas (Weiss, Shen et al. 2008).

Enfin, plusieurs délétions localisées aux positions 1q21.1, 15q11.2 et 15q13.3 ont été associées à la schizophrénie dans une étude cas/contrôles réalisée par le consortium international pour la schizophrénie et confirmées par d'autres études (Consort 2008; Stefansson, Rujescu et al. 2008). Une autre étude a identifié 90 CNVs chez 54 patients dont 13 étaient des CNVs rares pour des gènes associés à la schizophrénie comme *MYT1L*, *CTNND2* et *ASTN2* (Vrijenhoek, Buizer-Voskamp et al. 2008). Mais ces derniers restent à confirmer sur de plus larges cohortes.

b) Les CNVs et les Cancers

La compréhension des mécanismes génétiques du cancer et l'identification de tous les variants alléliques possibles, qui prédisposent à une ou plusieurs variétés de cancers, sont devenues une priorité (Almal and Padh 2012). Dans la recherche de CNVs fréquents associés au phénotype malin des tumeurs, une carte génétique a été réalisée qui liste tous les CNVs connus coïncidant avec des gènes associés au cancer. De cette analyse, 49 gènes du cancer contiennent ou chevauchent des CNVs identifiés chez plus d'une personne sur une large population de référence (Shlien and Malkin 2009).

Parmi les exemples de CNVs associés à un risque élevé de cancer il y a le gène *MTSUI* qui est localisé sur le chromosome 8p. Il est caractérisé par une délétion couvrant 1128pb englobant l'exon 4 du gène. L'étude de cette délétion dans la population Allemande a

révélé une association avec la progression ralentie de la maladie chez des familles et des patientes avec un risque très élevé du cancer du sein (Frank, Bermejo et al. 2007).

c) Les CNVs et les maladies autoimmunes

Il a été démontré que beaucoup de CNVs sont concentrés dans des régions du génome contenant des gènes impliqués dans la réponse immunitaire (Nguyen, Webber et al. 2006). Les CNVs dans les gènes à effet dosage-dépendant peuvent contribuer au développement de divers maladies autoimmunes (Tableau 4) (Schaschl, Aitman et al. 2009). Le locus FCGR (*Low-affinity Fc gamma Receptor*) a été largement étudié dans les maladies autoimmunes non-systémiques. Breunis et ses collaborateurs ont observé que les CNVs du gène *FCGR2C* prédisposent au Purpura Thrombocytopénique Idiopathique (ITP). Ces CNVs induisent une surexpression significative des récepteurs FCGR2C chez les patients atteints d'ITP (Breunis, van Mirre et al. 2008). Le gène *FCGR3A* a été associé, quant à lui, à la PR avec une forte association en particulier lorsque l'haplotype *FCGR3A-FCGR3B* 158V-NA2 est présent (Morgan, Griffiths et al. 2000; Morgan, Barrett et al. 2006).

Tableau 4: Maladies autoimmunes associées à des CNVs dans le génome humain (d'après Schaschl, Aitman et al. 2009)

Maladies autoimmunes	Gènes avec CNVs impliqués	Localisation	Risques associés aux CNVs	Références
SLE et AASV	<i>FCGR3B</i>	1q23	Faible	Aitman, Dong et al. 2006; Fanciulli, Norsworthy et al. 2007
SLE sans AASV	<i>FCGR3B</i>	1q23	Faible	Willcocks, Lyons et al. 2008
SLE	<i>C4A/C4B</i>	6q21	Faible	Yang, Chung et al. 2007
SLE	<i>CCL3L1</i>	17q12	Elevé/ Faible	Mamtani, Rovin et al. 2008
ITP	<i>FCGR2C</i>	1q23	Elevé	Breunis, van Mirre et al. 2008
Psoriasis	<i>DEFB</i>	8p23	Elevé	Hollox, Huffmeier et al. 2008
Maladie de Crohn	<i>DEFB</i>	8p23	Faible	Fellermann, Stange et al. 2006
PR et diabètes de type 1	<i>CCL3L1</i>	17q12	Faible	McKinney, Merriman et al. 2008

AASV: anticorps cytoplasmiques anti-neutrophil associés à une vascularite systémique ; FCGR : récepteur Fc gamma ; ITP : Purpura thrombocytopénique idiopathique ; PR : Polyarthrite Rhumatoïde ; SLE: Lupus systémique érythémateux.

Dans la région du Complexe Majeur d'Histocompatibilité (CMH), des CNVs du gène du Complément 4 (*C4*) ont été décrits dans la région du CMH de classe III au niveau du chromosome 6. Le complément *C4* est une protéine effectrice du système immunitaire et dont la délétion complète a été associée au Lupus Systémique Erythémateux (SLE) depuis 1974 (Hauptmann, Grosshans et al. 1974; Shen, Wu et al. 1994). Néanmoins une question reste encore en suspens concernant ce gène. En effet, puisque ce dernier se trouve dans une région très complexe, est-ce que l'influence des CNVs de *C4* est indépendante de la région CMH de classe III ou est-ce que sa délétion complète est en déséquilibre de liaison avec l'allèle *HLA-DRB1*0301* (Schaschl, Aitman et al. 2009)?

D'autres gènes avec des CNVs du système immunitaire sont impliqués dans les maladies autoimmunes, comme les gènes de la bêta-défensine (*DEFB*) qui codent pour la famille de protéines antimicrobiennes. Ces CNVs ont évolué au cours de l'histoire par des événements de duplications répétées d'un gène grâce à une sélection positive (Hollox, Armour et al. 2003; Semple, Rolfe et al. 2003; Hollox and Armour 2008). Cette famille de gènes se situe dans des positions chromosomiques différentes : 2 clusters sur le chromosome 20 et un sur le chromosome 8. Récemment, Hollox et collaborateurs ont mis en évidence une association entre le nombre de copies élevé de *DEFB* sur le chromosome 8p23 et le psoriasis (Hollox, Huffmeier et al. 2008), alors qu'une autre équipe a observé une association avec la maladie de Crohn (Fellermann, Stange et al. 2006).

Enfin, le dernier gène qui serait à citer dans les maladies autoimmunes, est le gène *CCL3L1* (*Chemokine C-C motif Ligand 3-like 1*). Plusieurs études ont suggéré que les récepteurs de chimiokines, qui induisent la réponse inflammatoire, auraient un rôle dans la physiopathologie des maladies autoimmunes (Godessart and Kunkel 2001; Ezerzer and Harris 2007). Ce gène, ayant des CNVs qui peuvent varier de 4 à 14 copies par génome diploïde en fonction des populations, a été associé au SLE ainsi qu'à la PR et le diabète de type 1 (Mamtani, Rovin et al. 2008).

Chapitre 3 : Les CNVs dans la PR

I. Les CNVs associés à la PR par étude du génome entier

Les études d'association sur génome entier (GWAS) ont permis avec succès de mettre en évidence des associations entre SNPs et les maladies fréquentes, mais ces SNPs associés représentent une faible fraction de la composante génétique de la plupart de ces maladies et notamment de la PR (Stahl, Raychaudhuri et al. 2010; Okada, Wu et al. 2014). Des nouvelles GWAS de CNVs ont donc vu le jour pour trouver les CNVs associés à la PR et contribuer à l'analyse de l'héritabilité manquante.

En 2010, le WTCCC a mené la première GWAS de CNVs sur huit maladies dont la PR. Les CNVs ciblés sont ceux déjà inventoriés, lors des études exploratoires des CNVs dans le génome (Conrad, Pinto et al. 2010), avec une taille supérieure à 500pb. L'aCGH a été utilisé pour l'identification de ces CNVs sur 2000 patients atteints de PR et 3000 contrôles. Un seul locus a été identifié comme associé à la PR. C'est le locus HLA (Craddock, Hurles et al. 2010). Des CNVs ont été décrits dans plusieurs haplotypes HLA (Horton, Gibson et al. 2008), mais en raison du fort DL dans cette région, il est peut être donc normal de trouver des CNVs dans ce locus associés à la PR (Craddock, Hurles et al. 2010).

Depuis cette étude, plusieurs autres ont suivi. Dans la même année, Chen et ses collaborateurs reprennent les données SNPs issues d'une précédente GWAS réalisée sur les SNPs dans les sept maladies citées précédemment, dont la PR (Consortium 2007). Pour l'identification des CNVs à partir de ces données SNPs, ils ont utilisé un algorithme basé sur le modèle des chaînes de Markov cachées. L'équipe a identifié 9 loci associés à la PR, qui regroupent chacun plusieurs gènes potentiellement à risque (Chen, Li et al. 2010).

Dans la même ligne d'étude, une équipe canadienne entreprend, à quelques détails près, l'étude déjà réalisée en 2010 en utilisant les mêmes données SNPs issus de l'étude WTCCC de 2007, mais se concentre uniquement sur l'identification de CNVs associés à la PR. Les chercheurs ont utilisé l'algorithme PennCNV qui se base également sur le modèle des chaînes de Markov cachées, et qui semble avoir une résolution de détection de CNVs à 1kb près avec un faible taux de faux positifs. De plus, pour affiner leur étude, ils ont fait une étape de contrôle qualité en excluant certains échantillons qui ne correspondaient pas à leurs critères (possible sources de contamination, ancêtre non caucasien...). Plusieurs CNVs ont été également exclus de l'étude, comme ceux détectés dans les chromosomes sexuels et dans les

régions codantes pour les immunoglobulines, sources de faux positifs. A partir de ces données, 11 CNVs ont été significativement associés à la PR avec un seuil de $p < 1 \times 10^{-4}$ (Tableau 5) (Uddin, Sturge et al. 2011).

Tableau 5: Résultats d'associations significatives de la plupart CNVs présents dans les régions indiquées. Les gènes contiennent, ou chevauchent des CNVs. Les positions sont issues de la base de données NCBI (d'après Uddin, Sturge et al. 2011)

Gènes	Chromosome	CNVR début-fin (pb)	Délétion		Duplication		Fisher Exact P	DGV ID*
			Cas	Contrôles	Cas	Contrôles		
ALOX5AP	13q12.3	30136207-30264780	0	0	35	2	5.06×10^{-12}	
SRGN	10q22.1	70423581-70542187	6	1	21	1	4.35×10^{-9}	48614
Désert de gène	7p21.3	11696007-11753538	20	0	0	0	1.60×10^{-8}	3665
LCP2	5q35.1	169605980-169668498	0	0	24	2	5.26×10^{-8}	
PREX1	20q13.13	46676571-46882578	0	0	18	0	9.70×10^{-8}	12195
B2M	15q21.1	42606873-42824209	0	0	26	4	2.71×10^{-7}	3961
LITAF	16p13.13	11506474-11695215	1	0	15	0	5.87×10^{-7}	
PRKCH	14q23.1	60986674-61130641	0	0	20	2	1.39×10^{-6}	49385
IRF1	5q31.1	131735192-131863294	2	0	18	2	1.39×10^{-6}	
TNFAIP3	6q23.3	138064233-138239517	0	0	11	0	5.25×10^{-5}	
TNIP1	5q23.3	150379418-150470998	2	0	18	2	7.59×10^{-5}	6455

*DGV ID : Identifiant dans la base « *Database of Genomic Variants* ».

Certains des CNVs identifiés, listés dans le tableau 5 comprennent des gènes impliqués dans les maladies autoimmunes. Parmi ces derniers, deux sortent du lot, *TNFAIP3* et *TNIP1*. En effet, les produits de ces deux gènes, A20 et ABIN/Nafla respectivement, interagissent ensemble pour influencer l'ubiquitination d'IKKg (IkB kinase-g) qui est essentiel pour la voie de signalisation NF- κ B (Elder 2009). De plus, la protéine A20 régule la dégradation de plusieurs composants de la voie du TNF (Tumor Necrosis Factor). Des polymorphismes de *TNFAIP3* ont été déjà associés à la PR, et les souris *knockout* pour ce gène développent des inflammations dans plusieurs organes et des arthrites (Plenge, Cotsapas et al. 2007). *IRF1* est également un candidat potentiel puisqu'il code pour le facteur de transcription des interférons de type 1 et est impliqué dans la réponse inflammatoire induite par les interférons α , β et γ (Uddin, Sturge et al. 2011).

L'expression du gène *ALOX5AP*, ayant le CNV associé avec la PR avec le plus de significativité, est régulée par la fixation de TNF- α sur le promoteur d'*ALOX5AP*, qui module sa propre expression (Reddy, Serio et al. 2003). Les études fonctionnelles ont montré que les souris déficientes pour ce gène développaient une arthrite moins sévère (Griffiths, Smith et al. 1997). Le gène *LITAF* a été décrit comme étant surexprimé chez les modèles de souris avec des arthrites induites (Inoue, Matsumoto et al. 2009). Pour le gène *PRKCH*, une association a été observée avec la PR dans la population Japonaise. Enfin, les deux gènes *LCP2* et *B2M* sont également des gènes de l'auto-immunité puisque *LCP2* est important dans le développement normal des lymphocytes T et que *B2M* appartient au CMH de classe I (Clements, Yang et al. 1998).

Enfin plus récemment, une étude a entrepris l'analyse d'association entre CNVs et PR dans le génome entier à partir des données SNPs obtenues dans une autre GWAS (Plenge, Cotsapas et al. 2007), en utilisant trois méthodes d'analyses différentes. L'identification des CNVs par la méthode d'analyse « SNP par SNP » a permis de détecter une forte association dans la région du CMH, avec un maximum d'accumulation de SNPs adjacents, qui englobe une délétion déjà décrite de *HLA-DRB1* et de *C4*. La deuxième méthode d'analyse de cluster de SNPs a permis d'identifier, outre les associations avec les délétions des deux précédents gènes, deux régions de 4,3 et 28kb contenant des clusters de SNPs significatifs au niveau des chromosomes 10p et 13q, respectivement. La troisième méthode d'analyse utilise l'algorithme PennCNV qui a permis de mettre en évidence des associations protectrices pour des délétions sur les chromosomes 2 et 14 et des associations à risque pour des délétions sur le chromosome 20 (Wu, Shete et al. 2013).

II. Les CNVs associés à la PR par étude de gènes candidats

Les résultats de détections de CNVs à grande-échelle sur génome entier ont été analysés par *Gene Ontology* (GO) et ont révélé qu'un grand nombre des gènes liés à des CNVs sont impliqués dans la réponse immunitaire, le métabolisme des xénobiotiques, la réponse aux stimuli extérieurs et dans d'autres mécanismes moléculaires (Feuk, Carson et al. 2006; de Smith, Walters et al. 2008). Dans cette partie, nous nous intéresserons aux gènes candidats impliqués dans la réponse aux stress oxydatifs et dans la réponse immunitaire.

1. La famille des Glutathion S-Transférase

L'augmentation du taux des espèces réactives de l'oxygène (ROS) qui induisent des altérations oxydatives des composants cellulaires comme l'ADN, est l'une des causes d'inflammation dans la PR (Halliwell 1995). La famille des Glutathion S-Transférase (GSTs) sont des isoenzymes qui catalysent des réactions de conjugaison avec plusieurs composés endogènes et exogènes (van Bladeren 2000). La classification de ces enzymes est basée sur la similarité des séquences protéiques, plusieurs classes existent donc (Alpha, Mu, Pi, Thêta et Sigma).

Les membres de la famille GST interviennent dans le processus de biotransformation des xénobiotiques pour faciliter leur élimination. Ce processus se divise en phase I et phase II. Les réactions de la première phase consistent en la transformation d'un composé en un métabolite plus polaire par formation d'un groupement fonctionnel (-OH, -NH₂ ou -SH). Cette transformation se fait surtout par les enzymes de la famille des Cytochromes P450. La deuxième phase implique des enzymes qui permettent de rendre le métabolite issu de la phase I encore plus facile à être excrété en le rendant plus hydrophile. Les enzymes de la famille GST interviennent au niveau de la phase II de la biotransformation. Les enzymes GST catalysent des réactions de conjugaison pour former un thio-ester entre un glutathion tripeptide endogène et un xénobiotique (Figure 27) (Jancova, Anzenbacher et al. 2010).

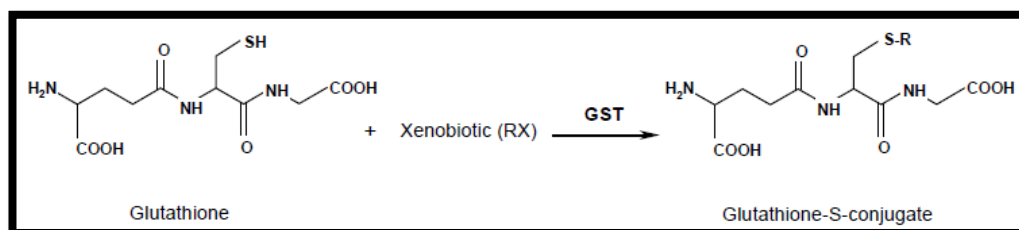


Figure 27: La réaction de conjugaison de la phase II de la détoxification des xénobiotiques (d'après Jancova, Anzenbacher et al. 2010).

Une augmentation de l'activité de ces enzymes a été observée dans le sang des patients atteints de PR, ce qui laisse suggérer qu'ils sont impliqués dans la réponse au stress oxydatif dans cette maladie (Ostalowska, Birkner et al. 2006; Surapneni and Chandrasada Gopan 2008). Il a été donc supposé qu'une variation dans l'efficacité de la détoxification des ROS et leurs produits, comme par exemple une variation génétique dans les enzymes impliqués, pourrait moduler le risque et la sévérité de la PR (Mattey, Hassell et al. 1999).

a) *Le gène Glutathion S-Transférase Mu 1*

(1) Fonction biologique du gène

Le gène Glutathion S-transférase de classe Mu 1 (*GSTM1*) code une enzyme qui est exprimée au niveau du cerveau, des lymphocytes, des muscles et enfin au niveau des gonades. Cette enzyme est importante dans la protection de la moelle osseuse contre des molécules génotoxiques telles que le benzène. L'expression de *GSTM1* est d'autre part corrélée avec la survie des cellules lymphoïdes dans la leucémie lymphoblastique (Wang, Groves et al. 2000).

(2) Les variations de nombre de copies du gène *GSTM1*

Parmi les polymorphismes du gène *GSTM1*, la délétion complète du gène, qui est un polymorphisme de nombre de copies, est présente dans environ 50% de la population Caucasienne (Garte, Gaspari et al. 2001). L'allèle *GSTM1-nul* ou *GSTM1*0* induit à l'absence complète de la protéine correspondante et a pour conséquence la dérégulation de la détoxification des xénobiotiques (McLellan, Oscarson et al. 1997).

La classe des glutathion S-transférase Mu est formée de cinq gènes, qui sont alignés sur le même chromosome 1p13.3 (Xu, Wang et al. 1998). Le gène *GSTM1* est localisé entre le gène *GSTM2* et le gène *GSTM5*, ce qui a permis de proposer une première hypothèse selon laquelle cette délétion était le résultat d'une recombinaison homologue inégale entre l'exon 8 du gène *GSTM2* et l'exon 8 du gène *GSTM1* (Vorachek, Pearson et al. 1991), ces deux régions étant identiques à plus de 99% sur 583 nucléotides. Une étude plus récente vient contredire cette hypothèse, elle démontre que le gène *GSTM2* restait intact chez des individus *GSTM1-nul* et que le gène *GSTM1* était entièrement délété. Cette même investigation a pu apporter une explication plausible à l'apparition de la délétion par la recombinaison entre deux séquences, presque identiques, flanquant le gène *GSTM1*. La séquence flanquante en 5' de *GSTM1* a un pourcentage d'identité égale à 99,5% avec la séquence de jonction de la délétion (*breakpoint*), alors que la région flanquante en 3' partage 98,2% d'identité avec cette même séquence. La recombinaison entre la région 3' et 5' induit donc le mécanisme de la délétion (Xu, Wang et al. 1998).

Une étude de 1997 menée par McLellan et collaborateurs a mis en évidence, par Southern Blot, une séquence qui se trouve dans le cluster des GST mu, entre *GSTM2* et *GSTM5*, qui contient le gène *GSTM1* dupliqué en tandem. L'auteur de ce travail émet

l'hypothèse que cette duplication est due à une NAHR. De plus, il suppose que, si la fréquence de délétion est de 50% dans la population Caucasienne, cela signifie que la fréquence de duplication, issue du même phénomène de recombinaison que la délétion, peut être donc d'une fréquence aussi élevée et induire un génotype 2 copies/0 copies ou +/- (Figure 28), (McLellan, Oscarson et al. 1997).

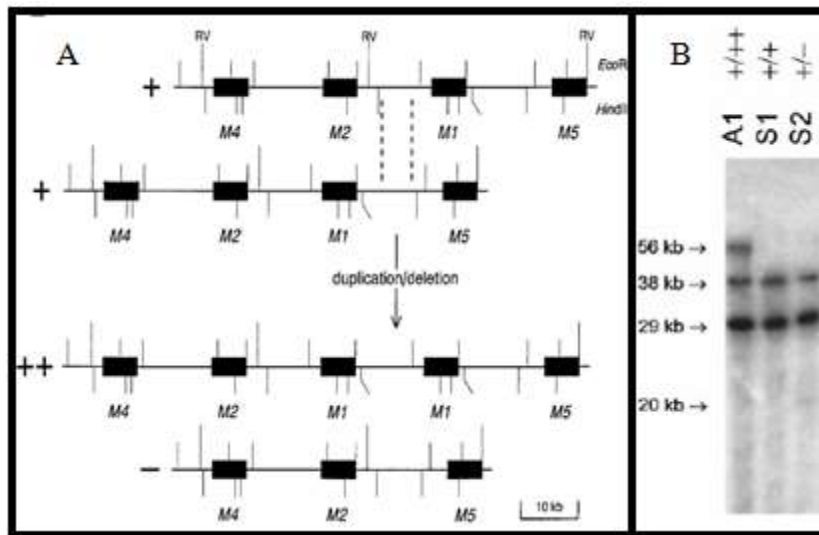


Figure 28: Mise en évidence de la duplication en tandem du gène *GSTM1*. A) Présentation schématique de la position du gène *GSTM1* au niveau du chromosome 1 avec le phénomène de recombinaison homologue inégale conduisant à la duplication/délétion. B) Résultat de Southern Blot montrant la bande de 56kb qui contient *GSTM1* dupliqué (A1) et qui n'est pas présente chez les autres individus (S1 et S2). La bande de 38kb représente la séquence qui contient une seule copie du gène *GSTM1*, et la bande de 29kb est une séquence de *GSTM4* qui est utilisée comme contrôle (d'après McLellan, Oscarson et al. 1997).

Grâce à un nouvel outil de quantification du nombre de copies basé sur la technologie de qPCR, plusieurs études se sont intéressées à l'analyse des CNVs de *GSTM1*. En effet, une étude faite en 2004, sur les CNVs des gènes *GSTM1* et *GSTT1* chez des familles trio dont l'enfant est atteint d'asthme, montre que les deux gènes peuvent avoir 0, 1 ou 2 copies. De ce fait, les auteurs supposent donc que deux copies représentent le génotype homozygote (+/+) pour le gène et qu'une copie représente le génotype hétérozygote (+/-) (Brasch-Andersen, Christiansen et al. 2004). En 2009, une autre équipe crée son propre outil de calcul de CNV afin d'analyser les données brutes issues de la PCR quantitative. Elle confirme les résultats précédemment trouvé en 2004 (Rose-Zerilli, Barton et al. 2009).

Plus récemment, des chercheurs chinois ont investigué l'association de nombre de copies de *GSTM1* et *GSTT1* avec la cataracte chez les personnes âgées. En prenant le nombre de copies égales à deux comme étant le nombre de copies le plus fréquent de *GSTM1* pour le

calcul de CNV, ils mettent en évidence des génotypes avec un nombre de copies supérieur à deux. Les fréquences des CNV sont de 61,3% pour les zéro copies, 24% pour une copie, 10% pour deux copies et 4,7% pour les copies supérieures à deux (Zhou, Hu et al. 2010). Enfin une équipe suédoise a utilisé le même outil que les chercheurs chinois pour étudier la relation de ces CNV avec la sévérité dans la PR. Les auteurs ont choisis pour leur analyse de CNV le nombre de copie égale à 1 comme étant le plus fréquent. Les fréquences des CNV trouvées dans leur étude diffèrent largement par rapport à celles trouvées par l'équipe de Zhou et collaborateurs : 53,7% pour les zéro copie, 38,3% pour une copie, 17,8% pour deux copies et 0,13% pour 3 copies (Lundstrom, Hartshorne et al. 2011). Le choix du nombre de copie le plus fréquent reste, dès lors, le point de divergence entre toutes ces études. Pourtant, l'étude des délétions fréquentes dans le génome entier a permis de conclure que les gènes ayant des délétions étaient soumis à une sélection purificatrice et que les duplications étaient anciennes (Locke, Sharp et al. 2006; McCarroll, Hadnott et al. 2006), ce qui suggère que les allèles délétés seraient les plus fréquents.

(3) L'expression du gène *GSTM1*

En 1997, l'équipe de McLellan et ses collaborateurs a étudié l'effet des CNVs sur l'expression du gène *GSTM1* et plus particulièrement sur l'activité enzymatique de la protéine codée par ce gène. Ils ont comparé l'activité de *GSTM1* chez un individu ayant 2 copies du gène avec celle d'individus ayant des CNVs. Ils ont montré que lorsqu'il y a une délétion complète du gène, il n'y a plus d'activité enzymatique. Lorsque le gène est délété sur un seul chromosome, l'activité était plus faible. Enfin, lorsqu'il y a 3 exemplaires du gène, l'activité était augmentée par rapport à la normale (McLellan, Oscarson et al. 1997). Ces résultats reflètent indirectement la variation de l'expression du gène en fonction des CNVs.

Plus récemment, des études ont analysé le profil transcriptomique du gène *GSTM1* en fonction des différents nombres de copies (McCarroll, Hadnott et al. 2006; Stranger, Forrest et al. 2007; Schlattl, Anders et al. 2011). Ces résultats concordent bien avec les résultats trouvés en 1997, puisque la délétion complète du gène n'induit pas de transcription du gène, la délétion hétérozygote induit une expression plus faible du gène que lorsqu'il y a deux copies (Figure 29).

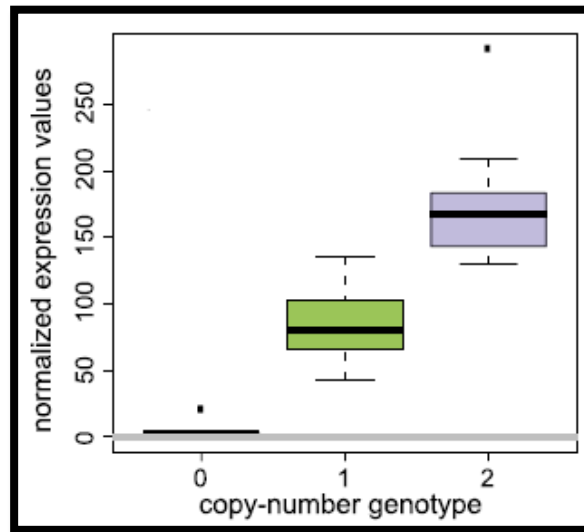


Figure 29: Expression différentiel du gène *GSTM1* en fonction du nombre de copies (d'après Schlattl, Anders et al. 2011).

b) Le gène glutathion *S*-Transférase Theta 1

(1) Fonction biologique de GSTT1

La protéine codée par ce gène a été purifiée la première fois par Meyer et collaborateurs en 1991 (Meyer, Coles et al. 1991). Le gène *GSTT1* est très fortement exprimé dans le foie où il constitue 4% des protéines présentes dans cet organe (Landi 2000; Wang, Groves et al. 2000). Comme *GSTM1*, il intervient dans la détoxification cellulaire des xénobiotiques, la protection contre les espèces réactives à l'oxygène, mais *GSTT1* ne semble pas intervenir dans la survie des lymphocytes T. En effet, *GSTT1* n'est pas du tout exprimé dans les cellules T, mais il est fortement exprimé dans les érythrocytes et dans la lignée myéloïde (Wang, Groves et al. 2000).

GSTT1 montre d'importantes différences au niveau de son activité catalytique par rapport aux autres classes de GST. En effet *GSTT1* catalyse faiblement le 1-chloro-2,4 dinitrobenzene contrairement aux autres classes (μ 1, α ...) qui sont, par contre inactives pour l'acide étacrynique. Enfin cette enzyme traite rapidement les substrats en utilisant une grande concentration de glutathion et libère le produit conjugué au glutathion plus facilement pour qu'il entame le cycle suivant de l'élimination des xénobiotiques (Landi 2000).

(2) Les variations de nombre de copies de GSTT1

Le gène *GSTT1* est localisé au niveau du chromosome 22 au niveau de la bande 22q11.2. (Webb, Vaska et al. 1996). Comme pour *GSTM1*, le gène *GSTT1* présente une complète délétion due à une recombinaison homologe entre deux régions hautement homologues et qui se trouvent de part et d'autre du gène (Sprenger, Schlagenhauser et al. 2000; Parl 2005). Il existe une hypothèse qui suggère qu'une duplication ancestrale du gène théta a eu lieu avant la spéciation des mammifères et qui a abouti à l'apparition des gènes *GSTT1* et *GSTT2* (Figure 28) (Landi 2000).

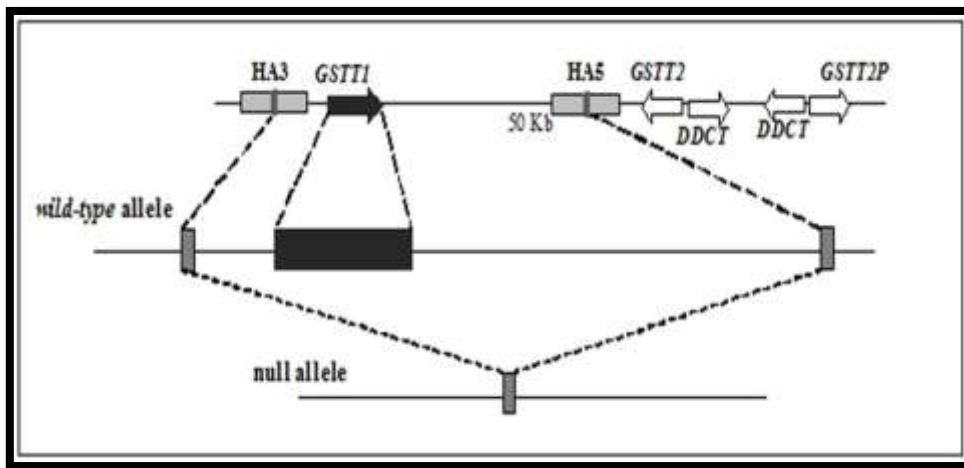


Figure 30: Localisation du cluster de gène de la famille GST theta sur le chromosome 22q11.2. La délétion de *GSTT1* se produit par recombinaison des régions répétées de part et d'autre du gène (d'après Parl 2005).

Les études portant sur les CNV présents dans le gène *GSTT1* montrent des résultats qui manquent de précision. En effet, dans le travail de Brasch-Andersen et collaborateurs, le nombre de copies du gène ne dépasse pas les deux copies (Brasch-Andersen, Christiansen et al. 2004) mais sans préciser si les deux copies sont sur le même chromosome ou sur deux chromosomes différents. L'étude réalisée en 2009 aboutit aux mêmes résultats précédents (Rose-Zerilli, Barton et al. 2009). Enfin, l'analyse effectuée par l'équipe chinoise montre qu'il peut y avoir un nombre de copies supérieur à deux, en prenant 2 copies comme le plus fréquent dans la population obtenant ainsi un nombre de copies qui varie entre 0 et plus de 2 copies (Zhou, Hu et al. 2010).

(3) Expression du gène *GSTT1*

Aucune étude sur l'expression du gène *GSTT1* en fonction des variations de nombre de copies dans la PR n'a été réalisée. Par contre, le travail de Butler et ses collaborateurs réalisé dans l'étude de l'expression des GSTs dans les poumons a montré que l'expression de *GSTT1* était cohérente avec les génotypes du gène puisque lorsque ce dernier est complètement délété (-/-) il n'y avait pas du tout d'ARNm du gène. Le gène à l'état hétérozygote est deux fois moins exprimé qu'à l'état homozygote (+/+) (Figure 31). Ce résultat est confirmé dans tous les tissus des poumons (Butler, Hackett et al. 2011).

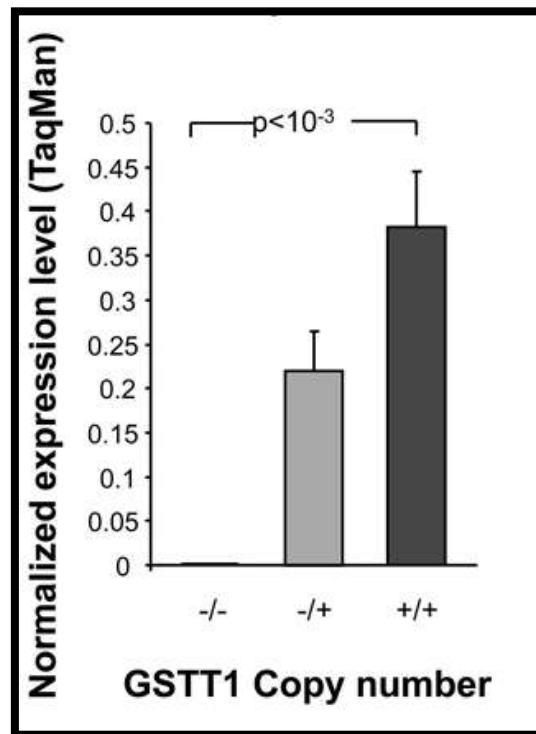


Figure 31: Expression du gène *GSTT1* en fonction des trois génotypes (d'après Butler, Hackett et al. 2011).

c) Associations entre gènes *GST* et la PR

Plusieurs études, basées sur une PCR spécifique de la présence du gène, ont produit des résultats divergents quant à l'association entre les polymorphismes de *GSTM1* et la PR. En effet, la première étude cas/contrôles faite sur la population caucasienne montre que l'allèle *GSTM1*-nul est significativement associé à un score moyen de *Larsen* élevé (qui permet une évaluation globale de chaque articulation) (Larsen 1995) par rapport aux patients ayant le gène *GSTM1* (89,9 et 74,7, respectivement) (Mattey, Hassell et al. 1999). Une autre

équipe Coréenne confirme ce résultat en observant que l'allèle *GSTM1*-nul augmentait significativement le risque de la PR avec un OR= 1,40 (IC [1,02 : 1,92]). Donc par conséquent, lorsque le gène *GSTM1* est présent, il serait associé à un risque moindre de la PR. De plus, l'absence du gène *GSTM1* semble être liée à la sévérité de la maladie (Yun, El-Sohemy et al. 2005).

Le travail de Bohanec Grabar et collaborateurs en 2009 vient controverser ces résultats. En évaluant le score d'activité de la maladie dans 28 articulations, aucune association entre la présence du gène *GSTM1* et la PR n'a été trouvée (Bohanec Grabar, Logar et al. 2009). Plus récemment, l'étude menée par une équipe suédoise ne trouve pas d'association entre les variations de nombre de copies de *GSTM1* et différents sous-groupes de la PR (présence ou absence d'HLA-SE, d'érosion, de FR) (Lundstrom, Hartshorne et al. 2011). Enfin, deux récentes méta-analyses ont vu le jour. La première, réalisée sur 2291 patients PR et 2713 contrôles, issus de plusieurs populations (Européenne et Asiatique) à partir de 12 études, n'a pas révélé d'association significative entre *GSTM1* et la PR, même après stratification selon les différents facteurs cliniques, biologiques et environnementaux associés à la PR (Song, Bae et al. 2012). La deuxième a été menée sur 4636 cas-index et 3916 contrôles issus de 8 études et montre une association significative uniquement dans la population de l'Asie de l'Est (Ji and Lee 2013).

Comme pour *GSTM1*, plusieurs études se sont intéressées à la relation entre les polymorphismes de *GSTT1* et la PR. Dans l'étude faite sur les patients atteints de PR en Corée, ce gène n'était associé ni au risque de développement de la PR (OR= 0,80, IC [0,59-1,10]) ni à la sévérité de la maladie (OR= 0,62, IC [0,26-1,47]) (Yun, El-Sohemy et al. 2005). Une autre investigation montre un résultat similaire (Mattey, Hassell et al. 1999). D'autres résultats viennent s'opposer aux précédents en montrant que l'allèle *GSTT1*-nul augmenterait significativement le risque de développer la maladie avec un OR=2,761, (IC [1,114-6,843]) (Bohanec Grabar, Logar et al. 2009). Une méta-analyse, réalisée en 2012 et consacrée à l'étude d'association entre *GSTT1* et la PR dans 2652 patients et 4117 contrôles à partir de 7 études, n'a pas mis en évidence une association, mais un risque plus élevé a été observé dans le groupe de patients séropositifs, pour les auto-anticorps, ayant la délétion complète du gène (Chen, Huang et al. 2012). Enfin, une dernière méta-analyse, réalisée sur 3174 cas-index et 2958 contrôles issus de 5 études, n'a pas révélé d'association significative avec la PR (Ji and Lee 2013).

d) Relation entre le tabac, les gènes *GSTM1*, *GSTT1* et la PR

A ce jour, ce point soulève encore des controverses. En effet, il s'avère que les diverses études menées dans le but de répondre à cette question « y-a-t-il une relation entre l'exposition au tabac, les polymorphismes présents dans *GSTM1* et la PR ? » aboutissent à des résultats divergents. Une équipe scientifique, en 2002, a entrepris l'étude de cette relation en incluant le FR dans la relation entre les variants de *GSTM1* et le tabac. Elle a pu établir en premier lieu que les patients qui étaient *GSTM1*-nul et qui avaient déjà fumé avaient un score de Larsen et de HAQ (*Health Assessment Questionnaire*) significativement plus élevé par rapport aux patients *GSTM1*-nul n'ayant jamais fumé (Mattey, Hutchinson et al. 2002). Cette relation entre l'allèle *GSTM1*-nul, tabac et PR est consolidée par la présence du FR puisqu'il existe une différence significative pour le pourcentage de présence de FR entre les individus fumeurs et les non fumeurs (75,5% et 51,2%, respectivement). En outre, cette association ne semble pas exister chez les patients fumeurs portant l'allèle *GSTM1*.

Un travail plus récent, mené sur des patientes PR, a obtenu un résultat tout à fait contradictoire avec le précédent, il suggère que la présence du gène *GSTM1* est associée à un risque plus élevé de la PR pour les patientes exposées au tabac avec un OR= 2,10, IC [1,13-3,89], (Criswell, Saag et al. 2006). Une étude plus large faite sur les variants des trois classes de la famille GST (*GSTM1*, *GSTT1* et *GSTP1*) vient contredire cette proposition en suggérant que la délétion de *GSTM1* combinée à l'exposition au tabac ne modifiait en rien l'activité de la maladie. Cette même étude met en évidence un effet combiné entre *GSTM1*-nul et *GSTT1*-nul chez les patients fumeurs avec un risque 4,84 fois plus élevé pour la PR (Bohanec Grabar, Logar et al. 2009). Un résultat similaire a été mis en évidence récemment par Keenan et collaborateurs qui montrent qu'il existe une relation entre le gène *GSTT1* et le tabac dans la PR et ainsi qu'une interaction entre le gène *GSTT1*, le tabac et la production d'auto-anticorps. Aucune relation n'a été trouvée entre *GSTM1* et ces différents facteurs dans cette étude (Keenan, Chibnik et al. 2010). L'équipe de Lundström rapporte que les copies supérieures à 1 de *GSTM1* semblent être un facteur à risque significatif pour la production des auto-anticorps dans la PR (OR 2, IC [1.07-3.74]), dans le sous-groupe de patientes non fumeuses dont l'âge est supérieur ou égale à 60 ans à l'apparition de la maladie (Lundstrom, Hartshorne et al. 2011).

2. La famille de gènes du Récepteur Fc Gamma

Les récepteurs Fc forment un groupe hétérogène de glycoprotéines de surface, sur les cellules hématopoïétiques, qui facilitent l'interaction entre le complexe immunitaire et les cellules effectrices du système immunitaire. Ces récepteurs régulent plusieurs réponses immunitaires à médiation humorale et cellulaire incluant la phagocytose, la dégranulation et la cytotoxicité cellulaire dépendante des anticorps (ADCC). Ils interviennent également dans la régulation transcriptionnelle des cytokines et de chimiokines. Ils reconnaissent spécifiquement les différents isotypes d'Immunoglobuline (IgA, IgG, IgD, IgE et IgM) (Li, Ptacek et al. 2009). Il a été démontré que des variations dans les gènes qui codent les récepteurs Fc pouvaient altérer l'efficacité des phagocytes mononucléaires à phagocyter les complexes immuns, ce qui peut avoir des conséquences sur certaines maladies autoimmunes (Salmon, Edberg et al. 1990; Salmon, Edberg et al. 1992). Le récepteur le plus largement étudié est celui spécifique aux IgG (Fc γ R).

De récentes études ont montré que les Fc γ Rs jouent un rôle important dans la coopération entre le système immunitaire innée et le système adaptatif. Chez l'être humain, la famille des Fc γ Rs se divise en trois sous-familles de récepteurs : Fc γ RI (CD64), Fc γ RII (CD32) et Fc γ RIII (CD16). Les récepteurs Fc γ RI, Fc γ RIIA/C et Fc γ RIII sont activateurs et Fc γ RIIB est inhibiteur notamment de la phagocytose, de la cytotoxicité et de la présentation de l'antigène. Ils agissent grâce à l'*immunoreceptor tyrosine activating* ou à *inhibitory motif* (ITAM et ITIM respectivement, Figure 32) (Li, Ptacek et al. 2009). Dans notre étude nous nous intéresserons à la sous-classe des récepteurs Fc γ RIII.

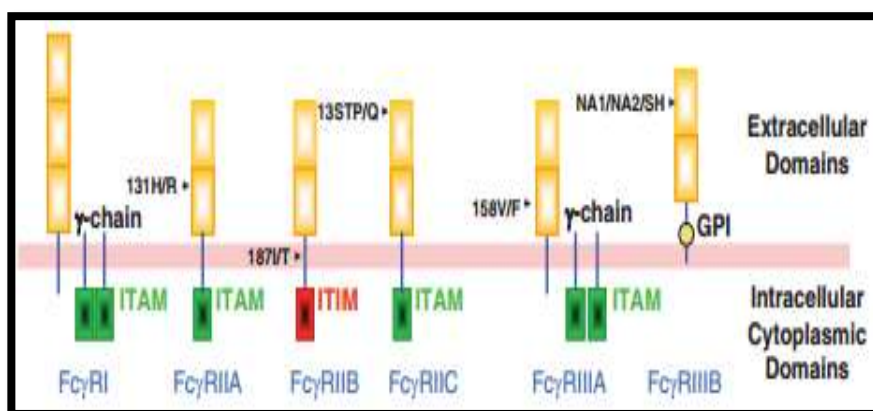


Figure 32: Structure des classes de Fc γ Rs avec la localisation des SNPs fonctionnels caractérisés (d'après Li, Ptacek et al. 2009).

a) Fonction des FcγRIII

Deux récepteurs, FcγRIIIa et FcγRIIIb, sont codés par deux gènes, *FCGR3A* et *FCGR3B* respectivement. FcγRIII est considéré comme un récepteur ayant moins d'affinité à fixer les IgG (*Fcγ low-affinity receptor*). Cependant FcγRIIIa fixe des IgG monomériques avec une moyenne affinité et les deux FcγRIIIa et FcγRIIIb se lient efficacement aux IgG multimériques et aux complexes immuns.

La protéine FcγRIIIa est exprimée à la surface des monocytes, des macrophages tissus-spécifique, des cellules dendritiques et des cellules *Natural killer* (NK) (Crocker and Kimberly 2005), alors que FcγRIIIb est exprimée à la surface des neutrophiles et des basophiles (Simmons and Seed 1988). FcγRIIIb est fortement exprimé avec 150 000 à 200 000 récepteurs par cellule et interagit avec la β2-integrine (CD11b/CD18) chez les neutrophiles (Li, Ptacek et al. 2009). Plusieurs polymorphismes des gènes FcγRIII ont été décrits comme étant associés à plusieurs maladies autoimmunes, mais dans la PR, ce sont les variants du gène *FCGR3B* qui ont été décrits comme étant associé à la susceptibilité de la maladie.

b) Les variations de nombre de copies de *FCGR3B*

Dans le cluster des gènes FcγRs, des duplications et des délétions d'environ 82kb ont été décrites pour le gène *FCGR3B* (de Haas, Kleijer et al. 1995; Koene, de Haas et al. 1996) (Figure 33). Ces duplications sont identiques à 98% (Bailey, Gu et al. 2002) et sont apparues depuis la séparation entre les grands singes et les petits singes actuels (Bailey, Gu et al. 2002; Nimmerjahn and Ravetch 2007). Une étude en 1993 a montré par Southern blot que ces CNVs étaient le résultat d'une recombinaison inégale entre *FCGR2A* et *FCGR2B* qui induisait la formation d'un pseudogène *FCGR2C* (Warmerdam, Nabben et al. 1993; Marques, Thabet et al. 2010).

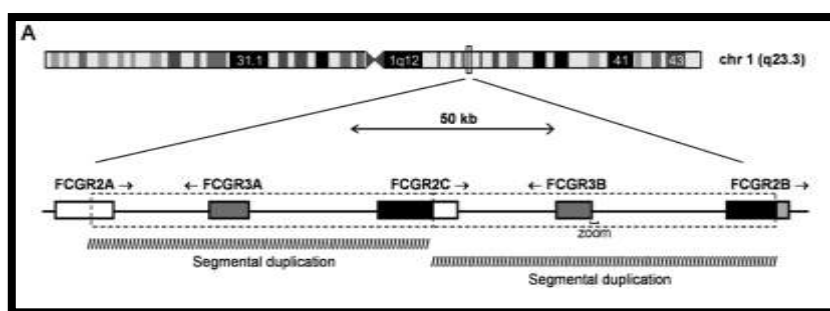


Figure 33: Cluster des gènes FcγRs au niveau du chromosome 1 (d'après Marques, Thabet et al. 2010).

Une récente étude a montré que, par ailleurs, il existe plusieurs *breakpoints*, distribués sur l'ensemble de la région FCGR, et regroupés dans des clusters de mutations créant un *hotspot* de NAHR. Ces derniers, favorisent l'apparition de délétions de différentes tailles entre *FCGR3B* et *FCGR3A*, comme il est illustré dans la Figure 34 (Machado, Hardwick et al. 2012).

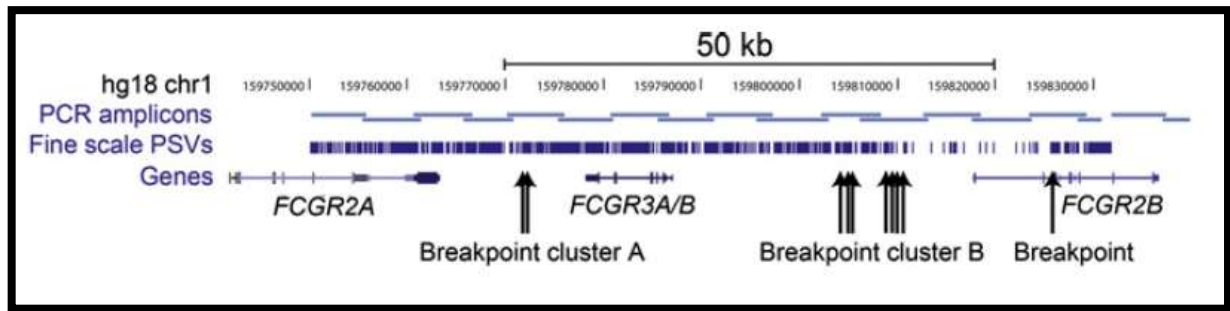


Figure 34: Région dupliquée de FCGR3 avec les différents clusters de *breakpoints*. PSVs: *Paraologue Sequences Variations* (d'après Machado, Hardwick et al. 2012).

Plusieurs techniques ont été utilisées pour quantifier les CNVs du gène *FCGR3B*. Un résumé des différentes études utilisant l'une de ces trois méthodes est donné dans le Tableau 6. La première technique à être proposée est la PRT par Hollox et collaborateurs en 2009. Elle a permis d'identifier un nombre de copies dans plusieurs populations contrôles qui varie entre 1 et 4 copies, avec des fréquences dans la population caucasienne de 3% pour une copie, 81% pour deux copies et 16% pour le nombre de copies supérieur à 2.

Tableau 6: Pourcentage des individus suivant le nombre de copie de *FCGR3B* et en fonction des différentes méthodes de quantification utilisées (d'après McKinney and Merriman 2012)

Study	Population	n	Detection method	CN < 2	CN = 2	CN > 2
Fanciulli	European	1052	SybrGreen, rounded CN	18.9	60.9	22.2
Bourzanos		221	SybrGreen, post-hoc CN assignment	11.3	69.7	19.0
McKinney		1309	Taqman, post-hoc CN assignment	6.0	82.0	12.0
Hollox		189	PRT/REDVR	4.2	87.8	7.9
Morris		365	PRT/REDVR	7.1	81.1	11.8
Neiderer		1817	PRT	7.8	81.7	10.6
Marques		304	MLPA	6.7	83.9	9.4
Breunis		129	MLPA	7.6	82.2	10.9
Mamtani		409	Taqman, rounded CN	~10	~82	~8
Yang		173	SybrGreen, Rounded CN	10.4	39.3	49.7
Zhou	146	Taqman, post hoc CN assignment	19.2	64.4	15.5	
Asano	2057	Taqman, post hoc CN assignment	7.0	71.1	21.9	
Hollox	64	PRT/REDVR	7.8	59.3	32.8	
Niederer	1874	PRT	10.0	76.1	13.9	
Hollox	African ^b	35	PRT/REDVR	14.3	85.7	0
Molokhia		450	PRT/REDVR	16.7	76.0	7.3
Niederer		152	PRT	17.2	75.2	6.6

^aMainland Chinese, Hong Kong Chinese, Japanese, Vietnamese.

^bYoruban, Afro-Caribbean, Kenyan.

Cette méthode a été par la suite largement reprise dans d'autres études (Morris, Roberts et al. 2010; Niederer, Willcocks et al. 2010). Presque en même temps, la MLPA et la qPCR commencent à être utilisées pour la quantification de *FCGR3B* et montrent des résultats similaires à ceux obtenus par PRT (Marques, Thabet et al. 2010).

c) Associations entre *FCGR3B* et la PR

Le gène *FCGR3B* est un bon gène candidat fonctionnel dans la PR de par ses fonctions essentielles dans la régulation du système immunitaire. L'hypothèse de l'implication de *FCGR3B* dans la pathogénicité de la PR a été investiguée la première fois en 2009 par McKinney et collaborateurs dans trois populations : de Nouvelle-Zélande, d'Allemagne et du Royaume Uni. Ces études cas/contrôles permettent d'observer une association significative entre le nombre de copies inférieur à 2 et la PR dans la population Allemande (OR 2,01, IC 95% [1,37-2,94]). Ils ont par la suite réalisé une méta-analyse avec les données CNVs de *FCGR3B* quantifiés dans une autre étude sur des patients PR (Breunis, van Mirre et al. 2009). Cette méta-analyse a permis de mettre en évidence une association significative entre *FCGR3B* et la PR en considérant les CNs 0 et 1 copie comme étant l'allèle mineur et les CNs 2 copies ou plus comme l'allèle majeur (McKinney, Fanciulli et al. 2010). Plus récemment, une autre étude met en évidence une association significative entre la PR et les CNs inférieurs à 2 de *FCGR3B* (OR 2,5, IC 95% [1,2-5,4]). La méta-analyse qui a suivi a confirmé cette association, avec un OR 1,47 (IC 95% [1,13-1,92]) (Graf, Lester et al. 2012). Ce résultat positif est confirmé par une autre étude réalisée dans la même année et qui a également mis en évidence une association entre la délétion de *FCGR3B* et la production de Facteur Rhumatoïde (P=0,01) (Robinson, Carr et al. 2012). Mais cette association reste tout de même non reproductible puisque dans un autre travail aucune association significative entre *FCGR3B* et la PR dans une population Allemande n'a été identifiée (Marques, Thabet et al. 2010). De plus, une méta-analyse réalisée sur des maladies autoimmunes, dont la PR, a montré que *FCGR3B* était associé au phénotype autoimmun en regroupant toutes les cohortes, mais aucune association significative n'a été identifiée spécifiquement entre la PR et *FCGR3B*. (McKinney and Merriman 2012).

3. La famille des « chemokine (C-C motif) ligand »

La famille des chimiokines est divisée en deux groupes, les CXCs (ou sous-famille α) codés par 15 gènes et les membres de C-C motifs (ou la sous-famille β) codés par 16 gènes,

localisés au niveau du chromosome 17q11.2-q12 et 8 gènes dispersés sur les chromosomes 2, 7, 9 et 16 (Zlotnik and Yoshie 2000). Parmi ces 16 gènes sur le chromosome 17, il existe 2 paires de gènes très proches *CCL3/CCL3L1* et *CCL4/CCL4L1* (Modi 2004). Cette région a été caractérisée par la présence de CNVs dont les mieux décrits sont ceux du gène *CCL3L1*. Ce gène a été proposé comme étant associé à la PR, mais à ce jour très peu d'études ont été réalisées. Nous nous intéresserons donc à ce gène pour la suite du travail.

a) **Fonction du gène *CCL3L1***

Les gènes *CCL3/CCL3L1* codent pour les *macrophages inflammatory protein* (MIP)-1 α , plus particulièrement les isoformes LD78 α et LD78 β , respectivement. Ce sont des β -chimiokines avec un faible poids moléculaire qui agissent comme des cytokines pro-inflammatoires en induisant l'activation d'une large variété de cellules immunitaires, et plus particulièrement les LTs CD8+, et les cellules dendritiques immatures. Elles sont inhibées par IL4, IL10 et IL13 (Standiford, Kunkel et al. 1993; Berkman, John et al. 1995). Ces protéines sont produites par des macrophages pour attirer les lymphocytes et amplifier la présence des macrophages dans le site de l'inflammation. L'isoforme LD78 β est deux fois plus efficace que l'isoforme LD78 α pour le chimiotactisme des monocytes et les lymphocytes (Figure 35) (Menten, Struyf et al. 1999).

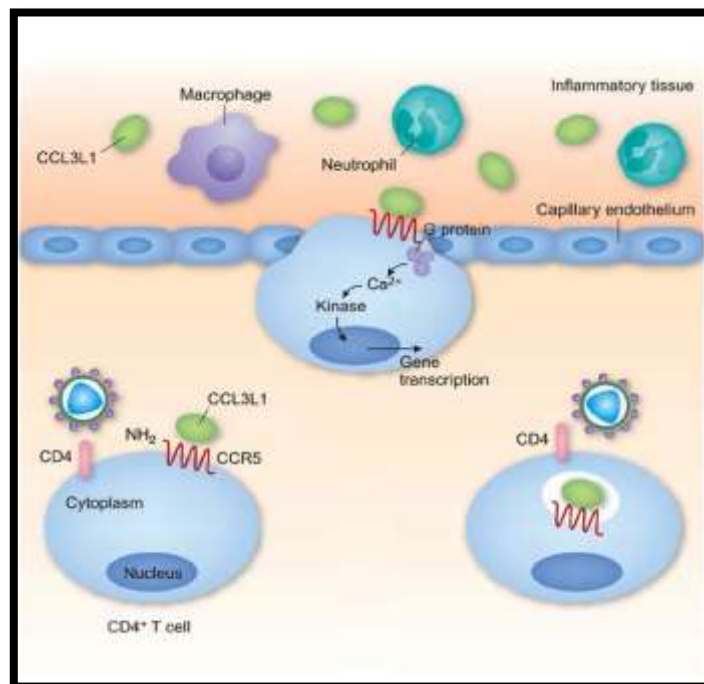


Figure 35: Interaction entre CCR5 et CCL3L1 lors de l'infection par le virus HIV (d'après Lederman and Sieg 2007).

De plus, l'isoforme LD78 β est l'agoniste le plus efficace du co-récepteur CCR5 qui, par sa fixation, permet une activité anti-virale et inhibe la réplication du virus HIV lors de l'infection (Menten, Struyf et al. 1999; Aquaro, Menten et al. 2001; Lederman and Sieg 2007). C'est pour cette raison que les CNVs du gène *CCL3L1* ont le potentiel d'influencer les désordres immunologiques et l'auto-immunité, ce qui fait de lui un gène candidat fonctionnel dans la PR.

b) Variants de nombre de copie du gène *CCL3L1*

Le gène *CCL3L1* est localisé au niveau du chromosome 17 dans un cluster de gènes homologues et peut subir des duplications ou des délétions de 90kb (Figure 36). Il partage environ 96% de séquences similaires avec son paralogue *CCL3*.

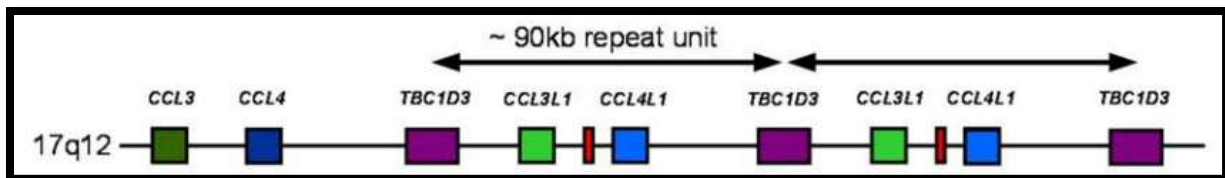


Figure 36: Région génomique de *CCL3L1* au niveau du chromosome 17 (d'après Walker, Janyakhantikul et al. 2009).

Contrairement aux deux précédents gènes étudiés, l'origine des CNVs dans *CCL3L1* n'est pas encore bien établie. Le nombre de copie du gène varie entre 0 et 4 copies dans la population Européenne et peut aller jusqu'à 14 copies dans la population Africaine (Gonzalez, Kulkarni et al. 2005). Mais ce nombre a été récemment revue à la baisse pour atteindre 10 copies (Carpenter, Farnert et al. 2012). Deux techniques ont été développées pour la quantification des CNVs de *CCL3L1*, en premier la qPCR pour le diagnostic lors des infections au virus HIV (Gonzalez, Kulkarni et al. 2005) puis la PRT, quatre ans plus tard (Walker, Janyakhantikul et al. 2009). Mais, rapidement, les scientifiques se rendent compte que les CNs trouvés par la qPCR décrite en 2005 étaient plus élevés par rapport à la PRT. C'est alors qu'en 2009 une étude comparative entre la PRT et la qPCR confirme que le nombre de copies estimé par qPCR était surestimé car les amorces et la sonde utilisées s'hybridaient sur *CCL3L1*, et sur son pseudogène *CCL3P1*. La PRT était, quant à elle, spécifique du gène *CCL3L1* et les résultats étaient concordants avec ceux trouvés par la méthodologie aCGH (Field, Howson et al. 2009).

a) Association entre *CCL3L1* et la PR

Malgré son intérêt potentiel dans la PR, peu d'études se sont intéressées à l'association entre les variants du gène *CCL3L1* et la PR. La première étude d'association a été réalisée en 2008 sur deux populations issues de la Nouvelle-Zélande et du Royaume Uni. Le nombre de copies supérieur à 2 a été associé à la PR avec un OR=1,34 (IC 95% [1,08-1,66]) ainsi que la délétion du gène (OR=1,30, IC 95% [1,00-1,54]) (McKinney, Merriman et al. 2008). Les deux études récentes réalisées sur la population Européenne n'ont pas mis en évidence d'associations entre les CNVs du gène et la PR (Carpenter, Walker et al. 2011; Nordang, Carpenter et al. 2012).

4. Les autres gènes avec des CNVs dans la PR

Parmi les autres gènes candidats fonctionnels, il y a également le gène *VPREB1*, localisé au niveau du chromosome 22q11.2 et caractérisé par une délétion. Il a été associé à différents troubles immunologiques tels que l'arthrite auto-immune chronique et l'arthrite juvénile. Ce gène intervient dans le développement de lymphocytes pre-B (LBs non matures) (Sullivan, McDonald-McGinn et al. 1997; Pelkonen, Lahdenne et al. 2002; McLean-Tooke, Spickett et al. 2007). Ce gène n'a été analysé qu'une seule fois dans la PR par Yim et collaborateurs dans la population Coréenne et une association de la délétion du gène *VPREB1* est observée avec un OR de 12,1 (CI 95% [2,8-51,6]) (Yim, Chung et al. 2011). Un deuxième gène dont la contribution à la physiopathologie de la PR a été confirmée par des études fonctionnelles chez les rongeurs, est *NCF1* (*Neutrophil Cytosolic Factor 1*). Il est très polymorphe et sujet à des duplications segmentaires (Olofsson, Holmberg et al. 2003; Antonell, de Luis et al. 2005). Il code pour une protéine qui intervient dans le complexe NOX2, qui produit les ROS dans les phagocytes et les CPA (Olofsson, Holmberg et al. 2003; Hultqvist, Olofsson et al. 2006). En 2011, une première étude sur les CNVs de *NCF1* a montré qu'il y a moins de patients PR avec un nombre de copies élevé par rapport au contrôle ($p=0,037$) (Olsson, Nerstedt et al. 2012). Enfin, deux gènes caractérisés par une délétion ont été associés à la PR et répliqués dans deux populations différentes (Espagnole et Chinoise) (Docampo, Rabionet et al. 2010; Lu, Guo et al. 2011). Ces gènes sont *LCE3B* et *LCE3C* (*Late Cornified Envelope (LCE) 3B and 3C*). Néanmoins, une étude très récente réalisée sur la

population Allemande n'a pas observé d'association significative entre la délétion de ces deux gènes et la PR (Bergboer, Umicevic-Mirkov et al. 2012).

Chapitre 4 : Les objectifs

La polyarthrite rhumatoïde (PR) est le plus fréquent des rhumatismes inflammatoires chroniques, avec une prévalence de 0,3 à 1% au sein de la population mondiale adulte. C'est une maladie complexe et multifactorielle. L'exposition au tabac est le facteur environnemental majeur établi dans la PR depuis les années 80. Le gène HLA-DRB1 est le facteur génétique majeur associé à la maladie. D'autres facteurs génétiques de risque ont été caractérisés, dont les variants associés présentent des polymorphismes de type SNP (Single Nucleotide Polymorphism). La composante génétique de la polyarthrite rhumatoïde n'est cependant pas encore complètement connue. Il est donc évident que les études de SNPs, ne permettent pas d'élucider à elles seules, entièrement l'étiologie génétique de la maladie.

Au cours de ces dernières années, de plus en plus d'études démontrent que les Variations de Nombre de copies (CNVs) sont associées à des maladies complexes, et notamment à la PR. Les études sur génome entier ont mis en évidence plusieurs gènes ayant des CNVs associés à la PR, mais ces derniers n'ont pas été répliqués sur plusieurs autres populations. De plus, les études gènes candidats menées sur des cas/contrôles, issus de différentes populations, ont montré des résultats d'association divergents pour plusieurs gènes candidats fonctionnels. Enfin, les méthodes de quantification de ces CNVs varient entre chaque étude et ne permettent pas forcément d'avoir toutes les informations moléculaires et génotypiques pour analyser correctement certains CNVs complexes.

Nous avons donc entrepris l'étude des CNVs de plusieurs gènes candidats fonctionnels dans la PR pour essayer d'apporter de nouvelles conclusions et élargir le champ de recherche du point de vue moléculaire. Pour cela nous avons choisi d'étudier quatre gènes candidats fonctionnels : les gènes Glutathion S-transférase de classe Mu1 (*GSTM1*) et Theta1 (*GSTT1*), le Récepteur Fc Gamma 3B (*FCGR3B*) et le gène *CCL3L1* de la famille *Chimiokine (C-C motif) Ligand*. Il était intéressant d'étudier le gène *GSTM1* car la fréquence de sa délétion complète était élevée (environ 50% dans la population Caucasienne). De plus, la majorité des travaux ayant investigué les CNVs du gène *GSTM1* ont analysé l'association entre la présence ou l'absence du gène et la PR, sans caractériser les génotypes. Enfin, le mécanisme moléculaire induisant ces CNVs décrits en 1997, n'a pas été reconfirmé avec les nouvelles techniques depuis 1997. Pour *GSTT1*, malgré la faible fréquence de sa délétion (17%), il joue un rôle important, comme pour *GSTM1*, dans la détoxification des ROS et des xénobiotiques. De plus, l'étude de ces deux gènes dans la PR n'a pas porté sur les génotypes de nombre de copies. Enfin, les gènes *FCGR3B* et *CCL3L1* sont des candidats de choix puisqu'ils codent pour des acteurs moléculaires prépondérants dans le déroulement de la réaction inflammatoire. Ces gènes sont également

caractérisés par des CNVs multialléliques pouvant atteindre 4 copies dans la population Caucasienne, et qui sont peu investigués dans la recherche de facteurs à risque de maladie complexe.

Avec les échantillons de familles trio (un patient ou cas index et ces deux parents) à notre disposition, les objectifs de notre étude seront donc divisés en deux axes :

- La mise au point de techniques moléculaires adaptées pour la caractérisation du génotype et la quantification des CNVs de chaque gène.
- L'étude de la transmission de ces variants et de leur association avec la PR ainsi que de la relation entre les données génomiques produites et les données cliniques, biologiques et environnementales connues pour les patients étudiés.

Chapitre 5 : Patients & Méthodes

I. Patients

Dans un premier temps, nous avons à disposition un set de 200 familles trio (un patient PR et ses deux parents), d'origine française (4 grands-parents français). Le diagnostic de la PR satisfait aux critères de *l'American College of Rheumatology*. Les différentes caractéristiques des patients sont regroupées dans le tableau ci-dessous. L'ADN a été extrait préalablement à partir de prélèvement de sang par une méthode d'extraction standard. Tous les individus participant à l'étude ont été informés et ont fourni un consentement éclairé signé. L'étude a été approuvée par les comités d'éthique de l'hôpital Bicêtre et de l'hôpital Saint Louis (AP-HP Paris, France).

Tableau 7: Caractéristiques des cas index PR

	200 familles trio	
	Nombre*	Pourcentage
Femmes	178/200	89%
Age < 40 years	148/200	74%
FR séropositifs	158/200	79%
ACPA séropositifs	149/193	77%
Présence d'érosion	170/200	85%
Présence de nodules	52/200	26%
TE	98/185	52.97%
Ayant au moins un allèle HLA- DRB1 shared epitope	158/200	79%

* Nombre de cas index avec des données connues / Nombre total de cas index

FR : Facteur Rhumatoïde ; ACPA : Antibodies to Citrullinated Protein Antigen ; TE : Tobacco Exposure

II. Caractérisation des CNVs du gène *GSTM1*

A. Identification de la présence/absence du gène

La première étape de la caractérisation des CNVs du gène *GSTM1* est une étape de *screening* qui consiste en l'identification de la présence ou de l'absence de ce gène. L'amplification spécifique fait intervenir trois amorces : P1, P2 et P3 (Figure 37) (Zhong, Wyllie et al. 1993). Les amorces P1 et P2 amplifient une séquence du gène *GSTM1* et une

séquence du gène *GSTM4* (153pb) qui sert de contrôle interne. Les amorces P1 et P3 vont permettre l'amplification d'une séquence spécifique de *GSTM1* (230pb).

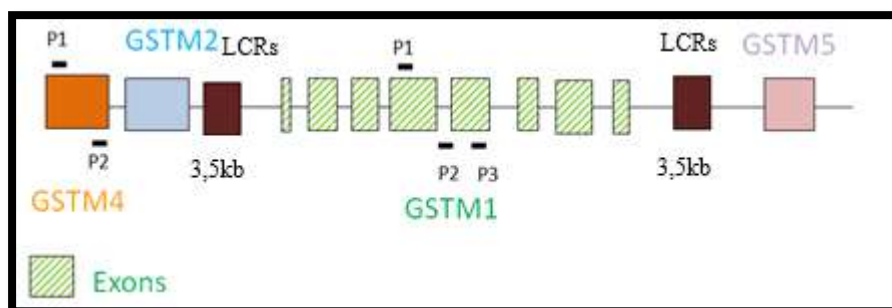


Figure 37: Position des amorces P1, P2 et P3 au niveau des gènes *GSTM1* et *GSTM4*.

Si le gène est délété, il y a seulement amplification avec P1 et P2 du gène *GSTM4* (153pb). Les séquences de ces amorces sont décrites dans le tableau 8.

Tableau 8 : Caractéristique des amorces d'amplification

Amorces	Pourcentages en GC%	Taille de l'amplicon
Amorces sens		
P1 5'-CGCCATCTTGTGCTACATTGCCCG-3'	58	
Amorces anti-sens		
P2 5'-ATCTTCTCCTCTTCTGTCTC-3'	45	153pb
P3 5'-TTCTGGATTGTAGCAGATCA-3'	40	230pb

Le mélange réactionnel (25µl de volume final) contient (en concentration finale) : 20ng d'ADN, 1X de tampon PCR, 1,5mM de MgCl₂, 800nM dNTP, 1U *RedTaq* Taq polymérase, et 0,8µM de chaque amorce. Les produits sont amplifiés avec le programme suivant : une étape de dénaturation initiale à 96°C pendant 1min, 37 cycles formés par : une deuxième dénaturation à 96°C (1min), une étape d'hybridation à 56°C (30s), une étape d'élongation à 72°C (30s). Une étape d'élongation finale est entamée après le dernier cycle et se fait à 72°C (5min). La migration des produits amplifiés se fait par électrophorèse dans un gel d'agarose 1,5% avec du Tris-Borate-EDTA (TBE) 1X. La révélation des produits amplifiés se fait sous UV avec une longueur d'onde de 312nm, mettant en évidence l'agent intercalant de l'ADN utilisé.

B. Identification du génotype par PCR classique

Pour accéder au génotype précis de chaque échantillon, une autre PCR a été mise au point par Buchard et collaborateurs en 2004. Cette réaction permet de co-amplifier une séquence du gène (de taille 625pb) et une séquence formée par la recombinaison des deux régions homologues de part et d'autre de *GSTM1* (de taille 4,7kb, Figure 20). Ainsi, lorsqu'un individu est homozygote pour la délétion, seule la séquence de 4,7kb est amplifiée. S'il est hétérozygote pour la délétion, il y a amplification simultanée de la séquence de délétion et de la séquence du gène. Enfin, si l'individu est homozygote pour le gène, seule la séquence du gène est amplifiée. Les amorces qui permettent l'amplification sont décrites dans le tableau 9 (Buchard, Sanchez et al. 2007). Seuls les échantillons positifs pour le gène *GSTM1* ont été génotypés lors de l'expérience.

Tableau 9 : Caractéristiques des amorces d'amplification du gène *GSTM1* et de la séquence de délétion

Amorces	Pourcentage en GC%	Taille de l'amplicon
Amorces de la délétion		
P1 sens 5'- AAGACAGAGGAAGGGTGCATTTGATA -3'	42	4,7kb
P2 anti-sens 5'-ACAGACATTCATTCCTCAAAGCGACCA-3'	46	
Amorces du gène		
P2 sens 5'-CAAATTCTGGATTGTAGCAGATCATGC-3'	40	625pb
P3 anti-sens 5'-CACAGCTCCTGATTATGACAGAAGCC -3'	50	

Le mélange réactionnel de volume final 25 μ l contient : 25ng à 100ng d'ADN, 1X de tampon PCR, 2mM MgCl₂, 960nM dNTP, 1U d'Ex-TaKaRa, 1,2 μ M de chaque amorce pour la délétion et 0,12 μ M de chaque amorce pour la région du gène. Le programme d'amplification est le suivant : une étape de dénaturation initiale à 94°C pendant 1min, 37 cycles formés par : une deuxième dénaturation à 94°C (30s), une étape d'hybridation et élongation à 66°C (5min), une étape d'élongation finale est entamée après le dernier cycle et se fait à 72°C (5min). La migration des produits amplifiés se fait par électrophorèse dans un gel d'agarose 0,8% avec du Tris-Borate-EDTA (TBE) 1X.

C. Quantification des CNVs

1. Quantification par PCR quantitative

L'estimation du nombre de copies a été évaluée en utilisant deux sondes fluorescentes portant chacune un fluorophore différent, selon la méthodologie TaqMan. La première sonde est spécifique du gène d'intérêt et la deuxième est spécifique d'un gène de référence dont le nombre de copies ne varie pas dans le génome. La réaction a été faite en triplicat sur les individus ayant au moins une copie du gène *GSTMI*. Les CNVs du gène d'intérêt ont été quantifiés avec une première sonde spécifique localisée au niveau de l'exon 1 (Hs02575461_cn, Applied Biosystems, Foster City, USA) et confirmés par une deuxième sonde localisée au niveau de l'intron 3 (Hs03352485_cn, Applied Biosystems). Une sonde spécifique du gène de référence RNase P (numéro du catalogue 4403326, Applied Biosystems) sert de normalisateur. Cinq contrôles positifs avec un CN connu ont été utilisés comme calibrateurs (NA10851, NA10852, NA10865, NA11893, NA11894, Coriell Institute <http://ccr.coriell.org/>, New Jersey, USA). La réaction a été réalisée dans un volume final de 20µl contenant 20ng d'ADN, 1X de Master Mix (contenant enzymes et autres réactifs nécessaires à l'amplification), 1X de chaque *assay* (sondes fluorescentes et amorces spécifiques du gène d'intérêt et du gène de référence respectivement). L'amplification et la détection se font dans la même machine (*Applied Biosystem 7500*), avec le programme suivant pour l'amplification : 95°C (10min), 40 cycles formés chacun d'une étape de dénaturation à 95°C (15sec) et d'une étape d'hybridation et d'extension à 60°C (60sec).

La valeur du « *cycle threshold* » Ct obtenue, permet de calculer le ΔC_t moyen (ΔC_{tm}) qui est la différence entre le Ct moyen du gène étudié et le Ct moyen du gène référence. Ce calcul repose sur le principe que, plus la quantité de matrice est importante, plus le nombre de cycles nécessaires pour atteindre la valeur seuil est faible et donc l'amplification se fait plus précocement.

Trois cas sont possibles (Figure 22) :

- $\Delta C_{tm} = 0$: le gène *GSTMI* et le gène de référence ont le même nombre de copies.
- $\Delta C_{tm} > 0$: le nombre de copies du gène est inférieur à celui de la référence : ceci implique que la fluorescence de la sonde spécifique de *GSTMI* met plus de temps à

arriver au cycle seuil. Donc, il n'y a pas assez d'ADN au départ de la réaction, l'absence du gène en est déduite.

- $\Delta C_{tm} < 0$: le gène est en plusieurs copies ce qui explique que la fluorescence atteint la C_t plus rapidement.

A chaque nombre de copies évalué est assignée une probabilité permettant d'assurer sa significativité statistique. L'analyse des données se fait grâce à un logiciel *Copy Caller*® d'*Applied Biosystem*, qui attribue pour chaque échantillon les valeurs de C_{tm} et ΔC_{tm} , le nombre de copies calculé et un indice de confiance relatif.

2. Quantification par le système *QuantStudio 12K flex*

Pour confirmer les génotypes obtenus par PCR classique et qPCR, nous avons tout d'abord testé un nombre limité d'échantillons avec une variante de PCR digitale (dPCR) sur puce, *QuantStudio 12K flex system digital PCR* (Life Technologies, Foster City, USA). Cette technologie se base sur le principe de la qPCR en séparant une réaction en 3072 événements distribués dans des nano-puits de 33nl. Les mêmes sondes que celles de la qPCR ont été utilisées pour la quantification des gènes *GSTM1* et *RNase P*. Pour la quantification du nombre de copies sur ces puces, il faut estimer le nombre d'événements en fonction du nombre de copies attendu : si le nombre de copies varie entre 0 et 2, la moitié de la puce (1536 événements) est suffisante, mais s'il est compris entre 2 et 4, une puce entière est recommandée (Figure 39).

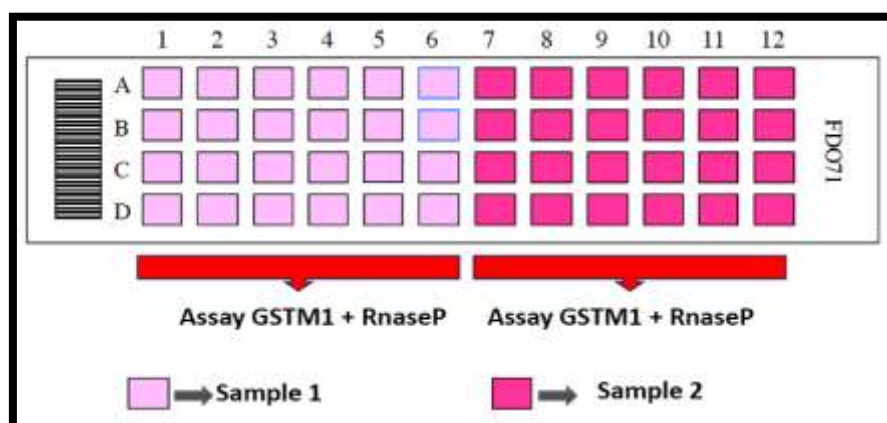


Figure 38 : Schéma de la distribution des échantillons sur les puces. Chaque carré contient 64 nano-puits de 33nl chacun.

Plus le nombre de copies augmente, plus le nombre de réplicats doit être important pour avoir une quantification précise de ces CNVs. Comme dans la littérature le nombre de copies de *GSTM1* dépasse très rarement 2, nous avons donc utilisé une puce pour deux échantillons. Un volume de 5µl est distribué dans chaque carré (64 réactions de 33nl par puits) contenant 40ng d'ADN, 2,5µl de *Digital master mix* et 1X de chaque assay (*GSTM1* et *RNaseP*). Le programme de la réaction est le même que pour la qPCR. L'analyse des résultats se fait par le logiciel Digital Suite™.

D. Mise en évidence de la duplication en tandem

1. Mise en évidence par la PCR *Long Range*

Pour mettre en évidence la duplication en tandem du gène *GSTM1* résultant de la NAHR décrite par McLellan et collaborateur en 1997, nous avons développé une PCR appelée *Long Range*, dont le protocole permet une amplification de fragments jusqu'à 40kb. Deux amorces (XLP1 et XLP2) ont été choisies à cet effet, une qui va s'hybrider à la région 3' de la première copie du gène et la deuxième va s'hybrider à la région 5' de la copie en tandem, ce qui va permettre d'amplifier une séquence de 17,8kb contenant successivement la partie terminale du gène, la région de jonction, et la partie initiale du gène dupliqué (Figure 38).

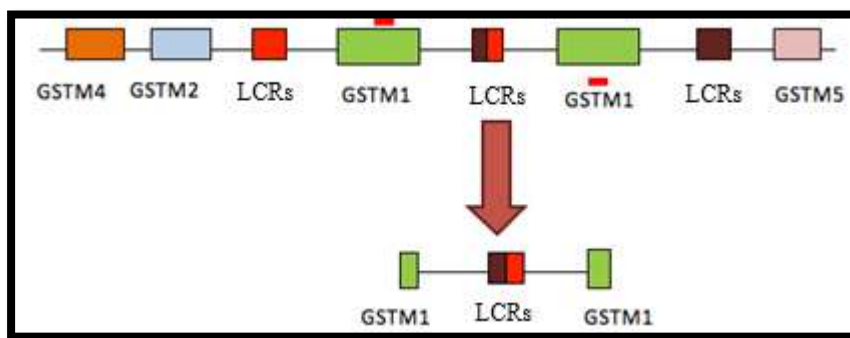


Figure 39 : Principe de l'amplification de la région dupliquée. Les deux traits rouges correspondent aux amorces XLP1 et XLP2.

La réaction est réalisée sur les individus ayant au moins 2 copies du gène *GSTM1*, dans 50µl de volume final avec 200ng d'ADN, 1X de tampon PCR, 2,5mM de MgCl₂, 2,5mM de dNTP, 2,5U de LA Taq TaKaRa, et 0,8mM de chaque amorce (la séquence correspondante est décrite dans le tableau 10).

Tableau 10 : Caractéristiques des amorces d'amplification de la région dupliquée

Amorces	Pourcentage en GC%	Taille de l'amplicon
Amorces de la délétion		
XLP1 sens 5'-GGGCCTTTATTCCCTCTCCCTCCTTCC-3'	58	17,5kb
XLP2 anti-sens 5'-TGTCACATGAACGAATGCAGGTGGA-3'	48	

L'amplification est réalisée selon le programme suivant : 94°C (1min), 30 cycles de 98°C (30sec) de dénaturation et d'hybridation/élongation à 68°C (15min), et enfin une élongation finale à 72°C (10min). La migration des produits amplifiés se fait par électrophorèse dans un gel d'agarose 0,6% avec du Tris-Borate-EDTA (TBE) 1X.

2. Mise en évidence par la PCR QX200 Droplet Digital

La quantification a été réalisée sur tous les individus ayant un nombre de copies égal à 2. La réaction a été réalisée avec la QX200™ Droplet Digital PCR (Bio-Rad Laboratories, California, USA) avec les mêmes contrôles positifs que ceux utilisés en qPCR. Des tests de contrôle qualité ont été réalisés suivant le protocole international dMIQE (*Minimum Information for the Publication of Digital PCR Experiment*) (Huggett, Foy et al. 2013).

a) Tests de qualité

(1) Test d'interférence des sondes

Les sondes fluorescentes utilisées pour le gène cible (*GSTM1*) et le gène de référence RNase P sont les mêmes que pour la qPCR. Des tests d'interaction entre la sonde du gène cible et celle de la référence ont été réalisés selon trois expériences : une première réaction en *simplex* avec seulement la sonde du gène cible, une deuxième réaction en *simplex* avec seulement la sonde du gène référence et enfin une réaction en *multiplex* avec les deux sondes. La réaction a été réalisée pour plusieurs échantillons contrôles, dans un volume final de 20µl avec 16ng d'ADN, 1X d'amorces et de sondes (concentrations finales de 900 et 250nmol/L, respectivement) pour le gène cible et le gène de référence et 2X de tampon *Droplet PCR Supermix*. Ensuite, la génération de *droplets* a été faite selon les recommandations du fabricant dans le générateur de *droplets*.

(2) Test de reproductibilité

Le deuxième test qui a été réalisé est un test de reproductibilité d'une même réaction de quantification dans une même plaque (test intra-plaque) et de la même réaction dans deux plaques réalisées avec les mêmes conditions et analysées séparément (test inter-plaques). Ce test a été réalisé avec un contrôle positif et un échantillon du set de trios avec les mêmes conditions utilisées que dans les tests d'interférence.

(3) Test de digestion

Afin de confirmer le génotype de nombre de copies (présence d'une copie du gène sur chaque chromosome ou deux copies dupliquées sur un même chromosome), la quantification a été précédée par une digestion enzymatique dans le but de séparer les éventuelles copies en tandem. En effet, si le gène est présent en copies dupliquées en tandem sur un chromosome, ces copies vont être présentes dans la même *droplet* et donc vont être comptabilisées comme un seul événement positif lors du comptage des *droplets*, ce qui va engendrer une quantification sous-estimée du nombre de copies. Par contre, si les deux copies sont séparées par une digestion, elles vont être distribuées dans deux *droplets* différentes et comptées comme deux événements positifs, donnant ainsi le bon nombre de copies (Figure 40).

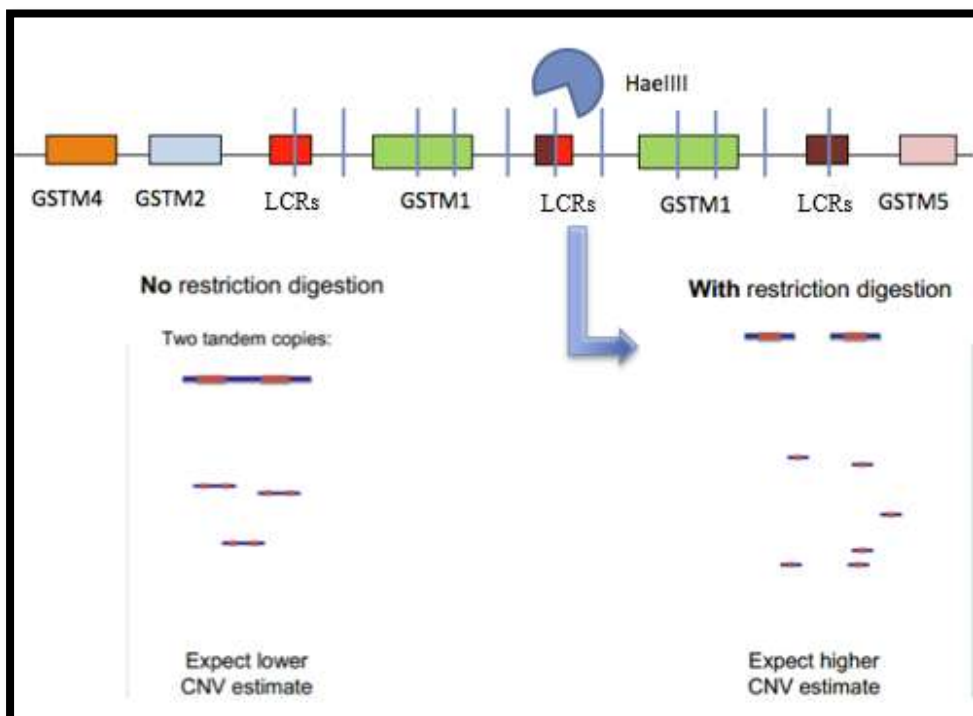


Figure 40 : Digestion avec l'enzyme HaeIII dans la région du gène GSTM1 et l'impact sur l'estimation des CNVs par la ddPCR. Les traits verticaux bleus indiquent les sites de coupure de l'enzyme HaeIII (d'après Bio-Rad Laboratories).

La digestion a été réalisée dans un volume final de 10µl avec un gradient d'enzyme (1U-20U pour trois enzymes, HaeIII, FokI et EcorV) avec 40ng d'ADN, 1X du tampon *CutSmart*TM.

b) Quantification

Après avoir validé ces tests, un protocole de ddPCR a été établi, et qui sera appliqué pour la quantification des CNVs des autres gènes candidats. Deux préparations sont réalisées dans un volume final de 20µl chacune, avec 16ng d'ADN non digéré et 25ng d'ADN digéré, respectivement et avec les réactifs décrits page 85. Les solutions sont ensuite vortexées pendant deux à trois minutes afin de fragmenter l'ADN, ce qui facilitera par la suite sa distribution lors de la génération de *droplets*. Après cette étape de génération, il y a amplification dans le thermocycleur C1000 TouchTM (Bio-Rad) suivant ces conditions : 95°C (10 min), 40 cycles de 94°C (30 sec), 60°C (60 sec) et une étape à 98°C (10 min). Les CNVs sont estimés avec le logiciel QuantaSoftTM (version 1.4, Bio-Rad) qui compte le nombre de *droplets* avec un signal fluorescent positif ou négatif. Les concentrations (copies/µl) du gène cible et de la référence sont calculées suivant la loi de Poisson et le rapport entre ces deux concentrations est calculé pour estimer le nombre de copies.

III. Caractérisation des CNVs du gène *GSTT1*

A. Identification de la présence/absence du gène

Pour le gène *GSTT1*, une amplification permettant la détection de la présence du gène, a été mise au point. La réaction se fait avec deux amorces qui amplifient une séquence du gène *GSTT1* et deux autres amorces qui amplifient une séquence du gène de la *β-globine*, utilisée comme contrôle interne (Tableau 11) (Chen CL, Liu Q et al. 1993).

La réaction se fait dans un volume final de 25µl avec : 20ng ADN, 1X de tampon PCR, 2,5mM de MgCl₂, 800nM de dNTP, 1U de *Red Taq* polymérase, 0,2µM de chaque amorce du gène *GSTT1* et de la *β-globine*. La réaction se fait selon le programme suivant : 96°C (1min), 37 cycles de 91°C (30sec), une étape d'hybridation 60°C (30sec) et 72°C (30s). Une élongation finale est entamée après le dernier cycle et se fait à 72°C (5min). La migration

des produits amplifiés se fait par électrophorèse dans un gel d'agarose 1,5% avec du Tris-Borate-EDTA (TBE) 1X.

Tableau 11 : Caractéristiques des amorces d'amplification du gène *GSTT1* et du gène de la β -globine

Amorces	Pourcentage en GC%	Taille de l'amplicon
Amorces <i>GSTT1</i>		
P1 sens 5'-TTCCTTACTGGTCCTCACATCTC -3'	39	480pb
P2 anti-sens 5'-TCACCGGATCATGGCCAGCA-3'	60	
Amorces β -globine		
P2 sens 5'-GAAGAGCCAAGGACGGT-3'	60	266pb
P3 anti-sens 5'-CAACTTCATCCACGTTACC -3'	50	

B. Identification du génotype

Comme pour *GSTM1*, nous avons utilisé une PCR qui permet la différenciation entre les génotypes homozygotes et hétérozygotes pour le gène *GSTT1*. Deux couples d'amorces ont été utilisés permettant l'amplification d'une séquence de 3,1kb correspondant à la région délétée, et une séquence de 1,6kb correspondant à la région du gène (Tableau 12) (Buchard, Sanchez et al. 2007). Les conditions de la réaction sont les mêmes que pour *GSTM1* (cf p78-79).

Tableau 12 : Caractéristiques des amorces d'amplification de la région délétée et du gène

Amorces	Pourcentage en GC%	Taille de l'amplicon
Amorces de la délétion		
P1 sens 5'-GAAGCCCAAGAATGGGTGTGTGTG -3'	55	3,1kb
P2 anti-sens 5'-TGTCCTCCATGGCCTCCAACATT-3'	54	
Amorces du gène		
P3 sens 5'-ACACCTTTGGGGAATGCTCTCG-3'	64	1,6kb
P4 anti-sens 5'-TGACCCCAGGTACCAGCCAGA -3'	55	

IV. Caractérisation des CNVs du gène *FCGR3B*

A. Quantification par qPCR

Pour le gène *FCGR3B*, nous avons utilisé la qPCR en premier lieu, puisqu'il n'y a pas de délétion complète du gène décrite dans la littérature et que la cartographie de la région du gène ne nous a pas permis de mettre au point une PCR classique permettant l'identification du

génotype. Comme pour les précédents gènes, une sonde spécifique a été utilisée pour le gène cible (Hs04211858_cn, *Applied Biosystem*) et une deuxième pour le gène de référence. La réaction a été réalisée dans les mêmes conditions que celle utilisée pour *GSTM1*. Deux calibrateurs avec un nombre de copies connu ont été utilisés, l'un avec deux copies et l'autre avec trois copies (NA10852 et NA10851, respectivement).

B. Quantification par ddPCR

Afin de déterminer les génotypes du nombre de copies du gène *FCGR3B*, nous avons utilisé la ddPCR sur les mêmes échantillons précédemment testés par qPCR. Nous avons réalisé les mêmes tests de contrôle qualité que ceux de l'étude du gène *GSTM1*. Pour cela, différentes enzymes ne coupant pas dans la séquence amplifiée (FokI, HaeIII, EcoRV, BsrI, RsaI et TspRI) et à des concentrations différentes ont été utilisées pour préciser la localisation de chaque copie au niveau des chromosomes homologues (par exemple si les 2 copies du gène sont chacune sur un chromosome ou dupliquées en tandem sur le même chromosome). Les tests de digestion ont été réalisés directement dans la solution d'amplification (20µl) selon les conditions suivantes : 33ng d'ADN, 1µl de solution de digestion (1x de tampon *CutSmart*TM, 1 à 10U d'enzyme et qspH₂O), 1X de chaque assay (*FCGR3b* et *RPP30*), et 1X de tampon *Supermix*. Une expérience de prédigestion (suivant les conditions décrites pour le gène *GSTM1*) a été réalisée pour s'assurer de la conformité avec les résultats obtenus par la digestion dans la solution d'amplification. Les CNVs ont été quantifiés en comparant la fluorescence du gène cible et celle du gène de référence *RPP30* en utilisant les mêmes sondes que celles décrites précédemment. Les trois contrôles positifs (NA10852, NA10851 et NA02587) ont été utilisés dans chaque expérience.

V. Caractérisation des CNVs du gène *CCL3L1*

A. Mise au point par qPCR

Les précédentes études ayant investigué les CNVs du gène *CCL3L1* avec la qPCR en utilisant une sonde spécifique n'ont pas tenu compte de la présence du pseudogène *CCL3P1* et du gène *CCL3L3*. Ce dernier a une similarité avec *CCL3L1* de 99%, ces deux gènes ne varient que par une seule paire de base. Nous avons donc choisi une amorce sens spécifique de *CCL3L1* avec le nucléotide spécifique à l'extrémité 3'. Nous avons également choisi la

séquence de la sonde spécifique du gène. Le détail des séquences utilisées est illustré dans le tableau 13.

Tableau 13 : Caractéristiques des amorces d'amplification et de la sonde spécifique du gène *CCL3L1*

Amorces	Pourcentage en GC%	Taille de l'amplicon
P1 sens 5'-TCTTCGGGGCTCTCAGGCCG -3'	70	
P2 anti-sens 5'-CCATCACCTGCTCCCAATCATGCA-3'	54	120pb
Sonde 5'-FAM_GGACTCACGTGGTGCAGAGAGGA-3'	61	

Puisque les amorces et la sonde ont été déterminées pour notre étude, nous avons testé leur spécificité. Pour cela, nous avons fait une PCR avec gradient de température pour chercher la meilleure température d'hybridation (volume final de 20µl avec 30ng d'ADN, 1X de *SYBR® Green Master Mix*, 900nM de chaque amorce). L'amplification a été réalisée dans le thermocycleur avec le programme : 95°C (10 min), 40 cycles à 94°C (30 sec), 56 à 66°C (60 sec) et une dernière étape à 98°C (10 min). L'analyse de la spécificité a été réalisée en qPCR avec la méthode de dissociation. En fonction de la taille et de la composition en nucléotide, la séquence amplifiée aura une température de fusion (T_m) spécifique. Lors de la réaction de dissociation, la température augmente jusqu'à 95°C. Plus la température approche la T_m , plus le nombre d'amplicons dénaturés augmente, ce qui entraîne, par conséquent, la libération de l'agent intercalant (SYBR Green) et l'augmentation de l'intensité de fluorescence. Puis, lorsque la température redescend et s'éloigne de la T_m , les séquences amplifiées s'hybrident de nouveau, emprisonnant ainsi le SYBR Green et induisant donc la diminution de l'intensité de fluorescence. Cette réaction se traduit par une courbe dite de dissociation (*melting-curve*) avec une phase ascendante correspondant à la dénaturation, un pic au niveau de la T_m , puis une phase descendante correspondant à l'hybridation. Chaque séquence a donc sa courbe de dissociation spécifique. Si une amplification non spécifique a lieu, nous devrions observer alors deux courbes de dissociation distinctes.

B. Mise au point de la ddPCR

Pour tester la spécificité de la ddPCR, nous avons analysé les *clusters* d'événements positifs. En présence d'amplification non spécifique, il devrait y avoir deux *clusters* d'événements positifs correspondants au gène cible. Ce test a été réalisé dans un volume final de 20µl contenant 30ng d'ADN, 1X de *Eva Green Super mix* et 900nM de chaque amorce. Le programme d'amplification est le même que pour la qPCR. Nous avons également refait ce

test mais cette fois-ci avec la sonde spécifique du gène *CCL3L1* et celle du gène de référence afin de vérifier la meilleure température d'hybridation de la sonde du gène. Cette expérience a été faite dans un volume final de 20µl avec 30ng d'ADN d'un échantillon contrôle (NA121354), 1X d'amorces et de sondes (concentrations finale de 900 et 250nmol/L, respectivement) pour le gène cible et le gène de référence et 2X de tampon *Droplet PCR Supermix*. Le programme est identique à celui de la qPCR.

1. Caractérisation des CNVs

Des tests de validation (interférence et reproductibilité) ont été réalisés selon les étapes décrites pages 83 et 84. La digestion a été faite dans la solution d'amplification et avant la réaction comme pour le gène *FCGR3B*. Plusieurs tests de digestion enzymatique ont été réalisés avec différentes enzymes (MseI, RsaI, TspRI, AluI, BsrI). La quantification a été réalisée suivant les mêmes conditions que pour les gènes *GSTM1* et *FCGR3B*. Un contrôle positif avec deux copies du gène a été utilisé (NA12264).

VI. Etude d'associations dans les familles trio

Les outils statistiques utilisés nous ont permis d'étudier l'association entre les gènes candidats et la maladie grâce à l'analyse de transmission des CNVs mis en évidence, dans les familles trio. L'analyse de risque a été investiguée entre ces variations et les facteurs cliniques, biologiques, génétiques et environnementaux associés à la PR.

A. Le Test de Déséquilibre de Transmission (TDT)

Ce test est utilisé pour comparer la transmission des allèles des parents hétérozygotes aux enfants malades. Ce test permet d'évaluer si la transmission allélique suit la loi de Mendel (c'est-à-dire 50% de chance pour chacun des allèles d'être transmis). Un Chi2 (X^2) de conformité à 1 degré de liberté (ddl) est calculé pour évaluer la significativité de l'écart à la transmission normale avec un risque d'erreur de 1^{ère} espèce de 5%.

$$Chi2 = \frac{(n1 - n2)^2}{n1 + n2}$$

n1= nombre d'allèle 1 transmis préférentiellement des parents hétérozygotes au descendant

n2= nombre d'allèle 2

H₀= n1=n2

H₁=n1≠n2

Si le Chi2 est supérieur à 3,84, un allèle est préférentiellement transmis des parents hétérozygotes aux patients. Cette méthode a été réalisée en utilisant le logiciel FBAT dans l'échantillon total et dans les différents sous-groupes de patients, stratifiés selon les données cliniques, biologiques, génétiques et environnementales. Pour la correction des tests multiples, un seuil de valeur de P de 0,05/nombre des tests effectués (soit 17) est considéré (correction de Bonferroni)

B. Test du Risque Relatif des Génotypes (GRR)

Le teste GRR permet de comparer les génotypes des contrôles virtuels (formés par les allèles non transmis) à ceux des patients. Puisque nous sommes en présence d'individus apparentés, nous avons donc utilisé le GRR de McNemar qui est le plus adapté (Schaid and Sommer 1994).

Le Chi2 à 2 ddl est calculé et la valeur seuil de significativité est dans ce cas de 5,81. Si la valeur du Chi2 est supérieure à 5,81, pour un risque d'erreur de 1^{ère} espèce de 5%, alors il y a une répartition des fréquences des génotypes significativement différente entre les contrôles virtuels et les patients.

Le GRR a été calculé pour l'échantillon total et pour les sous-groupes de patients formés à partir des données cliniques, biologiques, génétiques et environnementales. Le seuil considéré de la valeur de P est le même que pour le TDT après correction pour les tests multiples.

C. Association entre les facteurs associés à la PR et les CNVs

L'association entre l'exposition aux différents facteurs cliniques, biologiques, génétiques et environnementaux et la variable « CNVs » est mise en évidence par un Chi2 d'homogénéité. Le risque est ensuite quantifié par un calcul de l'Odds Ratio (OR).

Tableau 14 : Modèle de calcul de l'Odds Ratio.

	Pas d'exposition	Exposition
Variable à risque	a	b
Variable de référence	c	d

$$OR = \frac{(a \times d)}{(b \times c)} \text{ avec un Intervalle de Confiance (IC) 95\%}$$

Si l'exposition n'est pas associée à la présence de la variable, l'odds ratio est proche de 1. Si l'exposition est positivement liée à la présence de la variable, sa valeur est supérieure à 1, et si l'exposition est liée négativement (effet protecteur) sa valeur est inférieure à 1. Une valeur de P est calculée par le test de Fischer.

Ce test statistique nous a permis d'investiguer la relation entre le nombre de copies (la variable) et les facteurs (l'exposition). Par exemple, pour un gène donné, nous avons analysé la relation entre la présence d'au moins une délétion par rapport à l'exposition ou non au tabac. Comme pour les deux précédents tests, un seuil de la valeur P été considéré en tenant compte d'une correction relative aux tests multiples.

Chapitre 6 : Résultats

I. Caractérisation des CNVs du gène *GSTM1*

A. Identification de la présence/absence du gène

La première étape de screening nous a permis de déterminer la présence ou l'absence du gène *GSTM1* dans 96,4% des individus issus de 200 familles trio analysées. Nous avons observé soit l'amplification de deux bandes de 230pb et de 153pb correspondant à l'amplification d'une séquence du gène et une séquence du contrôle interne, soit et une seule bande de 153pb correspondant au contrôle interne (Figure 41).

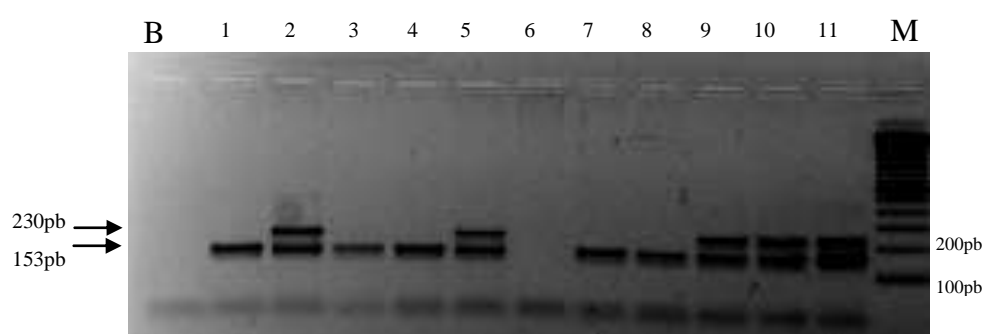


Figure 41 : Profil de migration de la PCR-présence/absence du gène *GSTM1* sur gel d'agarose pour 11 échantillons. Les puits 1, 3, 4, 7 et 8 correspondent aux échantillons homozygotes pour la délétion (-/-). Les puits 2, 5 et 9 à 11 représentent les échantillons ayant au moins une copie du gène (+/- ou +/+). Le puits 6 ne montre pas d'amplification. M : Marqueur 100pb Ladder (Invitrogen), B : Puits du contrôle négatif.

La moitié (50%) des échantillons testés ont au moins une copie du gène *GSTM1* (+/- et/ou +/+) montrant l'amplification des deux séquences du gène et du contrôle interne (exemples des échantillons 2 et 5, figure 41). L'autre moitié correspond aux échantillons homozygotes pour la délétion (-/-), avec une seule amplification du contrôle interne (exemple des échantillons 1 et 3, figure 41).

B. Identification du génotype

Les individus positifs pour le gène *GSTM1* ont été génotypés. Les résultats de cette PCR montrent trois profils de migration. Le premier profil de la région de délétion de 4,7kb correspond au génotype homozygote pour la délétion (-/-). Les deux amplifications de 4,7kb et 625pb correspondent à la région de délétion et à la séquence du gène, respectivement, indiquant un génotype hétérozygote pour la délétion. L'amplification seule de la séquence du

gène de 625pb correspond au génotype homozygote pour le gène (au moins une copie par chromosome) (Figure 42).

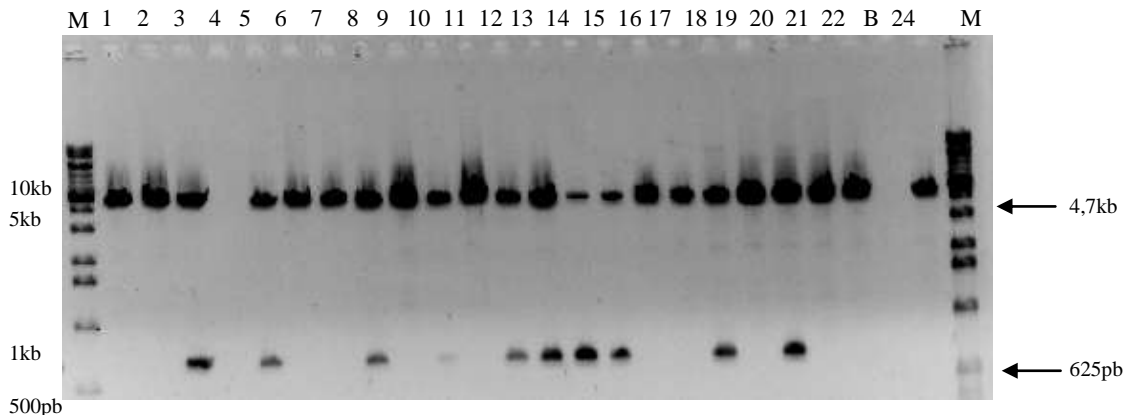


Figure 42 : Profil de migration des produits PCR issus de la séquence délétée (4,7kb) et de la séquence du gène (625pb) sur gel d'agarose pour 11 échantillons. Les puits 1, 2 et 6 par exemple correspondent à des échantillons homozygotes pour la délétion (-/-). Les puits 3, 5 et 8 sont des exemples d'échantillons hétérozygotes (+/-). Le puits 4 correspond à un échantillon qui ne présente pas d'amplification. M : Marqueur de taille moléculaire 1K ladder (Invitrogen), B : Puits du contrôle négatif.

Douze familles ayant au moins un membre ne montrant pas d'amplification, ont été exclues. La fréquence totale (résultats combiné avec ceux de la précédente PCR) du génotype (-/-) est de 54%, celle du génotype (+/-) est de 46% et aucun génotype homozygote (+/+) n'a été observé.

C. Quantification des CNVs

1. Quantification par PCR quantitative

Les CNVs des échantillons avec les génotypes +/-, ainsi que ceux n'ayant pas pu être amplifiés par PCR classique, ont été quantifiés par qPCR avec une première sonde (Hs02575461_cn). Tous les individus ayant un nombre de copies avec un indice de confiance inférieur à 90% ou non concordant avec les résultats des précédentes PCR ont été analysés avec une deuxième sonde du gène *GSTM* (Hs03352485_cn). La figure 43 ci-dessous montre le nombre de copies observé pour 18 individus.

Seize familles ont été exclues de l'analyse, 10 à cause d'absence d'amplification, 6 pour un indice de confiance inférieur à 90%, pour l'un des membres de la famille. Pour les individus génotypés, la fréquence du nombre de copies est de 95% pour 1copie et de 5% pour 2 copies. En comparant ces résultats à ceux issus de la précédente PCR, nous avons observé

que tous les échantillons ayant le génotype +/- ont un nombre de copies égale à deux. Nous avons donc posé l'hypothèse de la présence des 2 copies sur un même chromosome et la complète délétion du gène sur l'autre chromosome conduisant à un génotype 2/0.

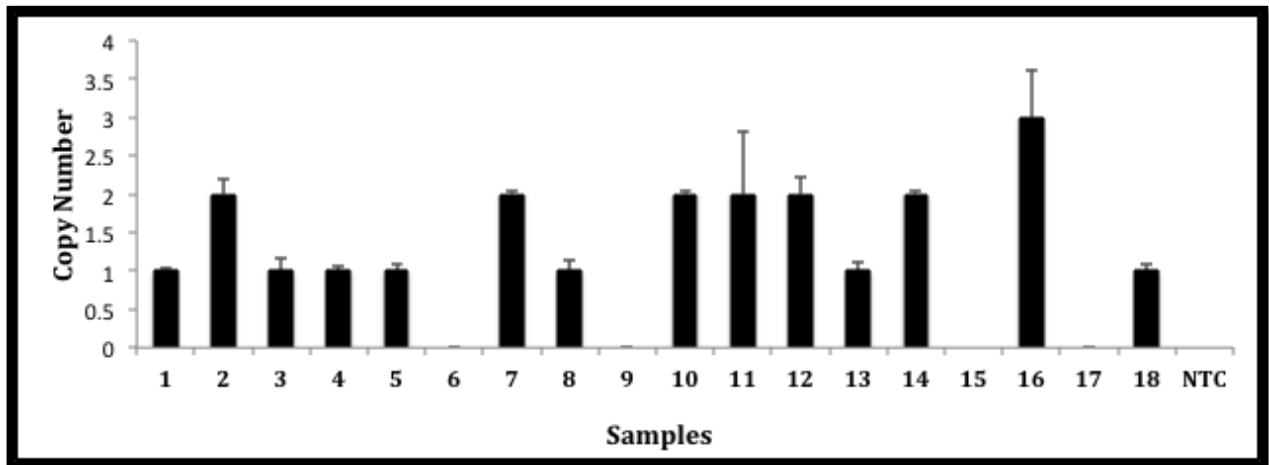


Figure 43 : Quantification des CNVs du gène GSTM1 par qPCR pour 18 échantillons. L'écart-Type est calculé par rapport au nombre de copies des triplicats. Les nombres de copies ont été estimés par le logiciel CopyCaller™. NTC : *No Template Control*.

2. Quantification par le système *QuantStudio 12K flex*

Un groupe de 26 échantillons, avec des nombres de copies allant de 0 à 2 obtenus en qPCR, ont été étudiés par dPCR avec la technologie *QuantStudio 12K flex system*. Après une étape de dilutions, si l'échantillon est encore trop concentré, il ne pourra pas être analysé. En effet, lorsque le taux d'évènements positifs devient largement supérieur à celui des évènements négatifs, la loi de Poisson ne peut plus s'appliquer et par conséquent il devient difficile d'estimer la concentration en copies/ μ l du gène cible et de la référence. Pour les échantillons ayant une bonne répartition des évènements positifs et négatifs sur la puce, le nombre de ces évènements a pu être calculé, ce qui a permis l'estimation de la concentration (en copies/ μ l) (Figure 44).

La figure 44a montre les évènements positifs et négatifs pour deux échantillons, l'échantillon 1 avec une copie et l'échantillon 2 avec deux copies. L'échantillon 1 montre bien un nombre d'évènements positifs inférieur à celui du deuxième échantillon (Figure 44a). La concentration en copies/ μ l du gène confirme cette différence entre les deux échantillons (Figure 44b).

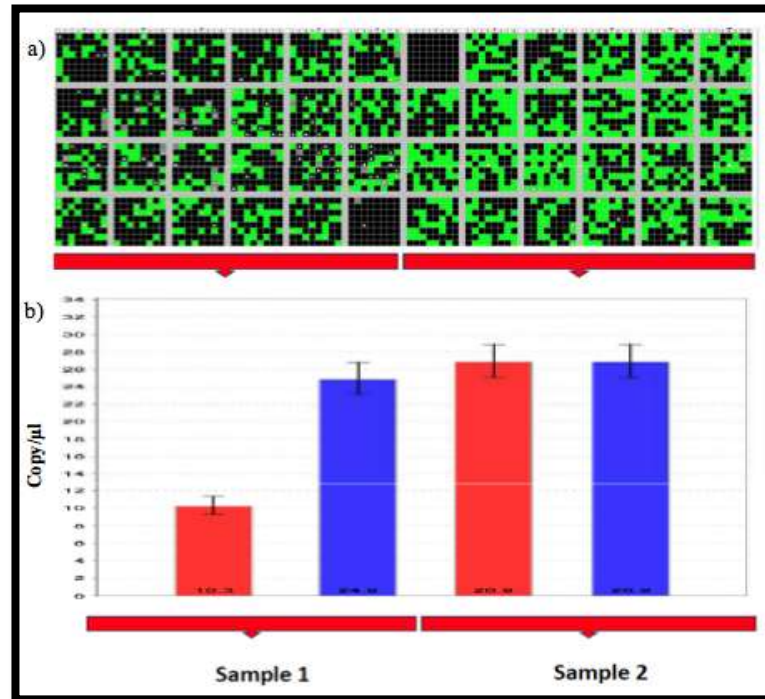


Figure 44 : Exemples de résultats obtenus par la dPCR pour deux échantillons avec un nombre de copies différents. a) Dénombrement des évènements positifs (en vert) et négatifs (en noir). b) Estimation de la concentration en copies/ μ l pour le gène cible (en rouge) et le gène de référence (en bleu) pour les deux échantillons.

Trois échantillons trop concentrés n'ont pas pu être analysés car la puce était saturée en évènements positifs. Pour les échantillons testés, les résultats en nombre de copies ne sont pas tous concordants avec ceux issus de la qPCR. La figure 45 illustre cette comparaison pour 11 échantillons. Les résultats sont discordants pour les échantillons 5 et 26.

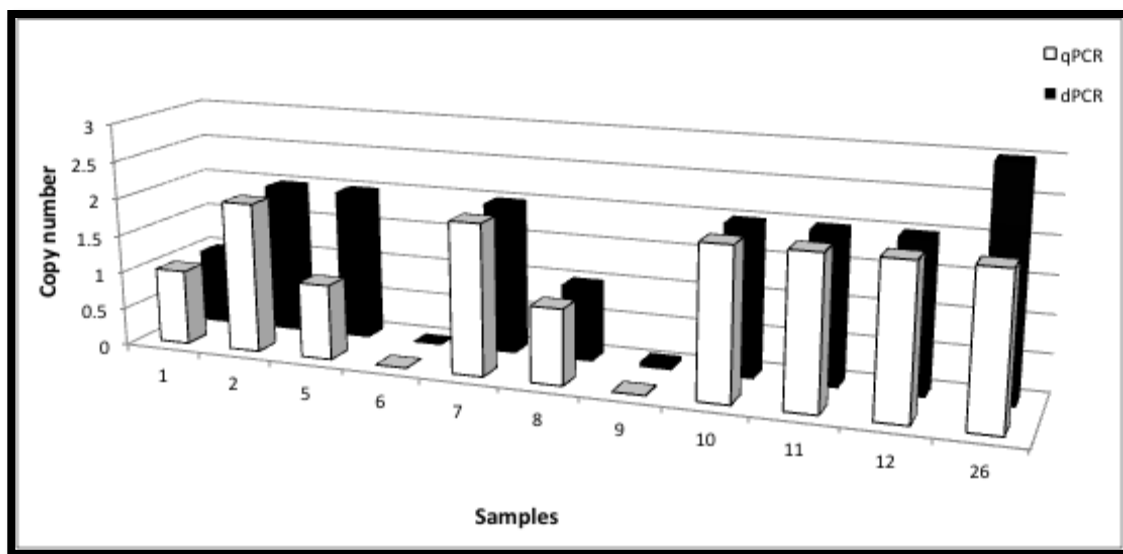


Figure 45 : Distribution du nombre de copies obtenus par qPCR et dPCR. Les 11 échantillons représentés ont été analysés à la fois par qPCR et dPCR.

D. Mise en évidence de la duplication en tandem

1. Mise en évidence par la PCR *Long Range*

La recherche de cette duplication en tandem a été réalisée sur tous les individus du set ayant un nombre de copies égal à 2. Sur 55 échantillons testés, seul 1 échantillon (le numéro 14, dont le nombre de copies identifié par qPCR est illustré dans la figure 42) montre une amplification d'une séquence avec la taille attendue (17,5kb, Figure 46).

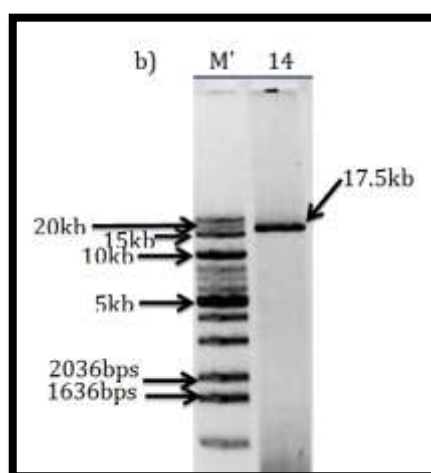


Figure 46 : Profil électrophorétique de la longue PCR sur gel d'agarose 0,6% obtenu pour l'échantillon 14. M: Lambda/HindIII digest DNA ladder, 100ng. M': 1KB DNA Extension Ladder.

Ce résultat indique la présence d'une duplication en tandem du gène *GSTM1* sur le même chromosome. Les autres échantillons testés ont donc un résultat discordant avec celui obtenu en combinant l'analyse de la PCR de génotype et de la qPCR.

2. Mise en évidence par la PCR QX200 *Droplet Digital*

a) Résultats des tests de validation

(1) Test d'interférence

Le test a montré une mauvaise séparation entre les événements positifs du gène cible et du gène *RNase P*, ce qui indique une interférence de la sonde de *RNase P* avec la sonde de *GSTM1* (Figure 47a). Nous avons alors testé une autre sonde d'un second gène de référence, le *RPP30* (Hindson, Ness et al. 2011), selon les mêmes conditions que pour la sonde *RNase P*.

Ce test a montré une nette séparation entre les *clusters* de *GSTM1* de *RPP30*, ce qui indique qu'il n'y a pas d'interférence entre ces deux sondes (Figure 47b).

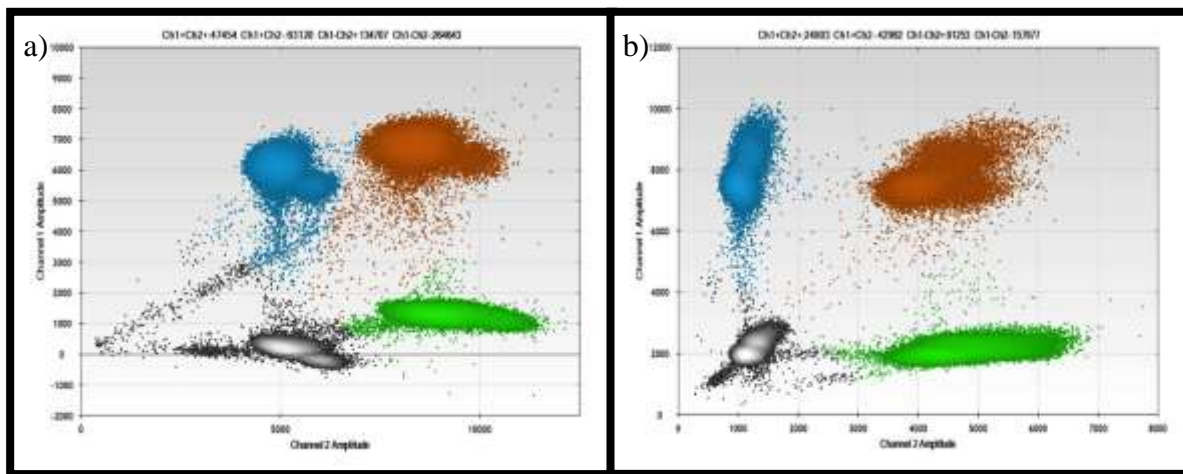


Figure 47 : Tests d'interférence des sondes de *GSTM1/RNase P* et *GSTM1/RPP30*. Les points bleus : événements positifs pour la sonde du gène cible. Les points verts : événements positifs pour la sonde de référence RNase P (a) et RPP30 (b). Les points oranges : événements positifs pour le gène cible et référence (RNase P et RPP30, respectivement). Les points noirs: événements négatifs.

(2) Test de reproductibilité

Le test de reproductibilité en intra et inter-plaques ne montre pas de différence significative dans la quantification du nombre de copies (Figure 48), ce qui montre qu'il n'est pas nécessaire de faire les réactions en plusieurs réplicats.

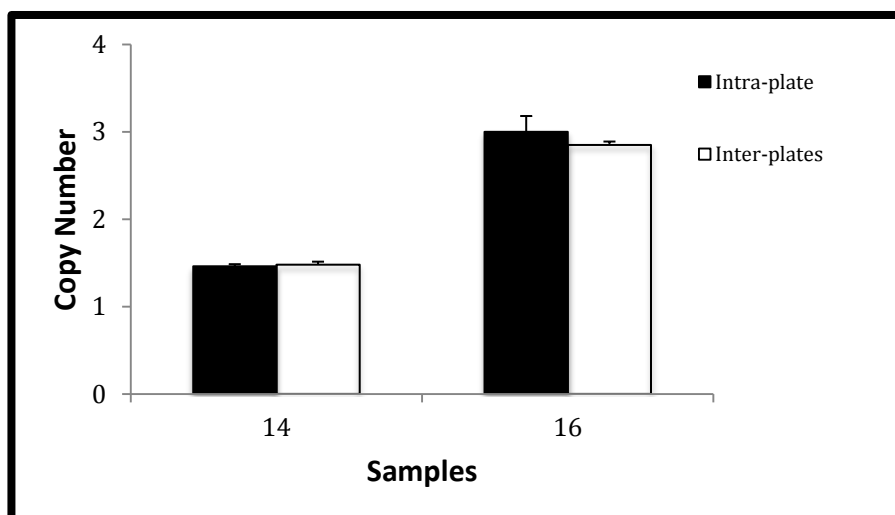


Figure 48 : Test de reproductibilité intra et inter-plaques. Le test en intra-plaque a été réalisé sur deux échantillons (14 et 16) en triplicat dans la même plaque alors que le test en inter-plaques a été réalisé une seule fois sur les mêmes échantillons dans deux plaques différentes.

(3) Test de digestion

Le test de digestion a été réalisé sur l'échantillon 14 présentant un génotype 2/0 au vu des résultats précédents. La valeur de nombre de copies obtenu par ddPCR est de 1,44. Nous avons tout d'abord testé l'enzyme EcoRV en prédigestion avec un gradient d'enzyme de 1 à 20U. Les résultats ne montrent aucun changement avec la digestion. Nous avons alors testé l'enzyme FoKI dans les mêmes conditions et le nombre de copies du gène de l'échantillon 14 a augmenté à 1,75 copies (1U), indiquant une digestion partielle de l'ADN. Enfin, avec l'enzyme HaeIII, nous avons observé une augmentation à 1,9 copies (Figure 49) indiquant une digestion presque complète de l'ADN et de la région contenant le gène. Cela souligne que la ddPCR ne peut pas mettre en évidence la duplication du gène *GSTM1* en tandem sur un même chromosome sans digestion. Ce résultat confirme celui obtenu par la *Long Range PCR*.

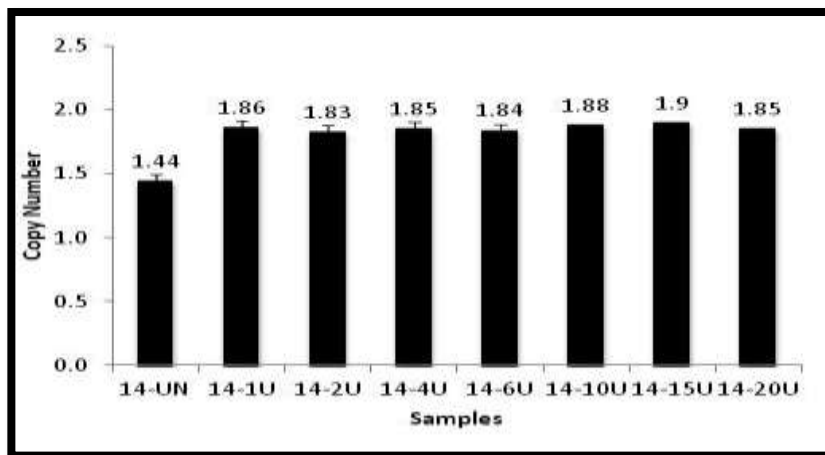


Figure 49: Gradient de HaeIII permettant la mise en évidence de la duplication en tandem. L'échantillon 14 a été digéré avec différentes quantités de HaeIII pour identifier les copies en tandem. L'écart-type a été calculé par la loi de Poisson pour une réaction partitionnée en environ 16 000 évènements. Le nombre de copies est estimé par le logiciel QuantaSoft™. Und: non digéré.

b) Caractérisation des CNVs

Nous avons fait une première quantification des CNVs des échantillons ayant au moins deux copies en qPCR ainsi que ceux dont les données PCR et qPCR étaient manquantes ou discordantes. Une seconde quantification a été réalisée avec une étape de digestion pour un sous-groupe d'échantillons afin de confirmer que les valeurs proches de 2 copies ne variaient pas avec la digestion (Figure 50).

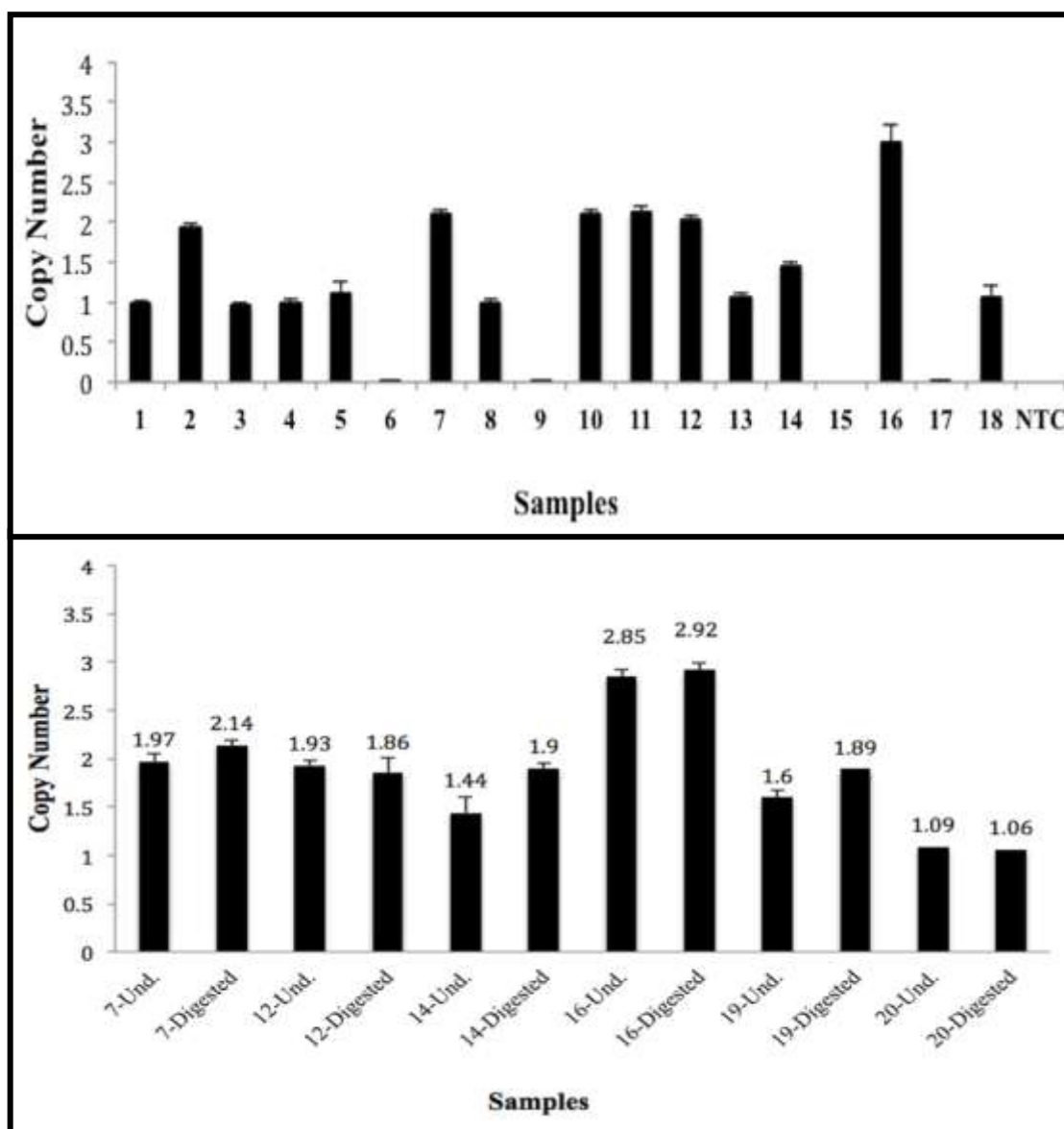


Figure 50: Résultats obtenus par ddPCR. A) Distribution des CNVS dans 18 échantillons, les mêmes qui ont été quantifiés par qPCR et ont été testés en triplicat. L'écart-type est calculé par rapport au nombre de copies de chaque triplicat. B) Comparaison des nombre de copies obtenus par ddPCR avec et sans digestion pour 6 échantillons. L'écart-type est calculé avec la loi de Poisson pour une réaction divisée en environ 16 000 évènements. Le nombre de copie est estimé par le logiciel QuantaSoft™. Und : ADN non digéré.

L'étape de digestion était nécessaire pour les deux échantillons 14 (précédemment décrit) et 19. L'ensemble des autres échantillons testés ont montré un nombre de copies proche à $\pm 0,25$ de la valeur 2 avec ou sans digestion. Nous avons donc considéré que le nombre de copies égal à 2, avec une variation de $\pm 0,25$ obtenue après une ddPCR sans digestion, était représentatif d'un génotype homozygote 1/1 (exemple des échantillons 7 et 12 de la figure 45).

3. Comparaison entre qPCR, dPCR et ddPCR

Les résultats de ddPCR avec et sans digestion ont permis de mettre en évidence la duplication en tandem du gène *GSTMI* (++/-) dans deux échantillons. Tous les autres échantillons avec deux copies sont homozygotes pour le gène (+/+), ce qui ne concorde pas avec les résultats obtenus après la PCR de génotype qui ne montre aucun génotype homozygote. De plus, l'estimation du nombre de copies par qPCR n'a pas permis de mettre en évidence la particularité des échantillons 14 et 19 et de conclure quant à la localisation des copies de *GSTMI* sur les chromosomes homologues (Figure 49). Par ailleurs, les CNVs quantifiés par la ddPCR montrent une barre d'erreur plus petite par rapport à ceux quantifiés par qPCR (Figure 43 et 50). En comparant les résultats de la ddPCR à ceux de sa variante, la dPCR, nous avons observé 4 discordances. En effet, le nombre de copie de l'échantillon 14 n'a pas pu être estimé à cause de la saturation de la puce, de même pour l'échantillon 19. L'échantillon 26 montre 3 copies avec la dPCR, alors qu'il est estimé à 2 copies avec et sans digestion en ddPCR. Enfin, l'échantillon 5 est estimé à 2 copies avec la dPCR, tandis qu'avec la ddPCR il est estimé à 1 copie (Figure 50). La figure 51 résume les nombres copies identifiés par chaque méthode.

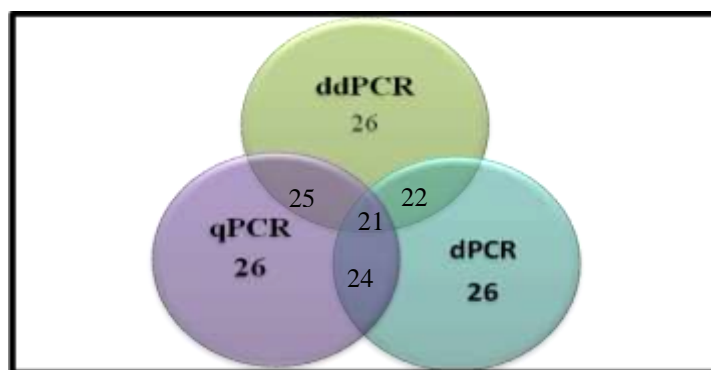


Figure 51: Diagramme de Venn indiquant le nombre d'échantillons ayant le même nombre de copies retrouvé par les trois méthodes. Le nombre total des échantillons dont les résultats sont comparés par les trois méthodes est 26.

E. Transmission des CNVs dans les familles

Nous avons combinés les résultats issus des méthodes de PCR présence/absence, qPCR et ddPCR, en retenant les CNs obtenus par ddPCR pour les résultats discordants entre les différentes méthodes. Ceci nous a permis de caractériser le génotype de 199 familles trio et ont montré une fréquence de 53,5% pour 0 copie, 37,5% pour 1 copie et 9% pour 2 copies. Des modèles de transmission ont été identifiés en utilisant le nombre de copie de chaque

membre de la famille (colonne 1) ainsi que leurs génotypes de nombre de copies (colonne 2). Les dix modèles mis en évidence sont détaillés dans le tableau 15.

Tableau 15: Modèles de transmission des CNVs dans 199 familles trio

Modèles familiaux	Génotypes	n	Fréquences
1-0x0=0	0/0x0/0=0/0	59	29,6%
2-0x1=0	0/0x1/0=0/0	38	19,1%
3-0x1=1	0/0x1/0=0/1	36	18,1%
4-1x1=0	1/0x1/0=0/0	8	4%
5-1x1=1	1/0x1/0=1/0	11	5,5%
6-1x1=2	1/0x1/0=1/1	10	5%
7-1x2=1	1/0x1/1=1/0	7	3,6%
8-1x2=2	1/0x1/1=1/1	7	3,6%
9-0x2=1	0/0x1/1=0/1	22	11%
10-0x2=2	0/0x0/2=0/2	1	0,3%

0= zéro copie, 1= une copie, 2= 2 copies du gène *GSTM1*. Une des 200 familles a été exclue pour un individu de la famille présentant une valeur de CN supérieure ou égale à un nombre entier $\pm 0,25$

Sur ces dix modèles, 9 montrent une transmission de l'allèle présentant la délétion du gène (0 copie), et/ou de l'allèle avec une copie. La famille du modèle 10 montre que la duplication est transmise d'un parent à sa descendance. De plus, lorsqu'on considère une combinaison de génotypes parentaux donnée, nous observons que les différents génotypes ont une fréquence comparable chez les cas. Par exemple, en analysant les génotypes des cas-index des modèles 2 et 3 (de même pour les modèles 4,5 et 6 et les modèles 7 et 8), la fréquence des individus ayant le génotype 0/0 et 0/1 est de 19,1% et 18,1%, respectivement. Ces observations indiqueraient qu'il n'y a pas de transmission préférentielle des CNVs du gène *GSTM1*.

F. Analyse familiale de l'association *GSTM1* et PR

Pour l'étude de l'association entre les allèles du gène *GSTM1* et la PR, nous avons considéré l'allèle (-) correspondant à la délétion du gène *GSTM1*, et l'allèle (+) correspondant à une copie du gène, permettant ainsi d'obtenir les trois génotypes (-/-, -/+ et +/+). Deux familles ont été exclues de l'analyse, la première ayant deux membres avec une duplication en tandem du gène, et la deuxième ayant un membre sans donnée de génotypage. Nous avons

investigué le TDT de chaque allèle dans 198 trios. Après correction de Bonferroni, nous n'avons pas observé de transmission préférentielle de l'allèle délété de *GSTM1* dans l'échantillon global, comme suggéré par le tableau 15. Le test du GRR n'a révélé aucune association entre la PR et les génotypes du gène (Tableau 16).

Tableau 16: Résultats du Test de Déséquilibre de Transmission (TDT) et du Risque Relatif des Génotypes (GRR) dans 198 familles trio de patients PR

Echantillons	Allèles	Nombre de famille	Familles informatives	P _{TDT**}	P _{GRR}
Total	+	198	118	0,68	0,837
	-			-	0,862
Age >40*	+	52	31	-	0,041
	-			0,02	0,2
Age <40*	+	146	87	0,07	0,151
	-			-	0,297
ACPA -	+	43	21	0,33	0,637
	-			-	0,739
ACPA +	+	148	95	0,85	0,816
	-			-	1
FR-	+	41	19	-	0,617
	-			0,29	0,059
FR +	+	157	99	0,36	0,651
	-			-	0,239
Erosion -	+	29	12	0,1	0,157
	-			-	0,705
Erosion +	+	169	106	-	0,829
	-			0,85	1
Nodules -	+	146	86	-	0,339
	-			0,33	0,394
Nodules +	+	52	32	0,016	0,041
	-			-	0,132
Exposition au tabac -	+	86	55	-	0,88
	-			0,9	1
Exposition au tabac +	+	97	22	0,41	0,639
	-			-	1
HLA SE -	+	44	61	0,9	0,513
	-			-	0,7
HLA SE +	+	154	26	0,73	0,907
	-			-	0,695

*: Age de début de la maladie ; **: Valeur de P issue du TDT ; "-" : Transmission non préférentielle. **Informations en Gras** : résultat inférieur ou égal au seuil de 5%. Deux des 200 familles ont été exclues, une avec des individus ayant la duplication en tandem et une pour un individu de la famille présentant une valeur de CN supérieure ou égale à un nombre entier $\pm 0,25$

Après stratification des familles selon les données cliniques, biologiques, environnementales et génétiques des cas-index, nous avons observé une transmission préférentielle de la délétion du gène avec les tests du TDT dans le sous-groupe des familles avec des patients ayant un âge supérieur à 40 ans ($P_{\text{TDT}} = 0,02$), mais le GRR montre une association avec l'allèle (+) ($P_{\text{GRR}} = 0,041$). Une autre association entre l'allèle présentant le gène et la présence de nodules, ($p_{\text{TDT}} = 0,016$ et $p_{\text{GRR}} = 0,041$) a été observée (Tableau 16). Néanmoins, après correction de Bonferroni, ces résultats ne sont plus significatifs.

G. Relation entre CNVs du gène et facteurs de la PR

Nous avons investigué la relation entre les facteurs associés à la PR et les CNVs du gène *GSTM1* pour 200 cas-index, par un calcul d'Odds Ratio. Nous avons considéré un modèle permettant d'analyser l'association de la délétion complète du gène par rapport aux copies 1 et 2, en fonction des différents facteurs associés à la PR.

Nous avons observé une association significative entre la délétion complète du gène et l'âge d'apparition de la maladie ($P = 0,015$). En effet le risque diminue lorsque le gène est délété chez les patients ayant un âge inférieur à 40 ans par rapport au groupe de patients ayant un âge supérieur à 40 ans. Une association a été observée également dans le sous-groupe de patients présentant des nodules ($p = 0,05$). Néanmoins, après correction de Bonferroni, ces associations ne sont plus significatives, mais confirment les résultats obtenus par les tests de TDT et GRR (Tableau 17).

Tableau 17: Analyse de la relation entre les CNVs du gène *GSTMI* et les facteurs associés à la PR (200 cas index)

Caractéristiques des patients	Nombre de copies		OR (IC)	P
	0	≠0		
Hommes	11	11	1,07 (0,46-2,7)	0,5
Femmes	94	84		
Age>40*	35	17	0,43 (0,2-0,8)	0,015
Age<40*	70	78		
FR+	83	75	0,99 (0,5-1,96)	1
FR-	22	20		
Erosion+	91	79	1,41 (0,65-3,12)	0,4
Erosion-	14	16		
Nodules+	21	31	0,51 (0,27-0,99)	0,05
Nodules-	84	64		
ACPA+	78	71	1,18 (0,6-2,33)	0,7
ACPA-	21	23		
Exposition au tabac +	49	49	0,84 (0,47-1,45)	0,6
Exposition au tabac -	48	39		
HLA-SE+	83	73	1,21 (0,57-2,19)	0,8
HLA-SE-	22	22		

*: Age de début de la maladie. **Information en Gras** : résultat inférieur ou égal au seuil de 5%.

II. Caractérisation des CNVs du gène *GSTT1*

A. Identification de la présence/absence

Nous avons discriminé la présence/absence du gène grâce à l'amplification des bandes de *GSTT1* et de la β -globine. La bande supérieure à la bande 400pb du marqueur de taille correspond à la séquence spécifique du gène *GSTT1* de 480pb et celle supérieure à la bande 200pb du marqueur de taille est la bande du contrôle interne de 266pb (Figure 52).

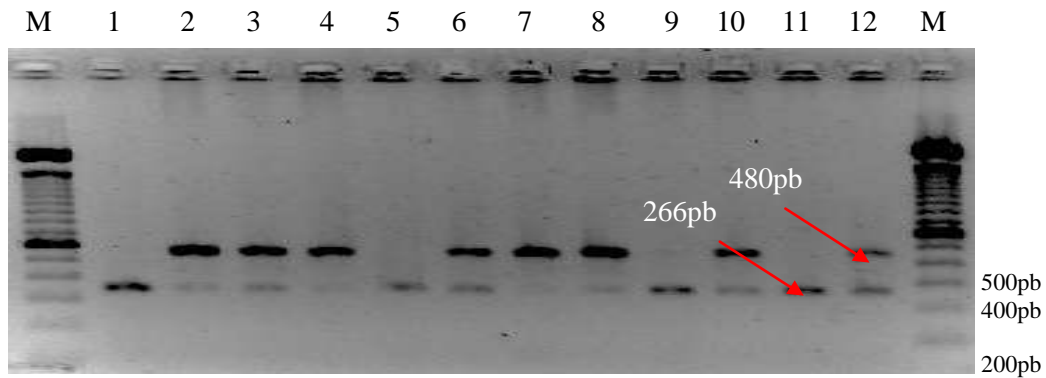


Figure 52 : Profil de migration des produits de PCR-Multiplex du gène *GSTT1* sur gel d'agarose pour 12 échantillons. Les échantillons 1, 5, 9 et 11 sont homozygotes pour la délétion (-/-), les autres échantillons ont au moins une copie du gène. M : 100pb Ladder (Invitrogen).

Sur les 200 familles étudiées, 188 familles ont pu être génotypés avec 14,2 % d'échantillons étant homozygotes pour la délétion (-/-) et 85,8% avec au moins une copie du gène *GSTT1* (+/- ou +/+).

B. Identification du génotype

Tous les échantillons ayant au moins une copie du gène *GSTT1* ou n'ayant pas pu être amplifiés lors de la précédente PCR, ont été génotypés. Les résultats de PCR montrent trois génotypes distincts. Le premier est un génotype homozygote pour la délétion de *GSTT1* (-/-) montrant une seule amplification de la séquence 3,1kb. Le deuxième génotype montre deux amplifications, l'une de 3,1kb et l'autre de 1,6kb correspondant respectivement à la séquence de la région délétée et à la séquence du gène, ce qui indique un génotype hétérozygote pour la délétion (+/-). Enfin, le troisième génotype homozygote pour le gène présente une seule amplification de la séquence du gène (1,6kb) (Figure 53).

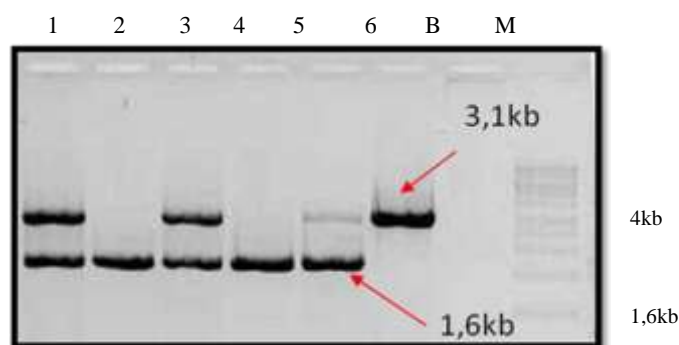


Figure 53: Profil de migration des amplifications de la région délétée et du gène *GSTT1* pour 6 échantillons. Les puits 1, 3 et 5 correspondent aux échantillons hétérozygote (+/-). Les puits 2 et 4

correspondent aux échantillons homozygotes pour le gène (+/+). L'échantillon du puits 6 est homozygote pour la délétion (-/-). M : 1kb Extension DNA Ladder, (Invitrogen). B : contrôle négatif.

Trois familles, ayant au moins un membre ne présentant pas d'amplification, ont été exclues de l'analyse. La fréquence des génotypes dans notre population est de 15,7% pour les homozygotes délétés (-/-), 69% pour les hétérozygotes (+/-) et 15,3% pour les homozygotes pour avec présence du gène (+/+).

C. Analyse familiale de l'association entre *GSTT1* et la PR

Comme pour *GSTM1*, nous avons considéré l'allèle (-) correspondant à la délétion du gène *GSTT1* et l'allèle (+) représentant la présence d'une copie du gène. Trois familles ont été exclues de l'analyse puisque le génotype de l'un des membres n'a pas pu être caractérisé. L'association a été investiguée par le test de TDT et par le GRR. Le TDT n'a pas révélé de transmission préférentielle dans la population totale. Par contre, le GRR a montré une association entre la présence d'au moins une copie du gène et la PR ($P=0,039$) mais, après correction de Bonferroni, cette dernière n'est plus significative. L'analyse après stratification selon les facteurs associés à la PR n'a pas montré de transmission préférentielle dans les sous-groupes. Il est cependant à noter qu'une association significative, après correction de Bonferroni, a été mise en évidence par le GRR entre la présence d'au moins une copie du gène et la présence de facteur rhumatoïde ($P=0,002$) (Tableau 18).

Tableau 18 : Résultats du Test de Déséquilibre de Transmission (TDT) et du Risque Relatif des Génotypes (GRR) 197 les familles trio de patients PR

Echantillons	Allèles	Nombre de famille	Familles informatives	P _{TDT**}	P _{GRR}
Total	+	197	178	0,76	0,039
	-			-	0,152
Age >40*	+	50	45	0,71	0,132
	-			-	0,527
Age <40*	+	147	133	0,88	0,116
	-			-	0,199
ACPA -	+	44	40	0,9	0,317
	-				0,346
ACPA +	+	147	133	0,83	0,086
	-			-	0,22
FR-	+	42	40	-	0,346
	-			0,611	1
FR +	+	155	138	0,53	0,002
	-			-	0,1
Erosion -	+	30	28	1	0,2
	-			1	0,157
Erosion +	+	167	150	0,739	0,09
	-			-	0,327
Nodules -	+	145	131	0,668	0,069
	-			-	0,274
Nodules +	+	52	47	-	0,317
	-			0,9	0,346
Exposition au tabac -	+	85	73	0,629	0,083
	-			-	0,433
Exposition au tabac +	+	98	92	-	0,57
	-			0,8	0,273
HLA SE -	+	43	35	1	0,317
	-			1	0,405
HLA SE +	+	154	143	0,734	0,069
	-			-	0,238

*: Age de début de la maladie ; **: Valeur de P issus du TDT ; "-": Transmission non préférentielle. **Information en Gras** : résultat inférieur ou égal au seuil de 5%. Trois des 200 familles ont été exclues puisque le génotype de l'un des membres n'a pas pu être caractérisé.

D. Relation entre les CNVs de *GSTT1* et les facteurs de la PR

Nous avons investigué l'association entre les CNVs du gène *GSTT1* les facteurs associés à la PR pour 199 patients. L'analyse a été réalisée dans les différents sous-groupes selon un modèle analysant l'association entre la délétion complète du gène par rapport à la présence d'au moins une copie du gène comme référence. L'analyse révèle une unique association entre la délétion complète du gène et la présence du facteur rhumatoïde (P=0,03), qui concorde avec l'association trouvée avec le GRR. Néanmoins, cette association n'est plus significative après la correction de Bonferroni (Tableau 19).

Tableau 19 : Analyse de la relation entre les CNVs du gène *GSTT1* et les facteurs associés à la PR (199 cas-index)

Caractéristiques des patients	Nombre de copies		OR (IC)	P
	-/-	+/+ et +/-		
Hommes	1	20	3,89 (0,5-30,15)	0,2
Femmes	29	149		
Age>40*	4	47	2,5 (0,82-7,56)	0,11
Age<40*	26	122		
FR+	19	138	0,38 (0,16-0,89)	0,03
FR-	11	31		
Erosion+	26	143	1,18 (0,38-3,66)	1
Erosion-	4	26		
Nodules+	5	47	0,54 (0,19-1,5)	0,2
Nodules-	25	122		
ACPA+	21	128	0,86 (0,34-2,1)	0,8
ACPA-	7	37		
Exposition au tabac +	15	83	0,86 (0,39-1,9)	0,8
Exposition au tabac -	15	72		
HLA-SE+	24	132	1,12 (0,42-2,9)	1
HLA-SE-	6	37		

*: Age de début de la maladie. **Information en Gras** : résultat inférieur ou égal au seuil de 5%. Un des 200 cas-index a été exclu puisque son génotype n'a pas pu être caractérisé.

III. Caractérisation des CNVs du gène *FCGR3B*

A. Quantification par qPCR

Les CNVs du gène *FCGR3B* ont été analysés par qPCR sur un premier lot de 56 familles. Comme pour *GSTM1*, nous avons sélectionné tous les échantillons ayant un indice de confiance $\geq 90\%$, ce qui exclut 19 familles ayant au moins un membre avec un nombre de copies attribué avec un indice de confiance $< 90\%$. Les CNVs quantifiés vont de 1 à 4 copies avec une fréquence de 15% pour une copie, 81% pour deux copies, 3% pour trois copies et 1% pour quatre copies (Figure 54).

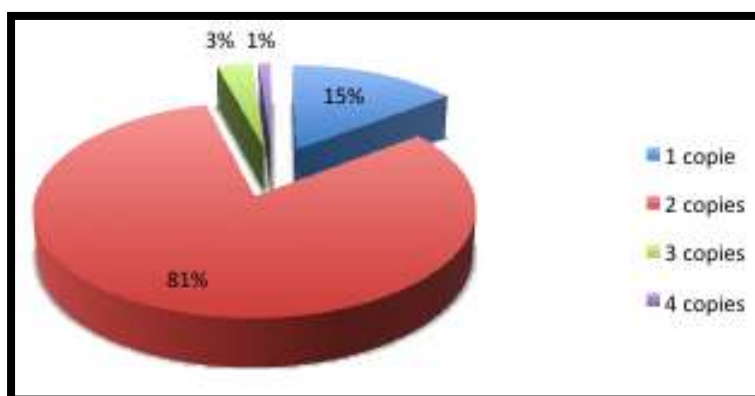


Figure 54 : Distribution des CNVs dans les 96 échantillons issus des familles trio.

Cinq échantillons n'ont pas eu d'amplification et ont été donc écartés de l'étude ainsi que leurs familles. Cependant, au vu des résultats présentant un nombre important d'échantillons avec des nombres de copies présentant un indice de confiance $< 90\%$, nous avons décidé de reprendre la quantification de ces CNVs par ddPCR sur l'ensemble des échantillons.

B. Quantification par ddPCR

1. Test d'interférence

Nous avons réalisé des tests en *simplex* pour la sonde du gène cible et le gène de référence, puis en *multiplex* avec les deux simultanément afin de vérifier la compatibilité de

ces deux sondes. La figure 55 montre une séparation correcte des *clusters* des événements correspondants à chaque sonde ce qui indique qu'il n'y a pas d'interférence entre elles.

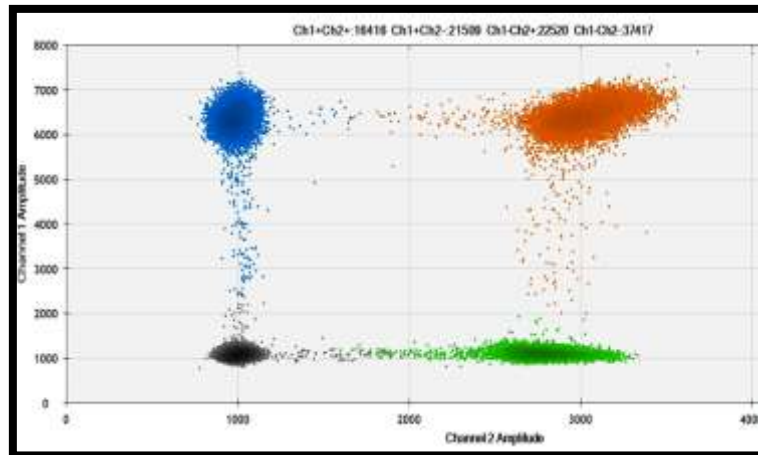


Figure 55 : Tests d'interférence des sondes de *FCGR3B/RPP30*. Les points bleus : les événements positifs pour la sonde du gène cible. Les points verts : événements positifs pour la sonde de référence RPP30. Les points oranges : événements positifs pour le gène cible et référence RPP30. Les points noirs: événements négatifs.

2. Test de reproductibilité

Deux contrôles positifs avec un nombre de copies connu ont été utilisés pour tester la reproductibilité des réactions de quantification. La quantification a été réalisée en duplicat dans la même plaque (test en intra-plaque) et une seule fois dans deux plaques différentes (test inter-plaque) (Figure 56).

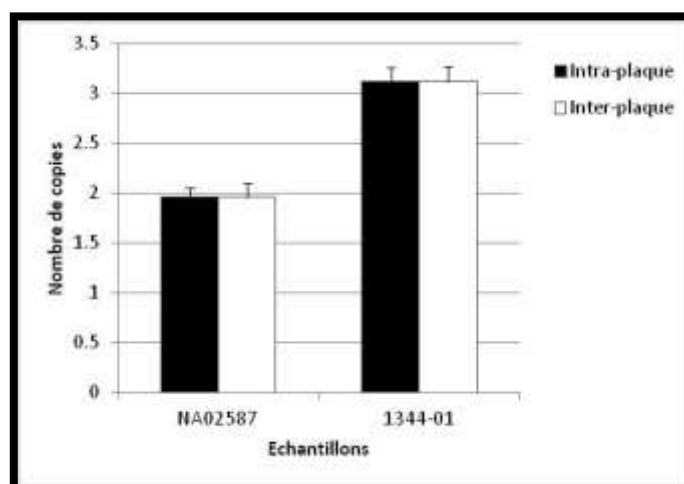


Figure 56: Résultats du test de reproductibilité intra et inter-plaque pour *FCGR3B*. Le test en intra-plaque a été réalisé sur deux échantillons en duplicat dans la même plaque alors que le test en inter-plaque a été réalisé sur les mêmes échantillons une seule fois dans deux plaques différentes.

Le résultat montre qu'il n'y a pas de différence significative entre le nombre de copies obtenu en réplicat dans la même plaque ou réalisé une seule fois sur deux plaques, ce qui indique que l'expérience est reproductible et que pour la suite de l'analyse, nous pouvons quantifier les CNVs de nos échantillons avec une seule réaction.

3. Tests de digestion

Au vu des premiers résultats obtenus sans digestion par ddPCR, 1,6% des échantillons montrent un nombre de copies égal à 1,6. Nous avons alors réalisé plusieurs tests de digestion avec différentes enzymes de restrictions en variant la quantité utilisée et le protocole de digestion (digestion avant ou pendant l'amplification). La figure 57 montre deux exemples de résultats obtenus par digestion avant l'amplification avec différentes quantité d'enzyme HaeIII ou FokI.

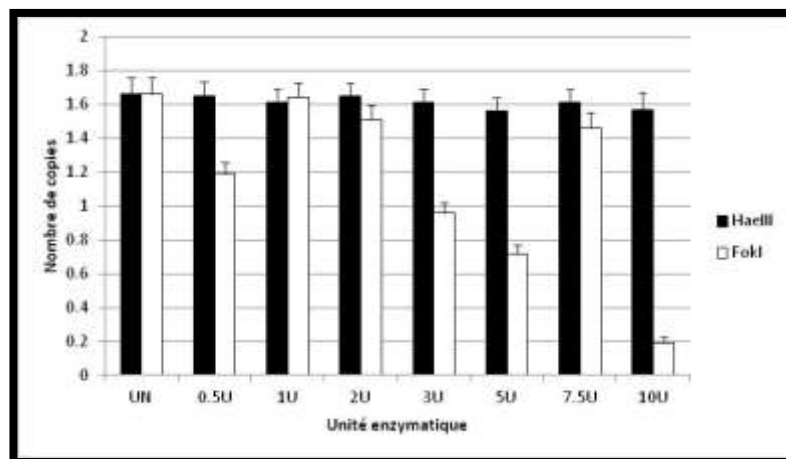


Figure 57: Tests de digestion avec un gradient d'unité enzymatique pour HaeIII et FokI réalisé sur un échantillon de 1,6 copies. L'écart-type est calculé avec la loi de Poisson pour une réaction divisée en environ 16 000 évènements. Le nombre de copie est estimé par le logiciel QuantaSoft™. Un : ADN non digéré. U : Unité d'enzyme.

Avec l'enzyme FokI, nous avons observé que le nombre de copies est très variable et tend à baisser, ce qui semble indiquer une dégradation partielle de l'ADN par cette enzyme. Avec HaeIII (ou avec d'autres enzymes comme BsRI, TspRI), aucun changement de la valeur du nombre de copies n'a été observé. Ce résultat reste le même avec le protocole de digestion avant ou pendant l'amplification. Pour les échantillons ayant une valeur autour de 1,6 copies, nous avons utilisé une autre source d'ADN et nous avons trouvé cette fois-ci une valeur proche de $2 \pm 0,25$, avec ou sans digestion. Ces résultats indiquent que la digestion, dans le cas

du gène *FCGR3B*, ne permet pas d'apporter une information quant à la localisation du gène sur les chromosomes homologues.

4. Caractérisation des CNVs

La quantification des CNVs a été réalisée sur les 200 familles trio. Douze familles ayant au moins un échantillon avec un écart au nombre de copies entier supérieure à $\pm 0,25$, ont été exclues de l'analyse. Les tests sans digestion montrent un nombre de copies allant de 0 à 4 (Figure 58).

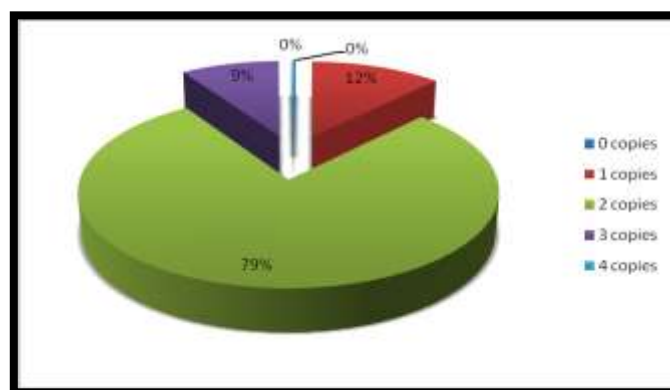


Figure 58 : Distribution des CNVs dans notre échantillon de familles.

La fréquence de ces CNVs est de 0,17% pour 0 copie, 12% pour 1 copie, 78,66% pour 2 copies, 9% pour 3 copies et 0,17% pour 4 copies. Les tests de digestion réalisés sur tous les individus ayant un nombre de copies supérieurs à 2 ne montrent pas de variations par rapport aux valeurs obtenues sans digestion. Ces résultats indiquent que dans le cas de copies de *FCGR3B* supérieures à 2 (3 et 4 copies), la ddPCR est capable de mettre en évidence les duplications en tandem sans avoir recours à la digestion. Cependant, l'hypothèse d'une duplication au niveau de deux régions chromosomiques différentes ne peut être éliminée.

C. Comparaison qPCR et ddPCR

Nous avons comparé la précision des deux méthodes sur un ensemble de 141 échantillons. Les fréquences des CNVs obtenues par qPCR et par la ddPCR pour les mêmes échantillons ne montrent pas de différence significative (établi par un test de Chi2 (X^2) de conformité à 1 ddl = 0,58). Néanmoins, le test de corrélation entre le nombre de copies trouvé

par qPCR et celui obtenu par ddPCR montre que plusieurs CNVs quantifiés par la qPCR ne sont pas en adéquation avec ceux obtenus par ddPCR (Figure 59).

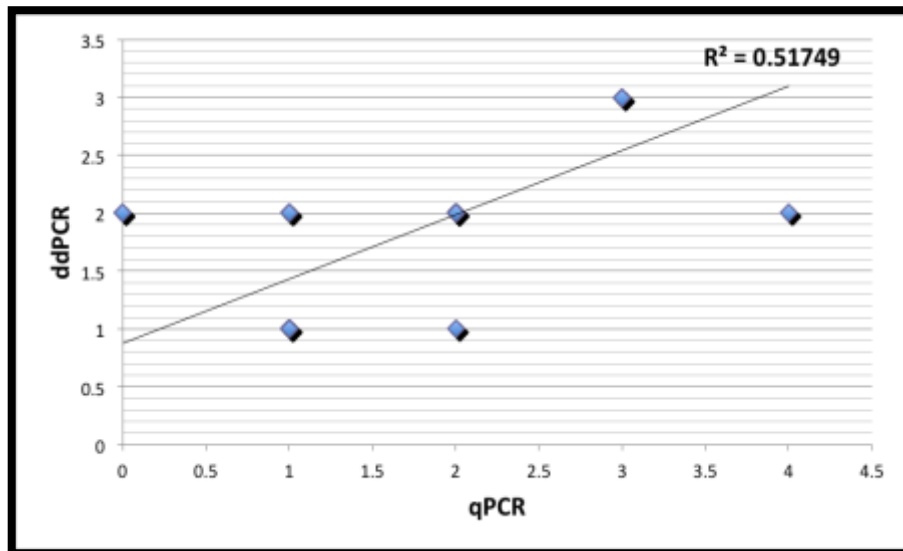


Figure 59 : Analyse de la corrélation entre les CNVs obtenus par qPCR et ddPCR. Les points bleus correspondent au nombre de copies trouvés par qPCR et ddPCR. Le R^2 indique qu'il y a 51% de corrélation entre les valeurs obtenues par les deux techniques.

D. Transmission des CNVs dans les familles

Puisque la ddPCR n'a pas permis d'identifier le génotype des échantillons ayant 2 copies, nous avons utilisé le contexte familial pour distinguer les différents génotypes. Onze modèles de familles ont été ainsi décrits. Aucun modèle ne permet de mettre en évidence un génotype ayant deux copies sur un même chromosome et une délétion sur l'autre (2/0). Par contre, les modèles 2, 5, 10 et 11 montrent que le génotype des échantillons ayant 2 copies ne pouvait être que 1/1 (Tableau 20). Nous avons donc considéré, pour la suite de l'analyse, que le génotype des copies 2 était 1/1. Dès lors, nous avons pu étudier la transmission des CNVs dans chaque famille. L'analyse de la transmission montre un événement de recombinaison *de novo* dans trois familles (modèles 2 et 10), induisant la délétion d'une copie du gène *FCGR3B*. Pour les autres modèles, nous avons observé qu'il n'y avait pas de transmission préférentielle entre le nombre de copies 0 et 1 (pour les modèles 3 et 4) et entre le nombre de copies 1 et 2 (pour les modèles 6 et 7).

Tableau 20 : Transmission des CNVs du gène *FCGR3B* dans 188 familles trio

Modèles familiaux	Génotypes	n	Hypothèse
1-2X2=2	1/1X1/1=1/1	114	
2-2X2=1	1/1X1/1=1/0	2	Délétion <i>de novo</i>
3-1X2=1	1/0X1/1=1/0	17	
4-1X2=2	1/0X1/1=1/1	18	
5-1X1=2	1/0X1/0=1/1	2	
6-2X3=3	1/1X1/2=1/2	15	
7-2X3=2	1/1X1/2=1/1	10	
8-1X3=2	1/0X1/2=1/1	6	
9-1X3=3	1/0X1/2=1/2	2	
10-4X2=2	2/2X1/1=1/1	1	Délétion <i>de novo</i>
11-0X2=1	0/0X1/1=1/0	1	

Douze des 200 familles ont été exclues puisqu'un individu de la famille présente une valeur de CN supérieure ou égale à un nombre entier $\pm 0,25$

E. Analyse familiale de l'association entre *FCGR3B* et la PR

Pour l'analyse de l'association des CNVs du gène *FCGR3B* dans la PR, nous avons considéré l'allèle (+) correspondant à la présence d'au moins une copie du gène et l'allèle (-) correspondant à la délétion du gène, permettant ainsi de considérer les 2, 3 et 4 copies ayant le génotype (+/+), le nombre de copie 1 ayant le génotype (-/+) et le 0 copie comme ayant le génotype (-/-). Ce modèle d'analyse nous permettra d'étudier l'association de la délétion du gène dans la PR. En effet, le nombre de copies inférieur à 2 a été décrit dans la littérature comme étant associé à la PR. Pour ce faire, nous avons réalisé les tests TDT et GRR dans 185 familles trio, en ne tenant pas compte des familles présentant une délétion *de novo*. Les tests TDT et GRR ne montrent pas d'association significative dans la population étudiée. Après stratification, une tendance non significative a été observée avec le TDT et le GRR dans le sous-groupe de patients ACPA séropositifs ($p=0,08$ et $p=0,09$ pour le TDT et GRR, respectivement) et une transmission préférentielle de l'allèle ayant le gène dans le sous-groupe de patients FR séropositifs ($P_{TDT}=0,08$) et dans le sous-groupe de cas-index positifs pour au moins un allèle (SE) du gène *HLA-DRB1* ($P_{TDT}=0,06$) (Tableau 21).

Tableau 21: Résultats du Test de Déséquilibre de Transmission (TDT) et du Risque Relatif des Génotypes (GRR) pour le gène *FCGR3B* dans 185 familles trio de patients PR

Echantillons	Allèles	Nombre de famille	Familles informatives	P _{TDT**}	P _{GRR}
Total	+	186	45	0,10	
	-			-	0,17
Age >40*	+	49	37	0,14	
	-			-	
Age <40*	+	139			
	-				0,23
ACPA -	+	40			0,65
	-				0,65
ACPA +	+	141	39	0,08	
	-			-	0,09
FR-	+	39			
	-				1
FR +	+	149	39	0,08	
	-			-	0,13
Erosion -	+	28			
	-				0,37
Erosion +	+	160	41	0,12	
	-			-	0,26
Nodules -	+	138	33	0,12	
	-			-	0,2
Nodules +	+	50	12	0,5	
	-			-	0,5
Exposition au tabac -	+	87	19	0,25	
	-			-	0,46
Exposition au tabac +	+	92	16	0,22	
	-			-	0,2
HLA SE -	+	41	22	1	
	-			1	1
HLA SE +	+	147	23	0,06	
	-			-	0,11

" " : Nombre de familles insuffisant pour le calcul statistique. *: Age de début de la maladie ; **: Valeur de P issue du TDT ; "- " : Transmission non préférentielle. Quinze des 200 familles ont été exclues puisqu'un individu de la famille présente une valeur de CN supérieure ou égale à un nombre entier à $\pm 0,25$ ou présente une délétion *de novo*.

F. Relation entre les CNVs de *FCGR3B* et les facteurs de la PR

La relation entre les variants de copies du gène *FCGR3B* et les caractéristiques cliniques, biologiques, génétiques et environnementaux de 200 patients a été investiguée.

Trois individus ayant des CNVs avec une variation supérieure de ± 0.25 n'ont pas été inclus dans l'étude. L'analyse est réalisée en prenant le nombre de copies 2 comme référence par rapport à laquelle l'association entre les CNVs et la présence de chaque facteur associé à la PR a été testée (Tableau 22).

Tableau 22: Analyse de la relation entre les CNVs du gène *FCGR3B* et les facteurs associés à la PR chez 197 cas-index

CNVs du gène <i>FCGR3B</i>	Caractéristiques des patients		OR (IC)	P
	Hommes	Femmes		
1	3	21	0.9 (0.24-3.34)	1
2	18	139		
3	1	15	1.94 (0.24-15.5)	0.7
	Age>40	Age<40		
1	6	18	1.03 (0.38-2.78)	1
2	40	116		
3	5	11	0.75 (0.24-2.31)	0.7
	FR-	FR+		
1	4	20	1.28 (0.4-4)	0.8
2	32	125		
3	4	12	0.76 (0.23-2.54)	0.7
	Erosion-	Erosion+		
1	2	22	2.12 (0.48-9.85)	0.4
2	26	131		
3	1	15	2.97 (0.37-23.53)	0.5
	Nodules-	Nodules+		
1	17	7	1.05 (0.41-2.75)	1
2	113	44		
3	15	1	0.17 (0.02-1.35)	0.07
	ACPA-	ACPA+		
1	5	17	0.97 (0.33-2.82)	1
2	34	119		
3	4	11	0.78 (0.23-2.62)	0.7
	Exposition au tabac -	Exposition au tabac +		
1	8	14	1.61 (0.63-4.07)	0.4
2	70	76		
3	8	6	0.69 (0.22-2.09)	0.5
	HLA-SE-	HLA-SE+		
1	7	17	0.62 (0.23-1.62)	0.4
2	32	125		
3	4	12	0.76 (0.23-2.54)	0.7

*: Age de début de la maladie. Trois des 200 cas-index ont été exclues puisqu'ils présentent une valeur de CN supérieure ou égale à un nombre entier ± 0.25 ou présente une délétion *de novo*.

Aucune association significative n'a été observée, seulement une tendance non significative entre la présence de 3 copies du gène et la présence de nodules chez les patients ($P=0,07$).

IV. Caractérisation des CNVs du gène *CCL3L1*

A. Mise au point par qPCR

En tenant compte de la taille de la séquence (184pb) et de la composition en nucléotides (57% de CG) du produit PCR, nous avons donc estimé la T_m à environ 84,4°C. Les courbes de dissociation obtenues avec les différentes températures d'hybridation des amorces montrent toutes un pic correspondant à la température de 84,4°C, mais avec des intensités de fluorescence variables. Ces résultats indiquent tout d'abord que les amorces sont bien spécifiques au gène *CCL3L1* et qu'il n'y a pas d'amplification non spécifique (Figure 60A). La température d'hybridation induisant l'intensité de fluorescence la plus élevée correspond à 60°C (Figure 60B). Nous avons donc poursuivi la caractérisation des CNVs par ddPCR avec ces conditions.

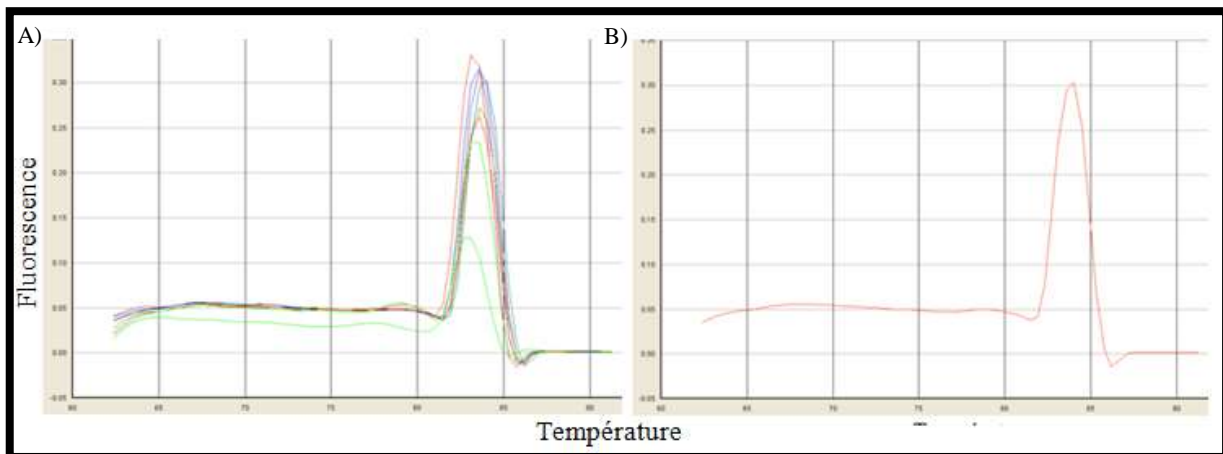


Figure 60 : Courbes de dissociation des amorces pour *CCL3L1* à différentes températures (entre 66°C et 56°C) d'hybridation (a) et à 60°C (b).

B. Caractérisation par ddPCR

1. Mise au point par ddPCR

Afin de s'assurer que l'*assay* de *CCL3L1* que nous avons choisi est adapté à la ddPCR, nous avons réalisé deux tests, un en utilisant l'EVA Green comme agent intercalant et l'autre avec la sonde spécifique du gène cible et celle du gène de référence *RPP30*, en utilisant un gradient de température d'hybridation. Le premier test confirme le résultat obtenu par qPCR puisque la séparation des évènements montre un seul *cluster* d'évènements positifs. Cela signifie que les amorces sont bien spécifiques du gène, et que la séparation est optimale à la température d'hybridation égale à 60°C (Figure 61a). Le 2^{ème} test confirme que la sonde que nous avons synthétisée est spécifique à *CCL3L1* uniquement (pas de double *cluster*) et que 60°C est la température optimale d'hybridation (Figure 61b).

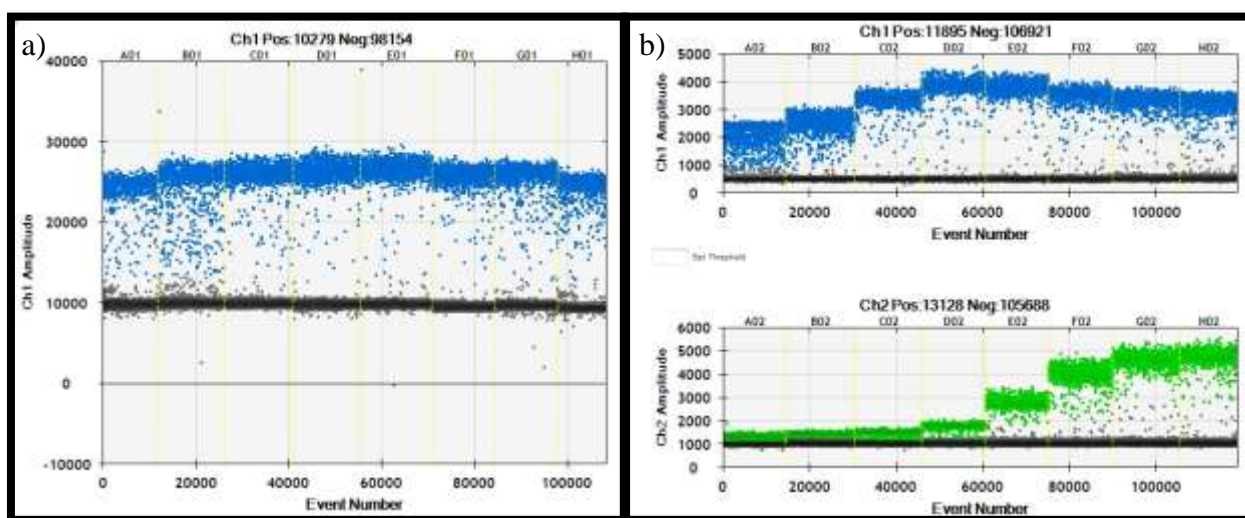


Figure 61: Séparation des *plots* positifs et négatifs pour le test en EVA Green (a) et le test avec les sondes (b). Les points bleus représentent les évènements positifs pour le gène *CCL3L1* avec l'EVA Green (a) et avec la sonde synthétisée (b). Les points verts correspondent aux évènements positifs pour le gène de référence (*RPP30*). Les points noirs sont les évènements négatifs. Les puits de A à H correspondent chacun à une température d'hybridation.

2. Test d'interférence

Comme pour les précédents gènes, nous avons investigué la présence ou pas d'interférence entre la sonde spécifique de *CCL3L1* et celle spécifique de *RPP30*. La figure 61

ne montre pas d'interférence entre les deux sondes puisque les évènements positifs correspondants sont bien séparés dans des *clusters* distincts (Figure 62).

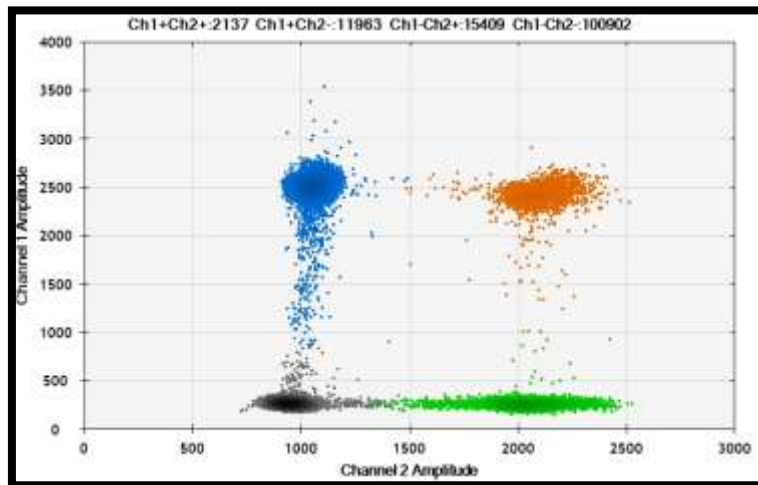


Figure 62: Tests d'interférence des sondes de *CCL3L1/RPP30*. Les points bleus : les évènements positifs pour la sonde du gène cible. Les points verts : évènements positifs pour la sonde de référence RPP30. Les points oranges : évènements positifs pour le gène cible et référence RPP30. Les points noirs: évènements négatifs.

3. Test de reproductibilité

Le test en intra et inter-plaques a été réalisé avec deux contrôles ayant un nombre de copies connu. La figure 63 illustre les résultats et montre que l'expérience est reproductible même lorsqu'elle n'est réalisée qu'une seule fois. La quantification des CNVs de *CCL3L1* a donc été réalisée en une seule réaction pour chaque échantillon.

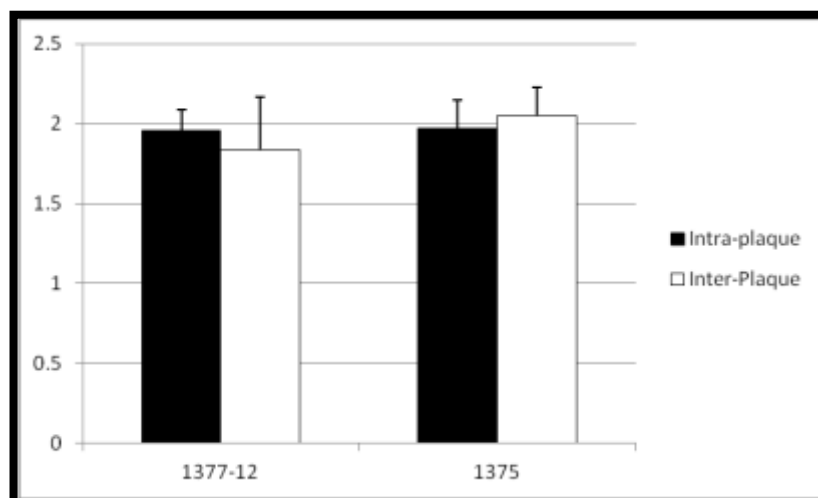


Figure 63: Résultats du test de reproductibilité intra et inter-plaque pour *CCL3L1*. Le test en intra-plaque a été réalisé sur deux échantillons en duplicat dans la même plaque alors que le test en inter-plaque a été réalisé sur les mêmes échantillons une seule fois dans deux plaques différentes.

4. Test de digestion

Les tests de digestion ont été réalisés sur un échantillon contrôle et montre un nombre de copies égal à 1,5. Plusieurs enzymes ont été utilisées sauf HaeIII qui, après vérification *in silico*, coupe dans la séquence amplifiée du gène. Nous avons réalisé l'expérience avec un gradient de 1 à 3U de chaque enzyme avec une digestion préalable ou pendant l'amplification. Il n'y a pas eu de modification du nombre de copies quelle que soit les conditions utilisées (Figure 64).

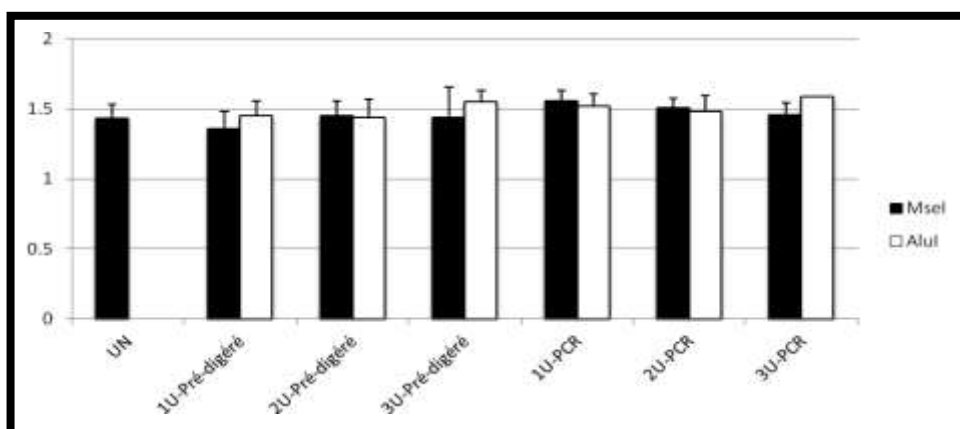


Figure 64: Résultats des tests de digestion avec deux enzymes (MseI et AluI) à différentes unités et avec deux protocoles. La digestion a été réalisée sur l'ADN du contrôle NA02587 soit avant l'amplification (Pré-digéré) ou pendant l'amplification (PCR). UN : correspond à l'ADN non digéré.

Par ailleurs, les résultats obtenus sans digestion pour plusieurs échantillons montrent un nombre de copies pouvant aller jusqu'à 4 copies. La digestion n'est pas donc nécessaire pour la caractérisation des duplications en tandem du gène *CCL3L1*. Pour la suite, nous avons fait la quantification sans digestion.

5. Caractérisation des CNVs

La quantification a été réalisée sur 100 familles. Sept échantillons, ne répondant pas à notre condition de validation (nombre de copies avec une variation supérieure à $\pm 0,25$), n'ont pas été pris en compte ainsi que leurs familles. Le nombre de copie obtenu varie entre 0 et 4 avec une fréquence de 3% pour 0 copie, 56% pour 2 copies, 23% pour 1 copie, 17% pour 3 copies et 1% pour 4 copies (Figure 65).

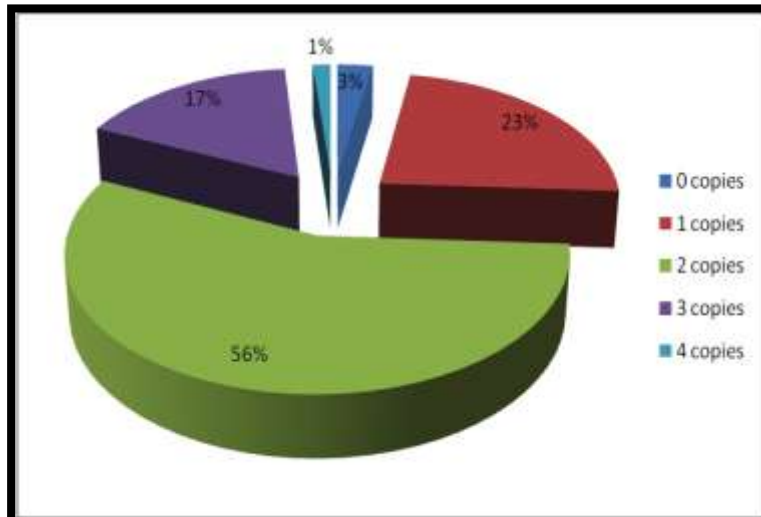


Figure 65: Distribution des CNVs du gène *CCL3L1* dans les échantillons de 93 familles trio.

Des échantillons ayant 2, 3 et 4 copies ont été digérés avec l'enzyme MseI suivant les conditions décrites dans la page précédente. Le nombre de copies de ces échantillons n'a pas été modifié, ce qui indique que la ddPCR peut distinguer chaque copie correctement qu'elle soit en tandem ou non.

C. Transmission des CNVs dans les familles

Nous avons essayé de déterminer le génotype en nombre de copies pour chaque individu en utilisant le contexte familial. Dix-huit modèles ont été obtenus (Tableau 23). Les CNVs des cas-index varient entre 1 et 3 contrairement aux parents dont le nombre de copies varie entre 0 et 4. Seuls 6 modèles de familles nous permettent de conduire à la caractérisation de génotypes. Dans les modèles 9 et 12, deux échantillons présentent un génotype 2/0. Dans le modèle 16, un des parents transmet 3 copies et l'autre transmet 0 copie conduisant au génotype 3/0 du cas index. Les familles du modèle 9 montrent que les copies 3 peuvent correspondre au génotype 1/2 et les familles du modèle 17 montrent que les 2 copies peuvent être à l'état homozygote 1/1. Pour les modèles 2 et 12 une délétion *de novo* peut être envisagée en fonction des génotypes que peuvent avoir les parents (2/0 pour un des parents du modèle 2 et 1/1 pour les parents du modèle 12). Au vu de ces résultats, il est donc difficile d'identifier les génotypes des individus des autres modèles et donc de faire une étude d'association en analysant la transmission des allèles de nombre de copies.

Tableau 23: Transmission des CNVs du gène *CCL3L1* dans 93 familles trio

	Modèles familiaux	Génotypes	n	observations
1	2X2=2		25	
2	2X3=1		2	délétion de novo possible
3	1X1=1	1/0X1/0=1/0	4	
4	0X1=1	0/0X1/0=1/0	6	
5	2X3=3		7	
6	1X3=2		5	
7	2X3=2		10	
8	2X1=1		10	
9	1X4=3	1/0X2/2=1/2	1	
10	2X1=2		10	
11	2X2=1		2	délétion de novo possible
12	0X3=2	0/0X1/2=2/0	1	
13	3X3=3		1	
14	4X2=3		2	
15	1X3=3		2	
16	0X4=3	0/0X3/1=3/0	1	
17	2X2=3	2/0X1/1=1/2	3	
18	3X1=1		1	

Sept des 100 familles ont été exclues puisqu'un individu de la famille présente une valeur de CN supérieure ou égale à un nombre entier $\pm 0,25$ ou présente une délétion *de novo*.

D. Relation entre les CNVs de *CCL3L1* et les facteurs de la PR

Nous avons investigué l'association entre les CNVs du gène *CCL3L1* et les facteurs à risque de la PR pour 98 cas. L'analyse a été réalisée en étudiant l'association du nombre de copies différents de 2 en fonction des caractéristiques des patients. Seule une tendance non significative a été observée dans le sous-groupe de patients avec un âge du début de la maladie inférieur à 40 ans et ayant 3 copies (Tableau 24).

Tableau 24: Analyse de la relation entre les CNVs du gène *CCL3L1* et les facteurs associés à la PR chez 98 cas-index

CNVs du gène <i>FCGR3B</i>	Caractéristiques des patients		OR (IC)	P
	Hommes	Femmes		
1	2	25	2,22 (0,43-11,28)	0.5
2	8	45		
3	2	16	1,42 (0,27-7,41)	1
	Age>40	Age<40		
1	9	18	1,11 (0,42-2,97)	1
2	19	34		
3	2	16	4,47 (0,98-21,56)	0.07
	FR-	FR+		
1	6	21	0,68 (0,21-2,17)	0.5
2	9	44		
3	4	14	0,68 (0,18-2,56)	0.7
	Erosion-	Erosion+		
1	1	26	3,95 (0,46-33,95)	0.2
2	7	46		
3	2	16	1,21 (0,22-6,47)	1
	Nodules-	Nodules+		
1	19	8	1,06 (0,38-2,95)	1
2	38	15		
3	11	7	1,61 (0,52-4,94)	0.5
	ACPA-	ACPA+		
1	6	21	0,73 (0,23-2,33)	0.7
2	9	43		
3	6	12	0,41 (0,12-1,41)	0.2
	Exposition au tabac -	Exposition au tabac +		
1	13	14	1,2 (0,47-3,05)	0.8
2	28	25		
3	7	11	1,76 (0,59-5,23)	0.4
	HLA-SE-	HLA-SE+		
1	6	21	1,02 (0,33-3,11)	1
2	12	41		
3	4	14	1,04 (0,28-3,69)	1

*: Age de début de la maladie. Deux des 100 cas-index ont été exclues puisqu'ils présentent une valeur de CN supérieure ou égale à un nombre entier $\pm 0,25$.

Chapitre 7 : Discussion

Au cours de ce travail nous avons étudié l'implication des CNVs de gènes candidats fonctionnels dans la Polyarthrite Rhumatoïde. Ces gènes sont impliqués soit directement dans la réponse immunitaire caractéristique de la PR, soit dans la survie et la détoxification des cellules actrices dans le déroulement de la maladie. Des CNVs ont été décrits au niveau de ces gènes induisant une délétion complète du gène ou sa duplication en tandem sur un même chromosome, ce qui pourrait modifier l'expression de ces gènes. Plusieurs études se sont intéressées à l'association de ces CNVs en utilisant différentes méthodes pour les caractériser. La première partie de notre travail a été de mettre en évidence les CNVs des gènes *GSTM1* et *GSTT1*, *FCGR3B* et *CCL3L1* en utilisant, à chaque fois, soit les trois variantes de la PCR (la PCR classique, la qPCR et la PCR digitale) soit deux d'entre elles. La deuxième partie a consisté en la caractérisation, après estimation des CNVs, des génotypes en nombre de copies et à étudier la transmission de ces CNVs dans notre échantillon de familles trio. Une fois les génotypes caractérisés, nous avons investigué l'association de ces CNVs avec la PR et avec les facteurs cliniques, biologiques, génétiques et environnementaux associés à la PR.

I. Caractérisation des CNVs du gène *GSTM1*

Notre travail a commencé par l'étude des CNVs du gène candidat *GSTM1*. Ce gène code pour une enzyme impliquée dans la phase II de la détoxification et joue un rôle dans la protection cellulaire contre le stress oxydatif (Jancova, Anzenbacher et al. 2010). Dans des études réalisées en 2006 et 2008, il a été démontré que l'activité des antioxydants, et plus particulièrement la Glutathion S-transférase, était augmentée dans le liquide synovial et le sang des patients PR (Ostalowska, Birkner et al. 2006; Surapneni and Chandrasada Gopan 2008). Nous avons donc émis l'hypothèse que les variations de copies de ce gène pourraient avoir un effet sur la physiopathologie de cette maladie. Pour étudier la transmission du gène *GSTM1* au sein des familles et son association avec la PR nous avons utilisé trois techniques. Tout d'abord, nous avons identifié la présence (+/+ ou +/-) ou l'absence (-/-) du gène grâce à la PCR classique, mais sans connaître le nombre de copies. Ensuite, nous avons utilisé une PCR décrite par Buchard et ses collaborateurs (Buchard, Sanchez et al. 2007) pour distinguer le génotype homozygote (+/+) du génotype (+/-). Cette variante de la PCR n'a permis d'identifier que des génotypes (-/-) et (+/-) dans la population étudiée. Cependant, dans les études réalisées dans les populations Danoises, Somaliennes et Groenlandaises, le génotype +/+ a été mis en évidence avec une fréquence de 7%, 16% et 8%, respectivement. Cette

différence pourrait s'expliquer par l'origine ethnique des populations étudiées. Nous avons par la suite utilisé la qPCR pour estimer le nombre de copies des individus positifs pour le gène, et confirmer donc les génotypes obtenus. Nous avons conclu de ces résultats que les génotypes de tous les échantillons ayant 2 copies étaient de génotype 2/0. Nous avons dès lors mis en place une PCR qui permet d'amplifier de longues séquences allant jusqu'à 40 kb, ce qui devait permettre la mise en évidence de la duplication en tandem de *GSTM1*. Cette PCR a montré que seul un échantillon présentait cette duplication et a confirmé donc que la fréquence des hétérozygotes (+/-) identifiés par PCR de génotype, discordante de celle décrite dans la littérature, n'est pas correcte. Pour confirmer ces résultats, nous avons tout d'abord utilisé la dPCR sur puce, mais cette dernière ne nous a pas apporté de réponse relative aux génotypes de nombre de copies. En effet, le protocole de cette technologie ne permet pas la distinction entre les configurations des génotypes ayant 2 copies (1/1 ou 2/0). Ensuite, nous avons testé tous les échantillons ayant 2 copies avec la *Droplet Digital PCR* (ddPCR) avec ou sans étape de digestion de l'ADN. Cette dernière a montré que, parmi les individus ayant 2 copies, seulement deux échantillons avaient finalement un génotype 2/0, les autres individus ayant un génotype 1/1. Ce résultat confirme celui obtenu par la longue PCR pour un seul échantillon. En effet, cette dernière n'a pas mis en évidence le deuxième échantillon ayant cette duplication en tandem. Ceci est sans doute dû au fait que la longue PCR est sensible à la quantité et à la qualité de l'ADN puisqu'elle amplifie de grands fragments. Donc si l'ADN de l'échantillon est partiellement dégradé ou n'est pas en concentration suffisante, l'amplification ne peut pas avoir lieu. Cette hypothèse est validée par la faible concentration en copies/ μ l de l'ADN de cet échantillon. Les résultats de la ddPCR confirment que les génotypes +/- identifiés pour les échantillons ayant 2 copies avec la PCR de génotypes, sont donc de faux hétérozygotes. Ceci peut s'expliquer par une amplification non spécifique de la région délétée de 4,7kb dans tous les échantillons génotypés. Par ailleurs, la qPCR n'a pas été capable de différencier entre le génotype homozygote (1/1) et hétérozygote (2/0). Plusieurs études ont utilisé la qPCR pour la caractérisation des CNVs du gène *GSTM1* dans les maladies complexes. Les fréquences des CNVs obtenues correspondent à celles que nous avons trouvées. Cependant, toutes ces études ont considéré que le génotype des échantillons avec 2 copies était homozygote (1/1) (Brasch-Andersen, Christiansen et al. 2004; Rose-Zerilli, Barton et al. 2009; Zhou, Hu et al. 2010; Lundstrom, Hartshorne et al. 2011).

La ddPCR a été la plus précise et la plus sensible pour l'analyse des nombres de copies de *GSTM1*, permettant ainsi la mise en évidence de deux copies dupliquées en tandem. En

effet, nous avons observé deux échantillons ayant 1,44 et 1,6 copies obtenues sans digestion et qui reflètent une sous-estimation des deux copies en tandem comptées comme un seul évènement dans la même *droplet*. Après digestion enzymatique de l'ADN des deux échantillons, le nombre de copies augmente pour atteindre 1,9 copies. Ceci confirme bien que les deux copies du gène *GSTMI* sont dupliquées en tandem. Pour les autres échantillons, nous avons observé que les barres d'erreurs calculées à partir des replicats sont plus faibles avec la ddPCR qu'avec la qPCR, permettant ainsi d'avoir des résultats plus précis.

Dans notre étude, 0,3% des échantillons étudiés montrent une duplication du gène *GSTMI* sur le même chromosome. Cette duplication a été décrite comme étant le résultat d'une NAHR entre deux chromosomes homologues, induisant simultanément la délétion de *GSTMI* sur un chromosome et sa duplication sur le deuxième chromosome (McLellan, Oscarson et al. 1997). Dans la littérature, cette duplication a été observée seulement chez les échantillons ayant 3 copies ou plus, avec une fréquence de 0,1% dans la population Suédoise (2386 patients et 1249 contrôles) (Lundstrom, Hartshorne et al. 2011), de 0,2% dans la population Danoise (10271 individus) (Norskov, Frikke-Schmidt et al. 2011) et 0,08% dans la population Allemande (1320 individus) (Timofeeva, Jager et al. 2009). Plus récemment, une étude réalisée sur la population Guadeloupéenne, descendante d'Afrique, a montré une fréquence plus élevée de la duplication (1,2%) avec un nombre allant de 0 à 4 copies. Mais les résultats de cette dernière étude, qui a été réalisée par qPCR, sont biaisés puisque, pour la quantification des CNVs, le nombre de copies 2 a été considéré comme étant le plus fréquent, ce qui ne concorde pas avec la littérature et avec ce que nous avons trouvé. En effet, la fréquence des 2 copies est d'environ 9% dans la population Caucasienne et Africaine alors que la copie 1 a une fréquence de 37%, ce qui suggère que la fréquence de la duplication est surestimée dans cette étude (Emeville, Broquere et al. 2014). De plus, la fréquence observée de la délétion dans notre étude (53,5%), et celle décrite dans la littérature (environ 52%), est beaucoup plus fréquente que celle de la duplication (0,3%), ce qui nous permet de confirmer l'hypothèse de McCarroll et ses collaborateurs, qui suggèrent qu'il y a une sélection positive de la délétion du gène *GSTMI* (McCarroll, Hadnott et al. 2006).

L'analyse des CNVs dans les familles trio montre plusieurs modèles de transmission des parents vers les patients PR, et souligne que la duplication en tandem est transmise d'un parent à sa descendance et qu'elle n'est pas le résultat d'une recombinaison *de novo*. Les

autres modèles obtenus indiquent qu'il n'y a pas de transmission préférentielle entre la délétion (0 copie) et la présence du gène (1 copie).

L'étude familiale des CNVs du gène *GSTM1* n'a pas révélé d'association significative après correction des tests multiples, même après stratification selon les facteurs cliniques, biologiques, génétiques et environnementaux. Cependant, une transmission préférentielle de la délétion et de la présence du gène a été observée dans le sous-groupe des patients ayant un âge de début de la maladie supérieur à 40 ans et dans le sous-groupe de patients présentant des nodules. Ce résultat est corroboré avec celui obtenu lors de l'analyse du risque entre les CNVs du gène et les différents facteurs. Des études réalisées sur des cas/témoins ont montré une association significative entre *GSTM1* et la PR. Une première étude a suggéré que *GSTM1* avait un rôle dans la sévérité de la maladie (Mattey, Hassell et al. 1999). Une autre étude a testé l'association entre l'exposition au tabac et la délétion du gène chez les patients PR et a montré un risque plus élevé chez les patients fumeurs ayant le génotype (-/-) que chez les patients non fumeurs, mettant en évidence ainsi l'influence de la délétion dans la relation entre l'exposition au tabac et la sévérité de la maladie (Mattey, Hutchinson et al. 2002). Ces résultats sont confortés par une étude récente réalisée sur la population Coréenne montrant une association significative entre la délétion et la sévérité de la PR, plus particulièrement chez les patients HLA-DRB1 SE- (Yun, El-Sohemy et al. 2005). Mais d'autres études n'ont pas trouvé d'association significative entre *GSTM1* et la PR ni dans le groupe de patients fumeurs ni dans d'autres sous-groupes (Criswell, Saag et al. 2006 ; Bohanec Grabar, Logar et al. 2009 ; Keenan, Chibnik et al. 2010). De plus, une étude récente utilisant la qPCR pour l'identification des CNVs de *GSTM1* dans une population de 2386 cas et 1249 témoins, n'a trouvé aucune association entre les variants du gène et la PR. Néanmoins, la stratification a permis de mettre en évidence une association entre la délétion du gène et la présence d'ACPA chez les patients positifs pour SE+ et une association entre la délétion et l'exposition au tabac dans le sous-groupe de patients séropositifs pour les ACPA. Un effet protecteur de la présence du gène a été mis en évidence chez les patients hommes fumeurs et séropositifs pour les ACPA. Ce résultat n'a pas pu être répliqué dans notre population car les effectifs issus de ces stratifications sont de tailles trop faibles. Deux récentes méta-analyses réalisées sur le gène *GSTM1* dans la PR dans deux populations, Européenne et Asiatique, ont révélé une association significative seulement dans la population Asiatique. Enfin, dans un projet collaboratif avec une équipe Tunisienne (Dr Abdellatif Maalej, Université de Sfax, Tunisie), les CNVs du gène *GSTM1* ont été quantifiés par ddPCR et les résultats préliminaires ne

montrent aucune association entre *GSTM1* et la PR dans la population Tunisienne. *Ce travail fait l'objet d'un manuscrit en cours de préparation (Achour, Ben Kilani, et collaborateurs).*

Cette hétérogénéité des résultats peut s'expliquer par le fait que les associations observées dans des petites populations (entre 115 et 277 individus) ne sont pas robustes statistiquement. Au vu de ces résultats, il semble que les CNVs du gène *GSTM1* n'ont pas d'effet sur la PR, mais qu'ils pourraient agir sur la physiopathologie puisqu'avec nos résultats et ceux observés dans la littérature, la délétion de *GSTM1* semble avoir une répercussion chez les patients ayant un âge de début de la maladie supérieur à 40 ans et les cas-index ACPA séropositifs et fumeurs. En effet, ayant un rôle dans la détoxification, nous pouvons supposer que la délétion de *GSTM1* pourrait avoir une conséquence sur la protection de l'organisme lors du vieillissement, puisqu'une étude fonctionnelle récente a montré chez le rat sénescent, par rapport à un jeune rat, une forte augmentation de l'activité des GSTs par rapport aux autres enzymes impliqués dans la détoxification (Vyskocilova, Szotakova et al. 2013). De plus, il a été démontré que l'exposition au tabac augmentait l'activité des GSTs (Ignatowicz, Wozniak et al. 2013). Et puisque l'exposition au tabac semble être impliquée dans le processus de production des ACPA (Klareskog, Ronnelid et al. 2008), nous pouvons donc penser que la délétion du gène pourrait déréguler la détoxification des xénobiotiques produits par le tabac et augmenter indirectement la production des ACPA.

Un article synthétisant les résultats obtenus pour la caractérisation des CNVs du gène GSTM1 et leur analyse dans les familles trio a été soumis (voir Annexe).

II. Caractérisation des CNVs du gène *GSTT1*

Le deuxième gène étudié, *GSTT1*, est également membre de la famille des GSTs et participe à la détoxification cellulaire. Même si ce dernier n'est pas exprimé dans les cellules immunitaires, il a été associé dans certaines études à l'exposition au tabac chez les patients PR et également dans le groupe de patients séropositifs pour les auto-anticorps (Bohanec Grabar, Logar et al. 2009; Chen, Huang et al. 2012). Mais, jusqu'à aujourd'hui, cette association reste controversée. De plus, la caractérisation des CNVs de ce gène a été souvent limitée à l'identification de la présence/absence de ce gène, ce qui pourrait biaiser l'analyse d'association. C'est pour cette raison que, dans notre étude, la PCR d'identification de la présence/absence du gène a été accompagné d'une PCR de génotype, permettant ainsi de faire

la distinction des trois génotypes (-/-, +/- et +/+). Contrairement à *GSTMI*, les fréquences génotypiques de *GSTTI* obtenues sont cohérentes avec ce qui a été décrit dans la littérature (Brasch-Andersen, Christiansen et al. 2004). Ces trois génotypes reflètent le nombre de copies du gène puisqu'aucun résultat de NAHR n'a été mis en évidence dans la littérature, et que les données sur génome entier obtenues par aCGH n'ont révélé aucune duplication (Conrad, Pinto et al. 2010). De plus, des études réalisées avec différents protocoles de qPCR et dans d'autres maladies complexes ont montré que le nombre de copies varie entre 0 et 2 copies (Norskov, Frikke-Schmidt et al. 2009; Rose-Zerilli, Barton et al. 2009; Norskov, Frikke-Schmidt et al. 2011). Enfin, l'étude investiguant l'expression de *GSTTI* montre également que les trois génotypes (-/-, +/- et +/+) sont corrélés avec les niveaux d'expression des copies 0, 1 et 2.

L'analyse de la transmission des allèles de *GSTTI* n'a pas révélé d'association significative avec la PR. Ce résultat confirme celui observé dans la littérature. En effet, la première étude réalisée en 1999 n'a pas révélé d'association entre la présence/absence du gène et la PR (Mattey, Hassell et al. 1999). Un résultat similaire a été obtenu en 2005 dans la population Coréenne (Yun, El-Sohemy et al. 2005). Mais une étude cas-index réalisée sur 213 patients PR, issus de la population Caucasienne, a montré que la délétion complète du gène *GSTTI* était associée à une sévérité accrue de la maladie. De plus, cette étude a montré que la délétion était également associée à un risque élevé dans le sous-groupe de patients fumeurs. Enfin, un effet combiné entre la délétion du gène *GSTMI* et celle du gène *GSTTI* a été observé dans le sous-groupe de patients fumeurs (Bohanec Grabar, Logar et al. 2009). Un travail plus récent réalisé sur 585 cas/contrôles affiliés, a mis en évidence une interaction entre la délétion du gène *GSTTI* et l'exposition au tabac, mais également avec la production d'ACPA chez les patients fumeurs. Comme pour *GSTMI*, nous pouvons supposer que la délétion du gène *GSTTI* perturbe la détoxification des xénobiotiques produit par l'exposition au tabac induisant ainsi une production d'auto-anticorps. Par contre, aucune interaction entre *GSTMI* et *GSTTI* dans les différents sous-groupes n'a été observée (Keenan, Chibnik et al. 2010). Les résultats des récentes méta-analyses investiguant l'association entre *GSTTI* et la PR sur plus de deux milles cas/contrôles, viennent s'opposer aux précédentes observations (Chen, Huang et al. 2012; Ji and Lee 2013). Cette hétérogénéité des résultats peut avoir deux explications : la première est que les associations significatives observées sont des résultats faux-positifs potentiellement dus à la taille restreinte des échantillons. La deuxième explication est que cette hétérogénéité est due à une différence des méthodologies utilisées pour la caractérisation des CNVs de *GSTTI*. En effet les premières études ont été réalisées par

la PCR présence/absence alors que celle de 2010 a utilisé la qPCR pour la quantification du nombre de copies.

Néanmoins, une association significative, même selon le seuil de significativité après correction des tests multiples, a été mise en évidence dans le sous-groupe de patients séropositifs pour le Facteur Rhumatoïde avec la présence d'au moins une copie du gène. Ce résultat a été confirmé par une tendance non significative observée par l'analyse de l'OR. Cette association concorde avec ce qui a été décrit dans la littérature, ce qui permet de supposer que la présence du gène *GSTT1* chez les patients PR pourrait favoriser la production de FR. Il est intéressant de noter que, lorsque chaque génotype est analysé séparément, la présence du FR est seulement associée au génotype hétérozygote, ce qui pourrait suggérer que la délétion sur un chromosome induit une diminution de l'expression du gène et, par conséquent, celle de l'activité de l'enzyme correspondante. Ceci pourrait perturber la détoxification cellulaire et induit la production d'auto-anticorps, alors que la délétion complète pourrait induire un système de compensation faisant intervenir une autre voie de détoxification. Cette dernière hypothèse serait confortée par le fait qu'aucune association avec la délétion complète n'a par ailleurs pas été observée. Malheureusement, aucune étude fonctionnelle n'a été faite pour investiguer cette hypothèse, ce qui pourrait être fait en analysant des modèles de souris atteint d'arthrite, le premier hétérozygote KO pour le gène *GSTT1*, le second homozygote pour la présence du gène (+/+) et le dernier avec la délétion complète de *GSTT1*.

III. Caractérisation des CNVs du gène *FCGR3B*

L'analyse a ensuite porté sur le gène *FCGR3B* impliqué dans la réponse inflammatoire et plus précisément dans la reconnaissance des IgGs par leur récepteur membranaire FcG des neutrophiles (Simmons and Seed 1988). En 2002, il a été suggéré que la voie de signalisation induite par le récepteur FcγRIIIb pouvait être une cible thérapeutique pour la PR. En effet, cette étude a montré, grâce à un modèle murin, que l'inhibition de ce récepteur par des anticorps monoclonaux diminuait l'inflammation des articulations (Fossati, Moots et al. 2002). Nous avons donc supposé que des variations génétiques modifiant l'expression de ce gène pourraient, par conséquent, avoir un impact sur la physiopathologie de la maladie.

Plusieurs études ont montré que le gène *FCGR3B* était localisé dans un *cluster* de gènes homologues sujets à de nombreuses recombinaisons permettant ainsi l'apparition de CNVs multi-alléliques (de Haas, Kleijer et al. 1995; Koene, de Haas et al. 1996; Machado, Hardwick et al. 2012). Ces différentes recombinaisons ne permettent pas de mettre au point une PCR pour l'identification des génotypes comme pour les deux précédents gènes. En effet, ces variations structurales induisent des CNVs de différentes tailles en fonctions des différentes régions de recombinaison. Donc, pour deux échantillons différents, nous pourrions avoir des produits PCR de différentes tailles, compliquant ainsi l'interprétation des résultats. Dans la littérature, seule une PCR de présence/absence a été réalisée amplifiant deux séquences, la première correspondant à celle du gène *FCGR3B* et la deuxième celle de *FCGR3A* (de Haas, Kleijer et al. 1995). Mais comme pour *GSTM1* et *GSTT1*, cette PCR ne permet pas d'identifier le génotype. Nous avons donc utilisé la qPCR comme première étape pour caractériser les CNVs du gène *FCGR3B*. Les résultats obtenus ont montré que les nombres de copies estimés étaient peu fiables puisque l'indice de confiance n'atteignait pas 90%. Et, même pour les échantillons ayant cet indice supérieur à 90%, les barres d'erreurs étaient élevées, comme ceux observées lors de l'analyse des CNVs du gène *GSTM1* par qPCR. Ce taux élevé d'échantillons qui ne répondent pas aux critères de qualité que nous avons choisis, peut s'expliquer par le fait que la PCR quantitative n'était pas assez sensible pour la détection des CNVs multi-alléliques atteignant les 4 copies. Weaver et ses collaborateurs ont démontré qu'il faut au minimum 18 réplicats pour améliorer la sensibilité de la qPCR afin qu'elle puisse détecter une variation entre 0 et 4 copies (Weaver, Dube et al. 2010). L'analyse de ces CNVs par ddPCR a permis de confirmer ce manque de sensibilité de la qPCR puisque plusieurs échantillons montrent un nombre de copies différents selon la technique utilisée. En effet, nous avons observé que, pour des échantillons donnés identifiés par qPCR comme ayant par exemple 0, 1 et 4 copies, la ddPCR identifie 2 copies. Par ailleurs, les fréquences des CNVs obtenues par ddPCR diffèrent légèrement avec ce qui a été décrit dans la littérature dans une population de même origine ethnique. Les résultats obtenus par qPCR, PRT ou par MLPA montrent une fréquence de la délétion du gène entre 6 et 10%, alors que dans notre étude, elle est de 12%. De plus, le pourcentage de copies supérieures à 2 dans les études réalisées avec les trois technologies citées précédemment se situe entre 7,9 et 12%, alors que dans notre population il est de 9% (McKinney and Merriman 2012). Ces divergences pourraient venir de l'attribution erronée de certaines valeurs de nombre de copies

avec des méthodologies moins sensibles que la ddPCR comme ce que nous avons observé avec nos résultats de qPCR.

Contrairement à *GSTMI*, la digestion préalable de l'ADN pour la ddPCR n'a pas permis de mettre en évidence les duplications en tandem du gène *FCGR3B*. Pourtant, les 3 et 4 copies indiquent qu'il y a forcément une duplication du gène pour pouvoir atteindre ces valeurs. Ceci peut s'expliquer par la distance estimée entre les *segmental duplications* (SD) du gène qui est d'environ 82kb. En effet, nous pouvons émettre l'hypothèse que le partitionnement, lors de la génération de *droplets*, fragmente l'ADN et par conséquent, lorsque la distance qui sépare les deux copies en tandem est grande, comme c'est le cas pour *FCGR3B*, les duplicats sont séparés dans deux *droplets* différentes permettant ainsi une estimation correcte du nombre de copies sans l'étape de digestion de l'ADN. Cette supposition est confortée par le fait que le nombre de copies obtenu par quantification absolue est proche de la valeur entière (par exemple 2,8 ou 2,9) et non une valeur arrondie, ce qui suggère que les copies ne sont pas séparées dans une faible proportion de *droplets*. Dans le cas du gène *GSTMI*, la distance entre les deux copies en tandem est d'environ 18kb, c'est pourquoi le protocole de digestion était nécessaire pour l'identification des génotypes en nombre de copies.

Pour attribuer les génotypes en fonction des CNVs identifiés, deux configurations existent : la première étant de prendre en considération la faible fréquence de duplication (9%) et de supposer que le génotype des échantillons ayant 2, 3 et 4 copies est respectivement 1/1, 2/1 et 2/2. La deuxième configuration implique que la duplication en tandem est fréquente et donc les échantillons ayant 2, 3 et 4 copies peuvent avoir le génotype 2/0, 3/0 et 3/1 ou 4/0. Pour départager ces deux hypothèses, nous avons étudié le contexte familial, et celui-ci confirme que la première hypothèse est la plus plausible puisqu'aucun des modèles de transmission établi ne montre de génotype 2/0. De plus, plusieurs modèles montrent que les échantillons ayant 2 copies sont homozygotes 1/1, ce qui conforte donc la première configuration. La transmission des CNVs dans les familles montre trois événements de délétion *de novo* dont deux ont été produites à partir d'un génotype 1/1. La troisième a eu lieu au niveau des copies en tandem rétablissant, pour la descendance, un nombre de copie normal sur chaque chromosome. Comme pour *GSTMI*, la duplication en tandem est transmise des parents vers les descendants et n'est pas le résultat d'une duplication *de novo*. Aucune transmission préférentielle n'a été observée entre le nombre de copies 0 et 1 et entre le nombre

de copies 1 et 2. Néanmoins, nous avons observé que dans les familles où un des parents a 3 copies, la duplication est plus souvent transmise lorsque le deuxième parent est homozygote 1/1 alors qu'elle sera moins transmise lorsque le deuxième parent est de génotype 1/0. Cette observation pourrait souligner une sélection positive de la duplication lorsque les génotypes parentaux ne présentent pas de délétion. Lorsque l'un des parents présente la délétion du gène *FCGR3B*, c'est le nombre de copie 1 qui est le plus transmis, comme pour *corriger* le génotype de la descendance.

Les analyses de transmission dans les familles étudiées n'ont pas permis de mettre en évidence une association significative entre les CNVs du gène *FCGR3B* et la PR. Mais la stratification selon les différents facteurs associés à la PR montre deux tendances non significatives, dans le sous-groupe de patients FR séropositifs et dans celui des patients ACPA séropositifs. Ces résultats n'ont pas été retrouvés par calcul de l'Odds Ratio dans les différents sous-groupes. Dans la littérature, les résultats d'association entre la délétion du gène *FCGR3B* et la PR divergent. En effet, dans la première étude réalisée en 2010 sur différentes populations, une association significative a été mise en évidence dans la population Allemande. Après regroupement des différentes cohortes, la méta-analyse confirme également l'association observée (McKinney, Fanciulli et al. 2010). La quantification des CNVs a été réalisée par qPCR et montre des fréquences des nombres de copies différentes par rapport à celles que nous avons obtenues. Une autre étude utilisant la qPCR, en prenant comme seuil de caractérisation des CNVs un indice de confiance >80% (bien inférieur à celui utilisé dans notre étude), a été réalisée sur un groupe limité d'échantillons (197 cas et 162 contrôles) et montre une association significative avec la PR. La méta-analyse réalisée avec ces résultats et les précédentes études, permet d'appuyer l'association trouvée (Graf, Lester et al. 2012). Dans la même année, Robinson et ses collaborateurs ont utilisé la PRT pour la quantification des CNVs du gène *FCGR3B* dans une large cohorte de 1115 patients et 654 contrôles et ont analysé l'association avec la PR en établissant deux modèles, l'un comparant les différents nombres de copies (1, 2 et 3 copies) entre les cas/contrôles, l'autre comparant la délétion *versus* les autres valeurs de CNVs entre les cas-index et les individus sains. Une association significative entre la délétion et la PR a été observé dans le deuxième modèle. De plus, les chercheurs ont mis en évidence une association significative dans le sous-groupe de patients FR+ dans le deuxième modèle également (Robinson, Carr et al. 2012). Mais ce résultat n'est pas répliqué dans une étude réalisée sur un large nombre d'échantillons (518 patients PR et 304 contrôles) dont les CNVs ont été quantifiés par la technique MLPA (Marques, Thabet et

al. 2010). Enfin, la méta-analyse réalisée par McKinney et collaborateurs, n'a pas révélé d'association significative entre la délétion du gène et la PR (McKinney and Merriman 2012).

L'origine de ces divergences peut être diverse. En effet dans les études réalisées sur des petits échantillons d'individus, les résultats d'associations décrits peuvent être des faux-positifs, induisant alors un biais lorsqu'ils sont inclus dans les méta-analyses. De plus, plusieurs méthodes ont été utilisées pour la quantification des CNVs et celles-ci ont montré certaines discordances dans l'attribution des CNVs, comme l'a souligné l'étude de McKinney et ses collaborateurs. En effet, ils ont observé une différence, non significative, entre les nombre de copies quantifiés par qPCR et les deux autres méthodes de quantification (PRT et MLPA), ce qui pourrait induire une modification dans la distribution des CNVs entre les cas-index et les contrôles (McKinney and Merriman 2012). La quantification des CNVs du gène *FCGR3B* que nous avons réalisée avec la ddPCR dans les familles trio n'a pas révélé d'association entre ce gène et la maladie. Néanmoins, la tendance non significative observée chez les patients produisant des auto-anticorps supporte ce qui a été décrit dans la littérature. Il serait donc intéressant de confirmer ces résultats dans une large cohorte de cas/contrôles en quantifiant les CNVs par ddPCR. Si l'association observée dans la littérature est validée, ceci pourrait suggérer que les CNVs du gène *FCGR3B* influencent la production d'auto-anticorps, mais puisque ces derniers ne sont pas spécifiques que de la PR, il est donc possible que la délétion soit associée à la présence de FR et non associée à la maladie. En effet, il a été démontré que lorsque le gène est délété ou dupliqué, l'expression membranaire des récepteurs FcγRIIIb sur lesquels se fixe le FR est modifiée en conséquence, ce qui apporte une validation fonctionnelle de l'effet des CNVs sur la capacité des neutrophiles, qui expriment ces récepteurs pour la fixation des complexes immuns (Robinson, Carr et al. 2012). De plus, une étude réalisée en 2008 a confirmé que la délétion du gène induisait une diminution de la régulation des complexes immuns solubles et une réduction de l'adhésion de ces complexes formés par des IgG au niveau des neutrophiles (Willcocks, Lyons et al. 2008). Ce résultat semble en concordance avec l'association décrite dans la littérature ainsi que la tendance observée dans notre étude entre la délétion du gène *FCGR3B* et la production de Facteur Rhumatoïde (anti-IgG). En effet, le FR induirait la formation de complexe immuns anti-IgG qui activeraient alors le système immunitaire en se fixant au récepteur FcγRIIIb au niveau des neutrophiles (van Boekel, Vossenaar et al. 2002). Nous pouvons donc supposer que la délétion du gène pourrait perturber la capacité des complexes immuns à se fixer sur leur récepteur au niveau des neutrophiles et par conséquent ces derniers ne seraient plus activés. Il

a cependant été démontré que les neutrophiles au niveau du liquide synovial semblent s'activer tout de même lors de la réponse inflammatoire et que les CNVs ne modifient pas la sécrétion des médiateurs pro-inflammatoires sécrétés par ces cellules (Wagner and Hansch 2004; Willcocks, Lyons et al. 2008). Ceci pourrait conforter l'hypothèse que cette délétion soit associée à la présence de FR et non associée à la maladie, le FR n'étant pas spécifique uniquement de la PR et une autre voie d'activation des neutrophiles dans la PR pouvant être utilisée. En conclusion, le gène *FCGR3B* pourrait alors jouer un rôle dans la régulation des complexes immuns mais pas dans l'inflammation.

IV. Caractérisation des CNVs du gène *CCL3L1*

Le dernier gène investigué dans ce travail est le gène *CCL3L1*. Il code pour deux isoformes de chimiokines qui sont sécrétées par les macrophages lors de la réponse inflammatoire. Elles interviennent dans le chimiotactisme des cellules immunitaires vers le site de l'inflammation. Ce gène a également un rôle dans la réponse anti-virale puisque l'une des isoformes sécrétée est un ligand pour le co-récepteur CCR5 (Menten, Struyf et al. 1999; Aquaro, Menten et al. 2001). Ce gène est connu pour sa grande variabilité et présente des CNVs multi-alléliques décrits dans la littérature (Gonzalez, Kulkarni et al. 2005). En 2005, des séquences d'amorces et d'une sonde ont été décrites pour la quantification des CNVs du gène *CCL3L1* dans le diagnostic lors de l'infection par le VIH (Gonzalez, Kulkarni et al. 2005). Mais, avec le développement d'un protocole de PRT dont la quantification ne tient pas compte du pseudogène, il s'est avéré que les nombres de copies quantifiés par la qPCR étaient surestimés par rapport à ceux trouvés par PRT (Field, Howson et al. 2009). Cependant, aucune des études citées n'a tenu compte du gène homologue *CCL3L3*, qui est identique à *CCL3L1* avec plus de 99% d'homologie. Dans les premiers tests menés avec la ddPCR, de nouvelles séquences d'amorces et de la sonde fluorescente ont été décrites par Sudmant et ses collaborateurs (Sudmant, Kitzman et al. 2010) et utilisées pour la quantification de *CCL3L1* dans des échantillons contrôles (Hindson, Ness et al. 2011). Cependant, ces séquences s'hybrident également au pseudogène. Pour la quantification de ce gène, nous avons donc choisi des amorces et une sonde fluorescente spécifiques qui ne s'hybrident pas avec le pseudogène *CCL3P1* et ni avec le gène *CCL3L3*.

La quantification par ddPCR a caractérisé des CNVs allant de 0 à 4. Comme pour le gène *FCGR3B*, la digestion de l'ADN n'a pas permis de mettre en évidence d'éventuels événements de recombinaison. Le test réalisé sur l'échantillon contrôle ayant un nombre de copies de 1,5 n'a pas montré de modification après digestion de l'ADN par plusieurs enzymes. Nous avons donc émis l'hypothèse que l'ADN de cet échantillon, issu d'une lignée cellulaire commercialisée, pourrait contenir des ADNs issus de deux lignées ayant un nombre de copies différents. Par ailleurs, les valeurs de nombres de copies ≥ 2 restent inchangées avec ou sans digestion, ce qui indique que la ddPCR peut quantifier chaque copie séparément. Dans ce cas, nous supposons que la grande distance séparant les copies en tandem (90kb) facilite la séparation des copies dupliquées sur le même chromosome, lors du partitionnement par la ddPCR. Ce résultat peut s'expliquer également par le fait que les copies dupliquées soit présentes dans deux régions chromosomiques différentes voir même sur deux chromosomes différents. Néanmoins, dans la littérature, aucune étude n'a pour l'instant mis en évidence cette duplication à distance malgré la fréquence élevée de la duplication (9% pour *FCGR3B* et 18 pour *CCL3L1*). Ce résultat indique donc que la quantification des CNVs dépend de la région du gène investigué. Les fréquences des CNVs obtenues dans la population Caucasienne (3% pour 0 copies, 23% pour 1 copie, 56% pour 2 copies, 17% pour 3 copies et 1% pour 4 copies) concordent avec celles décrites dans la littérature sauf pour le nombre de copies 4 où il y a une légère différence (3,9% chez les patients PR et 1,98% chez les contrôles) (Carpenter, Walker et al. 2011).

Pour l'attribution des génotypes en nombre de copies, nous avons étudié le contexte familial puisque la ddPCR n'a pas permis l'identification des copies du gène sur chaque chromosome. La transmission des CNVs dans les familles trio a permis l'identification de différents génotypes et de mettre en évidence plusieurs modèles de transmission. Néanmoins, les génotypes identifiés ne permettent pas d'élucider la transmission de CNVs dans tous les modèles. Comme nous l'avons observé pour le gène *FCGR3B*, il peut y avoir deux configurations des génotypes de nombre de copies ≥ 2 . Mais contrairement au précédent gène, les deux hypothèses sont vérifiées puisque plusieurs modèles montrent que les échantillons ayant 2 copies peuvent avoir un génotype homozygote 1/1 ou hétérozygote pour la délétion 2/0, de même que pour les échantillons ayant 3 copies (1/2 ou 3/0). Ces résultats permettent de souligner la limite de la ddPCR pour l'identification des génotypes mais également l'avantage des analyses familiales pour la caractérisation des CNVs multi-alléliques de gènes candidats. Effectivement, ces résultats n'ont pas été décrits dans la littérature puisque

l'analyse des CNVs du gène *CCL3L1* dans des cohortes de cas/contrôles ne permet pas d'identifier les génotypes des échantillons et de mettre en évidence la présence de duplication ou de triplication en tandem. Etant donné la présence de ces copies sur le même chromosome dans plusieurs modèles familiaux, cela pourrait indiquer que, contrairement au gène *GSTM1*, les duplications ne sont pas anciennes et que les événements de recombinaison induisant ces CNVs sont encore fréquents. Ceci pourrait expliquer la présence de nombre de copies élevé dans certaines populations, notamment dans la population Tunisienne, où nous avons observé une fréquence plus élevée des 4 copies et de 5 copies (14% et 0,67% chez des patients PR, 5% et 9,6% chez des contrôles respectivement) et atteignant même 7 copies (0,67%). Ces résultats concordent avec ceux décrits dans la littérature où les CNVs du gène *CCL3L1* peuvent aller jusqu'à 10 copies dans la population Africaine (Carpenter, Farnert et al. 2012). *Le travail de caractérisation des CNVs du gène CCL3L1 chez les familles trio PR de la population Caucasienne et chez les cas/contrôles Tunisiens fait l'objet d'un manuscrit en cours de rédaction (Ben Kilani, et al).*

La confirmation des deux configurations de génotypes dans notre population ne nous a pas permis d'étudier l'association des CNVs du gène *CCL3L1* avec la PR avec les tests de déséquilibre de transmission (TDT) et du risque relatif des génotypes (GRR) puisque nous ne pouvons pas attribuer des génotypes pour les nombres de copies ≥ 2 . Dans la littérature, une étude réalisée en 2012 a étudié l'association entre les CNVs du gène *CCL3L1* et la malaria dans un échantillon familial Africain. Les auteurs ont mis au point un protocole de PRT conjugué à deux PCR amplifiant deux microsatellites localisés dans la région de SD. Deux paires d'amorces sont couplées à deux fluorophores différents et spécifiques à chaque microsatellite. La variation de la taille de chaque séquence donne un signal fluorescent différent qui va correspondre à chaque allèle. Après amplification, les produits PCR obtenus par PRT et ceux obtenus pour les microsatellites sont différenciés par électrophorèse capillaire, et chaque signal fluorescent est quantifié. Ceci a donc permis aux chercheurs d'établir des haplotypes des CNVs de *CCL3L1* avec les allèles microsatellites. Ces haplotypes ont été par la suite analysés avec une variante du test TDT, le QTDT, qui analyse le déséquilibre de transmission d'un trait quantitatif (Carpenter, Farnert et al. 2012). Cette analyse ne peut s'appliquer dans notre cas puisque la ddPCR ne quantifie que la fluorescence correspondante au gène cible et au gène de référence. Il est à noter cependant que cette variante de la PRT, a un coût assez élevé puisque chaque échantillon est analysé en triplicat pour chaque expérience (3 fois pour la PRT et 3 fois pour chaque microsatellite). C'est pour

cela qu'il serait intéressant de développer un protocole qui regroupe la ddPCR et la PRT (dpPCR). Mais auparavant deux conditions doivent être validées : 1) le gène candidat doit présenter un gène paralogue dans la même région chromosomique qui servira de référence ; 2) il faut caractériser une paire d'amorces qui amplifie les deux gènes simultanément mais ayant chacun une taille différente de quelques paires de bases, permettant ainsi de les distinguer. La quantification pourra être réalisée par l'agent intercalant Eva Green (Bio-Rad) qui donnera donc deux signaux de fluorescents différents, puisque l'intensité de fluorescence dépendra de la taille des séquences amplifiées, et donnera donc deux *clusters* distincts d'évènements positifs. Ceci permettra de quantifier les CNVs du gène. Ensuite, pour la quantification spécifique d'un marqueur dans la région du CNV (microsatellites ou autre) nous pourrons comparer la fluorescence relative de chaque allèle du marqueur. L'analyse conjointe des deux résultats permettra d'identifier les haplotypes CNV/marqueur et donc de caractériser les génotypes de CNVs du gène étudié. Donc, pour un échantillon donné, il faudra réaliser deux réactions, au lieu de 9 comme ce qui a été décrit. Une autre solution peut être appliquée en marquant les amorces utilisées pour la quantification par un premier fluorophore et celles utilisées pour la quantification du microsatellite par un autre fluorophore et dans ce cas la quantification des deux variations se feront dans la même réaction.

Après analyse de l'Odds Ratio entre les CNVs du gène *CCL3L1* et les différents facteurs associés à la PR chez les patients, nous n'avons observé aucune association significative. Ce résultat concorde avec ceux observés dans la littérature. Sur les trois études réalisées sur la PR, deux ne montrent pas d'association entre les CNVs du gène *CCL3L1* et la PR. Ce résultat est également observé dans différents sous-groupes de patients (Carpenter, Walker et al. 2011; Nordang, Carpenter et al. 2012). Seule la première étude réalisée en 2008 a mis en évidence une association entre les CNVs du gène et la PR dans la population Néozélandaise. Mais, dans cette même étude, ce résultat n'a pas été répliqué dans la population Anglaise (McKinney, Merriman et al. 2008). Cette divergence peut être expliquée par le fait que, dans cette étude, la quantification des CNVs a été réalisée par la qPCR décrite en 2005, quantifiant également le pseudogène, ce qui induit une surestimation des valeurs de nombre de copies. En effet, comme l'a montré Walker et ses collaborateurs, le pseudogène est présent dans environ 25% de la population Européenne et sa présence est significativement associée à un nombre de copies élevé du gène *CCL3L1*. Il est donc présent dans environ 10% des individus ayant entre 1 et 2 copies du gène et dans 80% des échantillons ayant un nombre de copies entre 3 et 4 (Walker, Janyakhantikul et al. 2009). Donc si ce pseudogène est présent

chez certains patients, le nombre de copies sera surestimé par rapport aux contrôles ce qui introduirait un biais dans l'analyse d'association. Au regard des variabilités interethniques que nous avons observés, de nouvelles études d'association utilisant des protocoles spécifiques pour *CCL3L1*, comme celui décrit dans cette étude, doivent être réalisées dans d'autres populations.

L'analyse des CNVs des différents gènes candidats a montré que le choix de la méthode de quantification est déterminant puisque, comme nous l'avons observé pour les 4 gènes étudiés, il n'y a pas un protocole unique de quantification pour tous les CNVs. En effet, la ddPCR permet l'estimation précise des CNVs et l'identification des génotypes en nombre de copies pour les courtes *segmentals duplications*, mais elle ne peut caractériser les génotypes pour les longues duplications en tandem, et ceci malgré l'utilisation du protocole de digestion préalable de l'ADN. De plus, la ddPCR a la capacité de quantifier les CNVs de deux gènes d'intérêts en parallèle du gène de référence, ce qui apporte un gain de productivité et un coût moins important pour chaque réaction. C'est pourquoi il semble important d'établir un *process* de caractérisation des CNVs, dans une étude familiale, en fonction de leurs positions dans le génome et de l'évènement de recombinaison à l'origine de leur formation. Les différentes étapes de mise en évidence des CNVs choisies en fonction des différents cas de figures sont illustrées dans la figure 66.

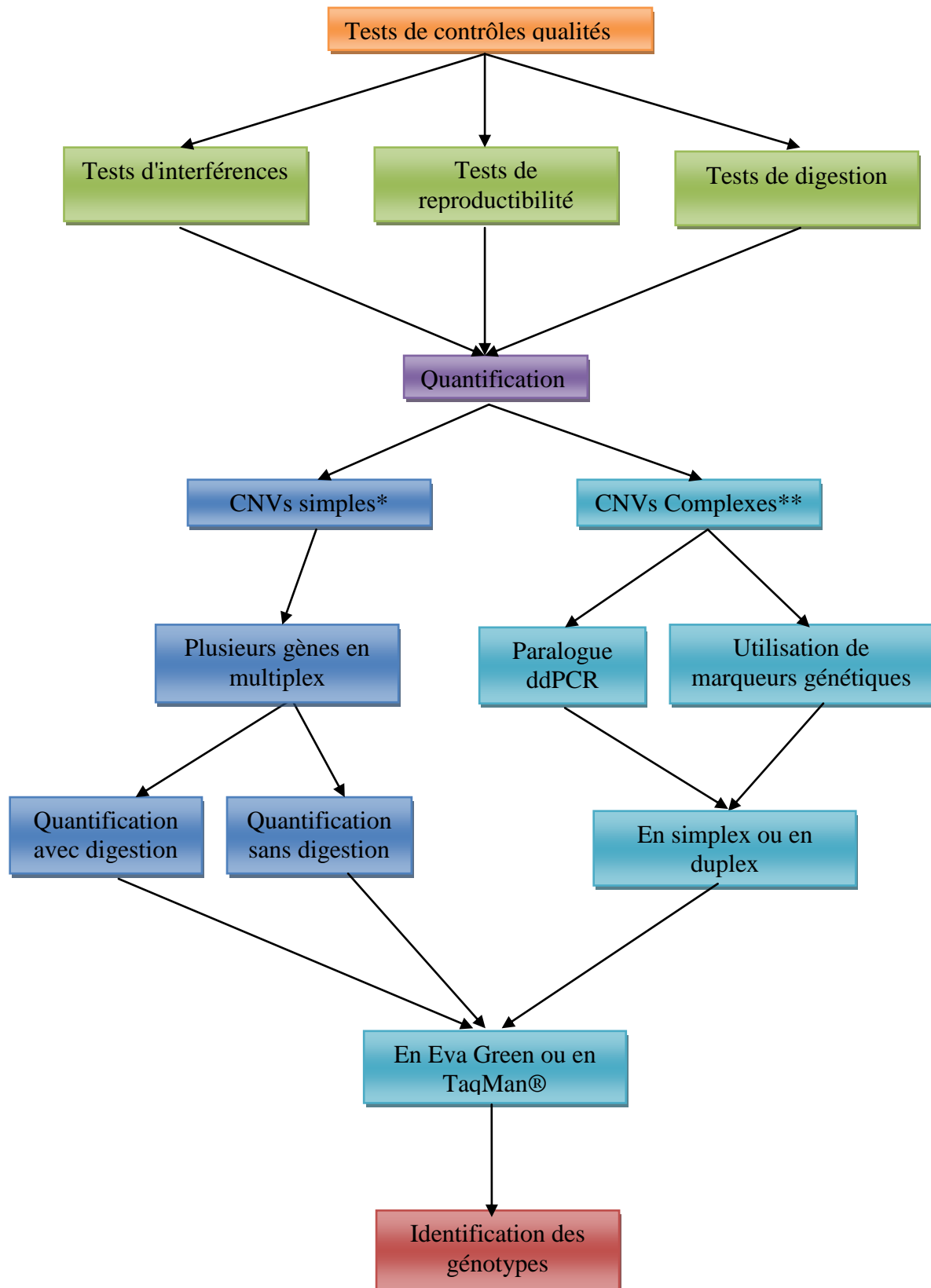


Figure 66: Schéma récapitulatif résumant les différentes étapes à suivre pour la quantification des CNVs par ddPCR en fonction de leurs caractéristiques. *: CNVs avec des SDs courtes (exemple du gène *GSTM1*). **: CNVs avec des SDs longues (exemple des gènes *FCGR3B* et *CCL3L1*).

Chapitre 8 : Conclusion

Au cours de ce travail nous avons investigué l'implication des CNVs de plusieurs gènes candidats dans la Polyarthrite Rhumatoïde en utilisant différentes méthodes de quantification. Nous avons démontré la limite de la PCR classique et de la qPCR pour la quantification des CNVs, qu'ils soient bi-alléliques ou multi-alléliques. Nous avons identifié la ddPCR comme étant la meilleure méthode à utiliser pour la quantification absolue des CNVs. Même si nous n'avons pas pu identifier les génotypes du gène *CCL3L1*, d'autres développements du protocole de la ddPCR peuvent être mis en place pour permettre cette identification, ce qui ne serait pas le cas avec la qPCR. Cette méthode présente plusieurs avantages par rapport à ses ancêtres en termes de productivité, de coût et d'applications diversifiées. En effet, au-delà de la caractérisation des CNVs, la ddPCR peut être utilisée pour la mise en évidence d'évènements rares, la caractérisation de mosaïques, l'identification de variations structurales de grandes tailles, et l'estimation du niveau d'expression des gènes.

L'apport de l'analyse familiale a également été déterminant dans notre étude car il nous a permis d'étudier la transmission de ces CNVs et de mettre en évidence la transmission de la duplication en tandem et les évènements de recombinaison comme les délétions *de novo*. De plus, cette analyse nous a permis d'identifier plusieurs génotypes avec des duplications et des triplications en tandem. Cela souligne la pertinence du choix d'un matériel familial lors de la caractérisation des CNVs.

Cette structure familiale de nos échantillons ont également permis d'investiguer l'association des CNVs avec la PR. Les résultats obtenus ne nous ont pas permis de mettre en évidence une implication des CNVs des gènes étudiés dans la maladie. Néanmoins, les CNVs de certains de ces gènes ont été associés à des facteurs cliniques et biologiques caractéristiques aux maladies autoimmunes, ce qui conforterait l'implication des CNVs dans la physiopathologie de ces maladies complexes.

Références bibliographiques

-A-

- Abecasis, G. R., D. Altshuler, et al. (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.
- Abel, L. and A. J. Dessein (1998). "Genetic epidemiology of infectious diseases in humans: design of population-based studies." Emerg Infect Dis **4**(4): 593-603.
- Alamanos, Y., P. V. Voulgari, et al. (2006). "Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review." Semin Arthritis Rheum **36**(3): 182-188.
- Aletaha, D., T. Neogi, et al. (2010). "2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative." Ann Rheum Dis **69**(9): 1580-1588.
- Almal, S. H. and H. Padh (2012). "Implications of gene copy-number variation in health and diseases." J Hum Genet **57**(1): 6-13.
- Alvaro-Gracia, J. M., N. J. Zvaifler, et al. (1991). "Cytokines in chronic inflammatory arthritis. VI. Analysis of the synovial cells involved in granulocyte-macrophage colony-stimulating factor production and gene expression in rheumatoid arthritis and its regulation by IL-1 and tumor necrosis factor-alpha." J Immunol **146**(10): 3365-3371.
- Antonell, A., O. de Luis, et al. (2005). "Evolutionary mechanisms shaping the genomic structure of the Williams-Beuren syndrome chromosomal region at human 7q11.23." Genome Res **15**(9): 1179-1188.
- Aquaro, S., P. Menten, et al. (2001). "The LD78beta isoform of MIP-1alpha is the most potent CC-chemokine in inhibiting CCR5-dependent human immunodeficiency virus type 1 replication in human macrophages." J Virol **75**(9): 4402-4406.
- Armour, J. A., R. Palla, et al. (2007). "Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats." Nucleic Acids Res **35**(3): e19.
- Arnett, F. C., S. M. Edworthy, et al. (1988). "The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis." Arthritis Rheum **31**(3): 315-324.
- Asquith, D. L. and I. B. McInnes (2007). "Emerging cytokine targets in rheumatoid arthritis." Curr Opin Rheumatol **19**(3): 246-251.
- Avouac, J., L. Gossec, et al. (2006). "Diagnostic and predictive value of anti-cyclic citrullinated protein antibodies in rheumatoid arthritis: a systematic literature review." Ann Rheum Dis **65**(7): 845-851.

-B-

- Bailey, A., W. Phillips, et al. (1996). "Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives." J Child Psychol Psychiatry **37**(1): 89-126.

- Bailey, J. A. and E. E. Eichler (2006). "Primate segmental duplications: crucibles of evolution, diversity and disease." Nat Rev Genet **7**(7): 552-564.
- Bailey, J. A., Z. Gu, et al. (2002). "Recent segmental duplications in the human genome." Science **297**(5583): 1003-1007.
- Balsa, A., A. Cabezón, et al. (2010). "Influence of HLA DRB1 alleles in the susceptibility of rheumatoid arthritis and the regulation of antibodies against citrullinated proteins and rheumatoid factor." Arthritis Res Ther **12**(2): R62.
- Bauters, M., H. Van Esch, et al. (2008). "Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair." Genome Res **18**(6): 847-858.
- Begovich, A. B., V. E. Carlton, et al. (2004). "A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis." Am J Hum Genet **75**(2): 330-337.
- Berek, C. and A. E. Schroder (1997). "A germinal center-like reaction in the nonlymphoid tissue of the synovial membrane." Ann N Y Acad Sci **815**: 211-217.
- Bergboer, J. G., M. Umicevic-Mirkov, et al. (2012). "A replication study of the association between rheumatoid arthritis and deletion of the late cornified envelope genes LCE3B and LCE3C." PLoS One **7**(2): e32045.
- Beri-Dexheimer, M., C. Bonnet, et al. (2007). "[Microarray-based comparative genomic hybridization in the study of constitutional chromosomal abnormalities]." Pathol Biol (Paris) **55**(1): 13-18.
- Berkman, N., M. John, et al. (1995). "Inhibition of macrophage inflammatory protein-1 alpha expression by IL-10. Differential sensitivities in human blood monocytes and alveolar macrophages." J Immunol **155**(9): 4412-4418.
- Bohanec Grabar, P., D. Logar, et al. (2009). "Genetic polymorphisms of glutathione S-transferases and disease activity of rheumatoid arthritis." Clin Exp Rheumatol **27**(2): 229-236.
- Brasch-Andersen, C., L. Christiansen, et al. (2004). "Possible gene dosage effect of glutathione-S-transferases on atopic asthma: using real-time PCR for quantification of GSTM1 and GSTT1 gene copy numbers." Hum Mutat **24**(3): 208-214.
- Brennan, P. and A. Silman (1994). "Breast-feeding and the onset of rheumatoid arthritis." Arthritis Rheum **37**(6): 808-813.
- Breunis, W. B., E. van Mirre, et al. (2008). "Copy number variation of the activating FCGR2C gene predisposes to idiopathic thrombocytopenic purpura." Blood **111**(3): 1029-1038.
- Breunis, W. B., E. van Mirre, et al. (2009). "Copy number variation at the FCGR locus includes FCGR3A, FCGR2C and FCGR3B but not FCGR2A and FCGR2B." Hum Mutat **30**(5): E640-650.
- Brighton, S. W., A. L. de la Harpe, et al. (1988). "The prevalence of rheumatoid arthritis in a rural African population." J Rheumatol **15**(3): 405-408.
- Brownlie, R. J., L. A. Miosge, et al. (2012). "Lack of the phosphatase PTPN22 increases adhesion of murine regulatory T cells to improve their immunosuppressive function." Sci Signal **5**(252): ra87.

- Buchard, A., J. J. Sanchez, et al. (2007). "Multiplex PCR detection of GSTM1, GSTT1, and GSTP1 gene variants: simultaneously detecting GSTM1 and GSTT1 gene copy number and the allelic status of the GSTP1 Ile105Val genetic variant." J Mol Diagn **9**(5): 612-617.
- Burmester, G. R., E. Feist, et al. (2014). "Emerging cell and cytokine targets in rheumatoid arthritis." Nat Rev Rheumatol **10**(2): 77-88.
- Butler, M. W., N. R. Hackett, et al. (2011). "Glutathione S-transferase copy number variation alters lung gene expression." Eur Respir J **38**(1): 15-28.

-C-

- Cabrejo, L., L. Guyant-Marechal, et al. (2006). "Phenotype associated with APP duplication in five families." Brain **129**(Pt 11): 2966-2976.
- Carames, B., M. J. Lopez-Armada, et al. (2008). "Differential effects of tumor necrosis factor-alpha and interleukin-1beta on cell death in human articular chondrocytes." Osteoarthritis Cartilage **16**(6): 715-722.
- Carpenter, D., A. Farnert, et al. (2012). "CCL3L1 copy number and susceptibility to malaria." Infect Genet Evol **12**(5): 1147-1154.
- Carpenter, D., S. Walker, et al. (2011). "Accuracy and differential bias in copy number measurement of CCL3L1 in association studies with three auto-immune disorders." BMC Genomics **12**: 418.
- Carter, N. P. (2007). "Methods and strategies for analyzing copy number variation using DNA microarrays." Nat Genet **39**(7 Suppl): S16-21.
- Carvalho, C. M., F. Zhang, et al. (2009). "Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching." Hum Mol Genet **18**(12): 2188-2203.
- Ceulemans, S., K. van der Ven, et al. (2012). "Targeted screening and validation of copy number variations." Methods Mol Biol **838**: 311-328.
- Charlesworth, B., M. T. Morgan, et al. (1993). "The effect of deleterious mutations on neutral molecular variation." Genetics **134**(4): 1289-1303.
- Chen CL, Liu Q, et al. (1993). "Higher frequency of glutathione S-transferase deletions in black children with acute lymphoblastic leukemia." Blood **89**(5): 1701-1707.
- Chen, J., F. Huang, et al. (2012). "Genetic polymorphism of glutathione S-transferase T1 and the risk of rheumatoid arthritis: a meta-analysis." Clin Exp Rheumatol **30**(5): 741-747.
- Chen, X., X. Li, et al. (2010). "Novel association strategy with copy number variation for identifying new risk Loci of human diseases." PLoS One **5**(8): e12185.
- Cheung, V. G., N. Nowak, et al. (2001). "Integration of cytogenetic landmarks into the draft sequence of the human genome." Nature **409**(6822): 953-958.
- Clements, J. L., B. Yang, et al. (1998). "Requirement for the leukocyte-specific adapter protein SLP-76 for normal T cell development." Science **281**(5375): 416-419.

- Coffey, G., F. DeGuzman, et al. (2012). "Specific inhibition of spleen tyrosine kinase suppresses leukocyte immune function and inflammation in animal models of rheumatoid arthritis." J Pharmacol Exp Ther **340**(2): 350-359.
- Colmegna, I., B. R. Ohata, et al. (2012). "Current understanding of rheumatoid arthritis therapy." Clin Pharmacol Ther **91**(4): 607-620.
- Conrad, D. F., T. D. Andrews, et al. (2006). "A high-resolution survey of deletion polymorphism in the human genome." Nat Genet **38**(1): 75-81.
- Conrad, D. F., D. Pinto, et al. (2010). "Origins and functional impact of copy number variation in the human genome." Nature **464**(7289): 704-712.
- Consort, T. I. S. (2008). "Rare chromosomal deletions and duplications increase risk of schizophrenia." Nature **455**(7210): 237-241.
- Consortium, I. H. (2003). "The International HapMap Project." Nature **426**(6968): 789-796.
- Consortium, I. H. (2005). "A haplotype map of the human genome." Nature **437**(7063): 1299-1320.
- Consortium, W. T. C. C. (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-678.
- Cooper, G. M., D. A. Nickerson, et al. (2007). "Mutational and selective effects on copy-number variants in the human genome." Nat Genet **39**(7 Suppl): S22-29.
- Cordell, H. J. and D. G. Clayton (2005). "Genetic association studies." Lancet **366**(9491): 1121-1131.
- Craddock, N., M. E. Hurles, et al. (2010). "Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls." Nature **464**(7289): 713-720.
- Criswell, L. A., K. G. Saag, et al. (2006). "Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women." Ann Rheum Dis **65**(9): 1163-1167.
- Croker, J. A. and R. P. Kimberly (2005). "Genetics of susceptibility and severity in systemic lupus erythematosus." Curr Opin Rheumatol **17**(5): 529-537.
- Cutolo, M., B. Seriolo, et al. (2002). "Androgens and estrogens modulate the immune and inflammatory responses in rheumatoid arthritis." Ann N Y Acad Sci **966**: 131-142.

-D-

- Daser, A., M. Thangavelu, et al. (2006). "Interrogation of genomes by molecular copy-number counting (MCC)." Nat Methods **3**(6): 447-453.
- de Haas, M., M. Kleijer, et al. (1995). "Neutrophil Fc gamma RIIIb deficiency, nature, and clinical consequences: a study of 21 individuals from 14 families." Blood **86**(6): 2403-2413.
- de Smith, A. J., A. Tsalenko, et al. (2007). "Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases." Hum Mol Genet **16**(23): 2783-2794.

- de Smith, A. J., R. G. Walters, et al. (2008). "Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease." Cytogenet Genome Res **123**(1-4): 17-26.
- De Vita, S., F. Zaja, et al. (2002). "Efficacy of selective B cell blockade in the treatment of rheumatoid arthritis: evidence for a pathogenetic role of B cells." Arthritis Rheum **46**(8): 2029-2033.
- Deighton, C. M., D. J. Walker, et al. (1989). "The contribution of HLA to rheumatoid arthritis." Clin Genet **36**(3): 178-182.
- del Gaudio, D., P. Fang, et al. (2006). "Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males." Genet Med **8**(12): 784-792.
- del Junco, D., H. S. Luthra, et al. (1984). "The familial aggregation of rheumatoid arthritis and its relationship to the HLA-DR4 association." Am J Epidemiol **119**(5): 813-829.
- Del Puente, A., W. C. Knowler, et al. (1989). "High incidence and prevalence of rheumatoid arthritis in Pima Indians." Am J Epidemiol **129**(6): 1170-1178.
- Despres, N., G. Talbot, et al. (1995). "Detection and expression of a cDNA clone that encodes a polypeptide containing two inhibitory domains of human calpastatin and its recognition by rheumatoid arthritis sera." J Clin Invest **95**(4): 1891-1896.
- Docampo, E., R. Rabionet, et al. (2010). "Deletion of the late cornified envelope genes, LCE3C and LCE3B, is associated with rheumatoid arthritis." Arthritis Rheum **62**(5): 1246-1251.
- Doran, M. F., G. R. Pond, et al. (2002). "Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period." Arthritis Rheum **46**(3): 625-631.
- Dunbar, S. A. (2006). "Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection." Clin Chim Acta **363**(1-2): 71-82.

-E-

- Elder, J. T. (2009). "Genome-wide association scan yields new insights into the immunopathogenesis of psoriasis." Genes Immun **10**(3): 201-209.
- Eliou, J. F., M. Andary, et al. (1988). "Increase of class II HLA molecules on the membrane of B lymphocytes from patients with rheumatoid arthritis." Autoimmunity **1**(3): 217-222.
- Emery, P., F. C. Breedveld, et al. (2008). "Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial." Lancet **372**(9636): 375-382.
- Emeville, E., C. Broquere, et al. (2014). "Copy Number Variation of GSTT1 and GSTM1 and the Risk of Prostate Cancer in a Caribbean Population of African Descent." PLoS One **9**(9): e107275.

Ezerzer, C. and N. Harris (2007). "Physiological immunity or pathological autoimmunity--a question of balance." Autoimmun Rev **6**(7): 488-496.

-F-

Fellermann, K., D. E. Stange, et al. (2006). "A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon." Am J Hum Genet **79**(3): 439-448.

Feuk, L., A. R. Carson, et al. (2006). "Structural variation in the human genome." Nat Rev Genet **7**(2): 85-97.

Field, S. F., J. M. Howson, et al. (2009). "Experimental aspects of copy number variant assays at CCL3L1." Nat Med **15**(10): 1115-1117.

Firestein, G. S. (2003). "Evolving concepts of rheumatoid arthritis." Nature **423**(6937): 356-361.

Firestein, G. S., J. M. Alvaro-Gracia, et al. (1990). "Quantitative analysis of cytokine gene expression in rheumatoid arthritis." J Immunol **144**(9): 3347-3353.

Flint, J., A. V. Hill, et al. (1986). "High frequencies of alpha-thalassaemia are the result of natural selection by malaria." Nature **321**(6072): 744-750.

Fossati, G., R. J. Moots, et al. (2002). "Differential role of neutrophil Fcgamma receptor IIIb (CD16) in phagocytosis, bacterial killing, and responses to immune complexes." Arthritis Rheum **46**(5): 1351-1361.

Francke, U., H. D. Ochs, et al. (1985). "Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa, and McLeod syndrome." Am J Hum Genet **37**(2): 250-267.

Frank, B., J. L. Bermejo, et al. (2007). "Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk." Carcinogenesis **28**(7): 1442-1445.

Freeman, J. L., G. H. Perry, et al. (2006). "Copy number variation: new insights in genome diversity." Genome Res **16**(8): 949-961.

-G-

Gabriel, S. E., C. S. Crowson, et al. (1999). "The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955-1985." Arthritis Rheum **42**(3): 415-420.

Gaffney, P. M., G. M. Kearns, et al. (1998). "A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families." Proc Natl Acad Sci U S A **95**(25): 14875-14879.

Garte, S., L. Gaspari, et al. (2001). "Metabolic gene polymorphism frequencies in control populations." Cancer Epidemiol Biomarkers Prev **10**(12): 1239-1248.

- Genovese, M. C., J. M. Bathon, et al. (2002). "Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiographic and clinical outcomes." Arthritis Rheum **46**(6): 1443-1450.
- Godessart, N. and S. L. Kunkel (2001). "Chemokines in autoimmune disease." Curr Opin Immunol **13**(6): 670-675.
- Gonzalez, E., H. Kulkarni, et al. (2005). "The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility." Science **307**(5714): 1434-1440.
- Gottar-Guillier, M., F. Dodeller, et al. (2011). "The tyrosine kinase BMX is an essential mediator of inflammatory arthritis in a kinase-independent manner." J Immunol **186**(10): 6014-6023.
- Graf, S. W., S. Lester, et al. (2012). "Low copy number of the FCGR3B gene and rheumatoid arthritis: a case-control study and meta-analysis." Arthritis Res Ther **14**(1): R28.
- Gregersen, P. K., J. Silver, et al. (1987). "The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis." Arthritis Rheum **30**(11): 1205-1213.
- Griffiths, R. J., M. A. Smith, et al. (1997). "Collagen-induced arthritis is reduced in 5-lipoxygenase-activating protein-deficient mice." J Exp Med **185**(6): 1123-1129.
- Gu, W., F. Zhang, et al. (2008). "Mechanisms for human genomic rearrangements." Pathogenetics **1**(1): 4.

-H-

- Halliwell, B. (1995). "Oxygen radicals, nitric oxide and human inflammatory joint disease." Ann Rheum Dis **54**(6): 505-510.
- Haraksingh, R. R., A. Abyzov, et al. (2011). "Genome-wide mapping of copy number variation in humans: comparative analysis of high resolution array platforms." PLoS One **6**(11): e27859.
- Hardwick, R. J., L. R. Machado, et al. (2011). "A worldwide analysis of beta-defensin copy number variation suggests recent selection of a high-expressing DEFB103 gene copy in East Asia." Hum Mutat **32**(7): 743-750.
- Harris, E., Ed. (2005). Clinical feature of rheumatoid arthritis. Philadelphia: WB Saunders.
- Harrison, P., L. Southam, et al. (2012). "Evidence of cis-acting regulatory variation in PTPN22 in patients with rheumatoid arthritis." Scand J Rheumatol **41**(4): 249-252.
- Harvey, J., M. Lotze, et al. (1981). "Rheumatoid arthritis in a Chippewa Band. I. Pilot screening study of disease prevalence." Arthritis Rheum **24**(5): 717-721.
- Hassfeld, W., G. Steiner, et al. (1989). "Demonstration of a new antinuclear antibody (anti-RA33) that is highly specific for rheumatoid arthritis." Arthritis Rheum **32**(12): 1515-1520.
- Hastings, P. J., J. R. Lupski, et al. (2009). "Mechanisms of change in gene copy number." Nat Rev Genet **10**(8): 551-564.

- Hauptmann, G., E. Grosshans, et al. (1974). "Lupus erythematosus syndrome and complete deficiency of the fourth component of complement." Boll Ist Sieroter Milan **53**(1): suppl:228.
- Hinds, D. A., A. P. Kloek, et al. (2006). "Common deletions and SNPs are in linkage disequilibrium in the human genome." Nat Genet **38**(1): 82-85.
- Hindson, B. J., K. D. Ness, et al. (2011). "High-throughput droplet digital PCR system for absolute quantitation of DNA copy number." Anal Chem **83**(22): 8604-8610.
- Hirayasu, K., J. Ohashi, et al. (2008). "Evidence for natural selection on leukocyte immunoglobulin-like receptors for HLA class I in Northeast Asians." Am J Hum Genet **82**(5): 1075-1083.
- Hollox, E. J. and J. A. Armour (2008). "Directional and balancing selection in human beta-defensins." BMC Evol Biol **8**: 113.
- Hollox, E. J., J. A. Armour, et al. (2003). "Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster." Am J Hum Genet **73**(3): 591-600.
- Hollox, E. J., U. Huffmeier, et al. (2008). "Psoriasis is associated with increased beta-defensin genomic copy number." Nat Genet **40**(1): 23-25.
- Hoogendijk, J. E., G. W. Hensels, et al. (1992). "De-novo mutation in hereditary motor and sensory neuropathy type I." Lancet **339**(8801): 1081-1082.
- Horton, R., R. Gibson, et al. (2008). "Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project." Immunogenetics **60**(1): 1-18.
- Houssien, D. A., T. Jonsson, et al. (1997). "Clinical significance of IgA rheumatoid factor subclasses in rheumatoid arthritis." J Rheumatol **24**(11): 2119-2122.
- Huggett, J. F., C. A. Foy, et al. (2013). "The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments." Clin Chem **59**(6): 892-902.
- Hultqvist, M., P. Olofsson, et al. (2006). "A new arthritis therapy with oxidative burst inducers." PLoS Med **3**(9): e348.
- Hutchinson, D. and R. Moots (2001). "Cigarette smoking and severity of rheumatoid arthritis." Rheumatology (Oxford) **40**(12): 1426-1427.

-I-

- Iafate, A. J., L. Feuk, et al. (2004). "Detection of large-scale variation in the human genome." Nat Genet **36**(9): 949-951.
- Ignatowicz, E., A. Wozniak, et al. (2013). "Exposure to alcohol and tobacco smoke causes oxidative stress in rats." Pharmacol Rep **65**(4): 906-913.
- Inoue, A., I. Matsumoto, et al. (2009). "Tumor necrosis factor alpha-induced adipose-related protein expression in experimental arthritis and in rheumatoid arthritis." Arthritis Res Ther **11**(4): R118.
- Itsara, A., H. Wu, et al. (2010). "De novo rates and selection of large copy number variation." Genome Res **20**(11): 1469-1481.

-J-

- Jancova, P., P. Anzenbacher, et al. (2010). "Phase II drug metabolizing enzymes." Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub **154**(2): 103-116.
- Jawaheer, D., W. Thomson, et al. (1994). "'Homozygosity" for the HLA-DR shared epitope contributes the highest risk for rheumatoid arthritis concordance in identical twins." Arthritis Rheum **37**(5): 681-686.
- Ji, J. D. and W. J. Lee (2013). "Association between the polymorphisms of glutathione S-transferase genes and rheumatoid arthritis: a meta-analysis." Gene **521**(1): 155-159.
- Johansson, I., E. Lundqvist, et al. (1993). "Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine." Proc Natl Acad Sci U S A **90**(24): 11825-11829.
- John, S. and J. Worthington (2001). "Genetic epidemiology. Approaches to the genetic analysis of rheumatoid arthritis." Arthritis Res **3**(4): 216-220.
- Jonsson, H., P. Allen, et al. (2005). "Inflammatory arthritis requires Foxo3a to prevent Fas ligand-induced neutrophil apoptosis." Nat Med **11**(6): 666-671.
- Jonsson, T., K. Steinsson, et al. (1998). "Combined elevation of IgM and IgA rheumatoid factor has high diagnostic specificity for rheumatoid arthritis." Rheumatol Int **18**(3): 119-122.

-K-

- Kallioniemi, A., O. P. Kallioniemi, et al. (1992). "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors." Science **258**(5083): 818-821.
- Keenan, B. T., L. B. Chibnik, et al. (2010). "Effect of interactions of glutathione S-transferase T1, M1, and P1 and HMOX1 gene promoter polymorphisms with heavy smoking on the risk of rheumatoid arthritis." Arthritis Rheum **62**(11): 3196-3210.
- Kidd, J. M., T. L. Newman, et al. (2007). "Population stratification of a common APOBEC gene deletion polymorphism." PLoS Genet **3**(4): e63.
- Klareskog, L., J. Ronnelid, et al. (2008). "Immunity to citrullinated proteins in rheumatoid arthritis." Annu Rev Immunol **26**: 651-675.
- Klinman, D. (2003). "Does activation of the innate immune system contribute to the development of rheumatoid arthritis?" Arthritis Rheum **48**(3): 590-593.
- Koene, H. R., M. de Haas, et al. (1996). "NA-phenotype-dependent differences in neutrophil Fc gamma RIIIb expression cause differences in plasma levels of soluble Fc gamma RIII." Br J Haematol **93**(1): 235-241.
- Korb, A., H. Pavenstadt, et al. (2009). "Cell death in rheumatoid arthritis." Apoptosis **14**(4): 447-454.
- Korbel, J. O., A. E. Urban, et al. (2007). "Paired-end mapping reveals extensive structural variation in the human genome." Science **318**(5849): 420-426.

Kryukov, G. V., L. A. Pennacchio, et al. (2007). "Most rare missense alleles are deleterious in humans: implications for complex disease and association studies." Am J Hum Genet **80**(4): 727-739.

-L-

Lander, E. S. and N. J. Schork (1994). "Genetic dissection of complex traits." Science **265**(5181): 2037-2048.

Landi, S. (2000). "Mammalian class theta GST and differential susceptibility to carcinogens: a review." Mutat Res **463**(3): 247-283.

Larsen, A. (1995). "to apply Larsen score in evaluating radiographs of rheumatoid arthritis in long term studies." J Rheumatol **22**: 1974-1975.

Ledbetter, D. H., V. M. Riccardi, et al. (1981). "Deletions of chromosome 15 as a cause of the Prader-Willi syndrome." N Engl J Med **304**(6): 325-329.

Lederman, M. M. and S. F. Sieg (2007). "CCR5 and its ligands: a new axis of evil?" Nat Immunol **8**(12): 1283-1285.

Lee, J. A., C. M. Carvalho, et al. (2007). "A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders." Cell **131**(7): 1235-1247.

Lee, J. H. and J. T. Jeon (2008). "Methods to detect and analyze copy number variations at the genome-wide and locus-specific levels." Cytogenet Genome Res **123**(1-4): 333-342.

Li, X., T. S. Ptacek, et al. (2009). "Fcgamma receptors: structure, function and role as genetic risk factors in SLE." Genes Immun **10**(5): 380-389.

Lieber, M. R. (2008). "The mechanism of human nonhomologous DNA end joining." J Biol Chem **283**(1): 1-5.

Lieber, M. R., H. Lu, et al. (2008). "Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: relevance to cancer, aging, and the immune system." Cell Res **18**(1): 125-133.

Lieber, M. R., Y. Ma, et al. (2003). "Mechanism and regulation of human non-homologous DNA end-joining." Nat Rev Mol Cell Biol **4**(9): 712-720.

Liu, H. and R. M. Pope (2004). "Apoptosis in rheumatoid arthritis: friend or foe." Rheum Dis Clin North Am **30**(3): 603-625, x.

Liu, P., M. Lacia, et al. (2011). "Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over." Am J Hum Genet **89**(4): 580-588.

Locke, D. P., A. J. Sharp, et al. (2006). "Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome." Am J Hum Genet **79**(2): 275-290.

Lu, X., J. Guo, et al. (2011). "Deletion of LCE3C_LCE3B is associated with rheumatoid arthritis and systemic lupus erythematosus in the Chinese Han population." Ann Rheum Dis **70**(9): 1648-1651.

- Lundstrom, E., T. Hartshorne, et al. (2011). "Effects of GSTM1 in rheumatoid arthritis; results from the Swedish EIRA study." PLoS One **6**(3): e17880.
- Lupski, J. R., R. M. de Oca-Luna, et al. (1991). "DNA duplication associated with Charcot-Marie-Tooth disease type 1A." Cell **66**(2): 219-232.

-M-

- MacGregor A, S. A., Ed. (2003). Classification and epidemiology. Mosby: London.
- MacGregor, A. J., H. Snieder, et al. (2000). "Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins." Arthritis Rheum **43**(1): 30-37.
- Machado, L. R., R. J. Hardwick, et al. (2012). "Evolutionary history of copy-number-variable locus for the low-affinity Fcγ receptor: mutation rate, autoimmune disease, and the legacy of helminth infection." Am J Hum Genet **90**(6): 973-985.
- Maine, C. J., E. E. Hamilton-Williams, et al. (2012). "PTPN22 alters the development of regulatory T cells in the thymus." J Immunol **188**(11): 5267-5275.
- Malemud, C. J. (2010). "Differential activation of JAK enzymes in rheumatoid arthritis and autoimmune disorders by proinflammatory cytokines—potential drug targets." Int J Interferon Cytokine Mediator Res **2**: 97-111.
- Malemud, C. J. (2011). "Apoptosis resistance in rheumatoid arthritis synovial tissue." J Clin Cell Immunol **S3**.
- Malemud, C. J. (2013). "Intracellular Signaling Pathways in Rheumatoid Arthritis." J Clin Cell Immunol **4**: 160.
- Malemud, C. J. and E. Pearlman (2009). "Targeting JAK/STAT signaling pathway in inflammatory diseases." Curr Signal Transduct Ther **4**: 201-221.
- Malemud, C. J. and S. K. Reddy (2008). "Targeting cytokines, chemokines and adhesion molecules in rheumatoid arthritis." Curr Rheum Res **4**: 219-234.
- Malemud, C. J., Y. Sun, et al. (2012). "Monosodium Urate and Tumor Necrosis Factor-α Increase Apoptosis in Human Chondrocyte Cultures." Rheumatology (Sunnyvale) **2**: 113.
- Mamtani, M., B. Rovin, et al. (2008). "CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythematosis." Ann Rheum Dis **67**(8): 1076-1083.
- Marian, A. J. (2012). "Molecular genetic studies of complex phenotypes." Transl Res **159**(2): 64-79.
- Marques, R. B., M. M. Thabet, et al. (2010). "Genetic variation of the Fcγ receptor 3B gene and association with rheumatoid arthritis." PLoS One **5**(10).
- Mattey, D. L., A. B. Hassell, et al. (1999). "Association of polymorphism in glutathione S-transferase loci with susceptibility and outcome in rheumatoid arthritis: comparison with the shared epitope." Ann Rheum Dis **58**(3): 164-168.

- Mattey, D. L., D. Hutchinson, et al. (2002). "Smoking and disease severity in rheumatoid arthritis: association with polymorphism at the glutathione S-transferase M1 locus." Arthritis Rheum **46**(3): 640-646.
- Mayer, B. A., M. Rehberg, et al. (2011). "Inhibitor of apoptosis proteins as novel targets in inflammatory processes." Arterioscler Thromb Vasc Biol **31**(10): 2240-2250.
- McCarroll, S. A., T. N. Hadnott, et al. (2006). "Common deletion polymorphisms in the human genome." Nat Genet **38**(1): 86-92.
- McCarroll, S. A., F. G. Kuruvilla, et al. (2008). "Integrated detection and population-genetic analysis of SNPs and copy number variation." Nat Genet **40**(10): 1166-1174.
- McCaughan, F. (2009). "Molecular copy-number counting: potential of single-molecule diagnostics." Expert Rev Mol Diagn **9**(4): 309-312.
- McCaughan, F. and P. H. Dear (2010). "Single-molecule genomics." J Pathol **220**(2): 297-306.
- McKinney, C., M. Fanciulli, et al. (2010). "Association of variation in Fcgamma receptor 3B gene copy number with rheumatoid arthritis in Caucasian samples." Ann Rheum Dis **69**(9): 1711-1716.
- McKinney, C., M. E. Merriman, et al. (2008). "Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis." Ann Rheum Dis **67**(3): 409-413.
- McKinney, C. and T. R. Merriman (2012). "Meta-analysis confirms a role for deletion in FCGR3B in autoimmune phenotypes." Hum Mol Genet **21**(10): 2370-2376.
- McLean-Tooke, A., G. P. Spickett, et al. (2007). "Immunodeficiency and autoimmunity in 22q11.2 deletion syndrome." Scand J Immunol **66**(1): 1-7.
- McLean, C. Y., P. L. Reno, et al. (2011). "Human-specific loss of regulatory DNA and the evolution of human-specific traits." Nature **471**(7337): 216-219.
- McLellan, R. A., M. Oscarson, et al. (1997). "Characterization of a human glutathione S-transferase mu cluster containing a duplicated GSTM1 gene that causes ultrarapid enzyme activity." Mol Pharmacol **52**(6): 958-965.
- McPherson, J. D., M. Marra, et al. (2001). "A physical map of the human genome." Nature **409**(6822): 934-941.
- Menten, P., S. Struyf, et al. (1999). "The LD78beta isoform of MIP-1alpha is the most potent CCR5 agonist and HIV-1-inhibiting chemokine." J Clin Invest **104**(4): R1-5.
- Merriman, T., R. Twells, et al. (1997). "Evidence by allelic association-dependent methods for a type 1 diabetes polygene (IDDM6) on chromosome 18q21." Hum Mol Genet **6**(7): 1003-1010.
- Meyer, D. J., B. Coles, et al. (1991). "Theta, a new class of glutathione transferases purified from rat and man." Biochem J **274** (Pt 2): 409-414.
- Meyer, K. (1922). "Immunitätsforsch." Exp Ther **34**: 229-234.
- Michou, L., V. H. Teixeira, et al. (2008). "Associations between genetic factors, tobacco smoking and autoantibodies in familial and sporadic rheumatoid arthritis." Ann Rheum Dis **67**(4): 466-470.

- Mills, R. E., K. Walter, et al. (2011). "Mapping copy number variation by population-scale genome sequencing." Nature **470**(7332): 59-65.
- Mimori, T. (2005). "Clinical significance of anti-CCP antibodies in rheumatoid arthritis." Intern Med **44**(11): 1122-1126.
- Mimori, T., K. Suganuma, et al. (1995). "Autoantibodies to calpastatin (an endogenous inhibitor for calcium-dependent neutral protease, calpain) in systemic rheumatic diseases." Proc Natl Acad Sci U S A **92**(16): 7267-7271.
- Miossec, P. (2003). "Interleukin-17 in rheumatoid arthritis: if T cells were to contribute to inflammation and destruction through synergy." Arthritis Rheum **48**(3): 594-601.
- Modi, W. S. (2004). "CCL3L1 and CCL4L1 chemokine genes are located in a segmental duplication at chromosome 17q12." Genomics **83**(4): 735-738.
- Morel, J., P. Miossec, et al. (2004). "Immunopatogenesis of rheumatoid arthritis." EMC-Rhumatologie Orthopédie **1**: 218-230.
- Morgan, A. W., J. H. Barrett, et al. (2006). "Analysis of Fcgamma receptor haplotypes in rheumatoid arthritis: FCGR3A remains a major susceptibility gene at this locus, with an additional contribution from FCGR3B." Arthritis Res Ther **8**(1): R5.
- Morgan, A. W., B. Griffiths, et al. (2000). "Fcgamma receptor type IIIA is associated with rheumatoid arthritis in two distinct ethnic groups." Arthritis Rheum **43**(10): 2328-2334.
- Morris, D. L., A. L. Roberts, et al. (2010). "Evidence for both copy number and allelic (NA1/NA2) risk at the FCGR3B locus in systemic lupus erythematosus." Eur J Hum Genet **18**(9): 1027-1031.
- Mottonen, T. T., P. J. Hannonen, et al. (1999). "Combination DMARD therapy including corticosteroids in early rheumatoid arthritis." Clin Exp Rheumatol **17**(6 Suppl 18): S59-65.
- Nakaoka, H., T. Cui, et al. (2011). "A systems genetics approach provides a bridge from discovered genetic variants to biological pathways in rheumatoid arthritis." PLoS One **6**(9): e25389.

-N-

- Nguyen, D. Q., C. Webber, et al. (2006). "Bias of selection on human copy-number variants." PLoS Genet **2**(2): e20.
- Niederer, H. A., L. C. Willcocks, et al. (2010). "Copy number, linkage disequilibrium and disease association in the FCGR locus." Hum Mol Genet **19**(16): 3282-3294.
- Nienhuis, R. L. and E. Mandema (1964). "A New Serum Factor in Patients with Rheumatoid Arthritis; the Antiperinuclear Factor." Ann Rheum Dis **23**: 302-305.
- Nimmerjahn, F. and J. V. Ravetch (2007). "Fc-receptors as regulators of immunity." Adv Immunol **96**: 179-204.
- Nishimura, K., D. Sugiyama, et al. (2007). "Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis." Ann Intern Med **146**(11): 797-808.

- Nobile, C., L. Toffolatti, et al. (2002). "Analysis of 22 deletion breakpoints in dystrophin intron 49." Hum Genet **110**(5): 418-421.
- Nordang, G. B., D. Carpenter, et al. (2012). "Association analysis of the CCL3L1 copy number locus by paralogue ratio test in Norwegian rheumatoid arthritis patients and healthy controls." Genes Immun **13**(7): 579-582.
- Norskov, M. S., R. Frikke-Schmidt, et al. (2011). "Copy number variation in glutathione-S-transferase T1 and M1 predicts incidence and 5-year survival from prostate and bladder cancer, and incidence of corpus uteri cancer in the general population." Pharmacogenomics J **11**(4): 292-299.
- Norskov, M. S., R. Frikke-Schmidt, et al. (2011). "Copy number variation in glutathione S-transferases M1 and T1 and ischemic vascular disease: four studies and meta-analyses." Circ Cardiovasc Genet **4**(4): 418-428.
- Norskov, M. S., R. Frikke-Schmidt, et al. (2009). "High-throughput genotyping of copy number variation in glutathione S-transferases M1 and T1 using real-time PCR in 20,687 individuals." Clin Biochem **42**(3): 201-209.

-O-

- Okada, Y., C. Terao, et al. (2012). "Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population." Nat Genet **44**(5): 511-516.
- Okada, Y., D. Wu, et al. (2014). "Genetics of rheumatoid arthritis contributes to biology and drug discovery." Nature **506**(7488): 376-381.
- Olofsson, P., J. Holmberg, et al. (2003). "Positional identification of Ncf1 as a gene that regulates arthritis severity in rats." Nat Genet **33**(1): 25-32.
- Olsson, L. M., A. Nerstedt, et al. (2012). "Copy number variation of the gene NCF1 is associated with rheumatoid arthritis." Antioxid Redox Signal **16**(1): 71-78.
- Oppenheimer-Marks, N., L. S. Davis, et al. (1991). "Differential utilization of ICAM-1 and VCAM-1 during the adhesion and transendothelial migration of human T lymphocytes." J Immunol **147**(9): 2913-2921.
- Orozco, G., D. Pascual-Salcedo, et al. (2008). "Auto-antibodies, HLA and PTPN22: susceptibility markers for rheumatoid arthritis." Rheumatology (Oxford) **47**(2): 138-141.
- Orozco, G., B. Rueda, et al. (2006). "Genetic basis of rheumatoid arthritis." Biomed Pharmacother **60**(10): 656-662.
- Ostalowska, A., E. Birkner, et al. (2006). "Lipid peroxidation and antioxidant enzymes in synovial fluid of patients with primary and secondary osteoarthritis of the knee joint." Osteoarthritis Cartilage **14**(2): 139-145.

-P-

- Pai, S., L. Pai, et al. (1998). "Correlation of serum IgA rheumatoid factor levels with disease severity in rheumatoid arthritis." Scand J Rheumatol **27**(4): 252-256.
- Panayi, G. S., V. M. Corrigall, et al. (2001). "Pathogenesis of rheumatoid arthritis. The role of T cells and other beasts." Rheum Dis Clin North Am **27**(2): 317-334.
- Pap, T., J. K. Franz, et al. (2000). "Activation of synovial fibroblasts in rheumatoid arthritis: lack of Expression of the tumour suppressor PTEN at sites of invasive growth and destruction." Arthritis Res **2**(1): 59-64.
- Park, S. S., P. Stankiewicz, et al. (2002). "Structure and evolution of the Smith-Magenis syndrome repeat gene clusters, SMS-REPs." Genome Res **12**(5): 729-738.
- Parl, F. F. (2005). "Glutathione S-transferase genotypes and cancer risk." Cancer Lett **221**(2): 123-129.
- Payami, H., J. Nutt, et al. (2003). "SCA2 may present as levodopa-responsive parkinsonism." Mov Disord **18**(4): 425-429.
- Pelkonen, P., P. Lahdenne, et al. (2002). "Chronic arthritis associated with chromosome deletion 22q11.2 syndrome." J Rheumatol **29**(12): 2648-2650.
- Pettit, A. R. and R. Thomas (1999). "Dendritic cells: the driving force behind autoimmunity in rheumatoid arthritis?" Immunol Cell Biol **77**(5): 420-427.
- Piotrowski, A., C. E. Bruder, et al. (2008). "Somatic mosaicism for copy number variation in differentiated human tissues." Hum Mutat **29**(9): 1118-1124.
- Plenge, R. M., C. Cotsapas, et al. (2007). "Two independent alleles at 6q23 associated with risk of rheumatoid arthritis." Nat Genet **39**(12): 1477-1482.
- Pope, R. M. (2002). "Apoptosis as a therapeutic tool in rheumatoid arthritis." Nat Rev Immunol **2**(7): 527-535.
- Potocki, L., W. Bi, et al. (2007). "Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype." Am J Hum Genet **80**(4): 633-649.

-Q-

- Qin, J., R. C. Jones, et al. (2008). "Studying copy number variations using a nanofluidic platform." Nucleic Acids Res **36**(18): e116.

-R-

- Raychaudhuri, S., E. F. Remmers, et al. (2008). "Common variants at CD40 and other loci confer risk of rheumatoid arthritis." Nat Genet **40**(10): 1216-1223.
- Razani, B., A. D. Reichardt, et al. (2011). "Non-canonical NF-kappaB signaling activation and regulation: principles and perspectives." Immunol Rev **244**(1): 44-54.
- Reddy, K. V., K. J. Serio, et al. (2003). "5-lipoxygenase-activating protein gene expression. Key role of CCAAT/enhancer-binding proteins (C/EBP) in constitutive and tumor

- necrosis factor (TNF) alpha-induced expression in THP-1 cells." J Biol Chem **278**(16): 13810-13818.
- Redon, R. and N. P. Carter (2009). "Comparative genomic hybridization: microarray design and data interpretation." Methods Mol Biol **529**: 37-49.
- Redon, R., S. Ishikawa, et al. (2006). "Global variation in copy number in the human genome." Nature **444**(7118): 444-454.
- Reiter, L. T., P. J. Hastings, et al. (1998). "Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients." Am J Hum Genet **62**(5): 1023-1033.
- Rhee, I. and A. Veillette (2012). "Protein tyrosine phosphatases in lymphocyte activation and autoimmunity." Nat Immunol **13**(5): 439-447.
- Robinson, J. I., I. M. Carr, et al. (2012). "Confirmation of association of FCGR3B but not FCGR3A copy number with susceptibility to autoantibody positive rheumatoid arthritis." Hum Mutat **33**(4): 741-749.
- Roeleveld, N., G. A. Zielhuis, et al. (1997). "The prevalence of mental retardation: a critical review of recent literature." Dev Med Child Neurol **39**(2): 125-132.
- Ronninger, M., Y. Guo, et al. (2012). "The balance of expression of PTPN22 splice forms is significantly different in rheumatoid arthritis patients compared with controls." Genome Med **4**(1): 2.
- Rose-Zerilli, M. J., S. J. Barton, et al. (2009). "Copy-number variation genotyping of GSTT1 and GSTM1 gene deletions by real-time PCR." Clin Chem **55**(9): 1680-1685.
- Rossin, E. J., K. Lage, et al. (2011). "Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology." PLoS Genet **7**(1): e1001273.
- Roth, D. B., T. N. Porter, et al. (1985). "Mechanisms of nonhomologous recombination in mammalian cells." Mol Cell Biol **5**(10): 2599-2607.
- Rovelet-Lecrux, A., D. Hannequin, et al. (2006). "APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy." Nat Genet **38**(1): 24-26.
- Russo-Marie, F. (1996). "La réaction inflammatoire dans la polyarthrite rhumatoïde." Méd Théor **2**: 261-266.

-S-

- Salmon, J. E., J. C. Edberg, et al. (1992). "Allelic polymorphisms of human Fc gamma receptor IIA and Fc gamma receptor IIIB. Independent mechanisms for differences in human phagocyte function." J Clin Invest **89**(4): 1274-1281.
- Salmon, J. E., J. C. Edberg, et al. (1990). "Fc gamma receptor III on human neutrophils. Allelic variants have functionally distinct capacities." J Clin Invest **85**(4): 1287-1295.
- Sany J, C. B., Jorgensen C., Ed. (1997). Polyarthrite rhumatoïde de l'adulte (I). Aspects cliniques. Appareil locomoteur. Paris, ElsevierSAS, Paris.

- Saraux, A., C. Guedes, et al. (1999). "Prevalence of rheumatoid arthritis and spondyloarthropathy in Brittany, France. Societe de Rhumatologie de l'Ouest." J Rheumatol **26**(12): 2622-2627.
- Schaid, D. J. and S. S. Sommer (1994). "Comparison of statistics for candidate-gene association studies using cases and parents." Am J Hum Genet **55**(2): 402-409.
- Schaschl, H., T. J. Aitman, et al. (2009). "Copy number variation in the human genome and its implication in autoimmunity." Clin Exp Immunol **156**(1): 12-16.
- Schellekens, G. A., B. A. de Jong, et al. (1998). "Citruiline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies." J Clin Invest **101**(1): 273-281.
- Schellekens, G. A., H. Visser, et al. (2000). "The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide." Arthritis Rheum **43**(1): 155-163.
- Scherer, S. W., J. Cheung, et al. (2003). "Human chromosome 7: DNA sequence and biology." Science **300**(5620): 767-772.
- Schlattl, A., S. Anders, et al. (2011). "Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions." Genome Res **21**(12): 2004-2013.
- Schmickel, R. D. (1986). "Contiguous gene syndromes: a component of recognizable syndromes." J Pediatr **109**(2): 231-241.
- Schouten, J. P., C. J. McElgunn, et al. (2002). "Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification." Nucleic Acids Res **30**(12): e57.
- Schwarz, K., Y. Ma, et al. (2003). "Human severe combined immune deficiency and DNA repair." Bioessays **25**(11): 1061-1070.
- Sebat, J., B. Lakshmi, et al. (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-528.
- Sebbag, M., M. Simon, et al. (1995). "The antiperinuclear factor and the so-called antikeratin antibodies are the same rheumatoid arthritis-specific autoantibodies." J Clin Invest **95**(6): 2672-2679.
- Seldin, M. F., C. I. Amos, et al. (1999). "The genetics revolution and the assault on rheumatoid arthritis." Arthritis Rheum **42**(6): 1071-1079.
- Semple, C. A., M. Rolfe, et al. (2003). "Duplication and selection in the evolution of primate beta-defensin genes." Genome Biol **4**(5): R31.
- Sennvik, K., J. Fastbom, et al. (2000). "Levels of alpha- and beta-secretase cleaved amyloid precursor protein in the cerebrospinal fluid of Alzheimer's disease patients." Neurosci Lett **278**(3): 169-172.
- Shai, R., F. P. Quismorio, Jr., et al. (1999). "Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families." Hum Mol Genet **8**(4): 639-644.
- Shaw, C. J. and J. R. Lupski (2004). "Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease." Hum Mol Genet **13 Spec No 1**: R57-64.

- Shen, L., L. C. Wu, et al. (1994). "Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication." *J Biol Chem* **269**(11): 8466-8476.
- Shengqi, W., W. Xiaohong, et al. (2002). "A new fluorescent quantitative polymerase chain reaction technique." *Anal Biochem* **309**(2): 206-211.
- Shlien, A. and D. Malkin (2009). "Copy number variations and cancer." *Genome Med* **1**(6): 62.
- Silman, A., A. Kay, et al. (1992). "Timing of pregnancy in relation to the onset of rheumatoid arthritis." *Arthritis Rheum* **35**(2): 152-155.
- Silman, A. J., A. J. MacGregor, et al. (1993). "Twin concordance rates for rheumatoid arthritis: results from a nationwide study." *Br J Rheumatol* **32**(10): 903-907.
- Silman, A. J., W. Ollier, et al. (1993). "Absence of rheumatoid arthritis in a rural Nigerian population." *J Rheumatol* **20**(4): 618-622.
- Simmons, D. and B. Seed (1988). "The Fc gamma receptor of natural killer cells is a phospholipid-linked membrane protein." *Nature* **333**(6173): 568-570.
- Simon, M., E. Girbal, et al. (1993). "The cytokeratin filament-aggregating protein filaggrin is the target of the so-called "antikeratin antibodies," autoantibodies specific for rheumatoid arthritis." *J Clin Invest* **92**(3): 1387-1393.
- Simonsson, M., S. Bergman, et al. (1999). "The prevalence of rheumatoid arthritis in Sweden." *Scand J Rheumatol* **28**(6): 340-343.
- Simpson, D. A., S. Feeney, et al. (2000). "Retinal VEGF mRNA measured by SYBR green I fluorescence: A versatile approach to quantitative PCR." *Mol Vis* **6**: 178-183.
- Singleton, A. B., M. Farrer, et al. (2003). "alpha-Synuclein locus triplication causes Parkinson's disease." *Science* **302**(5646): 841.
- Slager, R. E., T. L. Newton, et al. (2003). "Mutations in RAI1 associated with Smith-Magenis syndrome." *Nat Genet* **33**(4): 466-468.
- Smolen, J. S. and G. Steiner (2003). "Therapeutic strategies for rheumatoid arthritis." *Nat Rev Drug Discov* **2**(6): 473-488.
- Snijders, A. M., D. Pinkel, et al. (2003). "Current status and future prospects of array-based comparative genomic hybridisation." *Brief Funct Genomic Proteomic* **2**(1): 37-45.
- Solinas, A., L. J. Brown, et al. (2001). "Duplex Scorpion primers in SNP analysis and FRET applications." *Nucleic Acids Res* **29**(20): E96.
- Song, G. G., S. C. Bae, et al. (2012). "The glutathione S-transferase M1 and P1 polymorphisms and rheumatoid arthritis: a meta-analysis." *Mol Biol Rep* **39**(12): 10739-10745.
- Sprenger, R., R. Schlagenhauser, et al. (2000). "Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype-phenotype correlation." *Pharmacogenetics* **10**(6): 557-565.
- Stahl, E. A., S. Raychaudhuri, et al. (2010). "Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci." *Nat Genet* **42**(6): 508-514.

- Standiford, T. J., S. L. Kunkel, et al. (1993). "Gene expression of macrophage inflammatory protein-1 alpha from human blood monocytes and alveolar macrophages is inhibited by interleukin-4." Am J Respir Cell Mol Biol **9**(2): 192-198.
- Stankiewicz, P. and J. R. Lupski (2002). "Genome architecture, rearrangements and genomic disorders." Trends Genet **18**(2): 74-82.
- Stankiewicz, P. and J. R. Lupski (2006). "The genomic basis of disease, mechanisms and assays for genomic disorders." Genome Dyn **1**: 1-16.
- Stastny, P. (1976). "Mixed lymphocyte cultures in rheumatoid arthritis." J Clin Invest **57**(5): 1148-1157.
- Stefansson, H., D. Rujescu, et al. (2008). "Large recurrent microdeletions associated with schizophrenia." Nature **455**(7210): 232-236.
- Steinman, L. (2007). "A brief history of T(H)17, the first major revision in the T(H)1/T(H)2 hypothesis of T cell-mediated tissue damage." Nat Med **13**(2): 139-145.
- Steinmann, K., D. N. Cooper, et al. (2007). "Type 2 NF1 deletions are highly unusual by virtue of the absence of nonallelic homologous recombination hotspots and an apparent preference for female mitotic recombination." Am J Hum Genet **81**(6): 1201-1220.
- Stolt, P., C. Bengtsson, et al. (2003). "Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases." Ann Rheum Dis **62**(9): 835-841.
- Stranger, B. E., M. S. Forrest, et al. (2007). "Relative impact of nucleotide and copy number variation on gene expression phenotypes." Science **315**(5813): 848-853.
- Sudmant, P. H., J. O. Kitzman, et al. (2010). "Diversity of human copy number variation and multicopy genes." Science **330**(6004): 641-646.
- Sullivan, K. E., D. M. McDonald-McGinn, et al. (1997). "Juvenile rheumatoid arthritis-like polyarthritis in chromosome 22q11.2 deletion syndrome (DiGeorge anomalad/velocardiofacial syndrome/conotruncal anomaly face syndrome)." Arthritis Rheum **40**(3): 430-436.
- Surapneni, K. M. and V. S. Chandrasada Gopan (2008). "Lipid peroxidation and antioxidant status in patients with rheumatoid arthritis." Indian J Clin Biochem **23**(1): 41-44.
- Svensson, B., A. Boonen, et al. (2005). "Low-dose prednisolone in addition to the initial disease-modifying antirheumatic drug in patients with early active rheumatoid arthritis reduces joint destruction and increases the remission rate: a two-year randomized trial." Arthritis Rheum **52**(11): 3360-3370.
- Swedler, W., J. Wallman, et al. (1997). "Routine measurement of IgM, IgG, and IgA rheumatoid factors: high sensitivity, specificity, and predictive value for rheumatoid arthritis." J Rheumatol **24**(6): 1037-1044.
- Symmons, D. P. (2001). "Knee pain in older adults: the latest musculoskeletal "epidemic"." Ann Rheum Dis **60**(2): 89-90.
- Szekanecz, Z. and A. E. Koch (2007). "Macrophages and their products in rheumatoid arthritis." Curr Opin Rheumatol **19**(3): 289-295.

Szomor, Z., K. Shimizu, et al. (1995). "Appearance of calpain correlates with arthritis and cartilage destruction in collagen induced arthritic knee joints of mice." Ann Rheum Dis **54**(6): 477-483.

-T-

Tan, R., Y. Wang, et al. (2014). "An evaluation of copy number variation detection tools from whole-exome sequencing data." Hum Mutat **35**(7): 899-907.

Teo, S. M., Y. Pawitan, et al. (2012). "Statistical challenges associated with detecting copy number variations with next-generation sequencing." Bioinformatics **28**(21): 2711-2718.

Tighe, H., P. P. Chen, et al. (1993). "Function of B cells expressing a human immunoglobulin M rheumatoid factor autoantibody in transgenic mice." J Exp Med **177**(1): 109-118.

Timofeeva, M., B. Jager, et al. (2009). "A multiplex real-time PCR method for detection of GSTM1 and GSTT1 copy numbers." Clin Biochem **42**(6): 500-509.

Toffolatti, L., B. Cardazzo, et al. (2002). "Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene." Genomics **80**(5): 523-530.

Tuzun, E., A. J. Sharp, et al. (2005). "Fine-scale structural variation of the human genome." Nat Genet **37**(7): 727-732.

-U-

Uddin, M., M. Sturge, et al. (2011). "Autosome-wide copy number variation association analysis for rheumatoid arthritis using the WTCCC high-density SNP genotype data." J Rheumatol **38**(5): 797-801.

Uddin, M. J., C. Grosse-Brinkhaus, et al. (2010). "Mapping of quantitative trait loci for mycoplasma and tetanus antibodies and interferon-gamma in a porcine F(2) Duroc x Pietrain resource population." Mamm Genome **21**(7-8): 409-418.

Ulfgren, A. K., L. Grondal, et al. (2000). "Interindividual and intra-articular variation of proinflammatory cytokines in patients with rheumatoid arthritis: potential implications for treatment." Ann Rheum Dis **59**(6): 439-447.

-V-

van Bladeren, P. J. (2000). "Glutathione conjugation as a bioactivation reaction." Chem Biol Interact **129**(1-2): 61-76.

van Boekel, M. A., E. R. Vossenaar, et al. (2002). "Autoantibody systems in rheumatoid arthritis: specificity, sensitivity and diagnostic value." Arthritis Res **4**(2): 87-93.

- van der Heijde, D., L. Klareskog, et al. (2006). "Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis: two-year clinical and radiographic results from the TEMPO study, a double-blind, randomized trial." Arthritis Rheum **54**(4): 1063-1074.
- van der Helm-van Mil, A. H., J. Z. Wesoly, et al. (2005). "Understanding the genetic contribution to rheumatoid arthritis." Curr Opin Rheumatol **17**(3): 299-304.
- van der Woude, D., J. J. Houwing-Duistermaat, et al. (2009). "Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis." Arthritis Rheum **60**(4): 916-923.
- Van Esch, H., M. Bauters, et al. (2005). "Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males." Am J Hum Genet **77**(3): 442-453.
- Vang, T., M. Congia, et al. (2005). "Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant." Nat Genet **37**(12): 1317-1319.
- Viatte, S., D. Plant, et al. (2013). "Genetics and epigenetics of rheumatoid arthritis." Nat Rev Rheumatol **9**(3): 141-153.
- Vivar, N. and R. F. Van Vollenhoven (2014). "Advances in the treatment of rheumatoid arthritis." F1000Prime Rep **6**: 31.
- Vorachek, W. R., W. R. Pearson, et al. (1991). "Cloning, expression, and characterization of a class-mu glutathione transferase from human muscle, the product of the GST4 locus." Proc Natl Acad Sci U S A **88**(10): 4443-4447.
- Vrijenhoek, T., J. E. Buizer-Voskamp, et al. (2008). "Recurrent CNVs disrupt three candidate genes in schizophrenia patients." Am J Hum Genet **83**(4): 504-510.
- Vyskocilova, E., B. Szotakova, et al. (2013). "Age-related changes in hepatic activity and expression of detoxification enzymes in male rats." Biomed Res Int **2013**: 408573.

-W-

- Wagner, C. and G. M. Hansch (2004). "Genetic deficiency of CD16, the low-affinity receptor for immunoglobulin G, has no impact on the functional capacity of polymorphonuclear neutrophils." Eur J Clin Invest **34**(2): 149-155.
- Walker, S., S. Janyakhantikul, et al. (2009). "Multiplex Parologue Ratio Tests for accurate measurement of multiallelic CNVs." Genomics **93**(1): 98-103.
- Wang, L., M. J. Groves, et al. (2000). "Glutathione S-transferase enzyme expression in hematopoietic cell lines implies a differential protective role for T1 and A1 isoenzymes in erythroid and for M1 in lymphoid lineages." Haematologica **85**(6): 573-579.
- Warmerdam, P. A., N. M. Nabben, et al. (1993). "The human low affinity immunoglobulin G Fc receptor IIC gene is a result of an unequal crossover event." J Biol Chem **268**(10): 7346-7349.

- Weaver, S., S. Dube, et al. (2010). "Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution." Methods **50**(4): 271-276.
- Webb, G., V. Vaska, et al. (1996). "Chromosomal localization of the gene for the human theta class glutathione transferase (GSTT1)." Genomics **33**(1): 121-123.
- Weiss, L. A., Y. Shen, et al. (2008). "Association between microdeletion and microduplication at 16p11.2 and autism." N Engl J Med **358**(7): 667-675.
- Wellcome Trust Sanger Institute (2005). "The effect of genomic structural variation on gene expression and human disease" Hinxton, UK: 27-28.
- Weterings, E. and D. C. van Gent (2004). "The mechanism of non-homologous end-joining: a synopsis of synapsis." DNA Repair (Amst) **3**(11): 1425-1435.
- Weyand, C. M., T. G. McCarthy, et al. (1995). "Correlation between disease phenotype and genetic heterogeneity in rheumatoid arthritis." J Clin Invest **95**(5): 2120-2126.
- Wijbrandts, C. A., M. G. Dijkgraaf, et al. (2008). "The clinical response to infliximab in rheumatoid arthritis is in part dependent on pretreatment tumour necrosis factor alpha expression in the synovium." Ann Rheum Dis **67**(8): 1139-1144.
- Wilke, K., B. Duman, et al. (2000). "Diagnosis of haploidy and triploidy based on measurement of gene copy number by real-time PCR." Hum Mutat **16**(5): 431-436.
- Willcocks, L. C., P. A. Lyons, et al. (2008). "Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake." J Exp Med **205**(7): 1573-1582.
- Wu, C. C., S. Shete, et al. (2013). "Whole-genome detection of disease-associated deletions or excess homozygosity in a case-control study of rheumatoid arthritis." Hum Mol Genet **22**(6): 1249-1261.
- Wu, T. and C. Mohan (2009). "The AKT axis as a therapeutic target in autoimmune diseases." Endocr Metab Immune Disord Drug Targets **9**(2): 145-150.

-X-

- Xu, S., Y. Wang, et al. (1998). "Characterization of the human class Mu glutathione S-transferase gene cluster and the GSTM1 deletion." J Biol Chem **273**(6): 3517-3527.
- Xue, Y., D. Sun, et al. (2008). "Adaptive evolution of UGT2B17 copy-number variation." Am J Hum Genet **83**(3): 337-346.

-Y-

- Yamamoto, S., K. Shimizu, et al. (1992). "Calcium-dependent cysteine proteinase (calpain) in human arthritic synovial joints." Arthritis Rheum **35**(11): 1309-1317.
- Yim, S. H., Y. J. Chung, et al. (2011). "The potential role of VPREB1 gene copy number variation in susceptibility to rheumatoid arthritis." Mol Immunol **48**(11): 1338-1343.

- Yoon, S., Z. Xuan, et al. (2009). "Sensitive and accurate detection of copy number variants using read depth of coverage." Genome Res **19**(9): 1586-1592.
- Young, B. J., R. K. Mallya, et al. (1979). "Anti-keratin antibodies in rheumatoid arthritis." Br Med J **2**(6182): 97-99.
- Yun, B. R., A. El-Sohemy, et al. (2005). "Glutathione S-transferase M1, T1, and P1 genotypes and rheumatoid arthritis." J Rheumatol **32**(6): 992-997.

-Z-

- Zhang, J., N. Zahir, et al. (2011). "The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness." Nat Genet **43**(9): 902-907.
- Zhao, M., Q. Wang, et al. (2013). "Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives." BMC Bioinformatics **14 Suppl 11**: S1.
- Zhong, S., A. H. Wyllie, et al. (1993). "Relationship between the GSTM1 genetic polymorphism and susceptibility to bladder, breast and colon cancer." Carcinogenesis **14**(9): 1821-1824.
- Zhou, J., J. Hu, et al. (2010). "The association between copy number variations in glutathione S-transferase M1 and T1 and age-related cataract in a Han Chinese population." Invest Ophthalmol Vis Sci **51**(8): 3924-3928.
- Zlotnik, A. and O. Yoshie (2000). "Chemokines: a new classification system and their role in immunity." Immunity **12**(2): 121-127.

Annexes

Article soumis, novembre 2014

Copy Number Variant genotypes with tandem repeat copies for GSTM1 gene in Rheumatoid Arthritis families

Mohamed Sahbi Ben Kilani¹, François Cornélis², Robert Olaso³, Thomas Bardin⁴, Valérie Chaudru¹, Elisabeth Petit-Teixeira¹.

Corresponding author:

Elisabeth Petit-Teixeira

Laboratoire GenHotel-EA3886

Université d'Evry-Val d'Essonne

2, rue Gaston Crémieux - CP 5727

91057 EVRY cedex

elisabeth.teixeira@univ-evry.fr

Telephone number: +33 1 60 87 45 72

Fax number: +33 1 60 87 45 71 (Fax)

1: GenHotel-EA3886, Evry University, 91057 Evry-cedex, France.

2: GenHotel-Auvergne, EA4679, Auvergne University, Genetic department, CHU Clermont Ferrand, 63000 Clermont Ferrand, France.

3 : CEA-Institut de Génomique, Centre National de Genotypage, Evry-cedex, France.

4 : Fédération de Rhumatologie, Hôpital Lariboisière, Assistance Publique-Hôpitaux de Paris, 75010 Paris, France.

Abstract

Objectives Rheumatoid Arthritis, a common autoimmune disease leading to joints deformation and destruction, has a multifactorial etiology. The genetic component involved in RA is not entirely understood and exploring variants such as Copy Number Variants in candidate genes is currently a promising research field.

Methods We investigated Copy Number Variants of Glutathione S-Transferase Mu 1 gene. This gene was first identified using a multiplex PCR. For positive samples, we quantified Copy Number Variants thanks to quantitative PCR. Then we developed Droplet Digital PCR with a digestion protocol to identify Copy Number Variants genotypes. Finally, in order to investigate association risk for Glutathione S-Transferase Mu 1 Copy Number Variants in Rheumatoid Arthritis, we performed transmission tests in 200 trio families (on each patient and his/hers two healthy parents).

Results Droplet Digital PCR led to quantify Copy Number Variants with the highest accuracy. We were able to define genotypes and identify samples with tandem copies of Glutathione S-Transferase Mu 1 gene on the same chromosome. This rare event was inherited and not a result of *de novo* recombination. Furthermore, analysis of family genotypes highlighted a non-significant association of GSTM1 deletion with RA and showed different modes of Copy Number Variant transmission.

Conclusion This study suggested that Droplet Digital PCR is the best method to use for Copy Number Variant genotypes identification in familial sample and has to be developed in the field of new genomic variants identification in complex diseases such as Rheumatoid Arthritis.

Key words: Rheumatoid Arthritis; Copy Number Variation; Droplet Digital PCR; Glutathione S-Transferase Mu 1 gene; Familial transmission.

1. Introduction

Rheumatoid Arthritis (RA), a chronic autoimmune disease characterized by symmetrical inflammation of the peripheral joints resulting in a progressive destruction of cartilage and bone, has a multifactorial etiology. Tobacco exposure (TE) is the strongest environmental risk ever described (Costenbader et al 2006) and Shared Epitope (SE) alleles of *HLA-DRB1* gene play a prominent role (Stastny 1976; Gregersen et al 1987) in RA genomic component. Genome Wide Association Studies (GWAS) using SNPs (Stahl et al 2012 and Okada et al 2014, for the most significant) identified about one hundred genetic risk factors, but they explained only a modest fraction of trait heritability in RA.

Research findings in past few years indicated a strong association between Copy Number Variants (CNVs) and several complex and common diseases (Almal and Padh 2012). Gene Ontology analysis from several studies has revealed that a substantial amount of genes with CNV are involved in immune response, drug metabolism and responses to external biotic stimuli (Feuk 2006; de Smith et al 2008; Jancova et al 2010). Among them, Glutathione S-transferase (GST) M1 gene (*GSTM1*, OMIM 138350) codes one of the major phase II detoxification enzymes that are involved in the metabolism of xenobiotics and have an important role in cellular protection against oxidative stress (Jancova et al 2010). *GSTM1* gene is located on chromosome 1p13.3 and exhibits a CNV that leads to the complete deletion of the gene (*GSTM1*-null) in 52% of the Caucasian population (Garte et al 2001). This deletion is the result of a Non-Allelic Homologue Recombination (NAHR) between two high homologous box sequences on both sides of the gene (McLellan et al 1997; Xu et al 1998). This recombination leads to deletion on one chromosome and duplication on the second one of the homologous pair (Xu et al 1998). Previous studies used multiplex PCR (mPCR) to identify presence or absence of the gene (Mattey et al 1999; Mattey et al 2002; Yun et al 2005; Criswell et al 2006; Bohanec et al 2009), or used quantitative PCR (qPCR) to quantify *GSTM1* CNVs (Bohanec et al 2009; Keenan et al 2010; Lundström et al 2011). However, distinction between homozygous (1/1) and heterozygous (2/0) genotypes among individuals with two copies of the gene was not investigated. Digital PCR (Hindson et al 2011) and related protocol development highlighted opportunities to accurate CNV genotyping.

Due to its importance in detoxification, deletion of *GSTM1* gene has been considered as a potential risk for RA. We decided to investigate CNV *GSTM1* in 200 RA trio families using several methodologies: multiplex PCR, quantitative PCR and a Droplet Digital PCR (ddPCR). Our goal was to be able to accurately characterize CNV genotypes and to investigate CNV transmission in RA trios.

2. Methods

2.1. Patients

We studied 200 trio families, consisting in one RA patient and his/her two parents from French European origin. RA satisfied the revised criteria of the American College of Rheumatology (ACR) (Arnett et al 1988). Table 1 contains the following characteristics for each RA index: sex, age at RA onset, presence of bone erosions at x-ray examination,

presence of rheumatoid nodules, as well as seropositivity for rheumatoid factor (RF) and for Anti-Cyclic Citrullinated Peptide Antibodies (ACPA), presence of HLA-DRB1 shared epitope allele and tobacco exposure (TE) (Michou et al 2008). The Ethics Committees of Hôpital Bicêtre and Hôpital Saint Louis (Paris, France) approved the study (CPPRB 94-40) and all subjects provided informed written consent. Genomic DNA was extracted from whole blood samples by salting out or phenol/chloroform methods.

2.2. Screening presence/absence of *GSTM1* with multiplex PCR

We used a multiplex Polymerase Chain Reaction (mPCR) in order to screen the presence (of at least one copy) or the absence (the complete deletion) of *GSTM1* gene for 600 samples (200 trios). Hence, we used three primers (P1 5'-CGCCATCTTGTGCTACATTGCCCG-3', P2 5'-ATCTTCTCCTCTTCTGTCTC-3' and P3 5'-TTCTGGATTGTAGCAGATCA-3') for *GSTM1* gene and internal control gene amplifications, as described in a previous study (Zhong et al 1993). The first couple (P1 and P2) amplified a 153bp fragment corresponding to the internal positive control; and the second couple (P1 and P3) amplified a 230bp fragment of a specific *GSTM1* gene region.

2.3. Quantification of CNVs by qPCR

We performed quantification of CNVs for samples positive for *GSTM1* presence and for samples that failed any amplification with mPCR. We used two different hydrolysis probes for *GSTM1* CNV: the first probe was localized in the exon 1 and the second in the intron 3 (Hs02575461_cn and Hs03352485_cn respectively, Applied Biosystems, Foster City, USA). For Copy Number (CN) normalization, we used a probe of a reference gene with 2 copies, *RNase P* gene (OMIM 608513, catalog number 4403326, Applied Biosystems). The assay was performed according to the manufacturer's instructions and quantitative reactions were run on an Applied Biosystems 7500Real-Time PCR machine. We tested all samples in triplicates, and normalized fluorescence signals to ROX fluorescence. The CN was assigned from Cycle threshold (Ct) values using CopyCaller™ software (version 2.0; Applied Biosystems) and samples with known CN were used as calibrators (NA10851/NA10852/NA10865/NA11893/NA11894, <http://ccr.coriell.org/>, New Jersey, USA).

2.4. Genotyping of CNVs by ddPCR

We performed digital PCR method on samples from trio families with at least two copies detected by qPCR method, in order to characterize copy number genotype (0/2) or (1/1). We performed a reaction with QX200™ Droplet Digital PCR (Bio-Rad Laboratories, California, USA), following the Minimum Information for the Publication of Digital PCR Experiment (dMIQE) guidelines (Huggett et al 2013) and using positive controls with known CN positive, negative and no-template controls. For this ddPCR experiment, we used the same target probes as the ones used for qPCR. For the reference gene with two copies, we used a probe for *RPP30* gene (OMIM 606115). We prepared 20µl of reaction volume with 16ng of DNA template, primers and probes with final concentrations of, respectively, 900 and 250 nmol/L, and Droplet PCR Supermix (Bio-Rad) with final concentration of 1X. Then, we added 70µl of droplet generation oil and 13 000 to 16 000 droplets were generated in 40µl

(with at least one copy of the gene per droplet) with the Droplet Generator, according to the manufacturer's instructions. Next, we performed an amplification in a C1000 Touch™ Thermal Cycler (Bio-Rad) in the following conditions: 95°C (10 min), 40 cycles of 94°C (30 sec) and 60°C (60 sec), and 98°C (10 min). We used QuantaSoft™ (version 1.4, Bio-Rad) software to estimate CNVs by counting the positive and negative fluorescence signal events. In addition, we calculated the concentration (copies/μl) by Poisson law and normalized to the reference concentration using the same software.

For samples with a CN value far from an integer of more than 0.25, the quantification was preceded by a digestion with the restriction endonuclease *HaeIII* (New England Biolab, England), in order to separate possible tandem duplication of *GSTM1*. Digestion was done in a final volume of 10μl with one unit of *HaeIII* enzyme and 25ng of DNA, 1X of *CutSmart*™ buffer and qspH₂O. A quantity of 12.5 ng of digest DNA was used for ddPCR protocol described above.

2.5.Family-based association tests with CNV genotyping data

We first based the association analyses on the Transmission Disequilibrium Test (TDT), done with FBAT software (Horwath et al 2001), which compares for a given allele, its transmission from heterozygous parents to RA patients with the transmission expected from Mendel's law (i.e., 50 %). These analyses were performed considering genotypes regarding deletion allele (-) and one copy allele (+). We used an additive and a dominant model (allele (-) was considered as the dominant allele).

We also used Genotypes Relative Risk (GRR) according to McNemar chi-square, in order to compare the frequency of genotypes with at least one allele (-) between RA cases and pseudo-controls for which the genotypes derived from non-transmitted parental alleles (Scaid and Sommer 1994).

We applied these methods to the whole sample of trios and to the subgroups defined according to RA cases' characteristics, namely: more or less than the age of 40 years old at RA onset, presence or absence of erosion, nodules, Rheumatoid Factor (RF), Anti-Cyclic Citrullinated Peptide Antibodies (ACPA), HLA-DRB1-"Shared Epitope" (SE), and Tobacco Exposure (TE, smokers/ex-smokers or non-smokers). In order to correct for multiple testing, we considered a p-value threshold of 0.003 (0.05/17) after applying Bonferroni correction, as 17 tests were performed.

3. Results

3.1.Screening presence/absence of *GSTM1* gene with multiplex PCR

We obtained data from mPCR for 575 samples. Accordingly the results showed two configurations: one amplification of the internal control (153bp) indicating a homozygous deletion (-/-) with a frequency of 50% in samples, or two amplifications of 153bp and 230bp indicating a presence of *GSTM1* with at least one copy (-/+ or +/+, data not shown).

3.2.Quantification of CNVs by qPCR

We quantified with qPCR the CN for 284 samples positive for *GSTMI* presence and for 25 samples that failed any amplification with mPCR. Hence, for 13 samples showing a confidence in CN assignment less than 90% and for 18 samples with results discordant with mPCR, a second qPCR was performed with a second probe. Finally, we obtained quantification for 296 samples. qPCR identified 9.5% of samples with 0 copy, 73% of samples with 1 copy and 17,5% samples with 2 copies. Results for 4 samples were illustrated in Fig. 1a.

3.3. Genotyping of CNVs by ddPCR

First we tested ddPCR on two samples in triplicates in the same plate and in separate plates in order to test reproducibility of assigned CN. Results showed no significant difference between the two conditions (data not shown). Accordingly, ddPCR results of the 4 samples performed in triplicate are shown in Fig. 1b. Comparison of standard deviations between qPCR (Fig. 1a) and ddPCR (Fig. 1b) results highlighted the highest accuracy of ddPCR, leading to a best CN identification.

Then we performed a ddPCR reaction for each of the 68 samples with at least 2 copies detected with qPCR or without mPCR and qPCR amplification. Interestingly, two samples (number 3 and number 5) and one positive control (number 4 which was known to harbor 3 copies) showed an under-estimated CN: 1.44 for number 3, 1.6 for number 5 and 2.85 for number 4 (Fig. 2). Digestion of DNA *HaeIII* enzyme changed CN assignment from 1.44 to 1.9 copies for sample number 3, from 1.6 to 1.89 copies for sample number 5 and from 2.85 to 2.92 copies for sample number 4 (Fig. 2).

These results indicated that 2 copies are on tandem conformation for these samples (2/0 or 2/1). All samples showing 2 copies with undigested DNA, 25% were randomly tested with digestion and results remained the same. We then made the assumption that a Standard Deviation (SD) equal to ± 0.25 between samples predigested and not predigested as a non-significant variation. Accordingly, samples with 2 ± 0.25 copies are considered as homozygous (1/1) for *GSTMI*. Finally, among the 68 samples previously characterized with at least two copies, ddPCR identified 11.8% with 0 copy, 4.4% with 1 copy and 83.8% with 2 copies.

3.4. *GSTMI* CNVs inheritance in trio families

The combined results from all methods showed a frequency of 53.5% for 0 copy, 37.5% for 1 copy, and 9% for 2 copies in our studied population (597 samples, 199 trios, one excluded because of CN characterization failure). Hence, we created a number of modes of family transmission according to CNVs genotypes of each parent and offspring (Table 2). We identified 10 of them, 9 with transmission of a deleted allele and/or one copy allele. The remaining mode showed a transmission of tandem copies of *GSTMI* gene from one parent to his/her offspring. For a specific parental CNVs genotypes combination, frequencies of the different CNV genotypes in RA patients were similar (as an example, for modes 2 and 3, 19.1% of index were 0/0 and 18.1% were 0/1). Hence, these results indicated that no preferential transmission of *GSTMI* CNVs has been highlighted.

3.5. Family-based association tests with CNV genotyping data

In order to study the association between *GSTM1* and RA in trio families, we considered allele (-) as deletion of *GSTM1* and allele (+) as presence of one copy, accordingly, leading to three different genotypes (-/-, -/+, +/+). Two families were excluded from the analysis, one with duplicated copies, and the other because of one sample with CN characterization failure. We then investigated the TDT of each allele in the remaining 198 trios, using either an additive or a dominant model. After Bonferroni correction, we did not notice any preferential transmission of the deleted allele of *GSTM1* in the whole sample. And using GRR method, we did not find any association between *GSTM1* genotype and RA (Table 3). When stratifying trios according to clinical, biological, environmental and genetic factors for the index cases, we noticed an association of *GSTM1* deleted allele with RA with TDT and GRR tests in the subgroup of families with index having an age of onset higher than 40 years ($P_{TDT}=0.02$ and $P_{GRR}=0.041$) and in the subgroup of families with RA index carrying nodules ($P_{TDT}=0.016$ and $P_{GRR}=0.01$) (Table 3), although non-significant after Bonferroni correction.

4. Discussion

In the current study, we combined the results of multiplex PCR, qPCR, and the QX200™ Droplet Digital PCR to study the transmission of *GSTM1* CNVs in trio families of RA patients. The use of these technologies to identify CNVs for *GSTM1* gene was necessary to describe CNV genotypes and to analyze their transmission in RA families. First, we identified the presence and the absence of *GSTM1* using mPCR, without knowing the copy number. Second, we obtained the copy number of the gene by a relative qPCR without characterization of the genotype. Finally, thanks to digital PCR, we were able to pinpoint CNV values and to make a distinction between homozygous (1/1) and heterozygous (2/0) genotypes for samples with two copies. Two samples (0.3%) showed a duplication of the gene *GSTM1* on the same chromosome. This duplication event has been demonstrated to occur through inter chromosome NAHR, leading to the simultaneous deletion of *GSTM1* on one chromosome and a duplication on the second one (McLellan et al 1997). As the deleted genotype of *GSTM1* observed in our study (53.5%) and previously described in Caucasian population (52%, Garte et al 2001) is much more frequent than tandem copies, we are able to confirm the assumption that the deleted allele was positively selected, thus highlighting the purifying selection of *GSTM1* described previously (McCarroll et al 2006).

The analysis of CNVs in trio families showed different modes of transmission from parents to RA patients, and highlighted a transmitted duplication from one parent to his/her offspring, which was not the result of *de novo* event. The other modes of transmission showed a non-preferential transmission between the deletion and the presence of the gene. Family-based analyses showed a non-significant association between deletion of *GSTM1* gene and RA in trio families after correction for multiple testing. This association is observed in two subgroups: patients with age higher than 40 years and patients with nodules.

Some previous association studies found a significant association between *GSTM1* deletion and RA in case/control analysis. Indeed, a first study suggested a role of *GSTM1* on the severity of the disease, independently of *HLA-DRB1 SE* (Mattey et al 1999). A second study tested the association between smoking and RA in patients carrying *GSTM1* deletion,

and showed a higher risk in a group of smoker patients carrying the deleted genotype (-/-) than in non-smoker patients, thus suggesting that *GSTM1* could influence the relation between smoking and the severity of the disease (Mattey et al 2002). This result is supported by a recent study which analyzed the *GSTM1* genotypes of Korean RA cases versus controls and found a significant association of the (-) allele with RA severity especially in the subgroup of patients negative for HLA-DRB1 SE (Youn et al 2005). However, other studies observed no significant association between RA and *GSTM1* not even in smokers or in other sub-groups of patients (Criswell et al 2006; Bohanec Grabar et al 2009; Keenan et al 2010). Moreover, when *GSTM1* CNVs was analyzed by qPCR in a case/control study, no association between RA and the gene of interest was found, but a significant association with the severity of RA in ACPA positive patients was highlighted (Lundström et al 2011). These heterogeneous results can be explained by the differences among the studies regarding populations and samples' sizes. Our analysis investigating the transmission of *GSTM1* CNVs, using accurate genotyping data, showed that the deleted allele was not associated with RA. Performing the same methodology of CNV characterization in large case/controls association study could help to conclude about association of *GSTM1* gene deletion with RA.

In conclusion, digital PCR offers the most adequate methodology to accurately genotype CNVs. Analysis of familial sample leads to the identification of duplication events and to the characterization of genotypes, essential for candidate genes studies of complex diseases. As *GSTM1* gene was described as a biomarker in the clinical diagnosis of cancers (Jin et al 2014; Oliveira et al 2014), ddPCR protocol we described here could be used for this important issue. Furthermore, this methodology would be essential to replicate association of CNVs with complex diseases from genome screening studies.

Conflict of interest statement

None of the authors has any conflicts of interest to declare.

Acknowledgements

This work was supported by Association Française des Polyarthritiques, Société Française de Rhumatologie, Association Rhumatisme et Travail, Association Polyarctique, Genopole® and Evry-Val d'Essonne University.

We are grateful to RA patients, their family and rheumatologists for their participation in this study.

We thank the Bio-Rad team for the preliminary tests on the Droplet® Digital PCR.

We also thank Chadi Ben Youssef for helpful reading.

References

Almal SH, Padh H (2012) Implications of gene copy-number variation in health and diseases. *J Hum Genet* 57: 6-13.

Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, et al (1998) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 31: 315-24.

Bohanec Grabar P, Logar D, Tomsic M, Rozman B, Dolzan V (2009) Genetic polymorphisms of glutathione S-transferases and disease activity of rheumatoid arthritis. *Clin Exp Rheumatol* 27: 229-36.

Costenbader KH, Feskanich D, Mandl LA, Karlson EW (2006) Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *Am J Med* 503: 1-9.

Criswell LA, Saag KG, Mikuls TR, Mudano AS, Burma M, Folsom AR, Saag KG (2006) Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women. *Ann Rheum Dis* 65: 1163-67.

de Smith AJ, Walters RG, Froguel P, Blakemore AI (2008) Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenet Genome Res* 123: 17-26.

Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7: 85-97.

Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL, Baranova H, Bathum L, Benhamou S, Boffetta P, Bouchardy C, Breskvar K, Brockmoller J, Cascorbi I, Clapper ML, Coutelle C, Daly A, Dell'Omo M, Dolzan V, Dresler CM, Fryer A, Haugen A, Hein DW, Hildesheim A, Hirvonen A, Hsieh LL, Ingelman-Sundberg M, Kalina I, Kang D, Kihara M, Kiyohara C, Kremers P, Lazarus P, Le Marchand L, Lechner MC, van Lieshout EM, London S, Manni JJ, Maugard CM, Morita S, Nazar-Stewart V, Noda K, Oda Y, Parl FF, Pastorelli R, Persson I, Peters WH, Rannug A, Rebbeck T, Risch A, Roelandt L, Romkes M, Ryberg D, Salagovic J, Schoket B, Seidegard J, Shields PG, Sim E, Sinnet D, Strange RC, Stücker I, Sugimura H, To-Figueras J, Vineis P, Yu MC, Taioli E (2001) Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol Biomarkers Prev* 10: 1239-48.

Gregersen PK, Silver J, Winchester RJ (1987) The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 30: 1205-13.

Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, Kitano TK, Hodel MR, Petersen JF, Wyatt PW, Steenblock ER, Shah PH, Bousse LJ, Troup CB, Mellen JC, Wittmann DK, Erndt NG, Cauley TH, Koehler RT, So AP, Dube S, Rose KA, Montesclaros L, Wang S, Stumbo DP, Hodges SP, Romine S, Milanovich FP, White HE, Regan JF, Karlin-Neumann GA, Hindson CM, Saxonov S, Colston BW (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* 83: 8604-10.

Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype–phenotype associations. *Eur J Hum Genet* 9: 301–6.

Huggett JF, Foy CA, Benes V Emslie K, Garson JA, Haynes R, Hellemans J, Kubista M, Mueller RD, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT, Bustin SA (2013) The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. *Clin Chem* 59: 892-902.

Jancova P, Anzenbacher P, Anzenbacherova E (2010) Phase II drug metabolizing enzymes. *Biomed Pap Med Fac Univ Palacky Olomoucv Czech Repub* 154: 103-16.

Jin Y, Hao Z (2014) Polymorphisms of glutathione S-transferase M1 (GSTM1) and T1 (GSTT1) in ovarian cancer risk. *Tumour Biol* 35: 5267-72.

Keenan BT, Chibnik LB, Cui J Ding B, Padyukov L, Kallberg H, Bengtsson C, Klareskog L, Alfredsson L, Karlson EW (2010) Effect of interactions of glutathione S-transferase T1, M1, and P1 and HMOX1 gene promoter polymorphisms with heavy smoking on the risk of rheumatoid arthritis. *Arthritis Rheum* 62: 3196-3210.

Lundström E, Hartshorne T, Li K Lindblad S, Wick MC, Bengtsson C, Alfredsson L, Klareskog L, Padyukov L (2011) Effects of GSTM1 in rheumatoid arthritis; results from the Swedish EIRA study. *PLoS One* 6: e17880.

Mattey DL, Hassell AB, Plant M Dawes PT, Ollier WR, Jones PW, Fryer AA, Alldersea JE, Strange RC (1999) Association of polymorphism in glutathione S-transferase loci with susceptibility and outcome in rheumatoid arthritis: comparison with the shared epitope. *Ann Rheum Dis* 58: 164-68.

Mattey DL, Hutchinson D, Dawes PT, Nixon NB, Clarke S, Fisher J, Brownfield A, Alldersea J, Fryer AA, Strange RC (2002) Smoking and disease severity in rheumatoid arthritis: association with polymorphism at the glutathione S-transferase M1 locus. *Arthritis Rheum* 46: 640-46.

McCarroll SA, Hadnott TN, Perry GH Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM (2006) Common deletion polymorphisms in the human genome. *Nat Genet* 38: 86-92.

McLellan RA, Oscarson M, Alexandrie AK Seidegård J, Evans DA, Rannug A, Ingelman-Sundberg M (1997) Characterization of a human glutathione S-transferase mu cluster containing a duplicated GSTM1 gene that causes ultrarapid enzyme activity. *Mol Pharmacol* 52: 958-65.

Michou L, Teixeira VH, Pierlot C Lasbleiz S, Bardin T, Dieudé P, Prum B, Cornélis F, Petit-Teixeira E (2008) Associations between genetic factors, tobacco smoking and autoantibodies in familial and sporadic rheumatoid arthritis. *Ann Rheum Dis* 67: 466-70.

Okada Y, Wu D, Trynka G Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A, Ortmann W, Bhangale T, Denny JC, Carroll RJ, Eyler AE, Greenberg JD, Kremer JM, Pappas DA, Jiang L, Yin J, Ye L, Su DF, Yang J, Xie G, Keystone E, Westra HJ, Esko T, Metspalu A, Zhou X, Gupta N, Mirel D, Stahl EA, Diogo D, Cui J, Liao K, Guo MH, Myouzen K, Kawaguchi T, Coenen MJ, van Riel PL, van de Laar MA, Guchelaar HJ, Huizinga TW, Dieudé P, Mariette X, Bridges SL Jr, Zhernakova A, Toes RE, Tak PP, Miceli-Richard C, Bang SY, Lee HS, Martin J, Gonzalez-Gay MA, Rodriguez-

- Rodriguez L, Rantapää-Dahlqvist S, Arlestig L, Choi HK, Kamatani Y, Galan P, Lathrop M; RACI consortium; GARNET consortium, Eyre S, Bowes J, Barton A, de Vries N, Moreland LW, Criswell LA, Karlson EW, Taniguchi A, Yamada R, Kubo M, Liu JS, Bae SC, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Raychaudhuri S, Stranger BE, De Jager PL, Franke L, Visscher PM, Brown MA, Yamanaka H, Mimori T, Takahashi A, Xu H, Behrens TW, Siminovitch KA, Momohara S, Matsuda F, Yamamoto K, Plenge RM (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506: 376-81.
- Oliveira AL, Oliveira Rodrigues FF, Dos Santos RE, Rozenowicz RL, Barbosa de Melo M (2014) GSTT1, GSTM1, and GSTP1 polymorphisms as a prognostic factor in women with breast cancer. *Genet Mol Res* 13: 2521-30.
- Rose-Zerilli MJ, Barton SJ, Henderson AJ, Shaheen SO, Holloway JW (2008) Copy-number variation genotyping of GSTT1 and GSTM1 gene deletions by real-time PCR. *Clin Chem* 55: 1680-5.
- Scaid DJ, Sommer SS (1994) Comparison of statistics for candidate-gene association studies using cases and parents. *Am J Hum Genet* 55: 402-9.
- Stahl EA, Raychaudhuri S, Remmers EF (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 42: 508-14.
- Stastny P (1976) Mixed lymphocyte cultures in rheumatoid arthritis. *J Clin Invest* 57: 1148-57.
- Zhong S, Wyllie AH, Barnes D, Wolf CR, Spurr NK (1993) Relationship between the GSTM1 genetic polymorphism and susceptibility to bladder, breast and colon cancer. *Carcinogenesis* 14: 1821-4.
- Zhou J, Hu J, Guan H (2010) The association between copy number variations in glutathione S-transferase M1 and T1 and age-related cataract in a Han Chinese population. *Invest Ophthalmol Vis Sci* 51: 3924-28.
- Xu S, Wang Y, Roe B, Pearson WR (1998) Characterization of the human class Mu glutathione S-transferase gene cluster and the GSTM1 deletion. *J Biol Chem* 273: 3517-27.
- Yun BR, El-Sohemy A, Cornelis MC, Bae SC (2005) Glutathione S-transferase M1, T1, and P1 genotypes and rheumatoid arthritis. *J Rheumatol* 32: 992-97.

Table 1: Characteristics of RA index cases of 200 trio families

	Sample (n=200)	
	Number¹	Percentage
Female	178/200	89%
Age at onset less than 40 years	148/200	74%
RF positive	158/200	79%
ACPA positive	149/193	77%
Presence of Erosion	170/200	85%
Presence of Nodules	52/200	26%
TE	98/185	53%
Carrying at least one HLA-DRB1 shared epitope allele	158/200	79%

¹ Number of index cases / Number of index cases with data

^{RF} Rheumatoid Factor

^{ACPA} Anti-Cyclic Citrullinated Peptide Antibodies;

^{TE} Tobacco Exposure (smokers/ex-smokers)

**Table 2: Modes of GSTM1 CNVs transmission
in 199 trio families**

Family Mode	Genotypes	n	Percentage
1-0x0=0	0/0x0/0=0/0	59	29.6%
2-0x1=0	0/0x1/0=0/0	38	19.1%
3-0x1=1	0/0x1/0=0/1	36	18.1%
4-1x1=0	1/0x1/0=0/0	8	4%
5-1x1=1	1/0x1/0=1/0	11	5.5%
6-1x1=2	1/0x1/0=1/1	10	5%
7-1x2=1	1/0x1/1=1/0	7	3.6%
8-1x2=2	1/0x1/1=1/1	7	3.6%
9-0x2=1	0/0x1/1=0/1	22	11%
10-0x2=2	0/0x0/2=0/2	1	0.3%

⁰ zero copy ¹ one copy ² two copies of *GSTM1* gene

Table 3: Results of Transmission Disequilibrium Test (TDT) and Genotype Relative Risk (GRR) test in RA trio families

Samples	Number of Families	Informative families¹	Frequency of allele (-)	P_{TDT}^{2,3}	P_{GRR}^{3,4}
Whole Sample					
GSTM1 (+)	198	118	0.274	0.68	0.837
GSTM1 (-)	198	118	0.726	0.68	0.863
Age at onset >40 years					
GSTM1 (+)	52	31	0.260	0.02	0.041
GSTM1 (-)	52	31	0.740	0.02	0.183
Nodules (+)					
GSTM1 (+)	52	32	0.303	0.016	0.041
GSTM1 (-)	52	32	0.697	0.016	0.132

¹ Families with at least one heterozygous parent

² P value for Transmission Disequilibrium Test (TDT) using an additive model

³ Significant threshold was 0.003 after Bonferroni correction regarding the number of independent tests performed.

³ GRR: Genotype Relative Risk

(⁺) one copy allele

(⁻) deletion allele, the risk allele.

Fig. 1 GSTM1 Copy Numbers for 4 samples: **A)** Quantification with qPCR in triplicates. Standard deviation (SD) was calculated using replicates. CNs were calculated by CopyCaller™ software **B)** Quantification with ddPCR in triplicates. Standard deviation (SD) was calculated using replicates. CNs were calculated by QuantaSoft™ software. NTC: Non-Template Control

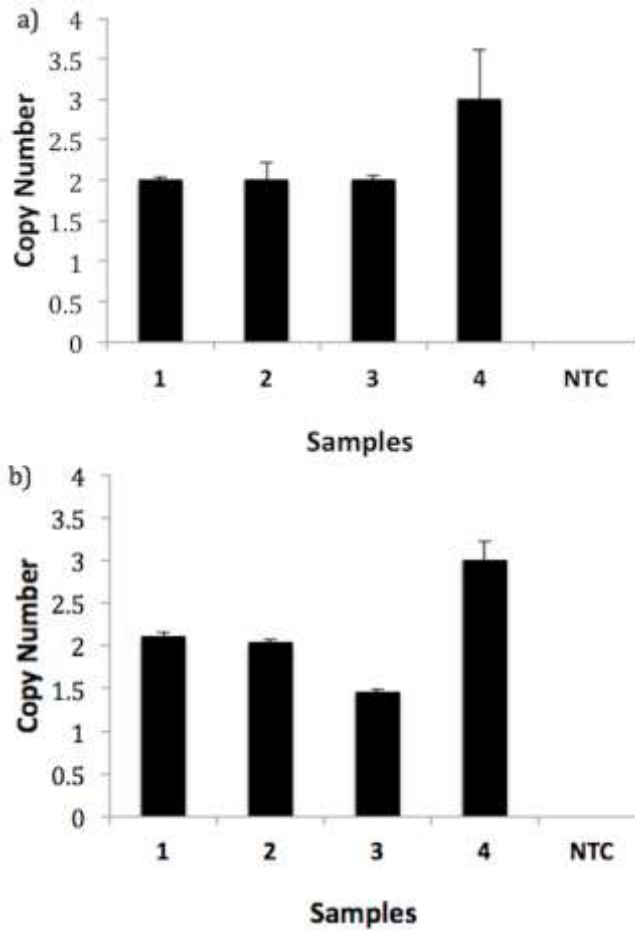


Fig. 2 ddPCR results for 5 samples after two independent experiments of digital PCR realized with or without enzymatic digestion. SD was calculated using Poisson law. CNs were calculated by QuantaSoft™ software.

