

*République Algérienne Démocratique Populaire*

*وزارة التعليم العالي و البحث العلمي*

*Ministère de l'Enseignement Supérieure et de la Recherche Scientifique*

*Université des Sciences et de la Technologie d'Oran « Mohamed Boudiaf »*



*Faculté des Sciences de la Nature et de la Vie*

*Département de Génétique Moléculaire Appliquée*

*Spécialité : Génétique moléculaire appliquée*

*Option : Biologie moléculaire*

## **THESE**

présentée par monsieur

**Mostefa FODIL**

en vue de l'obtention du

**DIPLOME DE DOCTORAT EN SCIENCES**

## **Thème**

---

**Etude d'association génétique et analyse de gènes candidats  
différentiellement exprimés au cours de la Polyarthrite Rhumatoïde**

---

*Soutenue le 12/05/2015, devant la commission d'examen composée de :*

<u>Président</u>	<i>M<sup>me</sup> Saïdi-Mehtar Nadhira</i>	<i>Professeur</i>	<i>Université USTO-MB</i>	<i>Algérie</i>
<u>Directeur de thèse</u>	<i>M<sup>r</sup> Boudjema Abdallah</i>	<i>Professeur</i>	<i>Université USTO-MB</i>	<i>Algérie</i>
<u>Co-directeur</u>	<i>M<sup>me</sup> Petit-Teixeira Elisabeth</i>	<i>Professeur</i>	<i>Université d'Evry Val d'Essonne</i>	<i>France</i>
<u>Examineur</u>	<i>M<sup>me</sup> Berrebbah-Alioua Amel</i>	<i>Maitre de Conférences A</i>	<i>Université USTO-MB</i>	<i>Algérie</i>
<u>Examineur</u>	<i>M<sup>me</sup> Boisson-Vidal Catherine</i>	<i>Directeur de recherche</i>	<i>CNRS</i>	<i>France</i>
<u>Examineur</u>	<i>M<sup>r</sup> Cornélis François</i>	<i>Professeur</i>	<i>Université d'Auvergne</i>	<i>France</i>
<u>Invité</u>	<i>M<sup>r</sup> Benzaoui Ahmed</i>	<i>Professeur</i>	<i>Université d'Oran</i>	<i>Algérie</i>
<u>Invité</u>	<i>M<sup>r</sup> Sahraoui Tewfik</i>	<i>Professeur</i>	<i>Université d'Oran</i>	<i>Algérie</i>

*Promotion : 2014/2015*



*« ... il pourrait même se faire que ce qui leur paraît des défauts fût comme les taches naturelles du visage, qui en relèvent quelquefois la beauté ...»*

*Extrait de « Don Quichotte »,  
Miguel de Cervantès*

***FINANCEMENTS &  
CONTRIBUTIONS***

Ce travail a été réalisé dans le cadre d'une thèse de cotutelle internationale entre l'Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf (USTO-MB) en Algérie et l'Université d'Evry Val d'Essonne (UEVE) en France.

Pour la partie pratique du travail, nous avons bénéficié d'un double financement :

- Pour la partie Algérienne : un projet PNR-USTO-MB «Projet national de Recherche, Université des Sciences et de la Technologie d'Oran - Mohamed Boudiaf» (appel d'offres 2011/2012).
- Pour la partie Française : par fonds de recherche de l'Association Française des Polyarthritiques, Société Française de Rhumatologie, Association Rhumatisme et Travail, Association Polyarctique, Génopole-Evry et Université d'Evry-Val d'Essonne.

Cette thèse a été soutenue par une bourse de mobilité internationale de doctorat attribuée par l'AUF «Agence Universitaire de la Francophonie» (79113FR033L) pour une durée de deux années et par un financement supplémentaire FDIR-UEVE «Fond de Développement pour l'internationalisation de la Recherche, Université d'Evry Val d'Essonne» (appels d'offres 2010/2011 et 2011/2012) également attribué pour une durée de 10 mois répartie sur deux périodes de 5 mois chacune.

*Nous témoignons ici de nos chaleureux remerciements et notre profonde reconnaissance aux patients atteints de Polyarthrite Rhumatoïde, à leurs familles respectives ainsi qu'aux rhumatologues pour avoir participé à cette étude.*

**REMERCIEMENTS**

Je remercie en premier lieu le Professeur **Nadhira SAIDI-MEHTAR** d'avoir accepté de présider le jury de ma soutenance. Aussi, je voudrais vous exprimer ma gratitude de m'avoir accueilli au sein de votre laboratoire de recherche à l'USTO-LGMC. Ce fut une grande chance et un immense honneur, madame, d'avoir reçu mes premières notions de génétique dans votre classe.

A madame **Catherine BOISSON-VIDAL**, pour votre soutien et vos remarques pertinentes lors de mon comité de thèse à Evry, et pour l'immense honneur que vous me faites d'avoir accepté de faire partie de ce Jury.

A madame **Amel BERREBBAH-ALIOUA**, je tiens à vous remercier très chaleureusement non seulement pour l'honneur que vous me faites en faisant partie de mon jury de thèse, mais aussi pour votre soutien et votre implication personnelle dans mes démarches. Que vous en soyez remerciée à travers ces quelques mots, vous assurant mon profond respect.

Je remercie le professeur **François CORNELIS** d'avoir accepté de faire partie de mon jury de thèse et de m'avoir accueilli, dès mon premier séjour au laboratoire GenHotel (stage de Magistère en 2008) avec sa grande sympathie, son expérience riche et ses conseils. Veuillez trouver ici l'expression de mon profond respect.

Mes chaleureux remerciements et mon profond respect vont au professeur **Ahmed BENZAOU**I pour sa magnifique contribution à ce travail, pour la qualité de ses remarques et son éternelle sympathie. Veuillez trouver ici, monsieur, ma sincère gratitude.

A monsieur **Tewfik SAHRAOUI** pour avoir accepté de faire partie de mon jury de thèse. Votre présence et vos suggestions m'honorent et contribuent grandement à la valorisation de ce travail. Encore une fois, un grand Merci.

Je voudrais par la suite exprimer, tout particulièrement, mon énorme gratitude à mes deux directeurs de thèse ; le Pr. **Elisabeth PETIT-TEIXEIRA** et le Pr. **Abdallah BOUDJEMA**. Au bout de ces 5 longues années de travail, nous avons passé des moments très chargés en réflexion, de discussions interminables et autres moments de détente mais aussi de vives émotions ...

Je tiens à vous rééditer, à travers ces quelques mots, ma profonde reconnaissance et ma grande fierté d'avoir été votre élève, mais pas seulement... Votre empreinte restera gravée à toujours aussi bien dans mon parcours personnel que professionnel.

Je tiens à remercier tout particulièrement M<sup>r</sup> **Lotfi LOUHIBI**, M<sup>r</sup> **Lahouari CHAA**, messieurs les vice-recteurs **Rachid BOUAMRANE** et **Rachid KESSAS** sans oublier M<sup>me</sup> **HOURIA** pour leur aide appréciable et pour leur réelle sympathie en dehors du cadre administratif officiel. Un grand merci à tout le staff administratif de l'Université des Sciences et de la Technologie d'Oran – Mohamed BOUDIAF « **USTO-MB** », ayant contribué de près ou de loin à l'aboutissement de mes démarches.

Tout pareil de l'autre côté de la méditerranée, un grand Merci à toute l'équipe de l'Ecole Doctorale **GAO** «Des Génomes aux Organismes» de l'Université d'Evry Val d'Essonne « **UEVE** » pour leur aide, accueil et sympathie à mon égard. Je remercie tout particulièrement monsieur le Professeur **Bernard PRUM**, M<sup>me</sup> **Véronique FOURNIER**, M<sup>me</sup> **Marité Quintin**, M<sup>me</sup> **Florence HERVY**, M<sup>me</sup> **Carole TROUSSIER** sans oublier l'éternelle souriante M<sup>me</sup> **Florence HAMON**. Encore une fois merci !

Bien évidemment, Merci à tous les membres du Laboratoire GenHotel-EA3886 au Génomôle d'EVRY de votre accueil, sympathie et amitié. Un grand Merci pour M<sup>me</sup> **Valérie CHAUDRU**, que j'ai eu le plaisir de connaître lors de mes deux dernières années de thèse. Votre constante amabilité et sincère sympathie resteront gravées dans ma mémoire. Merci tout particulièrement à **Mohamed-Sahbi** et **Julie-Améline** et à mon amie **Florence** sans oublier **les stagiaires** pour tous les beaux moments de détente et les pauses café-rigolades.

Merci à tous mes amis que j'ai rencontrés pendant ma résidence en France. A travers la colocation, j'ai eu le plaisir de nouer une amitié très forte avec eux sans oublier les soirées Cinoch et les folles parties de Badminton avec la bande des **RONGEURS** : **Benoit, Nicolas** et la très *British Eleanor*. Merci pour tout ! Je n'oublierai pas de remercier mes amis **Ibtissem** et **Sofiane** ainsi que leurs petits lutins **Anès** et **Moncef** ; Pour les parties de pêche au lac de Créteil, les couscous du Dimanche mais surtout pour votre amitié sincère et votre soutien. Un grand merci aussi à mes potes de toujours, **Omar (Moustache)**, **Abdelkader (Bakouka)** et **Mohammed (Moumouch)** ; Vous avez été d'un grand soutien pendant toutes ces années passées ensemble.

Je n'oublierai pas de remercier ma belle-famille (*de France et d'Algérie*) pour leur gentillesse et leur accueil, mais surtout pour avoir pris soin de mes petits chéris en mon absence.

Je remercie aussi toutes les personnes qui ont contribué de près ou de loin à la réalisation de cette thèse.

*Je ne saurai terminer ces lignes sans rendre hommage aux deux personnes par qui tout a commencé, **Maman** et **Papa** ! Vous avez su me prodiguer l'éducation et la formation initiale mais surtout l'amour conduisant à ce bel aboutissement... Je ne vous remercierai jamais assez... Veuillez trouver ici l'expression de mon amour et de mon éternelle reconnaissance. Que Dieu vous garde pour illuminer encore et toujours mon chemin.*

*Un grand merci à mon petit « motard » de frère **Nadjib** et mes deux petites sœurs **Nour El Houda** et **Latifa** sans oublier mon cher petit neveu **Akram** qui ont été et sont encore source de mon Bonheur.*

\* \* \*

*Je finirai par remercier ma meilleure Amie, mon épouse **Faouzia**, pour son immense contribution à cette thèse et à ma vie entière...*

*...Puisses-tu trouver ici le témoignage de mon total respect et ma sincère gratitude envers tout ce que tu as fait et ce que tu continues à faire sans te lasser. Sans toi, je n'aurai jamais pu tenir le coup, ni arriver au bout de cette thèse.*

\* \* \*

*Et enfin ... le Soleil de ma vie, mes Enfants !  
Ma Grande **WISSEM**, et Mon Petit **MOHAMMED** seront les dernières personnes que je remercierai...*

*Vous avez apporté un rayon de Lumière et de Bonheur dans ma vie, et votre image m'a été d'un grand réconfort pendant les longs mois de solitude en France loin de vous...*

*Vous êtes vraiment ma Meilleure récompense !*



# ***TABLE DES MATIERES***

LISTE DES FIGURES .....	i
LISTE DES TABLEAUX .....	ii
LISTE DES ABREVIATIONS .....	iii
<b>PREAMBULE</b> .....	<b>1</b>
<b>CHAPITRE 1 – INTRODUCTION</b> .....	<b>5</b>
<b>I. LES APPROCHES GENOMIQUES DES MALADIES COMPLEXES</b> .....	<b>6</b>
1 Les maladies complexes : définition.....	7
2 Caractérisation de la composante génétique.....	9
2.1 Analyses de liaison génétique.....	10
2.1.1 Différents types d’analyses de liaison .....	10
2.1.2 Marqueurs génétiques .....	11
2.1.3 Cartes génétiques et cartes physiques .....	17
2.2 Analyses d’association génétique .....	18
2.2.1 Notion de déséquilibre de liaison.....	19
2.2.2 Etudes d’association cas-témoins .....	21
2.2.3 Test du déséquilibre de transmission (TDT) .....	23
2.3 Choix de marqueurs et stratégies d’études : « Association <i>Versus</i> Liaison » .....	24
2.3.1 Approche gène candidat.....	24
2.3.2 Approche GWS “Genome Wide Scan” .....	26
2.3.3 Approche GWAS “Genome Wide Association Study” .....	29
2.3.4 Approche WES “Whole Exome Sequencing” .....	32
3 Analyse du Transcriptome.....	33
3.1 Définition du Transcriptome .....	33

3.2	Méthodologies d'étude.....	35
3.3	Relation Transcriptome / Variants génomiques .....	39
II.	<b>EXEMPLE DE MALADIE COMPLEXE : LA POLYARTHRITE</b>	
	<b>RHUMATOÏDE .....</b>	<b>42</b>
1	Définition et épidémiologie.....	43
2	La pathogenèse de la PR .....	46
2.1	La physiopathologie.....	46
2.2	La PR, une maladie auto-immune .....	48
2.2.1	Rupture de tolérance et mimétisme moléculaire.....	50
2.2.2	L'entretien de l'auto-immunité (Cellules et Cytokines).....	51
3	Les traitements .....	53
3.1	Les traitements conventionnels.....	53
3.2	Les Biothérapies.....	55
3.3	La pharmacogénétique (association de facteurs génétiques avec l'efficacité du traitement).....	57
4	Les facteurs de risque .....	59
4.1	Facteurs environnementaux.....	59
4.2	Facteurs individuels.....	62
5	La composante génétique de la PR .....	63
5.1	L'historique.....	63
5.2	Les études génétiques dans la PR.....	66
5.2.1	Principales GWAS et autres Méta-analyses .....	66
5.2.2	Facteurs de prédisposition génétique à la PR retenus pour notre étude	69
6	Données de Transcriptome dans la PR .....	83
6.1	Etudes de profils d'expression à large échelle .....	83

6.2	Etudes Pharmacogénomiques : Signatures de réponse aux traitements .....	85
III.	OBJECTIFS DE LA THESE .....	87
<b>CHAPITRE 2 - POPULATIONS &amp; MÉTHODES .....</b>		<b>90</b>
I.	Étude des facteurs de prédisposition génétique à la PR dans la population de l'Ouest Algérien .....	92
1	Population d'étude .....	93
2	Extraction de l'ADN .....	94
3	Choix des gènes et des SNPs.....	94
4	Génotypage moléculaire .....	95
5	Analyses statistiques .....	96
II.	Exploitation génétique des données du Transcriptome dans la PR dans des familles Européennes.....	99
1	Population d'étude .....	100
2	Choix des gènes et des SNPs.....	102
3	Génotypage moléculaire .....	103
4	Analyses statistiques .....	105
<b>CHAPITRE 3 – RESULTATS &amp; DISCUSSION.....</b>		<b>107</b>
I.	Étude des facteurs de prédisposition génétique à la PR dans la population de l'Ouest Algérien .....	109
1	Introduction .....	110
2	Résultats .....	111
3	Discussion .....	113
<b>ARTICLE 1 .....</b>		<b>117</b>

<b>II. Exploitation génétique des données du Transcriptome dans la PR dans des familles Européennes.....</b>	<b>128</b>
<b>1 Introduction .....</b>	<b>129</b>
<b>2 Résultats .....</b>	<b>130</b>
<b>2.1 Corrélation taux d'expression/génotypes .....</b>	<b>130</b>
<b>2.2 Etude d'association génétique des 10 SNPs étudiés avec la PR.....</b>	<b>132</b>
<b>3 Discussion .....</b>	<b>137</b>
<b>ARTICLE 2 .....</b>	<b>141</b>
<b>CHAPITRE 4 – CONCLUSION GENERALE &amp; PERSPECTIVES.....</b>	<b>148</b>
<b>CHAPITRE 5 – ANNEXES .....</b>	<b>152</b>
<b>REFERENCES BIBLIOGRAPHIQUES.....</b>	<b>189</b>

## **LISTE DES FIGURES**

**Figure 1** : Vue d'ensemble des plateformes en fonction du débit de polymorphismes SNP et de la taille de l'échantillon

**Figure 2** : Caractéristiques comparées des maladies héréditaires classiques mendéliennes et des maladies génétiques complexes

**Figure 3** : Représentation schématique des acteurs cellulaires et médiateurs inflammatoires impliqués dans la physiopathologie de la PR

**Figure 4** : Cytokines ciblées dans la polyarthrite rhumatoïde

**Figure 5** : Hypothèse physiopathologique de la PR ACPA+

**Figure 6** : Représentation schématique de l'épitope partagé codé par l'allèle HLA-DRB1\*0401

**Figure 7** : Contribution des principaux facteurs génétiques à la susceptibilité à la PR

**Figure 8** : Les multiples actions de la protéine LYP sur la signalisation cellulaire spécifique du lymphocyte T

**Figure 9** : Intervention de la famille des molécules STAT dans la réponse immunitaire

**Figure 10** : Coopération entre les voies de signalisation médiées par les TLRs et intervention d'IRF5

**Figure 11** : Fonctions biologiques des gènes de l'immunité mis en évidence dans l'étude du transcriptome dans la PR

**Figure 12** : Association entre le taux d'expression et les genotypes du polymorphisme RNASE2<sub>rs2013109</sub>

**Figure 13** : Schéma explicatif du principe de la technologie TaqMan®

**Figure 14** : Schéma résumant la procédure de génotypage utilisant la technologie TaqMan®

**Figure 15** : Capture d'écran d'un résultat graphique de génotypage par TaqMan® obtenu par le logiciel SDS

## **LISTE DES TABLEAUX**

**Tableau 1** : Critères de classification d'une PR selon l'ACR (1987)

**Tableau 2** : Nouveaux critères ACR/EULAR de la polyarthrite rhumatoïde (2010)

**Tableau 3** : Principaux facteurs de susceptibilité génétique impliqués dans le déterminisme de la polyarthrite rhumatoïde

**Tableau 4** : Caractéristiques de la population d'étude cas-témoins Algérienne.

**Tableau 5** : SNPs et gènes sélectionnés pour l'analyse dans la population Algérienne

**Tableau 6**: Caractéristiques des patients PR dans les trois groupes Européens

**Tableau 7** : Gènes et SNPs sélectionnés et leur MAF respective dans la population caucasienne de référence

**Tableau 8** : Caractéristiques des sondes TaqMan® utilisées dans cette étude

**Tableau 9** : Caractéristiques de l'échantillon cas/témoins étudié

**Tableau 10** : Analyse d'association des 5 SNPs étudiés chez les cas et les contrôles Algériens

**Tableau 11** : Test d'association familiale pour les 10 SNPs étudiés dans l'échantillon 2

**Tableau 12** : Résultats de la réplication du test d'association familiale pour PGLYRP1<sub>rs2041992</sub>

**Tableau 13** : Analyse du risque relatif des génotypes (GRR) dans l'échantillon 2

**Tableau 14** : Analyse haplotypique dans l'échantillon 2 pour les 4 SNPs du gène RUNX3

**Tableau 15** : Analyse haplotypique dans l'échantillon 2 pour les 2 SNPs du gène PGLYRP1

## LISTE DES ABREVIATIONS

**ACPA** : Anticorps anti-peptides citrullinés  
**ACR** : American College of Rheumatology  
**ADNc** : Acide DésoxyriboNucléique Complémentaire  
**AFBAC** : Affected Family-Based Controls  
**AINS** : Anti-Inflammatoires Non-Stéroïdiens  
**ARNm** : Acide Ribonucléique Messenger  
**ASP** : Affected Sib Pairs  
**CD** : Cluster of Differentiation  
**Chr** : chromosome  
**cM** : Centimorgans  
**CMH** : Complexe Majeur d’Histocompatibilité  
**CNV** : Copy Number Variant  
**C5** : Complement component 5  
**CTLA4** : Cytotoxic Lymphocyte Associated 4 protein  
**DL** : Déséquilibre de Liaison  
**DMARDs** : Disease Modifying Anti-Rheumatic Drugs  
**EBV** : Virus d’Epstein-Barr  
**eQTL** : expression Quantitative Trait Loci  
**EULAR** : European League Against Rhumatism  
**FDR** : False Discovery Rate  
**FR** : Facteur Rhumatoïde  
**GWS** : Genome Wide Scan  
**GWAS** : Genome Wide Association Study  
**HLA** : Human Leucocyte Antigen  
**IBD** : Inflammatory Bowel Disease  
**IC** : Intervalle de confiance  
**IFN** : Interféron  
**IL** : Interleukine  
**IL2RB** : Interleukin 2 receptor beta  
**IRF5** : Interferon regulatory Receptor  
**Kb** : Kilobases  
**LED** : Lupus Erythémateux Disséminé  
**LT** : Lymphocytes T  
**MAF** : Minor Allele Frequency



**MAI** : Maladies Auto-Immunes  
**Mb** : Mégabases  
**MTX** : Méthotrexate  
**MTHFR** : Méthylène Tetrahydrofolate Réductase  
**NARAC** : the North American Rheumatoid Arthritis Consortium  
**NCBI** : National Center for Biotechnology Information  
**NK** : Natural Killer  
**OR** : Odds Ratio  
**OLIG3** : Oligodendrocyte lineage transcription factor 3  
**PTPN22** : Protein tyrosine phosphatase non-receptor 22  
**PAD** : Peptidyl arginine déiminases  
**PBMCs** : Peripheral blood mononuclear cells  
**PCR** : Polymerase Chain reaction  
**RFLP** : Restriction Fragment Length Polymorphism  
**PGRLYP1** : Peptidoglycan recognition protein 1  
**PNN** : Polynucléaires neutrophiles  
**PR** : Polyarthrite rhumatoïde  
**RA** : Rheumatoid Arthritis  
**S100A8** : S100 calcium binding protein A8  
**RUNX3** : Runt-related transcription factor 3  
**STAT4** : Signal transducer and activator of transcription 4  
**SE** : Shared epitope  
**SNP** : Single Nucleotide Polymorphism  
**TGF** : Transforming growth factor  
**TNFAIP3** : TNF- $\alpha$  induced protein 3  
**TRAF1** : TNF receptor-associated factor 1  
**TDT** : Test du déséquilibre de transmission  
**TNF $\alpha$**  : Tumor Necrosis Factor  $\alpha$   
**T-reg** : Lymphocytes T Régulateurs  
**Th** : Lymphocytes T Helper  
**TLR** : Toll-like Receptor  
**UTR** : Untranslated region  
**VEGF** : Vascular Endothelial Growth Factor  
**WES** : Whole Exome Sequencing  
**WTCCC** : The Wellcome Trust Case Control Consortium

# ***PREAMBULE***

*Les premières lois de l'hérédité ont été découvertes en 1865 par le moine Gregor Mendel (1822-1884). Mendel a mené ses travaux dans le but ultime de résoudre la grande énigme des ressemblances familiales. Ses observations ont abouti à l'établissement des fondements de l'hérédité, devenus plus tard la base de la génétique moderne, mais ce n'est qu'au début du 20<sup>ème</sup> siècle, que l'œuvre de Mendel sera reconnue. Les différents chercheurs qui ont pris le relais, et les multiples études ayant été entreprises dans ce domaine ont révolutionné le monde de la biologie en introduisant de nouvelles disciplines d'études qui se sont adaptées au cours du temps, aux différentes découvertes et aux multiples progrès technologiques.*

*Les traits mendéliens, ou monogéniques, sont caractérisés par l'effet de mutation(s) rare(s) dans la population générale ayant des effets forts sur la variabilité du trait. Les traits complexes sont plutôt expliqués par plusieurs gènes interagissant entre eux et avec des facteurs de l'environnement. Ce ne sont pas des mutations délétères rares, mais plutôt des variants, généralement fréquents, dont la présence n'est ni nécessaire ni suffisante au développement de la maladie chez un individu. De ce principe, a priori très simple, est née l'épidémiologie génétique, dont l'un des objectifs est de caractériser la composante génétique de la variabilité de traits chez l'homme.*

*Les traits complexes tels que la polyarthrite rhumatoïde, certains cancers ou encore le diabète sont des problèmes de santé publique à cause de la gravité des symptômes et du grand nombre de personnes touchées dans la population générale.*

*Encouragé par le succès de la cartographie génétique pour des traits mendéliens, tels que la maladie de Huntington, certaines formes de cancers du sein ou de la maladie d'Alzheimer, les efforts se sont tournés vers la cartographie de gènes pour des traits complexes dont l'étiologie est multifactorielle. Il faudrait cependant rappeler que l'idée de base de la génétique d'un caractère sur une échelle continue est d'étudier sa variation pour expliquer les différences, et l'étude de cette variation nous amène alors à la diviser en composantes attribuables chacune à une cause différente.*

*A partir de ces différents concepts, nous nous sommes intéressés à l'étude d'une maladie remplissant tous ces critères, la polyarthrite rhumatoïde (PR).*

*Le présent document est le fruit de quatre années de travail de thèse, au cours desquelles j'ai eu l'occasion d'explorer de multiples aspects de la génétique humaine et de la biologie moléculaire. Le manuscrit est organisé en quatre parties :*

*Pour commencer, une grande introduction suivie d'une partie de description des patients, des populations et des méthodes utilisées lors de notre travail. Nous verrons en troisième lieu les résultats obtenus et la discussion qui en découle, puis nous clôturerons par une conclusion générale et les perspectives qui s'offrent à l'issue de cette thèse.*

*1/ Le premier chapitre est une grande introduction, elle même composée de trois parties. Dans la première, nous introduisons les éléments de base en génétique nécessaires à la compréhension de notre travail. Cette partie contient une brève description des approches génomiques des maladies complexes suivie d'un détail des différents types d'analyses génétiques, des outils utilisés pour ces analyses et aussi les multiples approches, dont les plus modernes d'entre elles, en terme de stratégies d'études d'association et/ou de liaison génétique.*

*Nous abordons par la suite, dans la seconde partie du chapitre 1, un exemple de maladie complexe et qui n'est autre que la PR. Nous avons regroupé dans cette partie un maximum d'informations nécessaires à la compréhension des différents mécanismes de la pathogénèse de la PR, de son développement et de ses traitements actuels. L'élément le plus important dans cette partie consiste en la description des facteurs de risque et notamment, la composante génétique, qui compose l'essentiel de notre travail.*

*La troisième et dernière partie de ce premier chapitre nous amène à entamer le côté pratique de notre travail puisqu'il s'agit des objectifs de la thèse, où nous détaillerons les différents points importants de nos travaux de recherche.*

*2/ Le deuxième chapitre de cette thèse est consacré à la description des populations impliquées dans notre étude, ainsi qu'aux différentes techniques utilisées aussi bien en génotypage qu'en analyse de données statistiques.*

*3/ Le troisième chapitre est réservé aux résultats obtenus et à la discussion de ces résultats. L'ensemble des travaux réalisés a fait l'objet de la publication de deux articles dans deux revues à comités de lecture internationaux : ACTA Reumatologica Portuguesa et Scandinavian Journal of Rheumatology.*

*4/ Le quatrième chapitre qui clôture notre travail est consacré à la conclusion générale que nous pouvons tirer à partir des résultats obtenus et aux différentes perspectives ouvertes par les travaux réalisés au cours de cette thèse.*

*5/ Le cinquième chapitre contiendra les annexes où figureront les protocoles expérimentaux détaillées des techniques citées dans le chapitre 2.*

# ***CHAPITRE 1 – INTRODUCTION***

**I. LES APPROCHES  
GENOMIQUES DES MALADIES  
COMPLEXES**

Les impacts socio-économiques des maladies humaines à caractère complexe sont considérables partout dans le monde. L'identification des facteurs responsables de ces maladies devrait permettre d'établir des moyens de prévention, de diagnostic et de traitement, nécessaires pour limiter la gravité des conséquences de ces pathologies sur l'être humain et sur la santé publique. Le défi à relever est donc de taille.

## **1 Les maladies complexes : définition**

Bien des facteurs contribuent à la santé et à la maladie chez l'humain. Notre environnement et notre biologie sont deux facteurs qui influencent fortement notre santé. On a longtemps cru que les maladies étaient causées exclusivement par notre environnement (comme la Tuberculose pulmonaire ou la Grippe) ou entièrement par notre biologie (comme les maladies héréditaires telles que la Fibrose Kystique ou la Dystrophie Musculaire de Duchenne). Nous constatons maintenant que bons nombres de maladies humaines proviennent d'une interaction complexe entre notre biologie et notre environnement, ainsi que de bien d'autres facteurs. C'est ce qu'on appelle les maladies à caractère complexe.

Ces maladies complexes sont à l'intersection entre la composante génétique polygénique et la composante environnementale (pouvant être aussi multiple), d'où leur nom de maladies multifactorielles. Cette complexité est à l'opposé de ce qui est observé dans les maladies monogéniques, dites mendéliennes ou à hérédité simple. Il en résulte donc une altération de la corrélation qui peut exister entre le phénotype et les génotypes, entre ce qui est visible et ce qui est caché.

Les gènes impliqués dans les maladies complexes sont appelés gènes de prédisposition, car chacun des variants de ces gènes, considérés isolément, n'est ni indispensable ni suffisant pour entraîner la maladie. Autrement dit, cette complexité peut déboucher sur une grande



hétérogénéité. En effet, il peut exister des individus atteints mais non porteurs des génotypes de prédisposition, mais aussi le contraire, des individus porteurs de ces facteurs mais non atteints de la pathologie en question. Cette description, aussi simple qu'elle puisse être, n'est pas toujours si évidente car dans certains cas de maladies, des facteurs environnementaux influencent l'âge d'apparition de la maladie ou même la gravité des symptômes.

L'avènement des outils modernes de biologie moléculaire au cours des années 1970, puis leur application à la génétique humaine dans les années 1980 ont conduit à une augmentation rapide et significative des connaissances, à la fois sur l'étiologie de plusieurs maladies héréditaires et sur le fonctionnement du vivant, au sens très large.

Ces connaissances ont par ailleurs permis une meilleure compréhension de l'étiologie de nombreuses maladies humaines, amenant ainsi pour certaines, à des possibilités de traitements voire à une guérison complète. Par exemple, le traitement ou la prévention efficace contre certaines maladies infectieuses est possible grâce au développement de l'hygiène, à la prise d'antibiotiques ou de vaccins.

Les outils de la biologie et de la génétique moléculaires ont été appliqués successivement à l'étude de la génétique des maladies monogéniques classiques (1-3), à la cartographie systématique du génome humain (4-6), puis à son séquençage (7, 8) et, enfin, à l'étude des maladies fréquentes à forte composante génétique (ou maladies génétiques complexes) (9-12). L'identification des mutations géniques, puis de la fonction du (des) produit(s) de ces gènes, permet de révéler la biologie normale et celle de la maladie. Les gènes impliqués dans les maladies complexes (ou polygéniques) sont pressentis comme ayant des effets individuels plus modestes que les gènes à forte pénétrance, impliqués dans les maladies héréditaires classiques mendéliennes (13).

En corollaire, l'identification des gènes associés aux maladies complexes demande d'adopter des stratégies et des méthodes différentes, aussi bien en terme de recrutement des sujets de recherche qu'en terme de méthodologies utilisées (9).

## **2 Caractérisation de la composante génétique**

Il existe deux approches méthodologiques, basées fondamentalement sur le même principe, permettant de détecter des facteurs génétiques de maladies : les études de liaison et les études d'association. Ces approches sont des outils puissants et complémentaires pour caractériser la composante génétique des maladies.

Généralement, on utilise dans un premier temps les études de liaison pour localiser des régions chromosomiques pouvant contenir un gène expliquant une part de la variabilité du trait. Puis, dans un deuxième temps, les études d'association pour préciser plus finement l'emplacement du gène ou du locus étudié. Un marqueur génétique peut apporter de l'information, de manière globale, à deux niveaux :

\* **au niveau familial**, si les allèles des deux loci se transmettent de façon non indépendante au cours des générations. On évalue l'indépendance de transmission des allèles à l'aide d'une **analyse de liaison** génétique. Les marqueurs utilisés pour l'analyse de liaison sont en général des marqueurs très polymorphes comme par exemple les microsatellites.

\* **au niveau de la population**, s'il existe une association préférentielle (déséquilibre de liaison) entre les allèles du variant génétique causal et les allèles de marqueurs génétiques. On évalue la force de cette association à l'aide d'une **analyse d'association**, qui est définie comme « l'analyse de l'association d'un variant avec le phénotype étudié par comparaison de la fréquence allélique du variant chez les cas et chez les témoins ». On utilise en général des SNPs (*Single Nucleotide Polymorphisms*).

## 2.1 Analyses de liaison génétique

Les analyses de liaison génétique testent la co-transmission d'une région chromosomique avec un phénotype. A partir de données familiales, elles permettent de tester l'indépendance de transmission entre des marqueurs polymorphes et la maladie dans les familles, et si ce n'est pas le cas, de situer le locus de maladie par rapport aux marqueurs. En effet, plus un marqueur est proche du locus de la maladie, plus la probabilité qu'il en soit séparé par un crossing-over est faible et celle qu'il soit co-hérité avec lui élevée. Montrer que la transmission d'une pathologie et d'un marqueur, dont la localisation est connue, se fait de façon non indépendante permet donc de positionner approximativement le locus de la maladie sur le génome.

### 2.1.1 Différents types d'analyses de liaison

Les méthodes statistiques utilisées pour tester l'hypothèse de liaison sont de deux types : les méthodes paramétriques et celles non-paramétriques, dites modèle-indépendantes.

Les méthodes paramétriques, comme le *Lodscore* (14) ont été développées pour l'analyse des maladies monogéniques, elles supposent la connaissance d'un modèle de transmission pour le caractère étudié (mode de transmission, fréquence de l'allèle de la maladie et pénétrance). Lorsque ce modèle est renseigné de façon incorrecte, il en résulte une perte de puissance pour détecter la liaison existante (15). Dans le cadre des maladies multifactorielles, le modèle de transmission sous-jacent est inconnu, les analyses de liaison paramétriques ne sont donc pas adaptées.

Les méthodes non-paramétriques ont été développées spécifiquement pour l'étude de maladies complexes. Ces méthodes recherchent la corrélation de transmission entre marqueurs et la maladie sans modéliser cette transmission *a priori*. Les premières méthodes non-paramétriques

s'intéressaient uniquement aux paires de germains atteints (16), la statistique a ensuite été étendue à des configurations familiales plus larges.

### **2.1.2 Marqueurs génétiques**

La notion de marqueur génétique a été introduite en 1980. Il s'agit d'une séquence d'ADN repérable spécifiquement. En cartographie génétique, le marqueur est utilisé pour "baliser" le génome. En contrôle du transfert de gène, le marqueur est un gène associé au gène d'intérêt, codant une caractéristique détectable facilement et précocement, facilitant le repérage des cellules au sein desquelles la transgénèse a réussi. La détection d'un marqueur génétique pouvait s'effectuer par hybridation avec une sonde complémentaire, ou par son expression phénotypique.

Les marqueurs peuvent siéger dans des régions codantes ou être "anonymes" en correspondant à des séquences non traduites (les variations alléliques de ces marqueurs ne sont pas décelables au niveau phénotypique mais au niveau de la séquence). Ils peuvent être de différentes natures : RFLP (*Restriction Fragment Length Polymorphism*) ou *polymorphismes de longueur de fragments de restriction*, microsatellites, SNPs, etc... . Nous en retiendrons pour la description les deux qui nous importent le plus : les microsatellites et surtout les SNPs.

#### **2.1.2.1 Marqueurs de type microsatellite**

Les microsatellites appartiennent à la famille des séquences répétées au sein du génome plus particulièrement ce que l'on appelle les VNTR (*Variable Number of Tandem Repeat*) ou encore les HVR (*Highly Variable Repeat*). Ils sont constitués de motifs nucléotidiques associés en tandem 10 à 50 fois. Ce motif est souvent un dinucléotide : C-A ; mais il existe de nombreux motifs tri, tétra, penta, voire hexa-nucléotidiques ou encore plus

complexes. Un allèle de microsatellites est un nombre de répétitions d'un motif simple, donc les allèles sont de longueurs différentes : c'est un cas de polymorphisme de longueur.

Des techniques d'amplification et de séparation en fonction de la taille du fragment étudié permettent de mettre en évidence des variations de longueur pour un même microsatellite. Ces variations constituent autant «d'allèles» très utilisés dans les études de liaison et dans la cartographie génétique. Dans les années 1990, c'est le polymorphisme des microsatellites qui a retenu l'attention des chercheurs. De tels marqueurs sont ubiquitaires dans les génomes des vertébrés. Ils existent en particulier dans le génome humain.

Les microsatellites, bien que plus denses au niveau centromérique et télomérique, sont bien répartis dans l'ensemble du génome. On estime en effet qu'il y a dans la région HLA (*Human Leucocyte Antigen*) un microsatellite tous les 30kb. Ce sont des marqueurs pertinents dans les études de cartographies de gènes impliqués dans une pathologie, et dans la génétique des populations, ceci étant dû à leur haut degré de polymorphisme et leur densité importante dans le génome (7).

Les études d'évolution moléculaire ont démontré, sur de courtes périodes de temps, le très important taux de mutation de ce type de séquence (17-19). Seuls sont exploitables les microsatellites polymorphiques porteurs d'informativité (ou communément appelés marqueurs informatifs).

Les marqueurs microsatellites bien que délaissés au profit des SNPs offrent encore une alternative intéressante aux SNPs dans les études d'associations génétiques. A l'origine de la première cartographie du génome humain, ces marqueurs ont également permis d'identifier les gènes impliqués dans un grand nombre de maladies monogéniques (19).

### 2.1.2.2 Polymorphismes mono-nucléotidiques (SNPs)

Les SNPs (*Single Nucleotide Polymorphisms*) constituent la forme la plus abondante de variations génétiques dans le génome humain. Ils représentent plus de 90% de toutes les différences entre individus. C'est un type de polymorphisme bi-allélique ponctuel de l'ADN au niveau d'une paire de bases.

Dans deux génomes humains tirés au hasard, 99,9% de la séquence d'ADN est identique. Les 0,1% restants contiennent des variations de séquence dont le type le plus commun est le polymorphisme pour un nucléotide (*SNP*).

Les SNPs sont stables, très abondants et distribués uniformément dans tout le génome. Ces variations peuvent être des marqueurs de la diversité entre populations ou individus, être associées à des différences de susceptibilité à des maladies, de sévérité des maladies et/ou à la réponse individuelle aux médicaments. De nombreux SNPs ont été mis en évidence lors de l'étude de sujets sains et malades portant des allèles différents d'un gène donné. Il existe aujourd'hui de nombreuses méthodes pour mettre en évidence des différences d'un nucléotide. Une des plus performantes, ayant émergé durant les dernières années, et qui offre de bons rendements c'est l'usage des puces à ADN.

Les SNPs sont des variations bi-alléliques d'un seul nucléotide dans la séquence d'ADN pour lesquels les différents allèles existent dans la population générale à une fréquence supérieure à 0,01. Ils constituent la majorité du polymorphisme génique (90% du polymorphisme du génome humain) (20). Les fréquences des différentes combinaisons de SNPs possibles dans le génome ne sont pas les mêmes. En effet, les SNPs impliquant une transition de C en T sont les plus fréquents, constituant les 2/3 des SNPs (20).

Les SNPs sont les marqueurs les plus utilisés dans les études génétiques actuellement et ceci pour plusieurs raisons : Ils sont très abondants dans le génome (environ  $3 \times 10^7$ , soit un SNP toutes les 100 pb en moyenne) et parce qu'ils sont potentiellement fonctionnels contrairement aux marqueurs STR, sauf exception. En effet, les SNPs peuvent affecter l'expression d'un gène (eQTL, expression Quantitative Trait Loci) ou d'un exon particulier (sQTL, splicing QTL) dans le cas de variants non codants. La stabilité de l'ARNm ou sa vitesse de traduction peuvent également être affectées (un codon stop prématuré induisant le mécanisme de NMD : « *non sens mediated decay* »).

Les SNPs codants peuvent également modifier la demi-vie d'une protéine (domaine PEST, riche en résidus proline P, acide glutamique E, sérine S et thréonine T), sa localisation (domaine NLS pour *nuclear localization sequence*), ou encore modifier sa fonction protéique. De plus, les avancées technologiques ont permis le génotypage d'un grand nombre de SNPs avec une efficacité plus grande et un coût de plus en plus faible permettant ainsi l'analyse des SNPs du génome entier (21).

Plusieurs consortiums ont été mis en place depuis 1999 tels que le SNP Consortium (TSC), le « *International Human Genome Project Consortium* » (HGP) et plus récemment le « *International HapMap Consortium* » pour identifier l'ensemble des SNPs existant dans le génome et ainsi créer des cartes génétiques de haute densité, facilement accessibles via des bases de données publiques telles que le NCBI (« *National Center for Biological Information* ») ou *HapMap*. On dénombre approximativement 3 millions de SNPs communs (fréquence de l'allèle mineure >20%), c'est-à-dire en moyenne un tous les 1 kb au sein du génome (22). Au total, on a identifié plus de 30 millions de SNPs chez l'Homme <http://www.ncbi.nlm.nih.gov/mailman/pipermail/dbsnpannounce/2010q4/000097.html>.

Néanmoins, il existe des variations de la répartition de ces SNPs au sein du génome car ils sont plus fréquents dans les introns et les régions

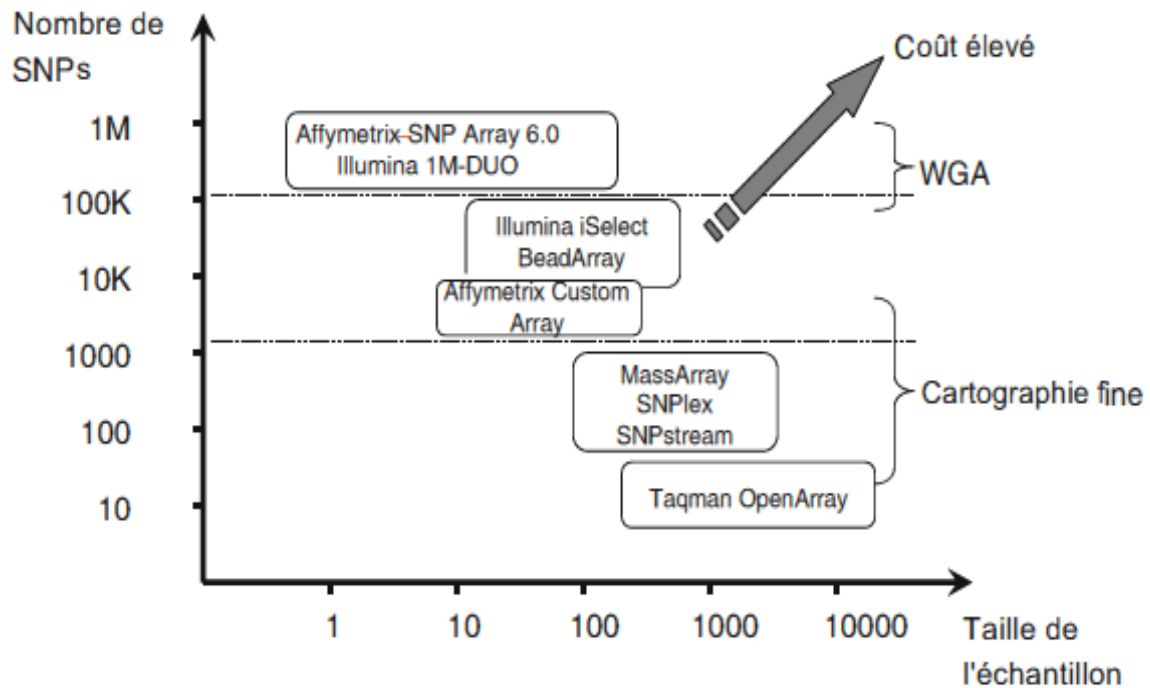
intergéniques que dans les exons et surtout les parties codantes. De plus, les SNPs synonymes sont plus fréquents que les SNPs non synonymes (faux-sens) (23). Dans près de 50% des cas, la fréquence de l'allèle minoritaire (MAF ou *Minor Allele Frequency*) dépasse 5%, indiquant que ces allèles n'exercent pas d'effet délétère fortement contre-sélectionné.

Le faible taux de mutations par génération ( $10^{-8}$ ) des SNPs ainsi que leur abondance dans le génome humain permettent leur utilisation dans différentes applications comme l'étude du déséquilibre de liaison et des haplotypes au sein de régions génomiques (projet *HapMap*) ou l'étude de la diversité génétique des populations humaines et des espèces. Ceci permettrait ainsi de suivre les migrations de populations, d'identifier des phénomènes de sélection positive et de préciser l'évolution des espèces. Ces polymorphismes constituent donc un outil exceptionnel pour l'identification des facteurs génétiques des maladies complexes par analyse d'association. Leur détection peut se faire par de nombreuses méthodes dont certaines à très haut débit et à coût de plus en plus faible.

Les SNPs sont bi-alléliques et donc moins polymorphes que les microsatellites. Des panels d'environ 10000 SNPs se sont avérés nécessaires pour caractériser par analyse de liaison le génome aussi précisément qu'une carte de 1000 microsatellites. Ainsi, du fait de l'automatisation et du faible coût de leur génotypage, les SNPs ont remplacé les microsatellites dans les analyses du génome entier.

La densité des panels de SNPs par puce proposés pour un même budget, augmente de façon exponentielle depuis quatre ans. Des panels de 100000 SNPs ont été proposés dès 2004 (24). En 2006 la compagnie *Illumina*® a commercialisé un panel de 300 000 SNPs et la compagnie *Affymetrix*® son premier panel de 500 000 SNPs. Les derniers panels proposés par ces deux compagnies comportent désormais plus d'un million de marqueurs (**figure 1**).





**Figure 1 : Vue d'ensemble des plateformes en fonction du débit de polymorphismes SNP et de la taille de l'échantillon. Adapté d'après Ding et al. (25) : Les plateformes sont sélectionnées sur la base des coûts de fonctionnement raisonnables.**

WGA : Whole Genome Array.

Les SNPs fréquents ne sont probablement pas les seuls polymorphismes génétiques impliqués dans la pathogenèse des maladies multifactorielles (26). En effet, des SNPs avec une MAF faible ou encore d'autres variants rares pourraient également jouer un rôle important dans ces maladies, tels que les mutations récurrentes du gène NOD2 associé à la maladie de *Crohn* (27).

Actuellement, il existe au moins deux aspects intéressants de la recherche en génomique. L'un est le développement constant de technologies plus abordables et de meilleure qualité. L'autre caractéristique est l'acquisition de plus en plus de nouvelles connaissances sur la structure et la fonction des gènes. Un exemple, les CNV (*Copy Number Variant*), qui sont des variations qui impliquent des segments d'ADN plus grands, allant de quelques kilobases à quelques mégabases. Ces variations sont beaucoup

moins fréquentes dans le génome humain que les SNPs, et leur rôle dans le développement de maladies monogéniques et multifactorielles a également été mis en évidence ces dernières années (28).

### **2.1.3 Cartes génétiques et cartes physiques**

Il existe deux types de cartes, ou plus exactement deux types de distances : distance physique et distance génétique.

La première se définit comme étant la longueur de la séquence d'ADN séparant deux points du génome, mesurée en nombre de paires de bases azotées (pb), kilobases (1 kb=1000 pb) ou mégabases (1 Mb =  $10^6$  pb). Elle nécessite l'identification moléculaire des locus et le séquençage du brin d'ADN les séparant. La généralisation de ces distances à l'ensemble du génome humain est très récente.

La seconde (entre deux locus liés) est mesurée par le taux de recombinaison méiotique, qui est par définition inférieur à 50%. Le taux de recombinaison n'est pas une mesure additive, du fait des recombinaisons multiples; différentes transformations de ce taux ont donc été proposées (par Haldane en 1919 et Kosambi en 1944) pour préciser la notion de distance génétique (29, 30). La plus communément utilisée est la mesure additive exprimée en centimorgans (cM). Une distance de 1 cM entre deux locus correspond à une probabilité de 1% pour eux d'être séparés par un crossing-over.

La recombinaison survenant de façon hétérogène le long du génome, la correspondance entre distance génétique et physique varie donc selon les régions considérées. En moyenne, 1 cM équivaut à 0,88 Mb (il est considéré grossièrement que 1 cM équivaut à 1 Mb) avec une variabilité selon le sexe. En effet, chez les femmes les recombinaisons sont deux fois plus fréquentes que chez les hommes, ce qui entraîne pour un même chromosome des tailles différentes en cM, mais pas en Mb.

## **2.2 Analyses d'association génétique**

La recherche d'associations, entre des marqueurs génétiques et des maladies, est une des voies possibles pour identifier des gènes de susceptibilité aux maladies. La technique d'étude est relativement simple : on compare la fréquence d'un marqueur génétique chez les sujets atteints et chez les témoins (étude cas-témoins). L'analyse statistique est fondée sur le tableau de contingence où figurent, pour les cas et les témoins, les fréquences de chaque génotype ou de chaque allèle au locus considéré. Le marqueur utilisé peut être un polymorphisme sans relation causale avec la pathologie ou bien un polymorphisme d'un gène candidat.

Un résultat positif (différence de distribution allélique entre les cas et les témoins) suggère que le marqueur est soit directement impliqué, soit en déséquilibre de liaison avec une ou plusieurs variation(s) génétique(s) causale(s).

Ces dernières années nous avons assisté à une véritable explosion des études d'association à grande échelle, permettant une recherche exhaustive des polymorphismes génétiques pouvant être impliqués dans les mécanismes biologiques à l'origine de traits complexes. Ceci s'est fait en parallèle avec les progrès techniques de la biologie moléculaire, qui ont largement contribué à la diminution à la fois du coût et du temps de génotypage.

Bien qu'une grande partie des régions chromosomiques identifiées ne contiennent pas de gènes codant des protéines (moins de 5% du génome code des protéines), ces études ont permis de révéler de nouveaux gènes candidats potentiels pour de nombreuses maladies.

### 2.2.1 Notion de déséquilibre de liaison

Un déséquilibre de liaison entre deux loci est défini par l'existence d'une combinaison d'allèles à ces loci plus fréquente que celle attendue sous l'hypothèse d'indépendance. Deux marqueurs génétiques en déséquilibre de liaison sont en général très proches sur le même chromosome (31), mais nous ne pouvons détailler le déséquilibre de liaison sans parler de l'équilibre de Hardy-Weinberg.

La théorie de l'équilibre de liaison a été développée au début du 20<sup>ème</sup> siècle par *Godfrey Hardy*, un mathématicien anglais, et *Wilhelm Weinberg*, un médecin allemand. Ils ont montré qu'à partir d'une population idéalisée, c'est-à-dire d'une population d'appariements aléatoires, en absence des forces de l'évolution, les fréquences d'allèle et les fréquences de génotype demeureront constantes de génération en génération. La loi de Hardy-Weinberg de l'équilibre génétique fournit un modèle mathématique permettant d'analyser des changements de fréquence d'allèle durant l'évolution dans une population, en examinant les relations entre des fréquences relatives de deux ou plusieurs allèles.

*A contrario*, on appelle déséquilibre de liaison (noté  $D$ ), le paramètre qui mesure l'écart à l'association aléatoire de deux allèles situés à des loci différents sur les gamètes. Il mesure le degré de dépendance statistique entre deux caractères, qui sont ici les allèles à deux loci différents au sein d'une même population et dans la même génération.

Toutefois, il est important de souligner que l'existence d'un déséquilibre de liaison entre deux loci n'implique pas une liaison génétique entre les deux loci, et que la mise en évidence d'une association n'est pas synonyme de relation de causalité, ce qui implique en général la démonstration cellulaire et biochimique du mécanisme d'action de la variation étudiée. Si la population témoin est suffisamment importante et représentative, il est possible de calculer le risque relatif de présenter la

pathologie en fonction de l'état homozygote ou hétérozygote de l'allèle en question.

Les études d'association sont particulièrement intéressantes lorsqu'il existe plusieurs loci impliqués et que l'effet quantitatif de chaque allèle délétère est trop faible pour être facilement analysé par une étude de liaison familiale. La probabilité de détecter une association dépend de l'intensité du déséquilibre de liaison entre le marqueur et la variation causale, de la fréquence de la variation dans la population, de l'effet observable de la variation (sa pénétrance) sur la pathologie, et enfin de la distance génétique entre les marqueurs et le locus impliqué dans la maladie lorsqu'on ne se réfère pas directement au gène candidat.

Dans ces conditions, un résultat négatif ne permet pas d'exclure le rôle d'un locus. Il se peut en effet qu'il ne soit pas en déséquilibre de liaison avec une variation morbide, même très proche. De même, une mutation à effet biologique fort et plus généralement une variation génétique, mais rarement représentée dans la population, peut aboutir à un test statistique négatif et passer inaperçue.

A partir du moment où, dans une étude, on se base sur des hypothèses, il faut se mettre des barrières à ne pas dépasser pour être vraisemblablement exact dans ses calculs. Ces barrières seront appelées dans le cas présent l'intervalle de confiance.

L'intervalle de confiance (IC) à 95% est un intervalle de valeurs qui a 95% de chance de contenir la vraie valeur du paramètre estimé. Avec moins de rigueur, il est possible de dire que l'IC représente la fourchette de valeurs à l'intérieur de laquelle nous sommes certains à 95% de trouver la vraie valeur recherchée. L'intervalle de confiance est donc l'ensemble des valeurs raisonnablement compatibles avec le résultat observé (l'estimation ponctuelle). Il donne une visualisation de l'incertitude de l'estimation.

Des intervalles de confiance à 99% ou à 90% sont parfois utilisés. La probabilité (degré de confiance) de ces intervalles de contenir la vraie valeur est respectivement de 99% et 90%. L'intervalle de confiance est constitué des valeurs qui ne sont pas statistiquement significativement différentes du résultat observé.

Les bornes supérieures et inférieures sont donc les valeurs les plus éloignées du résultat qui ne lui sont pas statistiquement différentes. Par contre, les valeurs situées à l'extérieur de l'intervalle sont statistiquement différentes du résultat observé. Ainsi, la borne supérieure est la plus grande valeur non significativement différente de la valeur observée.

### **2.2.2 Etudes d'association cas-témoins**

Parmi les interactions possibles entre phénomènes liés à la santé et processus démographique, il en est une que l'on peut rencontrer en épidémiologie génétique dans le cadre des études d'association entre un gène – ou un marqueur génétique – et une maladie.

Une méthode classique pour aborder ce type d'étude est la méthode cas-témoin qui consiste à comparer un échantillon de malades atteints de la maladie étudiée (pour laquelle on s'interroge et on recherche une éventuelle cause génétique) et un échantillon de personnes non malades comparables, ayant globalement les mêmes caractéristiques (d'âge, d'origine ethnique, etc.).

Si l'on décèle qu'un marqueur génétique spécifique est beaucoup plus représenté (voire systématiquement présent) chez les individus malades et peu représenté (voire absent) chez les individus sains, la mesure statistique peut conclure à une association entre ce marqueur génétique et la maladie. Ce type d'enquête et de méthodologie suppose implicitement que cas et témoins soient issus d'une population homogène.

Cette méthode a soulevé une polémique par rapport au risque de biais qui lui était lié. En effet, si jamais il existait un effet de « structure » non connu de la population (ou effet de « stratification » selon le vocabulaire des épidémiologistes), la population supposée homogène d'où sont tirés les échantillons étant, en fait, un mélange de deux ou plusieurs populations aux caractéristiques très différentes. Les résultats des tests statistiques d'association pourraient être biaisés et conclure à tort à une association, le facteur ethnique ou « sous-population » jouant dans ce cas comme un facteur de confusion dans l'étude de l'association gène-maladie. Afin de contourner ce problème, on peut avoir recours à des tests statistiques qui tiennent compte de la possible stratification des populations étudiées (32).

Ce problème est important en pratique dans des pays où les populations sont composées de nombreuses communautés différentes, restées relativement fermées, et où les échantillons qui en sont issus sont susceptibles d'être très hétérogènes. Statistiquement parlant, dans une étude cas-témoins, nous allons remplacer le risque relatif par l'Odds Ratio (OR). L'Odds Ratio (OR) ou « rapport des cotes » est le rapport de l'odds de l'événement (sa cote) dans le groupe traité divisé par l'odds de l'événement dans le groupe contrôle (33). L'Odds est égale à  $c = r / (1 - r)$  où  $r$  est la fréquence de l'événement. Ainsi un Odds est le rapport du nombre de patients  $n$  présentant l'événement,  $r \times n$ , divisé par le nombre de patients ne présentant pas l'événement,  $(1-r) \times n$ . Par exemple, un odds de 0,25 correspond au rapport 2/8 et signifie que pour 2 patients présentant l'événement, 8 ne le présentent pas ( $0,25 = 2/8 = r/(1-r)$ ). Dans la même situation, le risque est 0,20 (2/10). Un Odds peut aussi être interprété de la façon suivante : dans un groupe, pour 100 patients ne présentant pas l'événement étudié,  $100 \times c$  le présentent.

Il faut cependant prendre garde aux faux positifs qui découlent d'un biais de sélection des cas et des témoins, dont la sélection peut être inappropriée parce qu'ils diffèrent par l'âge, la répartition des sexes et des groupes ethniques, ou la présence d'une autre pathologie associée au critère

de choix - autant de paramètres susceptibles d'entraîner des différences de fréquences génotypiques non attribuables à la pathologie.

### **2.2.3 Test du déséquilibre de transmission (TDT)**

Tel que déjà mentionné, en génétique, une des problématiques est de trouver si la maladie qui fait l'objet de l'étude est d'ordre génétique ou pas, en plus de la recherche de l'éventuelle existence d'un ou plusieurs gènes responsables de la maladie. Pour vérifier si la maladie est d'ordre génétique, des études sur des jumeaux ou sur des familles sont faites dans un premier temps pour ensuite permettre, par des études de liaison par exemple, de recenser des gènes candidats, c'est-à-dire des gènes qui pourraient causer la maladie.

Comme les marqueurs génétiques ne sont pas toujours disponibles dans le gène candidat, on peut utiliser des marqueurs génétiques situés près du locus candidat et dont on connaît la position; alors il s'agit de montrer qu'il y a une liaison entre ces marqueurs génétiques et le statut de la maladie. Il existe plusieurs méthodes testant cette liaison, dont les plus connues sont celles basées sur l'étude des paires de germains atteints « ASP » (*affected sib pairs*), ou les méthodes appelées « AFBAC » (*affected family-based controls*) qui comparent, en utilisant le test du  $X^2$ , le nombre d'allèles transmis aux enfants malades (cas) au nombre d'allèles transmis aux enfants non malades (témoins).

Ces méthodes ont presque toujours échoué dans leur objectif de trouver un déséquilibre de liaison (34), car elles sont sensibles à la stratification de la population; c'est-à-dire qu'on peut trouver une association qui est due, non pas au déséquilibre de liaison, mais à la stratification de la population. En réponse à ce problème, *Spielman* et al. ont proposé le test du TDT (*Transmission Test for Linkage Disequilibrium*), qui est un test qui n'est nullement affecté par la présence d'une stratification de la population. Ce



dernier teste la liaison en présence d'une association allélique entre le marqueur et la maladie. Un des succès de ce test se réfère à l'étude génétique du diabète (34).

Le TDT s'applique aux traits binaires, c'est-à-dire au cas où l'état de l'individu est décrit par deux valeurs: malade ou non malade. Pour utiliser le TDT, il suffit de disposer de familles ayant au moins un enfant malade, contrairement aux tests cités ci-dessus, qui eux nécessitent des données sur des paires de frères et sœurs malades (dans le cas de la méthode ASP), ou des membres affectés et non affectés d'une même famille (dans le cas de la méthode AFBAC); ce type de données n'est pas toujours facile à collecter surtout lorsqu'il s'agit d'une maladie rare, d'où l'avantage du TDT.

## **2.3 Choix de marqueurs et stratégies d'études : « Association Versus Liaison »**

La recherche et la caractérisation des déterminants génétiques impliqués dans les maladies génétiques "complexes" ont été initiées dans la décennie 1990 par des travaux qui ont concerné, entre autre, le diabète de type 2. Compte tenu de la multiplicité des locus potentiellement impliqués, les méthodes employées dans les maladies monogéniques ont du être adaptées à l'étude de l'hérédité polygénique. On retiendra de cette époque deux méthodes majeures: la stratégie gène candidat et l'approche génome entier (*Genome Wide Scan*).

### **2.3.1 Approche gène candidat**

Cette stratégie consiste à rechercher et analyser les variants génétiques dans des gènes dont la fonction pourrait jouer un rôle dans la pathologie en question. Elle repose sur une bonne connaissance de la physiologie et de la fonction du gène mis en examen et de ses interactions

avec les autres partenaires impliqués dans la même voie métabolique qui pourrait être incriminée dans la pathologie. Par exemple le gène *GLUT4* codant un transporteur de glucose a été gène candidat dans un contexte de diabète de type 2.

Les gènes impliqués dans la réponse immunitaire et l'inflammation ont été candidats dans le cadre du psoriasis. Une fois le ou les variants génétiques identifiés (en général par séquençage direct de l'ADN de sujets atteints et de sujets non atteints) il reste à conclure sur chacun quant à son rôle (ou son absence de rôle) dans la maladie génétique complexe.

Comme dans une maladie complexe chaque variant génétique ne possède qu'un effet minime sur la variation du trait phénotypique et donc sur la survenue ou non de la maladie, cette phase nécessite une approche épidémiogénétique sur des groupes importants de patients atteints et de sujets témoins indemnes de la maladie (études cas-témoins).

La question posée étant : le variant génétique est-il plus fréquent chez les sujets présentant la maladie que chez les témoins ? La réponse à cette question ne peut être objectivée que par un test statistique de comparaison de fréquences au risque 5%, dans cet exemple, un test du  $\chi^2$ . Si la réponse est oui, on est en droit de conclure (au risque de 5%) que le variant en question est l'un des déterminants génétiques de la maladie, ou bien il est en déséquilibre de liaison génétique avec l'un des déterminants génétiques de la maladie.

N'ayant aucune information objective et biologique quant au retentissement du variant génétique sur la fonction du gène et sur la maladie, nous ne pouvons à ce stade que constater qu'il existe une association significative au sens statistique entre la présence du variant génétique et la présence de la maladie. C'est pourquoi ces analyses sont appelées analyses d'association. Ces analyses d'association peuvent aussi se réaliser par des études familiales: tests de déséquilibre de transmission (*Transmission Disequilibrium Test TDT*). On objective par un test statistique la réponse à la

question : le variant génétique est-il transmis, d'un parent hétérozygote à un enfant atteint, plus souvent que ne le voudrait le hasard ?

Ces approches gène candidat qui ont mis en évidence de nombreux variants génétiques impliqués dans des maladies génétiques complexes, présentent le défaut de ne pouvoir cibler et analyser que les gènes connus dont la fonction est évidemment connue et qui présentent un rapport *a priori* évident avec la pathologie.

### 2.3.2 Approche GWS “Genome Wide Scan”

L'approche génome entier consistait à génotyper chez les sujets d'un grand nombre de familles (présentant des patients atteints) un maximum ( $\pm 1000$ ) de marqueurs microsatellites très polymorphes, recouvrant la totalité du génome. C'est une méthode de génétique inverse et sans *a priori*, contrairement à l'approche gène candidat. L'étude est basée sur l'analyse de la liaison génétique entre chaque marqueur microsatellite et la maladie "complexe". Par analogie avec le calcul du *Lodscore* employé dans les maladies monogénique et qui quantifie la liaison génétique entre un locus et la maladie, la méthode *Genome Wide Scan* quantifie la liaison génétique entre chaque marqueur microsatellite et la maladie complexe.

Dans l'exemple suivant, extrait de *Styrkarsdottir et al.*, il a été étudié 1100 marqueurs microsatellites en relation avec l'ostéoporose (35). Les résultats montraient clairement que de nombreux loci présentaient une liaison génétique au moins "suggestive" avec l'ostéoporose, en accord avec la nature polygénique de la maladie. Le maximum de liaison génétique était observé sur le chromosome 20.

Une vue détaillée des résultats concernant le chromosome 20 montre que la zone de liaison génétique couvrait presque la moitié du chromosome 20, ce qui voudrait dire 50cM soit 50 Mégabases d'ADN. Les investigations

ont été restreintes au pic de liaison génétique compris entre les marqueurs D20S882 et D20S900, la zone à explorer à la recherche d'un (ou plusieurs) variant génétique en rapport avec la maladie restait de 1 726 485 bp (1,7Mb) et même si cela ne représentait plus que 3% du chromosome 20 et 0,05% du génome, cette zone à explorer était encore très grande à l'échelle du biologiste moléculaire puisque, compte tenu de la densité des gènes sur le chromosome 20, elle était censée contenir en approximation statistique 21 gènes (36).

Il n'est pas rare que la stratégie gène candidat soit utilisée en seconde intention pour restreindre le nombre de gènes dont l'investigation devrait être poussée (stratégie dite gène candidat positionnel). Dans l'exemple présenté, la région de 1.7 Mb contenait en fait six gènes répertoriés et connus en 2003 (*BMP2*, *CHGB*, *LOC51605*, *C20orf154*, *C20orf155*, et *C20orf42*).

Compte tenu de son rôle dans la formation de l'os et la différenciation des ostéoblastes, le gène *BMP2* (*Bone Morphogenetic Protein 2*) est apparu le meilleur gène candidat. Des investigations plus fines ont permis de mettre en évidence une mutation Ser37Ala du gène *BMP2*, fortement associée à l'ostéoporose mais qui n'expliquait que 30% de la liaison génétique sur le chromosome 20.

On perçoit ici les limites de la méthode qui sont :

(1) la relative grande taille des régions de liaison génétique (qui n'est pas sans rapport avec la distance entre les microsatellites) et donc le grand nombre de gènes à explorer en détail à la recherche de variants génétiques causals,

(2) la part de liaison restant inexpliquée par les variants génétiques associés à la maladie.

D'autre part, la stratégie « *Genome Wide Scan* » repose toujours sur la séquence « analyses de liaison puis clonage positionnel » :

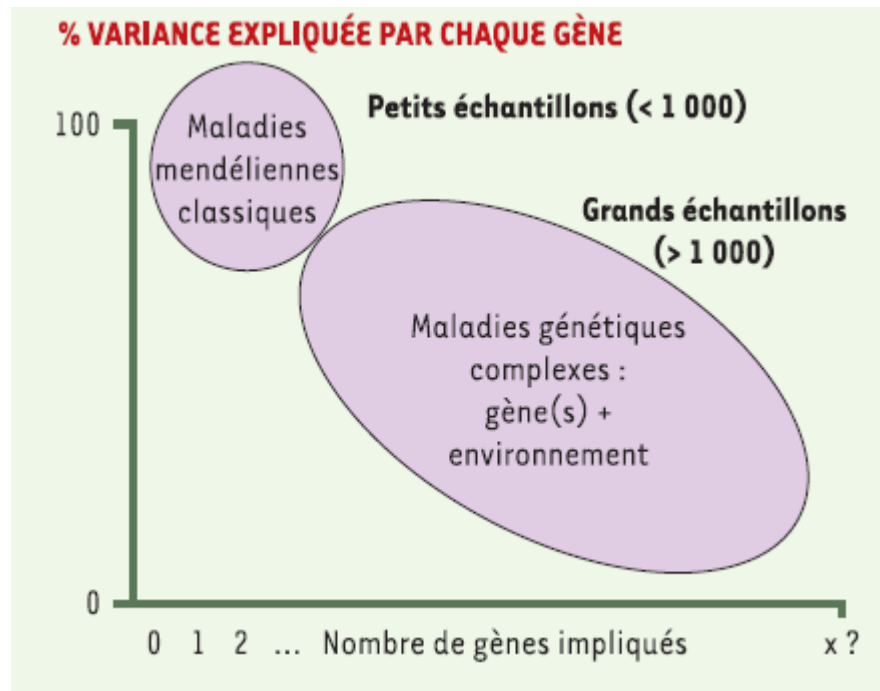
(1) identification de loci de susceptibilité par analyse de liaison génétique dans des familles,

(2) analyse de liaison avec des marqueurs plus rapprochés dans les loci précédemment identifiés ("*fine mapping*") pour restreindre la taille du locus de susceptibilité,

(3) puis analyse de gènes ou prédictions de gènes candidats positionnels, dont l'implication dans la maladie est validée ou écartée sur la base d'analyses d'association entre des variants génétiques et la maladie.

De ces études GWS, on peut déjà avancer l'idée que certaines maladies multifactorielles auront une base génétique avec un ou plusieurs variants génétiques à effets majeurs (comme dans la maladie de Crohn), alors que pour d'autres maladies multifactorielles il sera question d'une multitude de variants génétiques ayant tous un effet mineur (comme cela semble être le cas pour le diabète de type 2 et l'obésité).

Par extension il faut admettre que la dichotomie sémantique entre maladies monogéniques et maladies polygéniques (ou encore multifactorielles) est finalement artificielle. Il faudrait donc distinguer entre la mutation d'un gène qui explique toute la variance du trait phénotypique dans une maladie monogénique et une affection multifactorielle dont la composante génétique résulte d'une myriade de variants génétiques à effets faibles (**figure 2**). Entre les deux, il existe un *continuum* de situations qui expliquent les atteintes de sévérité variable en fonction du terrain ou « *background* » génétique dans des maladies monogéniques et les maladies polygéniques avec l'effet d'un ou plusieurs gènes majeurs (37). En effet, il est rapporté de nombreux cas de variants génétiques qui modulent la sévérité d'une maladie monogénique comme par exemple dans la mucoviscidose (38, 39).



**Figure 2. Caractéristiques comparées des maladies héréditaires classiques mendéliennes et des maladies génétiques complexes (37) :** L'abscisse représente le nombre estimé de gènes impliqués dans la maladie, tandis que l'ordonnée indique la proportion de la variance phénotypique expliquée par chacun de ces gènes. Le nombre habituel d'individus à étudier pour identifier ce ou ces gènes est aussi indiqué, montrant que les études visant à identifier les gènes impliqués dans des maladies génétiques complexes nécessitent l'analyse de plus de sujets, étant donné l'effet plus modeste de chaque gène impliqué.

### 2.3.3 Approche GWAS "Genome Wide Association Study"

Le développement des technologies de séquençage et de génotypage à haut débit à partir des années 2000, basées sur les "puces à ADN" (*DNA chips*) a ouvert la voie à l'obtention rapide d'un très grand nombre de génotypes. Il est alors devenu envisageable de génotyper un grand nombre de marqueurs génétiques chez un grand nombre de sujets.

Cette opportunité technologique a ouvert la voie aux GWAS qui se résument à l'étude d'association à très grande échelle: autant d'analyses d'association cas-témoin que de variants génétiques analysés. En effet la méthode consiste à génotyper un maximum de marqueurs génétiques (de 300 000 à 1 000 000 selon le type de puce à ADN) chez un grand nombre de

sujets atteints de la maladie génétique complexe et un grand nombre de sujets témoins.

Les variations d'un nucléotide (SNP), malgré leur nature bi-allélique qui contraste avec le caractère multi-allélique des microsatellites, ont été retenues compte tenu de leur très grand nombre sur le génome humain offrant ainsi une très bonne couverture du génome. En effet les SNPs constituent les variants génétiques les plus fréquents. Par ailleurs leur nature bi-allélique les rend plus accessibles au génotypage de masse au moyen des puces à ADN que les marqueurs microsatellites qui présentent un très grand nombre d'allèles.

La méthode *Genome Wide Scan* qui s'intéressait à la co-transmission entre des régions du génome (repérées par les marqueurs microsatellites) et le phénotype "malade" ou "non malade", reposait donc sur la liaison génétique entre les marqueurs microsatellites et les loci en rapport avec la maladie "complexe".

La méthode GWAS consiste en des études d'association entre un allèle donné d'un SNP et le phénotype "malade" ou "non malade." La méthode GWAS repose donc sur le déséquilibre de liaison génétique qui existe au sein du génome humain entre un allèle d'un marqueur génétique (SNP) et, à moins que le marqueur testé ne soit directement causal, un variant génétique potentiellement impliqué dans la maladie. Le variant étudié constitue une partie du déterminisme génétique de la maladie complexe.

Un des points forts de la méthode est qu'elle permet de cribler le génome avec une haute densité de marqueurs : en effet, pour une puce ADN 550K (qui génotype 550 000 SNPs), la densité moyenne est de 1 marqueur SNP tous les 60 kb.

L'association mise à profit dans la GWAS est également fonction :

(1) de la distance entre le locus du SNP analysé et le locus où réside le variant génétique de susceptibilité à la maladie (ce paramètre sera moins limitant compte tenu de la grande densité de marqueurs),

(2) mais surtout sera fonction du déséquilibre de liaison (DL) génétique qui existe entre un allèle donné du locus du marqueur SNP et l'allèle de susceptibilité au niveau du locus où réside le variant génétique de susceptibilité à la maladie. Comme le DL est variable d'une population et d'une ethnie à une autre, les approches GWAS devront prendre en compte cette spécificité et inclure des sujets homogènes quant au DL (40, 41).

La méthode GWAS pourrait sembler "la" méthode définitive susceptible de repérer tous les déterminants génétiques d'une maladie, compte tenu de sa haute densité de couverture du génome. Néanmoins la méthode souffre d'un certain nombre de limites. Comme elle se résume à une succession d'études d'association cas-témoin, l'association d'un SNP avec la maladie est testée comme habituellement, au risque statistique 5% (0,05). Une association sera réputée significative quand il n'y a que 5% de chance que cette affirmation soit fausse (ce qui somme toute laisse 95% de chance d'avoir raison), c'est le risque d'erreur de type I. En d'autres termes il y a 5% de chance que l'association soit trouvée positive par hasard.

Quand, après avoir analysé deux SNPs, on évalue la probabilité que l'un d'entre eux soit associé par hasard, le risque n'est plus de 0,05 mais de 0,1. Il sera de 0,15 pour 3 SNPs. On voit qu'à partir d'une centaine de SNPs testés la probabilité que l'un d'entre eux soit associé à la maladie par hasard (calculée par la loi binomiale) est quasiment de 1. Une méthode classique pour prendre en compte ce problème est de diminuer le seuil de significativité individuel de 0,05 à 0,025 dans le cas de deux SNPs et ainsi de suite. C'est le principe de la correction de *Bonferroni* (42).



La question qui se pose, c'est : à combien faut-il baisser le seuil de significativité individuel de chaque SNP pour que globalement une fois les 550 000 SNPs testés, le risque que l'un d'entre eux soit positif par hasard soit encore  $\leq 0,05$  ? Le calcul montre que si on teste chaque SNP au risque  $0,0000001$  ( $10^{-7}$ ), la probabilité pour que l'un d'entre eux soit déclaré comme étant associé à la maladie par hasard est voisine de  $0,05$ , ce qui devient statistiquement acceptable.

Il apparaît donc que seuls les SNPs présentant une association très fortement significative ( $p \leq 10^{-7}$ ), pourront être considérés comme étant très vraisemblablement associés à la maladie. C'est le cas de polymorphismes qui sont en déséquilibre de liaison ou qui sont physiquement très proches d'un déterminant génétique de la maladie, une proximité qui sera fonction de la densité des marqueurs génétiques. La méthode GWAS aura donc tendance à ne pouvoir détecter que les effets génétiques importants. C'est ce que l'on peut présenter comme étant un "bruit de fond statistique".

#### **2.3.4 Approche WES "Whole Exome Sequencing"**

Malgré les avancées spectaculaires dans la connaissance des facteurs génétiques de susceptibilité aux maladies multifactorielles réalisées entre autres via les GWAS, pour une maladie donnée, l'ensemble des variants identifiés n'explique qu'une faible partie de la variance du phénotype (héritabilité). En moyenne dans les maladies multifactorielles à peine 10% de l'héritabilité est expliquée par les variants génétiques connus. La question est alors « sur quoi reposent les 90% manquants ? » et quelles méthodes d'investigation employer pour caractériser cette part manquante et ainsi expliquer 100% de l'héritabilité?

Une approche récente du problème repose sur une opportunité technologique. Partant du principe que les séquences codantes ne représentent qu'une faible partie du génome mais concentrent 85% des

mutations potentiellement responsables des maladies, des méthodes permettant de "capturer" l'ensemble des séquences codantes ("*Whole Exome*") et d'en déterminer la séquence ("*Whole Exome Sequencing*") ont été développées (43).

Ces méthodes de "*Whole Exome Sequencing*" (WES) permettent d'identifier en théorie la quasi-totalité des variations de séquences qui existent au niveau des séquences codantes entre des sujets atteints de la maladie multifactorielle et des sujets indemnes. Il en ressort clairement que cette approche WES est bien adaptée à la caractérisation des variants rares qui étaient laissés de côté par la méthode GWAS précédemment décrite.

### **3 Analyse du Transcriptome**

#### **3.1 Définition du Transcriptome**

Le transcriptome est l'ensemble des ARNm présents à un instant donné dans une cellule donnée. La quantité d'ARN à un instant donné est le résultat d'un équilibre entre la synthèse (par les ARN polymérases) et la destruction (par les RNAses). La transcription des gènes est le premier mécanisme moléculaire qui permet d'expliquer pourquoi deux cellules d'un même organisme, ayant un génome identique, sont pourtant différentes. En effet, la transcription peut permettre d'obtenir des ARNm (ARN messagers) différents à partir d'un même gène, grâce à l'utilisation de sites d'initiation de la transcription alternatifs ou à une polyadénylation alternative. Si les séquences des ARNm peuvent être différentes, leur quantité varie également, sous l'influence de différents facteurs de transcription et de modifications épigénétiques (notamment la méthylation de l'ADN et les modifications de structure de la chromatine).

Après la transcription, des régulations post-transcriptionnelles permettent également de réguler qualitativement et quantitativement l'expression des gènes : qualitativement avec les mécanismes d'épissage

alternatif et d'édition, quantitativement à travers une stabilité variable d'un ARNm à l'autre. Il est également important de noter que toutes les études portant sur le transcriptome mesurent des niveaux d'expression sur des populations de cellules. On mesure ainsi la moyenne de cette quantité d'ARNm sur l'ensemble des cellules analysées. Si des techniques permettent actuellement d'étudier l'expression des gènes à l'échelle d'une seule cellule (44), elles restent encore lourdes à mettre en œuvre. Elles permettent néanmoins d'étudier l'effet stochastique de l'expression des gènes, qui serait particulièrement important dans différents phénomènes biologiques, dont la différenciation cellulaire (45).

A l'échelle des populations de cellules, l'analyse de l'expression de certains gènes a permis de montrer qu'ils étaient importants pour l'auto-renouvellement et/ou la différenciation des cellules souches par exemple. En effet, il a été démontré que le devenir des cellules souches embryonnaires (CSE) est différent selon le niveau d'expression du gène Oct-4, pour ne citer que cet exemple (46). Cependant, il ne s'agit que d'un exemple, et d'autres études ont montré que d'autres gènes étaient impliqués dans le maintien de l'auto-renouvellement des CSE ou dans leur différenciation dans certains lignages. De ce fait, il serait donc intéressant d'avoir une vision globale de l'expression des gènes.

Comme les techniques permettant d'identifier les cellules souches ont évolué parallèlement à la caractérisation de ces cellules, il en est de même pour les techniques d'analyse de l'expression des gènes : si l'on peut analyser l'expression, gène à gène, on peut maintenant également étudier simultanément l'expression d'un grand nombre de gènes.

Cette définition est théorique dans la mesure où toutes les méthodes qui existent actuellement présentent des limites et ne permettent donc pas d'étudier l'expression de tous les gènes en même temps. Par extension on parle encore de transcriptome, même s'il ne s'agit pas de l'ensemble des ARNm au sens strict du terme, mais plutôt d'un ensemble (particulièrement important) d'ARNm.

### 3.2 Méthodologies d'étude

Au cours des dernières années, des études de profils d'expression génique utilisant la technologie des puces à ADN ont été développées pour la recherche clinique à grande échelle et systématique afin d'identifier les gènes impliqués dans les maladies complexes. À l'heure actuelle, les puces couvrant l'ensemble du génome humain sont devenues disponibles dans le commerce et leur potentiel pour identifier l'activité d'un gène anormal dans une maladie est maintenant bien reconnu. L'analyse de l'expression par les puces est basée sur l'hybridation d'un jeu ordonné de sondes attachées à une surface avec une cible constituée d'ARNm isolé d'une population cellulaire donnée. En général, ce sont des ARNm isolés dans différentes situations biologiques, par exemple, avant et après un traitement ou bien chez des malades comparés à des sujets sains. L'hybridation reflète l'abondance relative de chaque ARNm et conduit à l'identification de gènes sur- ou sous-régulés par rapport à l'échantillon de référence. En regroupant des ensembles de gènes exprimés de manière différentielle selon leur fonction, des informations peuvent être obtenues sur les voies principales liées à une maladie ou à un traitement.

Jusqu'alors, l'étude des transcrits passait par une approche analytique d'un gène, et c'est par l'accumulation de ces expériences unitaires qu'une compréhension *a posteriori* de la cellule était envisagée. Les nouvelles approches globales quant à elles fournissent une image à un « temps T » du niveau d'expression de plusieurs milliers de gènes dans un type cellulaire et un contexte physiologique et/ou pathologique donné. Cette vision globale instantanée est très importante en particulier pour le transcriptome, pour lequel au temps T+1, l'image obtenue peut être très différente. Dans l'étude globale des transcrits, on distinguera les *microarrays* à une couleur (type *Affymetrix*<sup>®</sup>) où l'on mesure une valeur d'expression absolue et les puces deux couleurs (par exemple la plateforme *BeadChip* de *illumina*<sup>®</sup> ou encore *Agilent*<sup>®</sup>) qui sont basées sur l'expression relative d'un échantillon test par rapport à un échantillon référence.

Dans tous les cas, l'expérience commence par une extraction des ARN suivie de la synthèse des ADNc, leur marquage et l'hybridation sur les puces à ADN, lavages et séchage des lames puis lecture par un scanner. Le signal généré pour chaque *fluorochrome* est converti en fausses couleurs (généralement rouge pour la cyanine 5 et vert pour la cyanine 3). Les deux images ainsi obtenues sont superposées. Quant à l'analyse proprement dite des données d'expression, le but, dans la grande majorité des expériences, est de trouver les gènes dont l'expression varie entre deux groupes (par exemple un groupe témoin et un groupe traité par un médicament).

Initialement, la sélection des gènes « signatures » (gènes sous- ou/et sur-régulés) a été basée sur l'utilisation d'un seuil de « *Log ratio* ». Pour chaque gène, c'est le rapport entre ses niveaux d'expression mesurés dans des conditions différentes qui est informatif : on parle de facteur de régulation ou variation de ratio « *fold change* ». Avec cette méthode de variation de ratio, on se place à un seuil (souvent 2 en valeur absolue) et on sélectionne tous les gènes dont la variation de ratio est supérieure en valeur absolue à cette valeur seuil.

Cependant, ce n'est pas un test statistique et il n'intègre pas la variance (47). Par exemple, la variabilité des différences d'expression sur l'ensemble de la puce n'est pas prise en compte. Ainsi, un gène (par exemple un facteur de transcription) dont la variation n'est que de 1,5 fois peut avoir des conséquences drastiques.

D'autre part, une différence d'expression de 2 peut résulter d'un ratio d'intensité de 10 / 5 ou 10 000 / 5 000. Cette méthode ne prend pas en compte les réplicats biologiques dans le calcul des probabilités d'expression différentielle. Or, en utilisant les répétitions, il est possible de déterminer si un gène est ou n'est pas différentiellement exprimé, en utilisant les tests d'hypothèses (tests statistiques).

Il est important de noter que la comparaison des niveaux d'expression absolus ne s'effectue généralement pas directement entre les

gènes, mais se limite plutôt entre les différents échantillons hybridés. La raison généralement invoquée est que d'un gène à un autre, les unités de la mesure de l'expression ne sont pas les mêmes (48). Par conséquent, un *Microarray* n'est généralement pas utilisé pour répondre à la question « le gène A est-il plus exprimé que le gène B », mais pour répondre à des questions telles que « le gène A répond-il à la stimulation X? », « le gène A répond-il plus que le gène B à la stimulation X? » ou encore « l'expression du gène A est-elle corrélée à celle du gène B? ». Ces dernières portant sur les niveaux d'expression relatifs plutôt qu'absolus.

Il est aussi important de mentionner qu'en pratique, l'analyse de l'expression différentielle suit le paradigme *gène par gène*. En effet, bien que l'expression génique soit un phénomène coordonné, et que par conséquent les niveaux d'expression des gènes soient fortement dépendants, notre connaissance de cette dépendance (donc de la distribution jointe...) est à ce jour plutôt limitée.

L'analyse de l'expression différentielle suit habituellement la procédure suivante :

*Formuler un test d'hypothèse statistique.* On définit ici quantitativement ce que l'on entend par expression différentielle en formulant une hypothèse nulle et une hypothèse alternative. Par exemple, on testera l'hypothèse nulle que la différence entre les moyennes de deux groupes (ou le coefficient de régression d'un modèle linéaire) est égale à zéro. Cette hypothèse nulle est la plus courante, mais d'autres sont possibles, comme par exemple de tester si la différence de moyennes dépasse un certain seuil arbitraire ou encore un test sur la différence des médianes.

*Utilisation de réplicats* et ce pour apporter une correction de fond aux différents tests réalisés et aux données transcriptomes obtenues. Ces réplicats peuvent être soit *biologiques* (échantillons différents pour un même phénotype) ou encore *techniques* (plusieurs extraits d'un même échantillon).

Ces deux approches permettant l'amélioration de la sensibilité de l'analyse. Les résultats obtenus pourront être aussi validés par l'analyse en PCR (*Polymerase Chain reaction*) quantitative en temps réel d'un certain nombre de gènes pour une analyse de type contrôle de qualité, par exemple un échantillon minimum de 20 gènes.

*Calculer une statistique pour l'expression différentielle pour chaque gène.* Cette statistique peut être par exemple la simple différence des moyennes (le *fold-change*), le *t* de Student, la différence des médianes, une statistique non paramétrique, etc. Si l'un de ces tests est applicable, il faudra alors comparer *la valeur de la statistique obtenue précédemment* avec la distribution de cette dernière sous l'hypothèse nulle. Par exemple, le *t-test* de Student suit une distribution bien connue du même nom. Cette étape fournit habituellement une  $P_{value}$ , qui correspond à la probabilité d'erreur de type I sous l'hypothèse nulle.

*Ordonner les gènes selon la significativité de leur test d'hypothèse.* Il faut noter que dans le cas des *microarrays* d'*Affymetrix*, l'ordonnement par significativité est souvent le même que celui de la valeur absolue de la statistique puisque chaque test comporte le même nombre de degrés de liberté. On appelle aussi la statistique de l'expression différentielle « statistique d'ordonnement », même si à proprement parler, l'ordonnement s'effectue la plupart du temps sur une valeur de significativité comme la  $P_{value}$ .

*Recalculer la significativité pour tenir compte que plusieurs hypothèses ont été testées simultanément* (problème des tests multiples). L'analyste sera fréquemment intéressé à obtenir une sous-liste parmi les gènes les plus significatifs. Même si la valeur-P d'un test pris individuellement donne la probabilité d'erreur de type I (hypothèse nulle rejetée faussement, ou faux positif), évaluer la significativité d'une liste de plusieurs tests demande un calcul supplémentaire. L'idée se comprend assez facilement : le bruit expérimental fait inévitablement en sorte que

certains gènes, en réalité non différentiellement exprimés, résultent en des valeur-p significatives.

La significativité d'une liste sélectionnée par un simple seuil sur la valeur-p n'est donc manifestement pas quantifiée par ce seuil. Une nouvelle définition de la significativité, ainsi qu'une procédure pour l'évaluer sont donc nécessaires. Par exemple, les procédures de type *FDR (False Discovery Rate)* évaluent la *proportion* de faux positifs d'une sous-liste donnée.

### 3.3 Relation Transcriptome / Variants génomiques

On constate lorsqu'on étudie le même type cellulaire dans plusieurs populations humaines (asiatiques, africaines ou caucasiennes), une grande disparité des niveaux d'expression avec près de 30% des transcrits qui montreraient des différences d'expression entre populations (49). Cette disparité peut s'expliquer par l'influence du mode de vie mais également par la diversité génétique de ces populations. Ces différences se répercutent ensuite sur les protéines (ou le phénotype) et le fonctionnement de la cellule, donc en particulier sur le transcriptome.

Outre la régulation directe de l'expression des gènes par le génotype, l'effet des variants génotypiques sur les maladies, à titre d'exemple, peut également passer par l'activation de certaines voies de signalisation spécifiques, et l'exemple qui traduit au mieux cette image sont les eQTL (*expression Quantitative Trait Loci*). En effet, les variations de séquence de l'ADN peuvent avoir des effets en cis -et trans- régulateurs relatifs à la variation de l'expression d'un gène local ou d'un ou de plusieurs gènes lointains, respectivement. Le test, pour un effet cis-régulateur potentiel, implique une analyse de l'association des seules paires de SNPs du gène qui correspondent sur le même chromosome à proximité les uns des autres, alors que le test pour un effet trans - régulateur potentiel implique une analyse d'association globale du type *Genome Wide Scan* entre un SNP et



tous les gènes, conduisant à la multiplication des tests statistiques nécessaires pour l'analyse.

L'interprétation biologique des résultats d'associations pangénomiques (données GWAS) (50, 51) est très difficile puisque la plupart des loci candidats sont localisés soit dans des régions génomiques très pauvres (les déserts de gènes) ou alors dans des régions avec un grand nombre tout aussi plausible de gènes responsables. Les avancées réalisées dans la compréhension des mécanismes de régulation des gènes sont intervenues pour apporter un espoir d'éclaircissement. L'expression différentielle de gènes a été proposée comme une solution intermédiaire porteuse d'informations pour aider à l'interprétation de ces résultats (52). Le plus souvent, la combinaison des données SNP pour des associations significatives avec l'expression des gènes a été employée pour expliquer certains des résultats d'associations de GWAS (53, 54). Cependant, l'ubiquité de la variation de régulation à travers le génome humain (49, 55) rend très probables la coïncidence de chevauchement des eQTLs avec des loci de traits complexes. Cette probabilité est une conséquence directe de la structure de corrélation dans le génome (de déséquilibre de liaison), ce qui rend des variants fonctionnellement indépendants statistiquement corrélés (56).

Liang et al. ont réalisé une étude dans laquelle ils ont réussi à augmenter la puissance de la cartographie eQTL et à construire une carte plus complète des polymorphismes SNPs qui influent sur l'expression des gènes (57). Cette étude a révélé l'existence de 14177 eQTL élargissant ainsi la précédente analyse réalisée par la même équipe (55) en incluant des données générées à l'aide de puces d'expression « génome entier ». Une base de données téléchargeable gratuitement a été mise en ligne incluant les données récentes des eQTL identifiés <http://www.hsph.harvard.edu/liming-liang/software/eqt/>.

De manière globale, dans l'étude de la composante génétique des maladies, le transcriptome peut jouer un rôle de deux façons :

1. Tout d'abord, le transcriptome constitue un phénotype intermédiaire intimement lié à l'ADN et au fonctionnement de la cellule. Il offre donc souvent les premiers éléments de compréhension des mécanismes reliant les nouveaux loci mis en évidence par les GWAS aux maladies multifactorielles. L'étude au niveau du génome entier de la régulation génétique de l'expression des gènes est donc un axe majeur de développement de la génomique moderne (58).
2. Ensuite, le transcriptome peut fournir des biomarqueurs à relier aux informations sur l'état de santé du patient (59) et à certaines expositions environnementales comme la réponse à certains traitements par exemple (60). Ainsi à condition de savoir caractériser l'impact des divers facteurs environnementaux, il est possible d'utiliser le transcriptome soit comme variable d'ajustement, soit comme un trait phénotypique à part entière.

**II. EXEMPLE DE MALADIE  
COMPLEXE : LA  
POLYARTHRITE RHUMATOÏDE**

## **1 Définition et épidémiologie**

La polyarthrite rhumatoïde (PR) est une maladie auto-immune inflammatoire, caractérisée par une inflammation chronique et destructive à terme des articulations. L'étiologie de cette pathologie reste encore à ce jour mal caractérisée. L'évolution de la PR se fait par poussées avec le risque de voir se développer des destructions articulaires, osseuses et cartilagineuses irréversibles.

La PR est le plus fréquent des rhumatismes inflammatoires chroniques avec une prévalence estimée en moyenne à 0,8% dans les pays industrialisés, et on estime qu'il y aurait en France entre 300 000 et 500 000 patients atteints (61). En Algérie, il n'existe pas encore d'estimation officielle, mais le nombre serait situé en estimation entre 600 000 et 1 million de personnes atteintes.

La PR peut survenir à tout âge mais on l'observe surtout entre 40 et 60 ans. Elle est à cet âge, quatre fois plus fréquente chez la femme que chez l'homme, cette différence de sexe s'atténue progressivement au-delà de 70 ans (62). La PR est 2 à 3 fois plus fréquente chez les parents de sujets atteints; sa concordance chez les jumeaux monozygotes est de 15 à 30% et, chez les jumeaux dizygotes, de 5 à 10% (63).

C'est une pathologie qui a pour siège principal l'articulation qui est formée de deux extrémités (les épiphyses) reliées entre elles par un manchon rigide et étanche, la capsule articulaire. Les épiphyses et la capsule délimitent un espace clos, baigné par le liquide synovial. Les surfaces internes de cette cavité sont recouvertes soit par le cartilage coiffant l'os sous-chordal soit par la membrane synoviale qui tapisse la capsule articulaire. L'examen clinique d'un patient souffrant de PR est dominé par des douleurs articulaires d'incidence "inflammatoire", associées à un gonflement des articulations ou synovite (64).

L'atteinte est le plus fréquemment bilatérale et symétrique; elle concerne principalement les petites articulations du carpe, les articulations méta-carpo-phalangiennes et méta-tarso-phalangiennes. L'inflammation de la gaine des tendons (ou tenosynovite) est fréquente au niveau des muscles fléchisseurs de la main au stade initial de la maladie (65).

Les critères d'établissement du diagnostic reposent sur un certain nombre de points appelés critères de classification définis par l'*American College of Rheumatology* (ACR) (**tableau 1**).

**Tableau 1 : critères de classification d'une PR selon l'ACR (1987) (66).**

Signe clinique	Description
1. Raideur articulaire matinale	raideur matinale articulaire ou périarticulaire, durant au moins une heure avant l'amélioration maximale.
2. Arthrite d'au moins 3 groupes articulaires	gonflement des tissus mous ou épanchement (et non pas saillie osseuse isolée) d'au moins 3 groupes articulaires touchés simultanément, observés par un médecin. Les 14 groupes possibles sont, à droite ou à gauche, les IPP (interphalangiennes proximales), les MCP (métacarpo-phalangiennes), les poignets, les coudes, les genoux, les chevilles et les MTP (métatarso-phalangiennes).
3. Arthrite des articulations des mains	gonflement d'au moins un groupe articulaire (voir définition en 2) parmi poignets, IPP ou MCP.
4. Arthrite symétrique	atteinte simultanée des mêmes groupes articulaires (voir définition en 2) des deux côtés du corps (l'atteinte bilatérale des IPP, des MCP, ou des MTP est acceptable sans symétrie absolue).
5. Nodules rhumatoïdes	nodules sous-cutanés, sur les proéminences osseuses, les surfaces d'extension, ou dans les régions para-articulaires, observés par un médecin
6. Facteur rhumatoïde sérique	mise en évidence de quantités anormales de facteur rhumatoïde sérique par une méthode dont les résultats sont positifs chez moins de 5 % des sujets témoins normaux.
7. Modifications radiologiques	modifications radiologiques typiques de polyarthrite rhumatoïde sur les radiographies de face des mains et des poignets, avec obligatoirement des érosions ou une décalcification osseuse évidente localisée des articulations atteintes ou de façon plus nette dans les régions adjacentes à ces articulations (des modifications d'allure seulement arthrosique ne conviennent pas).

\* Au moins 4 des 7 critères sont exigés

Les critères 1 à 4 doivent être présents depuis au moins 6 semaines

Les critères diagnostiques de l'ACR de 1987 utilisés jusqu'à peu étaient insuffisants pour le diagnostic positif d'une PR au début (66). Dès lors l'ACR et l'EULAR (*European League Against Rheumatism*) ont établi en 2010 une nouvelle démarche diagnostique (67).

Dans cette nouvelle classification (**tableau 2**), est définie comme PR la présence de synovites dans au moins une articulation, l'absence d'autre diagnostic pouvant expliquer les synovites et un score d'au moins 6 sur les 4 items suivants : nombres et sites articulaires atteints, anomalies sérologiques (facteur rhumatoïde « FR », anticorps anti-peptides citrullinés « ACPA »), présence ou non d'un syndrome inflammatoire (vitesse de sédimentation « VS », la *C Reactive Protein* « CRP »), la durée d'évolution des symptômes (plus ou moins de six semaines).

**Tableau 2 : Nouveaux critères ACR/EULAR de la polyarthrite rhumatoïde (2010) (67)**

	Score
<b>Articulations atteintes</b>	
1 grosse articulation	0
2 à 10 grosses articulations (symétriques ou non)	1
1 à 3 petites articulations	2
4 à 10 petites articulations	3
>10 articulations (dont au moins une petite)	5
<b>Autoanticorps (FR et ACPA)</b>	
FR- et ACPA-	0
FR+ et/ou ACPA+ à faible taux (1 à 3 x normale)	2
FR+ et/ou ACPA+ à taux élevé (>3 x normale)	3
<b>Durée d'évolution des synovites</b>	
< 6 semaines	0
≥ 6 semaines	1
<b>Marqueurs biologiques de l'inflammation (VS et CRP)</b>	
VS et CRP normales	0
VS et/ou CRP anormale	1

**Score ≥ 6 → Polyarthrite rhumatoïde**

FR : facteurs rhumatoïde, ACPA : anticorps anti-peptides citrullinés,  
VS : vitesse de sédimentation, CRP : protéine C réactive

## **2 La pathogenèse de la PR**

### **2.1 La physiopathologie**

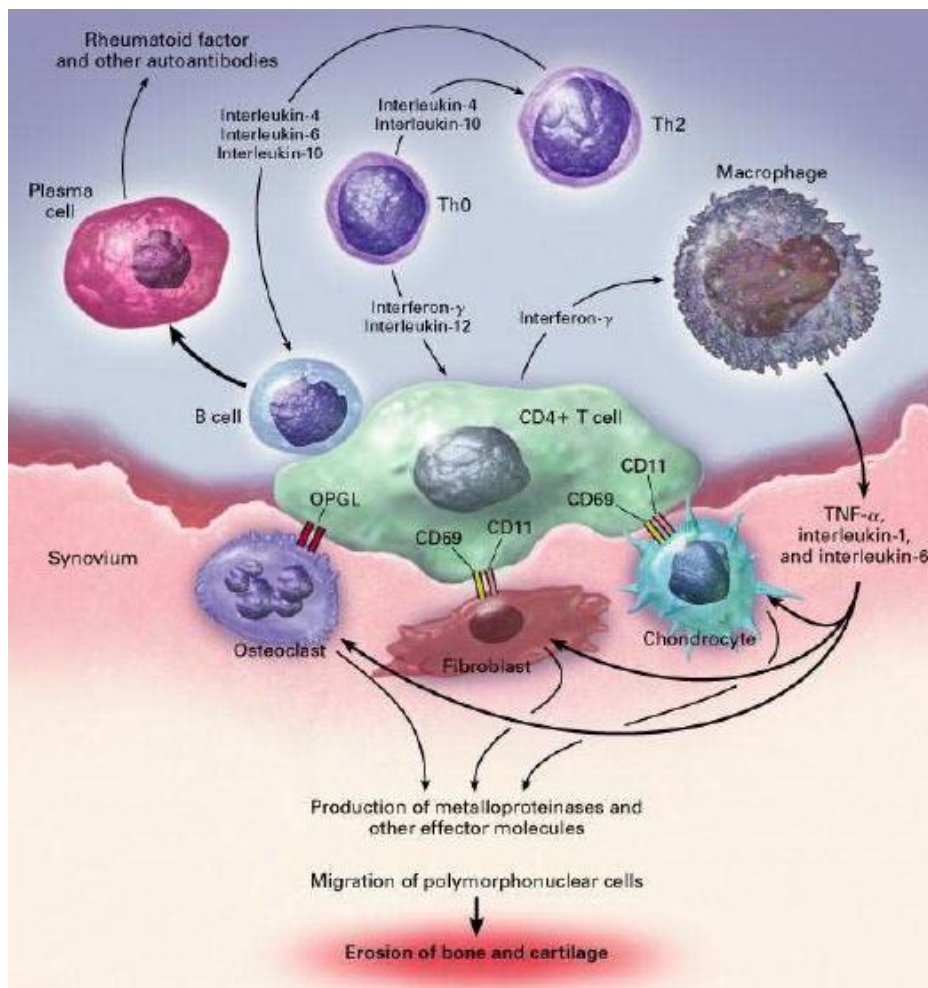
Les connaissances physiopathologiques de la polyarthrite progressent à petits pas, mais ceci n'est pas toujours utile pour le clinicien. Par exemple, pouvoir définir par avance (anticipation), une fois le diagnostic fait, les malades qui vont évoluer vers une forme grave (très destructrice, handicapante, systémique...) est capital mais encore impossible en clinique.

La physiopathologie de la PR est complexe et fait intervenir l'immunité innée, les cytokines et les différents acteurs de l'immunité acquise. D'autre part, on assiste à l'activation de cellules cibles de l'articulation telles que les synoviocytes, les chondrocytes et les ostéoclastes mais aussi à des anomalies des voies intracellulaires de signalisation.

Un des premiers mécanismes important est l'inflammation rhumatoïde, qui comme toutes les inflammations, est caractérisée par une vasodilatation, la formation d'un œdème et une infiltration cellulaire des cellules immunitaires (monocytes/macrophages, cellules dendritiques, lymphocytes T et B, neutrophiles et plasmocytes).

Ce qui nous permet de distinguer l'inflammation dans le cas précis de la polyarthrite rhumatoïde d'une autre inflammation (qui en général est un phénomène naturel de défense), c'est l'activation « incontrôlée » des deux types de réponses immunitaires, à la fois innée et acquise. Ceci va se traduire par une réaction inflammatoire « exagérée », en particulier de la membrane synoviale (68).

Au niveau cellulaire, on remarque au cours de l'évolution d'une PR que la synoviale rhumatoïde est infiltrée par des lymphocytes T CD4+, des lymphocytes B et des macrophages qui s'organisent en agrégats lymphoïdes avec parfois des centres germinaux (**figure 3**).



**Figure 3 : Représentation schématique des acteurs cellulaires et médiateurs inflammatoires impliqués dans la physiopathologie de la PR (69).**

L'évolution de la PR se caractérise également par une prolifération de la couche bordante de l'articulation qui est composée de synoviocytes et de macrophages mais aussi par une prolifération importante de néo-vaisseaux.

Les mécanismes moléculaires et cellulaires en jeu sont communs à l'ensemble des réactions inflammatoires. La phase initiale vasculaire permet un recrutement cellulaire et une activation rapide au niveau du site inflammatoire. Elle nécessite la mobilisation d'effecteurs pré-formés contenus dans les granules des mastocytes et des polynucléaires neutrophiles (PNNs). La dégranulation du mastocyte permet de sécréter immédiatement de puissants médiateurs vasodilatateurs (histamine, sérotonine et bradykinine) qui favorisent la domiciliation des cellules effectrices.



Les granulations du PNN renferment des substances de défense comme les protéines du complément et la NADPH oxydase dédiée à la production de formes réactives de l'oxygène (ou ROS pour *reactive oxygen species*).

Les métalloprotéases matricielles (ou MMPs pour "*Matrix Metalloproteinases*") favorisent la migration cellulaire. Dans le cas du PNN, la collagénase MMP-8 et la gélatinase MMP-9 dégradent respectivement les collagènes de type I et de type IV. Les protéases et les ROS peuvent aussi exercer des effets délétères par des dommages excessifs et incontrôlés de la matrice extracellulaire des tissus cibles.

Après une deuxième phase d'amplification faisant intervenir des facteurs chimiotactiques de nature différente et des cytokines inflammatoires comme le "*Tumor Necrosis Factor  $\alpha$* " (TNF $\alpha$ ), l'interleukine1 $\beta$  (IL-1  $\beta$ ) et l'IL-6, le macrophage permet la terminaison de la réaction inflammatoire par phagocytose et présente les peptides immunogènes afin d'initier une réponse immunitaire adaptative (**figure 3**).

La chronicité de ces phénomènes, *a priori* naturels, conduit aux conséquences destructrices et invalidantes de la PR que nous verrons plus loin dans la partie entretien de l'auto-immunité.

## 2.2 La PR, une maladie auto-immune

Le système immunitaire est une organisation de cellules et de molécules spécialisées ayant pour vocation principale la défense de l'organisme contre une grande variété de substances antigéniques étrangères telles que les agents infectieux (virus, bactéries ou parasites), et aussi contre les cellules tumorales.

Dans certains cas, il arrive que ce système immunitaire censé jouer un rôle protecteur vis-à-vis du *soi*, dirige une réponse immunitaire contre ce même *soi*, par un recrutement intensif de lymphocytes T (LT) auto-réactifs. Ces derniers sont des LT qui sont capables de réagir avec des peptides du *soi* présentés dans un contexte HLA bien précis, détaillé plus loin.

La discrimination entre le *soi* et le non-*soi* est une des questions les plus fondamentales de toute l'immunologie. Les fonctions du système immunitaire sont fondées sur la reconnaissance grâce à des récepteurs plus ou moins spécifiques d'antigènes ou tout au moins de certains motifs, déterminants antigéniques ou épitopes, propres aux agents exogènes.

Chez les patients atteints de PR, la proportion des LT auto-réactifs serait plus élevée que chez les sujets normaux et serait due à une anomalie de la sélection thymique. Dans le thymus, les LT qui reconnaissent les peptides du *soi* meurent en général par apoptose. Cette sélection clonale dite « négative » serait déficiente chez les malades atteints de PR et contribuerait à la survie d'un plus grand nombre de LT auto-réactifs. Cette hypothèse est soutenue par les différences observées entre le répertoire des LT des malades atteints de PR et celui des sujets normaux (70).

Il existe, entre autres, des indications en faveur de l'existence d'une activation par l'antigène de clones B ou T. La sélection des clones B auto-réactifs par l'antigène est démontrée par la présence de mutations somatiques dans les gènes codant les auto-anticorps au niveau des CDR (*complementarity determining regions*) (71). S'agissant des cellules T, on retrouve, par l'étude du répertoire  $V\beta$ , des restrictions d'utilisation de ces gènes qui témoignent d'une expansion oligoclonale plutôt que polyclonale (72).

Rappelons, en outre, l'association entre certains gènes de classe I ou de classe II du complexe majeur d'histocompatibilité (CMH) et la plupart des maladies auto-immunes, qui suggère la présentation de certains peptides par ces molécules de classe I ou II (73).

### 2.2.1 Rupture de tolérance et mimétisme moléculaire

Bien que l'auto-réactivité soit physiologique, le système de régulation de cette auto-immunité peut être défaillant. Apparaît alors une auto-immunité pathologique. Le fait que ces LT auto-réactifs soient activés lors d'un processus physiopathologique de la PR a fait l'objet de plusieurs études qui partent toutes d'un concept premier : la rupture de tolérance immunitaire.

Cette rupture de tolérance est bien sûr subordonnée au mécanisme d'échappement des LT auto-réactifs à la sélection clonale « négative », lors de la maturation thymique décrite plus haut. Les mécanismes impliqués dans cette activation erronée des cellules de l'immunité et la rupture de leur tolérance sont ceux qui vont causer le chamboulement total de l'activité du système immunitaire en favorisant la reconnaissance des antigènes du soi ou en augmentant l'immunogénicité de l'organe cible.

Par ailleurs, une homologie de séquence, ou « mimétisme moléculaire », entre l'auto-antigène et des antigènes étrangers viraux ou bactériens a pu être invoquée (74). Ceci est le cas du syndrome de Guillain-Barré, provoqué par une réaction croisée d'antigènes bactériens (*Campylobacter jejuni*) avec les gangliosides de la myéline des nerfs périphériques (75) ou du rhumatisme articulaire aigu dû à une infection par des streptocoques hémolytiques  $\beta$  de groupe A présentant des déterminants croisés avec l'endocarde (76). Peu d'exemples de pathologies humaines sont toutefois recensés, et ce mécanisme est considéré comme minoritaire.

En partant du principe de mimétisme moléculaire, différentes équipes de recherche se sont intéressées à l'étude de la présence de différents microorganismes dans les articulations de patients atteints de PR. A partir de ses constatations, le virus d'Epstein-Barr (EBV) a fait son apparition dans le monde de la PR (77-79).

Il existe de multiples relations entre le virus EBV et la polyarthrite rhumatoïde (PR) : les malades atteints de PR ont des titres d'anticorps

dirigés contre les antigènes d'EBV plus élevés par rapport à des sujets sains (80) ; la PR est caractérisée par un défaut de contrôle cellulaire de l'infection par EBV (81) ; des similitudes antigéniques entre EBV et des composants du soi ont été observées (séquences glycine-alanine d'EBNA-1 (*Epstein-Barr nuclear antigen-1*) et protéines de la synoviale (82); glycoprotéine gp110 et épitope partagé (83)); des réponses cellulaires dirigées contre des protéines du cycle répliatif et contre gp110 ont été observées au niveau du liquide articulaire de PR; EBV (ses antigènes ou son génome) a été mis en évidence au niveau de la synoviale de PR par hybridation in situ ou PCR, avec un rendement variable (84) ; la charge virale en EBV des cellules mononucléées du sang circulant est élevée dans la PR comparativement aux témoins; EBNA-1 est susceptible de subir une citrullination et des anticorps antipeptides citrullinés peuvent être induits par EBV (85) ; des lymphomes compliquent la PR et certains sont associés à EBV (86).

Ainsi, les liens EBV-PR sont nombreux et complexes, mais la relation de causalité directe n'a pas été formellement établie. Le virus d'Epstein-Barr pourrait agir indirectement dans le schéma physiopathologique de la PR, par l'intermédiaire du mauvais contrôle de l'infection et d'une exposition chronique à ses antigènes, sources de réaction inflammatoire chronique. L'effet des nouvelles thérapeutiques de la PR (biothérapies) sur la cohabitation EBV-hôte nécessite d'être analysé.

### **2.2.2 L'entretien de l'auto-immunité (Cellules et Cytokines)**

Nous avons vu plus haut que l'inflammation rhumatoïde, quelque soit la cause, résulte d'une infiltration de la synoviale par des cellules inflammatoires (lymphocytes T, lymphocytes B, monocytes macrophages) qui viennent renforcer le potentiel inflammatoire des synoviocytes résidants : type A d'origine macrophagique et type B d'origine mésenchymateuse fibroblastique.

Les cellules présentatrices d'antigène (telles les cellules dendritiques) produisent des cytokines telles que l'IL18, l'IL12 et l'IL1 qui activent les lymphocytes T. Le TNF $\alpha$  joue également un rôle important de chef d'orchestre des cytokines proinflammatoires.

Les monocytes/macrophages produisent des facteurs d'angiogénèse tel que le VEGF (*Vascular Endothelial Growth Factor*) qui favorisera l'apport de cellules inflammatoires. La réponse T spécifique est de type Th1 (*T Helper 1*) avec production d'IL2 et d'IFN-gamma, et plus accessoirement de type Th2 avec production d'IL4, d'IL10 et d'IL13.

Le rôle des cellules « T-reg » dans la PR a été démontré à la fois chez les patients et dans les modèles expérimentaux. Globalement, la maladie s'accompagne d'un déficit fonctionnel de ces cellules alors que l'étude de leurs variations numériques aboutit à des résultats variables. Ils participent du nouveau paradigme de la PR qui fait jouer un rôle essentiel à la balance entre les lymphocytes Th17, effecteurs et pro-inflammatoire, et les lymphocytes T-reg (87).

Les Th17 produisant de l'IL17A seraient un acteur important de la réaction immunologique et de l'ostéolyse produite par les ostéoclastes. A un stade plus évolué de la réaction inflammatoire, les monocytes/macrophages produisent de grandes quantités d'IL1 et de TNF $\alpha$  dont l'activité ne peut être suffisamment freinée par leurs inhibiteurs naturels, IL1-Ra et récepteurs solubles de l'IL1 d'une part, récepteurs solubles du TNF $\alpha$  d'autre part.

Les fibroblastes produisent également de l'IL1 et du TNF $\alpha$ , ainsi que des chimiokines (chimiotactiques) telles que l'IL8, des métalloprotéases et de l'oxide nitrique NO, des radicaux libres, des prostaglandines PGE2. Le phénotype de ces fibroblastes synoviaux est proche de celui de cellules transformées avec expression de divers oncogènes (C-myc).

Les chondrocytes participent aux phénomènes de chondrolyse en produisant de l'IL1 et des métalloprotéases et en diminuant ses capacités de synthèse des protéoglycanes. Enfin, les monocytes, stimulés par le TNF $\alpha$  et diverses molécules des lymphocytes T tel que RANK-ligand, se différencient en pré-ostéoclastes puis en ostéoclastes qui résorbent l'os sous chondral (**figure 3**).

La réaction inflammatoire synoviale libère les PAD (peptidyl arginine déiminases) monocytaires et des polynucléaires qui vont « citrulliner » diverses protéines du tissu conjonctif (vimentine), du cartilage (collagène type II) et de protéines exsudées du plasma (fibrine). Dégradés par les cellules présentatrices d'antigène, les épitopes citrullinés seront présentés aux lymphocytes « Th » synoviaux par les molécules HLA de classe II portant l'épitope partagé (DRB1\*04 et DRB1\*01). Cette présentation déclenche la production d'anticorps ACPA par les lymphocytes B et les complexes immuns formés vont ainsi entretenir la réaction inflammatoire synoviale.

### **3 Les traitements**

La polyarthrite rhumatoïde nécessite une prise en charge pluridisciplinaire qui peut faire intervenir, outre le rhumatologue et le médecin généraliste, l'orthopédiste, le psychologue, l'assistante sociale, l'ergothérapeute, le kinésithérapeute, le médecin de rééducation fonctionnelle.

#### **3.1 Les traitements conventionnels**

Dans les lignes qui vont suivre, nous appellerons traitement conventionnel tout traitement ne faisant pas appel aux biothérapies, beaucoup plus récentes. Le traitement de la polyarthrite rhumatoïde a trois

objectifs principaux : soulager les douleurs, stopper l'évolution de la maladie afin de prévenir le handicap.

La première catégorie, pour soulager la douleur, est le traitement symptomatique. Il repose essentiellement sur les antidouleurs et les anti-inflammatoires, soit non-stéroïdiens « AINS » ou même stéroïdiens « corticostéroïdes ». La toxicité digestive, mais également rénale, de tous les AINS justifie une surveillance étroite sur le plan clinique et biologique (mesure de la tension artérielle, contrôle de la fonction rénale). Par contre, la corticothérapie n'est utilisée qu'en alternative ou en complément des autres traitements et en cas de polyarthrite rhumatoïde active invalidante, insuffisamment soulagée par les anti-inflammatoires.

Le deuxième objectif du traitement de la polyarthrite consiste à réduire l'évolution de la maladie. Il repose sur la réduction de la fréquence, de la durée, de l'intensité des poussées, et vise l'arrêt de la progression des érosions et des destructions radiologiques. Ce type de traitement repose essentiellement sur ce qu'on appelle communément les « *traitements de fond* » ou encore « DMARDs » (*disease modifying anti-rheumatic drugs*).

Actuellement, le traitement de fond le plus fréquemment prescrit lors d'une polyarthrite rhumatoïde débutante est le Méthotrexate (MTX), traitement de fond de première intention, puis le Léflunomide qui possède, comme le MTX, une action rapide avec une efficacité clinique et radiologique significative et un bon taux de maintien thérapeutique.

L'hydroxychloroquine est prescrit aux PR potentiellement bénignes ou lorsque qu'il existe un doute diagnostique entre une PR et un autre rhumatisme (en particulier un lupus). Certaines équipes utilisent la combinaison de plusieurs traitements de fond, en particulier pour les PR potentiellement sévères avant le recours aux biothérapies, que nous détaillerons un peu plus loin.

La réduction du handicap est le troisième volet de la stratégie thérapeutique et repose essentiellement sur une hygiène de vie adaptée et en dernier recours le traitement chirurgical. Ce dernier vise surtout à soulager les malades, cependant, à un stade précoce, la chirurgie permet par ailleurs de prévenir les destructions articulaires (synovectomie). Le traitement chirurgical peut faire appel à plusieurs techniques.

Ces techniques sont de trois types : *La synovectomie* arthroscopique ou chirurgicale est indiquée en cas de synovite persistante malgré le traitement médical général et local ; *L'arthroplastie* (le plus souvent totale) permet d'apporter l'indolence et de rendre la fonction à une articulation détruite, telle que la hanche, le genou ou l'épaule ; *L'arthrodèse* arthroscopique ou chirurgicale permet d'apporter l'indolence et la stabilité à une articulation détruite lorsqu'une arthroplastie est difficilement réalisable (poignet, cheville, arrière-pied).

### **3.2 Les Biothérapies**

L'identification des communications entre les cellules a fourni les cibles thérapeutiques des biothérapies utilisées désormais largement dans le traitement de la polyarthrite rhumatoïde. Ces biothérapies et les études de stratégies thérapeutiques ont permis des progrès considérables dans la gestion de la PR, permettant même d'espérer d'obtenir de meilleurs résultats que les autres procédés thérapeutiques.

Les anti-TNF $\alpha$  (infliximab, étanercept, adalimumab) sont proposés actuellement dans les polyarthrites sévères, réfractaires (échec des autres traitements de fond) mais également, pour certains, en première intention, en fonction des facteurs pronostiques (**figure 4**). D'autres traitements sont très prometteurs comme le Rituximab (anticorps anti-CD20) ou l'Abatacept (CTLA4-Ig). L'inhibition des lymphocytes B par le Rituximab ou par d'autres



anticorps monoclonaux humanisés paraît une voie extrêmement intéressante.

Le Rituximab a clairement démontré son efficacité notamment en association au Méthotrexate, sur l'activité clinique et structurale de la PR, en particulier dans les populations de patients insuffisamment répondeurs aux anti-TNF $\alpha$ . L'inhibition des voies de co-stimulation des LT par des cellules présentatrices d'antigènes est une autre voie particulièrement intéressante ayant conduit à la commercialisation de l'Abatacept. Ce dernier a également montré une efficacité clinique dans la PR, surtout en association au Méthotrexate. Cette efficacité a été démontrée actuellement à la fois chez les patients insuffisamment répondeurs aux anti-TNF $\alpha$  et ceux insuffisamment répondeurs au Méthotrexate. Le Tocilizumab est un anticorps anti-récepteur de l'interleukine 6 qui, lui aussi, a démontré dans des études de phases II et III, une efficacité clinique indiscutable et particulièrement intéressante (**figure 4**).

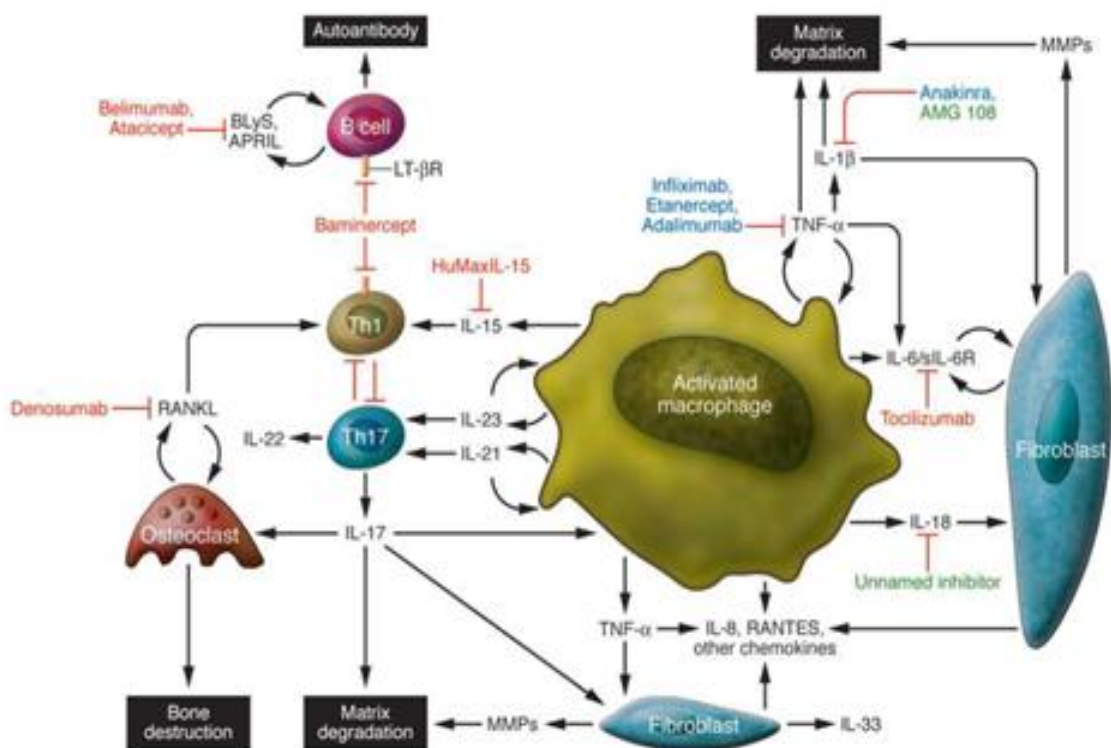


Figure 4 : Cytokines ciblées dans la polyarthrite rhumatoïde (88).

### **3.3 La pharmacogénétique (association de facteurs génétiques avec l'efficacité du traitement)**

La prise en charge de la PR est dans une phase de transition grâce à une meilleure compréhension conduisant à un grand nombre de nouvelles cibles. En parallèle, les progrès des technologies à haut débit sont maintenant transférés en clinique, produisant une très grande quantité de résultats sur l'hétérogénéité des gènes, des ARN transcrits et des protéines entre différents groupes de patients.

Dans ce contexte, la pharmacogénétique se concentre sur les associations d'expression d'un seul gène ou de plusieurs signatures génétiques avec la réponse aux médicaments. Plusieurs avancées majeures ont été réalisées dans ce domaine et pourtant, nous sommes encore très loin de trouver les biomarqueurs génétiques optimaux pour une vraie, ou du moins, une bonne prédiction des effets secondaires voire des réponses aux traitements selon un profil génétique donné (89).

Afin de développer de nouvelles thérapies, des initiatives sont axées sur le développement de nouveaux agents biologiques et de petites molécules ciblant les multiples voies de signalisation impliquées dans les différents processus physiopathologiques de la PR. Il sera nécessaire d'évaluer ces nouvelles molécules en combinaison avec les produits biologiques disponibles, ou chez les patients ayant des problèmes de tolérance. Avec plusieurs cibles moléculaires importantes (TNF $\alpha$ , IL-1, IL-6, IL-17 et beaucoup d'autres), la PR est maintenant considérée comme un syndrome clinique avec une pathogénie liée à différentes voies biologiques spécifiques et partagées.

La spécificité de chaque traitement peut dépendre du rôle de chaque cible dans les manifestations cliniques chez un patient donné à un stade particulier de sa maladie. Au cours des 20 dernières années, le nombre de médicaments enregistrés pour la PR a augmenté rapidement avec cinq médicaments anti-TNF $\alpha$  et d'autres ciblant d'autres cytokines et les cellules

T et B avec des outils spécifiques. Ces thérapies biologiques ont été une avancée majeure dans le contrôle des réponses cliniques et de la réduction, voire la prévention, des érosions osseuses (90).

Les premières analyses menées sur les études de SNPs ont été suivies par des analyses pharmacogénomiques plus détaillées sur plusieurs signatures de gènes. La plupart des études ont été menées sur le polymorphisme (-308 G>A), situé dans le promoteur du gène codant le TNF $\alpha$  et modulant sa transcription. Ce polymorphisme a été associé à une réponse clinique à l'infliximab, l'étanercept et l'adalimumab, et la plupart des études ont signalé une augmentation de l'efficacité du traitement en présence de l'allèle G (et surtout du génotype homozygote GG) de ce polymorphisme, le suggérant comme marqueur génétique prédictif de la réponse aux anti-TNF $\alpha$  (91, 92).

Les gènes qui ont été associées à des réponses à l'étanercept comprennent principalement des gènes de cytokines (TNF, IL10, TGF $\beta$ 1, IL1RN). En effet, quatre SNPs dans ces gènes ont été étudiés par rapport aux réponses à l'étanercept dans un groupe de 123 patients atteints de PR. Il en ressortait que certains SNPs des gènes TNF $\alpha$  et IL10 pourraient être associés à de meilleures réponses, tandis que la combinaison de IL1RN et TGF $\beta$ 1 (*transforming growth factor-beta 1*) était reliée à des réponses défavorables (93). D'autres polymorphismes ont été associés à l'évolution clinique au cours du traitement anti-TNF $\alpha$ , tels que le SNP A238G dans le gène du TNF $\alpha$ , le polymorphisme T196G dans le gène TNFRSF1B, la substitution Val158Phe dans le gène Fc $\gamma$ R ou encore un SNP dans le gène du récepteur de la tyrosine-protéine phosphatase C (PTPRC) (94, 95).

D'autre part, les réponses au Rituximab ont été liés à des SNPs dans le gène codant le Fc $\gamma$ RIIIA (changement d'acide aminé Val158Phe), à des SNPs du gène IL6 et à d'autres polymorphismes situés dans le gène codant le stimulateur des lymphocytes B (BLyS) (94-97).

Concernant les réponses aux thérapies habituelles de type DMARDs, plusieurs SNPs ont pu être associés à l'efficacité ou à la toxicité du méthotrexate (MTX). Le meilleur exemple est le groupe de gènes codant des protéines impliquées dans la voie du Folate, la synthèse des nucléotides, et la production de cytokines. Parmi les variantes de la voie associée au folate, les SNPs les plus étudiés sont C677T et la mutation A1298C du gène de la méthylène tétrahydrofolate réductase (MTHFR), le changement d'acide aminé G80A dans le gène RFC-1 et la mutation C1420T dans gène de la sérine hydroxyméthyltransférase (SHMT) (95, 98-100).

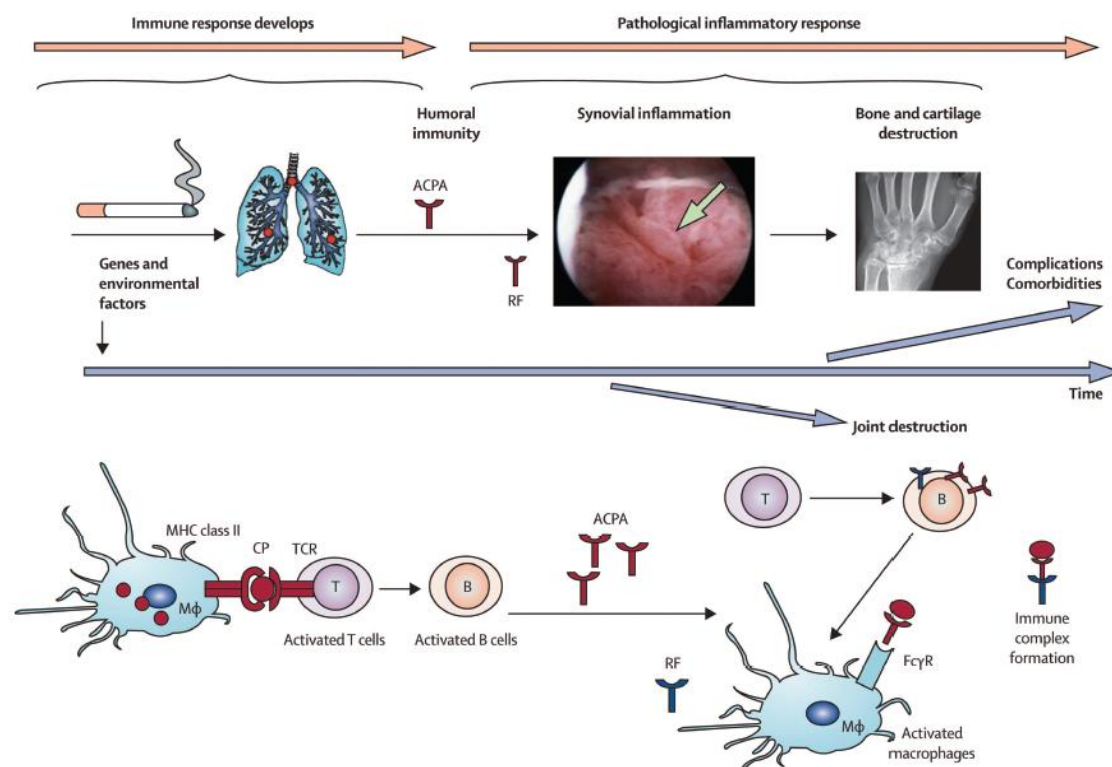
Ainsi, les études pharmacogénétiques peuvent nous rapprocher de la médecine personnalisée. Cependant, des études récentes ont donné des résultats non concluants, rapportant des caractéristiques génétiques sensiblement différentes dans différentes cohortes. En outre, bien que de nombreux gènes aient été mis en évidence pour prédire des réponses thérapeutiques, le rôle d'un grand nombre de ces gènes dans la pathogenèse de la polyarthrite rhumatoïde reste inconnu. D'autres études, éventuellement en utilisant des ensembles de gènes présélectionnés et bien caractérisés, sont nécessaires pour déterminer la valeur de la pharmacogénétique dans la PR (89, 95)

## **4 Les facteurs de risque**

### **4.1 Facteurs environnementaux**

Les principales études visant à l'identification des facteurs de risque de la PR se sont focalisées sur le régime alimentaire, le tabac et les hormones (101). À ce jour, peu de facteurs environnementaux ont été identifiés comme favorisant le développement de la PR. Le principal facteur de risque environnemental de PR est le tabac. Si cette notion est relativement ancienne (102), de récentes études ont bien souligné que ce

risque ne concernerait que la forme de PR exprimant des anticorps anti-peptides citrullinés (ACPA) (103-105) (**figure 5**).



**Figure 5. Hypothèse physiopathologique de la PR ACPA+ (105) :** Plusieurs facteurs interviennent dans le déclenchement de la PR : une prédisposition génétique (HLA-DRB1 et PTPN22 par exemple), des facteurs environnementaux qui favoriseraient la citrullination des peptides (tabac, infections) et des facteurs intrinsèques (hormonaux par exemple). Lorsque ces facteurs sont réunis, il s'en suit une double réponse immunitaire (innée et acquise) dirigée tout particulièrement contre la membrane synoviale. La synoviale rhumatoïde est alors infiltrée par de nombreuses cellules immunitaires : macrophages, lymphocytes TCD4+, lymphocytes B. De façon schématisée la pathogénie de la PR se déroule en trois phases : une phase de déclenchement rendant compte de la réponse immunitaire dirigée contre la synoviale, une phase d'inflammation de la membrane synoviale et la phase de destruction ostéoarticulaire.

Le risque de développer une PR est proportionnel à la durée et la quantité de tabac consommée (106). Ainsi, une étude prospective, réalisée sur une grande cohorte de patients (*Nurse's Health Study*), a montré une corrélation linéaire entre l'importance de la consommation de tabac et le risque de développer une PR : les fumeurs ayant une consommation supérieure à 40 paquets-années avaient un risque deux fois supérieur

comparé aux individus non fumeurs. Il est intéressant de noter que ce risque perdurait jusqu'à 20 ans après le sevrage tabagique (107).

Plusieurs études ont montré l'existence d'une interaction gène–environnement entre le tabac et l'épitope partagé ou « *shared epitope* » (*HLA-DRB1\*SE*), illustrant ainsi les mécanismes attendus d'une pathologie génétique complexe. Une étude de population cas/contrôle incluant des individus atteints de PR récente de moins d'un an d'évolution (cohorte EIRA ou *Epidemiological Investigation of RA*), a montré que les individus fumeurs ne portant pas d'allèle du *HLADRB1\*SE* (*SE* négatif) avaient 1,5 fois plus de risque de développer une PR-ACPA-positif que les individus non fumeurs et *SE* négatifs (108). Le risque de développer une PR-ACPA-positif pour les individus fumeurs portant deux copies du *SE* étaient 21 fois supérieur à celui des non fumeurs *SE* négatif (108). Cette interaction a depuis été observée dans de nombreuses cohortes indépendantes (109).

La consommation d'alcool semble avoir une action protectrice sur le développement de la PR. Si les mécanismes physiopathogéniques qui sous-tendent cette observation sont encore inconnus, une étude réalisée dans les populations suédoise (cohorte EIRA) et danoise (cohorte CACORA) ont bien montré qu'une consommation supérieure à 80mg d'éthanol/jour (environ cinq verres) était associée à une diminution de 40 à 50% du risque de PR comparée à une absence de consommation d'alcool. Encore une fois, dans le sous-groupe PR-ACPA-positif, le rôle protecteur de l'alcool était beaucoup plus prononcé chez les patients *SE* positif, soulignant une probable interaction alcool – *HLADRB1\*SE* (110).

Beaucoup d'autres facteurs environnementaux tels que la vitamine D, la contraception orale, le régime alimentaire riche en viande rouge ont été suspectés comme pouvant influencer la risque de développer une PR sans toutefois pouvoir emporter la conviction, les résultats de la plupart de ces études étant contradictoires (109). On signalera qu'une récente étude est venue confirmer les résultats d'une précédente sur le rôle protecteur de l'allaitement prolongé (supérieur à 12 mois) (111).

## **4.2 Facteurs individuels**

Un des plus importants facteurs individuels reste le facteur endocrinien ou hormonal. En effet, certaines observations cliniques permettent de soupçonner l'intervention de facteurs endocriniens au cours de la PR. Cette maladie à nette prédominance féminine, survient fréquemment en période péri-ménopausique. Une rémission est fréquente pendant la grossesse, une poussée très habituelle au décours de l'accouchement ou lors de l'allaitement (111) et les contraceptifs oestroprogestatifs diminuent sa sévérité sans en réduire véritablement l'incidence.

D'autre part, il a été suggéré aussi que des facteurs psychologiques, notamment le stress, pouvaient être importants dans l'apparition mais aussi dans le développement et le maintien de la pathologie. Toutefois, il n'existe pas de terrain psychologique particulier favorisant la survenue de la maladie. La pathologie elle-même, ou une poussée aiguë, peut être induite par un traumatisme affectif ou moins souvent par un traumatisme physique. Il n'est pas rare de voir apparaître la PR dans les semaines ou les mois qui suivent un événement de vie traumatisant. Il est possible que, dans ces cas, des médiateurs solubles sécrétés par le système nerveux central agissent sur l'axe hypothalamo-hypophyso-surrénalien et déclenchent des perturbations immunitaires, peut-être par l'intermédiaire de neuropeptides.

Enfin, un indice de masse corporelle élevé constituerait aussi un facteur de risque individuel de la PR (103).

En plus de tous ces facteurs, il reste le plus important facteur pour nous dans le cadre de cette étude, c'est le facteur génétique ou plus précisément la prédisposition génétique. Il reste de loin, l'élément le plus étudié parmi les facteurs de risque de la PR.

## 5 La composante génétique de la PR

Un point commun aux maladies auto-immunes est leur propension à apparaître dans des familles. L'intérêt pour la composante génétique de la PR n'est pas nouveau, puisqu'il a été d'actualité depuis 1958 (112). L'impact de la prédisposition génétique a été identifié par l'analyse des taux de concordances entre jumeaux monozygotes et dizygotes (113). Cette composante génétique est également révélée par le risque relatif entre apparentés du premier degré (frères-soeurs) de développer la maladie, comparé à sa prévalence dans la population (valeur de  $\lambda_s$ ). Les taux de concordance varient de 15% pour la polyarthrite rhumatoïde à 57% pour le lupus, avec des  $\lambda_s$  respectifs de 5-10 à 20-40 (114).

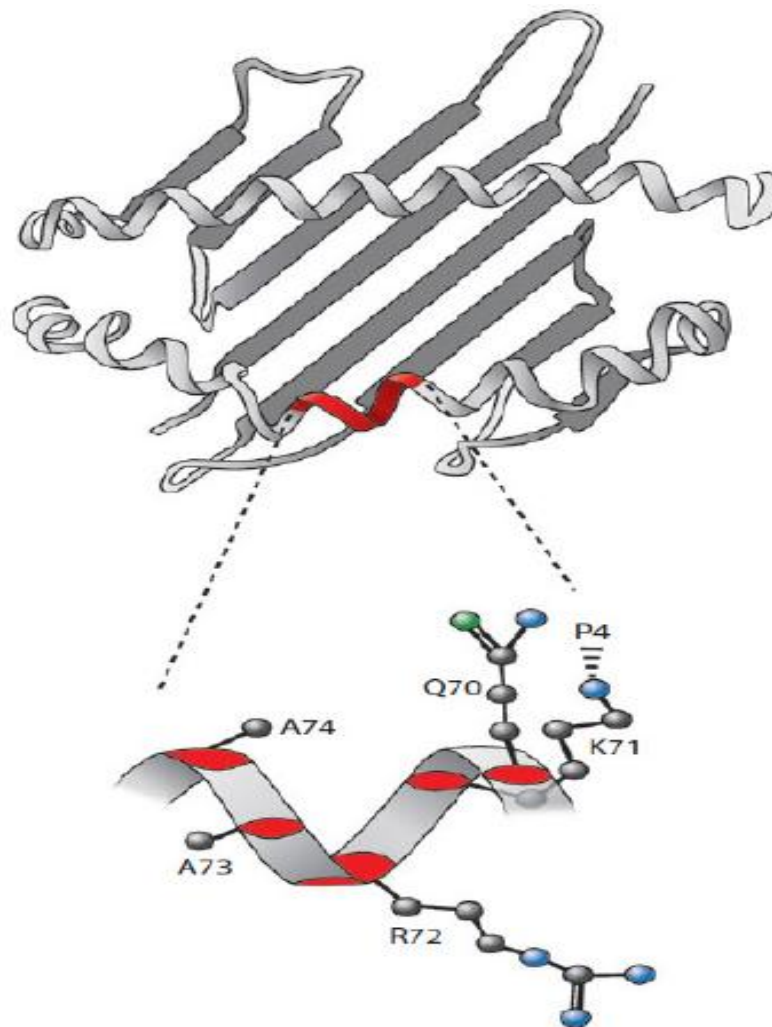
La recherche des gènes de la PR est en plein essor. Clairement, la maladie est polygénique, en plus d'être multifactorielle, mais il faut voir la polyarthrite rhumatoïde *-stricto sensu-* comme une affection dont l'éclosion demande à la fois la présence d'un terrain particulier, dans la composition duquel les gènes HLA de classe II en combinaison avec d'autres gènes à définir jouent probablement un rôle, et une réponse à un agent extérieur, lui aussi mal défini.

### 5.1 L'historique

Le premier gène confirmé, et de loin, le plus impliqué dans le déterminisme génétique de la PR est le gène HLA-DRB1, codant pour la chaîne  $\beta$  de la molécule HLA-DR et situé sur le chromosome 6 en 6p21. Les allèles de prédisposition HLA-DR représentent jusqu'à 50 % des facteurs génétiques (63, 113). L'association à la PR d'allèles de gènes localisés dans la région HLA a été proposée pour la première fois en 1976 par Stastny (115). Les allèles du gène codant pour les spécificités sérologiques DR4 et DR1 ont été identifiés plus tard comme étant associés à la PR (116, 117).



A la fin des années 1990, grâce au séquençage du locus HLA-DRB1, Gregersen et ses collaborateurs proposent l'hypothèse de l'épitope partagé (*Shared Epitope*) pour expliquer l'association de la région codant les molécules de classe II du CMH (complexe majeur d'histocompatibilité) avec la PR (**figure 6**).



**Figure 6 : Représentation schématique de l'épitope partagé codé par l'allèle HLA-DRB1\*0401 (118).**

Selon l'hypothèse formulée, les molécules HLA-DR seraient directement impliquées dans la physiopathologie de la PR et l'association entre HLA-DR et PR serait attribuable à des allèles de susceptibilité codant une séquence en acides aminés conservée, située dans la 3<sup>ème</sup> région

hypervariable de la chaîne  $\beta$  des molécules HLA-DR (119). Ceci serait dû à la présence de cette séquence « épitope partagé » (*Shared Epitope*) (QKRAA, QRRAA ou RRRAA) en position 70-74, codée par les allèles DRB1\*0101, \*0102, \*0401, \*0404 et \*0405 et \*1010, qui seraient associés à une plus grande susceptibilité à développer une PR (**figure 6**).

Plus récemment, une étude a affiné les signaux d'association pour la polyarthrite rhumatoïde dans la région HLA à cinq positions d'acides aminés codées par trois gènes HLA (trois positions d'acides aminés (11, 71 et 74) sur HLA-DR $\beta$ 1, un seul sur HLA-B (en position 9) et un autre sur HLA-DP $\beta$ 1 (en position 9), le tout étant situé dans des poches de présentation de peptides antigéniques (120). En adaptant les méthodes statistiques à partir d'études d'association pangénomique (GWAS) et l'utilisation de l'imputation d'un large panel de référence, cette étude a démontré le potentiel de cette approche pour identifier les variants fonctionnels dans les régions associées (53, 120).

Dans la population Algérienne, une première étude cas-témoin a été menée au sein du Laboratoire de recherche en Génétique Moléculaire et Cellulaire de l'Université USTO d'Oran (LGMC-USTO) sur un échantillon de l'Ouest Algérien. Cette étude a révélé l'existence d'une forte association entre l'allèle HLA-DRB1\*0405 et la susceptibilité au développement de la PR. De plus l'allèle HLA-DRB1\*0701 a été associé pour la première fois à un effet protecteur dans cette pathologie (121).

La présence d'un des allèles du gène HLA-DRB1 conférant la susceptibilité à la PR est décrite dans 40 % des individus dans la population générale et dans 70 % des patients atteints de la maladie (122). Ces données suggèrent donc que d'autres gènes non-HLA puissent également jouer un rôle dans la susceptibilité à la PR (123).

## 5.2 Les études génétiques dans la PR

Les études de liaison et/ou d'association génétique ont permis d'identifier de très nombreux gènes clés impliqués dans la PR et globalement dans les maladies génétiques, par l'analyse de la coségrégation de marqueurs génétiques et de la maladie dans les familles à cas multiples ; le meilleur exemple étant les gènes du système HLA. Depuis quelques années, grâce à une évolution des technologies moléculaires et des capacités de séquençage, nous assistons à un très fort développement des études d'association dites pangénomiques ou « *genome-wide association study* », le tout ayant conduit à l'identification des principaux facteurs de prédisposition génétique à la PR (**figure 7**). Dans le paragraphe suivant, nous allons en énumérer les principaux résultats.

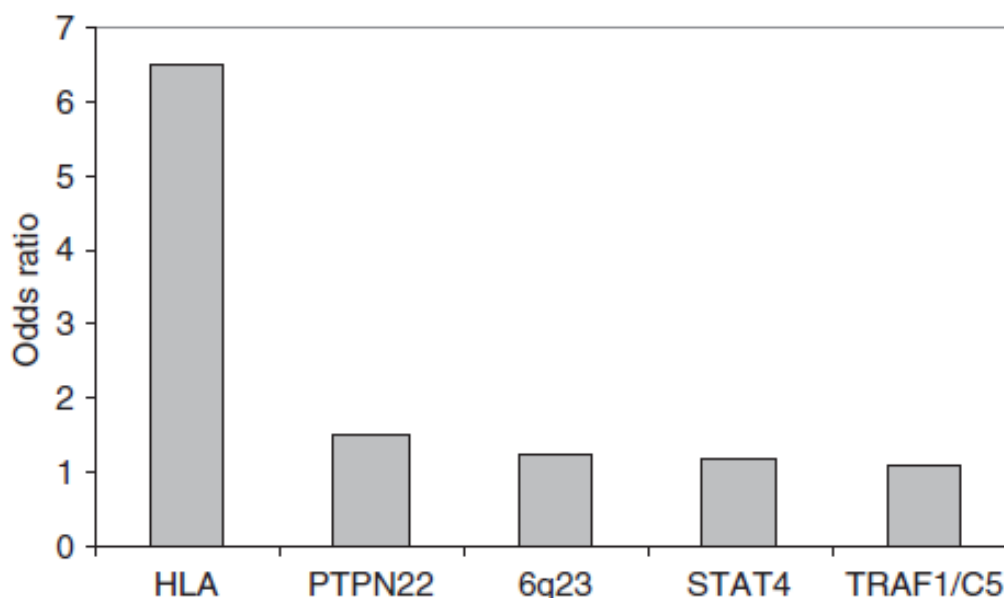


Figure 7 : Contribution des principaux facteurs génétiques à la susceptibilité à la PR (124).

### 5.2.1 Principales GWAS et autres Méta-analyses

Comme décrit précédemment, les GWAS correspondent à des études d'association cas-témoins, testant en une fois plusieurs centaines de milliers d'hypothèses sans nécessiter de pertinence biologique/physiopathologique.

Les études d'association GWAS sont intéressantes pour trouver des SNP candidats associés à des mécanismes physiopathologiques. Si cette méthode s'est avérée très efficace, elle pose le problème de l'identification d'un très grand nombre d'allèles de susceptibilité, pas toujours répliqués, dont le rôle fonctionnel n'est pas connu, et parfois même situés dans des régions du génome dont la fonction demeure inconnue (125). Les études génétiques de type GWAS et l'approche gènes candidats ont permis d'identifier à ce jour plusieurs dizaines de loci de susceptibilité impliqués dans la PR et de confirmer l'existence de facteurs génétiques communs entre la PR et d'autres maladies auto-immunes (MAI) (126).

Une des premières GWAS a été menée par le consortium anglais WTCCC (*The Wellcome Trust Case Control Consortium*), portant sur 1860 patients PR et 2938 contrôles. Cette GWAS a confirmé l'association des deux premiers loci de prédisposition HLA-DRB1 et PTPN22 (50). Neuf autres loci ont montré des résultats significatifs d'association avec la PR dans cette GWAS, dont le locus 6q23. L'association de ce dernier a été répliquée sans équivoque dans deux études au Royaume-Uni et aux États-Unis (127, 128).

Une autre GWAS, tout aussi importante que la précédente a été menée par Plenge et al. et a abouti à décrire l'association du locus TRAF1/C5 comme étant le troisième marqueur génétique confirmé dans la PR. En effet, l'étude a porté sur le génotypage de 317 503 SNPs dans une analyse combinée de type Cas-Contrôles regroupant 1522 cas PR ayant un profil ACPA positif (anticorps anti-peptides citrullinés) et 1850 sujets contrôles (129). Les échantillons ont été obtenus de deux bases de données : la première américaine (*the North American Rheumatoid Arthritis Consortium* ou « NARAC ») et la deuxième suédoise (*the Swedish Epidemiological Investigation of Rheumatoid Arthritis* ou « EIRA »).

Les résultats obtenus ont identifié un SNP sur le chromosome 9, rs3761847, qui a montré une association avec la polyarthrite rhumatoïde pour tous les échantillons testés (OR = 1,32 ; interval de confiance à 95% CI

]1,23 - 1,42[ ;  $P_{\text{Value}} = 4 \times 10^{-14}$ ) (129). Par ailleurs, comme dans l'étude GWAS du WTCCC (50), les associations les plus fortes ont été identifiées avec les gènes HLA-DRB1 et PTPN22.

Deux ans après la GWAS de Plenge et al, Gregersen et ses collaborateurs ont effectué une méta-analyse sur 2418 cas et 4504 témoins issus d'une population de l'Amérique du Nord suivie d'une étude de réplication sur 2604 cas et 2882 témoins (130). Ces études ont permis de mettre en évidence l'association de PTPN22 ( $p=1.62 \times 10^{-21}$ ) et la région TRAF1-C5 ( $p=4.09 \times 10^{-8}$ ) avec la PR, mais aussi de suggérer l'association d'un nouveau gène REL ( $p=6.01 \times 10^{-10}$ ) ainsi que les gènes CTLA4 ( $p=6.25 \times 10^{-9}$ ) et BLK ( $p=5.69 \times 10^{-9}$ ).

Kochi et al. (131) ont réussi, grâce à une méta-analyse établie sur 2303 cas et 3380 témoins et une étude de réplication sur 7069 cas et 20727 témoins dans la population japonaise, à confirmer l'association de HLA-DRB1 ( $p=1,9 \times 10^{-58}$ ) et de valider l'association du gène CCR6 avec la PR.

Stahl et al. (51) ont réalisé une méta-analyse GWAS sur 5839 patients positifs pour les auto-anticorps caractéristiques de la PR et 20169 témoins d'ancêtres européens. En raison du nombre élevé de comparaisons, une association a été considérée comme significative quand la valeur de  $p$  était inférieure à  $5 \times 10^{-8}$ . Les résultats obtenus confirment l'association de HLA<sub>rs6910071</sub>DRB1<sub>rs</sub> ( $p_{\text{GWAS}} = < 10^{-299}$ ), PTPN22<sub>rs2476601</sub> ( $p_{\text{GWAS}} = 9.1 \times 10^{-74}$ ), CTLA4<sub>rs3087243</sub> ( $p_{\text{GWAS}} = 1.2 \times 10^{-8}$ ), TNFAIP3<sub>rs6920220</sub> ( $p_{\text{GWAS}} = 8.9 \times 10^{-13}$ ) et CD40<sub>rs4810485</sub> ( $p_{\text{GWAS}} = 2.8 \times 10^{-9}$ ) avec la PR. La réplication des résultats a été établie sur 6768 patients et 8806 témoins et a permis d'identifier 7 nouveaux gènes de susceptibilité dont IL6ST, SPRED2, RBPJ, CCR6, IRF5, PDK et AFF3. Ces SNP sont situés dans ou près de gènes dont la fonction est liée d'une manière ou d'une autre à la réponse immunitaire.

Lee et al. (132) ont utilisé les données d'une GWAS faite par Freudenberg et al. (133) sur 801 cas et 757 témoins coréens, une population

restée très peu étudiée. Les résultats obtenus impliquent plus particulièrement quatre gènes candidats PADI4, MTR, PADI2, et TPH2 dans la susceptibilité à la PR.

Très récemment, Okada et al. ont réalisé une méta-analyse de GWAS totalisant plus de 100000 sujets d'ascendances européenne et asiatique (29880 cas de PR et 73758 contrôles) en analysant environ 10 millions de SNP (134). Les auteurs ont découvert 42 nouveaux loci à risque pour la PR, ce qui porte le nombre total de loci confirmés actuellement pour cette maladie à 101. Les 100 loci hors région HLA expliqueraient 5,5% et 4,7% de l'héritabilité chez les Européens et les Asiatiques, respectivement. Les auteurs ont par ailleurs démontré que les gènes identifiés sont des cibles de choix de thérapies approuvées et que les médicaments approuvés pour d'autres indications peuvent être utilisés également pour le traitement de la PR.

Ce travail complète les informations sur les gènes fondamentaux dans la PR, ainsi que sur les voies et les types de cellules qui contribuent à la pathogenèse de cette maladie.

### **5.2.2 Facteurs de prédisposition génétique à la PR retenus pour notre étude**

De manière générale, le risque relatif de développer la maladie lié aux gènes non-HLA reste faible, relativement à la région HLA ; bien qu'ils aient été confirmés dans des études indépendantes, leur association avec la PR n'est pas universelle (135) mais « ethnies-spécifiques ». Ceci n'empêche pas le fait qu'il existe actuellement une liste de gènes dits « confirmés » de par leur association à la PR, mais dont les mécanismes, voire les secrets de leur intervention en physiopathologie reste à éclaircir. A ce jour, seuls quelques gènes associés à la PR ont un variant fonctionnel dont le risque est bien

investigué par la conjonction d'études cas-témoins répliquées indépendamment et d'études de liaison génétique (136) (**tableau 3**).

**Tableau 3 : Principaux facteurs de susceptibilité génétique impliqués dans le déterminisme de la polyarthrite rhumatoïde**

Gène	Localisation	Fonction	Sous-type de PR
<b>Présentation de l'antigène</b>			
HLA-DRB1*	6p21	Présentation de l'antigène	
SE :			ACPA+
- 0401, 0402			
- 0404, 0405, 0408			
- 1001, 1402			
- 0101, 0102			
- 0301			ACPA-
<b>Transduction du signal et signalisation intra-cellulaire</b>			
PTPN22	1q13-3	Signalisation TCR et BCR	ACPA+
TNFAIP3/OLIG3	6q23	Signalisation TNFR/voie NF $\kappa$ B	ACPA+
TRAF1	9q33	Signalisation TNFR/voie NF $\kappa$ B	ACPA+
<b>Facteurs de transcription</b>			
IRF5	7q32	Voie INF type I	ACPA+
STAT4	2q32-2	Voie INF $\gamma$	ACPA+
c-Rel	2q13	Voie NF $\kappa$ B expression de CD40	
<b>Voies de co-stimulation et récepteurs membranaires</b>			
CTLA4	2q33	Co-stimulation T	
CD40	20q12	Co-stimulation T/ Voie NF $\kappa$ B	
<b>Enzymes de citrullination</b>			
PADI-4	1p36	Citrullination (peptidyl arginine déiminase)	PR asiatiques

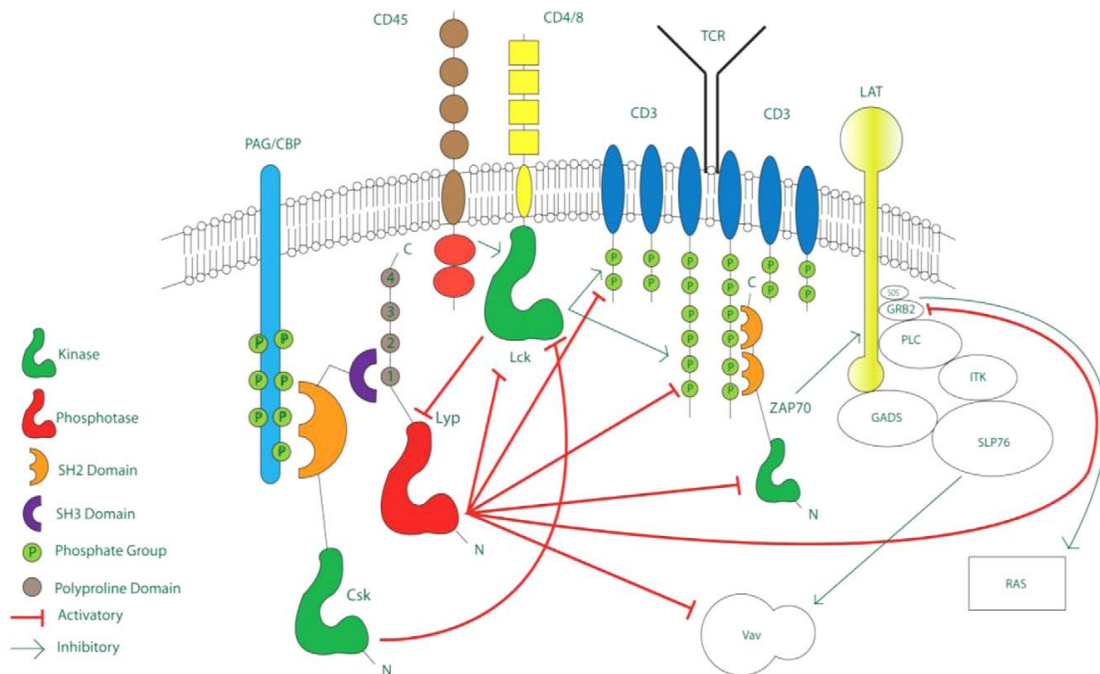
ACPA : anticorps anti-peptides citrullinés, PR : polyarthrite rhumatoïde

### 5.2.2.1 Gènes candidats associés au développement de la PR - Etude dans la population Algérienne

- **PTPN22**

Le gène PTPN22 (*protein tyrosine phosphatase non-receptor 22*) est localisé sur le chromosome 1p13 et code une phosphatase lymphoïde-spécifique (LYP), qui est une protéine intracellulaire inhibant la voie de signalisation du récepteur des lymphocytes T par une interaction physique sur un motif riche en proline du domaine SH3 de la kinase Csk (**figure 8**).

En raison de son implication dans la régulation des LT qui sont des acteurs majeurs de l'inflammation synoviale (136), le gène PTPN22, et plus précisément le polymorphisme rs2476601 (substitution C→T en position 1858), est le deuxième facteur confirmé de susceptibilité à la PR (137).



**Figure 8 : Les multiples actions de la protéine LYP sur la signalisation cellulaire spécifique du lymphocyte T (138).**

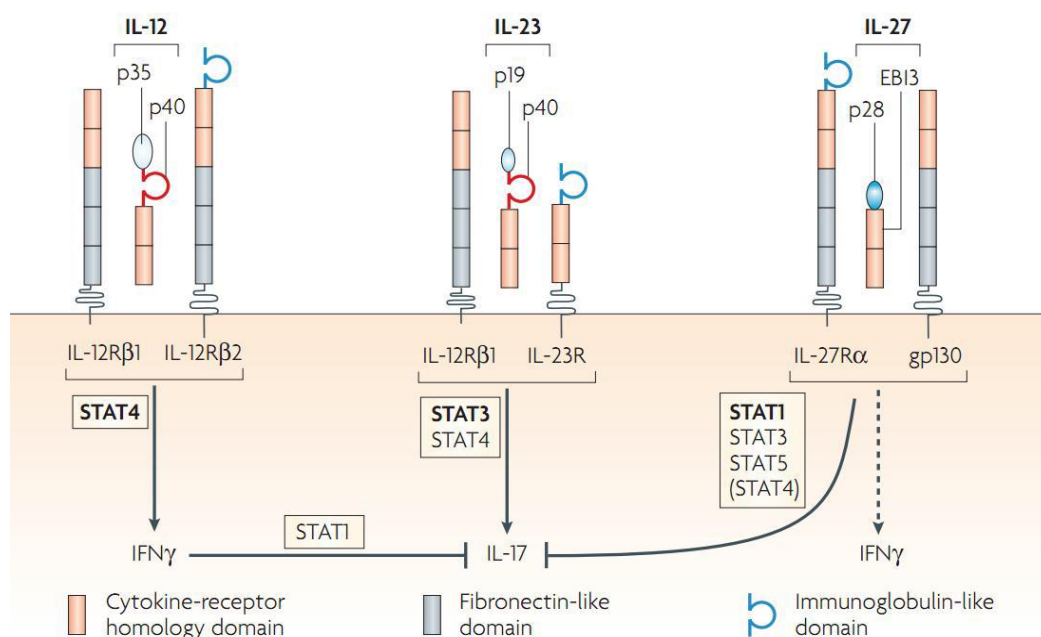
Ce polymorphisme fonctionnel de PTPN22 est responsable d'un changement d'Arginine en Tryptophane dans la protéine LYP en position 620 (R620W). Le variant à risque du polymorphisme C1858T de ce gène est caractérisé dans de nombreuses populations principalement caucasiennes (139), mais pas dans les populations asiatiques (140). La première association de ce gène avec une MAI a été rapportée au cours de l'année 2004 entre l'allèle T du variant C1858T et le diabète de type 1 (141). Ce polymorphisme du gène PTPN22 est associé à d'autres MAI en plus du diabète de type 1 et de la PR, comme le lupus érythémateux systémique, l'arthrite juvénile, la maladie de Basedow, la maladie d'Addison et le vitiligo. Toutefois, il n'y a pas d'association significative avec les maladies inflammatoires intestinales, la maladie cœliaque, ou encore la sclérose en



plaques (142, 143). Il apparaît maintenant que ce gène est le facteur génétique, en dehors de la région HLA, le plus important dans les MAI humaines (138).

### • STAT4

STAT4 (*signal transducer and activator of transcription 4*) localisé en 2q32 code une protéine impliquée dans la transmission du signal intracellulaire et dans la signalisation de récepteurs aux cytokines (**figure 9**).



**Figure 9: Intervention de la famille des molécules STAT dans la réponse immunitaire (144).**

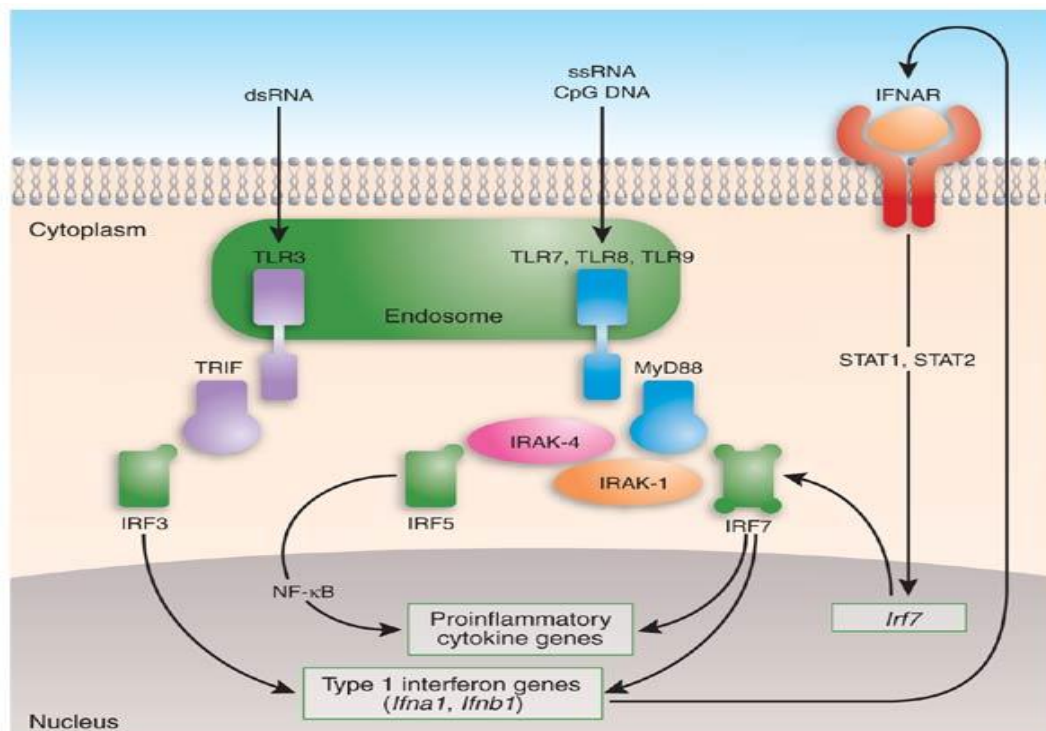
En effet, STAT4 est un facteur de transcription activé par l'IL-12, qui conduit à la sécrétion d'interféron de type II (IFN-γ), cytokine très importante dans la régulation des processus de réponse immunitaire de type TH1.

Le polymorphisme rs7574865 situé dans une région intronique est associé à la PR dans les populations asiatique et européenne (133). Plusieurs études se sont intéressées à cette association et la dernière ayant

conforté ces résultats est une très récente méta-analyse de 17 études cas-témoins regroupant 25 300 patients atteints de PR et 26 326 témoins sains (145). Cependant ce polymorphisme n'est pas spécifique de la PR puisqu'il est retrouvé aussi associé à plusieurs autres maladies auto-immunes comme le lupus érythémateux disséminé, la sclérodermie systémique, le syndrome de Sjögren primaire ou encore la maladie d'Addison (146-149) .

- **IRF5**

Ce gène situé en 7q32 code un membre de la famille IRF (Facteurs de Régulation des Interférons « IFNs »), un groupe de facteurs de transcription ayant une multitude d'actions cellulaires y compris l'activation de l'interféron, induite par virus, la modulation de la croissance cellulaire, la différenciation, l'apoptose et la modulation de l'activité du système immunitaire (**figure 10**).



**Figure 10 : Coopération entre les voies de signalisation médiées par les TLRs et intervention d'IRF5 (150).**

Les premiers travaux de Barnes *et al* s'intéressant au facteur IRF5 ont démontré sa participation dans la régulation de l'activité transcriptionnelle des gènes codant les IFNs de type 1 (151, 152). IRF5 est essentiel pour la production d'IFN de type I et de cytokines inflammatoires en aval des récepteurs *Toll-like* (TLRs), qui jouent, avec les IFNs de type I et les cytokines inflammatoires, un rôle primordial dans les maladies auto-immunes tel que le LED (Lupus Erythémateux Disséminé) (153, 154), mais aussi et surtout dans la PR.

En effet, Sigurdsson et son équipe (155) ont analysé 5 SNPs du gène IRF5 chez un échantillon de patients atteints de PR dans la population Suédoise, et ont trouvé que 4 sur les 5 SNPs étudiés étaient associés avec la PR, y compris le polymorphisme rs2004640. La plus forte association a été remarquée pour le polymorphisme rs3807306 ( $p = 0,00063$ ), particulièrement chez un sous-groupe de patients ACPA-négatifs ( $p = 0,000091$ ). Il est à noter aussi que ce dernier polymorphisme est en déséquilibre de liaison avec rs2004640 ( $r^2 = 0,67$ ).

L'association du gène IRF5 avec la PR a été répliquée par la même équipe dans une cohorte allemande mais aussi par d'autres études notamment de type GWAS (51). Plus récemment, certains polymorphismes d'IRF5 ont été associés à une évolution non destructrice de la PR (156).

- **Locus 6q23 (TNF-AIP3 / OLIG3)**

Une étude GWAS du *Wellcome Trust Case Control Consortium* (WTCCC 2007) a identifié neuf nouveaux SNPs comme étant associés à la polyarthrite rhumatoïde (50). L'association d'un SNP (rs6920220) a été répliquée dans une étude de validation par Thomson et al. où il a été démontré que ce SNP est cartographié en 6q23, entre les gènes OLIG3 (*oligodendrocyte lineage transcription factor 3*) et TNFAIP3 (*TNF- $\alpha$  induced protein 3*) (127).

Jusqu'à présent, toutes les études publiées concernant le premier gène, OLIG3, indiquent un rôle dans le développement des cellules neuronales, sans aucune intervention spécifique de la fonction immunitaire (157, 158). Le second gène, TNFAIP3, semble être le plus intéressant des deux gènes puisqu'il a un rôle direct dans la réponse immune. En effet, il est impliqué dans la signalisation du récepteur au TNF- $\alpha$  et inhibe la stimulation du facteur de transcription NF $\kappa$ B par le TNF- $\alpha$ .

Plenge et al. ont identifié un autre SNP en 6q23 (rs10499194 localisé à environ 150 kb de TNFAIP3 et OLIG3) qui a été associé de manière reproductible à la polyarthrite rhumatoïde. Ce polymorphisme est situé à environ 3,8 kb du premier SNP (rs6920220) identifié par Thomson et al. (127). Il a par ailleurs été démontré que les polymorphismes rs10499194 et rs6920220 situés en 6q23 sont corrélés avec une PR séropositive pour les anti-CCP (127, 129), et que le locus TNFAIP3 est impliqué dans plusieurs maladies auto-immunes (159, 160).

- **Locus TRAF1- C5**

TRAF1 (*TNF receptor-associated factor 1*) et C5 (*complement component 5*) sont situés sur le même locus TRAF1-C5 en 9q33 (129). Les gènes TRAF1 et TRAF2 forment un hétérodimère intervenant dans la voie de signalisation des MAP-kinases tandis que C5 est un élément de la cascade du complément. Ces deux gènes semblent être tous les deux des acteurs importants dans le contexte immunitaire de la PR de par leurs fonctions, mais les études d'association entre ce locus et la PR n'ont mis en évidence que des polymorphismes intergéniques, n'ayant apparemment aucune incidence fonctionnelle, comme c'est le cas pour les SNPs du locus 6q23.

Plusieurs études ont montré une association entre les SNPs rs3761847 et rs10818488 du locus TRAF1-C5 et la susceptibilité à la PR (161, 162), ainsi que pour d'autres maladies auto-immunes (163).

- **AUTRES GENES**

Outre les gènes précédemment décrits qui sont inclus dans notre étude, il existe un nombre important d'autres gènes qui ont été impliqués dans la prédisposition à la PR. Nous pouvons en citer dans la partie suivante quelques-uns des plus intéressants, et qui pourraient être envisagés dans la continuité de la première partie du présent travail pour une meilleure caractérisation génétique de la PR dans la population Algérienne.

TNFRII (ou TNFRSF1B) est situé en 1p36, une région génomique d'intérêt pour la PR, identifiée en 1998 par analyse de liaison (164). Ce gène code une protéine de 75kDa de 461 acides aminés (165). Ce récepteur est retrouvé à la surface d'un vaste panel de cellules dont les monocytes, les cellules endothéliales, les cellules de Langerhans et les macrophages (166). Il est capable de recruter les protéines adaptatrices TRAF1 et TRAF2 qui mènent à l'activation des voies NFκB et JNK (*Jun N-terminal Kinase*) (167).

Plusieurs polymorphismes ont été décrits au niveau de ce gène. Parmi eux, les polymorphismes A1663G (rs1061624), T1668G (rs5030792) et C1690T (rs3397) sont localisés au niveau de l'exon 10 (168), et un autre polymorphisme T676G (rs1061622) localisé au niveau de l'exon 6. Seuls les derniers polymorphismes C1690T et T676G ont été associés avec la PR (169, 170).

Le gène CTLA4 (*Cytotoxic Lymphocyte associated 4 protein*), localisé en 2q33, code une protéine co-stimulatrice exprimée à la surface des lymphocytes T. Cette dernière joue un rôle de régulation de la réponse immune en inhibant l'activation des LT suite à l'interaction entre le CTLA-4 et le CD86. Plusieurs études de cohortes ainsi que des méta-analyses ont montré une association entre le SNP rs3087243 du gène CTLA4 et la susceptibilité à la PR notamment chez des patients ACPA<sup>+</sup> (171, 172).

Le gène CD40, situé en 20q13, code une protéine de la superfamille du TNFR qui est impliquée dans la co-stimulation cellulaire. CD40 est

exprimé à la surface des CPA et des lymphocytes B qui se lie au CD40 ligand des LT. Il existe une association significative entre le polymorphisme rs4810485 du gène et les patients ACPA+ atteints de PR (173).

Le gène PADI4 (*Peptidyl arginine deiminase type IV*) est localisé au sein du même locus d'intérêt que TNFR2 situé en position 1p36. Ce locus contient le cluster comprenant les gènes codant les enzymes PAD qui sont impliqués dans la citrullination des résidus arginine. C'est un facteur de susceptibilité restreint aux populations asiatiques (174, 175). Le polymorphisme rs2240340 du gène localisé dans une région intronique n'a pas été confirmé dans la population caucasienne (176).

#### **5.2.2.2 Gènes candidats avec un différentiel d'expression dans la PR - Etude dans la population Française**

L'étude des polymorphismes des gènes présentant un différentiel d'expression dans la PR représente un complément d'information pour la compréhension du fonctionnement de ces gènes et de leur implication dans la maladie.

Dans le chapitre suivant, nous avons fait la synthèse des données bibliographiques concernant les gènes différentiellement exprimés de manière significative lors de l'étude du transcriptome entier dans la PR réalisée précédemment au laboratoire GenHotel EA-3886 par Teixeira et *al.* en 2009 (177). Une liste de 6 gènes prioritaires a été retenue parmi les 330 gènes mis en évidence et regroupés selon leurs fonctions biologiques au sein du système immunitaire (**figure 11**).

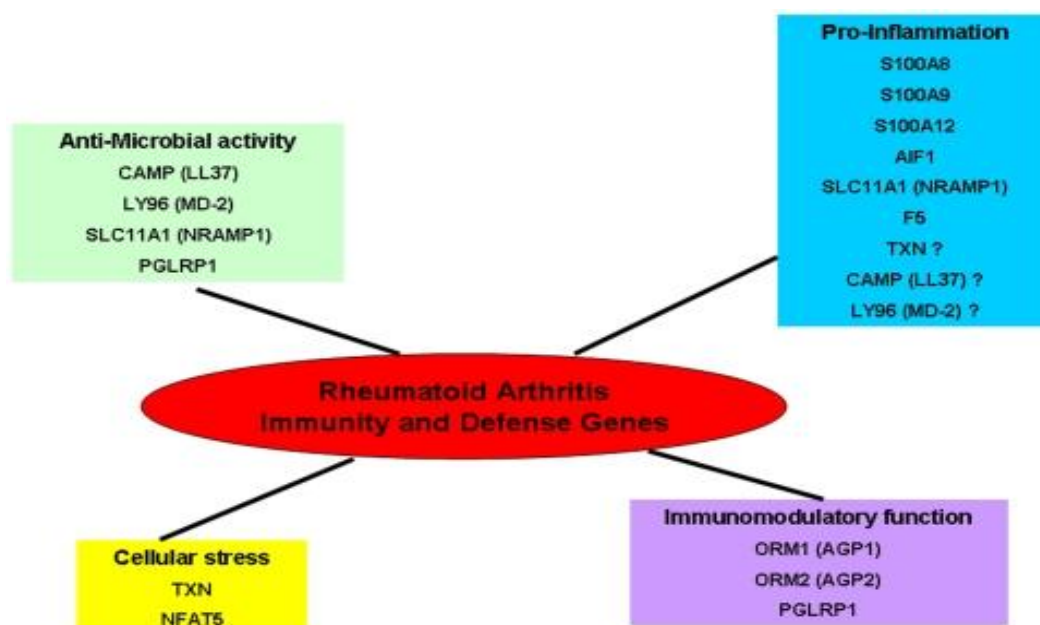


Figure 11 : Fonctions biologiques des gènes de l'immunité mis en évidence dans l'étude du transcriptome dans la PR (177).

- **LY96 (Lymphocyte antigen 96)**

Ce gène, situé en 8q21.11 et appelé aussi MD2, code une protéine qui agit comme une protéine adaptatrice extracellulaire dans l'activation de TLR4 (*Toll-like Receptor 4*) par le LPS (lipopolysaccharide) de bactéries à Gram négatif, un inducteur bien connu de la réponse immunitaire innée (178).

Des études chez des souris knock-out ont démontré que le MD-2 est indispensable pour les réponses LPS (179). Plusieurs études ont fait référence à l'importance potentielle de la signalisation TLR dans la pathogenèse de la polyarthrite rhumatoïde par le biais de la présence de différents ligands et de récepteurs fonctionnels de TLR dans les articulations enflammées provenant de patients atteints de PR (180).

Rappelons que le TLR4 est fortement exprimé dans la membrane synoviale à un stade précoce de la polyarthrite rhumatoïde ainsi que dans les stades avancés de la maladie. La stimulation *in vitro* des fibroblastes synoviaux de PR ayant le ligand de TLR4 (LPS) produit un large éventail de

cytokines pro-inflammatoires, des chimiokines, et des enzymes destructrices du tissu (181).

- **S100A8 (S100 calcium binding protein A8)**

La famille des gènes S100 comprend un groupe de gènes situés en 1q21 et qui codent des protéines à motif de liaison au calcium. Ces protéines, caractérisées par une forte prévalence dans les cellules d'origine myéloïde, sont impliquées dans une multitude de processus cellulaires comme la différenciation ou encore la progression du cycle cellulaire. Les phagocytes activés exprimant S100A8 et S100A9 sont parmi les premières cellules infiltrant les lésions inflammatoires dans la synoviale (182).

L'expression de S100A8 et S100A9 a été retrouvée comme étant la plus forte à la jonction cartilage - pannus, qui est le site principal de la destruction du cartilage et de l'érosion osseuse dans la polyarthrite (183). Par ailleurs, S100A12 est fortement exprimé dans le tissu synovial enflammé, alors qu'il est presque indétectable dans la synovie des témoins ou des patients après un traitement réussi (184).

D'autres études ont montré des niveaux accrus des concentrations de S100A8/S100A9 et S100A12 dans le liquide synovial et le sérum de patients atteints de PR (185, 186). Par conséquent, les protéines S100 pro-inflammatoires sont des cibles thérapeutiques intéressantes pour les interventions immunitaires dans le traitement de la PR.

- **PGLYRP1 (peptidoglycan recognition protein 1)**

Situé en 19q13.2-q13.3, PGLYRP1 est le premier gène de la famille des PGRPs qui codent des protéines de l'immunité innée, reconnaissant des peptidoglycanes (PG) bactériens et ayant un rôle dans l'immunité antibactérienne. La protéine PGLYRP-1 semble appartenir à la branche



immunitaire innée des molécules effectrices, telles que les peptides antimicrobiens et les lectines de type C. Cette protéine a été montrée comme étant presque exclusivement présente comme une protéine soluble dans les granules des polynucléaires neutrophiles (PNN) (187, 188).

Saha et ses collaborateurs (189) ont examiné les activités immunomodulatrices des PGRPs dans un modèle murin d'arthrite PG-induite. Ils ont montré qu'une injection systémique de PG ou muramyldipeptide (MDP) induit une arthrite aiguë des articulations chez les souris BalbC. Il a été remarqué, en outre, que les souris déficientes pour PGLRP1 souffrent d'arthrite MDP-induite qui dure plus longtemps que chez les souris WT (*Wild Type*).

La fonction anti-inflammatoire de PGLYRP1 ne se manifeste que dans les derniers stades de l'arthrite induite par le MDP, ce qui est cohérent avec la libération locale de PGLYRP1 à partir de granules PNN, après l'arrivée de ces derniers dans l'articulation de la souris.

Ces données indiquent que PGLYRP1 pourrait avoir un rôle spécialisé mais néanmoins significatif des événements de signalisation inflammatoire comme pendant l'arthrite chez les mammifères.

- **RUNX3 (runt-related transcription factor 3)**

Ce gène est localisé en 1p36.1 et est un membre de la famille RUNX (*runt-domain transcription factor family*). Une sous-unité de la protéine RUNX3 forme un complexe hétérodimérique qui se lie à l'ADN sur des séquences spécifiques de type *promoteur*. Le gène RUNX3 est impliqué dans la neurogenèse et la thymopoïèse, et fonctionne comme un suppresseur de tumeur (190-193). Dans ce sens, il a été rapporté que l'inactivation de RUNX3 était une condition cruciale de la prolifération de certains cancers (194), et *a contrario*, sa surexpression supprimerait le comportement « malin » des cellules du cancer de l'œsophage (195).

Dans un modèle de souris knock-out (KO) établi par Brenner et al. (196), les souris KO-RUNX3 ont développé des phénotypes d'IBD (*inflammatory Bowel Disease*) qui ressemblent aux manifestations IBD humaines. Des cellules dendritiques RUNX3-nulles dans les souris KO-RUNX3 ont montré une déficience de la réponse à TGF $\beta$  (197).

D'autres études ont suggéré que RUNX3 joue un rôle important dans la maturation, la migration et la fonction des cellules dendritiques en régulant l'expression de CD11a, CD49b, CD36 et CD103 (198-200). RUNX3 a également été impliquée dans la prolifération des cellules B, des cellules tueuses naturelles (NK), le développement des cellules *T helper* de type 1 (TH1) et de lymphocytes T cytotoxiques (201-204).

Sur le plan génétique, il a été rapporté que RUNX3 était associé avec la spondylarthrite ankylosante dans la population chinoise, et que certains SNPs du gène seraient même associés avec la sévérité de la maladie (205).

Compte tenu du rôle clé de RUNX3 dans le développement des cellules T et de la réponse immunitaire, il ne serait pas surprenant que le dysfonctionnement de RUNX3 puisse contribuer à des troubles immunitaires.

- **RNASE2 (Ribonuclease A family 2)**

Aussi connu comme EDN (*eosinophil-derived neurotoxin*), ce gène est situé en 14q11.2. Il code l'une des quatre principales protéines trouvées dans les granules cytoplasmiques de cellules importantes lors de la réponse immunitaire. Il s'agit des leucocytes éosinophiles humains, les cellules qui sont mobilisées à partir de la moelle osseuse vers le sang et les tissus en réponse à un stimulus (de type Th2) caractéristique de l'inflammation allergique et de l'infection par les helminthes parasites.

Dans ce sens, il a été prouvé qu'EDN avait aussi des propriétés chimiotactiques, ce qui vient enrichir son potentiel immunitaire alors que

cette activité est spécifique de cellules dendritiques, des acteurs de marque dans les processus inflammatoires (206). On notera en particulier la récente découverte qu'EDN peut initier la transduction du signal via une interaction avec le récepteur *Toll-like 2* (TLR2) (207), ce qui met directement en évidence le rôle important de cette protéine dans les mécanismes d'immunité innée et de reconnaissance des microorganismes.

Dans la polyarthrite rhumatoïde, Bovin et al. mettaient déjà en évidence en 2004 la différence d'expression d'un certain lot de gènes dont RNASE2 entre patients atteints de PR et sujets contrôles (208).

- **IL2RB (interleukin 2 receptor, beta)**

Le gène IL2RB est localisé en 22q12 et code le récepteur de l'interleukine 2 (ou encore CD122), qui est impliquée dans la réponse immunitaire T à médiation cellulaire, et qui est présent sous trois formes en fonction de la capacité de se lier à son ligand l'interleukine 2.

La forme de faible affinité est un monomère de la sous-unité alpha et n'est pas impliquée dans la transduction du signal. La forme d'affinité intermédiaire se compose d'une sous-unité hétérodimère alpha/beta, tandis que la forme de haute affinité se compose d'une sous-unité hétérotrimère alpha/beta/gamma. Les deux formes d'affinité intermédiaire et élevée du récepteur sont impliquées dans l'endocytose médiée par le récepteur et la transduction des signaux mitogéniques de l'interleukine 2 dans la réponse immunitaire cellulaire. La protéine codée par ce gène représente la sous-unité bêta et est une protéine membranaire de type I.

Ce gène a été rapporté comme étant associé avec la PR par plusieurs études notamment pour deux polymorphismes de types SNP : rs743777 et rs3218253 (50, 209). Dans une récente étude, l'haplotype « A-C » formé par ces mêmes SNPs a été associé avec la forme érosive de la PR débutante, indépendamment de la stratification selon le statut ACPA (210).

## **6 Données de Transcriptome dans la PR**

### **6.1 Etudes de profils d'expression à large échelle**

Malgré des études intensives, les mécanismes physiopathologiques et pathogéniques à l'origine de plusieurs maladies dont la PR sont encore mal compris. Chez les patients atteints de PR, plusieurs études de profil d'expression génique ont été faites dans les cellules du tissu synovial ou encore dans les cellules mononucléées du sang périphérique appelés aussi PBMCs (*Peripheral blood mononuclear cells*) à l'aide de la technologie des puces à ADN, dont la référence pour nous dans cette thèse reste celle réalisée par Teixeira et al. en 2009 (177). Dans le cas de la polyarthrite rhumatoïde, les différences dans les profils d'expression offrent des possibilités de stratifier les patients atteints de PR en fonction de critères moléculaires qui peuvent nécessiter différentes stratégies de traitement.

Les grands objectifs de l'étude du profil d'expression génique dans la PR se résument à (i) améliorer notre compréhension des mécanismes pathogéniques sous-jacents de la maladie, (ii) identifier de nouvelles cibles de médicaments, (iii) d'évaluer l'activité de la maladie, (iv) prévoir les résultats futurs, tels que la réactivité à un traitement, la gravité globale de la maladie, et le risque spécifique d'organes et finalement (v) développer de nouveaux tests de diagnostic (211), afin de permettre une meilleure prise en charge de la maladie.

L'approche synoviale est soulignée dans la mesure où la PR n'est pas une maladie systémique. Par exemple, les mécanismes inflammatoires ciblés par les anti-TNF $\alpha$  sont localisés dans le synovium et le profil d'expression génique du sang circulant de patients atteints de PR peut ne pas être représentatif de ces voies spécifiques du tissu synovial. Cette approche est généralement justifiée par son habilité à discriminer la PR des autres troubles articulaires (212), ainsi que par la capacité à identifier des variations du nombre de macrophages au cours du traitement.

L'étude du profil d'expression génique en utilisant les PBMC, permet d'identifier les processus à la fois pathogéniques et physiopathologiques comme cela a été démontré dans plusieurs types de maladies: cancer (213), asthme (214), LED (215), les maladies cardio-vasculaires (216) et des troubles psychiatriques (217). Les PBMCs, et plus exactement les lymphocytes, sont particulièrement pratiques pour la recherche médicale et autres applications diagnostiques car ils peuvent facilement et à plusieurs reprises être collectés en quantités suffisantes au cours de la maladie. La fonction altérée des lymphocytes dans les maladies, essentiellement auto-immunes, est le résultat d'une anomalie d'expression des gènes codant pour des cytokines, des composants de récepteurs, des voies de transduction du signal, et des modulateurs de la transcription et de la traduction. En outre, certains polymorphismes de gènes peuvent affecter leurs propriétés fonctionnelles ; c'est ce que l'on appelle un polymorphisme fonctionnel.

L'intérêt de l'expression des gènes à partir de PBMCs consiste à identifier, à la fois, les deux processus pathogénique et physiopathologique. Les processus pathogéniques sont principalement associés à la cause d'une maladie. Dans ce cas, l'étude du profil d'expression génique pourrait conduire à l'identification de gènes anormaux (de par leurs activités) qui peuvent ne pas être seulement limités aux PBMCs, mais dont la régulation pourrait être aussi affectée dans les cellules de l'organe pathologique. En revanche, les changements physiopathologiques de l'expression génique lymphocytaire sont considérés comme une réaction essentiellement normale du système immunitaire à un stimulus pathologique (infection bactérienne ou virale, inflammation, ...).

Par conséquent, et dans un contexte physiopathologique, les profils de gènes peuvent être partagés par une variété de maladies, alors que l'expression du gène pathogénique devrait être spécifique à la maladie (218).

Les études de profil d'expression dans la polyarthrite rhumatoïde peuvent être classées de manière globale en deux catégories:

- celles qui sont axées sur la recherche de « nouvelles pistes » des gènes candidats pour comprendre l'étiologie de la maladie et sa pathogénie,
- celles qui sont axées sur l'identification de profils d'expression typiques d'un état de la PR : en rapport avec la forme « érosive vs bénigne » ou par rapport à une réponse à un médicament « bonne réponse au traitement vs mauvaise réponse » ; c'est le cas des études de pharmacogénomique.

Dans la première catégorie, l'analyse de l'expression est souvent la première étape dans l'élucidation de la fonction des gènes, alors que dans la seconde, elle vise à réduire l'hétérogénéité phénotypique ou à identifier des profils d'expression qui peuvent servir d'outils de diagnostic de prédiction, issue de la maladie ou de la réponse à l'un des médicaments antirhumatismaux (tel que les anti-TNF $\alpha$ ) (219-221).

Des études de profils d'expression génique chez les patients atteints de PR ont non seulement révélé des gènes associés à la maladie elle-même (177), mais également identifié des sous-groupes distincts de patients atteints de PR que l'on peut désormais distinguer grâce à des profils d'expression génique différents (219, 220, 222-224).

## **6.2 Etudes Pharmacogénomiques : Signatures de réponse aux traitements**

L'expression des gènes avec une méthodologie haut débit peut contribuer à l'analyse et à l'identification de gènes qui sont impliqués dans la réponse au traitement ou dans les effets indésirables associés à la thérapie (225-229). Plusieurs études ont utilisé l'analyse de l'expression des gènes pour identifier les gènes « signatures » d'expression afin de prédire la réponse à un traitement par biothérapie chez des patients atteints de PR,

plus particulièrement l'infliximab et l'étanercept.(220, 221, 230). Cependant, leurs résultats sont hétérogènes.

Plus récemment, Mesko et al ont réalisé la première étude pharmacogénomique de réponse au tocilizumab chez 13 patients atteints de PR et traités par cette molécule. Les niveaux d'expression de 59 gènes au total ont montré des changements significatifs entre la ligne de base (avant début du traitement) et après 4 semaines de traitement. Quatre gènes (DHFR, CCDC32, EPHA, TRAV8) ont été identifiés comme étant associés avec une bonne réponse, et ce après correction pour tests multiples. Encore une fois, nous en savons très peu sur la fonction de ces gènes dans la PR. Les seules informations disponibles sur ces gènes dans la PR sont que la dihydrofolate reductase (DHFR) a été identifié comme facteur prédictif putatif de réponse MTX ; le T cell receptor variable d'alpha 8 (TRAV8) est impliqué dans les réponses des cellules T-CD8+ ; le récepteur de l'éphrine A4 (EPHA) joue un rôle dans le système nerveux, tandis que la fonction de CCDC32 reste inconnue dans la PR (231).

Les études décrites ci-dessus, menées dans des conditions extrêmement variables (type de tissu / cellule, type de puces, nombre de gènes et taille des échantillons), n'ont donc pas permis de définir un profil d'expression univoque, spécifique de la PR. Néanmoins, une dérégulation de la réponse immune semble ressortir clairement de ces travaux. La combinaison de ces différentes données dans des études globales de type méta-analyse permettrait potentiellement d'atteindre une puissance plus importante pour identifier une signature d'activité des gènes impliqués dans la PR. D'autre part, étudier le profil d'expression de cellules jouant directement un rôle dans la maladie, comme les cellules dendritiques (DCs) ou les LT CD4+, permettra de refléter davantage les changements d'activité des gènes jouant un rôle dans la PR.

### **III. OBJECTIFS DE LA THESE**



Nous nous sommes intéressés à l'étude génétique de la polyarthrite rhumatoïde (PR), maladie auto-immune, chronique et inflammatoire affectant principalement les articulations et entraînant leur destruction. La PR est une pathologie multifactorielle où des variants du locus HLA semblent être les principaux, mais non uniques, facteurs de prédisposition génétique.

Les analyses de gènes candidats dans des familles trios (constitué d'un patient atteint de PR et ses deux parents) ainsi que les analyses de polymorphismes du génome entier réalisées sur de grandes cohortes de cas comparés à des contrôles ont permis de dégager un certain nombre de facteurs génétiques mineurs. Cependant la totalité des facteurs génétiques n'est pas encore identifiée. Des analyses complémentaires du génome, telle que celle du transcriptome, sont des outils indispensables pour orienter la poursuite de l'étude des gènes candidats et plus largement, des biomarqueurs spécifiques de la PR.

Mon projet de thèse se divise en deux grandes parties tournant autour d'une plaque centrale, la génétique de la polyarthrite rhumatoïde, et qui se résume à ce qui suit :

- La première partie a été réalisée au laboratoire de Génétique Moléculaire et Cellulaire (Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf « USTO-MB » à Oran en Algérie, sous la direction du Pr. Abdallah Boudjema. L'étude génétique, de type cas-témoins, porte sur un total de 110 patients atteints de PR et de 197 témoins sains de l'Ouest Algérien. L'analyse d'association génétique a porté sur 5 polymorphismes déjà reconnus comme facteurs de prédisposition à la PR dans d'autres populations. Il s'agit de PTPN22<sub>rs2476601</sub>, STAT4<sub>rs7574865</sub>, IRF5<sub>rs2004640</sub>, 6q23<sub>rs6927172</sub> et TRAF1/C5<sub>rs10818488</sub>.

Ce travail a ensuite permis de comparer les résultats obtenus à ceux déjà produits dans des populations d'autres origines ethniques, et ainsi de pouvoir préciser l'impact de ces facteurs génétiques dans la PR.

- La seconde partie a été effectuée au laboratoire GenHotel-EA3886 (Université d'Evry Val d'Essonne « UEVE ») à Evry en France, sous la direction du Pr. Elisabeth Petit-Teixeira. Cette partie a porté sur l'analyse complémentaire des gènes différentiellement exprimés de manière significative lors de l'étude du transcriptome entier dans la PR réalisée précédemment au laboratoire par Teixeira et al. (177). Une liste de 6 gènes prioritaires a été retenue et des SNPs ont été choisis pour l'étude. L'association de ces polymorphismes avec la PR a été testée dans 440 familles trio (un patient atteint de PR et ses deux parents) ainsi que parmi les 14 patients PR et les 13 témoins pour lesquels le taux d'expression des gènes d'intérêt est connu. Ensuite, nous avons pu évaluer la relation entre les variants de chaque gène et leur taux d'expression.

Ce travail représente un complément d'information pour la compréhension du fonctionnement de ces gènes, la caractérisation de la relation entre leurs variants de séquence et leur expression, et enfin de leur implication dans la PR.

# **CHAPITRE 2 - POPULATIONS & MÉTHODES**

Au cours de ce travail de thèse, nous avons été amenés à travailler sur deux populations :

une première cohorte Algérienne de type cas-témoins, portant sur l'analyse d'association génétique entre des facteurs de prédisposition à la PR, déjà confirmés dans d'autres populations ;

et

une deuxième cohorte Européenne portant d'une part sur une étude familiale de type trio, ainsi que sur deux échantillons (cas et témoins) pour lesquels les données du transcriptome étaient disponibles et obtenues à partir d'une précédente étude réalisée en 2009 par Teixeira et al. au laboratoire GenHotel-EA3886 (177).

La présentation de cette partie Matériel et Méthodes se fera en deux volets :

1. Présentation de la cohorte Algérienne et des différentes étapes techniques de l'étude ;
2. Présentation de la cohorte Européenne ainsi que les différentes étapes techniques de l'étude.

L'ensemble des protocoles expérimentaux des techniques utilisées dans cette partie sont détaillés séparément dans le chapitre 5 (*Annexes*).

**I. ÉTUDE DES FACTEURS DE  
PREDISPOSITION GENETIQUE A  
LA PR DANS LA POPULATION DE  
L'OUEST ALGERIEN**

## **1 Population d'étude**

L'étude a été réalisée sur deux échantillons (cas atteints de PR et témoins sains) et la collection d'ADN a été établie au niveau du laboratoire de Génétique Moléculaire et Cellulaire de l'USTO-MB (Oran-Agérie). Les deux groupes sont originaires de l'Oranie (l'Ouest Algérien), une région qui regroupe les villes dont les coordonnées géographiques sont comprises entre latitude 36°20' nord, longitude 14°40' ouest et latitude 34°20' sud, longitude 1°15' ouest. L'origine ethnique « Ouest-Algérienne » de chaque individu a été définie par rapport au critère de naissance dans la région de l'Oranie en plus d'avoir une ascendance née dans cette région sur au moins deux générations (les parents et les quatre grands parents). Tous les individus participant à cette étude ont signé un consentement éclairé (**annexe 1**) conformément aux principes de la déclaration d'Helsinki (**annexe 2**).

Le premier échantillon « CAS-PR » a été constitué de 110 patients atteints de Polyarthrite Rhumatoïde, non apparentés et âgés de 18 à 78 ans (**tableau 4**). Le diagnostic a été posé en consultation au niveau du service de Rhumatologie du CHUO (Centre Hospitalo-Universitaire d'Oran, Algérie) selon les critères du collège américain de rhumatologie (ACR) de 1987 (66) et une fiche d'informations a été établie pour chaque malade (**annexe 3**).

Le deuxième échantillon a été constitué de 197 témoins sains non apparentés, âgés de 18 à 62 ans, et dont les parents ne présentent aussi aucun lien de parenté (absence de consanguinité). L'absence de toute maladie auto-immune chez les parents des témoins a été aussi retenue comme critère de sélection. Le recrutement des témoins sains s'est déroulé sur deux centres de prélèvement : le premier au secteur sanitaire de la commune d'Es-Sénia et le deuxième au centre de transfusion sanguine du Centre Hospitalo-Universitaire d'Oran (CHUO, Oran, Algérie).

Tableau 4 : Caractéristiques de la population d'étude cas-témoins Algérienne.

Caractéristiques	Cas PR	Témoins
Nombre de sujets	110	197
Femmes	99 (90%)	108 (54,82%)
Age (années)	48 ± 30,5	40 ± 22
Facteur Rhumatoïde +	31 (32%)	-

## 2 Extraction de l'ADN

Le protocole suivi est celui de la technique d'extraction au NaCl aussi appelée « *salting out* » qui consiste à traiter uniquement le lysat cellulaire par une solution saline afin d'éliminer les protéines fixées à l'ADN (232) (**annexe 4**). Le prélèvement d'un minimum de 5 ml de sang des individus se fait sur tube EDTA (Ethylène Diamine Tétra Acétique) et est stocké à une température de  $-20^{\circ}\text{C}$  jusqu'à l'extraction.

## 3 Choix des gènes et des SNPs

Le choix des SNPs des quatre premiers gènes (PTPN22<sub>rs2476601</sub>, STAT4<sub>rs7574865</sub>, IRF5<sub>rs2004640</sub> et TRAF1/C5<sub>rs10818488</sub>) et est basé sur les données publiées antérieurement dans les études d'association avec la PR de type « gènes candidats » ou encore des études dites génome entier « GWAS : *Genome Wide Association Study* » (50, 51). Le cinquième polymorphisme 6q23<sub>rs6927172</sub> a été choisi car il était en déséquilibre de liaison (DL) dans la population de caucasienne avec 6q23<sub>rs6920220</sub>, un autre SNP de

la même région génomique confirmé comme un facteur de risque de la PR (50, 127).

## 4 Génotypage moléculaire

Le génotypage des cinq polymorphismes a été réalisée avec un test de discrimination allélique par sondes fluorescentes en temps réel (TaqMan®, LifeTech) (**tableau 5**).

La réaction de polymérisation en chaîne des échantillons (PCR) est réalisée sur une machine de PCR en temps réel « ABI-7500 » (*Applied Biosystems*, maintenant *Life Technologies*, Foster City, CA, USA) selon le protocole du fabricant. Les conditions de PCR sont les mêmes quelque-soit le polymorphisme étudié, seul le mélange des sondes fluorescentes est spécifique de chaque polymorphisme (Assay). Le principe de la chimie TaqMan® est détaillé dans l'**annexe 5**.

**Tableau 5 : SNPs et gènes sélectionnés pour l'analyse dans la population Algérienne**

Gène/Locus	db SNP ID	Assay ID LifeTech	SNP	Localisation ( NCBI Build 37.2 * )	Polymorphisme
<b>PTPN22</b>	rs2476601	C__16021387_20	A/G	Chr.1: 114377568	Mutation faux sens
<b>STAT4</b>	rs7574865	C__29882391_10	G/T	Chr.2:191964633	Intron
<b>6q23 (TNFAIP3/OLIG3)</b>	rs6927172	C___1575580_10	C/G	Chr.6: 138002175	Intergénique
<b>IRF5</b>	rs2004640	C___9491614_10	G/T	Chr.7: 128578301	Intron
<b>TRAF1/C5</b>	rs10818488	C___2783655_10	A/G	Chr.9: 123705087	Intergénique

db : data base ; SNP : single nucleotide polymorphism ; Chr : chromosome ; NCBI: National Center for Biotechnology Information ; ID : Identification Number

\*<http://www.ncbi.nlm.nih.gov/projects/mapview/stats/BuildStats.cgi?taxid=9606&build=37&ver=2>



Le résultat de génotypage de chaque échantillon a été attribué automatiquement à l'aide du logiciel ABI PRISM® 7500 *Sequence Detection System* "SDS" pour la discrimination allélique. Les valeurs d'intensité de fluorescence spécifique de chaque allèle sont cependant vérifiées pour chaque classe de génotypes.

Pour le contrôle qualité des résultats produits, deux échantillons « références » issus des familles du CEPH « Centre d'Étude du polymorphisme Humain » ont été co-génotypés lors de chaque expérience de génotypage et 10% des échantillons choisis au hasard ont été génotypés une seconde fois. Aucun résultat discordant n'a été observé lors de ces tests, garantissant la reproductibilité de nos expériences. Par ailleurs, le taux de réussite de chaque génotypage est supérieur à 99%.

## **5 Analyses statistiques**

L'analyse statistique réalisée dans le cadre de notre étude a été principalement basée sur des comparaisons de fréquences des allèles des gènes étudiés entre patients PR et témoins sains, grâce à l'utilisation de tests du chi-carré ( $X^2$ ) en utilisant le logiciel *Epi-Info 6.0™* qui est un logiciel de statistiques appliqué à l'épidémiologie disponible en accès et téléchargement gratuit à l'adresse : <http://www.epiconcept.fr>.

Pour l'analyse cas-témoins, le calcul des fréquences alléliques pour chaque SNP est effectué pour le groupe des patients et le groupe contrôle après avoir vérifié au préalable que l'équilibre de Hardy-Weinberg est respecté chez les contrôles à l'aide d'un test du chi-carré ( $X^2$ ) standard. Cette évaluation classique du  $X^2$  est possible lorsque les effectifs sont supérieurs à 5. Dans le cas contraire, il est nécessaire d'utiliser le  $X^2$  corrigé, soit avec la correction de Yates (effectif inférieur à 5) soit avec la correction de Fisher (effectif inférieur à 3).

Pour rappel, la loi de Hardy-Weinberg permet le calcul de fréquences génotypiques théoriques à partir de fréquences alléliques. L'équilibre de Hardy-Weinberg est le modèle théorique central de la génétique des populations. Cette notion est soumise à certaines conditions, sous lesquelles, la diversité génétique de la population se maintient et doit tendre vers un équilibre stable de la distribution génotypique :

- la population est très grande et panmictique (les couples se forment au hasard et leurs gamètes se rencontrent au hasard),
- il ne doit y avoir ni sélection, ni mutation, ni migration,
- les générations successives sont discrètes (pas de croisement entre générations différentes).

Un écart à l'équilibre d'Hardy-Weinberg observé dans la population contrôle peut être également le reflet d'une erreur de génotypage conduisant à un excès d'hétérozygotes.

Les résultats de génotypage des cas PR et des témoins non apparentés ont été comparés en utilisant le test du chi-carré (tableaux 2x2 de contingence) pour la signification statistique. La méthode du risque relatif de génotype (GRR) (comparaison d'un seul génotype par rapport aux autres) a été utilisée pour comparer la distribution des génotypes dans les contrôles et des patients. Le test du GRR décrit par Lathrop (233) ajuste les fréquences génotypiques dans les contrôles aux proportions attendus de Hardy-Weinberg et donne des estimations plus précises des risques relatifs attendus.

L'évaluation du degré de significativité ( $P_{\text{value}}$ ) des différences de fréquence de chaque allèle entre patients et témoins correspond à la probabilité que l'écart global soit imputable seulement aux fluctuations du hasard. Lorsque la probabilité P est égale ou inférieure à 0,05 (5%) il y a moins de 5 chances sur 100 que la distribution résulte du hasard. Ainsi la différence de distribution entre les populations de malades et de témoins

pour un marqueur donné, est statistiquement significative et le marqueur peut être considéré comme étant associé à la maladie.

Les odds ratios (OR) et les intervalles de confiance à 95% (IC) ont été calculés avec la méthode de Woolf (234). L'OR a été calculé en tenant compte de l'allèle à risque de chaque SNP ou des génotypes contenant les allèles à risque. Ce paramètre statistique (OR) est utilisé dans les études de type cas-témoins où il permet l'appréciation de l'intensité d'une association entre un marqueur et une maladie. C'est une estimation du risque relatif dans le cas où il est impossible de mesurer les risques de la maladie chez les sujets exposés et non exposés au facteur étudié (235). Lors d'une étude d'association de type cas-témoins, seul l'OR peut être utilisé et non pas le risque relatif. L'OR, comme toute estimation réalisée sur un échantillon, doit être présenté avec son intervalle de confiance à 95 %, qui mesure la précision de l'estimation.

L'ajustement pour les tests multiples a été réalisé en utilisant la correction de Bonferroni (42). Le degré de significativité pour la valeur P a été réévalué à 0,01 (0,05/5 considérant 5 SNPs testés sur l'ensemble de l'étude). Pour rappel, cette méthode appelée aussi méthode de *Bonferroni-Holm*, consiste à réaliser les tests avec un seuil de signification plus petit que 5% et de choisir cette valeur de telle sorte qu'après l'inflation due aux comparaisons multiples, le niveau global atteint soit de 5%. Les tests sont donc réalisés avec un seuil de « 0.05/k » où k désigne le nombre de comparaisons effectuées

**II. EXPLOITATION GENETIQUE  
DES DONNEES DU  
TRANSCRIPTOME DANS LA PR  
DANS DES FAMILLES  
EUROPEENNES**

## 1 Population d'étude

L'étude réalisée au niveau du laboratoire européen de recherche sur la Polyarthrite Rhumatoïde (GenHotel-EA3886) a porté sur un total de 1347 individus (14 cas PR isolés, 13 témoins sains et 440 famille trios) présentés dans le **tableau 6** et répartis de la manière suivante :

- L'**échantillon 1** (*SAMPLE 1*) formé de 14 cas index atteints de PR et 13 sujets témoins sains, pour lesquels des données transcriptome ont été produites précédemment au laboratoire GenHotel (177).
  
- L'**échantillon 2** (*SAMPLE 2*) constitué lui-même de deux sous-groupes (Set-1 et Set-2) qui ont été regroupés pour les besoins de l'étude. Les individus dont les ADN forment ces deux sous groupes ont été recrutés sur la base des mêmes critères mais sur des périodes différentes :
  - *Set-1* : comportant cent (100) familles trio, composées chacune d'un patient atteint de PR et ses deux parents, d'origine française caucasienne (grands-parents également français).
  - *Set-2* : comportant cent (100) autres familles trio, composées chacune d'un patient atteint de PR et ses deux parents, d'origine française caucasienne (grands-parents également français).
  
- L'**échantillon 3** (*SAMPLE 3*) originairement appelé Set-3 comportant deux cent quarante (240) familles trio, composées chacune d'un patient atteint de PR et ses deux parents, d'origine européenne caucasienne (France, Italie, Portugal, Espagne, Belgique et Pays-Bas).

Les familles trios ont été recrutées en France entre 1994 et 1997 par auto-signallement du cas index au cours d'une campagne médiatique. Tous les patients avaient une PR répondant aux critères de 1987 du collège américain de rhumatologie (ACR) pour la PR (66) selon les données fournies par le rhumatologue en charge du patient, les diagnostics ayant ensuite été validés par deux rhumatologues. Tous les individus participant ont signé un consentement éclairé et ce projet a reçu l'avis favorable du CCPPRB (comité consultatif de protection des personnes dans la recherche biomédicale) de l'hôpital du Kremlin-Bicêtre en 1994.

Les caractéristiques des patients PR de ces trois groupes sont décrites ci-dessous dans le **tableau 6**.

**Tableau 6: caractéristiques des patients PR dans les trois groupes Européens**

	<b>Echantillon 1 (n = 14)</b>	<b>Echantillon 2 (n = 200)</b>	<b>Echantillon 3 (n = 240)</b>
Femmes, n (%)	12 (86)	177 (88,5)	212 (85)
Moyenne d'âge (Années) (± standard de déviation)	47 (±12)	30,5 (±8)	31 (±9)
Patients PR avec érosion osseuse, n (%)	10 (72)	168 (84)	167/234 (71)
Patients PR séropositifs pour le facteur rhumatoïde n (%)	13 (93)	153 (76,5)	170/231 (74)
Patients PR séropositifs pour les ACPA, n (%)	8/9 (89)	155 (77,5)	60/83 (72)
Patients PR porteurs d'au moins un allèle épitope partagé HLA-DRB1, n (%)	ND	159 (79,5)	ND

n: nombre de cas; ND : Non Disponible; ACPA : Anticorps anti-peptide citrullinés

## 2 Choix des gènes et des SNPs

Le choix des gènes pour notre étude a été déterminé à partir des résultats de la précédente étude réalisée par *Teixeira et al.* en 2009 au niveau du laboratoire GenHotel portant sur l'analyse du transcriptome dans la PR (177). Durant cette première étude, les puces d'expression *Illumina human-Expression 6V2 BeadChips* ont été utilisées pour l'établissement du profil de transcription complet du génome à partir de cellules mononucléées du sang périphérique (PBMC) de patients atteints de PR et des témoins sains. L'analyse différentielle gène par gène a été réalisée avec une analyse de variance (Anova) et les valeurs P ont été ajustées pour contrôler le *False discovery rate* « FDR 5% ». Une expression différentielle a été observée pour 339 gènes (238 sous-exprimés et 101 sur-exprimés) entre les deux groupes. Les données des puces à ADN ont été validées par PCR en temps réel dans un ensemble de quelques gènes présentant un haut degré de corrélation.

Pour la présente étude, à partir de la liste initiale de 339 gènes citée plus haut, nous avons sélectionné initialement 15 gènes ayant un *Fold Change* (facteur de variation d'expression) supérieur à 1,9. Le *Fold Change* est défini par le rapport de niveau d'expression comparé entre les patients PR et les contrôles. A partir de la liste initiale, six gènes ont été retenus pour la présente étude: S100A8, RNASE2, PGLYRP1, RUNX3, IL2RB et LY96 pour lesquels les mises au point techniques au niveau du laboratoire ont été les plus concluantes. Pour ces gènes, dix SNP ont été sélectionnés en tenant compte des données de déséquilibre de liaison, de la présence de tag-SNPs et des fréquences des allèles mineurs (*Minor Allele Frequencies* ou MAF) supérieures à 5% (**tableau 7**).

**Tableau 7: Gènes and SNPs sélectionnés et leur MAF respectives dans la population caucasienne de référence**

Gène	Position *	Fold change **	SNP	Allèles	MAF (allèle) ***
<b>S100A8</b>	1: 153362508-153363664	2.9	rs3006488	A/G	0.16 (G)
<b>LY96</b>	8:74903587-74941322	2.58	rs1991262	G/C	0.32 (C)
<b>RNASE2</b>	14: 21423611-21424595	2.55	rs2013109	C/G	0.24 (C)
<b>IL2RB</b>	22:37521878-37571094	2.15	rs3218253	G/A	0.17 (A)
<b>PGLYRP1</b>	19: 46522411-46526323	2	rs2072561	T/G	0.09 (G)
			rs2041992	C/T	0.41 (T)
<b>RUNX3</b>	1:25226002-25291612	1.91	rs12082398	C/T	0.45 (T)
			rs1507101	G/A	0.48 (A)
			rs2236851	C/T	0.11 (T)
			rs2282718	G/A	0.39 (A)

\* Position à partir de la base [www.ensembl.org](http://www.ensembl.org)

\*\* Fold Change : facteur de variation d'expression

\*\*\* MAF: minor allele frequency à partir de la population référence HapMap-CEU ([www.hapmap.org](http://www.hapmap.org))

Pour les neuf gènes non retenus, cinq avaient au moins un variant de type CNV (*Copie Number Variant*) décrit dans la base de données des variants génomiques (<http://dgv.tcag.ca/>). Les quatre autres gènes ont également été exclus, car leurs polymorphismes ne possédaient pas de fréquence de l'allèle mineur (MAF) supérieure à 5%.

### 3 Génotypage moléculaire

L'ADN génomique des échantillons a été précédemment isolé et purifié à partir de nouveaux leucocytes du sang périphérique selon un protocole standard. Pour le génotypage des 10 SNPs répartis sur les 6 gènes retenus pour cette partie, deux méthodes de génotypage ont été



utilisées :

Le génotypage du premier SNP rs1991262 (gène LY96) a été réalisée par la technique d'amplification par réaction de polymérisation en chaîne suivie par une digestion enzymatique (PCR-RFLP) (236). Les séquences des amorces conçues sont *Forward* : 5'-TAGGCTAGAGATTTCTTGGAGGA-3' et *Reverse* : 5'-GCAGACAGCCTGCAGCTGCC-3'. Le fragment d'amplification produit, qui fait une taille de 361bp, a été digéré avec l'enzyme de restriction ScrFI (*New England Biolabs Inc. Ipswich, MA, USA*), générant trois fragments pour l'allèle polymorphe C (183bp, 111bp et 67 pb) et deux fragments pour l'allèle ancestral G (250 pb et 111bp) (**annexe 6**).

Le génotypage des 9 SNPs restants (**tableau 8**) a été réalisé avec un test de discrimination allélique par sondes fluorescentes en temps réel (TaqMan®) selon le protocole décrit précédemment.

**Tableau 8 : Caractéristiques des sondes TaqMan® utilisées dans cette étude**

Assay ID	Gène	Référence NCBI - dbSNP	Localisation chromosomique	Polymorphisme
C__25741_10	S100A8	rs3006488	1q21.3	Intergénique
C__26538682_10	RNASE2	rs2013109	14q11.2	Intron
C__15871175_10	PGLYRP1	rs2072561	19q13.32	Intron
C__347749_10	PGLYRP1	rs2041992	19q13.32	Intron
C__31640679_10	PGLYRP1	rs13343537	19q13.32	Mutation Faux Sens
C__30928041_10	RUNX3	rs12082398	1p36.11	Intron
C__2470436_1_	RUNX3	rs1507101	1p36.11	Intron
C__2470432_1_	RUNX3	rs2236851	1p36.11	Intron
C__15957215_10	RUNX3	rs2282718	1p36.11	Intron

db : data base ; SNP : single nucleotide polymorphism ; NCBI: National Center for Biotechnology Information ; ID : Identification Number

Pour le contrôle de la qualité de nos données, la même démarche précédemment décrite a été adoptée pour les 10 SNPs sélectionnés. Aucun résultat discordant n'a été observé lors de ces tests, garantissant la reproductibilité de nos expériences. Par ailleurs, le taux de réussite de chaque génotypage est supérieur à 99%.

## **4 Analyses statistiques**

Pour chacun des dix SNP appartenant aux 6 gènes sélectionnés, le niveau d'expression moyen (obtenu à partir des données du transcriptome (177)) pour les différents génotypes ont été comparés entre les 14 patients PR et les 13 témoins (**échantillon 1**). Nous avons regroupé les génotypes homozygotes pour lesquelles la taille de l'échantillon était inférieure à 5 avec des génotypes hétérozygotes. Cette comparaison a été évaluée par des tests non paramétriques (*Kruskal-Wallis* et *Mann-Whitney*) en utilisant le logiciel XLSTAT (*Addinsoft, Paris, France, 2012 version [www.xlstat.com](http://www.xlstat.com)*).

Avant d'effectuer les tests d'association, l'équilibre de Hardy-Weinberg a été vérifié dans les contrôles (**échantillon 1**). Pour les deux autres **échantillons 2 et 3**, composés de famille trio, les contrôles sont constitués par les allèles non transmis des parents aux cas index PR des familles trio. Le test d'association en milieu familial (*Family Based Association Test ou FBAT*), dérivé du test original « test de déséquilibre de transmission » ou TDT (34), a été utilisé pour comparer la distribution des allèles des SNPs étudiés dans la descendance affectée avec la distribution attendue en utilisant la loi de Mendel. Cette analyse a été réalisée avec le logiciel gratuit « FBAT » (version 2.0.4 <http://www.biostat.harvard.edu/FBAT.htm>) (237).

Nous avons ensuite évalué le risque relatif de génotypes GRR (*Genotype Relative Risk*), qui compare le génotype de la descendance affectée avec le génotype du contrôle dérivée des chromosomes parentaux non-transmis, en utilisant la méthode décrite par M. Lathrop (233). Les

analyses des haplotypes des gènes RUNX3 et PGLYRP1 ont été réalisées en utilisant l'option HBAT (*Haplotype-Based Association Test*) dans le logiciel FBAT (237).

La stratification des familles trio a ensuite été effectuée en ce qui concerne la présence chez les patients atteints de PR d'au moins un des allèles de l'épitope partagé (SE) du gène HLA-DRB1 (01\*01, 01\*02, 04\*01, 04\*04, 04\*05, 04\*08, 04\*09, 10\*01 et 14\*02, associés à la PR dans notre population) (238). D'autres stratifications ont été faites selon des anticorps anti-peptides citrullinés (ACPA) et l'état de l'érosion osseuse.

Le degré de significativité P fixé initialement à 0,05 pour chaque analyse individuelle d'association a été réévalué à  $7 \times 10^{-4}$  en appliquant la correction de Bonferroni pour des tests multiples, compte tenu des 10 SNPs testés et des 6 stratifications réalisées.

Si au moins une analyse d'association montre un résultat significatif dans l'échantillon de découverte (**échantillon 2**) nous avons testé la même association dans l'échantillon de réplication (**échantillon 3**). Une analyse globale après regroupement des deux échantillons de familles trio (**échantillons 2 et 3**) a également été réalisée.

# ***CHAPITRE 3 – RESULTATS & DISCUSSION***

Ce chapitre *Résultats & Discussion* se présente pour les deux parties du travail (Algérienne et Française) comme suit :

- Une introduction pour rappeler le contexte, les gènes et les polymorphismes explorés, la population étudiée et un bref descriptif bibliographique ;
- Les résultats obtenus seront présentés de manière brève ;
- Une discussion générale pour chacune des parties reprenant la littérature scientifique et les principales études se rapprochant le plus de la notre ;
- La présentation de l'article produit pour chaque partie de l'étude

Les deux articles produits et présentés dans ce chapitre sont :

- ***“Association of PTPN22 (rs2476601) and STAT4 (rs7574865) polymorphisms with Rheumatoid Arthritis in the Western Algerian population”***

publié dans la revue *Acta Reumatologica Portuguesa*.

- ***“Relationship between SNPs and expression level for candidate genes in Rheumatoid Arthritis”***

publié dans la revue *Scandinavian Journal of Rheumatology*.

**I. ÉTUDE DES FACTEURS DE  
PREDISPOSITION GENETIQUE A  
LA PR DANS LA POPULATION DE  
L'OUEST ALGERIEN**

## **1 Introduction**

La polyarthrite rhumatoïde (PR) est une maladie auto-immune complexe caractérisée par des réponses immunitaires (innée et spécifique) modifiées provoquant la destruction auto-immune des tissus et des organes touchés. Bien que l'étiologie de la PR ne soit pas entièrement connue, il est aujourd'hui clairement démontré que des facteurs à la fois génétiques et environnementaux contribuent au développement de cette maladie auto-immune.

Des études génétiques et familiales antérieures indiquent une contribution génétique significative au développement de la PR (114). L'association du gène HLA-DRB1 avec la prédisposition à développer la PR a été bien établie dans plusieurs groupes ethniques. Cependant, la région HLA ne contribue qu'à hauteur de 30 à 50% de la composante génétique de la PR (124).

Les progrès de la technologie de génotypage ont facilité l'application d'approches globales portant sur l'ensemble du génome et ce afin d'identifier les variants causaux de la maladie. Des études pangénomiques récentes (GWAS) ainsi que d'autres études « gène candidat » ont rapporté des preuves supplémentaires de l'association de polymorphismes nucléotidiques simples (SNPs) dans plusieurs gènes candidats et / ou loci tels PTPN22, STAT4, OLIG3/TNFAIP3 (6q23), IRF5 et TRAF1/C5 (5q, 12q). Tous ces gènes sont impliqués dans la réponse immunitaire et/ou sa régulation.

Le but de cette partie de l'étude a été de tester, dans la population de l'Ouest Algérien, l'association entre ces cinq marqueurs génétiques confirmés (PTPN22<sub>rs2476601</sub>, STAT4<sub>rs7574865</sub>, 6q23<sub>rs6927172</sub>, IRF5<sub>rs2004640</sub> et TRAF1/C5<sub>rs10818488</sub>) et la PR. À notre connaissance, cette étude est la première à être réalisée dans la population Algérienne afin de trouver une association génétique entre ces polymorphismes et la PR.

## 2 Résultats

Les caractéristiques démographiques et cliniques de tous les sujets sont résumées dans le **tableau 9**. Le sexe ratio observé de 9 femmes pour 1 homme (9F / 1H) semble être une caractéristique ethnique de la population algérienne, compte tenu de la différence avec la population Française, et Européenne en général, qui varie entre 3 à 4 femmes pour un homme (3 à 4F / 1H).

Pour les cinq polymorphismes étudiés, nous n'avons pas observé d'écart à l'équilibre de Hardy-Weinberg dans le groupe témoin.

**Tableau 9: Caractéristiques de l'échantillon cas/témoins étudié**

Caractéristique	Cas PR (%)	Contrôles (%)	<i>P</i> <sub>value</sub>
Nombre de sujets	110	197	-
Homes	11 (10%)	89 (45.18%)	NS
Femmes	99 (90%)	108 (54.82%)	NS
Age (années)	48 ± 30.5	40 ± 22	NS
Facteur Rhumatoïde +	31 (32%)	-	-
Facteur Rhumatoïde -	66 (68%)	-	-

Les valeurs « âge » sont présentées en moyenne ± déviation standard.

NS : Non significatif

Pour l'étude d'association, nous avons observé une association significative avec la PR dans notre population pour deux SNPs (PTPN22<sub>rs2476601</sub> et STAT4<sub>rs7574865</sub>). Les trois autres SNPs (6q23<sub>rs6927172</sub>, IRF5<sub>rs2004640</sub> et TRAF1/C5<sub>rs10818488</sub>) n'ont pas atteint une signification statistique au niveau allélique (**tableau 10**).



**Tableau 10: Analyse d'association des 5 SNPs étudiés chez les cas et les contrôles Algériens**

Gène / SNP	Génotype/Allèle	Cas PR n=110 (%)	Contrôles n=197 (%)	X <sup>2</sup>	P <sub>value</sub>	OR (95% CI)
<b>PTPN22</b> <b>rs2476601</b>	TT	1 (1)	0 (0)	45.5	<b>1.31 x 10<sup>-10</sup></b>	9.83 (4.28 – 22.56)
	TC	33 (30)	7 (4)			
	CC	76 (69)	189 (96)			
	T	35 (16)	7 (2)	43.97	<b>3.32 x 10<sup>-11</sup></b>	
	C	185 (84)	385 (98)	45.38	<b>8.36 x 10<sup>-9</sup></b>	
	<b>TT + TC</b>	34 (31)	7 (4)			
CC	76 (69)	189 (96)				
<b>STAT4</b> <b>rs7574865</b>	GG	53 (48)	133 (68)	11.65	<b>2.9 x 10<sup>-3</sup></b>	1.75 (1.16 – 2.63)
	GT	53 (48)	57 (29)			
	TT	4 (4)	7 (4)			
	T	61 (28)	71 (18)	7.88	<b>4 x 10<sup>-3</sup></b>	
	G	159 (72)	323 (82)	11.05	<b>8 x 10<sup>-4</sup></b>	
	<b>GT + TT</b>	57 (52)	64 (32)			
GG	53 (48)	133 (68)				
<b>6q23</b> <b>rs6927172</b>	CC	58 (53)	135 (70)	11.39	<b>3 x 10<sup>-3</sup></b>	0.64 (0.42 – 0.97)
	CG	49 (45)	50 (26)			
	GG	3 (2)	9 (4)			
	G	55 (25)	68 (18)	4.86	<b>0.027 *</b>	
	C	165 (75)	320 (82)	8.61	<b>3 x 10<sup>-3</sup></b>	
	<b>CG + GG</b>	52 (47)	59 (30)			
CC	58 (53)	135 (70)				
<b>IRF5</b> <b>rs2004640</b>	GG	9 (8)	38 (19)	15.84	<b>3 x 10<sup>-4</sup></b>	1.01 (0.72 – 1.4)
	GT	80 (73)	97 (50)			
	TT	21 (19)	60 (31)			
	T	122 (55)	217 (56)	0.002	0.96	
	G	98 (45)	173 (44)	6.9	<b>8.6 x 10<sup>-3</sup></b>	
	<b>GT + TT</b>	101 (92)	157 (81)			
GG	9 (8)	38 (19)				
<b>TRAF1/C5</b> <b>rs10818488</b>	AA	7 (6)	38 (19)	11.63	<b>2.9 x 10<sup>-3</sup></b>	0.81 (0.57 – 1.13)
	AG	69 (63)	92 (47)			
	GG	34 (31)	66 (34)			
	A	83 (38)	168 (43)	1.53	0.21	
	G	137 (62)	224 (57)	0.24	0.62	
	<b>AA + AG</b>	76 (69)	130 (66)			
GG	34 (31)	66 (34)				

\* P<sub>value</sub> non significative après correction de Bonferroni (5 SNPs « alpha=0.01 »).

PR : polyarthrite rhumatoïde ; OR : Odds Ratio ; SNP: single nucleotide polymorphism

Le résultat d'association la plus significative a été obtenu avec PTPN22<sub>rs2476601</sub> avec une  $P_{\text{value}}$  allélique de  $3,32 \times 10^{-11}$  (OR = 9,83, IC 95% [4,28 à 22,56]). De plus, la présence d'au moins une copie de l'allèle T<sub>rs2476601</sub> présentait un risque significativement élevé ( $P = 8,36 \times 10^{-9}$ , OR = 12,08 et IC 95% [4,86 à 33,42]).

Une autre association significative a été obtenue avec STAT4<sub>rs7574865</sub> ( $P_{\text{allélique}} = 4 \times 10^{-3}$ ; OR = 1,75, IC 95% [1,16 à 2,63]). La présence d'au moins une copie de l'allèle T<sub>rs7574865</sub> a montré une différence significative entre les patients et les témoins ( $P = 8 \times 10^{-4}$ ; OR = 2,23 et IC 95% [1,35 à 3,71]).

Le troisième SNP, 6q23<sub>rs6927172</sub>, a montré une association significative ( $P_{\text{allélique}} = 0,027$ ; OR = 0,64, IC 95% [0,42 à 0,97]), mais la significativité n'a pas pu être reproduite selon nos critères après l'application de la correction pour tests multiples de Bonferroni. Cependant, au niveau génotypique, la valeur P pour 6q23<sub>rs6927172</sub> était toujours significative après correction de Bonferroni ( $P_{\text{génotypique}} = 3 \times 10^{-3}$ ).

Enfin, IRF5<sub>rs2004640</sub> et TRAF1/C5<sub>rs10818488</sub> ont montré une association significative seulement au niveau génotypique ( $P_{\text{Values}} = 3 \times 10^{-4}$  et  $2,9 \times 10^{-3}$  respectivement).

### **3 Discussion**

Une association significative avec la PR a été obtenue pour deux SNP (PTPN22<sub>rs2476601</sub> et STAT4<sub>rs7574865</sub>) dans notre population, les trois autres SNP (6q23<sub>rs6927172</sub>, IRF5<sub>rs2004640</sub> et TRAF1/C5<sub>rs10818488</sub>) n'ont pas atteint la signification statistique que nous nous sommes fixés dans les critères de l'étude (**tableau 10**). Cette étude préliminaire dans la population de l'Ouest algérien est largement en accord avec les études précédentes ayant rapporté les mêmes constatations dans d'autres populations, et que nous allons énumérer ci dessous.

Au cours de cette étude, la plus forte association avec la PR a été observée avec le polymorphisme PTPN22<sub>rs2476601</sub>. La première étude de l'association de ce polymorphisme avec la PR a été menée en 2004 (239), puis de nombreuses études ont confirmé cette association dans différentes populations (France (137) Royaume-Uni (240) Finlande (241), Suède (242), Allemagne (243), Pays bas (244), Espagne (245) et Canada (246). Cependant, une étude dans une population japonaise n'a pas pu confirmer l'association dans cette population et ce à cause d'une très faible fréquence de l'allèle mineur (247). Deux autres études chez les Tunisiens, une population ayant une origine ethnique proche de celle des Algériens, ont rapporté des résultats controversés sur l'association de PTPN22<sub>rs2476601</sub> avec la PR. En effet, Chabchoub et al. (248) ont décrit une absence d'association alors que Sfar et al. (249) ont confirmé l'association avec la polyarthrite rhumatoïde, à travers deux études cas-témoins indépendantes menées dans deux régions distinctes en Tunisie.

La deuxième plus importante association a été détectée avec STAT4<sub>rs7574865</sub> ( $P = 4 \times 10^{-3}$ , OR = 1,75, IC 95% [1,16 à 2,63]). De plus, la présence d'au moins une copie de l'allèle T<sub>rs7574865</sub> montré une différence significative entre les patients PR et les témoins sains ( $P = 8 \times 10^{-4}$ , OR = 2,23 IC à 95% [1,35 à 3,71]). Nos résultats sont en parfaite conformité avec les études antérieures sur l'association de STAT4<sub>rs7574865</sub> avec la PR dans différentes populations (148, 250-254).

Le troisième polymorphisme, 6q23<sub>rs6927172</sub>, a montré une tendance à l'association avec la PR dans notre étude, puisqu'il n'a pas atteint la signification statistique après application de la correction de Bonferroni ( $P_{\text{allélique}} 0,027 > P_{\text{corrigé}} 0.01$ ). Cela pourrait être dû au modèle de conception de l'étude (étude cas-témoins et taille d'échantillon relativement réduite), ou alors être une caractéristique ethnique de la population algérienne de l'Ouest. Cette observation n'est pas très différente des résultats d'une étude menée chez les espagnols, une autre population méditerranéenne, et qui fait état d'une faible association entre les SNPs situés dans cette région

génomique (6q23) et la PR. En effet, les auteurs de cette étude ont conclu à une faible preuve d'association à la fois dans la région intergénique 6q23 et dans le locus TNFAIP3 (255). Dans l'ensemble, l'association statistique a été mieux expliquée par la contribution (effet additif) de SNPs des deux loci TNFAIP3 et la région intergénique 6q23. Une autre étude chez les Tunisiens a aussi rapporté une tendance d'association entre le polymorphisme rs6920220 et la PR (256). Il est maintenant clair que le locus 6q23 est un candidat solide pour un rôle dans la polyarthrite rhumatoïde, cependant, notre étude ne représente qu'un premier élément suggérant l'implication de ce locus dans susceptibilité à la PR, dans la population Algérienne.

Concernant IRF5<sub>rs2004640</sub> une association significative a été obtenue seulement au niveau génotypique ( $P_{\text{génotypes}} = 3 \times 10^{-4}$ ) mais n'a pas atteint la signification statistique concernant les fréquences des allèles ( $P_{\text{allélique}} = 0,96$ ). Un résultat similaire a été obtenu dans la population Tunisienne par Maalej et al. où en effet, aucune différence significative n'a été observée au niveau allélique entre patients atteints de PR et témoins sains (257). Toutefois, au niveau génotypique, le génotype TT était plus fréquent chez les patients atteints de PR que chez les témoins (42,1% et 31,4% respectivement,  $P = 0,01$ ). Ceci pourrait être expliqué par le fait que l'effet d'association est apporté exclusivement par le génotype homozygote TT indépendamment de l'allèle T et de sa fréquence. Une autre étude rapporte une absence d'association avec la PR dans 100 familles trio caucasiennes Françaises pour ce polymorphisme ( $P = 0,76$ ) (258).

De la même manière que pour IRF5<sub>rs2004640</sub>, TRAF1/C5<sub>rs10818488</sub> a montré une association significative au niveau génotypique ( $P_{\text{génotypique}} = 2,9 \times 10^{-3}$ ) mais n'a pas atteint la signification statistique lorsque l'on compare les fréquences alléliques ( $P_{\text{allélique}} = 0,21$ ). Encore une fois, cela pourrait s'expliquer par la conception de l'étude ou alors représenter un vrai caractère ethnique de la population de l'Ouest algérien. Deux études menées respectivement par Zervou et al. (250) et Kurreeman et al. (161) ont signalé une association entre cette région et la PR dans les études d'association

génétiq ue à grande échelle, tandis qu'une étude sur l'ensemble du génome (GWAS) réalisée par le *Wellcome Trust Consortium* n'a pas réussi à identifier cette région en tant que région candidate pour la PR (50).

Enfin, l'étude cas-témoins présente certaines limites car elle ne peut pas éviter les biais existants entre les cas et les contrôles, et l'association peut être par conséquent sous ou surévaluée. Les études familiales de type trio ou encore les études de paires de germains (*sib-pairs*) sont connues pour être plus robustes, en évitant un appariement imparfait entre témoins et malades et en appliquant directement la loi de Mendel par le test de déséquilibre de transmission TDT.

Cependant, la présente étude est la première démonstration claire, dans la population de l'Ouest Algérien, concernant l'association de gènes non-HLA avec la susceptibilité à la PR. Il est utile de mener des études complémentaires basées sur la comparaison des résultats obtenus avec d'autres régions algériennes voire d'autres populations proches ethniquement (Maroc, Lybie...). Aussi, il serait intéressant de continuer l'étude en agrandissant l'échantillonnage à de grands groupes afin de confirmer ces résultats et d'analyser d'autres facteurs génétiques dans la population de l'Ouest Algérien, à commencer par les polymorphismes du locus HLA.

# ARTICLE 1

## **Association of PTPN22 (rs2476601) and STAT4 (rs7574865) polymorphisms with Rheumatoid Arthritis in the Western Algerian population**

Mostefa FODIL<sup>1,2</sup>, Ahmed BENZAOUÏ<sup>3</sup>, Faouzia ZEMANI-FODIL<sup>1</sup>, Meriem ABERKANE<sup>1</sup>, Wefa BOUGHRARA<sup>1</sup>, Nadhira SAIDI-MEHTAR<sup>1</sup>, Elisabeth PETIT-TEIXEIRA<sup>2</sup>, Abdallah BOUDJEMA<sup>1</sup>.

1. Laboratoire de Génétique Moléculaire et Cellulaire, Université des Sciences et de la Technologie d'Oran - Mohamed Boudiaf (USTO-MB), Algérie.
2. GenHotel-EA3886, Evry-Val d'Essonne University, Evry-Genopole, France.
3. Service de Rhumatologie, CHUO (Oran), Algérie.

## ABSTRACT

**Aim:** The aim of the present study was to replicate the association of five risk gene polymorphisms (PTPN22<sub>rs2476601</sub>, STAT4<sub>rs7574865</sub>, 6q23<sub>rs6927172</sub>, IRF5<sub>rs2004640</sub> and TRAF1/C5<sub>rs10818488</sub>) with RA in a specific population of the Western Algeria.

**Material and methods:** The study group comprised 110 patients with RA and 197 ethnically matched healthy control subjects. All polymorphisms were genotyped using predesigned TaqMan® assays. Allele and genotype frequencies in patients and control subjects were compared by chi-square test and odds ratios with 95% confidence intervals. Correction for multiple testing was carried out using the Bonferroni adjustment.

**Results:** Statistically significant associations with RA were detected. The strongest signal was obtained for PTPN22<sub>rs2476601</sub> with an allelic  $P_{value} = 3.32 \times 10^{-11}$  (OR = 9.83, 95% CI [4.28 – 22.56]). A second significant association was obtained with STAT4<sub>rs7574865</sub> (allelic  $P_{value} = 4 \times 10^{-3}$ ; OR = 1.75, 95% CI [1.16 – 2.63]). The third SNP, 6q23<sub>rs6927172</sub>, showed a significant result of association with RA, but missed our criteria for significance at allelic level after Bonferroni's correction (allelic  $P_{value} = 0.027$ ; OR = 0.64, 95% CI [0.42 – 0.97]). Finally, IRF5<sub>rs2004640</sub> and TRAF1/C5<sub>rs10818488</sub> showed a significant association only at genotypic level ( $P_{values} = 3 \times 10^{-4}$  and  $2.9 \times 10^{-3}$  respectively) but did not reach statistical significance when comparing allele frequencies ( $P_{values} = 0.96$  and  $0.21$  respectively).

**Conclusions:** From this initial study, we can conclude that PTPN22<sub>rs2476601</sub> and STAT4<sub>rs7574865</sub> polymorphisms are clearly associated with the risk of RA in the Western Algerian population.

**Key words:** Rheumatoid Arthritis, genetic association, gene polymorphisms, PTPN22, STAT4, Algerians.

## INTRODUCTION

Rheumatoid arthritis (RA) is a complex autoimmune disease characterized by altered inflammatory and impaired immune responses causing immune-mediated destruction of tissues and organs. Although the aetiology of this autoimmune disease is not completely known, it is nowadays clearly demonstrated that both genetic and environmental factors contribute to the development of RA.

Previous genetic and family studies indicated a significant genetic contribution to RA development<sup>1</sup>. Genetic association of HLA-DRB1 and RA has been well established in multiple ethnic groups. However, the HLA region contributes only 30–50% of genetic component for RA<sup>2</sup>.

Advances in genotyping technology have facilitated the application of whole genome association approaches to identify disease causal variants. Recent and several genome-wide and candidate-gene studies reported additional evidence for association of single nucleotide

polymorphisms (SNPs) markers in several candidate genes and/or loci such *PTPN22*, *STAT4*, *OLIG3/TNFAIP3* (6q23), *IRF5* and *TRAF1/C5* loci with RA <sup>3,4</sup>.

*All these genes are involved in the immune response and/or its regulation.*

*A non-synonymous single nucleotide polymorphism (SNP) in the gene encoding protein tyrosine phosphatase non-receptor 22 (PTPN22) is considered as the second risk factor of susceptibility to RA after HLA <sup>5</sup>. The STAT4 locus has been identified as a confirmed RA susceptibility locus in UK, Korean, Swedish, US, Greek, Colombian, Spanish and US populations <sup>6-11</sup>. A locus, mapped between OLIG3 and TNFAIP3 on chromosome 6q was identified in a genome-wide association study (GWAS) of seven common diseases, including RA, carried out by the WTCCC (The Wellcome Trust Case Control Consortium) <sup>4</sup>. Association with this locus 6q23 has been replicated in populations from the UK and USA <sup>12, 13</sup>. The recent finding of an association between the IRF5 gene and systemic lupus erythematosus SLE <sup>14</sup>, which has been replicated in multiple populations, as well as the association between IRF5 and RA <sup>15</sup> and inflammatory bowel diseases (IBD) <sup>16</sup>, provides additional support for the important role of IRF5 and the type I Interferon (IFN) system in autoimmune diseases. Finally, a GWAS in American and Swedish populations identified a novel locus mapping between TRAF1 and C5 loci associated with RA <sup>17</sup>. This association has been replicated in samples from Greece, Netherlands and North American populations <sup>17, 18</sup>.*

*This study was initiated to replicate, in a specific population of the Western Algeria, the association between these five defined genetic markers and RA. To our knowledge, this is the first study to be realized in the Western Algerian population in order to find a genetic association between these polymorphisms and RA.*

## **PATIENTS AND METHODS**

**Subjects.** *The study group comprised 110 RA patients from unrelated families living in Oran region (Western Algeria). A control group comprising 197 ethnically matched healthy volunteers have been included in the study.*

*All RA patients met the American Rheumatism Association 1987 revised criteria) <sup>19</sup> and were recruited from department of Rheumatology of Oran hospital "CHUO" (Centre Hospitalo-Universitaire d'Oran). Ethnic bias within the population studied was minimized by excluding patients that were not of western Algerian origin. Western Algerian origin was defined as having the four grandparents of each individual possessing a western Algerian ancestry. The control group was recruited on the same basis of ethnic and geographic characteristics. Healthy volunteers were recruited through a local media campaign followed by selection of individuals who fulfilled our criteria: having a Western Algerian origin, absence of consanguinity in parents and absence of any autoimmune disease for the volunteers or their parents.*



Informed consent was obtained from all subjects and the study was conducted according to the declaration of Helsinki Principles, and the ethics committee of Centre Hospitalo-Universitaire d'Oran (CHUO, Oran, Algeria) approved the study.

**DNA extraction.** Peripheral Blood was collected into 5 ml tubes containing EDTA and stored at  $-20^{\circ}\text{C}$  until analysis. Genomic DNA was purified from fresh peripheral blood leukocytes by standard DNA extraction methods<sup>20</sup>.

**Molecular genotyping.** Genes and SNPs choice was based on previous published data on RA association from genome-wide and candidate-gene studies<sup>3, 4</sup> (table 1). 6q23<sub>rs6927172</sub> was chosen because it was in Linkage Disequilibrium (LD) in the Caucasian population with 6q23<sub>rs6920220</sub>, another SNP in the same genomic region which was confirmed as a risk factor for RA<sup>4, 13</sup>.

Genotyping of the five gene polymorphisms was carried out with a Taqman® 5' allelic discrimination assay (table 1). It was performed on an ABI-7500 real-time polymerase chain reaction (PCR) machine (Applied Biosystems, now Life Technologies, Foster City, CA, USA) according to the manufacturer's protocol. Genotyping of each sample was automatically attributed using ABI PRISM® 7500 Sequence Detection System software "SDS" for allelic discrimination.

**Quality control.** For quality control, positive and negative controls (from CEPH "Centre d'Étude du polymorphisme Humain") were co-genotyped in each genotyping assay and additional 10% of randomly selected duplicates were included. No discrepancy between duplicates was observed in the genotyping data of the five selected SNPs. All control tests had greater than 99% genotype passing rates.

**Statistical analyses.** Prior to association tests, we checked the Hardy–Weinberg equilibrium in the control group using a standard chi-square test. Results from the control subjects and RA unrelated patients were compared using the chi-square test (2x2 contingency tables) for statistical significance.

The genotype relative risk (GRR) method (a single genotype vs. the others) was used to compare the genotype distribution in controls and patients. The GRR test adjusts the genotype frequencies in the controls to the expected Hardy–Weinberg proportions and yields more accurate risk estimates<sup>21</sup>.

$P_{\text{values}} \leq 0.05$  were considered statistically significant. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated with Woolf's method. Odds ratio was calculated considering the risk allele of each SNP or considering genotypes containing the risk alleles. Correction for multiple testing was carried out using the Bonferroni adjustment. The significance of the p value was assessed at 0.01 (0.05/5 considering 5 SNPs tested).

## RESULTS

Demographic and clinical characteristics of all subjects are summarized in table 2. The observed sex ratio of 9 females for 1 male seems to be an ethnic characteristic of the Algerian population as it has been reported in RA cohorts from other Algerian regions (data not published). There is a difference of sex ratio between RA and controls groups but it is not statistically significant.

We analysed five gene polymorphisms (*PTPN22*<sub>rs2476601</sub>, *STAT4*<sub>rs7574865</sub>, *TNFAIP3*<sub>rs6927172</sub>, *IRF5*<sub>rs2004640</sub> and *TRAF1/C5*<sub>rs10818488</sub>) with a prior evidence of association with RA in numerous genome-wide association studies and candidate-gene approaches<sup>3,4</sup>. For all five SNP polymorphisms no deviation from the Hardy-Weinberg equilibrium was observed in the control group (data not shown). We observed a clear significant association between two SNPs (*PTPN22*<sub>rs2476601</sub> and *STAT4*<sub>rs7574865</sub>) and RA in our population, since the other three SNPs (*6q23*<sub>rs6927172</sub>, *IRF5*<sub>rs2004640</sub> and *TRAF1/C5*<sub>rs10818488</sub>) did not reach statistical significance at allelic level (table 3).

The strongest result was obtained with *PTPN22*<sub>rs2476601</sub> with an allelic  $P_{value} = 3.32 \times 10^{-11}$  (OR = 9.83, 95% CI [4.28 – 22.56]) (table 3). Moreover, the presence of at least one copy of *T*<sub>rs2476601</sub> allele was significantly different between RA patients and healthy controls ( $P_{value} = 8.36 \times 10^{-9}$ ; OR = 12.08 and 95% CI [4.86 - 33.42]).

Another significant association was obtained with *STAT4*<sub>rs7574865</sub> (allelic  $P_{value} = 4 \times 10^{-3}$ ; OR = 1.75, 95% CI [1.16 – 2.63]). The presence of at least one copy of *T*<sub>rs7574865</sub> allele showed a significant difference between RA patients and healthy subjects ( $P_{value} = 8 \times 10^{-4}$ ; OR = 2.23 and 95% CI [1.35 – 3.71]).

The third SNP, *6q23*<sub>rs6927172</sub>, showed a significant association (allelic- $P_{value} = 0.027$ ; OR = 0.64, 95% CI [0.42 – 0.97]), but missed our criteria for significance after Bonferroni's correction. However, at the genotypic level, the  $P_{value}$  for *6q23*<sub>rs6927172</sub> was still significant after Bonferroni's correction (genotypic- $P_{value} = 3 \times 10^{-3}$ ) (table 3).

Finally, *IRF5*<sub>rs2004640</sub> and *TRAF1/C5*<sub>rs10818488</sub> showed a significant association only at genotypic level ( $P_{values}: 3 \times 10^{-4}$  and  $2.9 \times 10^{-3}$  respectively) but did not reach significance when comparing allele frequencies ( $P_{values}: 0.96$  and  $0.21$  respectively).

## DISCUSSION

The aim of the present study was to replicate, among the specific Western Algerian population, the association of these five risk gene polymorphisms with RA.

A clear significant association was obtained between RA and two SNPs (*PTPN22*<sub>rs2476601</sub> and *STAT4*<sub>rs7574865</sub>) in our population, since the other three SNPs (*6q23*<sub>rs6927172</sub>, *IRF5*<sub>rs2004640</sub> and *TRAF1/C5*<sub>rs10818488</sub>) did not reach statistical significance (table 3). This preliminary study in the Western Algerian population is in accordance with previous studies reporting the same findings in several populations.

The strongest association with RA was observed with *PTPN22*<sub>rs2476601</sub> polymorphism. The first study of the association of *PTPN22*<sub>rs2476601</sub> with RA was conducted in 2004<sup>5</sup>. Since, numerous studies confirmed this association in different populations including French<sup>22</sup> UK<sup>23</sup>, Finnish<sup>24</sup>, Swedish<sup>25</sup>, German<sup>26</sup>, Dutch<sup>27</sup>, Spanish<sup>28</sup> and Canadian<sup>29</sup> populations. However, a study in a Japanese population could not confirm for association as the causal variant was found to have a very low minor allele frequency<sup>30</sup>.

Two other studies in Tunisians, a population with a close ethnic origin to Algerians, reported controversial results about the association of *PTPN22*<sub>rs2476601</sub> polymorphism with RA. Indeed, Chabchoub et al.<sup>31</sup> reported a lack of association while Sfar et al.<sup>32</sup> confirmed the association with RA, through two independent case-control studies from distinct regions in Tunisia.

The second most important association was detected with *STAT4*<sub>rs7574865</sub> (allelic  $P_{\text{value}} = 4 \times 10^{-3}$ ; OR = 1.75, 95% CI [1.16 – 2.63]). Moreover, the presence of at least one copy of *T*<sub>rs7574865</sub> allele showed a significant difference between RA patients and healthy subjects ( $P_{\text{value}} = 8 \times 10^{-4}$ ; OR = 2.23 and 95% CI [1.35 – 3.71]). Our results are in accordance with all previous studies on association of *STAT4*<sub>rs7574865</sub> with RA in different populations<sup>6–11</sup>.

*6q23*<sub>rs6927172</sub> showed a trend of association with RA in our study. However, it did not reach statistical significance after multiple test correction when comparing allelic frequencies. This could be caused by our study design (case-control and the relatively reduced sample size) or represent a true ethnic characteristic of the Western Algerian population. This observation is not very different from a study from another Mediterranean population (Spanish) which reported a weak association between SNPs located in this genomic region and RA<sup>33</sup>. Indeed, Dieguez-Gonzalez et al. reported a weak evidence of association both in the *6q23* intergenic region and in the *TNFAIP3* locus. Overall, statistical association was best explained by the interdependent contribution of SNPs from the two loci *TNFAIP3* and the *6q23* intergenic region. Another study in Tunisians reported a trend of association between *rs6920220* and RA<sup>34</sup>. It is now clear that *6q23* locus is a demonstrated candidate for a role in RA. However, our study only provides a suggestive evidence of the involvement of this locus in RA susceptibility.

Concerning *IRF5*<sub>rs2004640</sub> a significant association was obtained only at genotypic level ( $P_{\text{values}} = 3 \times 10^{-4}$ ) but did not reach statistical significance when comparing allele frequencies ( $P_{\text{value}} = 0.96$ ). A similar result was obtained in the Tunisian population by Maalej et al.<sup>35</sup>; in this study, no significant difference was found at the allelic level between RA patients and Healthy controls. However, the genotype *TT* was more frequent in patients with RA (42.1%) than in the healthy controls (31.4%) ( $P_{\text{value}} = 0.01$ ). Another study reported no RA linkage in 100 French Caucasian trio families; there was no over-transmission of the *IRF5* *rs2004640-T* allele from heterozygotic parents to affected patients ( $P_{\text{value}} = 0.76$ )<sup>36</sup>. This could be explained by the fact that the association effect is brought by the homozygote genotype *TT* independently from the allele frequency.

Similarly to *IRF5*<sub>rs2004640</sub>, *TRAF1/C5*<sub>rs10818488</sub> showed a significant association at genotypic level ( $P_{\text{values}} = 2.9 \times 10^{-3}$ ) but did not reach statistical significance when comparing allele frequencies ( $P_{\text{value}} = 0.21$ ). Again, this could be explained, by the study design or represent a

true ethnic characteristic of the Western Algerian population. Zervou et al. and Kurreeman et al.<sup>6, 18</sup> reported an association between this region and RA in large-scale genetic association studies, while a genome-wide study performed by the Wellcome Trust failed to identify this region as a candidate for RA<sup>4</sup>.

Finally, the case-control approach cannot avoid imperfect matching between cases and controls, and the association may be consequently under- or overestimated. The family-based studies are known to be more robust, avoiding this imperfect matching and testing directly Mendel's law by the transmission disequilibrium test.

The current report is the first clear demonstration of non-MHC related susceptibility gene for RA that confers a risk among the Western Algerian population. It is worthwhile to conduct additional population-based studies, including larger groups to confirm these findings and to analyze other genetic factors in the Algerian population.

#### ACKNOWLEDGEMENTS

Mostefa FODIL's thesis work was supported by a doctoral scholarship from A.U.F. « Agence Universitaire de la Francophonie » (grant n° 79113FR033L) and additional funding from FDIR-UEVE « Fond de Développement pour l'Internationalisation de la Recherche, Université d'Evry Val d'Essonne » (appel d'offre 2011/ 2012), and from PNR-USTO-MB « Projet National de Recherche, Université des Sciences et de la Technologie d'Oran – Mohamed Boudiaf » (appel d'offre 2011/2012).

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

#### REFERENCES

1. Wandstrat A and Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nat Immunol* 2001; 2:802–809. doi:10.1038/ni0901-802.
2. Bowes J and Barton A. Recent advances in the genetics of RA susceptibility. *Rheumatology (Oxford)* 2008; 47:399–402.
3. Stahl EA, Raychaudhuri S, Remmers EF, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010; 42:508-514.
4. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447:661–678.
5. Begovich AB, Carlton VEH, Honigberg LA, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004; 75:330–337.

6. Zervou MI, Sidiropoulos P, Petraki E, et al. Association of a TRAF1 and a STAT4 gene polymorphism with increased risk for rheumatoid arthritis in a genetically homogeneous population. *Hum Immunol* 2008; 69: 567–571.
7. Barton A, Thomson W, Ke X, et al. Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Hum Mol Genet* 2008; 17: 2274–2279.
8. Lee HS, Remmers EF, Le JM, Kastner DL, Bae S-C, Gregersen PK. Association of STAT4 with rheumatoid arthritis in a Korean population. *Mol Med* 2007; 13: 455–460.
9. Palomino-Morales RJ, Rojas-Villarraga A, Gonzalez CI, Ramirez G, Anaya JM, Martin J. STAT4 but not TRAF1/C5 variants influence the risk of developing rheumatoid arthritis and systemic lupus erythematosus in Colombians. *Genes Immun* 2008; 9:379–382.
10. Remmers EF, Plenge RM, Lee AT, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 2007; 357:977–986.
11. Orozco G, Alizadeh BZ, Delgado-Vega AM, et al. Association of STAT4 with rheumatoid arthritis: a replication study in three European populations. *Arthritis Rheum* 2008; 58:1974–1980.
12. Plenge RM, Cotsapas C, Davies L, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007; 39:1477–1482.
13. Thomson W, Barton A, Ke X, et al. Rheumatoid arthritis association at 6q23. *Nat Genet* 2007; 39:1431–1433.
14. Sigurdsson S, Nordmark G, Goring HH, et al. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am J Hum Genet* 2005; 7:528–537.
15. Sigurdsson S, Padyukov L, Kurreeman FA, et al. Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. *Arthritis Rheum* 2007; 56: 2202–2210.
16. Dideberg V, Kristjansdottir G, Milani L, et al. An insertion-deletion polymorphism in the interferon regulatory factor 5 (IRF5) gene confers risk of inflammatory bowel diseases. *Hum Mol Genet* 2007; 16: 3008–3016.
17. Plenge RM, Seielstad M, Padyukov L, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis – a genome wide study. *N Engl J Med* 2007; 357:1199–1209.
18. Kurreeman FA, Padyukov L, Marques RB, et al. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med* 2007; 4:e278.
19. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; 31:315.
20. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988; 16:1215.
21. Lathrop GM: Estimating genotype relative risks. *Tissue Antigens* 1983; 22: 160-166.
22. Michou L, Lasbleiz S, Rat AC, et al. Linkage proof for PTPN22, a rheumatoid arthritis susceptibility gene and a human autoimmunity gene. *Proc Natl Acad Sci USA* 2007; 104:1649–1654.

23. Hinks A, Barton A, John S, et al. Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: further support that PTPN22 is an autoimmunity gene. *Arthritis Rheum* 2005; 52:1694–1699.
24. Seldin MF, Shigeta R, Laiho K, et al. Finnish case-control and family studies support PTPN22 R620W polymorphism as a risk factor in rheumatoid arthritis, but suggest only minimal or no effect in juvenile idiopathic arthritis. *Genes Immun* 2005; 6: 720–722.
25. Plenge RM, Padyukov L, Remmers EF, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. *Am J Hum Genet* 2005; 77:1044–1060.
26. Pierer M, Kaltenhauser S, Arnold S, et al. Association of PTPN22 1858 single nucleotide polymorphism with rheumatoid arthritis in a German cohort: higher frequency of the risk allele in male compared to female patients. *Arthritis Res Ther* 2006; 8:R75.
27. Zhernakova A, Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP. Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. *Genes Immun* 2005; 6:459–461.
28. Orozco G, Sanchez E, Gonzalez-Gay MA, et al. Association of a functional single nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Rheum* 2005; 52:219–224.
29. van Oene M, Wintle RF, Liu X, et al. Association of the lymphoid tyrosine phosphatase R620W variant with rheumatoid arthritis, but not Crohn's disease, in Canadian populations. *Arthritis Rheum* 2005; 52: 1993–1998.
30. Ikari K, Momohara S, Inoue E, et al. Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. *Rheumatology* 2006; 45:1345–1348.
31. Chabchoub G, Teixeira EP, Maalej A, et al. The R620W polymorphism of the protein tyrosine phosphatase 22 gene in autoimmune thyroid diseases and rheumatoid arthritis in the Tunisian population. *Ann Hum Biol* 2009; 36:342–349.
32. Sfar I, Dhaouadi T, Habibi I, et al. Functional polymorphisms of PTPN22 and FCGR genes in Tunisian patients with rheumatoid arthritis. *Arch Inst Pasteur Tunis* 2009; 86:51-62.
33. Dieguez-Gonzalez R, Calaza M, Perez-Pampin E, et al. Analysis of TNFAIP3, a feedback inhibitor of nuclear factor-kappaB and the neighbour intergenic 6q23 region in rheumatoid arthritis susceptibility. *Arthritis Res Ther* 2009; 2:R42. doi: 10.1186/ar2650.
34. Ben Hamad M, Cornelis F, Maalej A, Petit-Teixeira E. A Tunisian case-control association study of a 6q polymorphism in rheumatoid arthritis. *Rheumatol Int.* 2012; 32:1849-1850. doi: 10.1007/s00296-011-1996-6.
35. Maalej A, Hamad MB, Rebaï A, et al. Association of IRF5 gene polymorphisms with rheumatoid arthritis in a Tunisian population. *Scand J Rheumatol.* 2008; 37:414-418.
36. Garnier S, Dieudé P, Michou L, et al. IRF5 rs2004640-T allele, the new genetic factor for systemic lupus erythematosus, is not associated with rheumatoid arthritis. *Ann Rheum Dis* 2007; 66: 828–831.

**Table 1: Population characteristics**

Characteristics	RA Cases (%)	Healthy Controls (%)	<i>P value</i>
Number of Subjects	110	197	-
Men	11 (10%)	89 (45.18%)	NS
Women	99 (90%)	108 (54.82%)	NS
Age (years)	48 ± 30.5	40 ± 22	NS
Rheumatoid Factor +	31 (32%)	ND	-
Rheumatoid Factor –	66 (68%)	ND	-

Values are presented as the mean ± SD or number.

RA : Rheumatoid Arthritis

ND: Not Done

NS : Not significant

**Table 2: SNP markers used in analysis**

Gene/Locus	db SNP ID	Assay ID	SNP	Location (NCBI Build)	Gene/Function
PTPN22	rs2476601	C__16021387_20	A/G	Chr.1: 114377568	Missense mutation
STAT4	rs7574865	C__29882391_10	G/T	Chr.2:191964633	Intron
6q23 (TNFAIP3/OLIG3)	rs6927172	C___1575580_10	C/G	Chr.6: 138002175	Intergenic
IRF5	rs2004640	C___9491614_10	G/T	Chr.7: 128578301	Intron
TRAF1/C5	rs10818488	C___2783655_10	A/G	Chr.9: 123705087	Intergenic

db: data base ; SNP: single nucleotide polymorphism ; Chr: chromosome ; NCBI:

National Center for Biotechnology Information

Table 3: SNP association analysis in RA-HS case-control sample

Gene / SNP	Genotype/Allele	RA Patients n=110 (%)	Controls n=197 (%)	X <sup>2</sup>	P value	OR (95% CI)
PTPN22 rs2476601	TT	1 (1)	0 (0)	45.5	<b>1.31 x 10<sup>-10</sup></b>	9.83 (4.28 – 22.56)
	TC	33 (30)	7 (4)			
	CC	76 (69)	189 (96)			
	T	35 (16)	7 (2)	43.97	<b>3.32 x 10<sup>-11</sup></b>	
	C	185 (84)	385 (98)			
TT + TC	34 (31)	7 (4)	45.38	<b>8.36 x 10<sup>-9</sup></b>		
	CC	76 (69)	189 (96)			
STAT4 rs7574865	GG	53 (48)	133 (68)	11.65	<b>2.9 x 10<sup>-3</sup></b>	1.75 (1.16 – 2.63)
	GT	53 (48)	57 (29)			
	TT	4 (4)	7 (4)			
	T	61 (28)	71 (18)	7.88	<b>4 x 10<sup>-3</sup></b>	
	G	159 (72)	323 (82)			
GT + TT	57 (52)	64 (32)	11.05	<b>8 x 10<sup>-4</sup></b>		
	GG	53 (48)	133 (68)			
6q23 rs6927172	CC	58 (53)	135 (70)	11.39	<b>3 x 10<sup>-3</sup></b>	0.64 (0.42 – 0.97)
	CG	49 (45)	50 (26)			
	GG	3 (2)	9 (4)			
	G	55 (25)	68 (18)	4.86	<b>0.027 *</b>	
	C	165 (75)	320 (82)			
CG + GG	52 (47)	59 (30)	8.61	<b>3 x 10<sup>-3</sup></b>		
	CC	58 (53)	135 (70)			
IRF5 rs2004640	GG	9 (8)	38 (19)	15.84	<b>3 x 10<sup>-4</sup></b>	1.01 (0.72 – 1.4)
	GT	80 (73)	97 (50)			
	TT	21 (19)	60 (31)			
	T	122 (55)	217 (56)	0.002	0.96	
	G	98 (45)	173 (44)			
GT + TT	101 (92)	157 (81)	6.9	<b>8.6 x 10<sup>-3</sup></b>		
	GG	9 (8)	38 (19)			
TRAF1/C5 rs10818488	AA	7 (6)	38 (19)	11.63	<b>2.9 x 10<sup>-3</sup></b>	0.81 (0.57 – 1.13)
	AG	69 (63)	92 (47)			
	GG	34 (31)	66 (34)			
	A	83 (38)	168 (43)	1.53	0.21	
	G	137 (62)	224 (57)			
AA + AG	76 (69)	130 (66)	0.24	0.62		
	GG	34 (31)	66 (34)			

\* Pvalue not significant after Bonferroni's correction (alpha=0.01)

RA : rheumatoid arthritis ; OR : Odds Ratio ; SNP: single nucleotide polymorphism



**II. EXPLOITATION GENETIQUE  
DES DONNEES DU  
TRANSCRIPTOME DANS LA PR  
DANS DES FAMILLES  
EUROPEENNES**

## 1 Introduction

L'hétérogénéité de la polyarthrite rhumatoïde résulte de la complexité des mécanismes immunitaires et de la multiplicité des éléments qui interviennent dans le processus de la maladie, y compris un rôle crucial de la signalisation des cytokines et des molécules qui contribuent à la persistance des altérations physiopathologiques (259, 260).

L'étude de profils d'expression des gènes peut permettre d'identifier les mécanismes moléculaires impliqués dans la pathologie. Plusieurs études de profil d'expression de gènes ont identifié une expression remarquablement élevée d'un nombre considérable de gènes chez les patients atteints de PR par rapport aux témoins. Dans une étude récente réalisée par Teixeira et al. au laboratoire GenHotel-EA3886 sur les cellules mononucléaires du sang périphérique (177), une liste a été mise en évidence comprenant 339 gènes exprimés de manière différentielle au cours de la PR (238 gènes sous-régulés et 101 sur-régulés). La plupart des gènes nouvellement découverts lors de cette étude ont été impliqués essentiellement dans les différents mécanismes liés à l'immunité comme l'activité pro-inflammatoire, l'activité antimicrobienne, le stress cellulaire ou encore les fonctions immuno-modulatrices (211, 219, 261).

De récentes études d'association pangénomique (GWAS) ont clairement identifié de nombreux nouveaux marqueurs génétiques pour la PR (51, 251). Toutefois, ce qui est moins bien démontré c'est le rôle de ces marqueurs de susceptibilité dans la détermination de l'apparition de la maladie et sa persistance.

A la lumière de tous ces éléments, le but de notre étude a été d'abord d'analyser les gènes différentiellement exprimés en ce qui concerne la relation entre le niveau de données d'expression et de génotypage. Parmi les gènes montrant le plus fort facteur de variation d'expression (*Fold Change*) entre les patients PR et les contrôles sains dans notre précédente étude de profil d'expression (177), six gènes ont été choisis car aucune variation du

nombre de copies (CNV) n'y avait été décrite. Ces gènes sont S100A8, RNASE2, PGLYRP1, RUNX3, IL2RB et LY96.

Les polymorphismes nucléotidiques (SNPs) de ces gènes ont été choisis parmi ceux ayant une fréquence d'allèles mineurs supérieure à 5%, tout en tenant compte des données disponibles de déséquilibre de liaison de la région des gènes. Les polymorphismes ont été étudiés par un test d'association familial dans les familles trio. La stratification des familles trio a ensuite été effectuée concernant la présence chez les patients atteints de PR d'au moins un des allèles avec l'épitope partagé (*Shared Epitope* ou *SE*) du gène HLA-DRB1 (238), des anticorps anti-peptides citrullinés (ACPA) ainsi que suivant le statut d'érosion osseuse.

## 2 Résultats

L'exactitude des données de génotypage a été de 100%. Cette valeur est obtenue par le contrôle de la qualité des données qui consiste, comme nous l'avons vu précédemment, à refaire un génotypage pour des échantillons de référence ainsi que pour 10% des échantillons choisis de manière aléatoire.

Pour les tests d'association, tous les SNP étaient à l'équilibre de Hardy-Weinberg dans le groupe de témoins de l'**échantillon 1**. Dans les **échantillons 2 et 3** l'équilibre de Hardy-Weinberg était aussi respecté dans le groupe témoin, constitué par les chromosomes (allèles) parentaux non transmis par chacun des parents dans les familles trio.

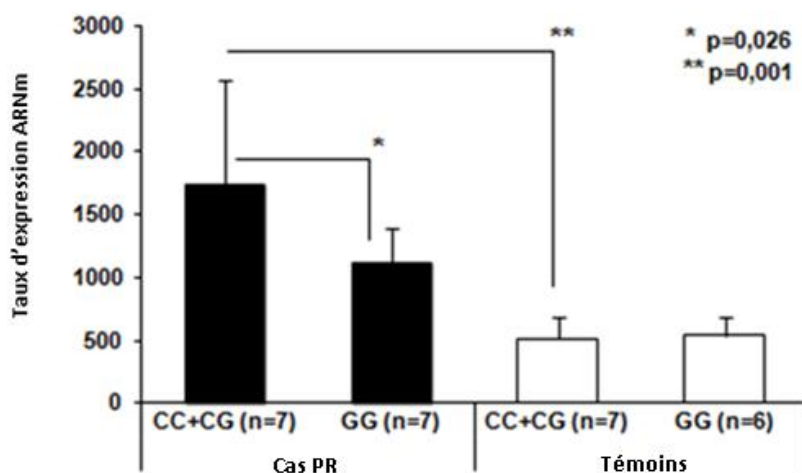
### 2.1 Corrélation taux d'expression/génotypes

Tout d'abord, les niveaux d'expression des gènes étudiés obtenus à partir de données de l'étude réalisée par Teixeira et al. (177) ont été comparés avec les génotypes respectifs pour chaque SNP dans

l'échantillon 1 constitué de 14 cas PR et de 13 témoins sains. Cette comparaison a été évaluée par des tests non paramétriques (*Kruskal-Wallis* et *Mann-Whitney*) en utilisant le logiciel XLSTAT (*Addinsoft, Paris, France, version 2012 [www.xlstat.com](http://www.xlstat.com)*).

Préalablement, nous avons vérifié la liste des eQTL disponibles pour chaque locus étudié en utilisant la base de données <http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl>. Aucun gène de notre étude n'a été concerné, à l'exception du gène LY96 pour qui l'existence d'un eQTL en *cis* a été décrite et ce concernant deux SNPs (rs10504554 et rs7822407) (262). Ces derniers n'étaient pas en déséquilibre de liaison avec le polymorphisme que nous avons testé au cours de notre étude (rs1991262), et ne constituaient, de ce fait, aucun biais pour l'analyse de nos résultats.

Un seul résultat positif a été obtenu pour le gène RNASE2, lorsque le taux d'expression de ce dernier a été étudié en tenant compte des génotypes du polymorphisme rs2013109G/C. Le niveau moyen d'expression du gène RNASE2 chez les patients PR dont les génotypes étaient (CC et GC) était significativement plus élevé que le taux d'expression chez les patients avec le génotype GG ( $P = 0,026$ ; **figure 12**). L'allèle rs2013109-C du gène RNASE2 est alors associé à une expression élevée du gène chez les patients PR.



**Figure 12 : Association entre le taux d'expression et les génotypes du polymorphisme RNASE2<sub>rs2013109</sub>.**  
RA : Rheumatoid arthritis patients ; HS : Healthy Subjects

## 2.2 Etude d'association génétique des 10 SNPs étudiés avec la PR

Par la suite nous avons réalisé une étude d'association génétique pour les 10 polymorphismes étudiés et la PR.

Comme indiqué dans notre protocole d'étude, chaque SNP devait être d'abord testé pour l'association dans l'**échantillon 2** (de découverte) composé de 200 familles trio caucasiennes Françaises. Si un résultat significatif était obtenu, nous procédions à une réplification dans l'**échantillon 3**.

Le test d'association en milieu familial, réalisé par le logiciel FBAT, a montré qu'il y avait une importante sur-transmission pour l'allèle T du polymorphisme PGLYRP1<sub>rs2041992</sub> dans l'**échantillon 2** ( $P = 0,019$ , **tableau 11**). Aucun autre résultat significatif n'a été observé pour les autres SNPs.

Nous avons alors effectué une stratification en sous-groupes dans l'**échantillon 2** selon la présence (SE+) ou absence (SE-) des allèles avec l'épitope partagé (ou *shared epitope*) de l'antigène HLA-DRB1. Nous avons pu observer l'existence d'un déséquilibre de transmission important pour le polymorphisme PGLYRP1<sub>rs2041992</sub> dans le sous-groupe SE+ ( $P = 0,0032$ ; **tableau 12**). Après application de la correction de Bonferroni, cette valeur restait supérieure au seuil de signification, donc non significative. Le degré de significativité  $P$  fixé initialement à 0,05 pour chaque analyse individuelle d'association a été réévalué à  $7 \times 10^{-4}$  en appliquant la correction de Bonferroni pour des tests multiples, compte tenu des 10 SNPs testés et des 6 stratifications réalisées.

Tableau 11: Test d'association familiale pour les 10 SNPs étudiés dans l'échantillon 2

Gène	SNP	Allèle	Afreq	n-fam	P <sub>value</sub>
<b>S100A8</b>	rs3006488	G	0.120	31	0.75
		A	0.880		
<b>RNASE2</b>	rs2013109	C	0.280	63	0.12
		G	0.720		
<b>PGLYRP1</b>	rs2072561	G	0.046	16	0.47
		T	0.954		
	<b>rs2041992</b>	<b>T</b>	<b>0.337</b>	<b>134</b>	<b>0.019 *</b>
		<b>C</b>	<b>0.663</b>		
<b>RUNX3</b>	rs12082398	C	0.427	63	0.65
		T	0.595		
	rs1507101	G	0.405	134	0.70
		A	0.583		
	rs2236851	T	0.189	81	0.55
C		0.811			
rs2282718	A	0.385	65	0.65	
		G	0.615		
<b>IL2RB</b>	rs3218253	A	0.247	121	0.93
		G	0.753		
<b>LY96</b>	rs1991262	G	0.483	95	0.93
		C	0.517		

**afreq**: fréquence allélique ; **n-fam** : nombre de familles informatives

\* P value non significative après correction de Bonferroni tenant compte de 10 SNPs et 6 stratifications ( $P > 7 \times 10^{-4}$ )

Compte tenu de cette tendance pour l'association, nous avons alors décidé de répliquer l'analyse sur l'échantillon 3 qui se composait de 240 familles trio européennes en utilisant les mêmes méthodes. Le résultat d'association n'a pas été confirmé pour PGLYRP1<sub>rs2041992</sub> dans l'échantillon de réplification ( $P = 0,6$ , **tableau 12**).

**Table 12: Résultats de la réplication du test d'association familiale pour PGLYRP1<sub>rs2041992</sub>**

Echantillon	Allèle	afreq	n familles	P <sub>value</sub>
Echantillon 2	T	0.337	134	0.019 *
	C	0.663		
Echantillon 2 / SE+	T	0.344	107	0.0032 *
	C	0.656		
Echantillon 3	T	0.313	132	0.6
	C	0.687		
Echantillons 2+3	T	0.321	260	0.238
	C	0.679		

**afreq**: allele frequency ; **n familles**: nombre de famille informatives; **SE+** : index PR présentant au moins un allèle HLA-DRB1 codant l'épitope partagé; **SE-** : index PR ne présentant pas d'allèle HLA-DRB1 codant l'épitope partagé

\* P value non significative après correction de Bonferroni tenant compte de 10 SNPs et 6 stratifications ( $P > 7 \times 10^{-4}$ )

Par ailleurs, les résultats de génotypage pour le gène HLA-DRB1 n'étaient pas disponibles pour effectuer le test après stratification en fonction de l'épitope partagé pour l'**échantillon 3** de réplication.

Le test d'association global réalisé dans l'échantillon combiné (**échantillon 2 + 3**) composé de 440 familles trio n'a pas atteint la signification et n'a pas confirmé de ce fait l'association de PGLYRP1<sub>rs2041992</sub> de PR ( $P = 0,238$ , **tableau 12**).

Concernant la stratification selon le statut ACPA des malades ou encore selon la présence de l'érosion osseuse, aucun résultat significatif n'a été observé pour tous les SNPs étudiés.

Le test de risque relatif des génotypes (GRR) n'a pas montré d'association significative d'un génotype avec la maladie dans les 200 familles de l'échantillon 2 (tableau 13).

Tableau 13 : Analyse du risque relatif des génotypes (GRR) dans l'échantillon 2

Gène	SNP	Génotypes	PR	Témoins	P <sub>value</sub> 1/1 *	P <sub>value</sub> 2/2 **
<b>S100A8</b>	rs3006488	AA	79	77	0.70860	0.50161
		AG	19	21		
		GG	2	2		
<b>RNASE2</b>	rs2013109	CC	11	6	0.071896	0.19201
		CG	40	36		
		GG	48	57		
<b>PGLYRP1</b>	rs2072561	GG	41	48	0.3894	0.4799
		GT	45	42		
		TT	12	8		
	rs2041992	CC	1	0	1	0.4350
CT	5	10				
TT	93	89				
<b>RUNX3</b>	rs12082398	CC	19	14	0.4445	1
		CT	45	51		
		TT	32	31		
	rs1507101	AA	35	26	0.218	0.5527
AG		51	56			
GG		13	17			
rs2236851	CC	67	62	0,492386627	0,359771944	
	CT	31	34			
	TT	2	4			
	AA	16	10			0,603425759
AG	47	55				
GG	37	35				
<b>IL2RB</b>	rs3218253	AA	8	7	0,127303603	0,161935327
		AG	39	28		
		GG	53	65		
<b>LY96</b>	rs1991262	CC	66	62	0,481862293	0,941548009
		CG	28	31		
		GG	4	5		

\* 1/1 génotype homozygote pour l'allèle 1

\*\* 2/2 génotype homozygote pour l'allèle 2



Concernant l'analyse d'association haplotypique, pour les deux gènes disposant de plusieurs polymorphismes inclus dans notre étude (RUNX3 et PGLYRP1) aucun résultat significatif reproductible n'a été obtenu (**tableaux 14 et 15**). Un seul haplotype (C-G-T-A) du gène RUNX3 a montré une valeur de P significative (0.04) au seuil de 5% mais pas après la correction de Bonferroni (**tableau 14**).

**Tableau 14 : Analyse haplotypique dans l'échantillon 2 pour les 4 SNPs du gène RUNX3**

Haplo / RUNX3	All - SNP1	All - SNP2	All - SNP3	All - SNP4	Afreq	n-fam	P
a1	T	A	C	G	0.379	63	0.614294
a2	C	G	C	A	0.260	50	0.553375
a3	T	A	T	G	0.127	32	0.875615
a4	T	G	C	G	0.049	19	0.304596
a5	C	A	C	A	0.042	15	0.167386
a6	<b>C</b>	<b>G</b>	<b>T</b>	<b>A</b>	<b>0.042</b>	<b>15</b>	<b>0.043288 *</b>
a7	C	G	C	G	0.041	17	0.653115

**Tableau 15 : Analyse haplotypique dans l'échantillon 2 pour les 2 SNPs du gène PGLYRP1**

Haplo / PGLYRP1	All - SNP1	All - SNP2	Afreq	n-fam	P
a1	T	C	0.673	68	0.243616
a2	T	T	0.281	66	0.112924
a3	G	T	0.045	16	0.491298

**afreq**: fréquence allélique ; **n-fam** : nombre de familles informatives

\* P value non significative après correction de Bonferroni

### 3 Discussion

Cette étude a été conçue pour analyser la relation entre les génotypes des SNP choisis et le niveau d'expression des gènes leur correspondant chez des patients atteints de PR. Le choix des gènes à explorer a été fait sur la base des résultats de l'étude d'expression différentielle chez les cas PR par rapport aux témoins réalisée préalablement au laboratoire GenHotel-EA3886 par Teixeira et al. (177). La liste des gènes à explorer comprenait quatre gènes sur-exprimés (S100A8, PGLYRP1, RNASE2, LY96) et deux gènes sous-exprimés (RUNX3, IL2RB). Par la suite, nous avons testé l'hypothèse de l'association génétique de ces gènes avec la polyarthrite rhumatoïde, en ayant recours à une approche familiale (étude des familles trio). Le principal avantage des données des familiales, par rapport à l'approche cas/témoins, est qu'elles fournissent des contrôles « internes » parfaitement adaptés pour chaque patient étudié, sans introduction de biais allélique dû à la diversité génétique de la population générale. En effet, chaque allèle, transmis par un parent donné à l'enfant malade, est en fait parfaitement adapté pour la population d'origine avec l'allèle non transmis de chaque parent hétérozygote.

En ce qui concerne la relation entre les données de transcriptome et de génotypage, des études antérieures ont fourni des résultats intéressants dans le contexte de la PR (206, 263-267). Dans notre étude, seul le gène RNASE2 montré une association entre le niveau de son expression et les données de génotypes pour le polymorphisme intronique rs2013109, dans **l'échantillon 2**.

Le gène RNASE2 (également connu sous le nom « *Eosinophil Derived Neurotoxin* ou *EDN* ») est l'une des quatre principales protéines trouvées dans les grands granules spécifiques des éosinophiles humains. Il est également un chimio-attractant sélectif pour les cellules dendritiques en culture (206). Dans le contexte exclusivement auto-immun de la PR, le gène RNASE2 pourrait être, de par son implication dans les processus

immunitaires innés et notamment inflammatoires, un des facteurs les plus importants impliqués dans la physiopathologie de la PR. Nos résultats sont en accord avec cette hypothèse, d'autant plus que nous suggérons que l'allèle C<sub>rs2013109</sub> du gène RNASE2 serait associé à une expression élevée du gène dans le contexte de la PR. Une surexpression de ce gène entrainerait une abondance de son produit, conduisant ainsi à l'instauration et/ou le maintien du dérèglement des mécanismes de défense immunitaire menant à l'inflammation et la destruction des articulations.

Cette étude préliminaire doit être confirmée en utilisant des méthodes alternatives. Tout d'abord, comme les données d'expression utilisées dans cette étude ont été produites sur des PBMCs, la même approche devrait être effectuée en utilisant des résultats de transcriptomique obtenus à partir d'échantillons de tissus spécifiques, comme le tissu synovial. L'étude transcriptomique basée sur les PBMCs permet l'identification des processus pathogéniques principalement associés à la cause de la maladie. Cette approche peut présenter quelques limites, notamment en passant « à côté » de l'identification de certains gènes anormaux ou dont l'activité est altérée. Par la suite, et en bénéficiant de l'évolution actuelle des méthodes de séquençage de nouvelle génération, l'utilisation de l'approche RNAseq permettrait des études futures pouvant prendre en compte la diversité du transcriptome plus largement que ne le fait la technologie des puces d'expression.

En ce qui concerne les SNPs sélectionnés, nous nous sommes intéressés aux polymorphismes se trouvant à l'intérieur de la séquence du gène (intragénique), pour lesquels nous avons testé l'association éventuelle avec le niveau d'expression des gènes. L'association des SNP situés à l'extérieur de chaque gène (intergéniques) n'a pas été étudiée et pourrait être utile pour compléter nos résultats préliminaires.

Concernant l'étude d'association génétique, seul le polymorphisme PGLYRP1<sub>rs2041992</sub> a montré dans un premier temps une preuve d'association avec la polyarthrite rhumatoïde dans notre échantillon de découverte.

Néanmoins, cette association n'a pas été confirmée lors de l'utilisation de l'échantillon de réplication de 240 familles trio européennes. Les échantillons utilisés ne sont pas significativement différents compte tenu des données épidémiologiques et cliniques. Cependant, et en vue de tester l'hypothèse d'hétérogénéité génétique entre nos échantillons 2 et 3, nous avons effectué une analyse de l'association dans le sous-groupe français de cet échantillon de familles européennes. Nous n'avons pas répliqué l'association trouvée dans l'échantillon de découverte. Pour expliquer l'absence de reproductibilité de ces résultats, nous suggérons que l'association était trop faible pour être reproduite dans cet échantillon qui, par ailleurs, ne permettait pas sa stratification sur la base du statut de l'épitope partagé du gène HLA-DRB1.

Néanmoins, le gène PGLYRP1 reste un gène candidat intéressant à explorer dans le contexte physiopathologique et autoimmunitaire de la PR. En effet, les PGRPs (protéines de reconnaissance du peptidoglycane) ont un rôle bien démontré dans l'activité antibactérienne (268, 269) en reconnaissant un peptidoglycane, qui est un composant de la paroi cellulaire de toutes les bactéries, absent de la surface des cellules eucaryotes (270). Des études ont démontré que les PGRPs influencent les interactions « hôte-pathogène » par le biais de leurs propriétés pro-inflammatoires, qui sont indépendantes de leurs activités antibactériennes, en modulant l'équilibre entre les cellules pro-inflammatoires Th17 et les lymphocytes T<sub>reg</sub> (ou LT régulateurs) (271). En revanche, PGLYRP1 a un effet inverse (anti-inflammatoire) sur un modèle murin d'arthrite aiguë induite par les peptidoglycanes (272). Cet effet inverse pourrait être expliqué à son tour soit par une grande hétérogénéité des variants impliqués différemment dans les processus physiopathologiques, soit par un défaut moléculaire et/ou cellulaire dans la réponse immune déclenchée, dû au caractère induit du modèle murin.

Compte tenu du rôle potentiel des bactéries dans la pathogenèse de la PR (273), et en présence de l'épitope partagé du gène HLA-DRB1, une surexpression du gène PGLYRP1 serait un facteur favorisant le dérèglement

des voies inflammatoires, de la production de cytokines et éventuellement de la présentation d'antigène. La tendance d'association que nous avons montrée pour PGLYRP1 intéressait particulièrement les patients PR portant l'épitope partagé HLA-DRB1 (SE+). Ce résultat doit être confirmé par d'autres études génétiques dans de plus grands échantillons disposant des données de génotypage pour le gène HLA-DRB1.

Sur un autre plan, sur les dix SNPs étudiés pour les six gènes retenus dans cette étude, un seul SNP (rs3218253 du gène IL2RB) a été récemment rapporté dans la dernière méta-analyse de GWAS comme étant associé à la PR chez les Caucasiens (51). Notre étude n'a toutefois pas permis de détecter une association significative dans l'échantillon de 200 familles trio caucasiennes Françaises dans lequel il a été testé.

Il est nécessaire que d'autres analyses systématiques soient développées dans des échantillons plus larges, en combinant les données de génotypage et de transcriptome, afin de mieux exploiter la capacité de cette approche pour identifier de nouveaux biomarqueurs potentiels de la PR.

## ARTICLE 2

### Relation between SNPs and expression level for candidate genes in Rheumatoid Arthritis

Mostefa Fodil <sup>1,2</sup> , Vitor Hugo Teixeira <sup>1,3</sup> , Valérie Chaudru <sup>1</sup> , Pascal Hilliquin <sup>4</sup> , Stefano Bombardieri <sup>5</sup> , Alejandro Balsa <sup>6</sup> , René Westhovens <sup>7</sup> , Pilar Barrera <sup>8</sup> , Helena Alves <sup>9</sup> , Paola Migliorini <sup>5</sup> , Thomas Bardin <sup>10</sup> , François Cornelis <sup>11\*</sup> , Abdallah Boudjema <sup>2</sup> , Elisabeth Petit-Teixeira <sup>1</sup>

1. GenHotel-EA3886, Evry-Val d'Essonne University, Evry-Genopole, France
2. Département de Génétique Moléculaire Appliquée, Université des Sciences et de la Technologie d'Oran Mohammed Boudiaf (USTO-MB), BP 1505, El Mnaouer, Algeria.
3. Centre for Respiratory Research, University College London, United Kingdom
4. Centre Hospitalier Sud Francilien, Corbeil-Essonnes, France.
5. Pisa University, 56126 Pisa, Italy
6. La Paz Hospital, 28046 Madrid, Spain
7. Rheumatology KU Leuven, Leuven, Belgium
8. Nijmegen University, 6500HB Nijmegen, the Netherlands
9. Porto San Joao Hospital, 4200 Porto, Portugal
10. Fédération de Rhumatologie, Pôle de l'Appareil Locomoteur, Lariboisière Hospital, AP-HP, Paris, France
11. GenHotel-Auvergne, CHU de Clermont-Ferrand, Auvergne University, France

\* ECRAF coordinator: The European Consortium on Rheumatoid Arthritis Families (ECRAF): F. Cornélis (coordinator), T. Bardin (France), P. Migliorini, S. Bombardieri (Italy), R. Westhovens, J. Dequeker (Belgium), A. Balsa, D. Pascuale-Salcedo (Spain), P. Barrera, L. Van de Putte, P. Van Riel, T.R. Radstake (The Netherlands), and H. Alves, A. Lopes-Vaz, M. Fernandes, C. Vaz (Portugal).

## Relationship between SNPs and expression level for candidate genes in rheumatoid arthritis

M Fodil<sup>1,2</sup>, VH Teixeira<sup>1,3</sup>, V Chaudru<sup>1</sup>, P Hilliquin<sup>4</sup>, S Bombardieri<sup>5</sup>, A Balsa<sup>6</sup>, R Westhovens<sup>7</sup>, P Barrera<sup>8</sup>, H Alves<sup>9</sup>, P Migliorini<sup>5</sup>, T Bardin<sup>10</sup>, F Cornelis<sup>11\*</sup>, A Boudjema<sup>2</sup>, E Petit-Teixeira<sup>1</sup>

5 <sup>1</sup>GenHotel-EA3886, Evry-Val d'Essonne University, Evry-Genopole, France, <sup>2</sup>Department of Applied Molecular Genetics, University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB), El M'naouer, Algeria, <sup>3</sup>Centre for Respiratory Research, University College London, UK, <sup>4</sup>South Francilien Hospital Centre, Corbeil-Essonnes, France, <sup>5</sup>University of Pisa, Italy, <sup>6</sup>La Paz Hospital, Madrid, Spain, <sup>7</sup>Department of Rheumatology, University of Leuven, Belgium, <sup>8</sup>Nijmegen University, The Netherlands, <sup>9</sup>Porto San Joao Hospital, Portugal, <sup>10</sup>Department of Rheumatology, Division of Musculoskeletal Diseases, Lariboisière Hospital, Paris, France, and <sup>11</sup>GenHotel-Auvergne, CHU Clermont-Ferrand, University of Auvergne, Clermont-Ferrand, France

**Objectives:** The study of polymorphisms of genes differentially expressed may lead to the identification of putative causal genetic variants in multifactorial diseases such as rheumatoid arthritis (RA). Based on preceding transcriptomic results, we genotyped 10 single nucleotide polymorphisms (SNPs) belonging to six genes (*SI00A8*, *RNASE2*, *PGLYRP1*, *RUNX3*, *IL2RB*, and *LY96*) showing the highest fold change (> 1.9) when level of expression was compared between RA patients and controls. These SNPs were then analysed to evaluate their role in RA.

**Method:** The relationship between gene expression and genotypes of SNPs was first investigated by Kruskal–Wallis and Mann–Whitney tests in RA patients and controls. The genetic association of these SNPs with RA were then analysed using family-based association tests in trio families.

**Results:** We found that *RNASE2* gene expression was related to rs2013109 genotypes in 14 RA patients ( $p = 0.030$ ). The association study in a discovery sample of 200 French trio families revealed a significant association with RA for one SNP, *PGLYRP1*-rs2041992 ( $p = 0.019$ ); this association was stronger in trios where RA patients carried the *HLA-DRBI* shared epitope (SE) ( $p = 0.003$ ). However, this association was not found in a replication sample of 230 European trio families ( $p = 0.6$ ).

**Conclusions:** Family-based association tests did not reveal an association between RA and any SNP of the candidate genes tested. However, *RNASE2* gene expression was differentially expressed in RA patients considering a sequence polymorphism. This result led us to highlight the potential disease-specific regulation for this candidate gene in RA.

Rheumatoid arthritis (RA) is a chronic inflammatory and autoimmune disease characterized by inflammation of the synovial membrane leading to the destruction of affected joints (1). The heterogeneity of the disease stems from the complexity of the immune mechanisms and the multiplicity of the elements involved in the disease process, including a crucial role for cytokines and signalling molecules contributing to the persistence of RA (2, 3). In addition, RA is a complex disease for which several studies have demonstrated

an association between genetic factors and response to therapy such as methotrexate (MTX) treatment (4).

The study of gene expression profiles may enable the identification of distinct molecular disease mechanisms involved in RA pathology. Several gene expression profiling studies have identified a remarkably elevated expression of a spectrum of genes in RA patients compared to controls. In our previous study performed in peripheral blood mononuclear cells (PBMCs) (5), we highlighted 339 differentially expressed genes (238 down-regulated and 101 up-regulated). Most of these newly discovered genes were involved essentially in different functional immunity and defence-related mechanisms such as pro-inflammation, anti-microbial activity, cellular stress, and immunomodulatory functions (6–8).

Recent genome-wide association studies (GWAS) clearly identified many new genetic markers for RA (9, 10). What is less clearly understood, however, is the role of these susceptibility markers in determining disease onset and its persistence.

Elisabeth Petit-Teixeira, GenHotel-EA3886, European Research Laboratory for Rheumatoid Arthritis, Rue Gaston Crémieux, 91057 Evry-Genopole Cedex, France.  
E-mail: epetit@polyartirite.net

\* Coordinator of the European Consortium on Rheumatoid Arthritis Families (ECRAF): T. Bardin (France), P. Migliorini, S. Bombardieri (Italy), R. Westhovens, J. Dequeker (Belgium), A. Balsa, D. Pascual-Salcedo (Spain), P. Barrera, L. Van de Putte, P. Van Riel, T. R. Radstake (The Netherlands), and H. Alves, A. Lopes-Vaz, M. Fernandes, C. Vaz (Portugal).

Accepted 22 April 2014

The aim of our study was first to analyse differentially expressed genes regarding the association between level of expression and genotyping data. Among genes showing the highest fold change between RA and control groups in our previous expression study (5), six genes were chosen as they do not exhibit copy number variations (CNVs). Single nucleotide polymorphisms (SNPs) were selected in genes with available data on linkage disequilibrium and having a minor allele frequency (MAF) higher than 5%. Second, polymorphisms were studied with a familial-based association test in trio families. Stratification of trio families was then performed regarding the presence in RA patients of at least one of the shared epitope (SE) alleles of the *HLA-DRB1* gene (11), anti-citrullinated peptide antibody (ACPA) status, and erosion.

This study is one way to discover new biomarkers for RA through specific expression variation, which is of a great interest for earlier diagnosis and better comprehension of molecular mechanisms of the disease.

**Method**

**Study population**

All subjects provided informed consent and the ethics committee of The Bicêtre Hospital (Kremlin-Bicêtre, Assistance Publique-Hôpitaux de Paris, France) approved the study. RA families were recruited through a national media campaign followed by selection of individuals who fulfilled the 1987 American College of Rheumatology criteria for RA according to the physicians treating the patients (12). A rheumatologist university fellow reviewed all clinical data. Sample 1 (Table 1) comprised DNA from 14 unrelated French patients. Thirteen unrelated French healthy controls were used for comparison. Sample 2 (Table 1) was the discovery sample for the familial-based association study and consisted of 200 French Caucasian unrelated trio families (one RA patient and both parents) with the four grandparents of French Caucasian origin. Sample 3 (Table 1) was the replication sample for the familial-based association study and was made up of the DNA from another 230 European trio families from France, Italy, Portugal, Spain, Belgium, and The Netherlands.

**Choice of genes and SNPs**

The choice of genes for this study was determined from the results of a whole transcriptome analysis in RA (5). In brief, Illumina Human-6v2 Expression BeadChips were used for a complete genome-wide transcript profiling of PBMCs from RA patients and healthy controls. Differential analysis for each gene was performed with one-way analysis of variance (ANOVA) and p-values were adjusted to control the false discovery rate (FDR 5%). A differential expression of 339 reference sequence genes (238 down-regulated and 101 up-regulated) between the two groups was observed. Microarray data were validated by real-time polymerase chain reaction (PCR) in a set of a few genes showing a high degree of correlation. From an initial list of 339 genes, we selected genes (n = 15) having a fold change higher than 1.9 when expression level was compared between RA and controls. Five genes were excluded from the list as they were reported to have at least one CNV described in the Database of Genomic Variants (<http://dgv.tcag.ca/>). Four genes were also excluded as they did not harbour SNPs or their SNPs did not have an MAF > 5%. Finally, six genes were included in this study: *SI00A8*, *RNASE2*, *PGLYRP1*, *RUNX3*, *IL2RB*, and *LY96*. For these genes, 10 SNPs (one SNP for four genes, two SNPs for one gene, and four SNPs for one gene) were retained for the study regarding disequilibrium data with tagSNPs and MAF > 5% (Table 2). For genes found to be associated with RA in the literature, we preferentially chose SNPs that showed a significant association.

**Molecular genotyping methods**

Genomic DNA of samples was isolated and purified from fresh peripheral blood leucocytes according to standard protocols (13).

Genotyping of nine SNPs [rs3006488 (*SI00A8* gene); rs2013109 (*RNASE2* gene); rs2072561 and rs2041992 (*PGLYRP1* gene); rs2282718, rs12082398, rs1507101 and rs2236851 (*RUNX3* gene); and rs3218253 (*IL2RB* gene)] was carried out with a Taqman® 5' allelic discrimination assay. This was performed on an ABI-7500 real-time PCR instrument (Applied Biosystems, now Life

Table 1. Characteristics of rheumatoid arthritis (RA) index cases from the investigated samples.

	Sample 1 (n = 14)	Sample 2 (n = 200)	Sample 3 (n = 240)
Females, n (%)	12 (86)	177 (88.5)	212 (85)
Age of disease onset (years), mean ± SD	47 ± 12	30.5 ± 8	31 ± 9
RA patients with bone erosions, n (%)	10 (72)	168 (84)	167 out of 234 (71)
RA patients seropositive for RF rheumatoid factor, n (%)	13 (93)	153 (76.5)	170 out of 231 (74)
RA patients seropositive for ACPA, n (%)	8 out of 9 (89)	155 (77.5)	60 out of 83 (72)
RA patients carrying at least one <i>HLA-DRB1</i> SE allele, n (%)	NA	159 (79.5)	NA

RF, Rheumatoid factor; SE, shared epitope; NA, not available; SD, standard deviation; ACPA, anti-citrullinated protein antibodies.



Table 2. Genes and SNPs selected and respective minor allele frequencies in HapMap-CEU reference population.

Gene	Position *	Fold change †	SNP	Alleles	MAF (allele)
<i>ST00A8</i>	1:153362508–153363664	2.9	rs3006488	A/G	0.16 (G)
<i>LY96</i>	8:74903587–74941322	2.58	rs1991262	G/C	0.32 (C)
<i>RNASE2</i>	14:21423611–21424595	2.55	rs2013109	C/G	0.24 (C)
<i>IL2RB</i>	22:37521878–37571094	2.15	rs3218253	G/A	0.17 (A)
<i>PGLYRP1</i>	19:46522411–46526323	2	rs2072561	T/G	0.09 (G)
			rs2041992	C/T	0.41 (T)
<i>RUNX3</i>	1:25226002–25291612	1.91	rs12082398	C/T	0.45 (T)
			rs1507101	G/A	0.48 (A)
			rs2236851	C/T	0.11 (T)
			rs2282718	G/A	0.39 (A)

MAF, Minor allele frequency from HapMap-CEU population data (www.hapmap.org).

\* Position from www.ensembl.org.

† Fold change from (5).

145 Technologies, Foster City, CA, USA) according to the  
 manufacturer’s protocol. Genotyping of each sample was  
 automatically attributed using ABI PRISM® 7500  
 Sequence Detection System software ‘SDS’ for allelic  
 discrimination. Genotyping of rs1991262 (*LY96* gene)  
 150 was performed by PCR followed by the restriction frag-  
 ment length polymorphism method (PCR-RFLP) (14).  
 The designed primers were: forward 5’-  
 TAGGCTAGAGATTTCTTGAGGA-3’ and reverse  
 5’-GCAGACAGCCTGCAGCTGCC-3’. The 361-bp  
 155 amplified fragment was digested with the restriction  
 enzyme ScrFI (New England Biolabs Inc, Ipswich, MA,  
 USA), generating three fragments for the C allele (183,  
 111, and 67 bp) and two fragments for the G allele (250  
 and 111 bp).

160 For quality control of our data, two reference samples  
 [from Centre d’Etude du Polymorphisme Humain  
 (CEPH) families] were genotyped with all our samples.  
 Moreover, 10% of randomly chosen samples were geno-  
 typed in an independent experiment.

#### 165 Statistical analysis

For each of the 10 SNPs belonging to selected genes, the  
 mean expression levels [obtained from array data (5)] for  
 the different genotypes were compared between RA  
 patients (sample 1) and 13 healthy controls. We pooled  
 170 homozygous genotypes for which sample size was lower  
 than five with heterozygous genotypes. This comparison  
 was assessed by non-parametric tests (Kruskal–Wallis  
 and Mann–Whitney) using XLSTAT version 2012  
 (Addinsoft, Paris, France; www.xlstat.com).

175 Before the association tests, Hardy–Weinberg equi-  
 librium (HWE) was checked in controls constituted by  
 parental untransmitted alleles to RA index cases from  
 the trio families. First, a family-based association test  
 was used to compare SNP alleles distribution in affected  
 180 offspring with the expected distribution using Mendel’s  
 law, derived from the original Transmission  
 Disequilibrium Test (15). This analysis was performed  
 by using FBAT version 2.0.4 (www.biostat.harvard.edu/

~fbat/default.html). Second, we evaluated the genotype  
 relative risk (GRR), which compares the affected off-  
 spring’s genotype with the control genotype derived  
 from non-transmitted parental chromosomes, using the  
 method described by Lathrop (16). Haplotype analyses  
 185 for the *RUNX3* and *PGLYRP1* genes were performed  
 using the haplotype-based association testing (HBT)  
 option in the FBAT program (17). Stratification of trio  
 families was then performed regarding the presence in  
 RA patients of at least one of the SE alleles of the *HLA-*  
 190 *DRB1* gene (\*0101, \*0102, \*0401, \*0404, \*0405, \*0408,  
 \*0409, \*1001, and \*1402 alleles, associated with RA in  
 our population) (11). Other stratifications were made  
 according to anti-citrullinated peptide antibodies  
 (ACPAs) and erosion status.

195 The significance of the p-value was assessed at  $7 \times 10^{-4}$   
 as the Bonferroni correction for multiple tests was applied  
 considering 10 SNPs and six stratifications. If at least one  
 association analysis had a significant result in the discov-  
 ery sample 2, it was tested in the replication sample. A  
 200 result after pooling the two trio families samples is also  
 provided.

#### Results

After quality control of data consisting of genotyping  
 reference samples and of genotyping redo for 10% of  
 205 randomly chosen samples, the accuracy of the genotyping  
 data was 100%.

First, the expression level from the array data (from 5) was  
 compared between genotypes for each SNP in sample 1.  
 When *RNASE2* mRNA expression was studied considering  
 210 rs2013109G/C genotypes, the mean level of expression for  
 (CC+CG) RA patients was significantly higher than that for  
 (GG) RA patients ( $p = 0.03$ ; Figure 1). Allele rs2013109-C  
 of *RNASE2* is thus associated with high expression of the  
 gene in the RA context. When looking at the eQTL loci using  
 215 <http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>, no gene of  
 our study was indicated, except for the *LY96* gene, for  
 which cis eQTL has been described for two SNPs,  
 rs10504554 and rs7822407 (18), that were not in linkage

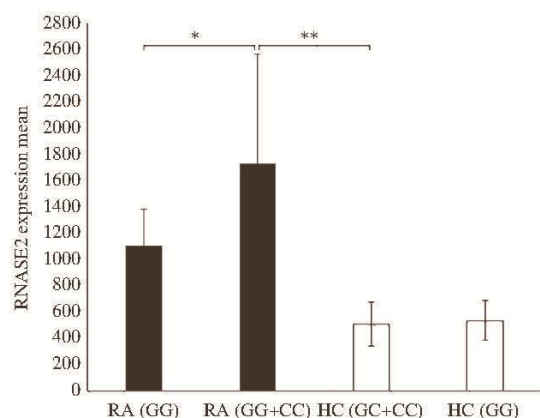


Figure 1. Comparison of expression level for the *RNASE2* gene considering rs2013109 genotypes in RA patients and healthy controls (HC). \*  $p = 0.026$  and \*\*  $p = 0.001$  with the Mann-Whitney test.

disequilibrium with the one we tested (rs1991262). We further studied the association with RA by comparing alleles and genotypes for each SNP between RA and control groups (from the expression study related in 3) using Fisher's exact test. Neither allele nor genotype was associated with RA (data not shown).

All SNPs in the investigated samples 2 and 3 were in HWE in the control group (made up of the non-transmitted parental chromosomes from the trio subjects) (data not shown). The family-based association test, using FBAT software, showed that there was a significant over-transmission for allele T of *PGLYRP1*-rs2041992 in sample 2 ( $p = 0.019$ , Table 3). No other significant result was observed for any SNP, or haplotype (data not shown for genes *RUNX3* and *PGLYRP1*). GRR analysis did not

show any significant association of any genotype (data not shown). The same tests were conducted in subgroups stratified according to the presence (SE+) or absence (SE-) of the *HLA-DRB1* SE alleles in RA patients from each trio family. Significant transmission disequilibrium for the *PGLYRP1*-rs2041992 within the SE+ subgroup ( $p = 0.0032$ ; Table 4) was observed again. No significant result was found after stratification considering ACPA status and the presence of erosion for any SNP.

The result of the association test regarding *PGLYRP1*-rs2041992 showed a p-value equal to 0.0032 but superior to the threshold of significance after Bonferroni correction. Considering this trend for association, we decided to study sample 3, which consisted of 230 European trio families, using the same methods. Association was not confirmed for *PGLYRP1*-rs2041992 ( $p = 0.6$ , Table 4). The test for *HLA-DRB1* genotyping was not available after stratification. A global association test realized in the pooled sample of 450 trio families did not reach significance and did not confirm an association of *PGLYRP1*-rs2041992 with RA ( $p = 0.238$ , Table 4).

### Discussion

This study was designed to analyse the relationship between genotypes of chosen SNPs and the level of expression of their related genes in RA patients. The list of genes to explore was established on the basis of the differential expression in RA vs. controls (5), and included four up-regulated genes (*S100A8*, *PGLYRP1*, *RNASE2*, and *LY96*) and two down-regulated genes (*RUNX3* and *IL2RB*). We tested the hypothesis that the differentially expressed genes were associated with RA,

Table 3. Familial-based association test for the 10 SNPs in sample 2 of 200 French Caucasian RA trio families.

Gene	SNP	Allele	Allele frequency	No. of informative families	p-value
<i>S100A8</i>	rs3006488	G	0.120	31	0.75
		A	0.880		
<i>RNASE2</i>	rs2013109	C	0.280	63	0.12
		G	0.720		
<i>PGLYRP1</i>	rs2072561	G	0.046	16	0.47
		T	0.954		
		T	0.337		
<i>RUNX3</i>	rs12082398	C	0.663	63	0.65
		C	0.427		
		T	0.595		
<i>IL2RB</i>	rs1507101	G	0.405	134	0.70
		A	0.583		
		T	0.189		
		C	0.811		
<i>LY96</i>	rs2236851	C	0.811	81	0.55
		A	0.385		
		G	0.615		
<i>IL2RB</i>	rs2282718	A	0.247	65	0.65
		G	0.753		
<i>LY96</i>	rs3218253	A	0.247	121	0.93
		G	0.753		
<i>LY96</i>	rs1991262	G	0.483	95	0.93
		C	0.517		

\* p-value not significant after Bonferroni correction considering 10 SNPs and six stratifications ( $p > 7 \times 10^{-4}$ ).

Table 4. Familial-based association test for *PGLYRP1*-rs2041992.

Sample	Allele	Allele frequency	No. of informative families	p-value
Sample 2	T	0.337	134	0.019*
	C	0.663		
Sample 2 SE+	T	0.344	107	0.0032*
	C	0.656		
Sample 3	T	0.313	132	0.6
	C	0.687		
Samples 2+3	T	0.321	260	0.238
	C	0.679		

SE+, RA index sharing at least one shared epitope allele of the *HLA-DRB1* gene; SE-, RA index not sharing an SE allele of the *HLA-DRB1* gene.

\* p-value not significant after Bonferroni correction considering 10 SNPs and six stratifications ( $p > 7 \times 10^{-4}$ ).

270 using family-based methodology. The main advantage of  
family trio data is that they provide perfectly matched  
controls for each patient investigated. Each patient's  
chromosome, transmitted by a given parent, was in fact  
perfectly matched for the population of origin with the  
untransmitted chromosome of each heterozygous parent.

275 Concerning the association between transcriptome and  
genotyping data, previous studies have provided some  
interesting results in the context of RA (19–24). In our  
study, the *RNASE2* gene showed an association between  
mRNA expression level and genotypes of one intronic  
polymorphism rs2013109 in RA patients. *RNASE2* (also  
known as the eosinophil-derived neurotoxin, EDN) is one  
of four major proteins found in the large specific granules  
of human eosinophilic leucocytes. It is also a selective  
chemoattractant for cultured dendritic cells (23). In the  
autoimmune context of RA, the *RNASE2* gene could be  
one of most important factors involved in the physio-  
pathology and the deregulation of mechanisms of defence  
leading to inflammation and joint destruction. Our results  
agree with this hypothesis by indicating that the  
rs2013109-C allele of *RNASE2* is associated with high  
expression of the gene in the context of RA. This pre-  
liminary study has to be confirmed using alternative  
methodologies. First, as expression data used in this  
study were produced on PBMCs, the same approach  
should be performed using transcriptome results obtained  
from tissue specific samples. PBMC gene expression  
profiling allows identification of pathogenetic process  
primarily associated with the cause of the disease. Then  
abnormal genes and gene activities, not only limited to  
PBMCs but also in cells of pathological tissue, could be  
identified. Nevertheless, our analysis should be tested for  
replication using samples for which transcriptome data  
from synovial tissue and polymorphism data are avail-  
able. Second, regarding selected SNPs, as we were  
focused on gene sequence, our study has investigated  
the potential association for some sequence polymorph-  
isms with expression level. The association of SNPs  
located outside each gene was not investigated and  
could be helpful to complete our preliminary results.  
Finally, benefiting from the current development of  
Next Generation Sequencing methodologies, use of the

RNAseq approach will allow future studies to consider  
transcription diversity more widely than array technology  
does.

Concerning the SNP association study, only one SNP  
(rs3218253 in the *IL2RB* gene) was reported in the last  
meta-analysis of GWAS as being associated with RA in  
Caucasians (9). Our study, however, did not detect a  
significant association in our French Caucasian popula-  
tion consisted of 200 trio families ( $p = 0.9$ , Table 3).  
*PGLYRP1*-rs2041992 showed a proof of association with  
RA in this sample. Nevertheless, this association was not  
confirmed when using a replication sample of 230  
European trio families. The samples used were not sig-  
nificantly different with regard to epidemiological and  
clinical data (data not shown). Furthermore, to test the  
hypothesis of genetic heterogeneity, we then performed  
an association analysis in the French subgroup of this  
sample of European families and still found no associa-  
tion (data not shown). We suggest that the association was  
too weak to be reproduced in this sample without strati-  
fication on the basis of the *HLA-DRB1* status.

Peptidoglycan recognition proteins (PGRPs) have a  
demonstrated role in the antibacterial activity (25, 26)  
by recognizing peptidoglycan, which is an essential cell  
wall component of all bacteria not present in eukaryotic  
cells (27). Recent studies have shown that PGRPs  
influence host-pathogen interactions through their pro-  
inflammatory or anti-inflammatory properties, which  
are independent of their antibacterial activities (28),  
modulating the balance between inflammatory T-helper  
(Th)17 cells and regulatory T cells. By contrast,  
*PGLYRP1* has the opposite anti-inflammatory effect in  
a model of peptidoglycan-induced acute inflammation  
and arthritis in mice (29). Taking into account the  
potential role of bacteria in the pathogenesis of RA  
(30), an up-regulation of *PGLYRP1* in the context of  
the *HLA-DRB1* SE may influence the onset of the  
disease affecting inflammatory circuits, cytokine pro-  
duction, and possibly antigen presentation. The associa-  
tion we have shown concerned particularly the RA  
index carrying the *HLA-DRB1* SE and should be con-  
firmed by others genetic studies in larger samples typed  
for *HLA-DRB1* alleles.

It is necessary for further systematic analyses, combining genotyping and transcriptome data, to be developed in a larger sample, so as to exploit the ability of this approach to identify novel potential biomarkers for RA.

360 Acknowledgements

This work was supported by the French Association of Polyarthritis, the French Society of Rheumatology, the Rheumatism Association, the Association of Polyarthritis, Genopole, and the University of Evry-Val d'Essonne. We are grateful to the RA patients, their families and rheumatologists for participation in this study.

References

1. Seldin MF, Amos CI, Ward R, Gregersen PK. The genetics revolution and the assault on rheumatoid arthritis. *Arthritis Rheum* 1999;42:1071–9.

370 2. Pope RM. Apoptosis as a therapeutic tool in rheumatoid arthritis. *Nat Rev Immunol* 2002;2:527–35.

3. Zhang Y, Qiu H, Zhang H, Wang L, Zhuang C, Liu R. Vascular endothelial growth factor A (VEGFA) polymorphisms in Chinese patients with rheumatoid arthritis. *Scand J Rheumatol* 2013;42:344–8.

375 4. Cáliz R, del Amo J, Balsa A, Blanco F, Silva L, Sanmarti R, et al. The C677T polymorphism in the MTHFR gene is associated with the toxicity of methotrexate in a Spanish rheumatoid arthritis population. *Scand J Rheumatol* 2012;41:10–14.

380 5. Teixeira VH, Olaso R, Martín-Magniette M-L, Lasbleiz S, Jacq L, Oliveira CR, et al. Transcriptome analysis describing new immunity and defense genes in peripheral blood mononuclear cells of rheumatoid arthritis patients. *PLoS One* 2009;4:e6803.

385 6. Toonen EJ, Barrera P, Radstake TR, van Riel PL, Scheffer H, Franke B, et al. Gene expression profiling in rheumatoid arthritis: current concepts and future directions. *Ann Rheum Dis* 2008;67:1663–9.

7. van der Pouw Kraan TC, Wijbrandts CA, van Baarsen LG, Voskuyl AE, Rustenburg F, Baggen JM, et al. Rheumatoid arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients. *Ann Rheum Dis* 2007;66:1008–14.

390 8. Baechler EC, Batliwalla FM, Reed AM, Peterson EJ, Gaffney PM, Moser KL, et al. Gene expression profiling in human autoimmunity. *Immunol Rev* 2006;210:120–37.

395 9. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010;42:508–14.

400 10. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, et al. Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Hum Mol Genet* 2008;17:2274–9.

405 11. Michou L, Croiseau P, Petit-Teixeira E, du Montcel ST, Lemaire I, Pierlot C, et al. Validation of the reshaped shared epitope HLA-DRB1 classification in rheumatoid arthritis. *Arthritis Res Ther* 2006;28;8:R79.

12. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.

13. Sambrook P, Fritsch E, Maniatis T, editors. A laboratory manual, vol. 1, 2nd edn. New York: Cold Spring Harbor Laboratory Press, 415 1989.

14. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980;32:314–31.

15. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;52:506–16.

16. Lathrop GM. Estimating genotype relative risks. *Tissue Antigens* 1983;22:160–6. 425

17. Lake SL, Blacker D, Laird NM. Family-based tests of association in the presence of linkage. *Am J Hum Genet* 2000;67:1515–25.

18. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond – the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010;5:e10693. 430

19. Maxwell JR, Gowers IR, Kuet KP, Barton A, Worthington J, Wilson AG. Expression of the autoimmunity associated TNFAIP3 is increased in rheumatoid arthritis but does not differ according to genotype at 6q23. *Rheumatology* 2012;51:1514–15.

20. Teixeira VH, Jacq L, Lasbleiz S, Hilliquin P, Oliveira CR, Cornelis F, et al. Genetic and expression analysis of CASP7 gene in a European Caucasian population with rheumatoid arthritis. *J Rheumatol* 2008;35:1912–18. 435

21. Muñoz-Valle JF, Ruiz-Quezada SL, Oregón-Romero E, Navarro-Hernández RE, Castañeda-Saucedo E, De la Cruz-Mosso U, et al. PAI-1 mRNA expression and plasma level in rheumatoid arthritis: relationship with 4G/5G PAI-1 polymorphism. *Rheumatol Int* 2012;32:3951–6. 440

22. Myouzen K, Kochi Y, Okada Y, Terao C, Suzuki A, Ikari K, et al. Functional variants in NFKBIE and RTKN2 involved in activation of the NF-κB pathway are associated with rheumatoid arthritis in Japanese. *PLoS Genet* 2012;8:e1002949. 445

23. Yang D, Rosenberg HF, Chen Q, Dyer KD, Kurosaka K, Oppenheim JJ. Eosinophil-derived neurotoxin (EDN), an antimicrobial protein with chemotactic activities for dendritic cells. *Blood* 2003;102:3396–403. 450

24. Chatzikiriakidou A, Voulgari PV, Georgiou I, Drosos AA. Altered sequence of the ETS1 transcription factor may predispose to rheumatoid arthritis susceptibility. *Scand J Rheumatol* 2013;42:11–14. 455

25. Lu X, Wang M, Qi J, Wang H, Li X, Gupta D, et al. Peptidoglycan recognition proteins are a new class of human bactericidal proteins. *J Biol Chem* 2006;281:5895–7.

26. Royet J, Dziarski R. Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defenses. *Nat Microbiol Rev* 2007;5:264–77. 460

27. Dziarski R, Gupta D. Mammalian peptidoglycan recognition proteins (PGRPs) in innate immunity. *Innate Immunity* 2010;16:168–74.

28. Rohatgi A, Ayers CR, Khera A, Mc Guire DK, Das SR, Matulevicius S, et al. The association between peptidoglycan recognition protein-1 and coronary and peripheral atherosclerosis: observations from the Dallas Heart Study. *Atherosclerosis* 2009;203:569–75. 465

29. Témoin S, Chakaki A, Askari A, El-Halaby A, Fitzgerald S, Marcus RE, et al. Identification of oral bacterial DNA in synovial fluid of patients with arthritis with native and failed prosthetic joints. *J Clin Rheumatol* 2012;18:117–21. 470

30. Hitchon CA, El-Gabalawy HS. Infection and rheumatoid arthritis: still an open question. *Curr Opin Rheumatol* 2011;23:352–7. 475

**CHAPITRE 4 – CONCLUSION  
GENERALE & PERSPECTIVES**

Ce travail de thèse a porté sur deux grandes parties s'articulant autour d'une thématique centrale, la génétique de la polyarthrite rhumatoïde.

Pour la première partie de notre étude, la recherche d'association avec la PR, dans la population Algérienne, pour des polymorphismes de gènes déjà confirmés dans d'autres populations a permis d'établir une preuve solide d'association pour les deux SNPs les plus importants PTPN22<sub>rs2476601</sub> et STAT4<sub>rs7574865</sub>. Une tendance à l'association a été toutefois retrouvée pour les trois SNPs restants (6q23<sub>rs6927172</sub>, IRF5<sub>rs2004640</sub> et TRAF1/C5<sub>rs10818488</sub>) mais reste insuffisante pour établir une preuve d'association. Ces résultats, préliminaires, convergent toutefois dans le même sens que la grande majorité des travaux publiés dans les populations proches géographiquement et/ou ethniquement de la population Algérienne. De plus amples études avec des échantillons élargis, de cas PR et de témoins sains, permettraient d'apporter une plus grande puissance dans les analyses statistiques. D'autre part, la constitution d'une cohorte familiale de PR, afin de pouvoir mener notamment une étude de type *trio*, permettrait de tester la réplication de l'association génétique des polymorphismes étudiés.

L'exploration d'autres polymorphismes situés dans d'autres gènes parmi la grande liste de gènes confirmés dans la PR serait l'aboutissement naturel de ce premier travail dans la population Algérienne. La priorité serait alors donnée au gène HLA-DRB1 qui permettrait d'établir, pour la population Algérienne, le profil « *Epitope Partagé* » caractéristique de la PR. La deuxième priorité serait la caractérisation du profil sérologique des patients PR concernant les anticorps anti peptides citrullinés (ACPA), ce qui permettrait d'élargir les analyses de stratification et l'établissement de groupes distincts de patients PR. La finalité serait alors la caractérisation des polymorphismes de gènes, ayant été mis en évidence par les études pangénomiques, comme CTLA4, CD40, IL2RB, PRKCQ, IL2/IL21 et REL. L'élargissement de cette première liste de gènes à tous les autres facteurs génétiques confirmés dans la PR, en plus des autres éléments déjà cités, permettrait une meilleure caractérisation génétique de notre population,

encore mal étudiée, et apporterai une contribution à la compréhension des mécanismes moléculaires et génétique de la PR. L'évolution naturelle de ce volet de l'étude serait d'être « à jour » avec ce qui se fait dans le domaine de la recherche en génétique de la PR.

Pour le deuxième volet de l'étude, l'analyse d'association génétique - dans des familles trio - des polymorphismes de gènes présentant un différentiel d'expression dans la PR représente un complément d'information pour la compréhension du fonctionnement de ces gènes et de leur implication dans la PR. Ce travail d'analyse d'association génétique a été complété par l'étude de la relation entre le génotype des SNPs étudiés et le taux d'expression du gène considéré pour le sous groupe d'échantillons ayant les deux données.

Les analyses familiales (*Trio ou Sib-Pairs*) de gènes candidats ainsi que les analyses de polymorphismes du génome entier réalisées sur de grandes cohortes de cas comparés à des contrôles ont permis de dégager un certain nombre de facteurs génétiques mineurs. Cependant la totalité des facteurs génétiques n'est pas identifiée. Des analyses complémentaires du génome, telles que l'étude du transcriptome, et la recherche de liens entre les données génétiques et transcriptomiques sont des outils indispensables pour orienter la poursuite de l'étude des gènes candidats et plus largement, des biomarqueurs spécifiques de la PR.

Cette étude est une façon de découvrir de nouveaux biomarqueurs de la PR par la variation d'expression spécifique, qui est d'un grand intérêt pour un diagnostic précoce et une meilleure compréhension des mécanismes moléculaires de la PR. Ces biomarqueurs pourraient avoir un intérêt clinique avéré en ouvrant la perspective d'identification de cibles thérapeutiques pour les biothérapies du futur. L'idéal pour la recherche concernant cette maladie incurable serait la mise en place de protocoles de thérapie génique permettant de l'éradiquer définitivement.

Sur un autre plan, les polymorphismes d'intérêts mis en évidence au cours de cette thèse représentent des cibles de choix pour l'analyse conjointe de la PR dans les deux populations algérienne et française en vue d'une comparaison ethnique des différents facteurs de prédisposition génétique à la PR. Ce travail serait une contribution aux études multicentriques réalisées dans d'autres populations et pourrait être une étape de plus dans la compréhension de la physiopathologie de la PR et une participation à l'étude génétique des populations.

Cette thèse de cotutelle a permis également d'effectuer un transfert de technologies du laboratoire GenHotel-EA3886 (Evry-France) vers le laboratoire LGMC-USTO (Oran-Algérie), notamment des tests statistiques applicables à l'analyse génétique mais surtout du génotypage par sondes fluorescentes en temps réel (TaqMan®). La mise en place de cette technique au niveau du laboratoire de génétique moléculaire et cellulaire à Oran (LGMC-USTO) présente alors un double atout, scientifique mais aussi pédagogique.



## **CHAPITRE 5 – ANNEXES**

# **ANNEXE 1**

**Consentement de prélèvement dans un but d'étude génétique et / ou de conservation dans une banque d'ADN chez une personne MAJEURE**

Identité du Médecin :

**Consentement de prélèvement dans un but d'étude génétique et / ou de conservation dans une banque d'ADN chez une personne MAJEURE**

Consentement établi en deux exemplaires dont un a été remis à l'intéressé(e).

Je soussigné(e).....  
né(e) le : .....  
demeurant à.....

accepte que soit effectué chez moi un prélèvement sanguin afin de faire réaliser une étude de génétique moléculaire, qui peut aider au diagnostic ou à la prévention de la maladie : .....  
pour moi-même ou des membres de ma famille, selon la proposition du Docteur .....

Dans certains cas, la connaissance de mon origine géographique est nécessaire pour orienter les études génétiques.

**J'en autorise le recueil.**

**Je n'en autorise pas le recueil.**

*(rayer la mention inutile)*

Cette étude sera faite à partir de l'ADN extrait du sang prélevé, qui sera conservé dans la banque d'ADN du *Laboratoire de Génétique Moléculaire et Cellulaire de l'Université des Sciences et de la Technologie d'Oran (LGMC-USTO)*. Dans certains cas, cette étude peut s'étendre sur plusieurs années. A tout moment je peux demander que mon ADN me soit restitué ou détruit.

Je déclare avoir été pleinement informé(e) de la nature des études qui seront effectuées et des conséquences éventuelles des résultats qui peuvent m'être donnés.

Ces résultats seront transmis uniquement au Docteur  
.....

**Je souhaite**

**Je ne souhaite pas**

**en être informé(e) par lui.**

*(rayer la mention inutile)*

Je consens au recueil, à la saisie et au traitement des données contenues dans mon dossier médical par des personnes tenues au secret professionnel. Les données qui me concernent resteront strictement confidentielles. Je n'en autorise la consultation qu'aux personnes qui collaborent à l'analyse. Je pourrai à tout moment demander que mon dossier soit détruit.

Conformément à la loi Informatique et Libertés, je peux exercer mon droit d'accès aux données qui me concernent par l'intermédiaire d'un médecin de mon choix.

L'utilisation éventuelle de mon ADN à des fins de recherche médicale, pour une étude non directement liée à celle à laquelle je consens ce jour, nécessitera un nouveau consentement de ma part.

Fait à :.....

Le :.....

**Signature du  
médecin recueillant  
le consentement**

**Signature de  
L'intéressé(e)**

# **ANNEXE 2**

## **Déclaration d'Helsinki**



**Déclaration d'Helsinki de L'AMM – Principes éthiques applicables à la  
recherche médicale impliquant des êtres humains**

Adoptée par la 18e Assemblée générale de l'AMM, Helsinki, Finlande, Juin 1964 et amendée par les :

29e Assemblée générale de l'AMM, Tokyo, Japon, Octobre 1975

35e Assemblée générale de l'AMM, Venise, Italie, Octobre 1983

41e Assemblée générale de l'AMM, Hong Kong, Septembre 1989

48e Assemblée générale de l'AMM, Somerset West, Afrique du Sud, Octobre 1996

52e Assemblée générale de l'AMM, Edimbourg, Ecosse, Octobre 2000

53e Assemblée générale de l'AMM, Washington DC, Etats Unis, Octobre 2002 (ajout d'une note de clarification)

55e Assemblée générale de l'AMM, Tokyo, Japon, Octobre 2004 (ajout d'une note de clarification)

59e Assemblée générale de l'AMM, Séoul, République de Corée, Octobre 2008

64e Assemblée générale de l'AMM, Fortaleza, Brésil, Octobre 2013.

## Préambule

L'Association Médicale Mondiale (AMM) a élaboré la Déclaration d'Helsinki comme un énoncé de principes éthiques applicables à la recherche médicale impliquant des êtres humains, y compris la recherche sur du matériel biologique humain et sur des données identifiables.

La Déclaration est conçue comme un tout indissociable. Chaque paragraphe doit être appliqué en tenant compte de tous les autres paragraphes pertinents.

Conformément au mandat de l'AMM, cette Déclaration s'adresse en priorité aux médecins. L'AMM invite cependant les autres personnes engagées dans la recherche médicale impliquant des êtres humains à adopter ces principes.

## Principes généraux

La Déclaration de Genève de l'AMM engage les médecins en ces termes : « la santé de mon patient prévaudra sur toutes les autres considérations » et le Code International d'Éthique Médicale déclare qu'un «médecin doit agir dans le meilleur intérêt du patient lorsqu'il le soigne».

Le devoir du médecin est de promouvoir et de sauvegarder la santé, le bien être et les droits des patients, y compris ceux des personnes impliquées dans la recherche médicale. Le médecin consacre son savoir et sa conscience à l'accomplissement de ce devoir.

Le progrès médical est basé sur la recherche qui, en fin de compte, doit impliquer des êtres humains.

L'objectif premier de la recherche médicale impliquant des êtres humains est de comprendre les causes, le développement et les effets des maladies et d'améliorer les interventions préventives, diagnostiques et thérapeutiques

(méthodes, procédures et traitements). Même les meilleures interventions éprouvées doivent être évaluées en permanence par des recherches portant sur leur sécurité, leur efficacité, leur pertinence, leur accessibilité et leur qualité.

La recherche médicale est soumise à des normes éthiques qui promeuvent et assurent le respect de tous les êtres humains et qui protègent leur santé et leurs droits.

Si l'objectif premier de la recherche médicale est de générer de nouvelles connaissances, cet objectif ne doit jamais prévaloir sur les droits et les intérêts des personnes impliquées dans la recherche.

Il est du devoir des médecins engagés dans la recherche médicale de protéger la vie, la santé, la dignité, l'intégrité, le droit à l'auto-détermination, la vie privée et la confidentialité des informations des personnes impliquées dans la recherche. La responsabilité de protéger les personnes impliquées dans la recherche doit toujours incomber à un médecin ou à un autre professionnel de santé et jamais aux personnes impliquées dans la recherche même si celles-ci ont donné leur consentement.

Dans la recherche médicale impliquant des êtres humains, les médecins doivent tenir compte des normes et standards éthiques, légaux et réglementaires applicables dans leur propre pays ainsi que des normes et standards internationaux. Les protections garanties par la présente Déclaration aux personnes impliquées dans la recherche ne peuvent être restreintes ou exclues par aucune disposition éthique, légale ou réglementaire, nationale ou internationale.

La recherche médicale devrait être conduite de sorte qu'elle réduise au minimum les nuisances éventuelles à l'environnement.

La recherche médicale impliquant des êtres humains doit être conduite uniquement par des personnes ayant acquis une éducation, une formation et des qualifications appropriées en éthique et en science. La recherche



impliquant des patients ou des volontaires en bonne santé nécessite la supervision d'un médecin ou d'un autre professionnel de santé qualifié et compétent.

Des possibilités appropriées de participer à la recherche médicale devraient être offertes aux groupes qui y sont sous-représentés.

Les médecins qui associent la recherche médicale à des soins médicaux devraient impliquer leurs patients dans une recherche uniquement dans la mesure où elle se justifie par sa valeur potentielle en matière de prévention, de diagnostic ou de traitement et si les médecins ont de bonnes raisons de penser que la participation à la recherche ne portera pas atteinte à la santé des patients concernés.

Une compensation et un traitement adéquats doivent être garantis pour les personnes qui auraient subi un préjudice en raison de leur participation à une recherche.

### **Risques, contraintes et avantages**

Dans la pratique médicale et la recherche médicale, la plupart des interventions comprennent des risques et des inconvénients.

Une recherche médicale impliquant des êtres humains ne peut être conduite que si l'importance de l'objectif dépasse les risques et inconvénients pour les personnes impliquées.

Toute recherche médicale impliquant des êtres humains doit préalablement faire l'objet d'une évaluation soigneuse des risques et des inconvénients prévisibles pour les personnes et les groupes impliqués, par rapport aux bénéfices prévisibles pour eux et les autres personnes ou groupes affectés par la pathologie étudiée.

Toutes les mesures destinées à réduire les risques doivent être mises en œuvre. Les risques doivent être constamment surveillés, évalués et documentés par le chercheur.

18. Les médecins ne peuvent pas s'engager dans une recherche impliquant des êtres humains sans avoir la certitude que les risques ont été correctement évalués et pourront être gérés de manière satisfaisante.

Lorsque les risques s'avèrent dépasser les bénéfices potentiels ou dès l'instant où des conclusions définitives ont été démontrées, les médecins doivent évaluer s'ils continuent, modifient ou cessent immédiatement une recherche.

### **Populations et personnes vulnérables**

Certains groupes ou personnes faisant l'objet de recherches sont particulièrement vulnérables et peuvent avoir une plus forte probabilité d'être abusés ou de subir un préjudice additionnel.

Tous les groupes et personnes vulnérables devraient bénéficier d'une protection adaptée.

La recherche médicale impliquant un groupe vulnérable se justifie uniquement si elle répond aux besoins ou aux priorités sanitaires de ce groupe et qu'elle ne peut être effectuée sur un groupe non vulnérable. En outre, ce groupe devrait bénéficier des connaissances, des pratiques ou interventions qui en résultent.

### **Exigences scientifiques et protocoles de recherche**

La recherche médicale impliquant des êtres humains doit se conformer aux principes scientifiques généralement acceptés, se baser sur une connaissance approfondie de la littérature scientifique, sur d'autres sources pertinentes d'informations et sur des expériences appropriées en laboratoire

et, le cas échéant, sur les animaux. Le bien être des animaux utilisés dans la recherche doit être respecté. La conception et la conduite de toutes les recherches impliquant des êtres humains doivent être clairement décrites et justifiées dans un protocole de recherche.

Ce protocole devrait contenir une déclaration sur les enjeux éthiques en question et indiquer comment les principes de la présente Déclaration ont été pris en considération. Le protocole devrait inclure des informations concernant le financement, les promoteurs, les affiliations institutionnelles, les conflits d'intérêts potentiels, les incitations pour les personnes impliquées dans la recherche et des informations concernant les mesures prévues pour soigner et/ou dédommager celles ayant subi un préjudice en raison de leur participation à la recherche.

Dans les essais cliniques, le protocole doit également mentionner les dispositions appropriées prévues pour l'accès à l'intervention testée après l'essai clinique.

### **Comités d'éthique de la recherche**

Le protocole de recherche doit être soumis au comité d'éthique de la recherche concerné pour évaluation, commentaires, conseils et approbation avant que la recherche ne commence. Ce comité doit être transparent dans son fonctionnement, doit être indépendant du chercheur, du promoteur et de toute autre influence indue et doit être dûment qualifié. Il doit prendre en considération les lois et réglementations du ou des pays où se déroule la recherche, ainsi que les normes et standards internationaux, mais ceux-ci ne doivent pas permettre de restreindre ou exclure l'une des protections garanties par la présente Déclaration aux personnes impliquées dans la recherche.

Le comité doit avoir un droit de suivi sur les recherches en cours. Le chercheur doit fournir au comité des informations sur le suivi, notamment

concernant tout évènement indésirable grave. Aucune modification ne peut être apportée au protocole sans évaluation et approbation par le comité. A la fin de la recherche, les chercheurs doivent soumettre au comité un rapport final contenant un résumé des découvertes et des conclusions de celle-ci.

### **Vie privée et confidentialité**

Toutes les précautions doivent être prises pour protéger la vie privée et la confidentialité des informations personnelles concernant les personnes impliquées dans la recherche.

### **Consentement éclairé**

La participation de personnes capables de donner un consentement éclairé à une recherche médicale doit être un acte volontaire. Bien qu'il puisse être opportun de consulter les membres de la famille ou les responsables de la communauté, aucune personne capable de donner un consentement éclairé ne peut être impliquée dans une recherche sans avoir donné son consentement libre et éclairé.

Dans la recherche médicale impliquant des personnes capables de donner un consentement éclairé, toute personne pouvant potentiellement être impliquée doit être correctement informée des objectifs, des méthodes, des sources de financement, de tout éventuel conflit d'intérêts, des affiliations institutionnelles du chercheur, des bénéfices escomptés et des risques potentiels de la recherche, des désagréments qu'elle peut engendrer, des mesures qui seront prises après à l'essai clinique et de tout autre aspect pertinent de la recherche.

La personne pouvant potentiellement être impliquée dans la recherche doit être informée de son droit de refuser d'y participer ou de s'en retirer à tout moment sans mesure de rétorsion. Une attention particulière devrait être accordée aux besoins d'informations spécifiques de chaque personne

pouvant potentiellement être impliquée dans la recherche ainsi qu'aux méthodes adoptées pour fournir les informations. Lorsque le médecin ou une autre personne qualifiée en la matière a la certitude que la personne concernée a compris les informations, il doit alors solliciter son consentement libre et éclairé, de préférence par écrit. Si le consentement ne peut pas être donné par écrit, le consentement non écrit doit être formellement documenté en présence d'un témoin.

Toutes les personnes impliquées dans des recherches médicales devraient avoir le choix d'être informées des conclusions générales et des résultats de celles-ci.

Lorsqu'il sollicite le consentement éclairé d'une personne pour sa participation à une recherche, le médecin doit être particulièrement attentif lorsque cette dernière est dans une relation de dépendance avec lui ou pourrait donner son consentement sous la contrainte. Dans ce cas, le consentement éclairé doit être sollicité par une personne qualifiée en la matière et complètement indépendante de cette relation.

Lorsque la recherche implique une personne incapable de donner un consentement éclairé, le médecin doit solliciter le consentement éclairé de son représentant légal. Les personnes incapables ne doivent pas être incluses dans une recherche qui n'a aucune chance de leur être bénéfique sauf si celle-ci vise à améliorer la santé du groupe qu'elles représentent, qu'elle ne peut pas être réalisée avec des personnes capables de donner un consentement éclairé et qu'elle ne comporte que des risques et des inconvénients minimes.

Lorsqu'une personne considérée comme incapable de donner un consentement éclairé est en mesure de donner son assentiment concernant sa participation à la recherche, le médecin doit solliciter cet assentiment en complément du consentement de son représentant légal. Le refus de la personne pouvant potentiellement être impliquée dans la recherche devrait être respecté.

La recherche impliquant des personnes physiquement ou mentalement incapables de donner leur consentement, par exemple des patients inconscients, peut être menée uniquement si l'état physique ou mental empêchant de donner un consentement éclairé est une caractéristique nécessaire du groupe sur lequel porte cette recherche.

Dans de telles circonstances, le médecin doit solliciter le consentement éclairé du représentant légal. En l'absence d'un représentant légal et si la recherche ne peut pas être retardée, celle-ci peut être lancée sans le consentement éclairé. Dans ce cas, le protocole de recherche doit mentionner les raisons spécifiques d'impliquer des personnes dont l'état les rend incapables de donner leur consentement éclairé et la recherche doit être approuvée par le comité d'éthique de la recherche concerné. Le consentement pour maintenir la personne concernée dans la recherche doit, dès que possible, être obtenu de la personne elle-même ou de son représentant légal.

Le médecin doit fournir des informations complètes au patient sur la nature des soins liés à la recherche. Le refus d'un patient de participer à une recherche ou sa décision de s'en retirer ne doit jamais nuire à la relation patient-médecin.

Pour la recherche médicale utilisant des tissus ou des données d'origine humaine, telles que les recherches sur tissus et données contenues dans les biobanques ou des dépôts similaires, les médecins doivent solliciter le consentement éclairé pour leur analyse, stockage et/ou réutilisation. Il peut se présenter des situations exceptionnelles où il est impraticable, voire impossible d'obtenir le consentement. Dans de telles situations, la recherche peut être entreprise uniquement après évaluation et approbation du comité d'éthique de la recherche concerné.

### Utilisation de placebo

Les bénéfices, les risques, les inconvénients, ainsi que l'efficacité d'une nouvelle intervention doivent être testés et comparés à ceux des meilleures interventions avérées, sauf dans les circonstances suivantes :

lorsqu'il n'existe pas d'intervention avérée, l'utilisation de placebo, ou la non intervention, est acceptable ; ou

lorsque pour des raisons de méthodologie incontournables et scientifiquement fondées l'utilisation de toute intervention moins efficace que la meilleure éprouvée, l'utilisation d'un placebo, ou la non intervention, est nécessaire afin de déterminer l'efficacité ou la sécurité d'une intervention,

et lorsque les patients recevant une intervention moins efficace que la meilleure éprouvée, un placebo, ou une non intervention, ne courent pas de risques supplémentaires de préjudices graves ou irréversibles du fait de n'avoir pas reçu la meilleure intervention éprouvée. Le plus grand soin doit être apporté afin d'éviter tout abus de cette option.

### Conditions de l'accès à l'intervention testée après l'essai clinique

En prévision d'un essai clinique, les promoteurs, les chercheurs et les gouvernements des pays d'accueil devraient prévoir des dispositions pour que tous les participants qui ont encore besoin d'une intervention identifiée comme bénéfique dans l'essai puissent y accéder après celui-ci. Cette information doit également être communiquée aux participants au cours du processus de consentement éclairé.

### Enregistrement des recherches, publication et dissémination des résultats

Toute recherche impliquant des êtres humains doit être enregistrée dans une banque de données accessible au public avant que ne soit recrutée la première personne impliquée dans la recherche.

Les chercheurs, auteurs, promoteurs, rédacteurs et éditeurs ont tous des obligations éthiques concernant la publication et la dissémination des résultats de la recherche. Les chercheurs ont le devoir de mettre à la disposition du public les résultats de leurs recherches impliquant des êtres humains. Toutes les parties ont la responsabilité de fournir des rapports complets et précis. Ils devraient se conformer aux directives acceptées en matière d'éthique pour la rédaction de rapports. Les résultats aussi bien négatifs et non concluants que positifs doivent être publiés ou rendus publics par un autre moyen. La publication doit mentionner les sources de financement, les affiliations institutionnelles et les conflits d'intérêts. Les rapports de recherche non-conformes aux principes de la présente Déclaration ne devraient pas être acceptés pour publication.

### Interventions non avérées dans la pratique clinique

Dans le cadre du traitement d'un patient, faute d'interventions avérées ou faute d'efficacité de ces interventions, le médecin, après avoir sollicité les conseils d'experts et avec le consentement éclairé du patient ou de son représentant légal, peut recourir à une intervention non avérée si, selon son appréciation professionnelle, elle offre une chance de sauver la vie, rétablir la santé ou alléger les souffrances du patient. Cette intervention devrait par la suite faire l'objet d'une recherche pour en évaluer la sécurité et l'efficacité. Dans tous les cas, les nouvelles informations doivent être enregistrées et, le cas échéant, rendues publiques.

*\*Les typos du paragraphe 26 et 32 ont été corrigés par le secretariat de l'AMM le 27 juin 2014.*

© *World Medical Association, Inc. - All Rights reserved.*

© *Asociación médica mundial - Todos los derechos reservados.* © *L'Association Médicale Mondiale - Tous droits réservés.*



## **ANNEXE 3**

**Fiche d'enquête (questionnaire) pour les patients atteints de  
Polyarthrite Rhumatoïde**

## Service de rhumatologie du CHU Oran

Étude d'association génétique de facteurs de prédisposition à la Polyarthrite Rhumatoïde PR  
dans la population de l'Ouest AlgérienFiche d'Identification du Patient

Date : ...../...../.....

Nom et prénom : .....

Sexe : homme  femme 

Date et lieu de naissance : ...../...../.....

N° du dossier : .....

N° du téléphone : .....

Médecin traitant : .....

Histoire de la maladie

Date de début de la maladie (PR): ...../...../.....

Symptomatologie : .....

Date du diagnostic de la PR : ...../...../.....

Date de la mise en route d'un traitement de fond (MTX) : ...../...../.....

Médicaments associés : .....

Mode de début : Aigue  Sub aigu  chronique Forme de début : Mono arthrite  Oligo arthrite  Poly arthrite Arthralgie  Symétrique  Asymétrique Signes associés : Fièvre  AEG 

Autre signes associés : .....

**Manifestations articulaires :**

Mono arthrite     oligo arthrite     poly arthrite     arthralgie

Localisation : .....

.....

.....

.....

Manifestations extra-articulaires :            Oui             Non

*Si oui*

préciser : .....

.....

.....

**Examens cliniques :**

Date de l'examen clinique : ...../...../.....

Poids : .....

Taille : .....

TA : .....

**Examens radiologiques :**

Date de l'examen radiologique : ...../...../.....

Score de Sharp modifié : .....

**Examens Biologique :**

Paramètres	valeurs	Date
VS		
CRP		
Plaquettes		
Créatinémie		
GB		
ALAT		
ASAT		
FR		
ACPA		

Evaluations de l'activité de la maladieAvant le traitement :DAS<sub>28</sub> : .....

N° de réveils nocturnes liés à la douleur : .....

Déroutillage matinal : .....

Après le traitement :

Doses du Méthotrexate : .....

La durée du traitement : .....

DAS<sub>28</sub> : .....

Nombre de réveils nocturnes liés à la douleur : .....

Déroutillage matinal : .....

En cas d'arrêt du traitement au Méthotrexate préciser la cause :
 Grossesse     allaitement     Intolérance     Inefficacité

Autre : .....

En cas d'intolérance :

Cité Le médicament remplaçant : .....

Quels sont les effets secondaires du ou Méthotrexate observé ?

Digestive		
Nausées :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Indigestion ou anorexie :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Diarrhée :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Ulcération de la bouche :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Douleurs abdominales :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>

Pulmonaire		
Présence de la toux :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Difficulté respiratoire :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Pneumopathie infectieuse :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Pneumopathie immunoallergique :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>

Autre : .....

Hépatique		
cytolysé hépatique :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Stéatohépatite métabolique :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Cirrhose :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Fibrose hépatique :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Taux de la transaminase :	.....	

Cutano-muqueuse		
Chute de cheveux :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Photosensibilité :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Erythème des extrémités :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
Eruption cutanée :	Oui <input type="checkbox"/>	Non <input type="checkbox"/>

Autre : .....

Rénale		Hématopoïétique (infectieux)	
insuffisance rénale :	Oui <input type="checkbox"/> Non <input type="checkbox"/>	Thrombopénie :	Oui <input type="checkbox"/> Non <input type="checkbox"/>
Autre :	.....	Leucopénie :	Oui <input type="checkbox"/> Non <input type="checkbox"/>
	.....	Autre :	.....

**Antécédents**

**Personnels :**

Autres Pathologies observées : .....

.....

.....

Traitements : .....

.....

.....

.....

Tabagisme : Oui  Non

**Familiaux :**

Parents atteints :  Aucun  Père  Mère

fratries atteintes  Aucun  Filles :{nombre} .....

Enfants atteints :  Autre  Filles :{nombre} .....

Garçons (nombre).....

Garçons (nombre).....

Autre : .....

.....

.....

**Signature du médecin**

# **ANNEXE 4**

## **Extraction d'ADN**

## 1 Protocole

L'extraction d'ADN est une étape incontournable pour chaque manipulation en biologie moléculaire.

Le protocole suivi est celui de la technique au NaCl aussi appelée « *salting out* » qui consiste à traiter uniquement le lysat cellulaire par une solution saline afin d'éliminer les protéines fixées à l'ADN (232). Cette technique se déroule en trois étapes :

- ***Lyse des hématies***

Au volume de sang (15 à 20 mL) prélevé sur EDTA, sont ajoutés 45mL de solution hypotonique du TE 10/10 [Tris/HCl (1M), EDTA (0,5M) ; pH=8]. Les tubes sont placés dans la glace pendant 30 minutes (choc thermique) et centrifugés à 3000 tours/minute durant 15 minutes.

Cette opération est renouvelée après élimination du surnageant mais le temps d'incubation dans la glace passe à 10 minutes au lieu de 30. Les lavages sont arrêtés lors de l'obtention d'un culot lymphocytaire blanchâtre.

- ***Lyse des lymphocytes***

Le culot lymphocytaire obtenu est additionné de 5mL de SLB (solution de lyse des globules blancs) [Tris/HCl (1M), EDTA (0,5M), SDS (10%) ; pH=8], ainsi que de 125µL d'enzyme protéinase K (20µg/mL). Cette dernière a pour rôle de digérer les protéines associées à l'ADN. Le mélange est mis en incubation au bain-marie à 37°C sous agitation douce pendant une nuit.

- **Précipitation de l'ADN**

Afin de précipiter l'ADN, 2mL de NaCl (5M) sont ajoutés, puis les tubes sont centrifugés pendant 15 minutes à 3000 tours/minute. Le surnageant obtenu, contenant la forme aqueuse de l'ADN, est récupéré dans un autre tube puis additionné de 2 à 3 volumes d'éthanol absolu froid. L'ADN condensé, est récupéré par une pipette pasteur scellée pour être lavé dans deux bains successifs d'éthanol (absolu puis 70%), puis resuspendu dans un tube *ependorf* contenant 200 à 500µL de solution TE 10/1.

Les tubes sont placés à température ambiante sous agitation afin de permettre une meilleure dissolution de l'ADN qui sera ensuite conservé à 4°C jusqu'à son utilisation.

## 2 Dosage de l'ADN

Le dosage est réalisé par un spectrophotomètre dans l'ultra violet, l'absorption doit être mesurée à 260nm afin de quantifier la concentration de l'ADN, ainsi qu'à 280nm dans le but d'évaluer le taux de protéines. Le rapport  $DO^{260}/DO^{280}$ , nous permettra de définir la pureté de l'ADN, qui doit être compris entre 1,5 et 2.

Une unité de densité optique correspond à l'absorption d'une solution d'ADN double brin à la concentration de 50µg/ml. L'évaluation de la concentration de l'ADN se fait selon la formule suivante : *Concentration de l'ADN (µg/ml) = facteur de dilution x 50 (1UDO) x DO260nm.*



# **ANNEXE 5**

**Génotypage par PCR en Temps Réel (TaqMan®)**

## 1 Principe

La PCR en temps réel repose sur la détection et la quantification d'un marqueur fluorescent au cours de la réaction d'amplification. Le signal fluorescent est directement proportionnel à la quantité de produits de PCR générés. En mesurant l'intensité de fluorescence émise à chaque cycle, il est possible de suivre la formation des produits de PCR pendant la phase exponentielle (phase au cours de laquelle la quantité de produits amplifiés est en corrélation directe avec la quantité initiale de matrice).

Les produits d'amplification peuvent être détectés selon deux grands principes : par marquage non spécifique avec des agents se liant à l'ADN double brin (*Sybr Green*) et par marquage spécifique du produit de PCR à l'aide d'une ou deux sonde(s) fluorescente(s) (Sondes *TaqMan*® ou balises moléculaires) (274).

Nous avons utilisé ici deux sondes discriminantes, VIC et FAM. Dans ce cas de figure, la PCR requiert deux amorces oligonucléotidiques pour amplifier la séquence d'intérêt et deux sondes *TaqMan*® : l'une s'hybridant à l'allèle sauvage, l'autre à l'allèle variant. Chaque sonde *TaqMan* renferme un fluorochrome à l'extrémité 5' spécifique de chaque type d'allèle (VIC ou FAM), un *quencher* non fluorescent (NFQ) à l'extrémité 3' et une molécule MGB (*Minor Groove Binder*) qui s'insère dans le petit sillon de la double hélice formée par le duplex ADN cible – sonde spécifique et le stabilise (**figure 13**).

Cette stabilisation du complexe ADN cible-sonde va de paire avec la température de fusion du complexe ( $T_m$ ). Ainsi, le  $T_m$  d'une sonde parfaitement hybridée avec le produit de PCR sera plus élevé que celui d'une sonde s'hybridant imparfaitement. Ce phénomène est accentué par l'utilisation du MGB qui va déstabiliser un duplex imparfait et donc faire chuter la  $T_m$ .

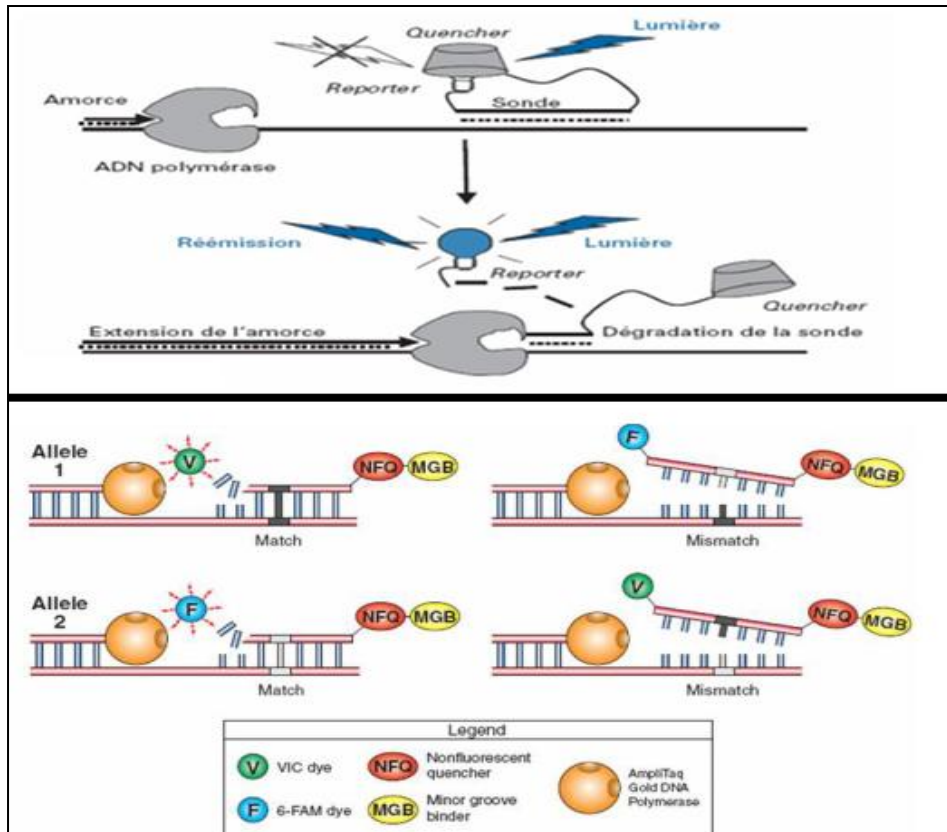


Figure 13 : Schéma explicatif du principe de la technologie TaqMan® (d'après le protocole TaqMan® SNP Genotyping Assays, Applied Biosystems).

Au début de l'élongation, les sondes spécifiques d'allèle s'hybrident spécifiquement à la matrice, entre les deux amorces nécessaires à l'amplification. La proximité du fluorochrome et du *quencher* empêche la détection de fluorescence. Au cours de l'étape d'élongation de la PCR, la Taq polymérase dégrade la sonde hybridée, ce qui entraîne une émission de fluorescence.

La discrimination allélique sera ensuite réalisée à l'aide d'un programme qui mesure la fluorescence à chaque cycle de la PCR et qui permet la discrimination entre les deux allèles présents dans les échantillons testés. Par exemple, si la sonde VIC est spécifique de l'allèle muté, une augmentation de la fluorescence VIC exclusivement indiquera une homozygotie pour l'allèle variant, une augmentation de FAM exclusivement rendra compte de la présence d'un homozygote sauvage ; et une

augmentation des 2 signaux de fluorescence marquera la présence d'un individu hétérozygote présentant chacun des deux allèles recherchés.

## 2 Protocole

### 2.1 Réactifs

- TaqMan Universal PCR Master Mix, Applied Biosystems, 2000 reactions, 50 ml
- Eau pour chromatographie, LiChrosolv, Merck
- Mixes de sondes TaqMan, TaqMan SNP Genotyping Assays (TaqMan MGB probes, FAM™ and VIC™ dye-labeled), spécifiques de chaque essai.

Les conditions de PCR sont les mêmes quelque soit le polymorphisme étudié. Seul le mixe avec les sondes est spécifique de chaque polymorphisme. Le protocole est résumé sur la **figure 14**.

### 2.2 Préparation des échantillons et du mixe de PCR

40 µL de chaque échantillon d'ADN (10 ng/µL) sont déposés dans des plaques 96 puits qui vont servir pour plusieurs PCR. Le mixe de PCR est préparé dans un godet de 40 ml. La composition pour un échantillon est :

- Mixe de sonde (40X) : 0,125 µL
- Mixe TaqMan (2X) : 2,5 µl
- Eau : 1,375 µL

On distribue ensuite de façon automatisée chaque échantillon d'ADN (1 µL) et le mixe de PCR (4 µL) dans une plaque 384 puits à l'aide d'un robot dispatcheur ( pipet robot, Biotek FX Stacker 10 Carousel, Beckman Coulter).

Les plaques sont ensuite fermées par une feuille thermocollante avant d'être placées dans le thermocycleur.

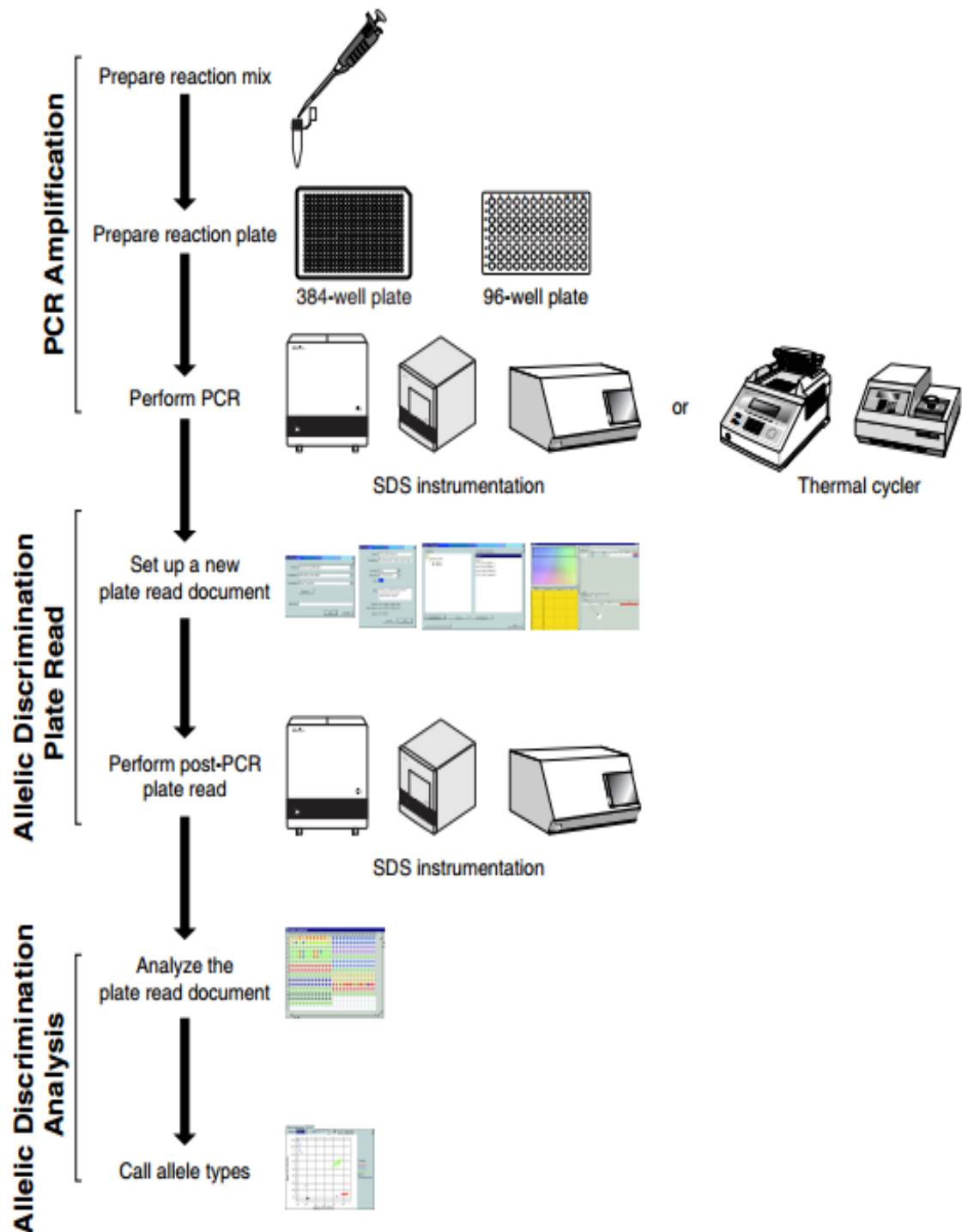


Figure 14 : Schéma résumant la procédure de génotypage utilisant la technologie TaqMan® (d'après le protocole TaqMan® SNP Genotyping Assays, *Applied Biosystems*).

### 2.3 Programme de PCR

Le programme de PCR est le même pour tous les polymorphismes, soit :

- 50°C, 2 min
- 95°C, 10 min
- Puis 55 fois : 95°C, 15 sec
- 60°C, 1 min
- 8°C, infini

La PCR est effectuée soit dans le thermocycleur 7500 HAT Abi Prism (*Applied Biosystems*), soit dans un thermocycleur Primus (*MWG AG Biotech*).

### 2.4 Lecture des résultats

La lecture de la fluorescence est faite dans le détecteur couplé au thermocycleur 7500 HAT Abi Prism et intégrée grâce au logiciel SDS 2.1 de l'ordinateur lié au détecteur (*Applied Biosystems*). Chaque échantillon est représenté par un point. On définit quel type de fluorescence est associé à chaque point (**figure 15**). Par exemple, le nuage de points exprimant majoritairement de la fluorescence FAM (bleu) correspond à des homozygotes pour l'un des deux allèles, le nuage rouge correspond à une fluorescence VIC majoritaire donc à des échantillons homozygotes pour l'autre allèle, tandis que le nuage vert correspond à une émission des deux fluorescences donc à des échantillons hétérozygotes pour le polymorphisme considéré. Les points noirs correspondent à une PCR qui n'a pas marché ou à un positionnement ambigu. Pour chaque mixe, le fournisseur précise à quel allèle, sauvage ou muté, est associé chacune des sondes VIC et FAM.

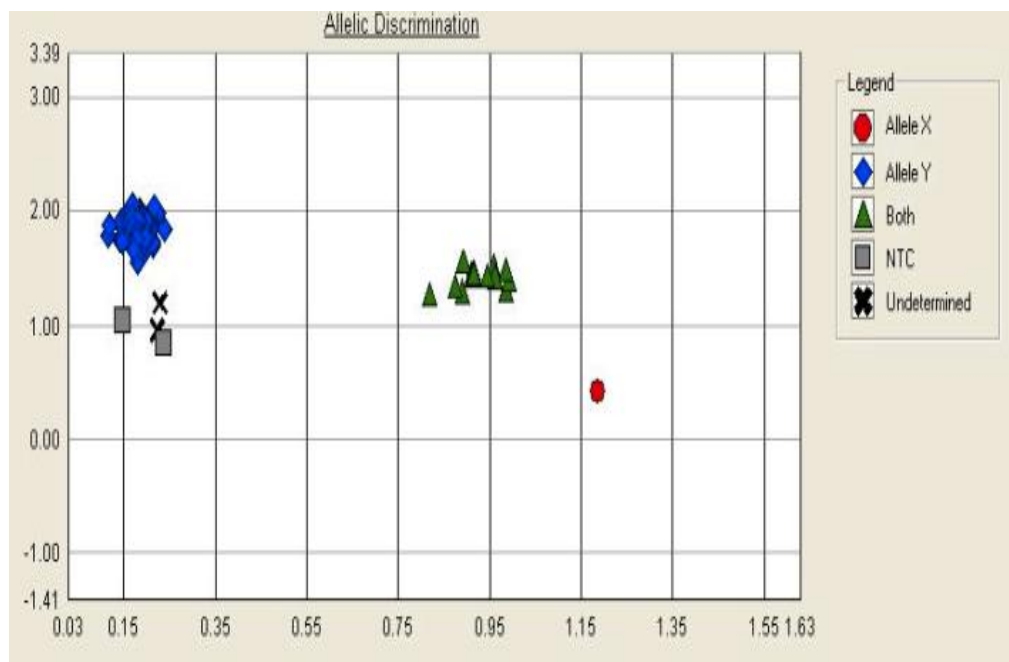


Figure 15 : Capture d'écran d'un résultat graphique de génotypage par TaqMan® obtenu par le logiciel SDS.

# **ANNEXE 6**

## **Génotypage par PCR-RFLP**



## 1 Principe

Cette technique est basée sur le fait que la présence de l'allèle muté crée ou entraîne la disparition (abolition) d'un site de coupure pour une enzyme de restriction donnée. On amplifie par PCR un segment d'ADN spécifique de la région d'intérêt contenant la mutation ponctuelle étudiée. Le produit d'amplification est ensuite digéré par une enzyme de restriction adéquate et les produits de digestion sont séparés par électrophorèse. Les profils de digestion permettent de constater la présence ou l'absence d'une coupure enzymatique du produit amplifié, révélant la présence ou l'absence du polymorphisme.

## 2 Choix des SNPs et des amorces

La condition pour retenir un SNP pour un génotypage par PCR-RFLP a été de faire en sorte que l'allèle fréquent de chaque SNP soit coupé au moins à deux reprises. Le choix des sites de restriction et de l'enzyme s'effectue sur le logiciel en ligne *Restriction Mapper* ([www.restrictionmapper.org](http://www.restrictionmapper.org)) alors que le choix des amorces se fait soit à partir des amorces décrites dans la littérature en cas de données publiées sur le gène, soit à partir de logiciels disponibles sur Internet (Primer3 : <http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>), soit manuellement. Dans ce cas, on choisit deux amorces sens et antisens d'environ 20 paires de bases chacune, flanquantes du polymorphisme avec pour objectif un fragment amplifié de 100 à 300 paires de bases, une température d'hybridation selon Wallace (4GC + 2AT) comprise entre 65 et 75°C et un taux de GC compris entre 45 et 55 %.

Pour chaque couple d'amorces, nous avons vérifié l'absence d'hybridation non spécifique en effectuant un alignement par l'utilitaire en ligne BLAST ([www.blast.ncbi.nlm.nih.gov](http://www.blast.ncbi.nlm.nih.gov)) sur l'ensemble du génome ainsi

qu'un alignement des 2 amorces l'une contre l'autre afin de s'assurer de l'absence d'auto-hybridation, les amorces non spécifiques étant éliminées.

L'optimisation des conditions d'amplification s'effectue à l'aide d'un gradient de PCR autour de la température d'hybridation soit  $\pm 3^{\circ}\text{C}$  (augmentation d'un demi-degré) soit  $\pm 6^{\circ}\text{C}$  (augmentation d'un degré). Une fois la PCR optimisée, on amplifie 3 couples de parents appartenant à des familles du CEPH et on génotype ces couples à la recherche du polymorphisme étudié. Si le polymorphisme n'est observé chez aucun de ces 6 individus, on en déduit que la fréquence de l'allèle mineur est inférieure à 16 % (1/12 chromosomes), donc l'étude de ce SNP n'est pas poursuivie. Si le polymorphisme est observé au moins une fois, on en déduit que le taux d'hétérozygotie est suffisant, alors le marqueur peut être testé sur les 200 familles trio de l'échantillon 1.

### 3 Protocole

#### 3.1 PCR

Pour le protocole expérimental, l'amplification par PCR est réalisée dans un volume total de 25  $\mu\text{l}$ . Le mélange réactionnel contient 1  $\mu\text{l}$  d'ADN génomique (10 ng), 0,75  $\mu\text{l}$  de  $\text{MgCl}_2$  (Ozyme, OZYA001, 50 mM), 0,5  $\mu\text{l}$  de Taq polymérase (Taq'Ozyme, 5U/ $\mu\text{l}$ , TAKARA), 2,5  $\mu\text{l}$  de dNTP à 2mM (SIGMA), 1  $\mu\text{l}$  de chaque amorce Forward (F)/Reverse (R) à 10 $\mu\text{M}$  (SIGMA), 2,5  $\mu\text{l}$  de tampon d'amplification (Ozyme, OZYA001, 10X) et un volume de 15,75  $\mu\text{l}$  d'eau distillée stérile.

Le protocole de PCR était le suivant : une première étape de dénaturation à  $96^{\circ}\text{C}$  pendant 10 minutes puis 35 cycles avec une étape de dénaturation à  $96^{\circ}\text{C}$  pendant 30 secondes, une étape d'hybridation à  $60^{\circ}\text{C}$  pendant 30 secondes et une étape d'élongation à  $72^{\circ}\text{C}$  pendant 1 minute, et, enfin, une étape finale d'élongation à  $72^{\circ}\text{C}$  pendant 10 minutes. La réaction d'amplification dure 1 heure et 53 minutes. Les amplifications sont exécutées

dans le thermocycleur (*Applied Biosystem® GeneAmp®*) sur un programme de PCR prédéfini.

### **3.2 Test d'amplification**

Les produits PCR obtenus sont déposés pour une migration sur un gel d'agarose à 2 %.

- **Préparation du gel**

Ce gel est préparé en pesant 1,0 g d'agarose qui est dissout dans 50 ml de tampon de charge TBE 1X (Tris : 0,089 M, EDTA : 0,25 mM, acide borique 0,089M, pH =8), puis chauffé pendant 2 minutes dans un four micro-onde. Il est refroidi et du bromure d'ethidium (BET), un intercalant de l'ADN qui fluoresce sous UV, est ensuite ajouté. Le mélange est coulé pour polymériser dans une cuve sur laquelle sont placés des peignes dont les empreintes sur le gel constitueront les puits de dépôt des échantillons. Après polymérisation complète, les peignes sont enlevés.

- **Dépôt des amplimères**

Le gel, ainsi polymérisé, est placé à l'intérieur d'une cuve d'électrophorèse et recouvert de tampon de migration TBE. Un mélange de 5 µl de produit d'amplification et de 5 µl de bleu de bromophénol (6X) et de sucrose 40% est déposé dans chaque puits. De plus, un marqueur de taille moléculaire de 100 paires de base est déposé sur le gel pour vérifier la taille de l'ADN amplifié.

- **Migration de l'ADN et visualisation**

La migration se fait à 120 Volts pendant 30 minutes. Le gel est déposé sur un transilluminateur et les fragments d'ADN amplifiés sont ainsi visualisés et ensuite photographiés. Cette visualisation est possible grâce au

BET qui est un cation s'intercalant entre les bases de l'ADN et ainsi le complexe ion éthidium/ADN devient fluorescent.

### **3.3 Digestion enzymatique (RFLP)**

Après avoir été amplifié, l'ADN est coupé par une enzyme de restriction au niveau d'un site reconnu spécifiquement. L'enzyme est choisie pour que le site de restriction soit situé au niveau de la mutation et permette ainsi de différencier l'ADN porteur de la mutation de l'ADN sauvage. Pour être un repère fiable, le site de restriction doit se trouver dans une région non répétée du génome. Une mutation modifiant la séquence primaire de l'ADN entraîne une modification de la carte de restriction (fragments supplémentaires ou absents).

- **Electrophorèse sur gel d'agarose**

Les fragments d'ADN obtenus sont séparés selon leur longueur par électrophorèse sur gel d'agarose à 2%. Ce dernier forme un maillage qui va être soumis à un courant électrique. Les fragments d'ADN vont migrer du négatif vers le positif. Plus ils sont gros, plus ils mettent longtemps à traverser le réseau d'agarose. Ces fragments sont ensuite révélés sous UV par mise en présence de BET (**figure 15**).

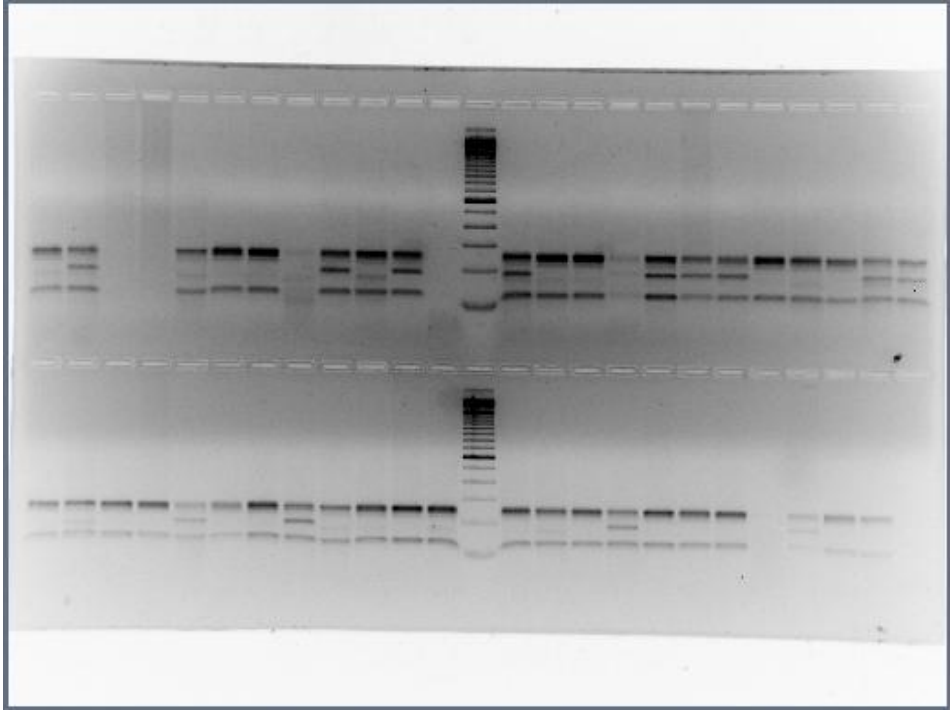


Figure 15 : Prise de vue du résultat de l'électrophorèse sur gel d'agarose à 2% pour les ADN de l'échantillon 1 concernant le gène LY96.

**REFERENCES  
BIBLIOGRAPHIQUES**

1. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*. 1987;50(3):509-17. Epub 1987/07/31.
2. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science*. 1989;245(4922):1073-80. Epub 1989/09/08.
3. Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, et al. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*. 1991;252(5009):1097-102. Epub 1991/05/24.
4. A comprehensive genetic linkage map of the human genome. NIH/CEPH Collaborative Mapping Group. *Science*. 1992;258(5079):67-86. Epub 1992/10/02.
5. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, et al. A genetic linkage map of the human genome. *Cell*. 1987;51(2):319-37. Epub 1987/10/23.
6. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*. 1996;380(6570):152-4. Epub 1996/03/14.
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921. Epub 2001/03/10.
8. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;291(5507):1304-51. Epub 2001/02/22.
9. Rao DC. Genetic dissection of complex traits: an overview. *Advances in genetics*. 2001;42:13-34. Epub 2000/10/19.
10. Giguere Y, Rousseau F. The genetics of osteoporosis: 'complexities and difficulties'. *Clinical genetics*. 2000;57(3):161-9. Epub 2000/04/27.
11. Giguere Y, Dodin S, Blanchet C, Morgan K, Rousseau F. The association between heel ultrasound and hormone replacement therapy is modulated by a two-locus vitamin D and estrogen receptor genotype. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2000;15(6):1076-84. Epub 2000/06/07.
12. Giguere Y, Dewailly E, Brisson J, Ayotte P, Laflamme N, Demers A, et al. Short polyglutamine tracts in the androgen receptor are protective against breast cancer in the general population. *Cancer research*. 2001;61(15):5869-74. Epub 2001/08/02.
13. Ponder BA. Cancer genetics. *Nature*. 2001;411(6835):336-41. Epub 2001/05/18.
14. Morton NE. Sequential tests for the detection of linkage. *American journal of human genetics*. 1955;7(3):277-318. Epub 1955/09/01.
15. Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J. Effects of misspecifying genetic parameters in lod score analysis. *Biometrics*. 1986;42(2):393-9. Epub 1986/06/01.
16. Risch N. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *American journal of human genetics*. 1990;46(2):242-53. Epub 1990/02/01.

17. Weber JL, Wong C. Mutation of human short tandem repeats. *Human molecular genetics*. 1993;2(8):1123-8. Epub 1993/08/01.
18. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, et al. A second-generation linkage map of the human genome. *Nature*. 1992;359(6398):794-801. Epub 1992/10/29.
19. Weissenbach J. Microsatellite polymorphisms and the genetic linkage map of the human genome. *Current opinion in genetics & development*. 1993;3(3):414-7. Epub 1993/06/01.
20. Brookes AJ. The essence of SNPs. *Gene*. 1999;234(2):177-86. Epub 1999/07/09.
21. Fan JB, Chee MS, Gunderson KL. Highly parallel genomic assays. *Nature reviews Genetics*. 2006;7(8):632-44. Epub 2006/07/19.
22. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nature genetics*. 2001;27(3):234-6. Epub 2001/03/10.
23. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature genetics*. 1999;22(3):231-8. Epub 1999/07/03.
24. Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E, et al. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature methods*. 2004;1(2):109-11. Epub 2005/03/23.
25. Ding C, Jin S. High-throughput methods for SNP genotyping. *Methods Mol Biol*. 2009;578:245-54. Epub 2009/09/22.
26. Ropers HH. New perspectives for the elucidation of genetic disorders. *American journal of human genetics*. 2007;81(2):199-207. Epub 2007/08/02.
27. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature*. 1996;379(6568):821-3. Epub 1996/02/29.
28. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444-54. Epub 2006/11/24.
29. Haldane J. The combination of linkage values and the calculation of distances between the loci of linked factors. *J Genet* 1919;8:299-309.
30. Kosambi D. The estimation of map distances from recombination values. *Ann Eugen*. 1944;12:172-5.
31. Gourraud PA, Barnette T, Vidan-Jeras B, Cambon-Thomsen A. Introduction to statistical analysis of population data in immunogenetics. *Transplant immunology*. 2005;14(3-4):245-53. Epub 2005/06/29.
32. Cochran WG. Methodological problems in the study of human populations. *Annals of the New York Academy of Sciences*. 1963;107:476-89. Epub 1963/05/22.
33. Szumilas M. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Académie canadienne de psychiatrie de l'enfant et de l'adolescent*. 2010;19(3):227-9. Epub 2010/09/16.
34. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American journal of human genetics*. 1993;52(3):506-16. Epub 1993/03/01.



35. Styrkarsdottir U, Cazier JB, Kong A, Rolfsson O, Larsen H, Bjarnadottir E, et al. Linkage of osteoporosis to chromosome 20p12 and association to BMP2. *PLoS biology*. 2003;1(3):E69. Epub 2003/12/24.
36. Deloukas P, Matthews LH, Ashurst J, Burton J, Gilbert JG, Jones M, et al. The DNA sequence and comparative analysis of human chromosome 20. *Nature*. 2001;414(6866):865-71. Epub 2002/01/10.
37. Rousseau F, Laflamme N. [Human molecular genetics: from monogenic to polygenic or complex disorders]. *Medecine sciences : M/S*. 2003;19(10):950-4. Epub 2003/11/13. *Genetique moleculaire humaine: des maladies monogeniques aux maladies complexes*.
38. Drumm ML, Konstan MW, Schluchter MD, Handler A, Pace R, Zou F, et al. Genetic modifiers of lung disease in cystic fibrosis. *The New England journal of medicine*. 2005;353(14):1443-53. Epub 2005/10/07.
39. Arkwright PD, Laurie S, Super M, Pravica V, Schwarz MJ, Webb AK, et al. TGF-beta(1) genotype and accelerated decline in lung function of patients with cystic fibrosis. *Thorax*. 2000;55(6):459-62. Epub 2000/05/19.
40. Salisbury BA, Pungliya M, Choi JY, Jiang R, Sun XJ, Stephens JC. SNP and haplotype variation in the human genome. *Mutation research*. 2003;526(1-2):53-61. Epub 2003/04/26.
41. Evans DM, Cardon LR. A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. *American journal of human genetics*. 2005;76(4):681-7. Epub 2005/02/19.
42. Sonnenberg A. [Bonferroni-Holm sequential test procedure]. *Zeitschrift fur Gastroenterologie*. 1985;23(12):703-4. Epub 1985/12/01. *Die sequentielle Testprozedur nach Bonferroni-Holm*.
43. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(45):19096-101. Epub 2009/10/29.
44. Brady G. Expression profiling of single mammalian cells--small is beautiful. *Yeast*. 2000;17(3):211-7. Epub 2000/10/12.
45. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nature reviews Genetics*. 2005;6(6):451-64. Epub 2005/05/11.
46. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature genetics*. 2000;24(4):372-6. Epub 2000/03/31.
47. Miller RA, Galecki A, Shmookler-Reis RJ. Interpretation, design, and analysis of gene array expression experiments. *The journals of gerontology Series A, Biological sciences and medical sciences*. 2001;56(2):B52-7. Epub 2001/02/24.
48. Gentleman R. *Bioinformatics and computational biology solutions using R and Bioconductor*. New York: Springer Science+Business Media; 2005. xix, 473 p. p.
49. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, et al. Population genomics of human gene expression. *Nature genetics*. 2007;39(10):1217-24. Epub 2007/09/18.

50. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-78. Epub 2007/06/08.
51. Stahl E, Raychaudhuri S, Remmers E, Xie G, Eyre S, Thomson B, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature genetics*. 2010;42(6):508-14.
52. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. Genetics of gene expression and its effect on disease. *Nature*. 2008;452(7186):423-8. Epub 2008/03/18.
53. Barrett JC. From HLA association to function. *Nature genetics*. 2012;44(3):235-6. Epub 2012/03/01.
54. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*. 2007;448(7152):470-3. Epub 2007/07/06.
55. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, et al. A genome-wide association study of global gene expression. *Nature genetics*. 2007;39(10):1202-7. Epub 2007/09/18.
56. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS genetics*. 2010;6(4):e1000895. Epub 2010/04/07.
57. Liang L, Morar N, Dixon AL, Lathrop GM, Abecasis GR, Moffatt MF, et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome research*. 2013;23(4):716-26. Epub 2013/01/25.
58. Luft FC. Twins in cardiovascular genetic research. *Hypertension*. 2001;37(2 Pt 2):350-6. Epub 2001/03/07.
59. Ding K, Kullo IJ. Genome-wide association studies for atherosclerotic vascular disease and its risk factors. *Circulation Cardiovascular genetics*. 2009;2(1):63-72. Epub 2009/09/15.
60. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM, Pajak A. Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation*. 1994;90(1):583-612. Epub 1994/07/01.
61. Guillemin F, Saraux A, Guggenbuhl P, Roux CH, Fardellone P, Le Bihan E, et al. Prevalence of rheumatoid arthritis in France: 2001. *Annals of the rheumatic diseases*. 2005;64(10):1427-30. Epub 2005/04/01.
62. Kaipainen-Seppanen O, Kautiainen H. Declining trend in the incidence of rheumatoid factor-positive rheumatoid arthritis in Finland 1980-2000. *The Journal of rheumatology*. 2006;33(11):2132-8. Epub 2006/10/03.
63. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis and rheumatism*. 2000;43(1):30-7. Epub 2000/01/22.
64. Ritchie DM, Boyle JA, McInnes JM, Jasani MK, Dalakas TG, Grieverson P, et al. Clinical studies with an articular index for the assessment

of joint tenderness in patients with rheumatoid arthritis. *The Quarterly journal of medicine*. 1968;37(147):393-406. Epub 1968/07/01.

65. Eshed I, Feist E, Althoff CE, Hamm B, Konen E, Burmester GR, et al. Tenosynovitis of the flexor tendons of the hand detected by MRI: an early indicator of rheumatoid arthritis. *Rheumatology (Oxford)*. 2009;48(8):887-91. Epub 2009/05/29.

66. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis and rheumatism*. 1988;31(3):315-24. Epub 1988/03/01.

67. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis and rheumatism*. 2010;62(9):2569-81. Epub 2010/09/28.

68. Firestein GS. Evolving concepts of rheumatoid arthritis. *Nature*. 2003;423(6937):356-61. Epub 2003/05/16.

69. Choy EH, Panayi GS. Cytokine pathways and joint inflammation in rheumatoid arthritis. *The New England journal of medicine*. 2001;344(12):907-16. Epub 2001/03/22.

70. Schmidt D, Goronzy JJ, Weyand CM. CD4+ CD7- CD28- T cells are expanded in rheumatoid arthritis and are characterized by autoreactivity. *J Clin Invest*. 1996;97(9):2027-37. Epub 1996/05/01.

71. van Esch WJ, Reparón-Schuijt CC, Hamstra HJ, van Kooten C, Logtenberg T, Breedveld FC, et al. Polyreactivity of human IgG Fc-binding phage antibodies constructed from synovial fluid CD38+ B cells of patients with rheumatoid arthritis. *Journal of autoimmunity*. 2002;19(4):241-50. Epub 2002/12/11.

72. Chini L, Bardare M, Cancrini C, Angelini F, Mancina L, Cortis E, et al. Evidence of clonotypic pattern of T-cell repertoire in synovial fluid of children with juvenile rheumatoid arthritis at the onset of the disease. *Scandinavian journal of immunology*. 2002;56(5):512-7. Epub 2002/11/02.

73. Falta MT, Fontenot AP, Rosloniec EF, Crawford F, Roark CL, Bill J, et al. Class II major histocompatibility complex-peptide tetramer staining in relation to functional avidity and T cell receptor diversity in the mouse CD4(+) T cell response to a rheumatoid arthritis-associated antigen. *Arthritis and rheumatism*. 2005;52(6):1885-96. Epub 2005/06/04.

74. Albert LJ, Inman RD. Molecular mimicry and autoimmunity. *The New England journal of medicine*. 1999;341(27):2068-74. Epub 1999/12/30.

75. Ang CW, Jacobs BC, Laman JD. The Guillain-Barre syndrome: a true case of molecular mimicry. *Trends in immunology*. 2004;25(2):61-6. Epub 2004/04/23.

76. Zabriskie JB, Freimer EH. An immunological relationship between the group A streptococcus and mammalian muscle. *The Journal of experimental medicine*. 1966;124(4):661-78. Epub 1966/10/01.

77. Balandraud N, Roudier J, Roudier C. Epstein-Barr virus and rheumatoid arthritis. *Autoimmunity reviews*. 2004;3(5):362-7. Epub 2004/08/04.

78. Lotz M, Roudier J. Epstein-Barr virus and rheumatoid arthritis: cellular and molecular aspects. *Rheumatology international*. 1989;9(3-5):147-52. Epub 1989/01/01.
79. Fox RI, Luppi M, Pisa P, Kang HI. Potential role of Epstein-Barr virus in Sjogren's syndrome and rheumatoid arthritis. *The Journal of rheumatology Supplement*. 1992;32:18-24. Epub 1992/01/01.
80. Inoue N, Harada S, Miyasaka N, Oya A, Yanagi K. Analysis of antibody titers to Epstein-Barr virus nuclear antigens in sera of patients with Sjogren's syndrome and with rheumatoid arthritis. *The Journal of infectious diseases*. 1991;164(1):22-8. Epub 1991/07/01.
81. Tosato G, Steinberg AD, Blaese RM. Defective EBV-specific suppressor T-cell function in rheumatoid arthritis. *The New England journal of medicine*. 1981;305(21):1238-43. Epub 1981/11/19.
82. Baboonian C, Venables PJ, Williams DG, Williams RO, Maini RN. Cross reaction of antibodies to a glycine/alanine repeat sequence of Epstein-Barr virus nuclear antigen-1 with collagen, cytokeratin, and actin. *Annals of the rheumatic diseases*. 1991;50(11):772-5. Epub 1991/11/01.
83. Roudier J, Petersen J, Rhodes GH, Luka J, Carson DA. Susceptibility to rheumatoid arthritis maps to a T-cell epitope shared by the HLA-Dw4 DR beta-1 chain and the Epstein-Barr virus glycoprotein gp110. *Proceedings of the National Academy of Sciences of the United States of America*. 1989;86(13):5104-8. Epub 1989/07/01.
84. Takeda T, Mizugaki Y, Matsubara L, Imai S, Koike T, Takada K. Lytic Epstein-Barr virus infection in the synovial tissue of patients with rheumatoid arthritis. *Arthritis and rheumatism*. 2000;43(6):1218-25. Epub 2000/06/17.
85. Anzilotti C, Merlini G, Pratesi F, Tommasi C, Chimenti D, Migliorini P. Antibodies to viral citrullinated peptide in rheumatoid arthritis. *The Journal of rheumatology*. 2006;33(4):647-51. Epub 2006/03/03.
86. Ekstrom K, Hjalgrim H, Brandt L, Baecklund E, Klareskog L, Ekbom A, et al. Risk of malignant lymphomas in patients with rheumatoid arthritis and in their first-degree relatives. *Arthritis and rheumatism*. 2003;48(4):963-70. Epub 2003/04/11.
87. Boissier MC, Assier E, Falgarone G, Bessis N. Shifting the imbalance from Th1/Th2 to Th17/treg: the changing rheumatoid arthritis paradigm. *Joint Bone Spine*. 2008;75(4):373-5. Epub 2008/06/24.
88. Brennan FM, McInnes IB. Evidence that cytokines play a role in rheumatoid arthritis. *J Clin Invest*. 2008;118(11):3537-45. Epub 2008/11/05.
89. Kurko J, Besenyei T, Laki J, Glant TT, Mikecz K, Szekanecz Z. Genetics of rheumatoid arthritis - a comprehensive review. *Clinical reviews in allergy & immunology*. 2013;45(2):170-9. Epub 2013/01/05.
90. Kooloos WM, Huizinga TW, Guchelaar HJ, Wessels JA. Pharmacogenetics in treatment of rheumatoid arthritis. *Current pharmaceutical design*. 2010;16(2):164-75. Epub 2010/03/09.
91. Cuchacovich M, Ferreira L, Aliste M, Soto L, Cuenca J, Cruzat A, et al. Tumour necrosis factor-alpha (TNF-alpha) levels and influence of -308 TNF-alpha promoter polymorphism on the responsiveness to infliximab in patients with rheumatoid arthritis. *Scandinavian journal of rheumatology*. 2004;33(4):228-32. Epub 2004/09/17.

92. Mugnier B, Balandraud N, Darque A, Roudier C, Roudier J, Reviron D. Polymorphism at position -308 of the tumor necrosis factor alpha gene influences outcome of infliximab therapy in rheumatoid arthritis. *Arthritis and rheumatism*. 2003;48(7):1849-52. Epub 2003/07/09.
93. Padyukov L, Lampa J, Heimbürger M, Ernestam S, Cederholm T, Lundkvist I, et al. Genetic markers for the efficacy of tumour necrosis factor blocking therapy in rheumatoid arthritis. *Annals of the rheumatic diseases*. 2003;62(6):526-9. Epub 2003/05/22.
94. Danila MI, Hughes LB, Bridges SL. Pharmacogenetics of etanercept in rheumatoid arthritis. *Pharmacogenomics*. 2008;9(8):1011-5. Epub 2008/08/07.
95. Davila L, Ranganathan P. Pharmacogenetics: implications for therapy in rheumatic diseases. *Nat Rev Rheumatol*. 2011;7(9):537-50. Epub 2011/08/10.
96. Cronstein BN. Pharmacogenetics in the rheumatic diseases, from pre-a-porter to haute couture. *Nature clinical practice Rheumatology*. 2006;2(1):2-3. Epub 2006/08/26.
97. Lee YH, Ji JD, Song GG. Associations between FCGR3A polymorphisms and susceptibility to rheumatoid arthritis: a metaanalysis. *The Journal of rheumatology*. 2008;35(11):2129-35. Epub 2008/10/10.
98. Berkun Y, Levartovsky D, Rubinow A, Orbach H, Aamar S, Grenader T, et al. Methotrexate related adverse effects in patients with rheumatoid arthritis are associated with the A1298C polymorphism of the MTHFR gene. *Annals of the rheumatic diseases*. 2004;63(10):1227-31. Epub 2004/09/14.
99. Dervieux T, Kremer J, Lein DO, Capps R, Barham R, Meyer G, et al. Contribution of common polymorphisms in reduced folate carrier and gamma-glutamylhydrolase to methotrexate polyglutamate levels in patients with rheumatoid arthritis. *Pharmacogenetics*. 2004;14(11):733-9. Epub 2004/11/27.
100. van Ede AE, Laan RF, Blom HJ, Huizinga TW, Haagsma CJ, Giesendorf BA, et al. The C677T mutation in the methylenetetrahydrofolate reductase gene: a genetic risk factor for methotrexate-related elevation of liver enzymes in rheumatoid arthritis patients. *Arthritis and rheumatism*. 2001;44(11):2525-30. Epub 2001/11/17.
101. Oliver JE, Silman AJ. Risk factors for the development of rheumatoid arthritis. *Scandinavian journal of rheumatology*. 2006;35(3):169-74. Epub 2006/06/13.
102. Heliövaara M, Aho K, Aromaa A, Knekt P, Reunanen A. Smoking and risk of rheumatoid arthritis. *The Journal of rheumatology*. 1993;20(11):1830-5. Epub 1993/11/01.
103. Pedersen M, Jacobsen S, Klarlund M, Pedersen BV, Wiik A, Wohlfahrt J, et al. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. *Arthritis research & therapy*. 2006;8(4):R133. Epub 2006/07/29.
104. Pedersen M, Jacobsen S, Garred P, Madsen HO, Klarlund M, Svejgaard A, et al. Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control

- study in Denmark. *Arthritis and rheumatism*. 2007;56(5):1446-53. Epub 2007/05/01.
105. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. *Lancet*. 2009;373(9664):659-72. Epub 2009/01/23.
106. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Annals of the rheumatic diseases*. 2003;62(9):835-41. Epub 2003/08/19.
107. Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *The American journal of medicine*. 2006;119(6):503 e1-9. Epub 2006/06/06.
108. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis and rheumatism*. 2006;54(1):38-46. Epub 2005/12/31.
109. Liao KP, Alfredsson L, Karlson EW. Environmental influences on risk for rheumatoid arthritis. *Current opinion in rheumatology*. 2009;21(3):279-83. Epub 2009/03/26.
110. Kallberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, Garred P, et al. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. *Annals of the rheumatic diseases*. 2009;68(2):222-7. Epub 2008/06/07.
111. Pikwer M, Bergstrom U, Nilsson JA, Jacobsson L, Berglund G, Turesson C. Breast feeding, but not use of oral contraceptives, is associated with a reduced risk of rheumatoid arthritis. *Annals of the rheumatic diseases*. 2009;68(4):526-30. Epub 2008/05/15.
112. Lawrence JS, Ball J. Genetic studies on rheumatoid arthritis. *Annals of the rheumatic diseases*. 1958;17(2):160-8. Epub 1958/06/01.
113. Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A, et al. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *British journal of rheumatology*. 1993;32(10):903-7. Epub 1993/10/01.
114. Wandstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nature immunology*. 2001;2(9):802-9. Epub 2001/08/30.
115. Stastny P. Mixed lymphocyte cultures in rheumatoid arthritis. *The Journal of clinical investigation*. 1976;57(5):1148-57.
116. Stastny P. Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. *The New England journal of medicine*. 1978;298(16):869-71.
117. Woodrow J, Nichol F, Zaphiropoulos G. DR antigens and rheumatoid arthritis: a study of two populations. *British medical journal (Clinical research ed)*. 1981;283(6302):1287-8.
118. Imboden JB. The immunopathogenesis of rheumatoid arthritis. *Annual review of pathology*. 2009;4:417-34. Epub 2008/10/29.

119. Gregersen P, Silver J, Winchester R. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis and rheumatism*. 1987;30(11):1205-13.
120. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature genetics*. 2012;44(3):291-6. Epub 2012/01/31.
121. Fodil M, Boudjema A, Benhamamouch S, Benzaoui M, Eliaou J, Sayer D, et al. Etude de l'association HLA-DRB1 avec la polyarthrite rhumatoïde dans la population de l'Ouest Algérien. *Algerian Journal of Medicine*. 2008;4:146-8.
122. Dieudé P, Cornélis F. Genetic basis of rheumatoid arthritis. *Joint Bone Spine*. 2005;72(6):520-6.
123. Stahl EA, Raychaudhuri S. Rheumatoid arthritis. Evidence for a genetic component to disease severity in RA. *Nat Rev Rheumatol*. 2012;8(6):312-3. Epub 2012/06/01.
124. Bowes J, Barton A. Recent advances in the genetics of RA susceptibility. *Rheumatology (Oxford, England)*. 2008;47(4):399-402.
125. Dieudé P, Dawidowicz K. Déterminisme des pathologies auto-immunes : facteurs génétiques et environnementaux. *Revue du Rhumatisme Monographies*. 2010.
126. Julia A, Marsal S. The Genetic Architecture of Rheumatoid Arthritis: From Susceptibility to Clinical Subphenotype Associations. *Current topics in medicinal chemistry*. 2013.
127. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis association at 6q23. *Nature genetics*. 2007;39(12):1431-3.
128. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nature genetics*. 2007;39(12):1477-82. Epub 2007/11/06.
129. Plenge R, Seielstad M, Padyukov L, Lee A, Remmers E, Ding B, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. *The New England journal of medicine*. 2007;357(12):1199-209.
130. Gregersen P, Amos C, Lee A, Lu Y, Remmers E, Kastner D, et al. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nature genetics*. 2009;41(7):820-3.
131. Kochi Y, Okada Y, Suzuki A, Ikari K, Terao C, Takahashi A, et al. A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nature genetics*. 2010;42(6):515-9.
132. Lee Y, Bae S-C, Choi S, Ji J, Song G. Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. *Molecular biology reports*. 2012;39(12):10627-35.
133. Freudenberg J, Lee H-S, Han B-G, Shin H, Kang Y, Sung Y-K, et al. Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis and rheumatism*. 2011;63(4):884-93.

134. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014;506(7488):376-81. Epub 2014/01/07.
135. Lee H-S, Korman B, Le J, Kastner D, Remmers E, Gregersen P, et al. Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations. *Arthritis and rheumatism*. 2009;60(2):364-71.
136. Cornélis F. Génétique de la polyarthrite rhumatoïde : un tournant décisif. *Revue du Rhumatisme Monographies*. 2010.
137. Michou L, Lasbleiz S, Rat A-C, Migliorini P, Balsa A, Westhovens R, et al. Linkage proof for PTPN22, a rheumatoid arthritis susceptibility gene and a human autoimmunity gene. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104(5):1649-54.
138. Burn GL, Svensson L, Sanchez-Blanco C, Saini M, Cope AP. Why is PTPN22 a good candidate susceptibility gene for autoimmune disease? *FEBS letters*. 2011;585(23):3689-98. Epub 2011/04/26.
139. Morgan A, Robinson J, Conaghan P, Martin S, Hensor E, Morgan M, et al. Evaluation of the rheumatoid arthritis susceptibility loci HLA-DRB1, PTPN22, OLIG3/TNFAIP3, STAT4 and TRAF1/C5 in an inception cohort. *Arthritis research & therapy*. 2010;12(2).
140. Kochi Y, Suzuki A, Yamada R, Yamamoto K. Ethnogenetic heterogeneity of rheumatoid arthritis-implications for pathogenesis. *Nat Rev Rheumatol*. 2010;6(5):290-5.
141. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nature genetics*. 2004;36(4):337-8. Epub 2004/03/09.
142. Orozco G, Rueda B, Martin J. Genetic basis of rheumatoid arthritis. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2006;60(10):656-62. Epub 2006/10/24.
143. Berthelot JM, Le Goff B. Rheumatoid arthritis and periodontal disease. *Joint Bone Spine*. 2010;77(6):537-41. Epub 2010/07/22.
144. Goriely S, Neurath MF, Goldman M. How microorganisms tip the balance between interleukin-12 family members. *Nature reviews Immunology*. 2008;8(1):81-6. Epub 2007/12/18.
145. Gu E, Lu J, Xing D, Chen X, Xie H, Liang J, et al. Rs7574865 polymorphism in signal transducers and activators of transcription 4 gene and rheumatoid arthritis: an updated meta-analysis of 28 case-control comparisons. *International journal of rheumatic diseases*. 2014. Epub 2014/04/23.
146. Mitchell AL, Macarthur KD, Gan EH, Baggott LE, Wolff AS, Skinningsrud B, et al. Association of autoimmune Addison's disease with alleles of STAT4 and GATA3 in European cohorts. *PloS one*. 2014;9(3):e88991. Epub 2014/03/13.
147. Zheng J, Yin J, Huang R, Petersen F, Yu X. Meta-analysis reveals an association of STAT4 polymorphisms with systemic autoimmune disorders and anti-dsDNA antibody. *Human immunology*. 2013;74(8):986-92. Epub 2013/05/01.



148. Remmers E, Plenge R, Lee A, Graham R, Hom G, Behrens T, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *The New England journal of medicine*. 2007;357(10):977-86.
149. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nature genetics*. 2009;41(11):1228-33. Epub 2009/10/20.
150. Malissen B, Ewbank JJ. 'TaiLoRing' the response of dendritic cells to pathogens. *Nature immunology*. 2005;6(8):749-50. Epub 2005/07/22.
151. Barnes BJ, Moore PA, Pitha PM. Virus-specific activation of a novel interferon regulatory factor, IRF-5, results in the induction of distinct interferon alpha genes. *The Journal of biological chemistry*. 2001;276(26):23382-90. Epub 2001/04/17.
152. Barnes BJ, Field AE, Pitha-Rowe PM. Virus-induced heterodimer formation between IRF-5 and IRF-7 modulates assembly of the IFNA enhanceosome in vivo and transcriptional activity of IFNA genes. *The Journal of biological chemistry*. 2003;278(19):16630-41. Epub 2003/02/26.
153. Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechler EC, et al. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104(16):6758-63. Epub 2007/04/07.
154. Rhodes B, Vyse TJ. The genetics of SLE: an update in the light of genome-wide association studies. *Rheumatology (Oxford)*. 2008;47(11):1603-11. Epub 2008/07/10.
155. Sigurdsson S, Padyukov L, Kurreeman FA, Liljedahl U, Wiman AC, Alfredsson L, et al. Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. *Arthritis and rheumatism*. 2007;56(7):2202-10. Epub 2007/06/30.
156. Dawidowicz K, Allanore Y, Guedj M, Pierlot C, Bombardieri S, Balsa A, et al. The interferon regulatory factor 5 gene confers susceptibility to rheumatoid arthritis and influences its erosive phenotype. *Annals of the rheumatic diseases*. 2011;70(1):117-21. Epub 2010/10/29.
157. Muller T, Anlag K, Wildner H, Britsch S, Treier M, Birchmeier C. The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. *Genes & development*. 2005;19(6):733-43. Epub 2005/03/17.
158. Ding L, Takebayashi H, Watanabe K, Ohtsuki T, Tanaka KF, Nabeshima Y, et al. Short-term lineage analysis of dorsally derived Olig3 cells in the developing spinal cord. *Developmental dynamics : an official publication of the American Association of Anatomists*. 2005;234(3):622-32. Epub 2005/09/08.
159. Musone S, Taylor K, Nititham J, Chu C, Poon A, Liao W, et al. Sequencing of TNFAIP3 and association of variants with multiple autoimmune diseases. *Genes and immunity*. 2011;12(3):176-82.
160. Ramos PS, Criswell LA, Moser KL, Comeau ME, Williams AH, Pajewski NM, et al. A comprehensive analysis of shared loci between systemic lupus erythematosus (SLE) and sixteen autoimmune diseases

- reveals limited genetic overlap. *PLoS genetics*. 2011;7(12):e1002406. Epub 2011/12/17.
161. Kurreeman F, Padyukov L, Marques R, Schrodi S, Seddighzadeh M, Stoeken-Rijsbergen G, et al. A Candidate Gene Approach Identifies the TRAF1/C5 Region as a Risk Factor for Rheumatoid Arthritis. *PLOS Medicine*. 2007.
162. Nishimoto K, Kochi Y, Ikari K, Yamamoto K, Suzuki A, Shimane K, et al. Association study of TRAF1-C5 polymorphisms with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in Japanese. *Annals of the rheumatic diseases*. 2010;69(2):368-73.
163. Xu K, Peng H, Zhou M, Wang W, Li R, Zhu KK, et al. Association study of TRAF1/C5 polymorphism (rs10818488) with susceptibility to rheumatoid arthritis and systemic lupus erythematosus: a meta-analysis. *Gene*. 2013;517(1):46-54. Epub 2013/01/17.
164. Shiozawa S, Hayashi S, Tsukamoto Y, Goko H, Kawasaki H, Wada T, et al. Identification of the gene loci that predispose to rheumatoid arthritis. *International immunology*. 1998;10(12):1891-5. Epub 1999/01/14.
165. Bradley J, Thiru S, Pober J. Disparate localization of 55-kd and 75-kd tumor necrosis factor receptors in human endothelial cells. *Am J Pathol*. 1995;146(1):27-32.
166. Galve-de Rochemonteix B, Nicod L, Dayer J. Tumor necrosis factor soluble receptor 75: the principal receptor form released by human alveolar macrophages and monocytes in the presence of interferon gamma. *American journal of respiratory cell and molecular biology*. 1996;14(3):279-87.
167. Dempsey P, Doyle S, He J, Cheng G. The signaling adaptors and pathways activated by TNF superfamily. *Cytokine & growth factor reviews*. 2003;14(3-4):193-209.
168. Pantelidis P, Lympany P, Foley P, Fanning G, Welsh K, du Bois R. Polymorphic analysis of the high-affinity tumor necrosis factor receptor 2. *Tissue antigens*. 1999;54(6):585-91.
169. Bayley JP, Bakker A, Kaijzel E, Huizinga T, Verweij C. Association of polymorphisms of the tumour necrosis factor receptors I and II and rheumatoid arthritis. *Rheumatology (Oxford, England)*. 2003;42(8):969-71.
170. Barton A, John S, Ollier W, Silman A, Worthington J. Association between rheumatoid arthritis and polymorphism of tumor necrosis factor receptor II, but not tumor necrosis factor receptor I, in Caucasians. *Arthritis Rheum*. 2001;44(1):61-5.
171. Daha N, Kurreeman F, Marques R, Stoeken-Rijsbergen G, Verduijn W, Huizinga T, et al. Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis. *Arthritis and rheumatism*. 2009;60(5):1255-60.
172. Kurreeman F, Liao K, Chibnik L, Hickey B, Stahl E, Gainer V, et al. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *American journal of human genetics*. 2011;88(1):57-69.
173. Orozco G, Eyre S, Hinks A, Ke X, Wellcome Trust Case Control consortium YC, Wilson A, et al. Association of CD40 with rheumatoid arthritis

confirmed in a large UK case-control study. *Annals of the rheumatic diseases*. 2010;69(5):813-6.

174. Suzuki A, Yamada R, Chang X, Tokuhira S, Sawada T, Suzuki M, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nature genetics*. 2003;34(4):395-402.

175. Lee Y, Rho Y, Choi S, Ji J, Song G. PADI4 polymorphisms and rheumatoid arthritis susceptibility: a meta-analysis. *Rheumatology international*. 2007;27(9):827-33.

176. Burr M, Naseem H, Hinks A, Eyre S, Gibbons L, Bowes J, et al. PADI4 genotype is not associated with rheumatoid arthritis in a large UK Caucasian population. *Annals of the rheumatic diseases*. 2010;69(4):666-70.

177. Teixeira VH, Olaso R, Martin-Magniette ML, Lasbleiz S, Jacq L, Oliveira CR, et al. Transcriptome analysis describing new immunity and defense genes in peripheral blood mononuclear cells of rheumatoid arthritis patients. *PloS one*. 2009;4(8):e6803. Epub 2009/08/28.

178. Viriyakosol S, Tobias PS, Kitchens RL, Kirkland TN. MD-2 binds to bacterial lipopolysaccharide. *The Journal of biological chemistry*. 2001;276(41):38044-51. Epub 2001/08/14.

179. Nagai Y, Akashi S, Nagafuku M, Ogata M, Iwakura Y, Akira S, et al. Essential role of MD-2 in LPS responsiveness and TLR4 distribution. *Nature immunology*. 2002;3(7):667-72. Epub 2002/06/11.

180. Brentano F, Kyburz D, Schorr O, Gay R, Gay S. The role of Toll-like receptor signalling in the pathogenesis of arthritis. *Cellular immunology*. 2005;233(2):90-6. Epub 2005/06/21.

181. Ospelt C, Brentano F, Rengel Y, Stanczyk J, Kolling C, Tak PP, et al. Overexpression of toll-like receptors 3 and 4 in synovial tissue from patients with early rheumatoid arthritis: toll-like receptor expression in early and longstanding arthritis. *Arthritis and rheumatism*. 2008;58(12):3684-92. Epub 2008/11/28.

182. Foell D, Roth J. Proinflammatory S100 proteins in arthritis and autoimmune disease. *Arthritis and rheumatism*. 2004;50(12):3762-71. Epub 2004/12/14.

183. Youssef P, Roth J, Frosch M, Costello P, Fitzgerald O, Sorg C, et al. Expression of myeloid related proteins (MRP) 8 and 14 and the MRP8/14 heterodimer in rheumatoid arthritis synovial membrane. *The Journal of rheumatology*. 1999;26(12):2523-8. Epub 1999/12/22.

184. Foell D, Kane D, Bresnihan B, Vogl T, Nacken W, Sorg C, et al. Expression of the pro-inflammatory protein S100A12 (EN-RAGE) in rheumatoid and psoriatic arthritis. *Rheumatology (Oxford)*. 2003;42(11):1383-9. Epub 2003/07/02.

185. Frosch M, Strey A, Vogl T, Wulffraat NM, Kuis W, Sunderkotter C, et al. Myeloid-related proteins 8 and 14 are specifically secreted during interaction of phagocytes and activated endothelium and are useful markers for monitoring disease activity in pauciarticular-onset juvenile rheumatoid arthritis. *Arthritis and rheumatism*. 2000;43(3):628-37. Epub 2000/03/23.

186. Kane D, Roth J, Frosch M, Vogl T, Bresnihan B, FitzGerald O. Increased perivascular synovial membrane expression of myeloid-related

- proteins in psoriatic arthritis. *Arthritis and rheumatism*. 2003;48(6):1676-85. Epub 2003/06/10.
187. Boneca IG. Mammalian PGRPs in the spotlight. *Cell host & microbe*. 2009;5(2):109-11. Epub 2009/02/17.
188. Cho JH, Fraser IP, Fukase K, Kusumoto S, Fujimoto Y, Stahl GL, et al. Human peptidoglycan recognition protein S is an effector of neutrophil-mediated innate immunity. *Blood*. 2005;106(7):2551-8. Epub 2005/06/16.
189. Saha S, Qi J, Wang S, Wang M, Li X, Kim YG, et al. PGLYRP-2 and Nod2 are both required for peptidoglycan-induced arthritis and local inflammation. *Cell host & microbe*. 2009;5(2):137-50. Epub 2009/02/17.
190. Cohen MM, Jr. Perspectives on RUNX genes: an update. *American journal of medical genetics Part A*. 2009;149A(12):2629-46. Epub 2009/10/16.
191. Subramaniam MM, Chan JY, Yeoh KG, Quek T, Ito K, Salto-Tellez M. Molecular pathology of RUNX3 in human carcinogenesis. *Biochimica et biophysica acta*. 2009;1796(2):315-31. Epub 2009/08/18.
192. Egawa T. Runx and ThPOK: a balancing act to regulate thymocyte lineage commitment. *Journal of cellular biochemistry*. 2009;107(6):1037-45. Epub 2009/05/30.
193. Inoue K, Shiga T, Ito Y. Runx transcription factors in neuronal development. *Neural development*. 2008;3:20. Epub 2008/08/30.
194. Lee YS, Lee JW, Jang JW, Chi XZ, Kim JH, Li YH, et al. Runx3 inactivation is a crucial early event in the development of lung adenocarcinoma. *Cancer cell*. 2013;24(5):603-16. Epub 2013/11/16.
195. Chen HX, Wang S, Wang Z, Zhang ZP, Shi SS. Overexpression of RUNX3 inhibits malignant behaviour of Eca109 cells in vitro and vivo. *Asian Pacific journal of cancer prevention : APJCP*. 2014;15(4):1531-7. Epub 2014/03/20.
196. Brenner O, Levanon D, Negreanu V, Golubkov O, Fainaru O, Woolf E, et al. Loss of Runx3 function in leukocytes is associated with spontaneously developed colitis and gastric mucosal hyperplasia. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(45):16016-21. Epub 2004/10/30.
197. Fainaru O, Shseyov D, Hantisteanu S, Groner Y. Accelerated chemokine receptor 7-mediated dendritic cell migration in Runx3 knockout mice and the spontaneous development of asthma-like disease. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(30):10598-603. Epub 2005/07/20.
198. Dominguez-Soto A, Relloso M, Vega MA, Corbi AL, Puig-Kroger A. RUNX3 regulates the activity of the CD11a and CD49d integrin gene promoters. *Immunobiology*. 2005;210(2-4):133-9. Epub 2005/09/17.
199. Woolf E, Brenner O, Goldenberg D, Levanon D, Groner Y. Runx3 regulates dendritic epidermal T cell development. *Developmental biology*. 2007;303(2):703-14. Epub 2007/01/16.
200. Brady G, Farrell PJ. RUNX3-mediated repression of RUNX1 in B cells. *Journal of cellular physiology*. 2009;221(2):283-7. Epub 2009/07/16.
201. Brady G, Whiteman HJ, Spender LC, Farrell PJ. Downregulation of RUNX1 by RUNX3 requires the RUNX3 VWRPY sequence and is essential

- for Epstein-Barr virus-driven B-cell proliferation. *Journal of virology*. 2009;83(13):6909-16. Epub 2009/05/01.
202. Setoguchi R, Tachibana M, Naoe Y, Muroi S, Akiyama K, Tezuka C, et al. Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development. *Science*. 2008;319(5864):822-5. Epub 2008/02/09.
203. Ohno S, Sato T, Kohu K, Takeda K, Okumura K, Satake M, et al. Runx proteins are involved in regulation of CD122, Ly49 family and IFN-gamma expression during NK cell differentiation. *International immunology*. 2008;20(1):71-9. Epub 2007/11/16.
204. Djuretic IM, Levanon D, Negreanu V, Groner Y, Rao A, Ansel KM. Transcription factors T-bet and Runx3 cooperate to activate Ifng and silence Il4 in T helper type 1 cells. *Nature immunology*. 2007;8(2):145-53. Epub 2007/01/02.
205. Lian Z, Chai W, Shi LL, Chen C, Liu J, Wang Y. Analysis of PPARGC1B, RUNX3 and TBKBP1 polymorphisms in Chinese Han patients with ankylosing spondylitis: a case-control study. *PloS one*. 2013;8(4):e61527. Epub 2013/05/03.
206. Yang D, Rosenberg HF, Chen Q, Dyer KD, Kurosaka K, Oppenheim JJ. Eosinophil-derived neurotoxin (EDN), an antimicrobial protein with chemotactic activities for dendritic cells. *Blood*. 2003;102(9):3396-403. Epub 2003/07/12.
207. Yang D, Chen Q, Su SB, Zhang P, Kurosaka K, Caspi RR, et al. Eosinophil-derived neurotoxin acts as an alarmin to activate the TLR2-MyD88 signal pathway in dendritic cells and enhances Th2 immune responses. *The Journal of experimental medicine*. 2008;205(1):79-90. Epub 2008/01/16.
208. Bovin LF, Rieneck K, Workman C, Nielsen H, Sorensen SF, Skjodt H, et al. Blood cell gene expression profiling in rheumatoid arthritis. Discriminative genes and effect of rheumatoid factor. *Immunology letters*. 2004;93(2-3):217-26. Epub 2004/05/26.
209. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nature genetics*. 2008;40(10):1156-9. Epub 2008/09/17.
210. Ruysseu-Witrand A, Lukas C, Nigon D, Dawidowicz K, Morel J, Sibilia J, et al. Association of IL-2RA and IL-2RB genes with erosive status in early rheumatoid arthritis patients (ESPOIR and RMP cohorts). *Joint Bone Spine*. 2013. Epub 2013/11/10.
211. Baechler EC, Batliwalla FM, Reed AM, Peterson EJ, Gaffney PM, Moser KL, et al. Gene expression profiling in human autoimmunity. *Immunological reviews*. 2006;210:120-37. Epub 2006/04/21.
212. Nzeusseu Toukap A, Galant C, Theate I, Maudoux AL, Lories RJ, Houssiau FA, et al. Identification of distinct gene expression profiles in the synovium of patients with systemic lupus erythematosus. *Arthritis and rheumatism*. 2007;56(5):1579-88. Epub 2007/05/01.
213. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403(6769):503-11. Epub 2000/02/17.

214. Brutsche MH, Brutsche IC, Wood P, Brass A, Morrison N, Rattay M, et al. Apoptosis signals in atopy and asthma measured with cDNA arrays. *Clinical and experimental immunology*. 2001;123(2):181-7. Epub 2001/02/24.
215. Mandel M, Achiron A. Gene expression studies in systemic lupus erythematosus. *Lupus*. 2006;15(7):451-6. Epub 2006/08/11.
216. Henriksen PA, Kotelevtsev Y. Application of gene expression profiling to cardiovascular disease. *Cardiovascular research*. 2002;54(1):16-24. Epub 2002/06/14.
217. Colangelo V, Schurr J, Ball MJ, Pelaez RP, Bazan NG, Lukiw WJ. Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *Journal of neuroscience research*. 2002;70(3):462-73. Epub 2002/10/23.
218. Gladkevich A, Nelemans SA, Kauffman HF, Korf J. Microarray profiling of lymphocytes in internal diseases with an altered immune response: potential and methodology. *Mediators of inflammation*. 2005;2005(6):317-30. Epub 2006/02/21.
219. Toonen EJ, Barrera P, Radstake TR, van Riel PL, Scheffer H, Franke B, et al. Gene expression profiling in rheumatoid arthritis: current concepts and future directions. *Annals of the rheumatic diseases*. 2008;67(12):1663-9. Epub 2008/02/06.
220. Toonen EJ, Gilissen C, Franke B, Kievit W, Eijsbouts AM, den Broeder AA, et al. Validation study of existing gene expression signatures for anti-TNF treatment in patients with rheumatoid arthritis. *PloS one*. 2012;7(3):e33199. Epub 2012/03/30.
221. Lindberg J, Wijbrandts CA, van Baarsen LG, Nader G, Klareskog L, Catrina A, et al. The gene expression profile in the synovium as a predictor of the clinical response to infliximab treatment in rheumatoid arthritis. *PloS one*. 2010;5(6):e11310. Epub 2010/07/02.
222. Soto H, Hevezi P, Roth RB, Pahuja A, Alleva D, Acosta HM, et al. Gene array analysis comparison between rat collagen-induced arthritis and human rheumatoid arthritis. *Scandinavian journal of immunology*. 2008;68(1):43-57. Epub 2008/05/20.
223. Qingchun H, Runyue H, LiGang J, Yongliang C, Song W, Shujing Z. Comparison of the expression profile of apoptosis-associated genes in rheumatoid arthritis and osteoarthritis. *Rheumatology international*. 2008;28(7):697-701. Epub 2008/02/16.
224. Sorensen LK, Havemose-Poulsen A, Sonder SU, Bendtzen K, Holmstrup P. Blood cell gene expression profiling in subjects with aggressive periodontitis and chronic arthritis. *Journal of periodontology*. 2008;79(3):477-85. Epub 2008/03/05.
225. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature*. 2006;439(7074):353-7. Epub 2005/11/08.
226. Gerhold DL, Jensen RV, Gullans SR. Better therapeutics through microarrays. *Nature genetics*. 2002;32 Suppl:547-51. Epub 2002/11/28.
227. Holleman A, Cheok MH, den Boer ML, Yang W, Veerman AJ, Kazemier KM, et al. Gene-expression patterns in drug-resistant acute

- lymphoblastic leukemia cells and response to treatment. *The New England journal of medicine*. 2004;351(6):533-42. Epub 2004/08/06.
228. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*. 2002;415(6870):436-42. Epub 2002/01/25.
229. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *The New England journal of medicine*. 2002;347(25):1999-2009. Epub 2002/12/20.
230. Tanino M, Matoba R, Nakamura S, Kameda H, Amano K, Okayama T, et al. Prediction of efficacy of anti-TNF biologic agent, infliximab, for rheumatoid arthritis patients using a comprehensive transcriptome analysis of white blood cells. *Biochemical and biophysical research communications*. 2009;387(2):261-5. Epub 2009/07/07.
231. Mesko B, Poliska S, Szamosi S, Szekanecz Z, Podani J, Varadi C, et al. Peripheral blood gene expression and IgG glycosylation profiles as markers of tocilizumab treatment in rheumatoid arthritis. *The Journal of rheumatology*. 2012;39(5):916-28. Epub 2012/04/03.
232. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research*. 1988;16(3):1215. Epub 1988/02/11.
233. Lathrop GM. Estimating genotype relative risks. *Tissue antigens*. 1983;22(2):160-6. Epub 1983/08/01.
234. Woolf B. On estimating the relation between blood group and disease. *Annals of human genetics*. 1955;19(4):251-3. Epub 1955/06/01.
235. Clerget-Darpoux F, Babron M, Prum B, Lathrop G, Deschamps I, Hors J. A new method to test genetic models in HLA associated diseases: the MASC method. *Annals of human genetics*. 1988;52(Pt 3):247-58.
236. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*. 1980;32(3):314-31. Epub 1980/05/01.
237. Lake SL, Blacker D, Laird NM. Family-based tests of association in the presence of linkage. *American journal of human genetics*. 2000;67(6):1515-25. Epub 2000/11/04.
238. Michou L, Croiseau P, Petit-Teixeira E, du Montcel ST, Lemaire I, Pierlot C, et al. Validation of the reshaped shared epitope HLA-DRB1 classification in rheumatoid arthritis. *Arthritis research & therapy*. 2006;8(3):R79. Epub 2006/05/02.
239. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *American journal of human genetics*. 2004;75(2):330-7. Epub 2004/06/23.
240. Hinks A, Barton A, John S, Bruce I, Hawkins C, Griffiths CE, et al. Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: further support that PTPN22 is an

- autoimmunity gene. *Arthritis and rheumatism*. 2005;52(6):1694-9. Epub 2005/06/04.
241. Seldin MF, Shigeta R, Laiho K, Li H, Saila H, Savolainen A, et al. Finnish case-control and family studies support PTPN22 R620W polymorphism as a risk factor in rheumatoid arthritis, but suggest only minimal or no effect in juvenile idiopathic arthritis. *Genes Immun*. 2005;6(8):720-2. Epub 2005/08/19.
242. Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. *American journal of human genetics*. 2005;77(6):1044-60. Epub 2005/12/29.
243. Pierer M, Kaltenhauser S, Arnold S, Wahle M, Baerwald C, Hantzschel H, et al. Association of PTPN22 1858 single-nucleotide polymorphism with rheumatoid arthritis in a German cohort: higher frequency of the risk allele in male compared to female patients. *Arthritis research & therapy*. 2006;8(3):R75. Epub 2006/04/26.
244. Zhernakova A, Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP. Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. *Genes Immun*. 2005;6(6):459-61. Epub 2005/05/06.
245. Orozco G, Sanchez E, Gonzalez-Gay MA, Lopez-Nevot MA, Torres B, Caliz R, et al. Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis and rheumatism*. 2005;52(1):219-24. Epub 2005/01/11.
246. van Oene M, Wintle RF, Liu X, Yazdanpanah M, Gu X, Newman B, et al. Association of the lymphoid tyrosine phosphatase R620W variant with rheumatoid arthritis, but not Crohn's disease, in Canadian populations. *Arthritis and rheumatism*. 2005;52(7):1993-8. Epub 2005/06/30.
247. Ikari K, Momohara S, Inoue E, Tomatsu T, Hara M, Yamanaka H, et al. Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. *Rheumatology (Oxford)*. 2006;45(11):1345-8. Epub 2006/05/13.
248. Chabchoub G, Teixeira EP, Maalej A, Ben Hamad M, Bahloul Z, Cornelis F, et al. The R620W polymorphism of the protein tyrosine phosphatase 22 gene in autoimmune thyroid diseases and rheumatoid arthritis in the Tunisian population. *Annals of human biology*. 2009;36(3):342-9. Epub 2009/04/04.
249. Sfar I, Dhaouadi T, Habibi I, Abdelmoula L, Makhoul M, Ben Romdhane T, et al. Functional polymorphisms of PTPN22 and FcγR genes in Tunisian patients with rheumatoid arthritis. *Archives de l'Institut Pasteur de Tunis*. 2009;86(1-4):51-62. Epub 2009/01/01.
250. Zervou MI, Sidiropoulos P, Petraki E, Vazgiourakis V, Krasoudaki E, Raptopoulou A, et al. Association of a TRAF1 and a STAT4 gene polymorphism with increased risk for rheumatoid arthritis in a genetically homogeneous population. *Human immunology*. 2008;69(9):567-71. Epub 2008/07/16.



251. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, et al. Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Human molecular genetics*. 2008;17(15):2274-9. Epub 2008/04/25.
252. Lee HS, Remmers EF, Le JM, Kastner DL, Bae SC, Gregersen PK. Association of STAT4 with rheumatoid arthritis in the Korean population. *Mol Med*. 2007;13(9-10):455-60. Epub 2007/10/13.
253. Palomino-Morales RJ, Rojas-Villarraga A, Gonzalez CI, Ramirez G, Anaya JM, Martin J. STAT4 but not TRAF1/C5 variants influence the risk of developing rheumatoid arthritis and systemic lupus erythematosus in Colombians. *Genes Immun*. 2008;9(4):379-82. Epub 2008/04/25.
254. Orozco G, Alizadeh BZ, Delgado-Vega AM, Gonzalez-Gay MA, Balsa A, Pascual-Salcedo D, et al. Association of STAT4 with rheumatoid arthritis: a replication study in three European populations. *Arthritis and rheumatism*. 2008;58(7):1974-80. Epub 2008/06/26.
255. Dieguez-Gonzalez R, Calaza M, Perez-Pampin E, Balsa A, Blanco FJ, Canete JD, et al. Analysis of TNFAIP3, a feedback inhibitor of nuclear factor-kappaB and the neighbor intergenic 6q23 region in rheumatoid arthritis susceptibility. *Arthritis research & therapy*. 2009;11(2):R42. Epub 2009/03/19.
256. Ben Hamad M, Cornelis F, Maalej A, Petit-Teixeira E. A Tunisian case-control association study of a 6q polymorphism in rheumatoid arthritis. *Rheumatology international*. 2012;32(6):1849-50. Epub 2011/08/13.
257. Maalej A, Hamad MB, Rebai A, Teixeira VH, Bahloul Z, Marzouk S, et al. Association of IRF5 gene polymorphisms with rheumatoid arthritis in a Tunisian population. *Scandinavian journal of rheumatology*. 2008;37(6):414-8. Epub 2008/08/30.
258. Garnier S, Dieude P, Michou L, Barbet S, Tan A, Lasbleiz S, et al. IRF5 rs2004640-T allele, the new genetic factor for systemic lupus erythematosus, is not associated with rheumatoid arthritis. *Annals of the rheumatic diseases*. 2007;66(6):828-31. Epub 2006/12/13.
259. Pope RM. Apoptosis as a therapeutic tool in rheumatoid arthritis. *Nature reviews Immunology*. 2002;2(7):527-35. Epub 2002/07/03.
260. Zhang Y, Qiu H, Zhang H, Wang L, Zhuang C, Liu R. Vascular endothelial growth factor A (VEGFA) polymorphisms in Chinese patients with rheumatoid arthritis. *Scandinavian journal of rheumatology*. 2013;42(5):344-8. Epub 2013/07/16.
261. van der Pouw Kraan TC, Wijbrandts CA, van Baarsen LG, Voskuyl AE, Rustenburg F, Baggen JM, et al. Rheumatoid arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients. *Annals of the rheumatic diseases*. 2007;66(8):1008-14. Epub 2007/01/16.
262. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS one*. 2010;5(5):e10693. Epub 2010/05/27.
263. Maxwell JR, Gowers IR, Kuet KP, Barton A, Worthington J, Wilson AG. Expression of the autoimmunity associated TNFAIP3 is increased in

- rheumatoid arthritis but does not differ according to genotype at 6q23. *Rheumatology (Oxford)*. 2012;51(8):1514-5. Epub 2012/05/25.
264. Teixeira VH, Jacq L, Lasbleiz S, Hilliquin P, Oliveira CR, Cornelis F, et al. Genetic and expression analysis of CASP7 gene in a European Caucasian population with rheumatoid arthritis. *The Journal of rheumatology*. 2008;35(10):1912-8. Epub 2008/09/12.
265. Munoz-Valle JF, Ruiz-Quezada SL, Oregon-Romero E, Navarro-Hernandez RE, Castaneda-Saucedo E, De la Cruz-Mosso U, et al. PAI-1 mRNA expression and plasma level in rheumatoid arthritis: relationship with 4G/5G PAI-1 polymorphism. *Rheumatology international*. 2012;32(12):3951-6. Epub 2011/12/28.
266. Myouzen K, Kochi Y, Okada Y, Terao C, Suzuki A, Ikari K, et al. Functional variants in NFKBIE and RTKN2 involved in activation of the NF-kappaB pathway are associated with rheumatoid arthritis in Japanese. *PLoS genetics*. 2012;8(9):e1002949. Epub 2012/10/03.
267. Chatzikiyakidou A, Voulgari PV, Georgiou I, Drosos AA. Altered sequence of the ETS1 transcription factor may predispose to rheumatoid arthritis susceptibility. *Scandinavian journal of rheumatology*. 2013;42(1):11-4. Epub 2012/10/30.
268. Lu X, Wang M, Qi J, Wang H, Li X, Gupta D, et al. Peptidoglycan recognition proteins are a new class of human bactericidal proteins. *The Journal of biological chemistry*. 2006;281(9):5895-907. Epub 2005/12/16.
269. Royet J, Dziarski R. Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. *Nature reviews Microbiology*. 2007;5(4):264-77. Epub 2007/03/17.
270. Dziarski R, Gupta D. Review: Mammalian peptidoglycan recognition proteins (PGRPs) in innate immunity. *Innate immunity*. 2010;16(3):168-74. Epub 2010/04/27.
271. Rohatgi A, Ayers CR, Khera A, McGuire DK, Das SR, Matulevicius S, et al. The association between peptidoglycan recognition protein-1 and coronary and peripheral atherosclerosis: Observations from the Dallas Heart Study. *Atherosclerosis*. 2009;203(2):569-75. Epub 2008/09/09.
272. Temoin S, Chakaki A, Askari A, El-Halaby A, Fitzgerald S, Marcus RE, et al. Identification of oral bacterial DNA in synovial fluid of patients with arthritis with native and failed prosthetic joints. *Journal of clinical rheumatology : practical reports on rheumatic & musculoskeletal diseases*. 2012;18(3):117-21. Epub 2012/03/20.
273. Hitchon CA, El-Gabalawy HS. Infection and rheumatoid arthritis: still an open question. *Current opinion in rheumatology*. 2011;23(4):352-7. Epub 2011/05/03.
274. Le Morvan V, Formento JL, Milano G, Bonnet J, Robert J. Techniques de recherche des polymorphismes génétiques. *Oncologie*. 2005;7(1):7-16.

## RÉSUMÉ

Nous nous sommes intéressés à l'étude génétique de la PR, maladie auto-immune, chronique et inflammatoire affectant principalement les articulations et entraînant leur destruction. La PR est une pathologie multifactorielle où le système HLA semble être le principal, mais non unique, facteur de prédisposition génétique.

Pour la première partie de notre étude, nous nous sommes intéressés à la recherche d'association avec la PR, dans la population Algérienne, pour des polymorphismes de gènes déjà confirmés dans d'autres populations. Cette approche a permis d'établir une preuve solide d'association pour les deux SNPs les plus importants PTPN22<sub>rs2476601</sub> et STAT4<sub>rs7574865</sub>. De plus amples études avec des échantillons élargis et l'inclusion de nouveaux polymorphismes de gènes candidats permettraient d'établir de nouvelles preuves d'association génétique avec la PR dans la population Algérienne.

Dans un second temps, nous nous sommes intéressés à l'analyse familiale d'association génétique de polymorphismes de gènes présentant un différentiel d'expression dans la PR. Cette partie de l'étude représente un complément d'information pour la compréhension du fonctionnement de ces gènes et de leur implication dans la PR. Ce travail d'analyse d'association génétique a été complété par l'étude de la relation entre les génotypes des SNPs étudiés et le taux d'expression des gènes considérés.

Les polymorphismes d'intérêts mis en évidence au cours de cette thèse représentent des cibles de choix pour l'analyse conjointe de la PR dans les deux populations algérienne et française en vu d'une comparaison ethnique des différents facteurs de prédisposition génétique à la PR.

**MOTS-CLÉS** : Polyarthrite Rhumatoïde, transcriptome, gènes candidats, PTPN22, PGLYRP1.