

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

CPIE : UN NOUVEL ALGORITHME STATISTIQUE CALCULANT LA
PROBABILITÉ DE TROUVER UN DONNEUR COMPATIBLE BASÉ SUR
L'APPARIEMENT DES ÉPITOPES DU COMPLEXE MAJEUR
D'HISTOCOMPATIBILITÉ EN TRANSPLANTATION RÉNALE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

JENNAT HOUDA LAMSATFI

FÉVRIER 2022

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je voudrais tout d'abord adresser toute ma gratitude au directeur de ce mémoire, Mr Karim Oualkacha, professeur au département de mathématiques à l'université du Québec à Montréal, pour sa grande patience, sa disponibilité, son support et surtout ses conseils judicieux, qui ont contribué à alimenter ma réflexion.

J'aimerais aussi remercier ma co-directrice, Mme Ruth Sapir-Pichhadze, professeur agrégée de médecine au département de médecine, à l'université de McGill et Dr en transplantation rénale au département de néphrologie au centre universitaire de santé McGill, pour sa disponibilité et son soutien.

Je désire également remercier tous les membres du laboratoire du Dr Sapir qui se sont engagés avec moi, en particulier Sahar Parto et Amélie Bourdieu, pour leur aide.

Je tiens à remercier sincèrement les membres du jury, Mme Marie-Hélène Descary et Mr Mamadou Yauck, professeurs au département de mathématiques à l'université de Québec à Montréal, qui me font le grand honneur d'évaluer ce travail.

Enfin, j'adresse mes plus sincères remerciements à mes chers parents et en particulier mon mari, qui a cru en moi, soutenu et encouragé tout au long de la réalisation de ce mémoire.

TABLE DES MATIÈRES

LISTE DES FIGURES	v
LISTE DES TABLEAUX	viii
RÉSUMÉ	x
INTRODUCTION	1
CHAPITRE I CONCEPTS DE GÉNÉTIQUE	3
1.1 L'ADN	3
1.2 Les gènes	5
1.3 Division cellulaire	6
1.4 Recombinaison génétique	8
1.5 Les marqueurs génétiques	10
1.6 L'équilibre de Hardy-Weinberg	11
1.7 Déséquilibre de liaison	13
CHAPITRE II IMMUNOLOGIE ET LE SYSTÈME CMH	18
2.1 Le système immunitaire	19
2.2 Le complexe majeur d'histocompatibilité (CMH)	20
2.3 Les principales cellules du système immunitaire	24
2.4 Réaction immunitaire adaptative	28
2.4.1 Réaction immunitaire à médiation humorale	28
2.4.2 Différenciation des lymphocytes B	30
2.4.3 Réaction immunitaire à médiation cellulaire	31
2.4.4 Différenciation des lymphocytes T	32
2.5 La transplantation rénale	35
2.5.1 La transplantation rénale et la compatibilité du groupe sanguin (ABO) du donneur-receveur	36

2.5.2	Nomenclature du système CMH	38
2.5.3	La compatibilité épitopique CMH (ou HLA) en transplantation rénale	39
2.5.4	Relation entre les allèles CMH et les épitopes	41
CHAPITRE III L'ALGORITHME DE L'APPARIEMENT DES ÉPITOPES (CPIE)		43
3.1	cPIE : Approche de compatibilité épitypique	43
3.1.1	Le calcul de la compatibilité épitypique en se basant sur les épitypes présents dans le bassin des donneurs	44
3.1.2	Le calcul de la compatibilité épitypique en se basant sur les fréquences des génotypes dans le bassin des donneurs	50
3.2	L'allocation d'organes et le cPIE	54
CHAPITRE IV SIMULATIONS ET RÉSULTATS		55
4.1	Simulations et préparation des données	55
4.2	Application de l'algorithme cPIE	56
4.2.1	Analyse des médianes	59
4.2.2	Analyse des médianes à l'aide du bootstrap non paramétrique	64
4.2.3	Analyse des fonctions de répartition des ethnicités	67
CONCLUSION		72
ANNEXE A INTRODUCTION À L'IMMUNOLOGIE		74
A.1	Relation entre les allèles HLA et les épitopes.	75
ANNEXE B L'ALGORITHME DE L'APPARIEMENT DES ÉPITOPES (CPIE)		84
B.1	Les étapes de l'algorithme cPIE	84
RÉFÉRENCES		86
RÉFÉRENCES		86

LISTE DES FIGURES

Figure	Page
1.1 Structure en double hélice de l'ADN.	5
1.2 La méiose I en haut et la méiose II en bas.	7
1.3 Recombinaison de gènes. Sur le gamète 2, A et b, b et C sont des recombinants, sur le gamète 3, a et B, B et c sont des recombinants. Mais A et C sur le gamète 2, a et c sur la gamète 3 ne sont pas des recombinants.	9
1.4 Exemple de trois SNPs sur trois individus.	11
2.1 Structure générale du complexe HLA.	20
2.2 Structure des molécules du CMH.	22
2.3 À gauche : aucune réaction immunitaire, à droite : déclenchement de réaction immunitaire.	23
2.4 Différence entre les classes I et II dans l'immunité.	24
2.5 Origine des lymphocytes.	25
2.6 Le macrophage phagocyte les complexes anticorps-antigène.	28
2.7 Structure de base d'un anticorps.	29
2.8 Mécanismes d'action des anticorps.	30
2.9 La sélection clonale. Cette figure illustre la sélection clonale des lymphocytes B. Après avoir reconnu un antigène, un lymphocyte B se divise et forme un clone de cellules. Les autres lymphocytes B ne réagissent pas, car leurs récepteurs antigéniques sont spécifiques à d'autres antigènes. Le clone de cellules issues du lymphocyte B activé donne naissance à des cellules mémoire B et à des plasmocytes producteurs d'anticorps.	31

2.10	Les principaux types de lymphocytes T. a) Différenciation des lymphocytes T. b) La destruction d'une cellule hôte infectée par des lymphocytes T cytotoxiques.	33
2.11	Rôle des lymphocytes T4.	34
2.12	Réaction immunitaire adaptative.	35
2.13	Nomenclature des allèles HLA 2010 (WHO), l'étoile indique l'identification par biologie moléculaire.	39
2.14	Les types d'épitopes.	41
4.1	Distribution de la probabilité de compatibilité par ethnicité, pour l'épitype au complet et le groupe <i>abv</i> de classe I, selon les trois seuils.	57
4.2	Distribution de la probabilité de compatibilité par ethnicité, pour les trois seuils, en considérant : 23 épitopes <i>abv</i> , 72 épitopes <i>abv</i> et 72 épitopes <i>oth</i>	59
4.3	Probabilités de compatibilité médianes par ethnicité de l'épitype au complet en considérant les trois seuils.	62
4.4	Probabilités de compatibilité médianes par ethnicité du groupe <i>abv</i> de classe I en considérant les trois seuils.	63
4.5	Distribution de la différence des probabilités de compatibilité médianes de l'épitype au complet selon les trois seuils. La couleur rouge représente le seuil 5 pour les 6 paires d'ethnicités, la couleur verte représente le seuil 8 et la couleur bleue représente le seuil 10.	66
4.6	Distribution de la différence des probabilités de compatibilité médianes des <i>abv</i> de classe I selon les trois seuils. La couleur rouge représente le seuil 5 pour les 6 paires d'ethnicités, la couleur verte représente le seuil 8 et la couleur bleue représente le seuil 10.	67
4.7	Les courbes de la fonction de répartition des quatre ethnicités, en considérant les <i>abv</i> de classe I au seuil 10.	68
4.8	Les courbes de la fonction de répartition des quatre ethnicités, en considérant l'épitype au complet aux trois seuils.	69
4.9	Les courbes de la fonction de répartition des quatre ethnicités, en considérant le groupe <i>abv</i> de classe I aux trois seuils.	69

B.1	Algorithme cPIE en se basant sur les épitypes présents dans le bassin des donneurs.	85
-----	---	----

LISTE DES TABLEAUX

Tableau	Page	
1.1	Génotype des individus de l'exemple présenté à la figure 1.4.	10
1.2	La distribution des génotypes	13
1.3	Les deux loci sont en équilibre de liaison.	14
1.4	Les deux loci sont en déséquilibre de liaison.	14
2.1	Comparaison des lymphocytes B et T.	27
2.2	Compatibilité du groupe sanguin entre donneur-receveur.	37
2.3	Exemple d'un génotype avec 5 loci.	42
2.4	Aperçu de quelques épitopes associés au génotype de l'exemple 2.3.	42
3.1	Épitype du receveur constitué de 10 épitopes. La présence d'un épitope est représentée par la valeur 1 et par la valeur 0 en cas de son absence.	48
3.2	Bassin des épitypes des donneurs. L'épitype est constitué de 10 épitopes. Pour chaque donneur, la présence d'un épitope est représentée par la valeur 1 et par la valeur 0 en cas de son absence. . .	48
3.3	Population des épitypes distincts. La présence d'un épitope est représentée par la valeur 1 et par la valeur 0 en cas de son absence.	49
3.4	Bassin des épitypes des donneurs avec les fréquences génotypiques correspondantes. L'épitype est constitué de 10 épitopes. Pour chaque donneur, la présence d'un épitope est représentée par la valeur 1 et par la valeur 0 en cas de son absence.	52
3.5	Population des épitypes distincts. La présence d'un épitope est représentée par la valeur 1 et par la valeur 0 en cas de son absence.	52
4.1	Probabilités de compatibilité médianes de l'épitype au complet et le groupe <i>abv</i> de classe I, selon les trois seuils, pour chaque ethnicité.	64

A.1	Épitopes associés au génotype (1).	75
A.2	Épitopes associés au génotype suite (2).	76
A.3	Épitopes associés au génotype suite (3).	77
A.4	Épitopes associés au génotype suite (4).	78
A.5	Épitopes associés au génotype suite (5).	79
A.6	Épitopes associés au génotype suite (6).	80
A.7	Épitopes associés au génotype suite (7).	81
A.8	Épitopes associés au génotype suite (8).	82
A.9	Épitopes associés au génotype suite (9).	83

RÉSUMÉ

Dans ce mémoire nous présentons un nouvel algorithme statistique cPIE (de l'anglais *calculated Panel of Incompatible Epitopes*). Il permet de calculer la probabilité de trouver des donneurs compatibles pour un candidat donné en transplantation rénale. Cet algorithme est basé sur l'appariement des épitopes du CMH (Complexe majeur d'histocompatibilité) en tenant compte du groupe sanguin, pour un candidat au niveau de la population. Cette approche permet de contribuer à l'augmentation de la survie du greffon et d'assurer une meilleure allocation du rein. Des simulations des génotypes, pour les donneurs et receveurs de différentes ethnicités (auto-identifiées), sont effectuées pour comparer la distribution de la probabilité de compatibilité épitypique. Nos résultats montrent l'effet de l'appartenance ethnique auto-identifiée qui peut affecter l'accès à la transplantation rénale.

MOTS-CLÉS : CMH, HLA, immunologie, structure d'épitopes, compatibilité, épitopes, épélets, transplantation rénale.

INTRODUCTION

La transplantation rénale reste le traitement de choix pour la survie des patients atteints d'insuffisance rénale. Une des grandes difficultés de la greffe consiste à éviter le rejet. La compatibilité entre donneur-receveur au niveau génétique est un facteur important pour déterminer si le corps acceptera ou rejettera le rein d'un donneur. Actuellement, l'attribution des organes repose sur l'absence d'incompatibilité telle que déterminée par les anticorps anti-Complexe majeur histocompatibles (anti-CMH) préformés. Plus précisément, on utilise les informations du génotypage du complexe majeur d'histocompatibilité (CMH) pour déduire l'incompatibilité entre donneur-receveur au niveau de l'épitope. Ce dernier, appelé aussi déterminant antigénique, représente une région d'antigène possédant le site de fixation aux anticorps ou aux récepteurs lymphocytaire. On parle ainsi de la compatibilité/incompatibilité au niveau du système du CMH ou HLA (de l'anglais *Human Leukocyte Antigens*). Ce système représente une composante importante du système immunitaire qui pourrait interagir avec des antigènes internes (le soi) et externes (le non-soi) et les présenter à d'autres composantes du système immunitaire. On dit alors que le donneur et le receveur sont compatibles si leurs tissus sont immunologiquement compatibles les uns avec les autres. Cependant, plusieurs obstacles rendent la compatibilité entre donneur-receveur difficile. L'hétérogénéité des variants génétiques CMH dans la population, due à la recombinaison génétique que subissent les chromosomes, introduit une variation génétique importante au niveau des allèles CMH. Ceci présente un grand défi au niveau de la correspondance des antigènes. Toutefois, la compatibilité au niveau de l'épitope permet de contourner ce problème. Cette approche, basée sur la compatibilité au niveau de

l'épitope, permet de considérer quelques centaines d'épitopes au lieu des milliers d'allèles. De plus, certains épitopes sont partagés entre les différentes molécules de classe I du CMH.

Ce mémoire a pour objectif de présenter un modèle statistique et un algorithme, nommé cPIE (de l'anglais *calculated Panel of Incompatible Epitopes*). Il permet de déterminer la compatibilité donneur-receveur au niveau des épitopes, en considérant différentes stratégies de compatibilité (ou appariement). Dans ce travail, nous nous limiterons à l'application du cPIE exclusivement à la classe I CMH et en considérant deux stratégies de compatibilité. Les deux stratégies de compatibilité ou d'appariement des épitopes CMH sont l'épitope au complet de classe I et les épitopes *abv* (abréviation du terme anglais *Antibody verified*) de classe I. Les résultats de ce projet pourraient aider l'organisation d'approvisionnement en organes à l'attribution des organes. Le premier chapitre de ce mémoire est consacré à la présentation des concepts de base en génétique permettant au lecteur de se familiariser avec la terminologie génétique. Ensuite, une introduction à l'immunologie et au système CMH sera décrite au chapitre II, afin d'assurer une bonne compréhension du concept de compatibilité sur lequel se base la conception de notre algorithme. Le chapitre III est consacré à la présentation et au développement du nouvel algorithme statistique, basé sur la compatibilité d'épitope, en considérant les deux approches décrites précédemment. Enfin, le dernier chapitre de ce mémoire est consacré à l'application de l'algorithme à travers une analyse par simulation, décrivant la distribution de la compatibilité donneur-receveur pour différentes ethnicités auto-identifiées et les deux stratégies de compatibilité. Pour ce faire, on utilise la base de données des haplotypes du programme *NMDP* (National Marrow Donor Program, 2007) qui représente les fréquences alléliques du CMH, dans plus de 20 populations principalement américaines.

CHAPITRE I

CONCEPTS DE GÉNÉTIQUE

La génétique est la science qui étudie les caractères héréditaires des individus, leur transmission au fil des générations et leurs variations. L'histoire de cette science a commencé avec le travail, publié en 1866, du frère Augustin Gregor Johann Mendel, sur la théorie d'hérédité sur les plantes de pois.

Ce chapitre est consacré à l'introduction des concepts de base génétiques essentiels à la compréhension du modèle statistique présenté dans ce mémoire. Les concepts présentés dans ce chapitre proviennent essentiellement du livre *Statistical Methods for QTL Mapping* (Chen, 2014), et des articles (Gross, 2014), (Baxter-Lowe, 2021) et (Edwards, 2008).

1.1 L'ADN

Le génome est l'ensemble des chromosomes d'un organisme. Le génome humain contient 46 chromosomes, 22 paires de chromosomes homologues et une paire de chromosomes sexuels non homologues. Un chromosome est une double hélice de deux brins de séquences d'acide désoxyribonucléique (ADN). Chaque brin est composé d'une succession de nucléotides accrochés les uns aux autres par des liaisons phosphodiester. Chaque nucléotide est composé par l'une des quatre bases azotées : l'adénine (A), la thymine (T), la guanine (G) et la cytosine (C) qui

codent l'information. Les deux brins sont appariés de manière complémentaire pour former la double hélice. La base A dans un brin est toujours appariée avec T dans l'autre brin, et la base C dans un brin est toujours appariée avec G dans l'autre brin. Par exemple, si GATCCT est un segment sur un brin, le segment complémentaire sur l'autre brin est CTAGGA. La structure en double hélice est illustrée à la figure 1.1¹. C'est la succession des bases azotées qui détermine l'information contenue dans l'ADN selon un code appelé le **code génétique**, et les 4 bases composent l'alphabet du programme génétique. Chaque mot de la séquence d'ADN est formé par trois lettres appelés **codons**. Par exemple, GCC est un codon. Il y a en tout $4^3 = 64$ mots ou codons. Chaque mot code l'information d'un acide aminé particulier. Il existe 20 acides aminés distincts et chacun d'eux est codé par plus d'un mot. Les codons forment des segments contenant des informations particulières. Il existe deux types de tels segments : les **exons** qui codent les informations pour la production de protéines et les **introns** qui ne codent pas directement pour les protéines.

1. Source : <https://www.cea.fr/comprendre/Pages/sante-sciences-du-vivant/1-ADN-dechiffrer-pour-mieux-comprendre-le-vivant.aspx?Type=Chapitre&numero=1>

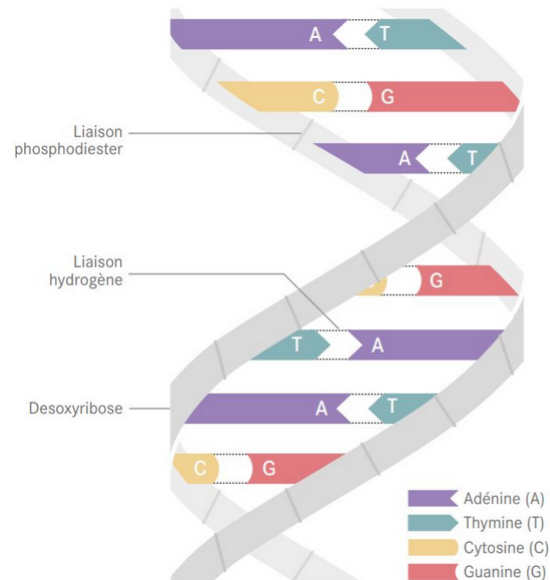


FIGURE 1.1 – Structure en double hélice de l’ADN.

1.2 Les gènes

Le gène est l’unité de l’hérédité. Un gène est une séquence d’exons et d’introns qui sont transmis de génération en génération. Le gène a toujours deux formes homologues sauf s’il est situé sur des chromosomes sexuels. Ces formes homologues d’un gène sont appelées **allèles**. Deux allèles constituent un gène chez l’individu et la combinaison de ces deux allèles est appelée **génotype**. Si les deux allèles sont identiques, le génotype est dit **homozygote** et si les deux allèles sont distincts le génotype est dit **hétérozygote**.

Le système ABO (Gross, 2014), par exemple, est l’un des systèmes de groupes sanguins les plus importants en médecine transfusionnelle. Il se compose d’**antigènes**²

2. Une matière étrangère à l’organisme capable de déclencher un mécanisme de défense dans le but de l’éliminer.

A et B et d'**anticorps**³ contre ces antigènes. Il existe 4 groupes principaux dans le système ABO (A, B, AB et O) qui résultent de 3 allèles majeurs (A, B et O) du gène ABO. Les antigènes A et B sont des antigènes glucidiques plutôt que protéiques. L'étape finale de leur biosynthèse est catalysée par les glycosyltransférases (abrégé GT) A et B, qui sont respectivement codées par les allèles A et B du gène ABO. Les individus de groupe sanguin O ne produisent pas de GT A ou B fonctionnelles et manquent donc d'antigènes A et B. La présence d'anticorps naturels contre les antigènes A et B, chez les individus qui n'expriment pas ces antigènes, provoque un résultat néfaste lors de la première transfusion incompatible.

Étant donné que les antigènes A et B existent dans des cellules autres que les globules rouges, la correspondance ABO est également importante dans la transplantation de cellules, de tissus et d'organes.

1.3 Division cellulaire

La division de la cellule permet de conserver le matériel génétique et le transmettre dans chaque cellule fille. Cette division cellulaire permet de former des cellules sexuelles. Grâce à ce processus, une cellule **diploïde**, c'est-à-dire une cellule qui contient des paires de chromosomes identiques, permet la formation de 4 cellules **haploïdes**⁴, les **gamètes**. La cellule diploïde est formée par l'union de deux gamètes de sexe différent. La division cellulaire permettant de former des gamètes est la **méiose** et a lieu dans les organes génitaux (testicules, ovaires).

Au cours de la méiose, la cellule mère est divisée en quatre cellules de gamètes, chacune contenant le même type de chromosomes, mais deux fois moins que les chromosomes de la cellule mère. La méiose se produit après l'interphase du cycle

3. Une matière de défense produite par l'organisme en réaction à une matière étrangère.

4. Des cellules contenant des chromosomes en simple exemplaire.

cellulaire au cours de laquelle les chromosomes sont répliqués, et elle consiste en deux divisions cellulaires consécutives. Les deux processus de la division cellulaire sont appelés **méiose I** ou division réductionnelle et **méiose II** ou division équationnelle (figure 1.2⁵). Après les deux phases de la méiose, la cellule mère se divise finalement en quatre cellules de gamètes, chacune contenant la moitié du matériel de la cellule mère. Pendant la prophase de la méiose I, les chromosomes homologues s'apparient. Ils sont comprimés ensemble dans le cadre d'un processus appelé **synapsis**, pour former des **bivalents** ou **tétrades**. Au cours de ce processus, les chromatides d'un bivalent échangent leur matériel en des endroits appelés **chiasmata** (points de chevauchement de deux chromatides). Ce phénomène appelé **brassage intrachromosomique** (ou en anglais *crossing-over*) est important car il permet d'engendrer une grande diversité génétique.

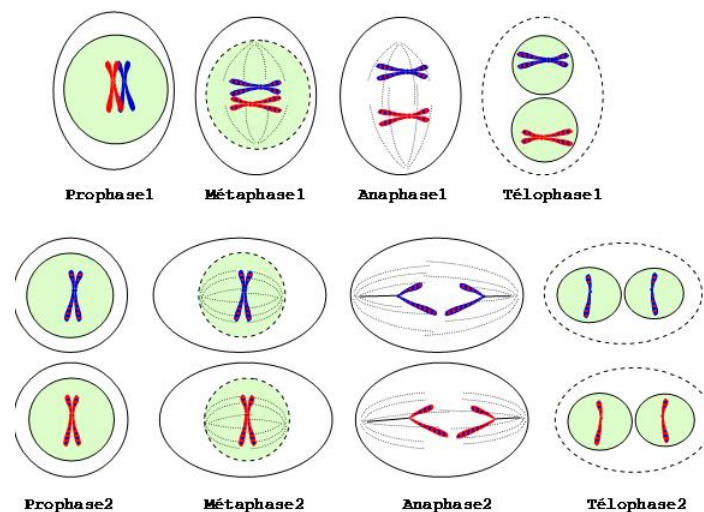


FIGURE 1.2 – La méiose I en haut et la méiose II en bas.

Après la synapsis, les chromosomes homologues sont décomprimés, mais restent

5. Source : http://ressources.unisciel.fr/DAEU-biologie/P1/co/P1_chap3_c3.html

attachés au niveau du chiasmata. Au cours de la métaphase de la méiose I, les chromatides soeurs s'alignent le long de la plaque de métaphase sur deux rangées. Au cours de l'anaphase et de la télophase, le même nombre et type de paires de chromatides soeurs migrent aux deux pôles pour former deux noyaux. Chacun de ces nouveaux noyaux contient la moitié des chromatides d'origine, et ces deux noyaux subissent une division supplémentaire pendant la méiose II. Le processus de la méiose II est identique à celui de la mitose, sauf qu'il n'est pas précédé d'une réplication de l'ADN, et les chromatides soeurs ne sont pas forcément génétiquement identiques à cause des brassages intrachromosomiques. À la fin de la méiose II, la cellule mère d'origine est divisée en quatre gamètes.

1.4 Recombinaison génétique

Dans le processus de la méiose, les chromosomes dans les cellules filles (gamètes) ne sont plus obtenus par paires homologues, il n'y a qu'un chromosome (gamètes haploïdes). En raison du croisement au cours de la méiose, certains chromosomes dans les gamètes sont des combinaisons de chromosomes homologues paternel et maternel. La figure 1.3⁶ illustre un faisceau de 4 chromatides soeurs homologues avec deux chiasmata et ses quatre gamètes. Les parties noires représentent les chromosomes paternels, avec des allèles en lettres majuscules. Les parties blanches représentent les chromosomes maternels, avec des allèles en lettres minuscules. Au niveau des deux chiasmata, les mêmes chromatides ont participé au croisement, ce qui a donné lieu aux deux gamètes 2 et 3. Ces derniers sont composés de matériel provenant, à la fois, de chromosomes paternel et maternel. Quant aux gamètes 1 et 4 restent purs à la source de leur matériel génétique. La recombinaison ne concerne que deux **loci** (une position fixe d'un gène ou d'un marqueur génétique sur un

6. Exemple tiré de : (Chen, 2014)

chromosome) sur un gamète. Si les allèles de deux loci ne proviennent pas de la même source parentale, on dit qu'il y a une recombinaison de ces deux loci. Les deux loci sont appelés un **recombinant**. Une autre façon, de déterminer si deux loci sont un recombinant, consiste à compter le nombre de croisements entre eux. Si le nombre de croisements entre eux est impair, les deux loci sont un recombinant, sinon ils ne le sont pas. Pour deux loci fixés, la proportion de recombinants parmi tous les gamètes produits par la méiose dans la population est appelée fraction de recombinaison entre eux. Ainsi, pour un gamète sélectionné au hasard, la fraction de recombinaison est la probabilité que le gamète soit un recombinant.

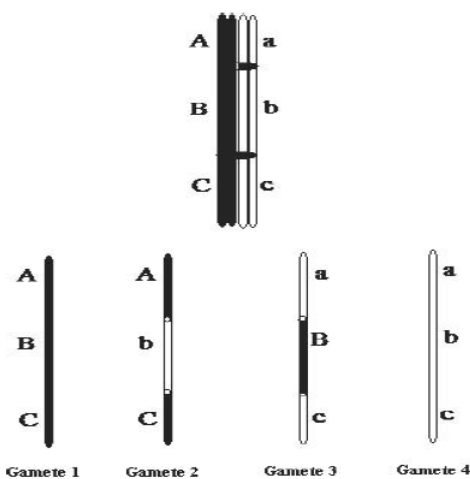


FIGURE 1.3 – Recombinaison de gènes. Sur le gamète 2, A et b, b et C sont des recombinants, sur le gamète 3, a et B, B et c sont des recombinants. Mais A et C sur le gamète 2, a et c sur la gamète 3 ne sont pas des recombinants.

Les gènes d'histocompatibilité⁷ sont caractérisés par un polymorphisme⁸ excep-

7. Les gènes responsables sur la défense de notre corps.

8. Différentes formes que peuvent prendre un même gène.

tionnel résultant de mécanismes actifs de génération de la diversité. L'analyse de la variation du CMH (Complexe Majeur d'histocompatibilité) a révélé des preuves de recombinaison intra et intergénique et une variation à plus de 40% des nucléotides situés dans les régions codantes (Baxter-Lowe, 2021). Bien que les événements de recombinaison aient contribué à la diversité du CMH, les taux de recombinaison dans cette région du génome sont relativement faibles et montrent des sites de recombinaison spécifiques à la population. Pour cette raison, les gènes d'histocompatibilité sont souvent hérités sous forme d'haplotypes (un ensemble d'allèles) sur un seul chromosome maternel ou paternel.

1.5 Les marqueurs génétiques

Un marqueur génétique est une séquence d'ADN repérable grâce à un emplacement connu sur un chromosome. Le marqueur génétique est en fait un drapeau qui identifie une section d'ADN associée à un critère de performance recherché. Il existe différentes sortes de marqueurs et les SNP (de l'anglais *Single Nucleotide Polymorphism*) - prononcé **snip** - constituent la forme la plus abondante de variations génétiques dans le génome humain. C'est un type de polymorphisme de l'ADN dans lequel deux chromosomes diffèrent sur un segment donné par une seule paire de bases (figure 1.4). Chaque orthographe d'une région de l'ADN représente un allèle. Un **haplotype** est défini alors par une séquence d'allèles reçue du même parent. Chaque individu se caractérise, pour un chromosome donné, par deux haplotypes, provenant l'un du père et l'autre de la mère. Ces deux haplotypes reçus définissent le génotype de l'individu aux deux loci. Par exemple, dans le tableau 1.1, l'individu 1 possède le génotype GC sur le SNP1, le génotype AC sur le SNP2 et le génotype TG sur le SNP3. La catégorie des SNPs ayant subi une mutation sont appelés des **mutants** et possèdent l'allèle **mineur**, c'est-à-dire l'allèle le moins fréquent dans la population.

		SNP1	SNP2	SNP3
		↓	↓	↓
Individu 1	Maternel	T A G T C ... A A T G ... C T T G		
	Paternel	T A C T C ... A C T G ... C G T G		
Individu 2	Maternel	T A C T C ... A A T G ... C G T G		
	Paternel	T A C T C ... A A T G ... C T T G		
Individu 3	Maternel	T A G T C ... A A T G ... C T T G		
	Paternel	T A C T C ... A C T G ... C C T G		

FIGURE 1.4 – Exemple de trois SNPs sur trois individus.

TABLEAU 1.1 – Génotype des individus de l'exemple présenté à la figure 1.4.

Individu	Génotype SNP1	Génotype SNP2	Génotype SNP3
1	G/C	A/C	T/G
2	C/C	A/A	G/T
3	G/C	A/C	T/T

1.6 L'équilibre de Hardy-Weinberg

L'équilibre de Hardy-Weinberg, encore appelé équilibre panmictique, a été proposé indépendamment par le mathématicien anglais Hardy (Edwards, 2008) et le médecin allemand Weinberg et reste le modèle théorique central de la génétique des populations. Cette notion d'équilibre est assujettie à différentes hypothèses qui permettent de montrer que la diversité génétique de la population se maintient, et doit tendre vers un équilibre stable de la distribution génotypique. Ces hypothèses sont les suivantes :

- La population est infinie (pour minimiser les variations d'échantillonnage) ;
- La population est panmictique (union aléatoire des individus et gamètes) ;
- Il ne doit y avoir dans la population ni sélection, ni mutation, ni migration (aucun apport d'allèle extérieur) ;
- Les générations successives sont discrètes (pas de superposition de générations dans les croisements).

En particulier, au bout de quelques générations, les proportions génotypiques se fixent autour d'un équilibre spécifié par une simple fonction des fréquences alléliques. Considérons l'exemple d'un locus bi-allélique d'allèles A , a et fréquences à la génération t

$$f_t(A) = p_t, \quad f_t(a) = q_t = 1 - p_t.$$

La distribution des génotypes, sous les hypothèses ci-dessus, à la génération $t + 1$ est donnée par ce qui suit :

- Pour que l'individu soit AA , il faut qu'il ait reçu de ses deux parents l'allèle A et la probabilité de cet événement est p_t^2 (car les croisements se font aléatoirement). Ainsi la fréquence de AA est $f_{t+1}(AA) = p_t^2$;
- En suivant le même raisonnement on obtient $f_{t+1}(aa) = q_t^2$;
- Pour que l'individu soit Aa , il y a deux possibilités. Soit il reçoit A du père et a de la mère avec probabilité $p_t q_t$, ou bien il reçoit a du père et A de la mère avec probabilité $q_t p_t$. La fréquence de Aa est $f_{t+1}(Aa) = p_t q_t + q_t p_t = 2p_t q_t$.

Sous les hypothèses du modèle de Hardy-Weinberg, à la génération $t + 1$, la distribution des génotypes dans la population est décrite dans le tableau ci-dessous.

TABLEAU 1.2 – La distribution des génotypes

AA	Aa	aa
p_t^2	$2p_tq_t$	q_t^2

Sous les hypothèses d'absence de sélection, mutation, migration et taille infini, on peut trouver, à partir des fréquences génotypiques, les fréquences alléliques à la génération $t + 1$:

$$f_{t+1}(A) = f_{t+1}(AA) + \frac{1}{2}f_{t+1}(Aa) = p_t^2 + p_tq_t = p_t,$$

$$f_{t+1}(a) = f_{t+1}(aa) + \frac{1}{2}f_{t+1}(Aa) = q_t^2 + p_tq_t = q_t.$$

Les fréquences alléliques, sous les hypothèses du modèle de HW, restent stables au cours des générations.

1.7 Déséquilibre de liaison

Le déséquilibre de liaison (Feingold, 1991) (en anglais *Linkage disequilibrium*, *LD*) dans une population, se définit comme l'association non aléatoire d'allèles à des loci différents. On dit qu'il y a déséquilibre de liaison si la fréquence des gamètes porteurs de deux gènes A et B est différente du produit des fréquences de ces deux derniers, ce produit définit l'équilibre. Prenons un exemple pour illustrer les deux situations d'équilibre et du déséquilibre de liaison. Deux loci G et H sont en équilibre de liaison s'ils sont indépendants, c'est-à-dire si la probabilité d'observer un couple d'allèles dans les deux loci est égale au produit des probabilités d'observer ces allèles individuellement. Cette situation est présentée dans le tableau 1.3. On suppose G , H biallélique et on dénote : A , a et B , b leurs allèles respectifs. p_A et p_B les fréquences de A et B , p_{AB} , p_{Ab} , p_{aB} et $p_{ab} = 1 - p_{AB} - p_{Ab} - p_{aB}$, les fréquences des 4 haplotypes possibles.

TABLEAU 1.3 – Les deux loci sont en équilibre de liaison.

	A	a
B	$p_{AB} = p_A p_B$	$p_{aB} = p_a p_B$
b	$p_{Ab} = p_A p_b$	$p_{ab} = p_a p_b$

Deux loci G et H sont en déséquilibre de liaison s'ils ne sont pas indépendants, c'est-à-dire si la probabilité d'observer un couple d'allèles dans G et H sur un chromosome n'est pas égale au produit des probabilités d'observer ces allèles individuellement. Ce cas de figure est présenté dans le tableau 1.4.

TABLEAU 1.4 – Les deux loci sont en déséquilibre de liaison.

	A	a
B	$p_{AB} \neq p_A p_B$	$p_{aB} \neq p_a p_B$
b	$p_{Ab} \neq p_A p_b$	$p_{ab} \neq p_a p_b$

La mesure du déséquilibre de liaison, D , est défini par la relation

$$D = p_{AB} - p_A p_B.$$

En développant la mesure D pour le cas du déséquilibre de liaison on obtient

$$\begin{aligned}
D &= p_{AB} - p_A p_B \\
&= p_{ab} - p_a p_b \\
&= p_A p_b - p_{Ab} \\
&= p_a p_B - p_{aB} \\
&\neq 0.
\end{aligned}$$

Puisque $D \neq 0$, on dit qu'il y a déséquilibre de liaison. La mesure D du déséquilibre de liaison est très sensible aux fréquences marginales p_A et p_B ; si une des fréquences alléliques est faible alors D sera également faible. On peut montrer que

$$\begin{aligned}
D &\leq \min(p_A p_b, p_a p_B) \quad \text{si } D \geq 0, \\
-D &\leq \min(p_A p_B, p_a p_b) \quad \text{si } D \leq 0.
\end{aligned}$$

On a donc

$$D_{min} = -\min(p_A p_B, p_a p_b) \leq D \leq \min(p_A p_b, p_a p_B) = D_{max}.$$

Si $p_a \simeq 0$, alors $D_{max} \simeq 0$, $D_{min} \simeq 0$ et donc $D \simeq 0$. Par conséquent, si $D \simeq 0$, on ne peut pas conclure que les loci sont en équilibre de liaison car cette faible valeur pourrait être le résultat des faibles fréquences alléliques. Pour contourner ce problème, une autre mesure D' de LD plus robuste (Lewontin, 1964) aux fréquences alléliques faibles a été introduite et est défini par

$$D' = \begin{cases} D/D_{max} & \text{si } D \geq 0, \\ D/D_{min} & \text{si } D < 0. \end{cases}$$

On a alors $0 \leq D' \leq 1$: Si $D' \simeq 1$ cela indique un fort LD et si $D' \simeq 0$ on a un faible LD. Il faut souligner que le fait d'avoir D ou $D' = 1$ ne permet pas de conclure que les deux loci portent exactement la même information. Afin de

mesurer cette quantité d'information on utilise le coefficient de corrélation (Hill et al., 1968) noté r^2 . En considérant toujours notre exemple des deux loci G et H , on attribue aux allèles A, B la valeur 1 et à a, b la valeur 0 et on calcule le coefficient de corrélation entre les loci :

$$\begin{aligned} r &= \frac{\text{cov}(G, H)}{\sqrt{V(G)V(H)}} = \frac{E[GH] - E[G]E[H]}{\sqrt{V(G)V(H)}} \\ &= \frac{p_{AB} - p_A p_B}{\sqrt{p_A p_a p_B p_b}} \\ &= \frac{D}{\sqrt{p_A p_a p_B p_b}}. \end{aligned}$$

Ainsi, nous avons

$$r^2 = \frac{D^2}{p_A p_a p_B p_b}.$$

Si $r^2 = 1$, on dit que la dépendance est parfaite et donc l'allèle sur un locus détermine l'allèle sur l'autre locus.

En pratique, seuls les génotypes des individus sont observés et non leurs haplotypes. Pour les individus ayant, par exemple, les génotypes Aa à un locus et Bb à un autre, il est impossible d'en déduire leur phase, c'est à dire l'appartenance des allèles à chaque haplotype (AB et ab , ou Ab et aB). Pour estimer les fréquences haplotypiques et la mesure D correspondante, un algorithme EM (Dempster et al., 1977) est souvent utilisé. Afin de minimiser le temps de calcul, on quantifie habituellement le LD à l'aide du coefficient r^2 empirique, sans passer par l'estimation de D .

Dans le contexte du CMH, que nous allons décrire plus en détails dans le chapitre suivant, les loci CMH-A, -B, -C, -DR, -DQ, -DP sont distincts mais étroitement

liés sur le même chromosome. La transmission des gènes se fait donc en blocs, des parents aux enfants. Ainsi, plusieurs grappes d'acides aminés polymorphes sont partagés à travers les loci CMH.

CHAPITRE II

IMMUNOLOGIE ET LE SYSTÈME CMH

L'immunologie est la branche de biologie qui étudie le système de défense d'un organisme. Son origine est généralement attribuée à Edward Jenner (Riedel, 2005), qui a découvert un vaccin qui induisait une protection contre la variole humaine, une maladie souvent mortelle. Louis Pasteur (Berche, 2012) a mis au point un vaccin contre le choléra chez les poulets et a mis au point un vaccin contre la rage qui s'est avéré un succès spectaculaire. Ensuite, Emil von Behring et Shibasaburo Kitasato (Kaufmann, 2017) ont découvert que le sérum des individus vaccinés contenait des substances qu'ils appelaient des anticorps.

Ce chapitre commence par donner une introduction à l'immunologie pour faciliter la compréhension de la réponse immunitaire. Les informations concernant la partie immunologie proviennent essentiellement des livres « *Immunobiology* » (Janeway et al., 2001), « *Immunologie* » (Male, 2019) et « *Biologie* » (Faucher et al., 2011). Ensuite, nous présenterons une introduction sur la transplantation rénale. Par la suite, nous allons expliquer la compatibilité du groupe sanguin, la nomenclature du système du complexe majeur d'histocompatibilité (CMH), et la compatibilité épitopique du donneur-receveur. Finalement, nous présenterons un exemple expliquant la relation entre les allèles d'un génotype et les épitopes associés.

2.1 Le système immunitaire

Le système immunitaire a pour mission de protéger l'organisme contre des envahisseurs étrangers ou dangereux (par ex : bactéries, virus, tissus greffés, cellules cancéreuses). Afin de défendre l'organisme contre ces envahisseurs, le système immunitaire doit être capable de distinguer ce qui appartient au corps (soi) de ce qui n'en fait pas partie (exogène ou étranger). Le système immunitaire se présente sur deux aspects : l'immunité spécifique et l'immunité non spécifique.

L'immunité non spécifique ou innée est un système de défense immédiat qui fait intervenir des cellules responsables de la **phagocytose**¹. Elle utilise les fonctions d'exclusion du revêtement cutanéomuqueux et la mise en jeu de signaux d'activation entre des ensembles de molécules plasmatiques et des cellules (par ex : macrophages et cellules NK). Cette immunité conduit notamment à la réaction inflammatoire. Quant à l'immunité spécifique ou adaptative, elle se développe en quelques jours et dépend de la reconnaissance spécifique du corps étranger et prélude à sa destruction. Cette immunité possède une mémoire et garde le souvenir de la rencontre. Elle implique la reconnaissance de constituants appelés antigènes (Ag) par des molécules complémentaires très diversifiées (les récepteurs d'antigènes) appelés anticorps (Ac). Les récepteurs antigéniques sont aussi présents sur des cellules de l'immunité spécifique ; les **lymphocytes** T et B appelés respectivement TCR (*T cell receptor*) et BCR (*B cell receptor*). Cependant la reconnaissance antigénique dans le cas des TCR et BCR ne peut se faire seulement si l'antigène est présenté par une molécule de CMH.

1. Mécanisme par lequel une cellule englutit et digère des particules.

2.2 Le complexe majeur d'histocompatibilité (CMH)

Le système du complexe majeur d'histocompatibilité est caractérisé par son extraordinaire polymorphisme génétique. Le polymorphisme CMH génère une grande diversité de molécules CMH représentant pour l'individu la capacité de présenter un grand nombre de peptides et donc une meilleure défense contre les agents pathogènes. Il est à l'origine du déclenchement d'une réponse proliférative et cytotoxique très violente entre individus CMH incompatibles lors d'une **allogreffe**². Chez l'homme le système CMH est appelé complexe HLA (de l'anglais *Human Leukocytes Antigen*). C'est un ensemble de gènes situés sur un segment du bras court du chromosome 6, qui représente environ $\frac{1}{1000}$ ème du génome humain (figure 2.1³). Ces gènes sont responsables de la discrimination entre le soi et le non soi et de la reconnaissance intercellulaire dans les réponses immunitaires humorale et cellulaire.

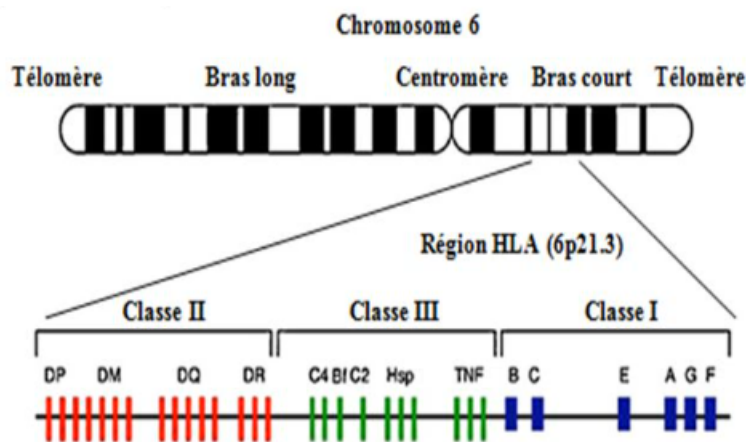


FIGURE 2.1 – Structure générale du complexe HLA.

2. Opération chirurgicale par laquelle sont transmis, pour une personne malade (receveur), un organe ou tissu provenant d'une autre personne (donneur).

3. Source : <https://slideplayer.fr/slide/5052916/>

Le CMH est une région du génome dont les gènes codent pour les molécules d'histocompatibilité qui sont présentes à la surface de cellules présentatrices d'antigène, et qui assurent la présentation des antigènes aux lymphocytes T afin de les activer. Les gènes CMH sont regroupés en trois classes dont les produits diffèrent selon leur structure, leur expression et leurs fonctions. Les protéines de la région CMH classe I sont présentes sur la membrane de toutes les cellules nucléées, et les plus importants gènes de cette classe sont les gènes HLA-A, HLA-B et HLA-C ; codant pour la chaîne α des molécules HLA de classe I. Les protéines de classe II sont présentes sur la membrane de certaines cellules du système immunitaire comme les **macrophages**⁴ et les lymphocytes B, et les gènes principaux de cette classe sont les gènes DRA, DRB, DQA, DQB, DPA et DPB ; codant pour les chaînes α et β des molécules du même nom. Entre ces deux régions existe une troisième région, appelée HLA classe III, dont les principaux gènes ne sont pas impliqués dans le processus de présentation de peptides antigéniques. Seulement les gènes de classe I et II qui codent pour les antigènes d'histocompatibilité. Les molécules du CMH présentent deux particularités structurales. D'une part, elles sont des protéines membranaires présentant un développement important de leur domaine extra-cellulaire qui jouent un rôle important dans les phénomènes de reconnaissance. D'autre part, chaque molécule du CMH délimite un profond sillon dans lequel peut se loger un peptide (figure 2.2⁵).

4. Des cellules impliquées dans la défense de l'organisme contre les substances étrangères.

5. Source : <https://corpus.ulaval.ca/jspui/handle/20.500.11794/17461>

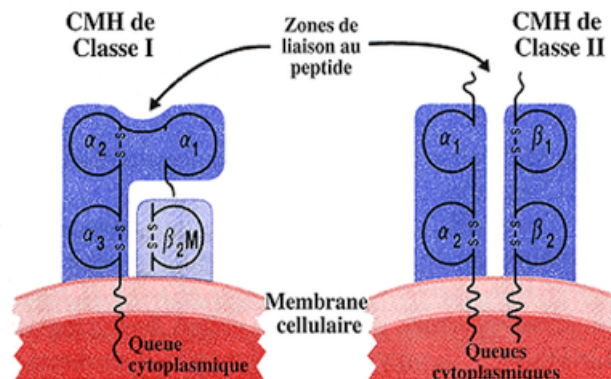


FIGURE 2.2 – Structure des molécules du CMH.

Les molécules du CMH sont des marqueurs du soi, permettant au système immunitaire de vérifier que la cellule fait bien partie de l'organisme. Leur rôle est de présenter des déterminants antigéniques ou **épitopes**⁶ du non soi à des cellules immunitaires afin de déclencher une réaction immunitaire. Prenons un exemple pour comprendre le rôle du CMH. La figure 2.3 présente deux situations possibles de réaction immunitaire. Dans le premier cas, à gauche, à l'intérieur de la cellule, la protéine normale se découpe en fragments (peptide), par des systèmes enzymatiques dans le cytoplasme. Ce peptide se lie à une molécule du CMH (le soi), en se fixant sur le sillon, formant un complexe peptide-CMH qui migre à la surface de la cellule. Cette dernière expose son contenu peptidique et le système immunitaire ne déclenche aucune réaction immunitaire car le contenu fait bien partie du soi. Dans le deuxième cas, à droite, la cellule fabrique une protéine anormale (par exemple une cellule infectée par un virus) qui se dégrade en peptide anormal. De la même manière, ce peptide s'associe au CMH et la cellule présente le contenu peptidique anormal (appelé aussi **soi modifié**) qui va déclencher une réaction

6. Un épitope est une partie d'un antigène qui peut déclencher une réponse immunitaire lors de la rencontre d'un agent pathogène envahissant.

immunitaire. Ainsi la cellule expose, en permanence, un échantillonnage de son contenu peptidique pour assurer la protection de l'organisme en cas d'intrus.

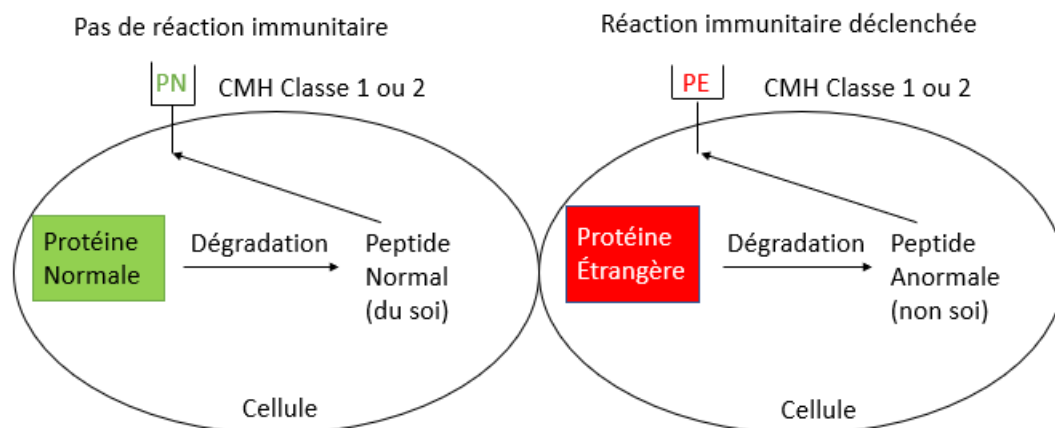


FIGURE 2.3 – À gauche : aucune réaction immunitaire, à droite : déclenchement de réaction immunitaire.

Les deux types du CMH présentent des fragments provenant de la dégradation de protéines cellulaires normales. Cependant, il existe une différence fonctionnelle entre ces deux classes. Les CMH de classe I présentent des peptides issus d'antigènes **endogènes**⁷. La cellule présentant ce complexe peptide-CMH-I déclenche une réaction immunitaire conduisant à la destruction de cette cellule. Contrairement aux molécules CMH de classe II qui présentent des fragments d'antigènes **exogènes**⁸. La cellule présentant ce complexe peptide-CMH-II est appelé **cellule présentatrice d'antigène** (CPA). La présentation de ce complexe déclenche une réaction immunitaire n'aboutissant pas à la destruction de la cellule présentatrice d'antigène mais à la neutralisation de l'élément étranger (figure 2.4).

7. Protéines virales ou anormales synthétisées dans la cellule.

8. Antigènes étrangers phagocytés et dégradés dans la cellule.

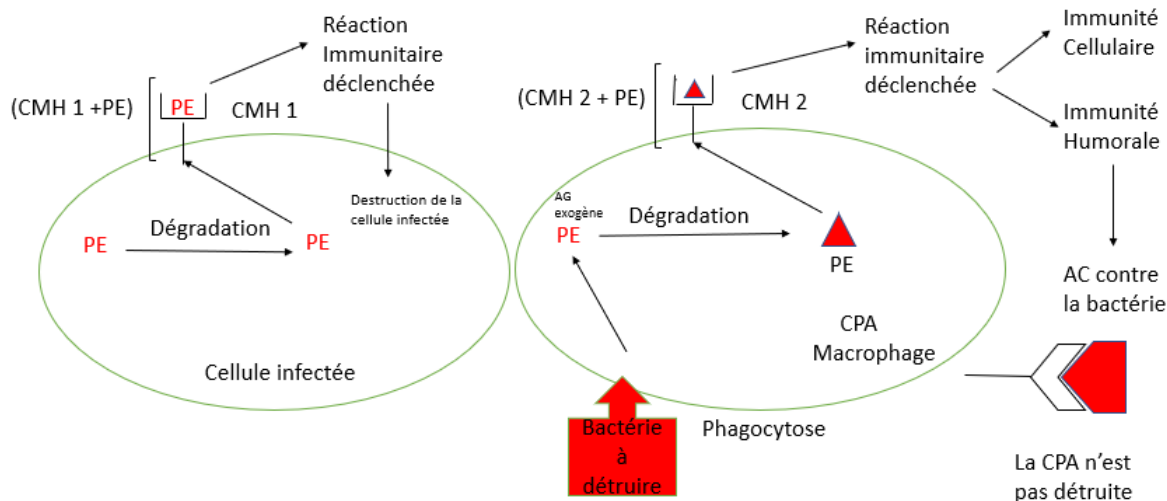


FIGURE 2.4 – Différence entre les classes I et II dans l'immunité.

2.3 Les principales cellules du système immunitaire

Le système immunitaire se compose de cellules et d'organes qui travaillent ensemble pour protéger le corps et réagir aux infections et aux maladies. Il existe trois principaux types de cellules du système immunitaire : les lymphocytes T (LT) qui sont impliqués dans l'immunité cellulaire, les lymphocytes B (LB) qui produisent des anticorps et sont responsables de l'immunité humorale et les cellules présentatrices d'anticorps (macrophages, cellules dendritiques et LB) qui digèrent les antigènes étrangers et présentent les épitopes liés au CMH-II.

Les lymphocytes qui migrent de la moelle osseuse au thymus, pour y subir leur maturation, deviennent des lymphocytes T. Les lymphocytes qui subissent leur maturation dans la moelle osseuse deviennent des lymphocytes B (figure 2.5⁹).

9. Source : <https://www.lymphoma.ca/fr/le-lymphome/lymphome-101/systeme-lymphatique/>

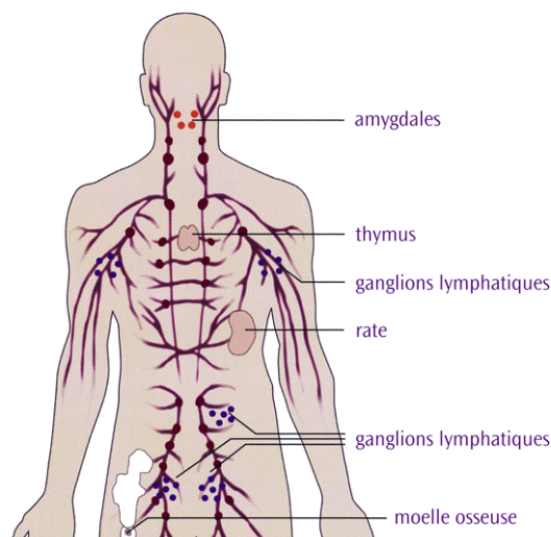


FIGURE 2.5 – Origine des lymphocytes.

Les précurseurs des lymphocytes T expriment chacun un type spécifique de récepteur appelé TCR¹⁰. Les précurseurs des lymphocytes B expriment chacun un type spécifique de récepteur appelé BCR¹¹, qui représente un anticorps membranaire. Les TCR et BCR sont alors des récepteurs des épitopes. Un lymphocyte ne peut donc reconnaître qu'un seul type d'épitope. Les TCR et BCR sont d'une immense variété si bien qu'il existe, chez un même sujet, un très grand nombre de lymphocytes B et de lymphocytes T différents par leurs récepteurs.

Les lymphocytes T doivent être capables de reconnaître les protéines du complexe majeur d'histocompatibilité du soi (CMH du soi) et donc doivent passer par un système de sélection. La première sélection se fait dans le thymus. Les lymphocytes T possédant des TCR aptes à reconnaître le CMH, exprimés par des cellules

10. T cell receptor.

11. B cell receptor.

du thymus, sont conservés et les autres sont éliminés par **apoptose**¹². C'est un apprentissage de reconnaissance du soi. Cette sélection est appelée **sélection positive**. Ensuite, les cellules survivantes seront soumises à la deuxième **sélection négative**. La reconnaissance d'un **auto-antigène**¹³ entraîne l'apoptose du lymphocyte T en développement, ce qui élimine les lymphocytes T qui réagissent au soi et risqueraient de causer une maladie auto-immune. L'incapacité de se lier solidement à un antigène du soi permet la survie et la poursuite de la maturation du lymphocyte T. Seuls survivent les lymphocytes T capables de reconnaître les antigènes du non soi présentés par le CMH, c'est-à-dire les lymphocytes T capables de reconnaître le **soi-modifié**.

Quant aux lymphocytes B, la sélection se fait dans la moelle osseuse. Les lymphocytes B auto-réactifs dont le BCR est capable de fixer des molécules du soi portées par des cellules de la moelle osseuse sont éliminés. Les lymphocytes B qui survivent sont ceux dont le BCR ne reconnaît pas les antigènes du soi et qui sont donc capables de reconnaître le non soi. Ces lymphocytes B et lymphocytes T sont appelés **immunocompétents**. Après avoir acquis leur immunocompétence, les lymphocytes sont amenés par le sang dans les ganglions lymphatiques. Ces ganglions présentent une position stratégique car cela permet de rencontrer un grand nombre d'antigènes amenés par les vaisseaux lymphatiques traversant les ganglions et permet de favoriser les communications à courte distance entre les cellules immunitaires. Une fois que les lymphocytes se lient aux antigènes, ils achèvent leur différenciation en lymphocytes B et lymphocytes T fonctionnels. Le tableau 2.1 ci-dessous présente une comparaison des lymphocytes B et T.

12. Mort par suicide cellulaire.

13. Antigène du soi.

TABLEAU 2.1 – Comparaison des lymphocytes B et T.

	Lymphocytes B	Lymphocytes T
Type de réponse immunitaire	Humorale	Cellulaire
Sécrétion d'anticorps	Oui	Non
Principales cibles	Agents pathogènes extracellulaires (ex : bactéries, champignons, parasites, certains virus dans le liquide interstitiel)	Agents pathogènes intracellulaires (ex : cellules infectées par un virus) et cellules cancéreuses
Origine	Moelle osseuse rouge	Moelle osseuse rouge
Lieu de maturation	Moelle osseuse rouge	Thymus
Cellules effectrices	Plasmocytes	Lymphocytes T cytotoxiques, lymphocytes T auxiliaires, lymphocytes T régulateurs
Formation de cellules mémoires	Oui	Oui

Finalement, les cellules présentatrices d'antigènes jouent aussi un rôle important dans la réaction immunitaire spécifique. Après phagocytose, les macrophages présentent des épitopes sur leur membrane puis ils phagocytent en fin de réaction immunitaire humorale les complexes anticorps-antigène (figure 2.6¹⁴).

14. Source : Le système immunitaire (Muller, 2015).

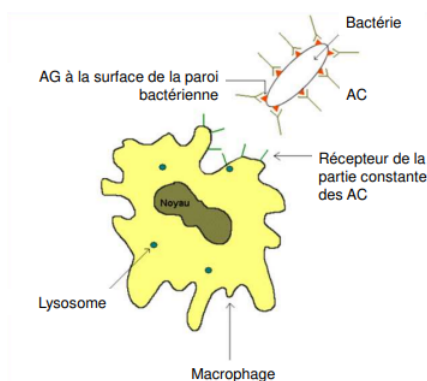


FIGURE 2.6 – Le macrophage phagocyte les complexes anticorps-antigène.

2.4 Réaction immunitaire adaptative

2.4.1 Réaction immunitaire à médiation humorale

L'immunité humorale est un type de réaction immunitaire spécifique provoquée par l'activation du lymphocyte B par l'antigène qui consiste en la sécrétion d'anticorps spécifiques. Les agents effecteurs de cette immunité sont les anticorps. Il existe cependant une différence entre l'immunité humorale active et l'immunité humorale passive. L'immunité humorale active est acquise naturellement par exposition aux antigènes de notre environnement ou artificiellement par vaccination. Le système immunitaire est donc actif et une mémoire immunitaire est établie. Lors d'une prochaine rencontre de pathogène, la réponse est plus rapide et la protection est de longue durée. Dans l'immunité humorale passive, les anticorps sont issus soit du sérum d'un donneur soit de la mère vers son enfant. Le système immunitaire est passif et aucune mémoire immunitaire n'est établie. La protection est de courte durée car les anticorps du donneur seront dégradés.

Il est important à ce stade d'introduire la notion des anticorps car ils sont les produits caractéristiques de l'immunité humorale. Les anticorps sont des protéines appelées **immunoglobulines** (Ig) sécrétées par les lymphocytes B en réponse à

l'exposition à un antigène et se liant spécifiquement à cet antigène. La structure de base des anticorps est constituée de quatre chaînes : deux chaînes longues identiques ou chaînes lourdes H (*heavy*) et deux chaînes courtes identiques ou chaînes légères L (*light*). L'ensemble de ces quatre chaînes polypeptidiques constitue une molécule en forme de Y ; un monomère d'anticorps (figure 2.7¹⁵, à gauche). Chaque des quatre chaînes possède à une extrémité une petite région variable d'anticorps (région V), alors que le reste forme une grande région constante (région C). Les régions V des chaînes H et L de chaque moitié s'associent pour former un site de fixation à l'antigène dont la forme lui permet de lier spécifiquement un épitope. Chaque monomère possède donc deux sites de fixation à l'antigène (figure 2.7, à droite).

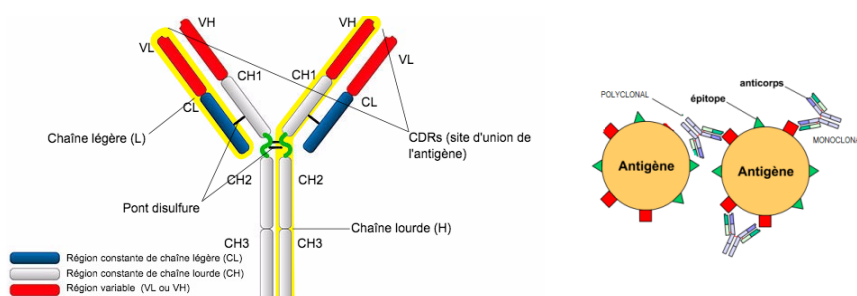


FIGURE 2.7 – Structure de base d'un anticorps.

Il existe environ 30000 gènes CMH différents notés dans le système CMH (Robinson et al., 2015). Les anticorps sont des protéines et sont donc codés par des gènes d'où la très grande variété d'anticorps dépassant le milliard de types différents. Les cellules souches des lymphocytes B contiennent quelques centaines de segments d'ADN dont certains sont réarrangés pour constituer le gène de l'anticorps. Ce processus est appelé **recombinaison somatique** et est la source de cette diversité.

15. Source : <https://www.3trois3.com/articles/le-systeme-immunitaire/>

La diversité de ces anticorps permet de protéger l'organisme contre plusieurs pathogènes. Il existe plusieurs mécanismes permettant aux anticorps de défendre l'organisme contre les antigènes. La figure 2.8¹⁶ résume les mécanismes de destruction des antigènes par l'intermédiaire des anticorps.

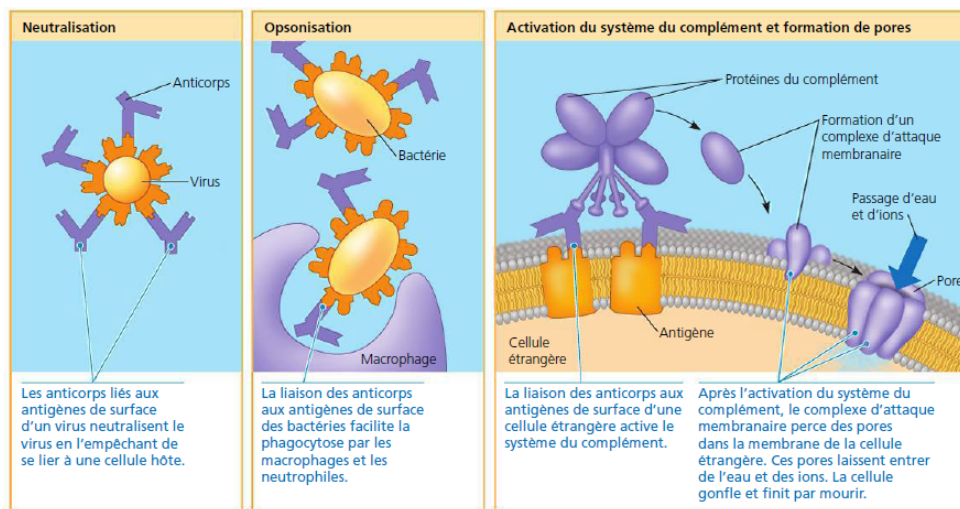


FIGURE 2.8 – Mécanismes d'action des anticorps.

2.4.2 Différenciation des lymphocytes B

Le grand nombre des anticorps résulte en un grand nombre de populations différentes de lymphocytes B, prêtes à rencontrer l'antigène. Le premier contact d'un lymphocyte B avec un antigène étranger induit le processus de **sélection clonale** des lymphocytes B. La figure 2.9¹⁷ résume la prolifération d'un lymphocyte B en un clone après sa liaison avec un antigène.

16. Source : Tiré du livre de Biologie (Faucher *et al.*, 2011).

17. Source : Tiré du livre de Biologie (Faucher *et al.*, 2011).

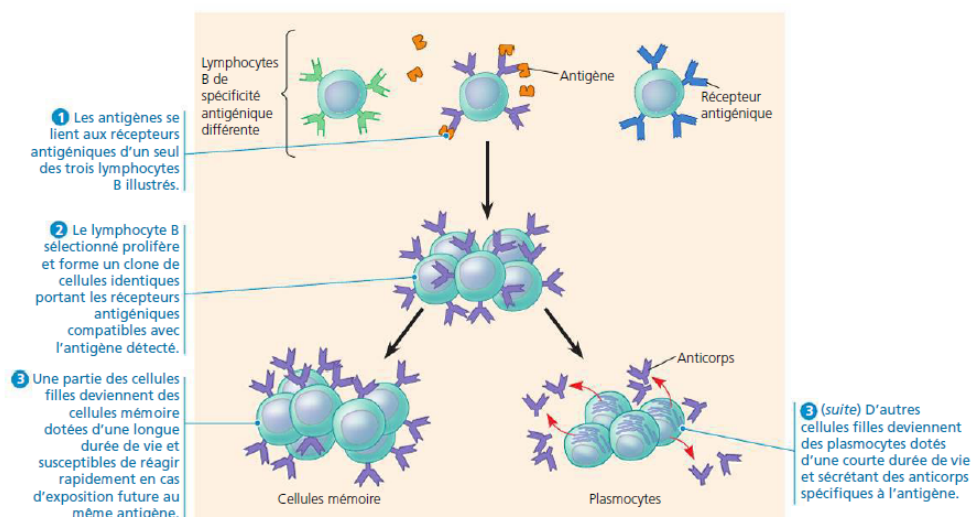


FIGURE 2.9 – La sélection clonale. Cette figure illustre la sélection clonale des lymphocytes B. Après avoir reconnu un antigène, un lymphocyte B se divise et forme un clone de cellules. Les autres lymphocytes B ne réagissent pas, car leurs récepteurs antigéniques sont spécifiques à d'autres antigènes. Le clone de cellules issues du lymphocyte B activé donne naissance à des cellules mémoire B et à des plasmocytes producteurs d'anticorps.

2.4.3 Réaction immunitaire à médiation cellulaire

L'immunité cellulaire est un type de réaction immunitaire spécifique dont les effecteurs sont des cellules lymphocytes T **cytotoxiques**. Elle s'exerce sur des cellules infectées, anormales ou greffées. Elle aboutit à la lyse des cellules cibles après la reconnaissance du non-soi. On distingue deux catégories principales des lymphocytes T : les lymphocytes T4, ou lymphocytes T auxiliaires, qui jouent un rôle essentiel dans l'immunité spécifique et les lymphocytes T8, ou lymphocytes T cytotoxiques, qui sont les cellules effectrices de l'immunité à médiation cellulaire, elles vont tuer les cellules cibles étrangères.

2.4.4 Différenciation des lymphocytes T

Tout comme les lymphocytes B, les lymphocytes T passent par un processus de sélection clonale afin de les différencier. Pour les lymphocytes T4, la cellule présentatrice d'antigène présente l'épitope lié à son CMH-II. Le TCR du lymphocyte T4 reconnaît le soi (CMH-II) et le non-soi (épitope). Ce contact entraîne une activation du lymphocyte T4 et donc la sélection clonale des lymphocytes T4 possédant un TCR reconnaissant l'épitope. Dans le cas des lymphocytes T8, la cellule anormale présente l'épitope lié à son CMH-I. Le TCR du lymphocyte T8 reconnaît le CMH-I (soi) et l'épitope (non-soi). Ce contact entraîne une activation du lymphocyte T8 et la sélection clonale des lymphocytes T8 possédant un TCR reconnaissant l'épitope. Par conséquent, un lymphocyte T activé se divise pour former un clone de lymphocytes T fonctionnels (réaction primaire) dont la durée de vie est courte. Une partie des cellules du clone deviennent des lymphocytes T mémoires responsables des réactions secondaires lors d'une nouvelle exposition au même antigène (figure 2.10¹⁸, a). Quand le lymphocyte T8 est activé, il libère des perforines qui s'insèrent dans la membrane de la cible et des granzymes. Les perforines forment des pores permettant aux granzymes d'entrer et de tuer la cellule cible (figure 2.10, b).

18. Source : Biologie (Faucher *et al.*, 2011).

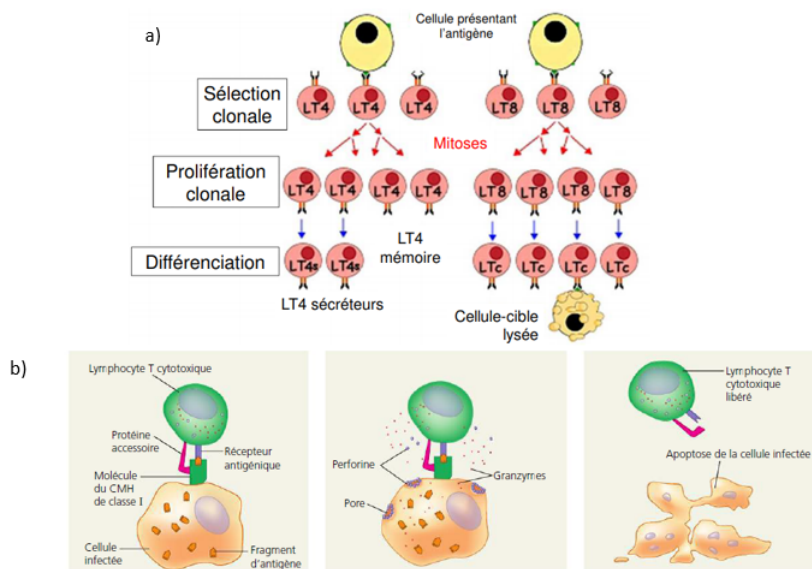


FIGURE 2.10 – Les principaux types de lymphocytes T. a) Différenciation des lymphocytes T. b) La destruction d'une cellule hôte infectée par des lymphocytes T cytotoxiques.

Les lymphocytes T4 activés libèrent l'**Interleukine-2** (IL-2) qui contribuent aussi à la prolifération des lymphocytes T8. Non seulement les lymphocytes T4 jouent un rôle dans la réaction immunitaire cellulaire, mais aussi dans la réaction immunitaire humorale. En effet, les lymphocytes T4 activent aussi les lymphocytes B par contact et par l'IL-2. Autrement dit, les lymphocytes T4 sont les cellules « *Maestro* » de toute l'immunité spécifique, sans elles il n'y aurait aucune réaction immunitaire spécifique (figure 2.11¹⁹).

19. Source : Le système immunitaire (Muller, 2015).

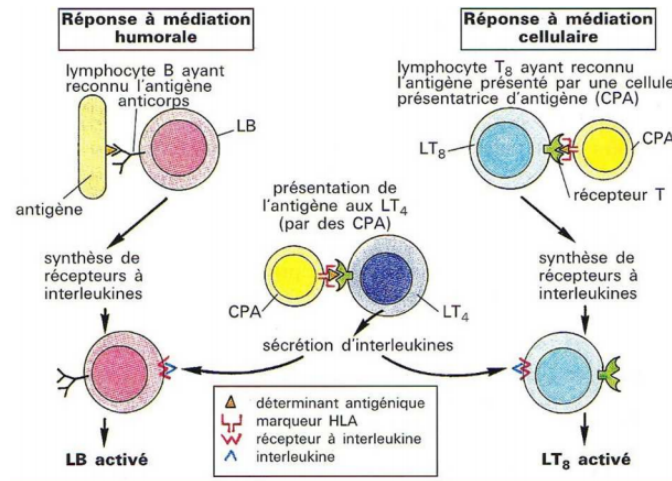


FIGURE 2.11 – Rôle des lymphocytes T4.

L'immunité adaptative qui prolonge l'immunité innée se fait par médiation humorale (action des anticorps) ou par médiation cellulaire (action des lymphocytes T8). Les cellules effectrices de cette immunité sont mises en place après exposition à un élément étranger grâce aux processus de sélection, d'amplification et de différenciation clonale. Enfin, l'immunité innée intervient à nouveau à la fin de l'immunité adaptative, soit pour participer à la destruction de l'élément étranger soit pour éliminer les produits de sa destruction. Il y a une coopération constante entre les deux immunités, la figure 2.12 ci-dessous, résume la réaction immunitaire adaptative.

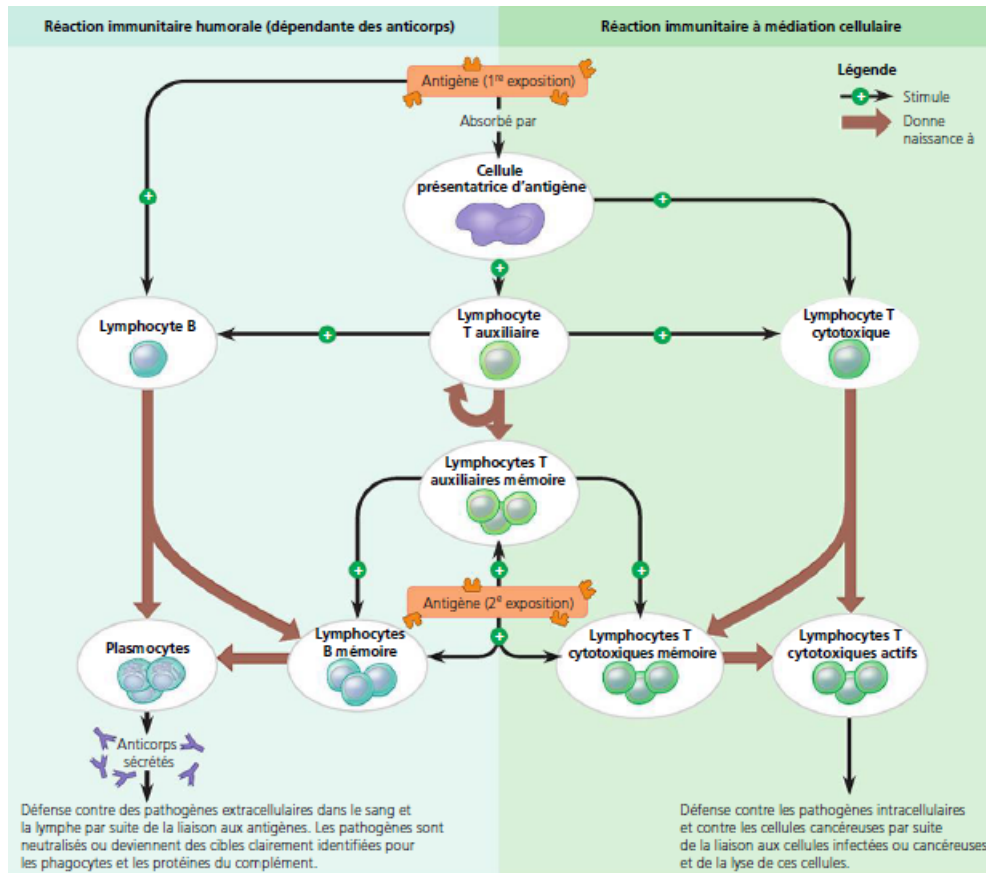


FIGURE 2.12 – Réaction immunitaire adaptative.

2.5 La transplantation rénale

Dans le cas d'insuffisance rénale terminale, les patients risquent de mourir s'il n'y a pas d'intervention. Cette dernière peut se faire par dialyse ou en poursuivant une transplantation rénale, ce qui améliore la survie et la qualité de vie des patients jugés éligibles pour subir une transplantation. Le principal obstacle physiologique de la transplantation est le rejet de l'organe greffé qui résulte d'une réaction immunitaire protectrice normale du sujet receveur, reconnaissant l'organe greffé comme un non soi.

En 1906, Mathieu Jaboulay, un médecin lyonnais spécialisé en chirurgie vasculaire, est le premier à essayer d'appliquer à l'homme les expérimentations animales. Deux tentatives de **xénogreffes**²⁰ de reins chez deux femmes atteintes d'insuffisance rénale se soldent par des échecs (Mordant, 2007). Ces échecs ont permis de découvrir un des principaux obstacles à la greffe : le rejet, dont l'origine a été attribuée à un problème immunologique dès les années 1950. Les premières tentatives pour contourner ce problème consistaient à réaliser une irradiation de l'organisme receveur, conduisant à la destruction quasi totale des cellules immunocompétentes et permettant une acceptation prolongée du greffon. Différents médicaments anti-rejets ont été testés par la suite, dont les corticoïdes, permettant d'effectuer des transplantations chez l'homme sans avoir recours à l'irradiation. C'est en 1952 qu'a eu lieu la première transplantation rénale à partir d'un donneur vivant, sur le jeune Marius Renard par l'équipe de Louis Michon à l'Hôpital Necker à Paris, les suites néphrologiques étant assurées par Jean Hamburger. Le jeune homme est décédé 21 jours plus tard. Au Canada, la première transplantation rénale entre deux vraies jumelles a été réalisée avec succès en 1958 par le chirurgien canadien John Dossetor à l'Hôpital Royal Victoria de Montréal (Transplant Québec, 2013). Le succès de cet événement a permis d'identifier les antigènes des systèmes ABO et CMH comme étant les antigènes majeurs de la transplantation, présents dans la plupart des tissus de l'organisme.

2.5.1 La transplantation rénale et la compatibilité du groupe sanguin (ABO) du donneur-receveur

Les antigènes ABO sont largement exprimés sur les membranes des globules rouges et des cellules tissulaires, ainsi que dans la salive et les fluides corporels. Il existe un risque plus élevé de perte du greffon dans la transplantation rénale ABO-

20. greffes entre espèces différentes.

incompatible puisque les anticorps stimulés peuvent se lier directement aux antigènes des groupes sanguins sur la surface endothéliale rénale et provoquer un rejet aigu. Cela peut également provoquer une agrégation plaquettaire (en anglais *platelet aggregation*) ultérieure, conduisant à un rejet chronique, et finalement à un échec de l'allogreffe (en anglais *allograft failure*). Le groupe sanguin du donneur doit être compatible avec celui du receveur, si c'est le cas l'analyse du typage tissulaire sera faite. Le tableau 2.2, explique à quel group sanguin le rein du receveur doit appartenir pour être compatible avec celui du donneur (🔴) et lors de quelle combinaison une réaction immunitaire se produit (croix ✖). Les organes de personnes de groupe sanguin O peuvent être attribués à tous les patients, quel que soit leur groupe sanguin ; inversement, ces personnes ne peuvent être transplantées qu'avec des organes appartenant au groupe sanguin O. Les personnes de groupe sanguin AB ne peuvent donner leurs organes qu'à des receveurs du même groupe sanguin.

TABLEAU 2.2 – Compatibilité du groupe sanguin entre donneur-receveur.

Groupe sanguin des donneurs	A	B	AB	O
Groupe sanguin des receveurs				
A	🔴	✖	✖	🔴
B	✖	🔴	✖	🔴
AB	🔴	🔴	🔴	🔴
O	✖	✖	✖	🔴

2.5.2 Nomenclature du système CMH

Originellement, la diversité antigénique du CMH a été analysée par la technique sérologique²¹, mais depuis l'apparition de la technique PCR (de l'anglais *polymerase-chain reaction*), la diversité génétique a pu être étudiée au niveau de l'ADN génomique par des techniques de biologie moléculaire. Notons que cette section est tirée de l'article « *The changing landscape of HLA typing : Understanding how and when HLA typing data can be used with confidence from bench to bedside* » (Baxter-Lowe, 2021).

La nomenclature HLA (ou CMH) doit respecter des critères précis qui permettent son homogénéité entre les résultats au niveau international. Elle est tenue à jour par le comité international « *WHO nomenclature committee* ». Le système du Comité de nomenclature des allèles CMH est mis à jour au fur et à mesure que la connaissance du polymorphisme CMH s'améliore. Le système de nomenclature actuel utilise deux à quatre champs de nombres pour nommer chaque allèle (figure 2.13²²). Les deux premiers champs de nombres sont utilisés pour nommer une protéine CMH avec une séquence d'acides aminés unique. Le premier champ provient des assignations du typage sérologique, et s'est développé pour définir une famille d'allèles qui présentent une certaine similitude avec les allèles initialement assignés au champ. Le deuxième champ distingue les allèles au sein de la famille qui codent pour des séquences d'acides aminés uniques. Le troisième champ est utilisé pour nommer les allèles qui codent pour la même séquence d'acides aminés, mais ont des mutations silencieuses dans la région codante. Le quatrième champ est utilisé pour nommer les variations des introns. Les allèles avec une expression anormale

21. L'étude des antigènes à la surface des cellules.

22. Source : <http://clement.ad.free.fr/fac/immuno/roneo.cours2.cmh.pdf>

ont un suffixe constitué d'une lettre qui décrit les variantes d'expression.

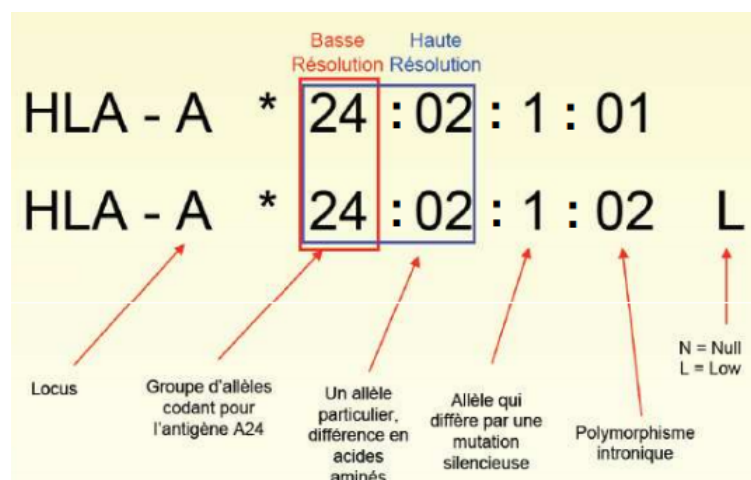


FIGURE 2.13 – Nomenclature des allèles HLA 2010 (WHO), l'étoile indique l'identification par biologie moléculaire.

2.5.3 La compatibilité épitopique CMH (ou HLA) en transplantation rénale

La mise en évidence de l'histocompatibilité ou incompatibilité entre donneur et receveur permet une réduction du risque de réactions immunitaires délétères pour le greffon, pouvant conduire à son rejet. L'analyse du typage tissulaire permet de déterminer la compatibilité donneur-receveur, c'est-à-dire s'ils ont des similarités génétiques qui indiquent que la greffe pourrait réussir. On parle alors du système CMH. On dit que deux individus sont compatibles s'ils partagent le même CMH, c'est-à-dire que leurs tissus sont immunologiquement compatibles les uns avec les autres. Par conséquent, chez la grande majorité des receveurs, la greffe d'un tissu ou d'un organe déclenche une réaction de rejet parce que quelques molécules du CMH sur le tissu du donneur sont étrangères au receveur. Pour atténuer le risque de rejet dans le cas de transplants non identiques, il faut donc utiliser le tissu d'un donneur dont les molécules du CMH présentent un maximum de compatibilité

avec celles du receveur. Puisque nous ne pouvons pas obtenir une compatibilité parfaite du CMH, pratiquement tous les patients ont besoin d'une suppression immunitaire. La disponibilité d'un panel d'anticorps monoclonaux anti-CMH a permis de mieux comprendre le concept d'épitopes. Ces épitopes dits structurels sont composés de 15 à 25 résidus d'acides aminés qui interagissent avec six régions déterminant la complémentarité (CDR pour *complementary determining regions*), situées sur les segments variables des chaînes lourdes et légères des immunoglobulines. Si la plupart des résidus peuvent être remplacés par d'autres sans influencer le lien épitope-anticorps, un changement d'une partie fonctionnelle de l'épitope peut être responsable d'un changement de la faculté de reconnaissance de l'anticorps. Ces épitopes fonctionnels, correspondant de 1 à 5 résidus d'acides aminés sont nommés **éplets** et sont essentiellement situés au niveau des domaines $\alpha 1$ et $\alpha 2$ pour les molécules CMH de classe I, et $\alpha 1$ et $\beta 1$ pour les molécules CMH de classe II. Ils sont très variables selon les allèles et sont responsables de l'entraînement de la majorité du polymorphisme des allèles (Lemieux *et al.*, 2021).

On distingue les épitopes continus ou linéaires formés de plusieurs acides aminés consécutifs et les épitopes discontinus ou conformationnels constitués d'acides aminés discontinus rapprochés par le repliement de la protéine (figure 2.14). Il existe un registre (*HLA Epitope Registry*) contenant la liste des épitopes connus à ce jour. Il différencie les épitopes exposés ou non à la reconnaissance anticorps, et les épitopes vérifiés ou non comme étant capable de lier des anticorps spécifiques.

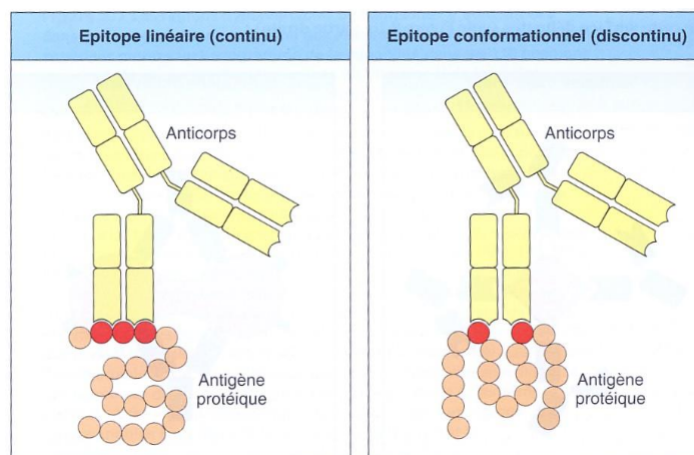


FIGURE 2.14 – Les types d'épitopes.

L'identification de ces épitopes se fait directement dans leur conformation native par les lymphocytes B par l'intermédiaire de leur immunoglobuline de surface formant leur récepteur BCR. Le complexe antigène-anticorps du donneur est appréhendé pour être chargé sur les molécules CMH du receveur, afin d'être présenté aux lymphocytes T CD4 effecteurs. La coopération de lymphocytes T et B mène à la production d'anticorps spécifiquement dirigés contre les épitopes reconnus du donneur.

2.5.4 Relation entre les allèles CMH et les épitopes

Dans cette section, nous présentons un exemple qui explique la relation entre les allèles d'un génotype et les épitopes associés. Nous allons montrer qu'il existe des épitopes partagés inter-intra locus. Considérons un génotype homozygote pour le locus B et hétérozygote pour les loci A et C (tableau 2.3). En se référant au registre CMH des épitopes (*HLA Epitope Registry*), ce génotype est associé à 451 épitopes. Le tableau 2.4 montre un aperçu de quelques épitopes associés au génotype. La description détaillée de tous les épitopes se trouve en annexe A.1.

La présence d'un épitope dans ce génotype est représenté par la valeur 1 et par 0 en cas de son absence. *C1* et *C2* représentent respectivement les classe I et II, *abv* désigne que l'anticorps a été vérifié cliniquement et qu'il est important dans la transplantation rénale, et *oth* désigne autre (l'anticorps n'a pas été vérifié). Les épitopes partagés entre A*01 :02 et A*36 :03 sont bv.44KM, abv.62QE, oth.80T, on appelle ce genre de partage **un partage intra-locus**. Pour ce même génotype il existe deux types de partage **inter-locus** soient les épitopes abv.41T, oth.99Y et oth.77N pour les locus B*13 :08 et C*07 :69, et les locus A*01 :02 et C*17 :05 respectivement.

TABLEAU 2.3 – Exemple d'un génotype avec 5 loci.

A*1	A*2	B*1	B*2	C*1	C*2	DRB1*1	DRB1*2	DQB1*1	DQB1*2
A*01 :02	A*36 :03	B*13 :08	B*13 :08	C*17 :05	C*07 :69	DRB1*11 :01	DRB1*11 :01	DQB1*03 :01	DQB1*03 :01

TABLEAU 2.4 – Aperçu de quelques épitopes associés au génotype de l'exemple 2.3.

C1.abv.21H	C1.abv.41T	C1.abv.44KM	C1.abv.44RMA	C1.abv.44RT
0	1	1	1	0
C1.oth.9H	C1.oth.9S	C1.oth.9T	C1.oth.9Y	C1.oth.11AV
0	0	0	1	1
C2.abv.25R	C2.abv.28H	C2.abv.31FYY	C2.abv.31I	C2.abv.37L
1	0	1	0	0
C2.DR.oth.26L	C2.oth.26Y	C2.DR.oth.28D	C2.oth.28DHF	C2.oth.28DY
0	0	1	0	1

CHAPITRE III

L'ALGORITHME DE L'APPARIEMENT DES ÉPITOPES (CPIE)

Dans ce chapitre, nous présentons un nouvel algorithme statistique cPIE (*calculated Panel of Incompatible Epitopes*), basé sur la compatibilité d'épitype (une série d'épitopes). Il permet de déterminer la probabilité de trouver un donneur compatible, au niveau des épitypes pour un candidat donné, au niveau de la population. Nous allons nous baser, dans un premier temps, seulement sur les épitypes présents dans le bassin des donneurs pour le calcul de la compatibilité épitypique. Nous allons, par la suite, décrire la modification de l'approche dans le cas où nous avons accès aux fréquences des génotypes.

3.1 cPIE : Approche de compatibilité épitypique

La reconnaissance des épitopes qui diffèrent entre donneur et receveur, en termes de positions polymorphes pouvant entraîner la formation d'anticorps anti-HLA, peut être déterminée par le logiciel nommé HLAMatchmaker ((Duquesnoy, 2002), (Duquesnoy et Marrari, 2002) et (Duquesnoy, 2007)). Ce dernier détermine quels épitopes sont partagés ou différents entre les molécules CMH d'un donneur et d'un receveur. En complément du HLAMatchmaker, nous présentons le développement mathématique de l'algorithme statistique cPIE, basé sur l'appariement (ou en anglais *matching*) d'épitopes, qui nous permet de déterminer la probabilité de trouver

un donneur dit ABO-épitype¹ compatible, au niveau de la population pour un candidat donné.

Afin d'illustrer le concept de cet algorithme nous allons introduire quelques notations. Soient G_R le génotype du receveur et G_D le génotype du donneur. On note E_R et E_D l'épitype du receveur et l'épitype du donneur, respectivement. Nous définissons l'incompatibilité des épitopes entre donneur et receveur par :

$$m = E_D \setminus E_R = \{e_i \mid e_i \in E_D \wedge e_i \notin E_R\} = \begin{cases} \{e_1, e_2, \dots, e_k\} & \text{si } k > 0, \\ \emptyset, & \text{sinon,} \end{cases}$$

où k est le nombre d'épitopes dans lesquels une incompatibilité des épitopes s'est produite. Ainsi, m représente tous les épitopes du donneur qui ne sont pas partagés avec le receveur lorsque l'on considère à la fois la comparaison intra et inter locus.

3.1.1 Le calcul de la compatibilité épitypique en se basant sur les épitypes présents dans le bassin des donneurs

Dans cette section, nous présentons le développement mathématique et les étapes de l'algorithme cPIE. Nous considérons des épitopes intra-locus et inter-locus partagés pour produire l'épitype pour chaque individu. L'algorithme d'appariement d'épitype se résume, dans ce cas-ci, en 5 étapes principales, à savoir, 1) le calcul de la fréquence d'épitype, 2) le calcul d'incompatibilité d'épitopes dans le bassin des donneurs en fonction de l'épitype, 3) le calcul de compatibilité dans le bassin des donneurs en fonction de l'épitype pour un seuil d'incompatibilité donné, 4) le calcul de la probabilité de compatibilité étant donné un certain seuil d'incompatibilité. Enfin, il faut tenir compte dans ce calcul de la probabilité de la fréquence

1. Source : Un épitype représente une combinaison d'éplets déduits du génotype d'un individu.

du groupe sanguin dans la population. L'algorithme est implémenté dans le logiciel R et est décrit plus en détails en annexe B.1. Les étapes principales sont les suivantes :

— **Estimation de la fréquence épitypique :**

Nous estimons la fréquence d'épitype dans le bassin des donneurs comme suit :

$$f_E = \frac{n_e}{N}, \quad (3.1)$$

où n_e représente le nombre de sujets qui possèdent l'épitype E dans le bassin des donneurs et N constitue la taille du bassin des donneurs.

— **Probabilité d'appariement d'un candidat :**

Pour un candidat avec un épitype E_R pour lequel on veut calculer la probabilité de compatibilité, étant donné un seuil t d'incompatibilité, la probabilité de compatibilité est donnée par :

$$P_{E_R}^{(t)} = \sum_{i=1}^n f_{E_D} \mathbb{1} \left\{ \sum_{j=1}^l ((E_D^i - E_R)_j)_+ \leq t \right\}, \quad (3.2)$$

où

- f_{E_D} : la fréquence de l'épitype E_D dans la population des donneurs ;
- n : le nombre d'épitypes distincts dans le bassin des donneurs ;
- l : le nombre de l'ensemble de tous épitopes dans l'épitype ;
- $(E_D^i - E_R)_j$: Le $j^{\text{ième}}$ élément du vecteur $(E_D^i - E_R)$;
- $\mathbb{1} \{x \leq t\}$: la fonction indicatrice définie par :

$$\mathbb{1} \{x \leq t\} = \begin{cases} 1 & \text{si } x \leq t, \\ 0 & \text{si } x > t; \end{cases}$$

- $(x)_+$: la fonction définie par :

$$(x)_+ = \begin{cases} 0 & \text{si } x < 0, \\ x & \text{si } x \geq 0. \end{cases}$$

Ainsi, un candidat est compatible avec un donneur du bassin des donneurs lorsqu'il inclut l'épitope du donneur, ou lorsque la distance (nombre d'incompatibilité d'épitopes avec le donneur) est inférieure ou égale au nombre toléré d'incompatibilité (un seuil t). En utilisant la fonction de compatibilité du package cPIE, nous pouvons trouver la distance entre donneur-receveur (E_D, E_R).

— **Considération du groupe sanguin :**

Pour atteindre également la compatibilité au niveau du groupe sanguin, la probabilité sera réduite d'un facteur de w qui représente la fréquence du groupe sanguin du receveur dans la population. Plus précisément, soit b_r l'un des quatre groupes sanguins du candidat, alors on peut définir $w_{b_r} = \mathbb{P}(b_r)$. On suppose que l'appartenance à n'importe quel groupe sanguin est indépendante d'avoir un épitype spécifique dans n'importe quelle population. Cela signifie pour les receveurs que

$$\mathbb{P}(b_r \cap E_R) = \mathbb{P}(b_r) * \mathbb{P}(E_R).$$

En tenant compte du groupe sanguin, la probabilité de compatibilité (ou d'appariement) d'un candidat est donnée par :

— Candidat avec le groupe sanguin AB (compatible avec tous les groupes sanguins, tableau 2.2) :

$$P_{E_R, AB}^{(t)} = (w_A + w_B + w_{AB} + w_O) * P_{E_R}^{(t)}, \quad (3.3)$$

où w_A est la probabilité que le groupe sanguin A soit dans la population et il en va de même pour w_B, w_{AB} et w_O .

— Candidat avec le groupe sanguin A (compatible avec les groupes sanguins A et O , tableau 2.2) :

$$P_{E_R, A}^{(t)} = (w_A + w_O) * P_{E_R}^{(t)}; \quad (3.4)$$

- Candidat avec le groupe sanguin B (compatible avec les groupes sanguins B et O , tableau 2.2) :

$$P_{E_R, B}^{(t)} = (w_B + w_O) * P_{E_R}^{(t)}; \quad (3.5)$$

- Candidat avec le groupe sanguin O (compatible seulement avec le groupe sanguin O , tableau 2.2) :

$$P_{E_R, O}^{(t)} = w_O * P_{E_R}^{(t)}. \quad (3.6)$$

Notons que par soucis de simplification, par exemple, pour l'équation 3.6, nous avons considéré pour les receveurs du groupe sanguin O seulement des donneurs du même groupe sanguin. Il existe des donneurs du groupe sanguin $A2$ (Elkhalidy et Clarke, 2020) (un sous groupe du groupe sanguin A) qui peuvent potentiellement faire un don au groupe sanguin O .

Nous avons implémenté les cinq étapes à l'aide du logiciel R, et pour ne pas épuiser le bassin des donneurs d'un groupe sanguin particulier, nous avons intégré une fonctionnalité dans notre algorithme appelée **compatibilité du groupe sanguin unique**. Étant donné qu'un donneur de groupe sanguin O est compatible avec tous les autres groupes sanguins, et qu'un receveur de groupe sanguin O ne peut être compatible qu'avec un donneur du même type, le bassin des donneurs va s'épuiser rapidement de donneurs du groupe sanguin O . Ainsi, les receveurs appartenant au groupe sanguin O auront alors moins de chance de trouver des donneurs s'ils sont loin dans la liste d'attente. En considérant cette fonctionnalité, les organes des donneurs, de chacun des groupes sanguins, ne peuvent être attribués qu'à des receveurs du même groupe sanguin. Ainsi, les probabilités de compatibilité d'un candidat, en tenant compte des compatibilités du groupe sanguin uniques deviennent :

- Pour le groupe sanguin AB : $P_{E_R, U_{AB}}^{(t)} = w_{AB} * P_{E_R}^{(t)}$;
- Pour le groupe sanguin A : $P_{E_R, U_A}^{(t)} = w_A * P_{E_R}^{(t)}$;

- Pour le groupe sanguin B : $P_{E_R, U_B}^{(t)} = w_B * P_{E_R}^{(t)}$;
- Pour le groupe sanguin O : $P_{E_R, U_{AB}}^{(t)} = w_O * P_{E_R}^{(t)}$.

Exemple illustratif 1 :

Pour un candidat donné, ce qui est observé en pratique est son groupe sanguin et son génotype pour la classe I et la classe II. Dans cet exemple, nous considérons un candidat avec les caractéristiques suivantes :

- Groupe sanguin O ;
- Le génotype transformé en épitype E_R , pour la classe I, donné au tableau 3.1.

Nous prenons un bassin de sept donneurs (tableau 3.2) avec les E_D correspondants et nous fixons le seuil d'incompatibilité à $t = 4$.

TABLEAU 3.1 – Épitype du receveur constitué de 10 épitopes. La présence d'un épitype est représentée par la valeur 1 et par la valeur 0 en cas de son absence.

Épitype du receveur										
id	C1.abv.107W	C1.abv.127K	C1.abv.131S	C1.abv.138K	C1.abv.138MI	C1.abv.143S	C1.abv.144K	C1.abv.144KR	C1.abv.144QL	C1.abv.144TKH
1	0	1	1	0	0	0	1	0	1	1

TABLEAU 3.2 – Bassin des épitypes des donneurs. L'épitype est constitué de 10 épitopes. Pour chaque donneur, la présence d'un épitype est représentée par la valeur 1 et par la valeur 0 en cas de son absence.

Épitype du donneur										
id	C1.abv.107W	C1.abv.127K	C1.abv.131S	C1.abv.138K	C1.abv.138MI	C1.abv.143S	C1.abv.144K	C1.abv.144KR	C1.abv.144QL	C1.abv.144TKH
1	0	1	1	0	0	0	0	1	1	1
2	0	1	1	0	0	0	0	1	1	1
3	1	1	0	1	1	0	1	1	0	1
4	1	1	0	0	1	1	0	0	1	1
5	1	1	1	1	1	1	0	0	0	0
6	0	0	0	1	1	1	0	0	0	1
7	1	0	0	1	1	1	0	1	0	0

TABLEAU 3.3 – Population des épitypes distincts. La présence d'un épitype est représentée par la valeur 1 et par la valeur 0 en cas de son absence.

épitype distinct													
E_D^i	Cl.abv.107W	Cl.abv.127K	Cl.abv.131S	Cl.abv.138K	Cl.abv.138MI	Cl.abv.143S	Cl.abv.144K	Cl.abv.144KR	Cl.abv.144QL	Cl.abv.144TKH	Incompatibilité	f_E	
1	0	1	1	0	0	0	0	1	1	1	1	0.2857	
2	1	1	0	1	1	0	1	1	0	1	4	0.1428	
3	1	1	0	0	1	1	0	0	1	1	3	0.1428	
4	1	1	1	1	1	1	0	0	0	0	4	0.1428	
5	0	0	0	1	1	1	0	0	0	1	3	0.1428	
6	1	0	0	1	1	1	0	1	0	0	5	0.1428	

Le tableau 3.2, représente le bassin des donneurs et leurs épitypes correspondants. Dans le tableau 3.3, chaque ligne représente un épitype distinct dans la population des épitypes, et f_{E_D} représente la fréquence empirique de l'épitype (équation 3.1). Puisque les donneurs 1 et 2 ont exactement le même épitype (tableau 3.2), ce dernier va être comptabilisé une seule fois dans le tableau 3.3. La colonne incompatibilité, représente la quantité $\sum_{j=1}^l ((E_D^i - E_R)_j)_+$ de l'équation 3.2. Par exemple, pour l'épitype 1, on a $i = 1$, avec $l = 10$ épitypes. Ainsi, $\sum_{j=1}^{l=10} ((E_D^1 - E_R)_j)_+ = 1$, ce qui correspond à une seule incompatibilité. Notons que les fréquences empiriques de l'épitype sont élevées car nous avons seulement considéré un bassin de 6 épitypes distincts, pour illustration. Ceci n'est pas vrai en générale. Les lignes vertes représentent les épitypes qui sont admissibles au seuil de 4 et donc elles sont considérées dans le calcul. C'est-à-dire la colonne d'incompatibilité qui satisfait à la fonction indicatrice de l'équation 3.2. En prenant toujours le premier épitype du tableau 3.3, on a $\sum_{j=1}^{l=10} ((E_D^1 - E_R)_j)_+ = 1$ et donc $\mathbb{1} \left\{ \sum_{j=1}^{l=10} ((E_D^1 - E_R)_j)_+ \leq 4 \right\} = 1$. Par contre, l'épitype 6 ne satisfait pas à cette condition et ne sera pas pris en compte. En effet, $\sum_{j=1}^{l=10} ((E_D^6 - E_R)_j)_+ = 5$ et donc $\mathbb{1} \left\{ \sum_{j=1}^{l=10} ((E_D^6 - E_R)_j)_+ > 4 \right\} = 0$. Ainsi, la probabilité de trouver un donneur

compatible au niveau de l'épitype en utilisant l'équation 3.2 est :

$$\begin{aligned}
 P_{E_R}^{(4)} &= \sum_{i=1}^5 f_{E_D} \mathbb{1} \left\{ \sum_{j=1}^{10} ((E_D^i - E_R)_j)_+ \leq 4 \right\} \\
 &= 0.2857 + 0.1428 + 0.1428 + 0.1428 + 0.1428 \\
 &= 0.8569.
 \end{aligned}$$

Si l'on suppose que la population sous étude a une fréquence de 0.4 pour le groupe sanguin O et que le receveur possède ce groupe sanguin, alors on a $w_O = 0.4$. En considérant ceci dans les calculs (équation 3.6), nous trouvons

$$P_{E_R, O}^{(4)} = 0.4 * P_{E_R}^{(4)} = 0.34276.$$

Donc la probabilité de trouver, pour ce patient, un donneur compatible, en supposant que les fréquences du groupe sanguin et de l'épitype soient indépendants, est de 34.27%, pour un seuil d'incompatibilité égal à 4.

3.1.2 Le calcul de la compatibilité épitypique en se basant sur les fréquences des génotypes dans le bassin des donneurs

Dans cette section, nous considérons avoir accès aux fréquences des génotypes estimées d'une quelconque façon à l'extérieur des données analysées. Les étapes de l'algorithme sont similaires aux étapes décrites à la section 3.1.1. Le point différent, dans ce cas-ci, est le calcul de la fréquence épitypique qui sera en fonction de la fréquence du génotype.

Les principales étapes de l'algorithme se résument comme suit :

— **Estimation de la fréquence épitypique :**

Soit E un épitype donné, dans la population des donneurs. Nous estimons la fréquence d'épitype dans le bassin de donneurs correspondant en fonction de la fréquence du génotype que nous supposons connue. Ainsi la fréquence

de l'épitope est donnée par :

$$f_E = \sum_{G \in J_E} f_G, \quad (3.7)$$

où J_E représente l'ensemble des génotypes menant à E , c'est-à-dire $J_E = \{G | E_G = E\}$, et f_G représente la fréquence du génotype.

— **Probabilité d'appariement d'un candidat :**

Pour un receveur avec un épitype E_R pour lequel on veut calculer la probabilité de compatibilité, étant donné un seuil t d'incompatibilité, l'expression de la probabilité de compatibilité demeure la même :

$$P_{E_R}^{(t)} = \sum_{i=1}^n f_{E_D} \mathbb{1} \left\{ \sum_{j=1}^l ((E_D^i - E_R)_j)_+ \leq t \right\}, \quad (3.8)$$

où f_{E_D} représente la fréquence de l'épitope E du donneur telle que définie par l'équation 3.7.

— **Considération du groupe sanguin :**

Pour tenir compte du groupe sanguin, la probabilité d'appariement d'un receveur reste la même que celle de la section 3.1.1. Il en va de même pour les probabilités en considérant la compatibilité du groupe sanguin unique.

Exemple illustratif 2 :

Considérons le même exemple qu'à la section 3.1.1, avec les mêmes caractéristiques des donneurs et du receveur, en termes d'épitopes, du groupe sanguin et du seuil d'incompatibilité. Cette fois-ci la fréquence de l'épitope est différente comme le montre le tableau 3.5.

TABLEAU 3.4 – Bassin des épitypes des donneurs avec les fréquences génotypiques correspondantes. L'épitype est constitué de 10 épitopes. Pour chaque donneur, la présence d'un épitope est représentée par la valeur 1 et par la valeur 0 en cas de son absence.

Épitype du donneur											
id	Cl.abv.107W	Cl.abv.127K	Cl.abv.131S	Cl.abv.138K	Cl.abv.138MI	Cl.abv.143S	Cl.abv.144K	Cl.abv.144KR	Cl.abv.144QL	Cl.abv.144TKH	f_G
1	0	1	1	0	0	0	0	1	1	1	0.00012
2	0	1	1	0	0	0	0	1	1	1	0.00023
3	1	1	0	1	1	0	1	1	0	1	0.00054
4	1	1	0	0	1	1	0	0	1	1	0.00063
5	1	1	1	1	1	1	0	0	0	0	0.00034
6	0	0	0	1	1	1	0	0	0	1	0.00051
7	1	0	0	1	1	1	0	1	0	0	0.00069

TABLEAU 3.5 – Population des épitypes distincts. La présence d'un épitope est représentée par la valeur 1 et par la valeur 0 en cas de son absence.

Épitype distinct													
id	Cl.abv.107W	Cl.abv.127K	Cl.abv.131S	Cl.abv.138K	Cl.abv.138MI	Cl.abv.143S	Cl.abv.144K	Cl.abv.144KR	Cl.abv.144QL	Cl.abv.144TKH	Incompatibilité	f_G	f_E
1	0	1	1	0	0	0	0	1	1	1	1	0.00012	0.00035
2	1	1	0	1	1	0	1	1	0	1	4	0.00054	0.00054
3	1	1	0	0	1	1	0	0	1	1	3	0.00063	0.00063
4	1	1	1	1	1	1	0	0	0	0	4	0.00034	0.00034
5	0	0	0	1	1	1	0	0	0	1	3	0.00051	0.00051
6	1	0	0	1	1	1	0	1	0	0	5	0.00069	0.00069

Rappelons que le 1^{er} et le 2^{ème} donneurs du tableau 3.4 ont exactement le même épitype. Parfois des épitypes similaires peuvent résulter de différents génotypes, alors que d'autres épitypes peuvent être uniques à des génotypes particuliers. Ceci est dû au partage des épitopes inter- et intra-loci. Ainsi, dans le tableau 3.5, la fréquence de l'épitype associée sera la somme de la fréquence génotypique de ces deux épitypes, d'où $f_E = 0.00012 + 0.00023 = 0.00035$ (équation 3.7). Pour le reste des épitypes, ils sont distincts, et dans ce cas-ci, la fréquence de l'épitype est la même que la fréquence du génotype. Étant donné un seuil $t = 4$, l'épitype 6 n'est pas considéré et seules les lignes vertes satisfont à la condition $\sum_{j=1}^l ((E_D^i - E_R)_j)_+ \leq 4$ de l'équation 3.8. Ainsi, la probabilité de trouver un

donneur compatible au niveau de l'épitope en utilisant l'équation 3.8 est :

$$\begin{aligned}
 P_{E_R}^{(4)} &= \sum_{i=1}^5 f_{E_D} \mathbb{1} \left\{ \sum_{j=1}^{10} ((E_D^i - E_R)_j)_+ \leq 4 \right\} \\
 &= 0.00035 + 0.00054 + 0.00063 + 0.00034 + 0.00051 \\
 &= 0.00237,
 \end{aligned}$$

et en considérant le groupe sanguin (équation 3.6), nous trouvons

$$P_{E_R,O}^{(4)} = 0.4 * P_{E_R}^{(4)} = 0.000948.$$

Alors, pour le bassin de donneurs considéré, nous avons 0.0948% de chance de trouver un donneur compatible, pour ce receveur, étant donné un seuil d'incompatibilité de 4.

Nous avons présenté les étapes principales de l'algorithme dans les deux approches. Il existe d'autres fonctionnalités, intégrées dans l'algorithme, concernant ce que nous avons appelé des **motifs** ou des stratégies d'appariement ou de compatibilité. Ces motifs caractérisent différents sous-groupes d'épitopes. Par exemple, on peut comparer l'épitope au complet du candidat ou on peut considérer un sous-groupe d'épitopes d'intérêt pour lequel on veut trouver les donneurs compatibles davantage. Ces sous-groupes d'épitopes peuvent contenir seulement les *abv* (des épitopes importants pour la transplantation vérifiés cliniquement), de classe I ou de classe II, ou encore contenir des épitopes, appelés *High risk epitopes* ((Mohammadhasanzadeh et al., 2021) dont je suis co-auteur), qui augmentent le risque de rejet du rein. Après avoir décrit le cPIE, dans la section suivante, nous allons souligner l'avantage du cPIE en allocation des organes.

3.2 L'allocation d'organes et le cPIE

Présentement, en attribution d'organes, on considère ce qu'on appelle un test des **anticorps réactifs au panel** (en anglais *Panel Reactive Antibody* PRA) (Cecka, 2010). C'est un test de routine de dépistage qui mesure le degré de sensibilisation d'un receveur potentiel de rein, à la suite d'une exposition antérieure à des antigènes CMH externes lors de transfusions sanguines, de grossesses ou de transplantations d'organes antécédentes. En suivant cette procédure, les candidats qui peuvent avoir des anticorps préformés contre le CMH du donneur ne se verront pas attribuer un organe. Ainsi, en considérant l'approche de la compatibilité au niveau de l'éplet à l'aide du cPIE permettra d'éviter les anticorps préformés et pourrait également empêcher la formation de nouveaux anticorps post-greffe.

CHAPITRE IV

SIMULATIONS ET RÉSULTATS

Afin de décrire la manière dont la probabilité de trouver un donneur compatible diffère selon l'origine ethnique auto-identifiée, pour différentes stratégies d'appariement issues de notre algorithme, nous avons effectué une analyse par simulation. Cette étude de simulation a pour but d'estimer l'impact de la présence d'une structure de population au niveau des ethnicités, dans l'échantillon sous étude, sur la probabilité de trouver un donneur compatible. Cette analyse nous a permis également de voir si en moyenne, les receveurs d'une ethnicité spécifique ont plus ou moins de chance de trouver un donneur compatible au niveau des épitypes.

Nous allons d'abord décrire la base de données des allèles à partir de laquelle nous avons généré les génotypes. Ensuite, nous allons illustrer l'application du cPIE selon deux stratégies d'appariement. Finalement, nous effectuerons une analyse des données issues de l'algorithme.

4.1 Simulations et préparation des données

Nous avons utilisé la base de données des haplotypes provenant du *NMDP* (National Marrow Donor Program, 2007). Elle présente les allèles et haplotypes CMH dans la population américaine. Ces fréquences ont été mises à jour en décembre 2007. On y trouve, pour chaque ethnicité auto-identifiée, les fréquences alléliques

pour 5 loci ; HLA-A, HLA-B, HLA-C pour la classe I, et HLA-DRB1 et HLA-DQB1 pour la classe II. Les fréquences utilisées du groupe sanguin de la population américaine, pour chaque ethnicité, proviennent du *American Red Cross* (American Red Cross, ARC).

Nous allons simuler une paire d'haplotypes pour produire un génotype d'un donneur. Pour simuler les génotypes d'un patient, nous allons simuler une paire d'haplotypes en tenant compte de l'ethnicité du patient. Ensuite, ce processus est répété plusieurs fois afin de générer notre bassin de donneurs et de patients. Pour chacune des ethnicités (Hispanique, Européenne, Île du pacifique asiatique et Africaine-américaine), un échantillon de 10000 génotypes a été simulé, ce qui a donné un bassin de donneurs de 40000. Pour simuler le bassin des candidats, nous avons généré, pour chacune des ethnicités, un échantillon de 1000 génotypes. Une fois les génotypes simulés, ils sont convertis en épitopes avant d'appliquer le cPIE.

4.2 Application de l'algorithme cPIE

Pour faciliter l'illustration de notre étude, nous considérons seulement la classe I, avec les 3 loci A, B et C. Pour chacune des 4 ethnicités, nous considérons deux stratégies de compatibilité. L'épitope au complet, qui contient 223 épitopes, et le groupe des *abv* de la classe I qui consiste en 72 épitopes. Mentionnons que, dans la littérature, le choix du seuil du nombre d'incompatibilité est subjectif pour chaque étude et est souvent spécifique à chaque étude (*ad hoc*). Nous utilisons plutôt des choix heuristiques de seuils, pour améliorer la capacité de trouver des donneurs compatibles. Nous avons choisi trois valeurs de seuils : 5, 8 et 10. Rappelons que notre variable d'intérêt est la probabilité de compatibilité $P_{E_R}^{(t)}$ en tenant compte du groupe sanguin du candidat.

La figure 4.1 illustre la distribution de la probabilité de compatibilité pour chaque

ethnicité selon les trois seuils d'incompatibilité choisis. Dans le cas de l'épitype au complet, bien que les probabilités de compatibilités médianes des trois seuils semblent fixes, la médiane augmente en augmentant le seuil pour toutes les ethnicités. Cette observation est plus claire dans le groupe *abv* de classe I, en tolérant plus d'incompatibilité, la probabilité de compatibilité médiane augmente. Notons aussi que la variabilité de la probabilité de compatibilité au seuils 8 et 10, dans le cas du groupe *abv* de classe I, semble légèrement plus grande dans l'ethnicité européenne par rapport aux autres ethnicités.

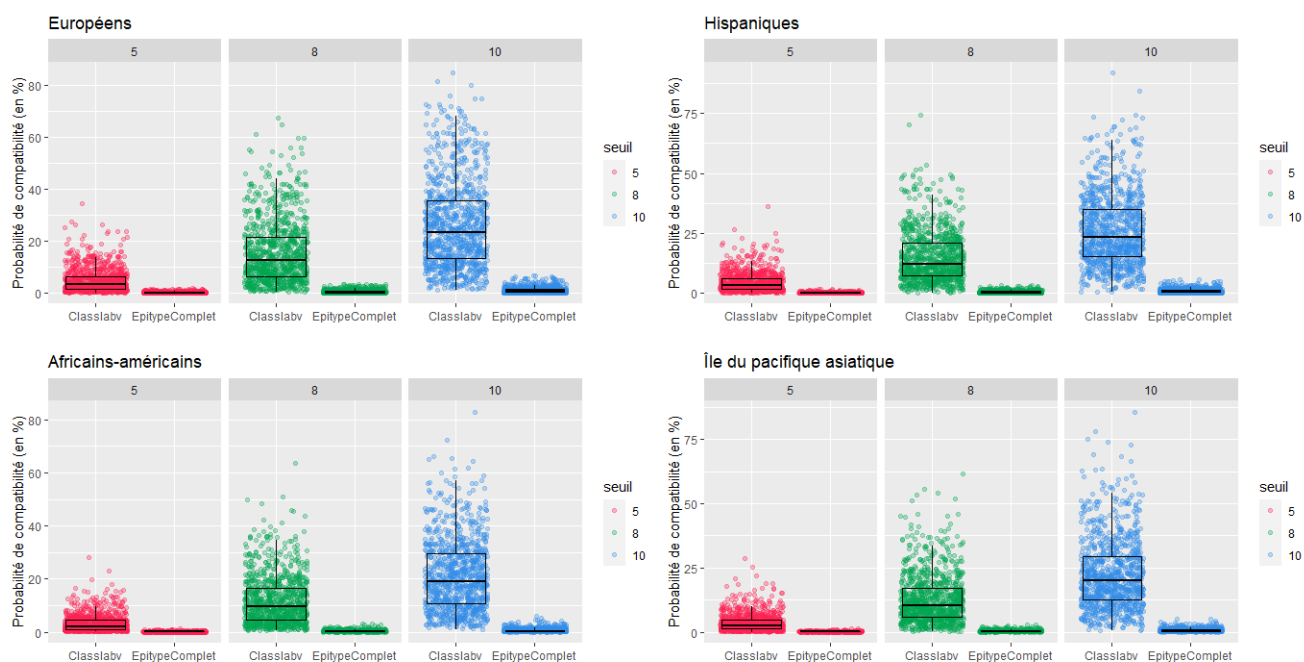


FIGURE 4.1 – Distribution de la probabilité de compatibilité par ethnicité, pour l'épitype au complet et le groupe *abv* de classe I, selon les trois seuils.

Dans l'ensemble, on remarque qu'il y a une grande variabilité lorsqu'on considère la stratégie *abv* de classe I en comparaison avec l'épitype au complet. On pense que cette variabilité est due aux épitopes *abv*. Pour vérifier cette hypothèse, nous avons choisi aléatoirement, dans un premier temps, un sous ensemble d'épitopes de taille

72; dont 23 épitopes sont des *abv*. Ensuite, nous avons choisi aléatoirement un sous ensemble d'épitopes de taille 72 qui ne contient aucun épitope *abv*. La figure 4.2 montre que dans le cas où nous avons 23 épitopes *abv*, la distribution a une allure différente. Pour chaque ethnicité, on remarque la formation de 4 groupes et ces groupes se distinguent plus en augmentant le seuil. Ces quatre groupes correspondent aux fréquences des groupes sanguins et la formation de ces groupes est dû au fait d'avoir une probabilité de compatibilité élevée. En effet, en tolérant un grand nombre d'incompatibilité entre donneur et receveur et en considérant un ensemble petit d'épitopes (23), la probabilité de trouver un donneur compatible augmente et s'approche de 1, et en ajustant pour le groupe sanguin du patient les données reflètent les équations 3, 4, 5 et 6 de la section 3.1.1. Dans le cas où l'on considère le sous ensemble d'épitopes de 72 ne contenant aucun épitope *abv*, l'allure de la distribution de la probabilité de compatibilité n'est pas différente de celle lorsqu'on considère 72 épitopes *abv*. Ainsi la variabilité de la distribution dans le cas des épitopes *abv* n'est pas dû à la présence de ces épitopes en particulier, mais plutôt à la taille des épitopes considérés dans l'appariement d'un donneur et receveur.

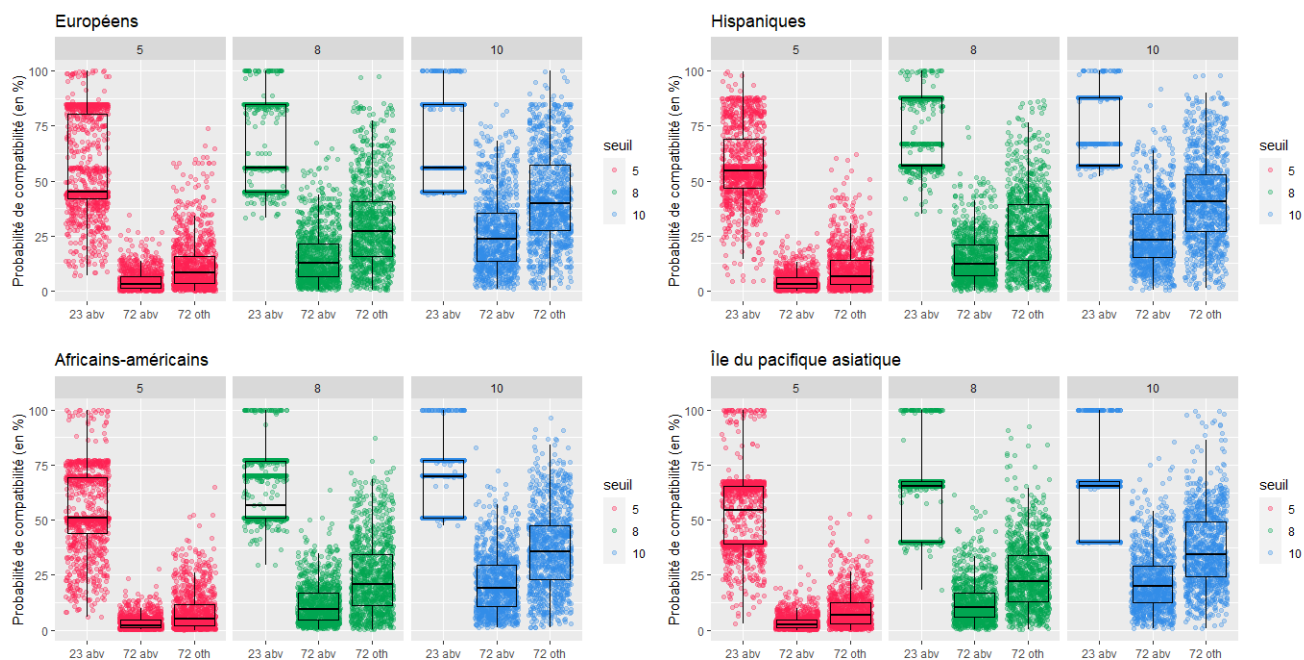


FIGURE 4.2 – Distribution de la probabilité de compatibilité par ethnicité, pour les trois seuils, en considérant : 23 épitopes *abv*, 72 épitopes *abv* et 72 épitopes *oth*.

4.2.1 Analyse des médianes

Dans cette section, puisque la distribution de la compatibilité est asymétrique à droite, nous avons décidé de comparer les probabilités de compatibilité médianes des différentes ethnicités. On s'attendait à observer des distributions différentes. Cette analyse va nous informer aussi dans quelle mesure la considération des épitopes (qui sont plus susceptibles d'être partagés que les allèles) peut avoir un impact sur la probabilité de compatibilité selon différentes ethnicités. Afin d'alléger la notation, on pose :

$$\left\{ \begin{array}{l} 1 = \text{Européens (EUR)} \\ 2 = \text{Africains-américains (AFA)} \\ 3 = \text{Île du pacifique asiatique (IPA)} \\ 4 = \text{Hispaniques (HIS)} \end{array} \right.$$

Nous allons procéder en deux étapes :

- Première étape : On utilise le test de *Kruskal Wallis* pour k échantillons, pour tester si les ethnicités ont des caractéristiques identiques au sens de la médiane. On désigne par $\tilde{\mu}_i$ la médiane de l'ethnicité i , alors notre hypothèse s'écrit sous la forme :

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \dots = \tilde{\mu}_k,$$

$$H_A : \tilde{\mu}_i \neq \tilde{\mu}_j \text{ pour au moins } i \neq j.$$

Nous avons alors toutes les observations $n_1 + n_2 + \dots + n_k = N$ classées de 1 à N , où $k = 1, 2, 3, 4$. La statistique F_{KW} du test de *Kruskal Wallis*¹ est défini par :

$$F_{KW} = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1).$$

La quantité R_i étant la somme des rangs de la $i^{\text{ème}}$ ethnicité. Ainsi, l'hypothèse nulle sera rejetée, à un seuil de signification de $\alpha = 5\%$, lorsque

$$F_{KW} \geq \chi_{1-\alpha}^2(k-1).$$

Si l'hypothèse H_0 est rejetée on passe à la deuxième étape.

- Deuxième étape : Si l'hypothèse H_0 est rejetée alors au moins une ethnicité est différente d'une autre. Pour identifier quelle paire est significativement

1. Les tailles des échantillons sous étude sont assez grandes ($n_k > 900, \forall k$), la distribution de F_{KW} peut être approximée par une loi du χ^2 à $(k-1)$ degrés de liberté lorsque H_0 est vrai.

différente, nous utilisons le test de *Wilcoxon* avec un seuil de signification de $\alpha = 5\%$. Nous avons $\frac{k(k-1)}{2}$ comparaisons à réaliser et nous utilisons la correction de *Bonferroni* avec $\alpha^* = \frac{2\alpha}{k(k-1)}$. Pour chaque paire, le paramètre Δ représente la différence des médianes entre deux ethnicités. Le test s'écrit sous la forme :

$$H_0 : \Delta = 0,$$

$$H_A : \Delta \neq 0.$$

Soit $R(Y_i)$ le rang de Y_i parmi les échantillons combinés, par exemple la paire $(1, 2)$, c'est-à-dire parmi $X_1, X_2, \dots, X_{n_1}, Y_1, Y_1, \dots, Y_{n_2}$. Alors la statistique du test de *Wilcoxon* est

$$W = \sum_{i=1}^{n_2} R(Y_i).$$

Le test rejette H_0 si $W \geq w_{\frac{\alpha^*}{2}}$ ou si $W \leq n_2(n_1 + n_2 + 1) - w_{\frac{\alpha^*}{2}}$. Puisque les tailles des échantillons sous étude sont grandes, l'approximation normale avec une correction de continuité de 0.5 est appliquée. Ainsi on rejette H_0 si

$$W \geq n_2(n_1 + n_2 + 1)/2 + \frac{1}{2} + z_{\frac{\alpha^*}{2}} \sqrt{n_1 n_2 (n_1 + n_2 + 1)/12},$$

ou bien si

$$W \leq n_2(n_1 + n_2 + 1)/2 - \frac{1}{2} - z_{\frac{\alpha^*}{2}} \sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}.$$

- L'épitype au complet :

Pour les trois seuils, on observe que les ethnicités européenne et hispanique ont une probabilité de compatibilité médiane plus grande que celle des ethnicités africaine-américaine et de l'île du pacifique asiatique (figure 4.3 et

2. Le même principe du rejet de H_0 s'applique pour les autres paires considérées.

tableau 4.1). Le test non paramétrique de *Kruskal Wallis* a été conduit pour comparer les probabilités de compatibilité médianes. Le test est significatif (on rejette H_0), il existe au moins une paire pour laquelle les médianes sont différentes. Ensuite, les comparaisons multiples par paires à l'aide du test de *Wilcoxon*, aux seuils 5 et 8, ont montré que les probabilités de compatibilité médianes de toutes les ethnicités sont significativement différentes. Au seuil 10, le test pour la paire (2, 3) est non significatif, la probabilité de compatibilité médiane des patients d'origine africaine-américaine n'est pas statistiquement différente de celle des patients d'origine de l'île du pacifique asiatique.

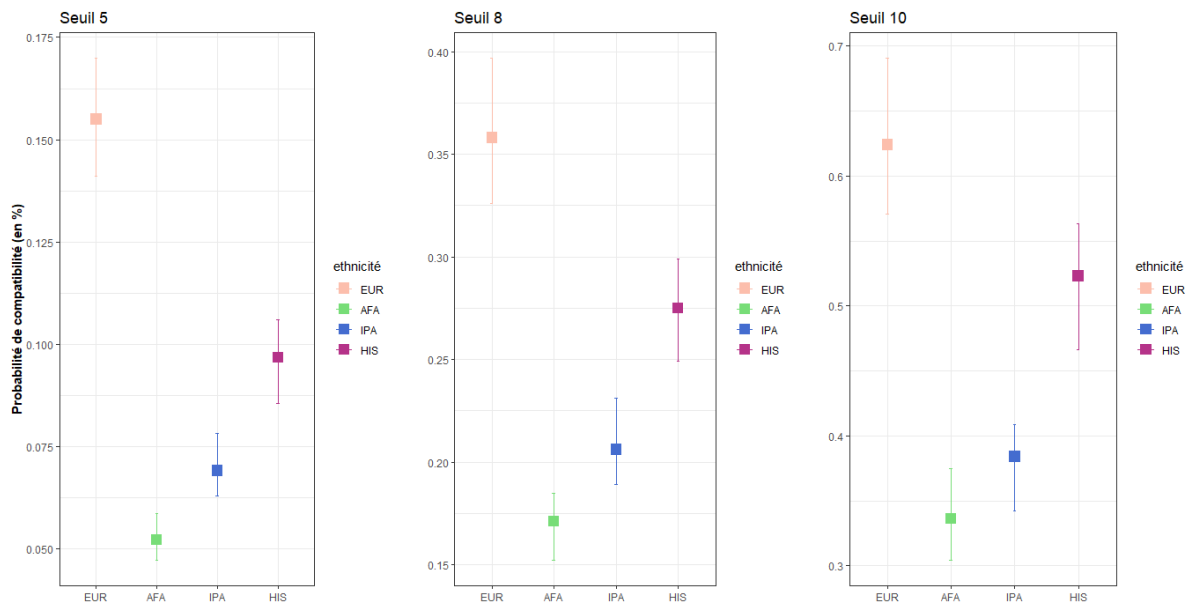


FIGURE 4.3 – Probabilités de compatibilité médianes par ethnicité de l'épitype au complet en considérant les trois seuils.

- Groupe *abv* de classe I :

Le test de *Kruskal Wallis* rejette H_0 , et le test de *Wilcoxon* est non significatif pour les deux paires (1, 4) et (2, 3) pour les trois seuils. On a toujours

la même observation concernant la médiane de probabilité de compatibilité élevée pour les ethnicités européenne et hispanique par rapport aux deux autres ethnicités (figure 4.4).

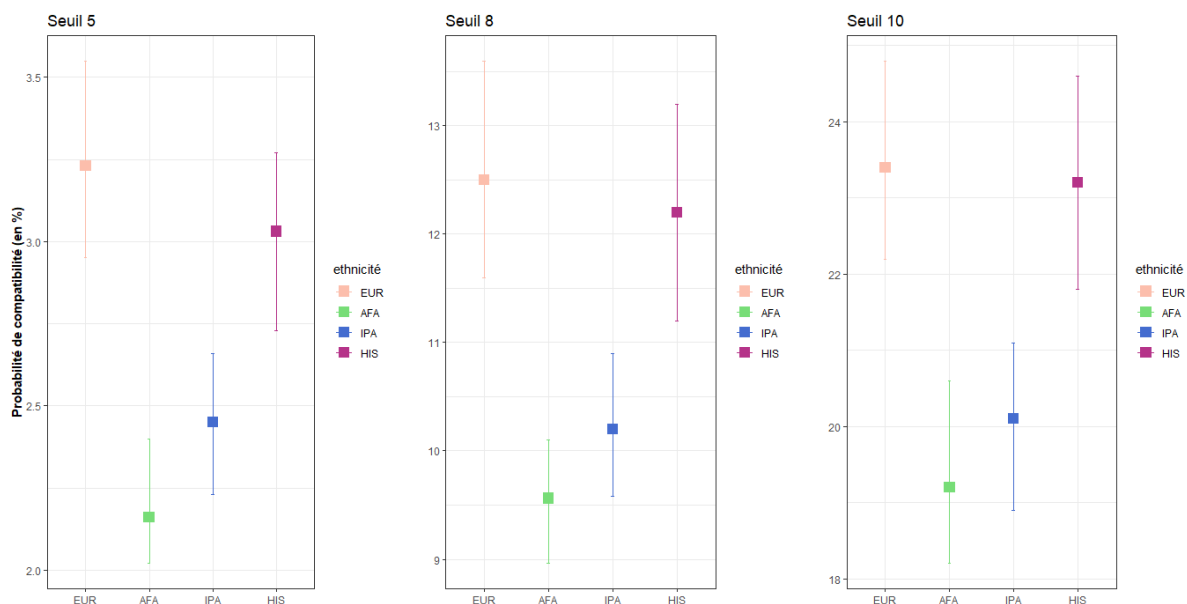


FIGURE 4.4 – Probabilités de compatibilité médianes par ethnicité du groupe *abv* de classe I en considérant les trois seuils.

Pour résumer (tableau 4.1), les candidats d'origine européenne et hispanique ont plus de chance de trouver un donneur compatible par rapport aux autres candidats, peu importe la stratégie d'appariement ou le seuil considérés. Notons que ces résultats, issus des données simulées, ont pour objectif principal l'illustration de la validité de notre algorithme.

TABLEAU 4.1 – Probabilités de compatibilité médianes de l'épitype au complet et le groupe *abv* de classe I, selon les trois seuils, pour chaque ethnicité.

		Seuils		
		5	8	10
AFA	Épitype Complet	0.05%	0.17%	0.33%
	Classe I <i>abv</i>	2.16%	9.56%	19.2%
EUR	Épitype Complet	0.15%	0.36%	0.62%
	Classe I <i>abv</i>	3.23%	12.50%	23.40%
HIS	Épitype Complet	0.10%	0.27%	0.52%
	Classe I <i>abv</i>	3.03%	12.20%	23.20%
IPA	Épitype Complet	0.07%	0.20%	0.38%
	Classe I <i>abv</i>	2.45%	10.20%	20.10%

4.2.2 Analyse des médianes à l'aide du bootstrap non paramétrique

Rappelons que les tests utilisés dans la section sont basés sur les rangs et en pratique d'autres hypothèses³ sont souvent faites sur la forme des populations sous-jacentes. Dans cette section, nous appliquons la méthode du bootstrap non paramétrique pour confirmer les résultats de ces tests.

On s'intéresse à tester, pour chaque paire d'ethnicité et pour chaque seuil, si la différence des probabilités de compatibilité médianes est nulle. L'hypothèse s'écrit

3. Par exemple, une hypothèse commune est appelée le modèle de position (en anglais *the location model*). Ce modèle suppose que deux population *A* et *B* ont la même forme et la même variance et qu'ils diffèrent seulement d'un changement (ou décalage) dans la quantité inconnue de $\tilde{\mu}$ par exemple.

sous la forme :

$$H_0 : \tilde{\mu}_i - \tilde{\mu}_j = 0,$$

$$H_A : \tilde{\mu}_i - \tilde{\mu}_j \neq 0.$$

La quantité $\tilde{\mu}_i - \tilde{\mu}_j$, est estimée par $\tilde{X}_i - \tilde{X}_j = D$, où \tilde{X}_i représente la probabilité de compatibilité médiane de l'ethnicité i à partir de l'échantillon, et D représente la différence des médianes de la probabilité de compatibilité entre les patients d'origine i et ceux d'origine j . Puisque la distribution de la différence des médianes de la probabilité de compatibilité est inconnue, on utilise la méthode du bootstrap pour trouver des approximations des bornes inférieures et supérieures de l'intervalle de confiance (IC). Notons I^* , la borne inférieure approximée de l'IC et S^* la borne supérieure approximée de l'IC. Soient $B = 10000$ le nombre d'échantillons bootstrap générés avec remise, à partir de l'échantillon initial, de taille n_i , où n_i représente la taille de l'échantillon de l'ethnicité i , et M_i^* la probabilité de compatibilité médiane de l'ethnicité i . On obtient le vecteur D_i^* de longueur B . On cherche à trouver la distribution bootstrap de D_i en utilisant $D_{i_{obs}}$ de la paire considérée. En prenant les quantiles 0.025 et 0.975 de la quantité $D_i^* - D_{i_{obs}}$, on trouve $I^* \approx I$ et $S^* \approx S$, ainsi l'IC de bootstrap à 95% est de la forme $(D_{i_{obs}} - S^*, D_{i_{obs}} - I^*)$.

- Différence des probabilités de compatibilité médianes de l'épitype au complet :

En considérant un seuil de 10 (figure 4.5), l'IC à 95% n'inclut pas la valeur 0 et donc pour toutes les paires, la différence des médianes de probabilités de compatibilité est significativement différente de 0, sauf dans le cas où l'on considère un seuil de 10, la différence de probabilités médianes pour la paire (2, 3) est non significative. Ces résultats sont en concordance avec le test non paramétrique de *Wilcoxon* appliqué dans la section précédente (4.2.2).

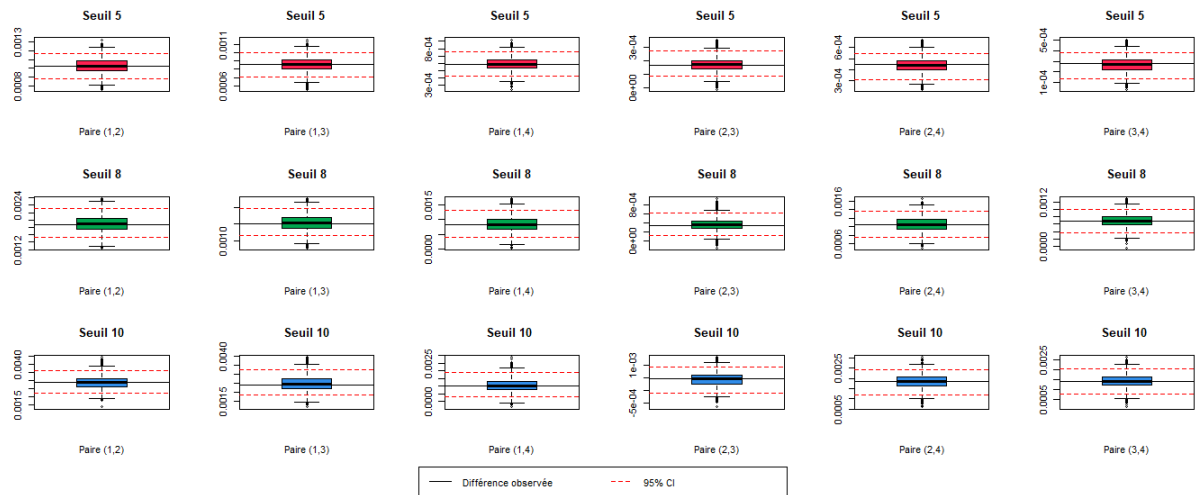


FIGURE 4.5 – Distribution de la différence des probabilités de compatibilité médianes de l'épitype au complet selon les trois seuils. La couleur rouge représente le seuil 5 pour les 6 paires d'ethnicités, la couleur verte représente le seuil 8 et la couleur bleue représente le seuil 10.

- Différence des probabilités de compatibilité médianes du groupe *abv* de classe I :

Peu importe le seuil choisi, on ne rejette pas H_0 pour les paires (1,4) et (2,3) (figure 4.6), ceci est encore une fois en concordance avec le test non paramétrique de *Wilcoxon*. La probabilité de compatibilité médiane des patients d'origine européenne n'est pas statistiquement différente de celle des patients d'origine hispanique et il en va de même pour les patients d'origine africaine-américaine versus les patients d'origine de l'île du pacifique asiatique.

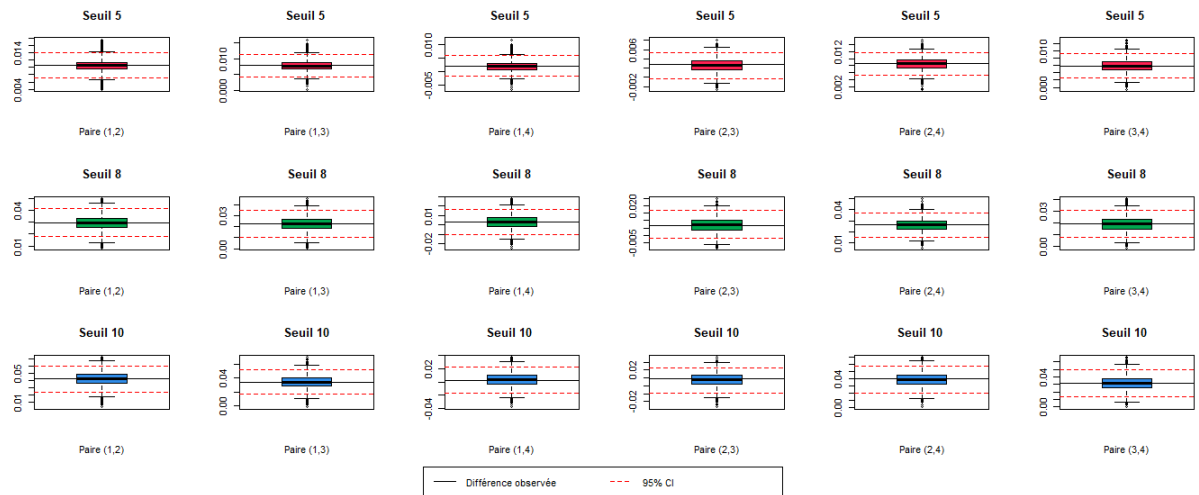


FIGURE 4.6 – Distribution de la différence des probabilités de compatibilité médianes des abv de classe I selon les trois seuils. La couleur rouge représente le seuil 5 pour les 6 paires d’ethnicités, la couleur verte représente le seuil 8 et la couleur bleue représente le seuil 10.

4.2.3 Analyse des fonctions de répartition des ethnicités

Comme mentionné dans la section précédente 4.2.2, la comparaison des probabilités de compatibilité médianes a été effectuée car nous avons l’hypothèse que les distributions des quatre ethnicités proviennent de la même distribution. Dans cette section nous testons cette hypothèse. Nous allons d’abord tracer la fonction de répartition des quatre ethnicités sur le même graphique pour visualiser les courbes, et nous allons présenter un seul exemple pour expliquer l’interprétation du graphique. Ensuite, le test d’*Anderson-Darling* est conduit pour tester notre hypothèse. Le lecteur est invité à consulter *Anderson-Darling k-Sample Test* (Composites Materials Handbook, 2012) pour plus de détails sur la procédure de

ce test.

Considérons, par exemple, le point $x = 0.25$, qui représente une probabilité de compatibilité de 25%, dans la figure 4.7. À ce point, environ 54% des patients d'origine européenne et hispanique ont une probabilité de compatibilité moins de 25%. Pour les ethnicités africaine-américaine et de l'île du pacifique asiatique, environ 65% des patients ont une probabilité de compatibilité moins de 25%. Dans l'ensemble, on remarque que les fonctions de répartition des patients d'origine européenne et hispanique sont proches l'une de l'autre et se distinguent des autres qui, aussi, sont proches l'une de l'autre ce qui est en concordance avec nos observations sur la médiane de la probabilité de compatibilité des ethnicités dans la section précédente 4.2.2. Les figures 4.8 et 4.9 représentent les fonctions de répartition des quatre ethnicités, en considérant l'épitype au complet et le groupe *abv* de classe I respectivement.

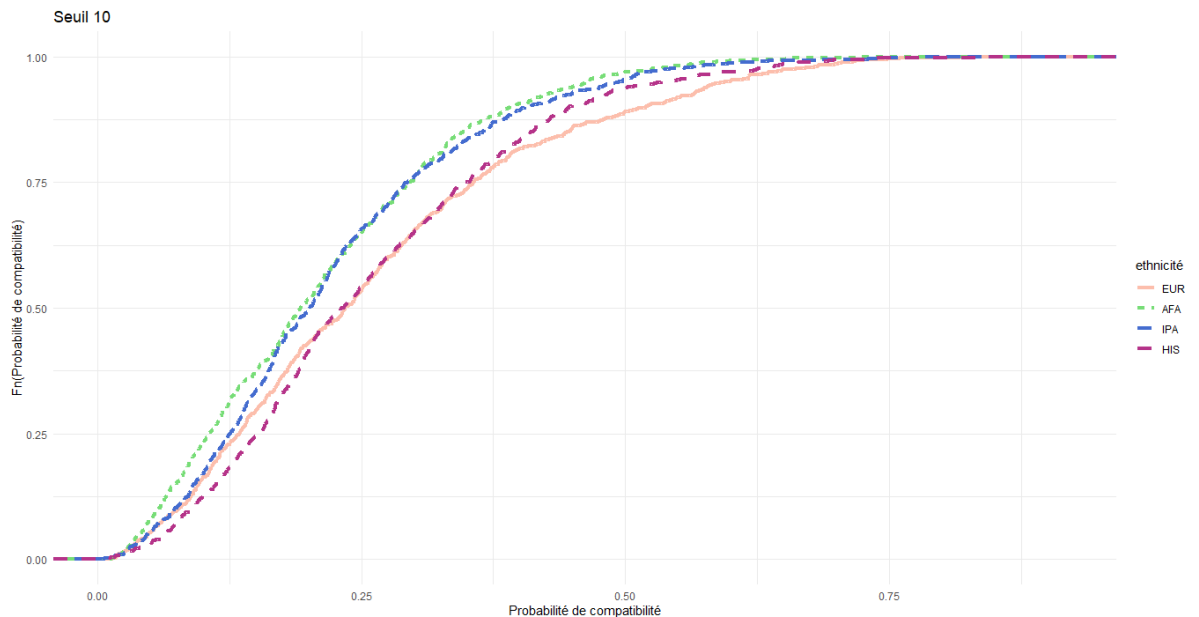


FIGURE 4.7 – Les courbes de la fonction de répartition des quatre ethnicités, en considérant les *abv* de classe I au seuil 10.

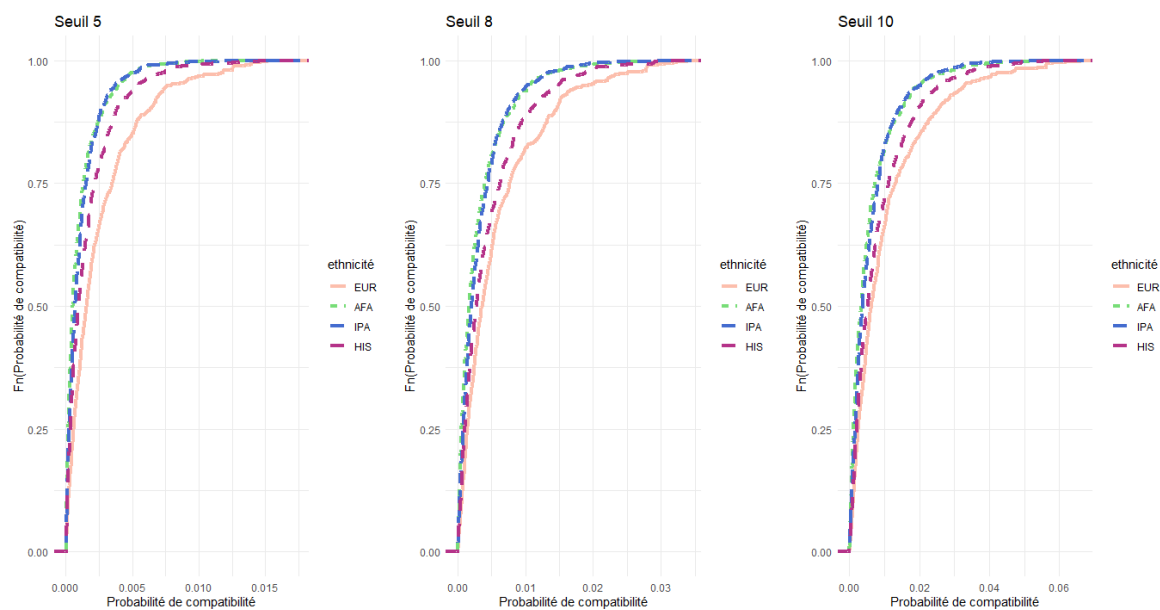


FIGURE 4.8 – Les courbes de la fonction de répartition des quatre ethnicités, en considérant l'épitype au complet aux trois seuils.

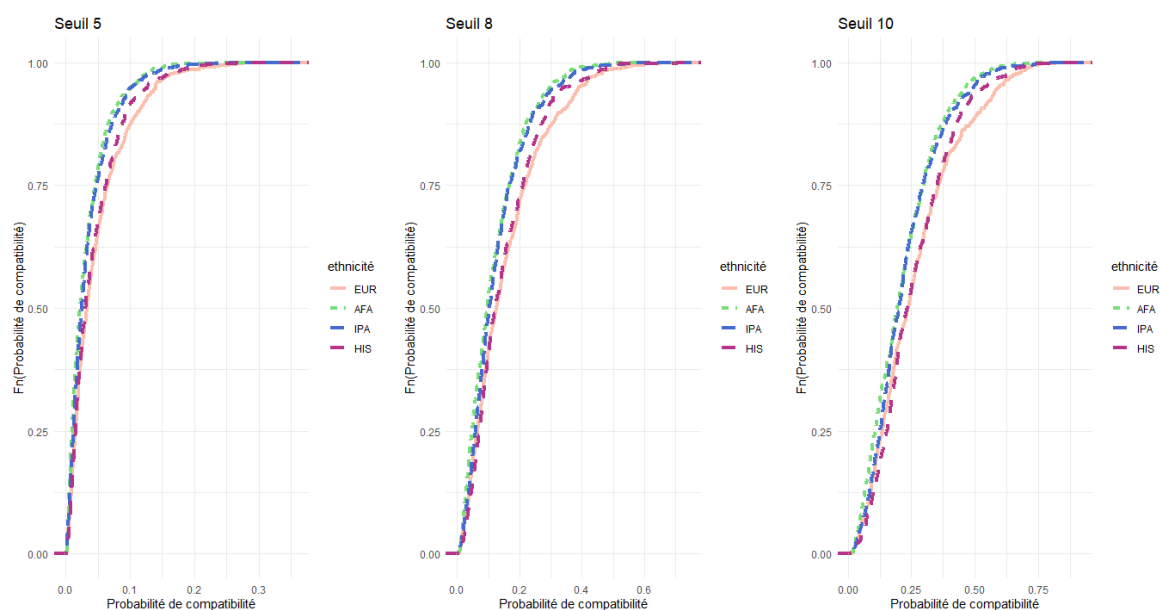


FIGURE 4.9 – Les courbes de la fonction de répartition des quatre ethnicités, en considérant le groupe *abv* de classe I aux trois seuils.

Pour voir si les échantillons des quatre ethnicités proviennent de la même distribution, on utilise le test d'*Anderson-Darling* (AD) pour k -échantillons. L'hypothèse nulle, pour 4-échantillons, s'écrit sous la forme :

H_0 : Les quatre échantillons proviennent de la même distribution,

H_A : Les échantillons ne proviennent pas de la même distribution.

Avant d'introduire la statistique du test on pose quelques notations. On définit par x_{ij} l'observation j dans l'échantillon i et par $N = n_1 + n_2 + \dots + n_k$ le nombre totale de toutes les observations. Les valeurs distinctes dans l'ensemble de données combinées, classées du plus petit au plus grand, sont notées z_1, z_2, \dots, z_L , où L sera inférieur à N s'il existe des observations *ex aequo*. La statistique simplifiée du test d'*Anderson-Darling* pour k -échantillons est

$$ADK = \frac{N-1}{N^2(k-1)} \sum_{i=1}^k \left[\frac{1}{n_i} \sum_{j=1}^L h_{ij} \frac{(NF_{ij} - n_i H_j)^2}{H_j(N - H_j) - N h_j / 4} \right],$$

avec

- h_j : le nombre de valeurs dans les échantillons combinés égal à z_j ;
- H_j : le nombre de valeurs dans les échantillons combinés inférieur à z_j plus la moitié du nombre de valeurs dans les échantillons combinés égal à z_j ;
- F_{ij} : le nombre de valeurs dans le $i^{\text{ème}}$ groupe (ethnicité) qui sont inférieures à z_j plus la moitié du nombre de valeurs dans ce groupe qui sont égales à z_j .

Sous l'hypothèse d'absence de différence dans les populations, la moyenne et la variance de la statistique ADK sont d'environ 1. On rejette H_0 avec un seuil de signification de $\alpha = 5\%$ si $ADK > ADC_\alpha$, et la valeur critique est défini par

$$ADC_\alpha = 1 + \sigma_n \left[b_0 + \frac{b_1}{\sqrt{k-1}} + \frac{b_2}{k-1} \right],$$

où les quantités b_0 , b_1 et b_2 représentent des constantes qui dépendent du choix de seuil α , et σ_n représente la variance de la statistique ADK .

Le test AD rejette H_0 et donc on peut conclure que les échantillons ne proviennent pas de la même distribution. On procède alors aux comparaisons multiples en utilisant le cas spécial du test AD pour 2-échantillons avec un seuil de signification α^* qui correspond à la correction de *Bonferroni*⁴. Le test pour chaque paire d'ethnicité s'écrit sous la forme :

H_0 : Les deux échantillons ont la même distribution,

H_A : Les deux échantillons n'ont pas la même distribution.

Le test rejette H_0 si $ADK > ADC_{\alpha^*}$. Pour toutes les paires, l'hypothèse H_0 est rejetée pour le cas de l'épitype au complet et celui groupe *abv* de classe I pour les trois seuils. La probabilité de compatibilité est significativement différente selon l'ethnicité. Rappelons que le test AD est très puissant et sensible aux tailles d'échantillons. Il est capable de détecter une différence assez petite (Engmann et Cousineau, 2011). Cette conclusion est encore une fois statistiquement significative mais non cliniquement validée.

4. $\alpha^* = \frac{2\alpha}{k(k-1)}$ avec $k = 4$.

CONCLUSION

Dans ce mémoire, nous avons pour objectif d'explorer et d'implémenter un nouvel algorithme statistique cPIE, calculant la compatibilité donneur-receveur au niveau des épitopes en considérant différents seuils d'incompatibilité, en transplantation rénale. Ce dernier, basé sur la compatibilité d'épitopes, nous a permis de déterminer la probabilité de trouver un donneur compatible au niveau de l'épitype et en considérant le groupe sanguin, pour des candidats de différentes ethnicités auto-identifiées, au niveau de la population, dans le but d'éviter le rejet précoce du greffon.

Nous avons estimé la fréquence épitypique en absence et en présence des fréquences des génotypes. En tenant compte de la compatibilité du groupe sanguin du patient et indépendamment de l'épitype et des seuils d'incompatibilité considérés, nous avons remarqué que la probabilité de compatibilité médiane, chez les patients d'origine européenne et hispanique était plus élevée par rapport aux autres patients d'autres ethnicités. Ceci nous permet d'avoir une idée sur les candidats qui seraient désavantagés en allocation du rein, et donc d'explorer de nouvelles stratégies pour les accommoder afin d'établir une allocation plus égalitaire. Nous n'avons pas pu trouver une référence dans la littérature qui suggère un seuil d'incompatibilité optimale permettant d'obtenir une probabilité de compatibilité la plus optimale. Dans les données simulées, lorsqu'on considère les épitopes *abv* de classe I, on remarque qu'un seuil de 10 pourrait donner un meilleur compromis entre la probabilité de compatibilité et la restriction quant au nombre d'incompatibilité des épitopes alloué entre donneur-receveur. Si une telle situation se reproduit en analysant les données réelles, le choix de ce seuil devrait être vérifié

et validé, par des modèles de prédiction du risque du rejet du greffon, en faisant varier le seuil. En effet, un seuil de 10 représente une stratégie pour éviter les inégalités (en anglais *inequity*), mais engendre un risque immunitaire associé qui est inévitablement plus important.

Nous avons remarqué, aussi, un grand nombre de patients éloignés de la probabilité de compatibilité médiane, indépendamment de l'ethnicité, qui possèdent des caractéristiques particulières. Ces patients partagent plus d'épitopes avec les donneurs et ils ont donc beaucoup plus de chance de trouver un donneur compatible. Si ces patients particuliers sont sur la liste d'attente, le cPIE peut nous aider à les identifier et on pourrait leur attribuer plus de poids dans la queue de la distribution de la compatibilité.

Quoiqu'il existe environ de 30000 gènes CMH, le cPIE proposé nous permet de réduire l'hétérogénéité entre receveur-donneur et le temps de calcul des probabilités de compatibilités. Il permet aussi de calculer la compatibilité de plusieurs stratégies d'appariement. Cette possibilité de choix de stratégie intégrée dans l'algorithme, ajustée par le groupe sanguin, va permettre d'identifier de nouvelles méthodes pour l'expansion du bassin du donneur. Par ailleurs, l'application du cPIE sur différentes ethnicités auto-identifiées va permettre de distinguer les patients qui peuvent être défavorisés en allocation.

ANNEXE A

INTRODUCTION À L'IMMUNOLOGIE

A.1 Relation entre les allèles HLA et les épitopes.

TABLEAU A.1 – Épitopes associés au génotype (1).

C1.abv.21H	C1.abv.41T	C1.abv.44KM	C1.abv.44RMA	C1.abv.44RT
0	1	1	1	0
C1.abv.45KE	C1.abv.56R	C1.abv.62EE	C1.abv.62GE	C1.abv.62GK
0	0	0	0	0
C1.abv.62GRN	C1.abv.62LQ	C1.abv.62QE	C1.abv.62RR	C1.abv.65GK
0	0	1	0	0
C1.abv.65QIA	C1.abv.65QKR	C1.abv.65RNA	C1.abv.69AA	C1.abv.69TNT
0	1	0	0	1
C1.abv.70IAQ	C1.abv.71ATD	C1.abv.71SA	C1.abv.71TTS	C1.abv.73AN
0	0	0	0	1
C1.abv.73TVS	C1.abv.76ANT	C1.abv.76EG	C1.abv.76ESI	C1.abv.76ESN
0	0	0	0	0
C1.abv.76VRN	C1.abv.79GT	C1.abv.80I	C1.abv.80K	C1.abv.80N
1	0	0	1	1
C1.abv.80TLR	C1.abv.82LR	C1.abv.90D	C1.abv.107W	C1.abv.127K
1	1	1	0	0
C1.abv.131S	C1.abv.138K	C1.abv.138MI	C1.abv.143S	C1.abv.144K
1	0	0	1	0
C1.abv.144KR	C1.abv.144QL	C1.abv.144TKH	C1.abv.145KHA	C1.abv.145RT
0	1	0	0	0
C1.abv.149TAH	C1.abv.150AAH	C1.abv.151AHA	C1.abv.156DA	C1.abv.158T
0	0	0	0	0
C1.abv.161D	C1.abv.163EW	C1.abv.163LS/G	C1.abv.163LW	C1.abv.163R
0	1	0	0	0
C1.abv.163RG	C1.abv.163RW	C1.abv.166DG	C1.abv.173K	C1.abv.177KT
0	0	0	0	0

TABLEAU A.2 – Épitopes associés au génotype suite (2).

C1.abv.180E	C1.abv.193PL	C1.abv.193PV	C1.abv.219W	C1.abv.248M
0	1	1	0	0
C1.abv.253Q	C1.abv.267QE	C1.oth.1C	C1.oth.9D	C1.oth.9F
1	1	1	1	0
C1.oth.9H	C1.oth.9S	C1.oth.9T	C1.oth.9Y	C1.oth.11AV
0	0	0	1	1
C1.oth.12M	C1.oth.14W	C1.oth.16S	C1.oth.17S	C1.oth.24S
1	0	0	0	1
C1.oth.24T	C1.oth.30G	C1.oth.32L	C1.oth.35Q	C1.oth.43R
1	0	0	1	0
C1.oth.44RM	C1.oth.44RME	C1.oth.45EE	C1.oth.56E	C1.oth.59H
1	0	0	0	0
C1.oth.62RN	C1.oth.63NI	C1.oth.65RA	C1.oth.65RK	C1.oth.66I
0	0	0	0	1
C1.oth.66IC	C1.oth.66IF	C1.oth.66IS	C1.oth.66IY	C1.oth.66K
0	0	1	0	1
C1.oth.66KA	C1.oth.66KH	C1.oth.66N	C1.oth.66NH	C1.oth.66NM
0	0	1	0	0
C1.oth.66NV	C1.oth.69RA	C1.oth.69RT	C1.oth.71HS	C1.oth.71KA
0	1	0	0	0
C1.abv.21H	C1.abv.41T	C1.abv.44KM	C1.abv.44RMA	C1.abv.44RT
0	1	1	1	0
C1.abv.45KE	C1.abv.56R	C1.abv.62EE	C1.abv.62GE	C1.abv.62GK
0	0	0	0	0
C1.abv.62GRN	C1.abv.62LQ	C1.abv.62QE	C1.abv.62RR	C1.abv.65GK
0	0	1	0	0
C1.abv.65QIA	C1.abv.65QKR	C1.abv.65RNA	C1.abv.69AA	C1.abv.69TNT
0	1	0	0	1
C1.abv.70IAQ	C1.abv.71ATD	C1.abv.71SA	C1.abv.71TTS	C1.abv.73AN
0	0	0	0	1

TABLEAU A.3 – Épitopes associés au génotype suite (3).

C1.abv.73TVS	C1.abv.76ANT	C1.abv.76EG	C1.abv.76ESI	C1.abv.76ESN
0	0	0	0	0
C1.abv.76VRN	C1.abv.79GT	C1.abv.80I	C1.abv.80K	C1.abv.80N
1	0	0	1	1
C1.abv.80TLR	C1.abv.82LR	C1.abv.90D	C1.abv.107W	C1.abv.127K
1	1	1	0	0
C1.abv.131S	C1.abv.138K	C1.abv.138MI	C1.abv.143S	C1.abv.144K
1	0	0	1	0
C1.abv.144KR	C1.abv.144QL	C1.abv.144TKH	C1.abv.145KHA	C1.abv.145RT
0	1	0	0	0
C1.abv.149TAH	C1.abv.150AAH	C1.abv.151AHA	C1.abv.156DA	C1.abv.158T
0	0	0	0	0
C1.abv.161D	C1.abv.163EW	C1.abv.163LS/G	C1.abv.163LW	C1.abv.163R
0	1	0	0	0
C1.abv.163RG	C1.abv.163RW	C1.abv.166DG	C1.abv.173K	C1.abv.177KT
0	0	0	0	0
C1.abv.180E	C1.abv.193PL	C1.abv.193PV	C1.abv.219W	C1.abv.248M
0	1	1	0	0
C1.abv.253Q	C1.abv.267QE	C1.oth.1C	C1.oth.9D	C1.oth.9F
1	1	1	1	0
C1.oth.9H	C1.oth.9S	C1.oth.9T	C1.oth.9Y	C1.oth.11AV
0	0	0	1	1
C1.oth.12M	C1.oth.14W	C1.oth.16S	C1.oth.17S	C1.oth.24S
1	0	0	0	1
C1.oth.24T	C1.oth.30G	C1.oth.32L	C1.oth.35Q	C1.oth.43R
1	0	0	1	0
C1.oth.44RM	C1.oth.44RME	C1.oth.45EE	C1.oth.56E	C1.oth.59H
1	0	0	0	0

TABLEAU A.4 – Épitopes associés au génotype suite (4).

C1.oth.62RN	C1.oth.63NI	C1.oth.65RA	C1.oth.65RK	C1.oth.66I
0	0	0	0	1
C1.oth.66IC	C1.oth.66IF	C1.oth.66IS	C1.oth.66IY	C1.oth.66K
0	0	1	0	1
C1.oth.66KA	C1.oth.66KH	C1.oth.66N	C1.oth.66NH	C1.oth.66NM
0	0	1	0	0
C1.oth.66NV	C1.oth.69RA	C1.oth.69RT	C1.oth.71HS	C1.oth.71KA
0	1	0	0	0
C1.oth.71QS	C1.oth.71TD	C1.oth.71TN	C1.oth.73AS	C1.oth.73ID
0	0	1	1	0
C1.oth.74Y	C1.oth.76ED	C1.oth.76EN	C1.oth.76ES	C1.oth.76ET
1	0	1	0	1
C1.oth.76VDT	C1.oth.76VS	C1.oth.77D	C1.oth.77N	C1.oth.77NGT
0	1	0	1	0
C1.oth.77S	C1.oth.77SRN	C1.oth.80T	C1.oth.80TA	C1.oth.80TL
1	1	1	1	0
C1.oth.81ALR	C1.oth.91R	C1.oth.94I	C1.oth.95F	C1.oth.95I
1	0	0	0	1
C1.oth.95L	C1.oth.95V	C1.oth.95W	C1.oth.97I	C1.oth.97M
1	0	1	0	0
C1.oth.97N	C1.oth.97R	C1.oth.97S	C1.oth.97T	C1.oth.97V
0	1	0	1	0
C1.oth.97W	C1.oth.99F	C1.oth.99S	C1.oth.99Y	C1.oth.102H
0	0	0	1	0
C1.oth.103L	C1.oth.103M	C1.oth.105S	C1.oth.109F	C1.oth.113H
1	0	0	0	1
C1.oth.113HD	C1.oth.113HN	C1.oth.113YD	C1.oth.113YN	C1.oth.114H
0	1	1	1	0

TABLEAU A.5 – Épitopes associés au génotype suite (5).

C1.oth.114Q	C1.oth.114R	C1.oth.116D	C1.oth.116F	C1.oth.116L
0	0	0	1	1
C1.oth.116S	C1.oth.116Y	C1.oth.145HT	C1.oth.147L	C1.oth.149AH
1	0	0	1	0
C1.oth.150AH	C1.oth.151AHE	C1.oth.151AHV	C1.oth.151ARV	C1.oth.151H
0	0	0	1	0
C1.oth.152A	C1.oth.152E	C1.oth.152HA	C1.oth.152RA	C1.oth.152RE
1	1	0	1	1
C1.oth.152RR	C1.oth.152T	C1.oth.152V	C1.oth.152W	C1.oth.156L
0	0	1	0	1
C1.oth.156QA	C1.oth.156R	C1.oth.156RA	C1.oth.156WA	C1.oth.162DLS
0	0	0	0	0
C1.oth.162GLS	C1.oth.163E	C1.oth.163L	C1.oth.163LE	C1.oth.163LG
0	1	0	0	0
C1.oth.163T	C1.oth.166ES	C1.oth.170RH	C1.oth.177DK	C1.oth.177DT
1	0	0	0	0
C1.oth.184A	C1.oth.184H	C1.oth.184R	C1.oth.186R	C1.oth.193AV
0	0	1	0	0
C1.oth.193LV	C1.oth.193PI	C1.oth.194V	C1.oth.199V	C1.oth.207S
0	1	1	0	0
C1.oth.211T	C1.oth.245AS	C1.oth.245TA	C1.oth.245V	C1.oth.270C
0	0	0	0	1
C1.oth.275EL	C1.oth.275G	C1.oth.275K	C2.abv.4Q	C2.abv.4R
0	0	1	0	1
C2.abv.11STS	C2.abv.13FE	C2.abv.13FEY	C2.abv.16Y	C2.abv.25Q3
1	0	0	0	0
C2.abv.25R	C2.abv.28H	C2.abv.31FYY	C2.abv.31I	C2.abv.37L
1	0	1	0	0

TABLEAU A.6 – Épitopes associés au génotype suite (6).

C2.abv.37YV	C2.abv.37S	C2.abv.40YD2	C2.abv.47F	C2.abv.48Q6
1	0	0	1	0
C2.abv.51R2	C2.abv.57DE	C2.abv.57DEDP	C2.abv.57V	C2.abv.57S
0	1	1	0	0
C2.abv.70DA	C2.abv.70D	C2.abv.70QT	C2.abv.70R	C2.abv.70RE
1	1	0	0	0
C2.abv.71A	C2.abv.71K	C2.abv.73A	C2.abv.73GQ	C2.abv.74R
0	0	1	0	0
C2.abv.74L	C2.abv.77N	C2.abv.77T	C2.abv.78V2	C2.abv.85A
0	1	1	0	0
C2.abv.96EV	C2.abv.96HK	C2.abv.96Y2	C2.abv.96H	C2.abv.96Q
0	1	0	1	0
C2.abv.96QK2	C2.abv.98Q	C2.abv.98E	C2.abv.104A	C2.abv.108T
0	0	0	0	0
C2.abv.112Y	C2.abv.120N	C2.abv.140A	C2.abv.140TV	C2.abv.142M3
0	0	0	1	0
C2.abv.149H	C2.abv.181M	C2.oth.6C2	C2.oth.13FEL	C2.oth.13SE
1	0	0	0	1
C2.oth.13SEF	C2.oth.13SEY	C2.oth.16H	C2.oth.25RF	C2.oth.26F
1	0	1	1	1
C2.DR.oth.26L	C2.oth.26Y	C2.DR.oth.28D	C2.oth.28DHF	C2.oth.28DY
0	0	1	0	1
C2.DR.oth.28E	C2.oth.28EH	C2.oth.28EY	C2.oth.30G	C2.DR.oth.30H
0	0	0	0	0
C2.DR.oth.30Y	C2.oth.31F	C2.oth.31FH	C2.oth.31FY	C2.oth.31FYN
1	1	0	1	1
C2.oth.32H	C2.oth.32Y	C2.oth.32YN	C2.oth.33N	C2.oth.37D2
0	1	1	1	0

TABLEAU A.7 – Épitopes associés au génotype suite (7).

C2.oth.37F	C2.oth.37FL2	C2.oth.37FV	C2.oth.37N	C2.oth.37Y
0	0	0	0	1
C2.oth.38A	C2.oth.38L	C2.oth.38V	C2.oth.40F	C2.oth.47Y
0	0	1	1	0
C2.oth.47YR	C2.oth.57A	C2.oth.57D	C2.oth.57DA	C2.oth.58A
0	0	1	0	0
C2.oth.58AY	C2.oth.58EEDP	C2.oth.60Y	C2.oth.67F	C2.oth.67I
0	1	1	1	0
C2.oth.67L	C2.oth.70DK	C2.oth.70DR	C2.oth.70DRA	C2.oth.70Q
0	0	1	1	0
C2.oth.70QA	C2.oth.70QE	C2.oth.70QK	C2.oth.70QKA	C2.oth.70QQ
0	0	0	0	0
C2.oth.70QRA	C2.oth.71E	C2.oth.71R	C2.oth.73G	C2.oth.74A
0	0	1	0	1
C2.oth.74E	C2.oth.77TY	C2.oth.78Y	C2.oth.85V	C2.oth.85VV
0	1	1	1	0
C2.oth.86G	C2.oth.86V	C2.oth.96EN3	C2.oth.96ES2	C2.oth.98K
1	0	0	0	1
C2.oth.98KN	C2.oth.98KS	C2.oth.104AK	C2.oth.104S	C2.oth.108P
0	1	0	1	1
C2.oth.112H	C2.oth.120S	C2.oth.133RS	C2.oth.140AV	C2.oth.149Q
1	1	0	0	0
C2.oth.164FR	C2.oth.180V	C2.oth.180VMP	C2.oth.180VTP	C2.oth.181T
0	1	0	1	1
C2.oth.189R	C2.oth.189S	C2.abv.40GR3	C2.abv.45EV	C2.abv.45GE3
1	0	0	1	0
C2.abv.45GV	C2.abv.46VY3	C2.abv.47KHL	C2.abv.52PL3	C2.abv.52PQ2
0	1	0	1	0

TABLEAU A.8 – Épitopes associés au génotype suite (8).

C2.abv.52PR	C2.abv.55PP	C2.abv.55RL3	C2.abv.74SR3	C2.abv.74SV2
0	1	0	0	0
C2.abv.75S3	C2.abv.77R	C2.abv.84QL3	C2.abv.125SQ	C2.abv.140A2
0	0	1	0	0
C2.oth.2D2	C2.oth.2G2	C2.oth.3P4	C2.oth.3S	C2.oth.9F
0	0	0	1	0
C2.DQ.oth.9Y	C2.oth.13GM	C2.oth.23L	C2.oth.23R	C2.oth.25FT
1	0	0	1	0
C2.oth.25YT	C2.DQ.oth.26L	C2.DQ.oth.30H	C2.DQ.oth.30Y	C2.oth.37YA
0	0	0	1	1
C2.oth.37YV	C2.oth.40ERV	C2.oth.41KA2	C2.oth.41RA2	C2.oth.45G
0	0	0	0	0
C2.oth.47QL5	C2.oth.52SK5	C2.oth.55PPA	C2.oth.55PPD	C2.oth.55RPD
0	0	0	1	0
C2.oth.56P	C2.oth.56PA	C2.oth.56PD	C2.oth.56PS	C2.oth.56PV
1	0	1	0	0
C2.oth.61FT4	C2.oth.66D	C2.oth.66DR	C2.oth.66ER	C2.oth.66EV
0	0	0	1	1
C2.oth.66IL	C2.oth.66IT	C2.oth.67VG	C2.oth.67VT	C2.oth.70GT
0	0	0	1	0
C2.oth.70RT	C2.oth.74EL	C2.oth.75V	C2.oth.75I	C2.oth.75IL
1	1	0	0	0
C2.oth.76L	C2.oth.85VA	C2.oth.85VY	C2.oth.86G2	C2.oth.87F
0	0	0	0	0
C2.oth.116V	C2.oth.125G	C2.oth.129H	C2.oth.130R	C2.oth.135D
1	0	0	1	1
C2.oth.135G	C2.oth.160A	C2.oth.160AD	C2.oth.160AE	C2.oth.160D
0	0	0	0	0

TABLEAU A.9 – Épitopes associés au génotype suite (9).

C2.oth.160S	C2.oth.167H2	C2.oth.167R	C2.oth.175E	C2.oth.185I
0	1	0	0	0
C2.oth.185T				
1				

ANNEXE B

L'ALGORITHME DE L'APPARIEMENT DES ÉPITOPES (CPIE)

B.1 Les étapes de l'algorithme cPIE

```

1 Pour chaque receveur du bassin des receveurs ;
2 Calculer la fréquence épitypique  $f_{E_D}$ , en considérant seulement les épitypes présents dans le bassin des donneurs ;
3 Calculer l'incompatibilité des épitopes, dans le bassin des donneurs, basé sur l'épitope,  $(E_D^i - E_R)$  ;
4 Calculer la compatibilité, selon un seuil fixé d'incompatibilité défini,  $\mathbb{1} \left\{ \sum_{j=1}^l ((E_D^i - E_R)_j)_+ \leq t \right\}$ , ;
5 Si : on ne considère pas le groupe sanguin du candidat ;
6 Calculer la probabilité d'appariement d'un candidat, au seuil fixé au niveau de l'épitope seulement,  $P_{E_R}^{(t)}$  ;
7 Si : on considère le groupe sanguin du candidat ;
8   Si : le groupe sanguin est  $O$  ;
9   Calculer la probabilité de compatibilité  $P_{E_R,O}^{(t)} = w_O * P_{E_R}^{(t)}$  ;
10  Sinon : le groupe sanguin est  $A$  ;
11  Calculer la probabilité de compatibilité  $P_{E_R,A}^{(t)} = (w_A + w_O) * P_{E_R}^{(t)}$  ;
12  Sinon : le groupe sanguin est  $B$  ;
13  Calculer la probabilité de compatibilité  $P_{E_R,B}^{(t)} = (w_B + w_O) * P_{E_R}^{(t)}$  ;
14  Sinon : le groupe sanguin est  $AB$  ;
15  Calculer la probabilité de compatibilité  $P_{E_R,AB}^{(t)} = (w_A + w_B + w_{AB} + w_O) * P_{E_R}^{(t)}$  ;
16 Sinon : on considère le groupe sanguin du candidat avec compatibilité sanguine unique ;
17  Si : le groupe sanguin est  $O$  ;
18  Calculer la probabilité de compatibilité  $P_{E_R,U_{AB}}^{(t)} = w_O * P_{E_R}^{(t)}$  ;
19  Sinon : le groupe sanguin est  $A$  ;
20  Calculer la probabilité de compatibilité  $P_{E_R,U_A}^{(t)} = w_A * P_{E_R}^{(t)}$  ;
21  Sinon : le groupe sanguin est  $B$  ;
22  Calculer la probabilité de compatibilité  $P_{E_R,U_B}^{(t)} = w_B * P_{E_R}^{(t)}$  ;
23  Sinon : le groupe sanguin est  $AB$  ;
24  Calculer la probabilité de compatibilité  $P_{E_R,U_{AB}}^{(t)} = w_{AB} * P_{E_R}^{(t)}$  ;
25 Retourner une table contenant le nombre d'incompatibilités d'épitopes, selon le seuil fixé et les probabilités de
    compatibilités associées pour tout le bassin des receveurs ;

```

FIGURE B.1 – Algorithme cPIE en se basant sur les épitypes présents dans le bassin des donneurs.

RÉFÉRENCES

- American Red Cross (ARC). ARC. *blood services. blood types*. Récupéré de <https://www.redcrossblood.org/donate-blood/blood-types.html>
- Baxter-Lowe, L. A. (2021). The Changing Landscape of HLA Typing : Understanding How and When HLA Typing Data can be Used With Confidence from Bench to Bedside. *American Society for Histocompatibility and Immunogenetics*, 82(7), 466–477.
<http://dx.doi.org/https://doi.org/10.1016/j.humimm.2021.04.011>
- Berche, P. (2012). Louis Pasteur, from Crystals of Life to Vaccination. *Clinical Microbiology and Infection*, 18(Suppl.5), 1–6. <http://dx.doi.org/https://doi.org/10.1111/j.1469-0691.2012.03945.x>
- Cecka, J. M. (2010). Calculated PRA (CPRA) : The New Measure of Sensitization for Transplant Candidates. *American Journal of Transplantation*, 10, 26–29.
<http://dx.doi.org/10.1111/j.1600-6143.2009.02927.x>
- Chen, Z. (2014). *Statistical Methods for QTL Mapping*. Chapman & Hall/CRC Mathematical and Computational Biology Series.
- Composites Materials Handbook. (2012). *Composites Materials Handbook, CMH-17, Volume 1. Polymer Matrix Composites Guideline for Characterization of Structural Materials* (no. cmh-17-1 revision g éd.). SAE International.
- Dempster, A. P. et al. (1977). Maximum Likelihood from Incomplete Data via The em Algorithm. *Journal of the Royal Statistical Society*, 39(1), 1–38.
- Duquesnoy, R. J. (2002). HLAMatchmaker : A Molecularly Based Algorithm for Histocompatibility Determination. I. Description of The Algorithm. *American Society for Histocompatibility and Immunogenetics*, 63, 339–352.
- Duquesnoy, R. J. (2007). Update on HLAMatchmaker : A Molecularly Based Algorithm for Histocompatibility Determination at The Epitope Level. *ASHI Quarterly*, 74–79.

Duquesnoy, R. J. et Marrari, M. (2002). HLAMatchmaker : A Molecularly Based Algorithm For Histocompatibility Determination. II. Verification of The Algorithm and Determination of The Relative Immunogenicity of Amino Acid Triplet-Defined Epitopes. *Human Immunology*, 63, 353–363.

Edwards, A. W. F. (2008). G. H. Hardy (1908) and Hardy–Weinberg Equilibrium. *Genetics*, 179(3), 1143–1150.
<http://dx.doi.org/https://doi.org/10.1534/genetics.104.92940>

Elkhalidy, Y. et Clarke, G. (2020). Sous-groupes du Système ABO. Récupéré de <https://profedu.blood.ca/fr/transfusion/bonnes-pratiques/pratiques-suivre-en-matiere-de-serologie/sous-groupes-du-systeme-abo>

Engmann, S. et Cousineau, D. (2011). Comparing Distributions : The Two-Sample Anderson-Darling Test as an Alternative to The Kolmogorov-Smirnoff Test. *Journal of applied quantitative methods*, 6, 1–17.

Faucher, J., Lachaîne, R. et al. (2011). *Biologie* (4th éd.). PEARSON, CAMPBELL BIOLOGIE.

Feingold, J. (1991). Le Déséquilibre de Liaison. *médecine/sciences*, 7(2), 161–168.

Gross, M. B. (2014). Blood group, ABO System. Récupéré de <https://omim.org/entry/616093>

Hill, W. et al. (1968). The Effects of Inbreeding at Loci With Heterozygote Advantage. *Genetics*, 60(3), 615–628.

Janeway, C. A. et al. (2001). *Immunobiology : The Immune System in Health and Disease* (5th éd.). Garland Science.

Kaufmann, S. H. E. (2017). Remembering Emil Von Behring : from Tetanus Treatment to Antibody Cooperation with Phagocytes. *American Society For Microbiology*, 8(1).
<http://dx.doi.org/https://doi.org/10.1128/mBio.00117-17>

Lemieux, W., Sapir-Pichhadze, R. et al. (2021). Matchmaker, Matchmaker Make me a Match : Opportunities and Challenges in Optimizing Compatibility of HLA Eplets in Transplantation. *International Journal of Immunogenetics*, 48(2), 135–144.
<http://dx.doi.org/https://doi.org/10.1111/iji.12525>

Lewontin, R. C. (1964). The Interaction of Selection and Linkage. i. general considerations; heterotic models. *Genetics*, 49(1), 49–67.

Male, D. (2019). *Immunologie, Aide-mémoire Illustré* (4th éd.). DE BOECK SUP.

Mohammadhasanzadeh, H. et al. (2021). On Path to Informing Hierarchy of Eplet Mismatches as Determinants of Kidney Transplant Loss. *Kidney International Reports*, 6, 1567–1579.
<http://dx.doi.org/https://doi.org/10.1016/j.ekir.2021.03.877>

Mordant, P. (2007). La Transplantation d'organe Petite Histoire d'une Grande Aventure. Récupéré de
https://sofia.medicalistes.fr/spip/IMG/pdf/La_transplantation_d_organe_petite_histoire_d_une_grande_aventure.pdf

Muller, Y. (2015). Le Système Immunitaire. Récupéré de
<https://moodle.umontpellier.fr/enrol/index.php?id=233#section-3>

National Marrow Donor Program (2007). National Marrow Donor Program, high-resolution hla alleles and haplotypes in the us population. Récupéré de
<https://bioinformatics.bethematchclinical.org/hla-resources/haplotype-frequencies/high-resolution-hla-alleles-and-haplotypes-in-the-us-population/>

Riedel, S. (2005). Edward Jenner and The History of Smallpox and Vaccination. *Baylor University Medical Center Proceedings*, 18(1), 21–25.
<http://dx.doi.org/https://doi.org/10.1080/08998280.2005.11928028>

Robinson, J. et al. (2015). The IPD and IMGT/HLA Database : Allele Variant Databases. Récupéré de
<http://hla.alleles.org/misc/citing.html>

Transplant Québec (2013). Transplant Québec - Historique. Récupéré de
<https://www.transplantquebec.ca/mission>